Check for updates

International Journal of Osteoarchaeology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Next-Generation Paleopathology: Using Commercial AI in Bioarchaeological Diagnosis

Jessica Mongillo[1,2] | Giulia Vescovo[1] | Giulia Manzo[1] | Nicoletta Zedda[1] | Barbara Bramanti[1]

[1]Department of Environmental and Prevention Sciences, University of Ferrara, Ferrara, Italy | [2]Department of Humanities, University of Ferrara, Ferrara, Italy

Correspondence: Giulia Vescovo (giulia.vescovo@unife.it) | Giulia Manzo (giulia.manzo@unife.it)

Received: 4 December 2025 | **Revised:** 6 May 2026 | **Accepted:** 18 May 2026

Keywords: AI chatbots | artificial intelligence | differential diagnosis | paleopathology

ABSTRACT

Artificial intelligence encompasses computational systems capable of performing cognitive functions such as learning, reasoning, and problem-solving. Within this domain, generative AI and large language models such as ChatGPT, Gemini, and Copilot have shown significant potential in clinical diagnostics. Their capacity to generate structured differential diagnoses, integrate radiological criteria, and synthesize complex information underscores their transformative potential. Despite extensive validation in modern medical disciplines, their application in paleopathology, defined as the study of ancient disease through skeletal bones, remains unexplored. This study assessed the diagnostic performance of three commercial AI systems (ChatGPT 5.0 Plus, Gemini 2.5 Flash, and Copilot 365) in two archaeological cases presenting distinct osseous lesions. Diagnostic outcomes were compared with interpretations by three expert anthropologists to evaluate concordance and reliability. Complete agreement was achieved for the diagnosis of a spinal lesion, whereas partial agreement was observed for a patellar lesion. The findings demonstrate that AI systems can generate plausible differential diagnoses but also highlight variability in interpretive reasoning and prompt sensitivity. Differences among platforms suggest distinct conceptual frameworks in diagnostic inference. Although current AI tools cannot replace expert assessment, their integration into paleopathology may enhance diagnostic reproducibility, encourage interdisciplinary dialogue, and support the development of domain-specific datasets to improve future model accuracy and reliability.

1 | Introduction

Artificial intelligence (AI) refers to systems capable of learning, reasoning, and problem-solving (Raisch and Krakowski 2021). An advanced subset of AI, generative AI (GenAI), stands out for its ability to produce novel outputs in response to user-defined prompts (De Cremer and Narayanan 2023). Among GenAI tools, large language models (LLMs) play an increasingly prominent role in clinical medicine. Recent models, such as OpenAI's Generative Pretrained Transformer 5 (GPT-5), integrate cutting-edge performance in various fields such as mathematics, programming, writing, health-related questions, and even visual comprehension (OpenAI 2025; Uddin et al. 2026). Its diagnostic performance has been extensively validated across multiple disciplines (Takita et al. 2025; Younis et al. 2024). ChatGPT, trained on extensive textual datasets, has attracted substantial attention for its ability to emulate human-like responses and to assist in diagnostic interpretation, including imaging-based fields (Hosseini et al. 2023; Currie et al. 2023). However, AI outputs are inherently shaped by the data on which the models were trained, potentially introducing biases and inaccuracies, thereby raising ethical concerns when recommendations are interpreted as equivalent to human clinical judgment. Nonetheless, one of the key strengths of AI lies in its capacity to rapidly synthesize

Nicoletta Zedda and Barbara Bramanti have equal contributions.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *International Journal of Osteoarchaeology* published by John Wiley & Sons Ltd.

extensive bodies of literature, which may assist researchers in avoiding low-probability or nonproductive experimental directions (Dave et al. 2023).

Of other AI tools, such as Microsoft's Copilot and Google Gemini, the diagnostic utility remains underexplored (Takita et al. 2025; Bayala et al. 2025), whereas specialized medical chatbots targeted primarily at healthcare providers (e.g., Ada Health and Babylon Health) have been developed to support symptoms interpretation using contemporary clinical data and interaction with the patients (Alduhailan et al. 2025).

