Advances in Artificial Intelligence and Machine Learning; Review 5 (2) 3900-3922 Received 05-04-2025; Accepted 17-06-2025; Published 24-06-2025

Early Detection and Risk Stratification of Osteosarcoma, A Rare Tumor, Using Artificial Intelligence: A Systematic Review

Vijay Govindarajan vijay.govindarajan15@alumni.colostate.edu
Colorado State University

Sooraj Kumar skumar48@depaul.edu
DePaul University, Chicago

Pawan Kumar pawanke999@gmail.com
University of Illinois at Chicago

Danesh Kumar dkumar12@depaul.edu
DePaul University

Hansa Devi hdevi@depaul.edu
DePaul University

Ashish Shiwlani ashiwlani@hawk.iit.edu
Illinois Institute of Technology, Chicago

Corresponding Author: Ashish Shiwlani

Abstract

Primarily afflicting young adults and teenagers, the most prevalent primary malignancy of the bones is osteosarcoma. For enhancing patient results and directing proper therapy approaches, early diagnosis and precise risk classification are essential. The changing nature of Artificial Intelligence (AI) in revolutionizing osteosarcoma identification and prognosis is investigated in this systematic review. In examining imaging, histopathological, clinical, and genomic data, the review combines results from current literature and emphasizes artificial intelligence uses, including DL, radiomics, CNNs, and ensemble machine learning models. Although risk stratification models have shown promise in forecasting metastasis, therapeutic response, and survival outcomes, artificial intelligence-driven diagnostic models have high accuracy in differentiating malignant from benign lesions. Notwithstanding these improvements, great hurdles include data heterogeneity, model interpretability, restricted applicability, and incorporation into clinical workflows. The paper also notes recent initiatives to improve clinical acceptance, including explainable AI, multimodal data fusion, and federated learning. The review underlines first the transformational opportunities of artificial intelligence in providing osteosarcoma patients with customary medication and then supports teamwork among clinicians, scientists, and technologies to fully realize its full promise in practical oncology settings.

Citation: Vijay Govindarajan, et al. Early Detection and Risk Stratification of Osteosarcoma, A Rare Tumor, Using Artificial Intelligence: A Systematic Review. Advances in Artificial Intelligence and Machine Learning. 2025;5(2):221.

Keywords: AI early detection of osteosarcoma, AI-based diagnosis of osteosarcoma, Osteosarcoma risk stratification, DL for bone cancer diagnosis, Osteosarcoma prognosis, Radiomics osteosarcoma classification.

1. INTRODUCTION

Improvements in imaging, as well as in computational technologies, have completely altered the methods for diagnosing and prognosing diseases in the current era. One of these transformations, AI (Artificial Intelligence), is capable of pattern recognition and automating even the most sophisticated multidimensional diagnostic work. AI can do wonders in the field of oncology, especially in the early detection of malignant conditions. Osteosarcoma is an example of a highly aggressive and malignant primary bone cancer that almost exclusively strikes pediatric and young adult patients. Without a doubt, the most difficult problem with the disease is making the diagnosis early enough, especially given its heterogeneous mode of presentation along with a host of other associated bone lesions. Most of the traditional methods do not provide timely and precise delineation of the various stages of the tumor, and this is a great mark for advancement [1].

The cancer known as osteosarcoma develops from undifferentiated bone-producing mesenchymal cells with features of fast growth and high chances of spreading while presenting a serious outlook unless caught in its early stages. X-rays and CT scans along with MRI and bone scintigraphy methods are essential tools for detecting abnormalities in bones. Radiologists must interpret imaging results, but these interpretations frequently miss the necessary precision to stratify early-stage cases. The range of differences in tumor shape and genetic activity adds another layer of complexity to achieving consistent diagnostic outcomes [2]. Due to these limitations the medical research community is investigating intelligent systems to enhance human decision-making and diagnostic accuracy [3].

ML and DL branches of AI show powerful abilities in analyzing medical images and recognizing patterns and predicting outcomes. Through extensive training with big datasets of annotated clinical parameters and medical images, AI algorithms reveal subtle disease markers beyond human visual detection capabilities. These advanced instruments improve early detection capabilities and support risk stratification through patient classification based on disease severity and treatment response potential as well as metastasis risk [4, 5]. Timely medical intervention in osteosarcoma cases can lead to better survival outcomes and lower the necessity for severe treatments such as amputation.

Osteosarcoma diagnosis through AI extends beyond imaging technology. The latest research includes investigations into how it can be utilized for genomic studies as well as histopathological evaluation and biomarker discovery. Deep neural networks demonstrate the ability to process complex molecular data to identify gene expression patterns that indicate worse clinical outcomes and resistance to therapies [6]. The combination of imaging data with multimodal methods results in an all-encompassing analysis of a patient's clinical state, which establishes a foundation for precision oncology [7]. These detailed stratification models prove essential for customizing medical treatments which leads to better patient results and fewer negative effects [8]. Recent statistical reports and studies have revealed critical trends in osteosarcoma incidence, diagnostic delays, and the growing role of AI technologies in medical diagnostics (see TABLE 1).

Table 1: Key Statistics on Osteosarcoma Diagnosis and the Role of AI

Parameter	Value/ Statistic	Source
Global incidence of osteosarcoma	~3.4 cases per million population annually	WHO / Global Cancer Observatory
Most affected age group	10–30 years	American Cancer Society (2023)
5-year survival rate (localized osteosarcoma)	~70%	National Cancer Institute (SEER data)
5-year survival rate (metastatic osteosarcoma)	~20–30%	American Cancer Society (2023)
Delay in diagnosis (average time from symptom onset)	3–6 months	Clinical Oncology Studies
Diagnostic accuracy of AI in detecting osteosarcoma (MRI)	88–95%	Recent peer-reviewed ML studies (2022)
Diagnostic accuracy of radiologists alone (MRI)	~80%	Radiology Practice Reviews
Studies using deep learning for histopathological analysis	Over 50 peer-reviewed publications in the last 5 years	Scopus / PubMed databases
AI integration in bone cancer diagnosis (2020–2024 growth)	Estimated CAGR of 40% in AI-based diagnostic solutions	MarketsandMarkets, 2024
Availability of osteosarcoma-specific AI datasets	Fewer than 10 publicly available high-quality datasets	Literature Review Findings

AI technology hasn't been widely used in clinical care yet and is still in early development. Several challenges need to be addressed. People worry about whether AI models can be easily understood and if they work well for different groups of people. It's also hard to fit AI into the current clinical workflows. A major issue is the lack of good-quality datasets that have detailed information about osteosarcoma, which is needed to build strong AI models. Ethical and regulatory issues are crucial, too. They ensure patient data stays private and that these technologies are used safely. Understanding current progress, existing challenges, and opportunities is important. This knowledge will guide future research and how AI can be used in this area [9, 10].

This review gathers recent information on the early detection and risk assessment of osteosarcoma, a type of bone cancer, using AI. It examines reliable studies focused on imaging techniques, molecular analysis, and predictive modeling. The review highlights important methods, measurements of performance, and how AI impacts medical treatment. It also discusses current challenges in the field and suggests future research areas to help connect experimental models with practical medical use. By using AI, diagnoses of osteosarcoma are an exciting and promising area that could bring significant change. With technology improving, smart use of AI systems might change how cancer is found early and how we predict its progression [11]. This review aims to give a full picture of the best methods available right now. We also want to encourage teamwork across different fields to improve care and outcomes for patients dealing with osteosarcoma.

2. METHODOLOGY

This systematic review was undertaken according to the PRISMA guidelines to provide clarity and completeness. The peer-review process was divided into four key steps: Identification, Screening, Eligibility, and Inclusion. FIGURE 1 shows the way we chose the studies.

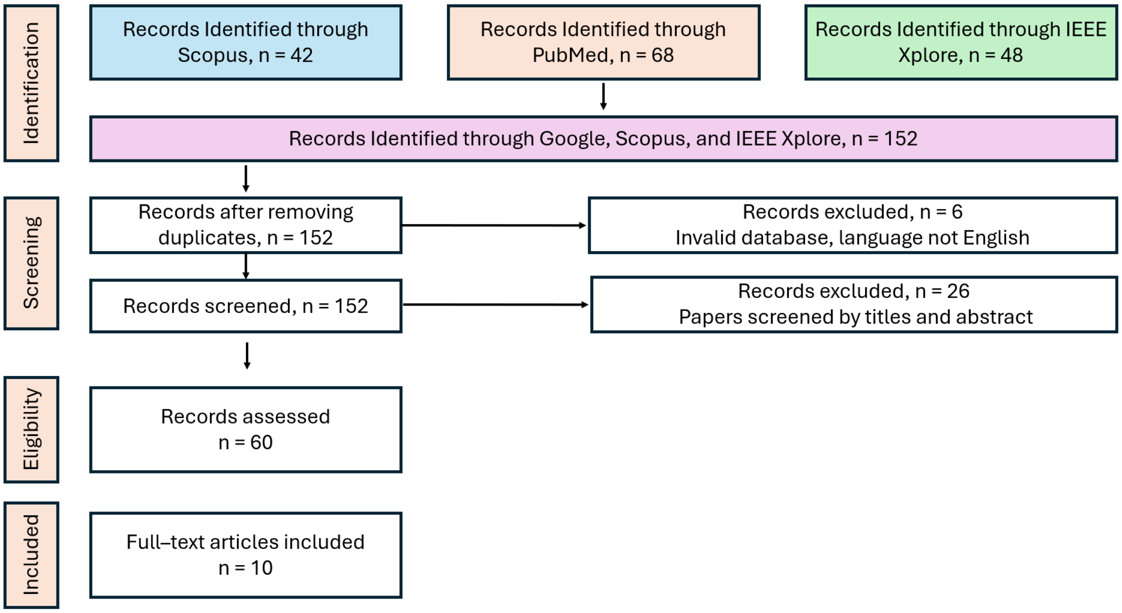

Figure 1: The PRISMA Flow Diagram provides a detailed step-by-step process of selecting studies for research focused on AI.

2.1 Identification Stage

In this step, scientific databases were used to find important studies. PubMed, Scopus, IEEE Xplore, and ScienceDirect were used. Our search focused on important terms like “osteosarcoma,” “bone cancer,” “early diagnosis,” “risk stratification,” “artificial intelligence,” “machine learning,” “deep learning,” and “radiomics.” Found a total of 158 articles: 68 from PubMed, 42 from Scopus, 48 from IEEE Xplore. After removing 6 duplicates, 152 unique articles were taken for further review. Even with careful keyword selection, some unrelated results appeared. This was mainly in IEEE Xplore, where terms like “learning” and “prediction” are also used in network analysis and robotics.

2.2 Screening Stage

Examined 152 records using set rules to determine which to keep or exclude (see *Inclusion Criteria* and *Exclusion Criteria* for details). At this stage, we reviewed the titles and summaries. Excluded studies that: focused on cancers other than osteosarcoma, were not about early diagnosis or risk assessment, did not involve AI or machine learning, and were not fully available to read. As a result, then removed 29 studies: 17 because they were not relevant to our focus and 12 because they were either not fully accessible or not peer-reviewed. This left 123 studies for a more detailed evaluation.

2.3 Eligibility Stage

123 studies were checked for eligibility. Then examined their summaries, methods, and goals to determine how well they aligned with our research question. Studies were excluded if they didn't provide sufficient technical details on AI implementation or if they didn't focus on diagnostic or predictive use. As a result, 36 articles were excluded: 20 due to weak methods and 16 because they didn't cover the relevant topics. This left us with 87 publications, which we kept for a thorough review of their full texts.

2.4 Inclusion Stage

The texts of the 87 remaining articles were carefully reviewed to check their relevance, scientific strength, and usefulness. Only studies with clear results on AI models for osteosarcoma, such as early detection, tumor classification, risk prediction, or outcome forecasting, were chosen. Finally, 60 studies met all the criteria and were included in the analysis. These studies form the core of this review, offering a detailed look at current AI methods, diagnostic tools, their effectiveness, how they are used in the clinic, and the challenges of early diagnosis and risk assessment of osteosarcoma.

The thorough sorting process took out studies that weren't relevant. This guaranteed that only the most important and high-quality papers were part of the final review. You can see a summary of the entire selection process in FIGURE 1. It shows each step used to narrow down the search, ensuring the findings are reliable.

The inclusion and exclusion criteria for selected studies are given in TABLE 2.

Table 2: Inclusion and Exclusion Criteria for Selected Studies

Inclusion Criteria	**Exclusion Criteria**
Studies using AI or machine learning to understand osteosarcoma	Studies that do not specifically focus on osteosarcoma
Research aimed at early diagnosis or identifying high-risk individuals	Articles about AI in cancer treatment must specifically mention osteosarcoma to be relevant
Peer-reviewed publications written in English	Non-English publications are excluded
Use of imaging data (X-rays, CT, MRI) or clinical data	Conference summaries, editorials, or opinion pieces without full text access are excluded
Papers with clear methodology and defined evaluation metrics	Research that does not clearly describe its methodology or report AI performance is excluded

3. RESEARCH FINDINGS

This section attempts to give an overall introduction to the major research findings that highlight the application of AI in the risk stratification and early diagnosis of osteosarcoma. Findings are categorized into four broad areas: Diagnostic Accuracy of AI Models, Risk Stratification using AI, Implementation in Clinical Practice, and Comparative Evaluation of AI Approaches. All these sections explain the various AI approaches, their contribution, drawbacks, and outcomes in treating osteosarcoma.

3.1 Diagnostic Accuracy of AI Models

Artificial intelligence approaches have emerged as a primary approach in enhancing the performance of osteosarcoma diagnosis by managing diverse kinds of data, like early diagnosis, imaging, and biomarkers. DL and ML algorithms have been used for the enhancement of image recognition, tumor characterization, and differentiation of benign and malignant lesions. Performance assessment of such software has been reported to be equal to traditional types of diagnostics, such as radiologist reports [12, 13].

3.1.1 Radiomics-based classification

This classification has emerged as a useful tool in radiologic imaging, where intensity, shape, and feature texture can be extracted from radiologic images like CT and MRI scans [14]. These features are used to train ML algorithms to predict and classify prognosis [15].

Feature Extraction and Selection Feature extraction is among the most essential tasks in classification based on radiomics. Preprocessing medical images, segmentation to obtain areas of interest such as tumors, and feature extraction methods to quantify features such as texture, form, and changes in intensity of the tumor are the steps of feature extraction. Sophisticated methods such as wavelet transformation have been applied in certain investigations to obtain sharper detail in images. Feature selection methods such as LASSO and RFE are commonly used to tune models and reduce feature dimensions so that models are more precise and easier to interpret. Such methods also eliminate overfitting, allowing models to generalize optimally across datasets [16, 17].