In forensic medicine, AI was recognized as a transformative development, with the potential to reshape traditional practices (Fang et al. 2020). Human forensic assessments, particularly morphological analyses of skeletal material, are inherently susceptible to fatigue, emotional influence, and cognitive bias (Nakhaeizadeh et al. 2014). AI offers a means to mitigate such constraints, providing consistent and objective analyses, unaffected by human fallibility (Kotěrová et al. 2018).

Considering the growing application of AI in diagnostics, we decided to test its effectiveness in paleopathology, which seeks to reconstruct past human health and disease from archaeological remains. Skeletal pathologies present particularly intricate diagnostic challenges due to the complex, multifactorial skeletal responses to disease and the frequent absence of pathognomonic lesions (Klaus 2017; Lorenzo et al. 2015; Ortner 2003). The discipline must adopt meticulous descriptive protocols and engage in structured differential diagnosis. Modern paleopathology now integrates evolutionary, epidemiological, and sociocultural frameworks in the interpretation of disease (Angel 1981; Ubelaker 1982; Zuckerman et al. 2011).

Critical to this evolution was the establishment of comprehensive reference works and diagnostic guides for skeletal diseases, grounded in systematic observation and documentation (Aufderheide et al. 1998; Buikstra and Ubelaker 1994; Ortner 2003; Steinbock 1976). The diagnostic protocol in paleopathology was most comprehensively formalized by Ortner (2003, 2012), who described differential diagnosis as a tiered, problem-solving approach. The process begins with an understanding of bone cell biology (Ragsdale and Lehmer 2012), followed by detailed lesion characterization and a skeleton-wide distributional analysis. Subsequently, lesions are categorized, and differential diagnoses are carefully weighed against one another.

Although this diagnostic framework significantly strengthened the scientific foundation of disease identification in historical populations, there is still room for further improvement. AI, especially LLMs, has not yet been systematically applied to paleopathological datasets. Given the effectiveness of these models in modern medical diagnostics, their integration into archaeological research may help address long-standing interpretive challenges.

This study aims to evaluate the capacity of commercial AI systems to contribute to the evaluation of paleopathological diagnoses by testing their performance on two skeletonized individuals presenting distinct lesions. Two cases were deliberately selected: one relatively easy to diagnose and one more challenging to investigate how diagnostic difficulty affects the outcome.

2 | Materials and Methods

To explore the capacity of commercial AI systems to contribute to differential diagnosis in paleopathology, two archaeological cases were selected to represent lesions of differing etiology and interpretative complexity: (i) a 19th-century individual with a spinal lesion and (ii) an adolescent from a 15th-century cemetery with a patellar lesion.

Firstly, we used and adapted a pragmatic framework for selecting optimal prompts to maximize the capabilities of AI applications in healthcare by Meskó (2023) (Figure 1).

A rigorous differential diagnosis was undertaken following the instructions of Lawler (2017), in accordance with the principal reference works in bioarchaeology (Appleby et al. 2015; Appleby 2022; Brothwell 2012; Buikstra 2010; Grauer 2012; Ortner 2003, 2012). Full skeletal analyses were performed, as a comprehensive examination is essential for establishing inclusion and exclusion criteria in paleopathological diagnosis.

As an initial step in differential diagnosis, the biological profile (sex, age at death, and body size) was reconstructed (Lawler 2017). Age at death was estimated using epiphyseal fusion, dental eruption, and morphology of the auricular surface, whereas sex was assessed through morphological analysis of the pelvis, supplemented by humeral, mandibular, and iliac traits in younger individuals. Body size (stature and body mass) was reconstructed following Ruff's methods for adults and subadults, and contextual data were obtained from excavation records and reports.