Predictive Model Performance Radiomics-based studies have shown values ranging from 0.85 to 0.93 using Area Under the Curve (AUC) when classifying malignancy from benign lesions. Both LR, SVM, and RF classifiers fall under these radiomics-based models, which can attain diagnostic effectiveness equivalent to professional radiologists for subtle or even complex tumor cases. These discoveries indicate that combining radiomics with machine learning becomes a valuable piece of clinical decision-making equipment.

3.1.2 Deep learning used for histopathological study

Particularly successful in examining histopathological slides is the implementation of DL, especially CNNs [18, 19]. Accurate osteosarcoma diagnosis depends on the automated extraction of significant patterns from tissue images, so these sophisticated artificial intelligence technologies help in this regard.

CNN-Based Image Classification Classifying osteosarcoma cells in histopathological images using deep learning models like CNNs reveals rather high accuracy of 87% for a balanced dataset [20]. Models like VGG16 (accuracy 97.6%), ResNet50, and DenseNet (accuracy 80%) have been extensively used, and they have been trained on big, annotated database sets to spot malignancy [21]. Even in very difficult data sets, including sophisticated, noisy, or low-resolution pictures, these CNN models have performed well.

Interpretability and Visualization The "black box" characteristics of deep learning models, which make it hard to understand model decisions, are one of their problems. Methods including Gradient-weighted Class Activation Mapping (GradCAM) and saliency maps have been used to solve this problem. These techniques show which elements of an image the model concentrates on when it makes predictions, therefore offering openness to doctors. Especially in demanding contexts such as cancer diagnosis, this interpretability is first in building confidence in AI systems [22, 23].

FIGURE 2 illustrates various AI models employed in osteosarcoma detection and classification across different studies.

3.2 Risk Stratification Using AI

Apart from enhancing diagnostic accuracy, artificial intelligence has shown strong potential to stratify individuals according to their risk of metastasis, relapse, and therapy failure. More customized treatment plans and preventive management of osteosarcoma patients have become possible thanks to artificial intelligence models that use both medical and imaging data to produce individual risk profiles.

3.2.1 Predicting metastasis risk

Especially in the lungs and other bones, osteosarcoma is famed for its strong metastatic potential [24, 25]. For beginning aggressive treatment methods, including chemotherapy and surgical operations, early identification of patients at high risk is imperative.

Ensemble Learning Approaches Particularly powerful in metastasis risk prediction are ensemble learning techniques like Gradient Boosting Machines (GBM), AdaBoost, and XGBoost. These models merge several weak learners to boost the general predictive quality. Ensemble models have reached remarkable sensitivity ratings of up to 92% by combining radiomic features from imaging

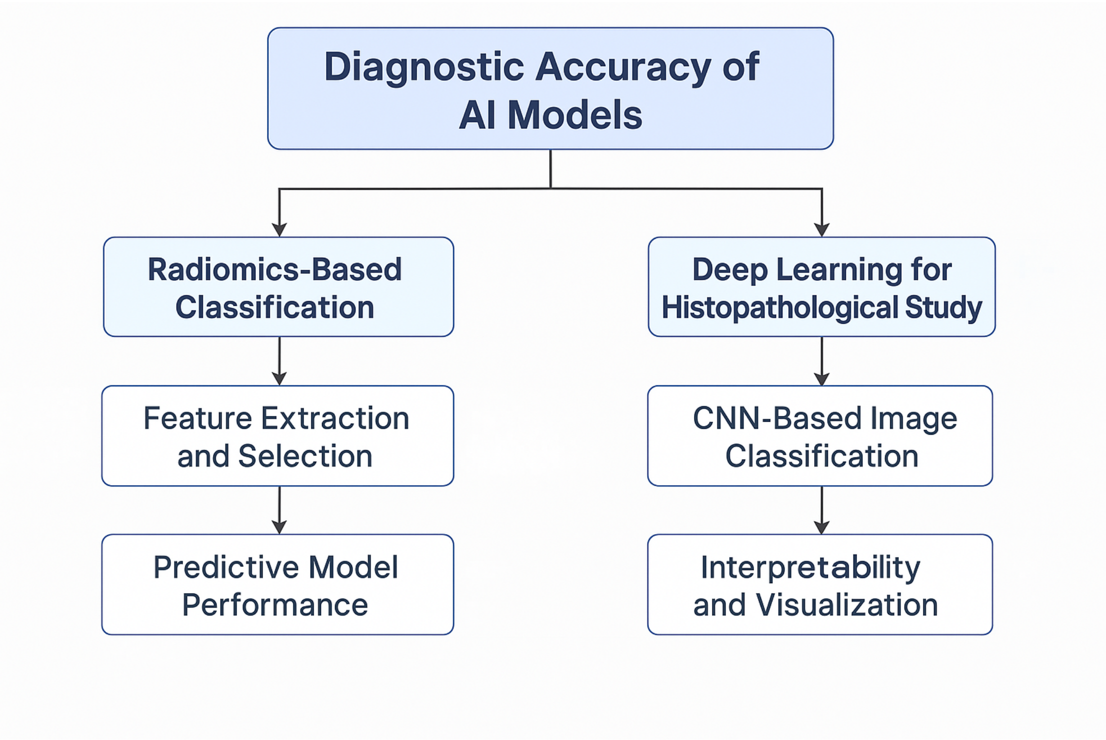

Figure 2: Diagnostic Accuracy of AI Models in Osteosarcoma Detection and Classification

data with clinical features (e.g., tumor size, patient age, serum biomarkers), therefore significantly outperforming standard methods [26].

Clinical Utility Flagging patients with high metastasis risk, AI models have been incorporated into clinical management. This enables medical providers to customize their management systems, giving more active observation and treatment for patients likely to have metastasis. Also, by means of artificial intelligence integrated into tumor board talks, doctors have been able to make better-informed judgments and thereby enhance patient outcomes [27].

3.2.2 Prognostic biomarker discovery

Identifying possible prognostic biomarkers, essential for estimating patient results and forecasting disease progression, has also been significantly aided by artificial intelligence methods.

Integration of Genomic Data To forecast patient results, machine learning algorithms have been trained on multi-omics data comprising gene expression profiles, DNA methylation patterns, and mutation data. AI systems could find gene signatures and molecular markers linked with bad prognosis, such as TP53 mutations and high VEGF expression, using extensive genomics and imaging datasets. In customized treatment planning, these biomarkers can be very useful aids [28, 29].

Personalized Prognosis Models Using artificial intelligence models combining molecular, radiographic, and clinical data, custom prognosis systems have been developed. By producing personalized risk profiles, these systems help doctors forecast a patient's chances of relapse and survival. In this way, these models offer important knowledge on the best long-term monitoring plans and ideal therapy techniques [30].

3.3 Integration with Clinical Practice

While AI-based models' performance in osteosarcoma diagnosis and risk stratification has been encouraging, incorporating them into actual clinical practices remains a formidable task. Some of the major concerns are the absence of clinical validation, variability in reporting standards, limited interpretability, and regulatory pathways for ensuring safety and efficacy. Implementation depends on overcoming the experimental-clinical utility gap by creating transparent, reproducible, and clinically validated AI systems.

3.3.1 Clinical Decision Support Systems (CDSS)

Emerging as powerful instruments for helping radiologists, pathologists, and oncologists by providing real-time insights for diagnosis, treatment planning, and risk stratification are AI-powered Clinical Decision Support Systems (CDSS). Still, for these systems to be reliable and legally usable, they have to follow regulatory frameworks, including the FDA's Software as a Medical Device (SaMD) and European CE marking requirements, and undergo strict clinical validation via multicenter trials.

Enhancing Workflow AI-based CDSS, when integrated with systems such as Picture Archiving and Communication Systems (PACS) or Electronic Health Record (EHRs), can highlight abnormalities in real-time, cue high-risk cases first, and minimize diagnostic delay. In the studies reviewed, such system-level improvements were suggested but not yet validated to the full extent within clinical environments. For example, Papalia et al. (2024) [31], and Wang et al. (2024) [32], had encouraging diagnostic precision (AUC > 0.8 and 92.5% accuracy, respectively), but did not have clinical deployment trials to confirm their suitability in actual hospital environments. Validation within such processes would involve prospective cohort studies, inter-observer agreement metrics, and impact assessment on diagnosis-to-treatment intervals [33, 34].

Clinician Acceptance Trust among clinicians is important for the adoption of AI. Interpretability strategies like saliency maps, Grad-CAM, and SHAP values have been successful in increasing clinician insight into AI predictions. For instance, Qu et al. (2024) [35], and Khan et al. (2024) [36], utilized interpretable biomarkers and DNA methylation profiles for risk stratification, maximizing their potential for clinical application. Nevertheless, without usability testing, clinician feedback loops, and training programs, these models are still restricted to research purposes. User-centered design approaches, incorporating feedback from radiologists and oncologists during development, should be a standard practice [37, 38].

3.3.2 Performance in multimodal settings

AI systems that fuse multi-modal data—integrating imaging, clinical information, and omics—exhibit improved prediction performance and more informative patient understandings. Such models have a large potential for clinical decision-making in practice but need to be validated clinically and externally.

Fusion of Imaging and Omics Data Several reviewed studies investigated multimodal data fusion. For example, Cè et al. (2025) [39], used imaging analytics on MRI, CT, and PET scans for survival stratification, whereas Aziz et al. (2023) [40], and Qu et al. (2024) [35], used RNA-seq and gene expression profiles. While these models enhanced performance measures like AUC and accuracy, none of them mentioned clinical trial-based validation. To translate the above findings into practice, there is a need to perform retrospective validation on independent hospital datasets and follow it with prospective observational trials. Phrases like "external validation," "real-world evidence (RWE)," and "clinical outcome assessment" are commonly employed to denote successful translation steps [41].

End-to-End Learning Systems End-to-end learning models streamline the AI pipeline by automatically extracting features and learning from raw inputs to diagnostic outcomes. Ogbonna et al. (2024) [42], and Wang et al. (2024) [32], employed deep learning for histopathological classification and MRI interpretation with high accuracy and specificity. These systems, though efficient, are black-box models in many instances, which brings with it concerns for clinical deployment. Regulatory bodies today advise Explainable AI (XAI) capabilities and open documentation to facilitate approval. Furthermore, model calibration (e.g., through Brier score or Expected Calibration Error), decision curve analysis, and prospective trial registration (e.g., ClinicalTrials.gov) should be undertaken to enable such models to work consistently across various patient populations [43].

3.3.3 Clinical validation status and recommendations

A thorough analysis of the reviewed studies reveals the following (see TABLE 3):

3.3.4 Pathways for Clinical Validation and Regulatory Approval

To move from research prototypes to clinically adopted systems, the following steps should be implemented:

Standardized Reporting

- Follow the **TRIPOD-AI**, **CONSORT-AI**, or **SPIRIT-AI** reporting guidelines for transparent and reproducible AI studies.
- Ensure datasets, preprocessing steps, and performance metrics are **fully disclosed and open for peer benchmarking**.

Table 3: Clinical Validation Status and Recommendations for Reviewed AI Models in Osteosarcoma Diagnosis and Risk Stratification

Study	Clinical Validation Status	Recommendation
Muhammad Ainul Fikri et al., 2025	No external or clinical validation	Needs a prospective clinical trial with patient-level data
Papalia et al., 2024	Meta-review; no direct validation	Recommend real-world pilot implementation and external validation
Ogbonna et al., 2024	Multimodal accuracy reported; no clinical validation	Validation using standardized DXA benchmarks and patient trials
Qu et al., 2024	Biomarker-based model validated on GSE21257	Validated model; ready for clinical translation study
Ong et al., 2024	Single-center studies; limited generalization	Multi-center external validation is required
Khan et al., 2024	High accuracy in a small cohort	Needs scale-up and validation across diverse genetic backgrounds
Cè et al., 2025	Multimodal model; performance metrics reported	Require clinical outcome-based validation
Wang et al., 2024	CNN with high accuracy; no clinical validation	Needs application in histopathology labs with patient follow-up
Aziz et al., 2023	Good performance; no clinical trial	Suggest integration with clinical genomics systems for RWE collection
Aziz et al., 2023	NLP-based model; F1-score reported	Needs external validation using annotated clinical reports

External and Prospective Validation

- Apply models to unseen, real-world hospital datasets from different institutions.
- Conduct prospective observational trials or pragmatic clinical trials under ethically approved settings.

Regulatory and Ethical Compliance

- Align model development with FDA SaMD or EU MDR requirements.
- Implement Bias and Fairness Audits to ensure generalization across age, gender, and ethnicity.

Model Interpretability and Clinician Involvement

- Incorporate explainable AI methods like SHAP, LIME, or attention heatmaps.
- Use co-design methodologies involving continuous feedback from domain experts.

Through the focus on standardized validation, regulatory compliance, and clinician-driven design, the studies reviewed here can shift from promising research findings to clinically relevant solutions. The reinforcement of these avenues will considerably advance reproducibility, credibility, and the potential for real-world value in managing osteosarcoma.

3.4 Comparative Analysis of AI Approaches

Early diagnosis, prognosis, and risk stratification in osteosarcoma have been greatly aided by artificial intelligence (AI) advances. Still, the successful clinical translation of artificial intelligence (AI) systems depends on critically evaluating their performance, interpretability, robustness, and adaptability to real-world limitations in addition to merely enumerating algorithms. The examined models are examined here in depth with regard to classical machine learning (ML), deep learning (DL), hybrid methods, dataset features, and model-specific benefits and constraints.

3.4.1 Classical ML vs. Deep learning

Strengths and Weaknesses Owing to their lesser computational demands, interpretability, and efficacy on smaller datasets, classical machine learning algorithms include Support Vector Machines (SVM), Random Forests (RF), and Logistic Regression have consistently shown strong performance on structured clinical datasets. For example, Fikri et al. [44], Used XGBoost, SVM, and RF on a balanced clinical dataset in 2025 to achieve high accuracies (94.7% for XGBoost and 94.39% for SVM), implying these models are successful when handling tabular or clinical data with engineered features [45].

By contrast, DL models—particularly Convolutional Neural Networks (CNNs)—thrive at processing high-dimensional and unstructured data such as histology slides and medical images. Wang et al., for instance, achieved a 92.5% accuracy in histopathological classification with CNNs in 2024; Cè et al. [39], Combined multimodal radiomics with deep learning in 2025 to improve survival stratification and treatment response prediction. Still, deep learning methods frequently suffer from the “black-box” phenomenon, need massive labeled datasets, and costly computing resources. Their immediate interpretability and application in resource-limited clinical environments are constrained by this [46, 47]. A comparative overview of the technical specifications, input features, model types, and performance metrics of the AI algorithms used across the reviewed studies is presented in TABLE 4.

Use-Case Suitability The reviewed studies illustrate that **classical ML models** are particularly well-suited for use cases involving structured, numerical, or categorical data, such as gene expression profiles (e.g., Qu et al., 2024 [35]) and clinical variables. Their interpretability makes them preferable for regulatory compliance and physician trust.