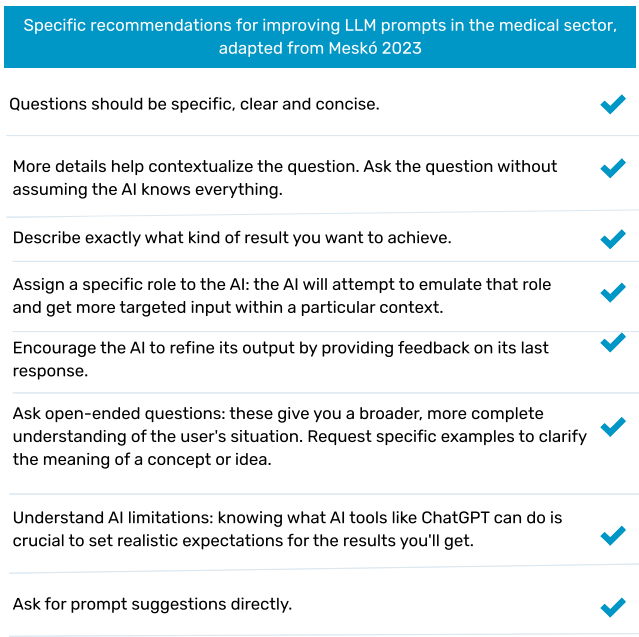


FIGURE 1 | Framework for selecting optimal prompts in healthcare, adapted from Meskó (2023). [Colour figure can be viewed at wileyonlinelibrary.com]

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Three biological anthropologists independently investigated the remains to complete the expanded information set (Lawler 2017) with physical examination and problem list, diagnostic imaging, and differential diagnosis. Their consensus was used to generate a possible diagnosis of diffuse idiopathic skeletal hyperostosis (DISH) for the first case and of osteoid osteoma (OO) or dorsal defect of the patella (DDP) for the second case. Sections 2.1 and 2.2 provide a detailed account of the nature of the lesions.

These cases were then submitted to three commercial AI chatbots (Gemini 2.5 Flash, Google; ChatGPT 5.0, OpenAI; and Copilot 365, Microsoft). Each system was asked to generate and refine prompts for differential diagnosis, incorporating both textual data and imaging (2D bone photographs, radiographs, and CT scans). For the second case, we submitted a query to refine the diagnosis and to understand the behavior of AI after our suggestion. AI outputs were required to present plausible differential diagnoses ranked by likelihood, and results were compared with the anthropologists' consensus using a 5-point matrix.

2.1 | DISH

DISH is diagnosed radiographically by the presence of flowing ossification along the anterolateral aspect of at least four contiguous vertebrae, often accompanied by extraspinal enthesophytes (Ortner 2003; Resnick et al. 1975). A defining feature is the exuberant ossification of the anterior longitudinal ligament, typically observed on the right side of the thoracic spine, because the anatomical position of the descending aorta on the left may inhibit ossification in that region (Ezra et al. 2025; Kuperus et al. 2020; Ortner 2003). Although DISH is generally straightforward to identify in clinical settings, its diagnosis in archaeological remains can be complicated by postmortem damage or incomplete pathological development (see also Table S1.1).

2.2 | Benign Bone Tumors/Tumor-Like Lesions

Benign bone tumors are neoplasms that do not produce metastases; they typically develop between childhood and young adulthood and grow slowly over the individual's lifetime (Hakim et al. 2015; Siek 2019). Benign bone tumors can be classified according to the region or bone involved (Hakim et al. 2015; Siek 2019) and the nature of the lesion (Anderson et al. 2021). Their diagnosis can be challenging, particularly when tumor-like lesions mimic neoplastic bone changes. They are usually the result of congenital, inflammatory, ischemic, and traumatic disorders (Alyas et al. 2007) and include a variety of lesions (Atif et al. 2018; Singh et al. 2009). The patella is a skeletal site rarely affected by the development of tumors or tumor-like lesions.

In the present study, we consider the patellar lesion as highly consistent with OO (see Section 3), a benign bone tumor first recognized by Jaffe (1935) as a distinct pathological entity and later described by Lichenstein as "a small round tumour like nidus composed of osteoid and trabeculae of newly formed bone deposited within a highly vascularized osteogenic connective tissue" (Lichtenstein 1965). OO is usually less than 1 cm in diameter and is classified into three different categories: cortical, articular, and subperiosteal. It mainly affects young, male individuals.