On the other hand, **DL models** are more effective in imaging-based use cases (e.g., histopathology, CT, MRI), as shown by Cè et al. (2025) [39], and Wang et al. (2024) [32]. They outperform classical models in feature abstraction and end-to-end learning from raw data. In genomics, **hybrid models** like autoencoder + RF [40], provide a middle ground by leveraging unsupervised feature extraction and interpretable classifiers.

Table 4: Comparative Technical Summary of AI Algorithms in Reviewed Studies

Study	Model Type	Input Type	Dataset Size	Strengths	Limitations	Accuracy/ AUC
Fikri et al. (2025)	Classical ML (XGBoost, SVM, RF)	Clinical Data	Custom (SMOTE Balanced)	High accuracy, interpretability	Limited to structured data	94.7%, 94.39%
Wang et al. (2024)	DL (CNN)	Histopathology	Public OS Dataset	Good at unstructured data	Black-box, data-intensive	92.5%
Aziz et al. (2023)	Hybrid (Autoencoder + RF)	RNA-Seq	TARGET-OS	Feature reduction + classification	Model complexity	AUC: 0.87
Papalia et al. (2024)	DL & ML (Meta-review)	SPECT Images	Meta-review (59 studies)	Broad coverage	Lack of implementation uniformity	AUC > 0.8
Hua et al. (2024)	NLP + SVM	Radiology Reports	Custom Clinical	Language-based prediction	Needs annotation	F1: 0.84

In real-world practice, **hybrid and ensemble models** that integrate DL for feature extraction and classical ML for decision-making are gaining traction, offering both accuracy and interpretability.

3.4.2 Challenges in model generalizability

Dataset Bias and Overfitting The shared limitation in the majority of studies is the limited size of the dataset and diversity of the sources. Numerous models (e.g., Ong et al., 2024 [48]; Khan et al., 2024 [36]) are trained with single-institution or small cohort data, some having fewer than 40 diagnosed cases. This can result in possible overfitting and poor generalization to populations with various demographics, imaging protocols, or clinical processes. For example, Khan et al. [36], have reported almost perfect accuracy in the classification of subtypes, but generalizability is doubtful because the sample size was too small.

Imbalanced datasets and a lack of external validation were also observed in a number of the studies, which hindered reproducibility.

Strategies for Improvement Many solutions have been suggested for generalizability and overfitting problems, including federated learning, which enables models to be trained on data from many sources without damage patient confidentiality. Cross-validation methods and transfer learning strategies also offer hope for increasing model robustness and improving generalizability among a range of kinds of patients [49]. To overcome these limitations, several strategies have been suggested or applied:

- SMOTE-based balancing (e.g., Fikri et al., 2025 [44]) helped mitigate class imbalance.
- Multimodal data fusion (e.g., Ogbonna et al. [42], 2024; Cè et al., 2025 [39]) increased model robustness by incorporating different imaging and clinical modalities.
- Transfer learning and cross-domain adaptation, though underutilized in reviewed studies, remain promising for leveraging large public datasets and fine-tuning on local data.
- Federated learning represents a future direction to train models across institutions without direct data sharing, addressing data privacy and diversity challenges.
- Benchmarking frameworks and multi-center validation studies are urgently needed to standardize evaluation metrics and improve reliability.

3.4.3 Comparative insights based on dataset scale and model performance

Among the reviewed studies:

- Fikri et al. [44], and Aziz et al. used structured and transcriptomic data and achieved strong binary classification performance (AUC: 0.87–94.7%) on modest-sized datasets.
- Wang et al. and Cè et al. demonstrated the utility of DL on imaging data with high accuracy, but lacked details on external validation.
- Khan et al. achieved exceptionally high accuracy in sarcoma subtype classification using methylation data, but results may be inflated due to the small dataset (36 cases).
- Papalia et al. performed a meta-review, offering general insights into model types but lacking implementation depth.

Thus, models that balance accuracy, interpretability, and data efficiency (like XGBoost or hybrid autoencoder + RF) are currently more adaptable to clinical deployment. In contrast, DL models remain more powerful but data-hungry and less explainable—requiring further improvement for regulatory approval and clinical trust.

The comparative analysis underscores the need for technical rigor, dataset diversity, and clinically meaningful outcomes in osteosarcoma AI research. Future studies must go beyond reporting accuracy and instead focus on algorithm transparency, external validation, and real-world applicability. Emerging approaches like explainable AI, multimodal learning, and federated architectures are likely to shape the next generation of AI tools in oncology. The overview of the key AI contributions in osteosarcoma diagnosis and risk stratification is depicted in FIGURE 3.

TABLE 5 provides a comprehensive summary of the reviewed studies focusing on AI-based approaches for the diagnosis and risk stratification of osteosarcoma, highlighting the methodologies, datasets used, performance metrics, and key findings across the literature.

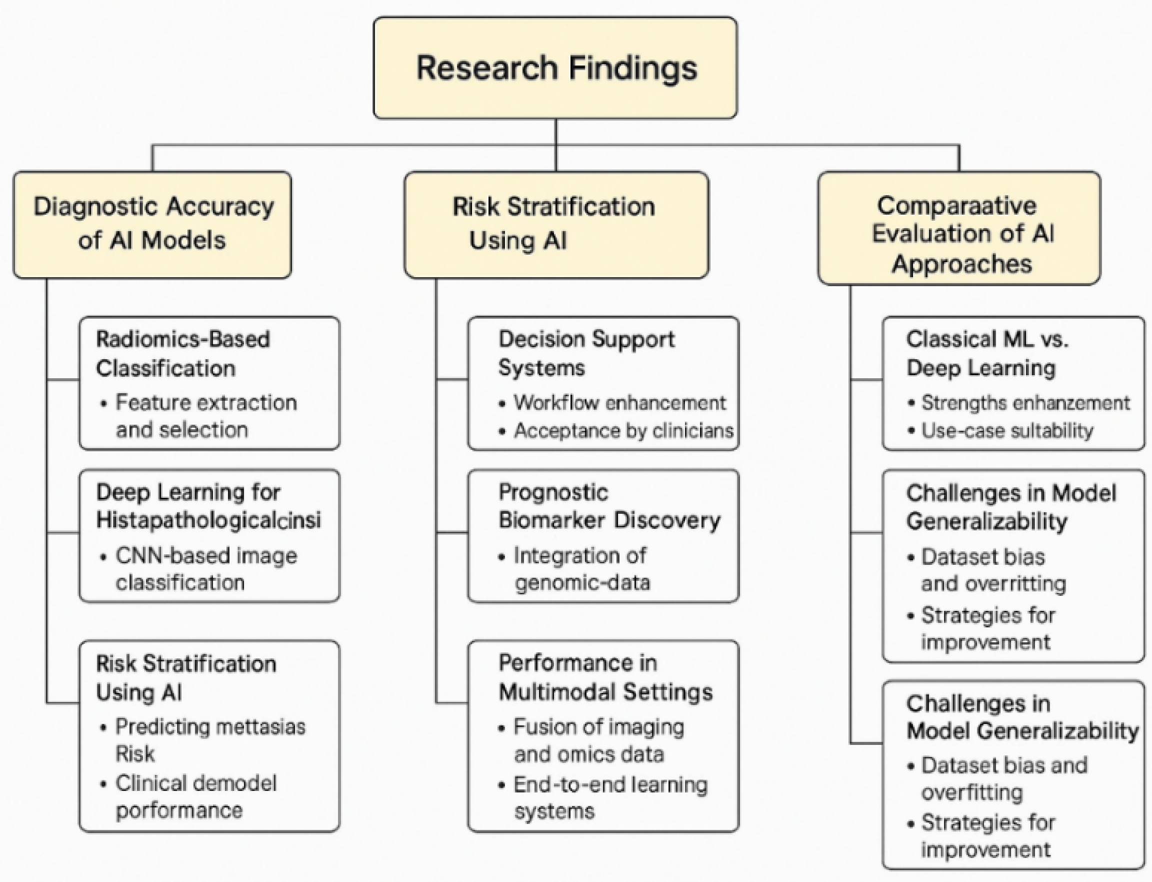

Figure 3: AI contributions in osteosarcoma diagnosis and risk stratification

4. LIMITATIONS AND FUTURE DIRECTIONS

Notwithstanding the encouraging developments in artificial intelligence (AI) for early detection and risk stratification of osteosarcoma, many constraints remain that impede the smooth incorporation of these technologies into clinical use. Understanding these difficulties is crucial for putting present results into context, but also for plotting a direction toward stronger, generalizable, and ethically applicable artificial intelligence solutions in osteosarcoma treatment.

4.1 Limitations

The current limitations in the application of AI in osteosarcoma prediction are as follows:

Table 5: Summary of Reviewed Studies on AI-Based Diagnosis and Risk Stratification in Osteosarcoma

Author, Year	**Target Variable**	**Input**	**Architecture**	**Pre-Processing**	**Dataset**	**Outcome**	**Output**	**Result**
Muhammad Ainul Fikri et al., 2025 [44]	Osteosarcoma Detection	Clinical Data	XGBoost, SVM, RF	SMOTE Balancing	Custom (Balanced via SMOTE)	Risk of OS	Binary	Acc.: 94.7% (XGB), 94.39% (SVM)
Papalia et al., 2024 [31]	Bone Metastasis Detection & Prognosis	Bone SPECT Images	DL & ML Models	Image Normalization	59 Studies (Meta-Review)	Metastasis Hotspot Detection	Binary	AUC > 0.8 (most models), RF & LR < 0.8
Ogbonna et al., 2024 [42]	Early Osteoporosis Detection	MRI, CT, HR-pQCT	Deep Learning	Multimodal Image Fusion	Studies (2014–2024)	Fracture Risk Prediction	Multi-class	Higher Accuracy, Sensitivity, Specificity than DXA
Qu et al., 2024 [35]	Prognosis in OS Patients	Gene Expression Profiles	SVM, RF	Biomarker Selection (SERPINE2, CPT1B)	GSE21257	High/Low Risk Stratification	Binary	Validated, Effective Prognostic Model
Ong et al., 2024 [48]	Spinal Cancer Detection & Management	CT Spine Images	ML, DL, Radiomics	Tumor Region Annotation	Multiple Single-Center Studies	Detection, Classification, Prognosis	Multi-class	Performance Variable; Generalization Limited
Khan et al., 2024 [36]	Sarcoma Subtype Classification	DNA Methylation Profiles	ML Classifier	Methylation Pattern Extraction	36 Confirmed Cases	OS, EWS, SS Classification	Multi-class	Acc.: ~100% (SS), 98.84% (OS), 93.33% (EWS)
Cè et al., 2025 [39]	Diagnosis & Prognosis of OS	MRI, CT, PET	Imaging Analytics + Radiomics	Multimodal Imaging Analysis	Multi-modal Clinical Data	Treatment Response Prediction	Multi-class	Improved Survival Stratification
Wang et al., 2024 [32]	Histopathological Classification	Histology Slides	CNN	Image Labeling	Public Osteosarcoma Histology Dataset	Cell Type Classification	Multi-class	Acc.: 92.5%
Aziz et al., 2023 [40]	Genomic Risk Stratification	RNA-seq Data	Deep Learning (Autoencoder + RF)	Gene Selection, Dimensionality Reduction	TARGET-OS	High/Low Risk Patient Prediction	Binary	AUC: 0.87
Hua et al., 2024 [50]	Osteosarcoma Metastasis Prediction	Radiological Reports, MRI	NLP + SVM	Entity Recognition, Tokenization	Clinical Report Dataset	Metastasis Prediction	Binary	F1: 0.84

4.1.1 Limited and imbalanced datasets

Given the paucity of big, varied, and well-annotated data sets particular to osteosarcoma, this is among the most significant restrictions in the current scenario. Existing databases usually have little sample sizes and class imbalances—usually are more benign or healthy samples than malignant ones—due to the uncommon nature of the condition. This restricts how well machine learning and deep learning models can generalize across populations, hence overfitting and decreased performance upon being presented with unknown clinical data. Most datasets come from one institution or geographical area; this reduces the models' relevance across several patient groups [51].

4.1.2 Lack of standardization in methodologies

The approaches used vary markedly across studies. Variations in image preprocessing approaches, feature extraction pipelines, model architectures, and performance evaluation criteria complicate the reproduction of results and result comparison. The lack of standardized practices for data handling, model validation, and reporting impedes the impartial benchmarking of artificial intelligence models and, hence, delays regulatory approval and medical translation [52].

4.1.3 Interpretability and Trust Issues

Although DL algorithms—particularly CNNs have shown excellent diagnostic accuracy, their "black box" nature presents problems in a medical setting. Many medical practitioners are reluctant to depend on AI-driven judgments devoid of explicit, interpretable justification. Although tools like GradCAM and SHAP have been introduced to improve transparency, they are not yet sufficient to fully bridge the gap between AI outputs and clinical reasoning [53].

4.1.4 Integration into clinical workflows

Transforming research models of AI systems into adaptable clinical tools still presents a major challenge. Not only will integrating AI models into current healthcare infrastructure, such as radiology information systems (RIS) and EHR, require technical adjustments, but also medical staff training, realignment with institutional policies, and changes in clinical workflow. Furthermore, varied by region and still changing are the regulatory structures governing the clinical implementation of artificial intelligence tools [54].

4.1.5 Ethical, legal, and privacy concerns

Especially sensitive imaging and genetic data, the use of patient data raises important legal and ethical concerns into play. Still difficult is making data anonymous, obtaining patient consent, and adhering to privacy policies, including GDPR and HIPAA. Training data biases can also cause unfair results, perhaps harming underrepresented groups in diagnosis and therapy planning [55].

4.2 Future Directions

Using conscientious, collaborative, multidisciplinary approaches, future studies should tackle the restrictions listed above to fully exploit AI's potential to improve osteosarcoma results.

4.2.1 Creation of large, multicenter datasets

Hospitals, research facilities, and data repositories must urgently cooperate to produce vast, multicenter databases. In this respect, federated learning could be very significant in allowing model training across distributed data sets without sacrificing patient confidentiality. Such a method would guarantee data variety and enhance model generalization among groups [56].

4.2.2 Development of standardized evaluation frameworks

The AI research industry should give top priority to establishing uniform standards for data preprocessing, model building, validation approaches, and performance metrics to help reproducibility and relevant cross-study comparisons. For AI-based medical diagnostics, programs designed especially for such assistance as CONSORT and TRIPOD might be invaluable in establishing standards and guaranteeing methodological accuracy [57].

4.2.3 Advancements in Explainable AI (XAI)

Future AI systems should pay attention to improving model transparency and explainability. Integrating explainable AI methods—attention mechanisms, decision trees inside deep models, and causal inference frameworks—can help healthcare professionals obtain a straightforward understanding of the decision-making process. This would help in medical acceptance and build confidence [58, 59].

4.2.4 Real-world clinical trials and validation

Most modern research supports AI models on synthetic or retrospective data sets. Next, possible clinical studies are needed to assess the real-world safety and effectiveness of artificial intelligence systems used for osteosarcoma diagnosis and prognosis. Apart from the accuracy of diagnosis, these studies should evaluate patient results, doctor satisfaction, and workflow efficiency.

4.2.5 Integration of multimodal and longitudinal data

For a more thorough knowledge of osteosarcoma biology, artificial intelligence models should be created to combine multimodal data, including imaging, histopathology, genomics, proteomics, and clinical history [60]. The inclusion of longitudinal data might also let artificial intelligence systems monitor the development of illnesses and forecast metastasis or relapse over time, therefore helping long-term therapy planning.

5. CONCLUSION

This systematic review highlights the changing possibilities of AI in the early detection and risk staging of osteosarcoma. Through examination of many AI approaches, including machine learning, deep learning, radiomics, and histopathological image analysis, this review shows how these tools are aiding more precise, timely, and data-driven diagnosis. Integrating artificial intelligence into clinical workflows has proven to be beneficial for improving diagnostic accuracy, mitigating human errors, and enhancing prognostic modeling. The reviewed uses show notable progress toward more individualized and effective osteosarcoma treatment despite existing limits in data availability, model interpretability, and clinical validation. Moreover, this analysis stresses the need for cross-disciplinary work, including pathologists, clinicians, bio informaticists, and data scientists, to overcome problems delaying the complete clinical application of artificial intelligence in oncology. Future work should center on creating ethical, transparent, clinically validated artificial intelligence tools that can easily interface with current medical systems. AI is set to become a vital component of osteosarcoma treatment as research develops, since it enables early prevention, directs therapy choices, and eventually raises patient quality of life and survival. Continuing advancement of AI applications offers the potential of changing cancer diagnosis and prognosis, therefore significantly advancing precision oncology worldwide.

References

[1] Borji A, Kronreif G, Angermayr B, Hatamikia S. Advanced Hybrid Deep Learning Model for Enhanced Classification of Osteosarcoma Histopathology Images. 2024. ArXiv Preprint: https://arxiv.org/pdf/2411.00832.

[2] Spaanderman DJ, Marzetti M, Wan X, Scarsbrook AF, Robinson P, et al. AI in Radiological Imaging of Soft-Tissue and Bone Tumours: A Systematic Review Evaluating Against Claim and Future-Ai Guidelines. EBiomedicine. 2025;114.

[3] Nguyen MD, Nguyen DT, Nguyen TV, Yamada H, Pham HH, et al. Bridging Classification and Segmentation in Osteosarcoma Assessment via Foundation and Discrete Diffusion Models. 2025. ArXiv Preprint: http://arxiv.org/abs/2501.01932.

[4] Capobianco E, Tao Z, Huang F. Risk Stratification System and Web-Based Nomogram Constructed for Predicting the Overall Survival of Primary Osteosarcoma Patients After Surgical Resection. Front Public Health. 2022;10:949500.

[5] Hasei J, Nakahara R, Otsuka Y, Nakamura Y, Hironari T, et al. High-Quality Expert Annotations Enhance Artificial Intelligence Model Accuracy for Osteosarcoma X-Ray Diagnosis. Cancer Sci. 2024;115:3695-3704.

[6] Hasei J, Nakahara R, Otsuka Y, Nakamura Y, Ikuta K, et al. The Three-Class Annotation Method Improves the AI Detection of Early-Stage Osteosarcoma on Plain Radiographs: A Novel Approach for Rare Cancer Diagnosis. Cancers (Basel). 2025;17:29.

[7] Rodriguez-Merchan EC. Some Artificial Intelligence Tools May Currently Be Useful in Orthopedic Surgery and Traumatology. World J Orthop. 2025;16:102252.

[8] Kawaguchi K, Miyama K, Endo M, Bise R, Kohashi K, et al. Viable Tumor Cell Density After Neoadjuvant Chemotherapy Assessed Using Deep Learning Model Reflects the Prognosis of Osteosarcoma. NPJ Precis Oncol. 2024;8:16.

[9] Varghese J. Artificial Intelligence in Medicine: Chances and Challenges for Wide Clinical Adoption. Visc Med. 20204;36:443-449.

[10] Stafie CS, Sufaru IG, Ghiciuc CM, Stafie II, Sufaru EC, et al. Exploring the Intersection of Artificial Intelligence and Clinical Healthcare: A Multidisciplinary Review. Diagnostics. 2023;13:1995.

[11] Deepa R, Arunkumar S, Jayaraj V, Sivasamy A. Healthcare's New Frontier: Ai-Driven Early Cancer Detection for Improved Well-Being. AIP Adv. 2023;13.

[12] Vezakis IA, Lambrou GI, Matsopoulos GK. Deep Learning Approaches to Osteosarcoma Diagnosis and Classification: A Comparative Methodological Approach. Cancers. 2023;15:2290.

[13] Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The Use of Artificial Intelligence, Machine Learning and Deep Learning in Oncologic Histopathology. J Oral Pathol Med. 2020;49:849-856.

[14] Gou F, Wu J. An Attention-Based Ai-Assisted Segmentation System for Osteosarcoma MRI Images. In: Proceedings IEEE International Conference on Bioinformatics and Biomedicine. 2022;1539-1543.

[15] M'Sabah CE, Bouziane A, Ferdi Y. A Survey on Deep Learning Methods for Cancer Diagnosis Using Multimodal Data Fusion. In: 9th E-Health and Bioengineering Conference. IEEE. 2021;1-4.

[16] Zhang W, Guo Y, Jin Q. Radiomics and Its Feature Selection: A Review. Symmetry. 2023;15:1834.

[17] Demircioğlu A. Benchmarking Feature Selection Methods in Radiomics. Invest Radiol. 2022;57:433-443.

[18] Jayachandran A, Ganesh S, Kumar SR. Multi-Stage Deep Convolutional Neural Network for Histopathological Analysis of Osteosarcoma. Neural Comput Appl. 2023;35:20351-20364.

[19] Wu Y, Cheng M, Huang S, Pei Z, Zuo Y, et al. Recent Advances of Deep Learning for Computational Histopathology: Principles and Applications. Cancers. 2022;14:1199.

[20] Ahmed I, Sardar H, Aljuaid H, Alam Khan FA, Nawaz M, et al. Convolutional Neural Network for Histopathological Osteosarcoma Image Classification. Comput Mater Continua. 2021;69:3365-3381.

[21] Chen W, Ayoub M, Liao M, Shi R, Zhang M, et al. A Fusion of VGG-16 and Vit Models for Improving Bone Tumor Classification in Computed Tomography. J Bone Oncol. 2023;43:100508.

[22] He J, Bi X. Automatic Classification of Spinal Osteosarcoma and Giant Cell Tumor of Bone Using Optimized Densenet. J Bone Oncol. 2024;46:100606.

[23] Shao J, Lin H, Ding L, Li B, Xu D, et al. Deep Learning for Differentiation of Osteolytic Osteosarcoma and Giant Cell Tumor Around the Knee Joint on Radiographs: A Multicenter Study. Insights Imaging. 2024;15:35.

[24] Sheen H, Kim W, Byun BH, Kong CB, Song WS, et al. Metastasis Risk Prediction Model in Osteosarcoma Using Metabolic Imaging Phenotypes: A Multivariable Radiomics Model. PLOS One. 2019;14:e0225242.

[25] Bai BL, Wu ZY, Weng SJ, Yang Q. Application of Interpretable Machine Learning Algorithms to Predict Distant Metastasis in Osteosarcoma. Cancer Med. 2023;12:5025-5034.

[26] Jiang J, Pan H, Li M, Qian B, Lin X, et al. Predictive Model for the 5-Year Survival Status of Osteosarcoma Patients Based on the SEER Database and Xgboost Algorithm. Sci Rep. 2021;11:5542.

[27] Senthil Kumar K, Miskovic V, Blasiak A, Sundar R, Pedrocchi AL, et al. Artificial Intelligence in Clinical Oncology: From Data to Digital Pathology and Treatment. Am Soc Clin Oncol Educ Book. 2023;43:e390084.

[28] Cai Z, Poulos RC, Liu J, Zhong Q. Machine Learning for Multi-Omics Data Integration in Cancer. iScience. 2022;25:103798.

[29] Wunder JS, Gokgoz N, Parkes R, Bull SB, Eskandarian S, et al. TP53 Mutations and Outcome in Osteosarcoma: A Prospective Multicenter Study. J Clin Oncol. 2005;23:1483-1490.

[30] Gaur K, Jagtap MM. Role of Artificial Intelligence and Machine Learning in Prediction Diagnosis and Prognosis of Cancer. Cureus. 2022;14:e31008.

[31] Papalia GF, Brigato P, Sisca L, Maltese G, Faiella E, et al. Artificial intelligence in detection, management, and prognosis of bone metastasis: a systematic review. Cancers. 2024;16:2700.

[32] Wang Y, Wang Z, Zhang B, Yang F. Comprehensive Diagnostic Model for Osteosarcoma Classification Using CT Imaging Features. J Bone Oncol. 2024;47:100622.

[33] Carrasco Ramírez JG. AI in Healthcare: Revolutionizing Patient Care With Predictive Analytics and Decision Support Systems. J Artif Intell Gen Sci. 2024;1:5-10.

[34] Wang L, Chen X, Zhang L, Li L, Huang Y, et al. Artificial Intelligence in Clinical Decision Support Systems for Oncology. Int J Med Sci. 2023;20:79-86.

[35] Qu H, Jiang J, Zhan X, Liang Y, Guo Q, et al. Integrating Artificial Intelligence in Osteosarcoma Prognosis: The Prognostic Significance of SERPINE2 and cpt1B Biomarkers. Sci Rep. 2024;14:4318.

[36] Khan AA, Kumar R N, Chakma S, Das S. Sarcoma Diagnosis by Dna Methylation Classifier: A Systematic Review Current Status and Future Prospects. Pathol Res Pract. 2024;263:155634.

[37] Lambert SI, Madi M, Sopka S, Lenes A, Stange H, et al. An Integrative Review on the Acceptance of Artificial Intelligence Among Healthcare Professionals in Hospitals. NPJ Digit Med. 2023;6:111.

[38] Shamszare H, Choudhury A. Clinicians' Perceptions of Artificial Intelligence: Focus On Workload Risk Trust Clinical Decision Making and Clinical Integration. Healthcare. 2023;11:2308.

[39] Cè M, Cellina M, Ueanukul T, Carrafiello G, Manatrakul R, et al. Multimodal Imaging of Osteosarcoma: From First Diagnosis to Radiomics. Cancers. 2025;17:599.

[40] Aziz MT, Mahmud SM, Elahe MF, Jahan H, Rahman MH, et al. A Novel Hybrid Approach for Classifying Osteosarcoma Using Deep Feature Extraction and Multilayer Perceptron. Diagnostics. 2023;13:2106.

[41] Huang W, Tan K, Zhang Z, Hu J, Dong S. A Review of Fusion Methods for Omics and Imaging Data. IEEE ACM Trans Comput Biol Bioinform. 2023;20:74-93.

[42] Ogbonna C, Onuiri EE. Predictive Diagnostic Model for Early Osteoporosis Detection Using Deep Learning and Multimodal Imaging Data: A Systematic Review and Meta-Analysis. Asian J Eng Appl Technol. 2024;13:28-35.

[43] J. Hasei *et al.*, "The Three-Class Annotation Method Improves the AI Detection of Early-Stage Osteosarcoma on Plain Radiographs: A Novel Approach for Rare Cancer Diagnosis," *Cancers (Basel)*, vol. 17, no. 1, Jan. 2025, doi: 10.3390/cancers17010029.

[44] Fikri MA, Wardhana AK, Riwanto Y, Partiwi IYR, Sekar Ningrum FSA, et al. Improving Osteosarcoma Detection Through Smote-Driven Machine Learning Approaches. Int J Inform Dev. 2025;13:517-529.

[45] Dinesh P, Vickram AS, Kalyanasundaram P. Medical Image Prediction for Diagnosis of Breast Cancer Disease Comparing the Machine Learning Algorithms: Svm Knn Logistic Regression Random Forest and Decision Tree to Measure Accuracy. ECS Trans. 2022;107:12681-12691.

[46] Latif J, Xiao C, Imran A, Tu S. Medical Imaging Using Machine Learning and Deep Learning Algorithms: A Review. In: 2nd International Conference on Computing Mathematics and Engineering Technologies iCoMET. IEEE. 2019:1-5.

[47] Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, et al. Machine Learning (ML) in Medicine: Review Applications and Challenges. Mathematics. 2021;9:2970.

[48] Ong W, Lee A, Tan WC, Fong KT, Lai DD, et al. Oncologic Applications of Artificial Intelligence and Deep Learning Methods in CT Spine Imaging—A Systematic Review. Cancers. 2024;16:2988

[49] Tang H, Sun N, Shen S. Improving Generalization of Deep Learning Models for Diagnostic Pathology by Increasing Variability in Training Data: Experiments on Osteosarcoma Subtypes. J Pathol Inform. 2021;12:30.

[50] Hua W, Xu B, Zhang X, Chen T. Ai-Assisted Diagnostic Potential of CT in Bone Oncology and Its Impact on Clinical Decision-Making for Intensive Care. J Bone Oncol. 2024;48:100639.

[51] Tasci E, Zhuge Y, Camphausen K, Krauze AV. Bias and Class Imbalance in Oncologic Data—Towards Inclusive and Transferrable AI in Large Scale Oncology Data Sets. Cancers. 2022;14:2897.

[52] Stanzione A, Cuocolo R, Ugga L, Verde F, Romeo V, et al. Oncologic Imaging and Radiomics: A Walkthrough Review of Methodological Challenges. Cancers. 2022;14:4871.

[53] Dhar T, Dey N, Borra S, Sherratt RS. Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust. IEEE Trans Technol Soc. 2023;4:68-75.

[54] Juluru K, Shih HH, Keshava Murthy KN, Elnajjar P, El-Rowmeim A, et al. Integrating AI Algorithms Into the Clinical Workflow. Radiol Artif Intell. 2021;3:e210013.

[55] Mohammad Amini MM, Jesus M, Fanaei Sheikholeslami D, Alves P, Hassanzadeh Benam A, et al. Artificial Intelligence Ethics and Challenges in Healthcare Applications: A Comprehensive Review in the Context of the European GDPR Mandate. Mach Learn Knowl Extr. 2023;5:1023-1035.

[56] Dang TK, Lan X, Weng J, Feng M. Federated Learning for Electronic Health Records. ACM Trans Intell Syst Technol. 2022;13:1-17.

[57] Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JF. Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. J Med Internet Res. 2022;24:e36823.

[58] Sheu RK, Pardeshi MS. A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System. Sensors. 2022;22:8068.

[59] Lötsch J, Kringel D, Ultsch A. Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients. BioMedInformatics. 2021;2:1-17.

[60] Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, et al. Artificial Intelligence for Multimodal Data Integration in Oncology. Cancer cell. 2022;40:1095-110.