Alternatively, the observed patellar lesion is also highly consistent with DDP, a tumor-like lesion found in about 1% of the population, typically in preadolescents. DDP is defined as a lucent, round lesion with well-defined margins, located in the superolateral quadrant of the patella; the lesion is always identical in location and in radiographic appearance (Goergen et al. 1979; Van Holsbeeck et al. 1987). DDP is considered the result of a stress-induced anomaly caused by a traction of the vastus lateralis muscle (see Table S1.2).

3 | Results

Table 1 presents the differential diagnosis form of the skeletal individual with spinal lesion, divided into an initial information set and a second section dedicated to the list of problems, in which the level of agreement among the three anthropologists (A1, A2, and A3) is reported. The symbol ✓ indicates that the lesion was observed and included in both skeletal analyses and imaging, whereas the symbol ✗ denotes that it was not observed or not considered during the analysis. Regarding the differential diagnosis, each pathology was evaluated on a 5-point likelihood scale, and a corresponding list was compiled based on the consensus between the three anthropologists.

In the 'list of problems', only one instance of marked disagreement (two out of three in agreement) is reported, and two instances in which one of the three observers did not record the lesion are reported. However, complete concordance was achieved for the differential diagnosis of DISH. Degenerative spinal disease was considered moderately likely by two of the three evaluators, whereas AS was excluded. OA was judged highly likely by two of the three.

Table 2 shows the consensus among the anthropologists regarding the patella's lesion. The problem list also includes other skeletal alterations, as a comprehensive evaluation of the entire skeleton is essential to determine whether any relationship exists among the various pathological manifestations. The anthropologists show broad consensus concerning the patellar lesion. However, the other skeletal alterations were not uniformly considered, owing to differences in the interpretation of pitting observed on the endocranial surface and, in one of the three assessments, the omission of porotic hyperostosis. Moreover, a lytic lesion on the mastoid process was noted but not consistently included in the evaluations.

The three anthropologists agreed on the likelihood of OO for its high consistency. Of the remaining pathologies, DDP was assessed as highly consistent by two of the three anthropologists. The likelihood list, therefore, includes only those conditions for which at least two of the three observers agreed.

3.1 | Prompting

Appendix S1 describes the prompt used in this study for each AI chatbot. Copilot performs best when provided with a discursive

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Differential diagnosis form of the skeletal individual with a spinal lesion and consensus among the anthropologists.

Initial information set			
Specimen US 14 (Tb. 101)			
Species	*Homo sapiens*		
Ethnicity/provenance	N-W Italy (Emilia-Romagna)		
Sex	Female (Acsádi and Nemeskéri 1970; WEA 1980)		
Age at death	50+ (Lovejoy et al. 1985)		
Stature	160.30 ± 2.92 cm (Ruff et al. 2012)		
Body mass	55.84 ± 7.8 kg (Ruff et al. 2012)		
Initial evaluation			
Datation	1800–1900		
Burial circumstances	Inhumation in soil, single burial		
State of preservation	Good		
Taphonomic observations	Partially fragmented		
Ecology			
Environment	Rural		
Social context	Low classes		
Archaeological context	Cemetery of a parish church (Santa Maria di Toano, Reggio Emilia, Italy)		
Expanded information set			
Problem list			
	A1	**A2**	**A3**
Ossification along right anterolateral T5–T11 thoracic bodies	✓	✓	✓
Curved osteophyte at T4–T5 with horizontal and vertical growth	✓	✓	✓
Osteoarthritis of the cervical zygapophyseal joint and bodies	✓	✓	✓
Eburnation of the odontoid process	✓	✗	✗
Other lesions			
Bilateral femoropatellar osteoarthritis with eburnation	✓	✓	✓
Marginal lipping of distal femoral condyles and patellae	✓	✓	✓
Marginal lipping of femoral heads	✓	✓	✓
Scapular articular lipping bilaterally	✓	✗	✓
Clavicular enthesopathies	✓	✓	✓
Iliac crest osteophytes	✗	✓	✓
Imaging (RX)			
T5–T11 thoracic bodies with flowing ossifications at the anterolateral aspect of multiple contiguous levels on the right side	✓	✓	✓
Preservation of disc height	✓	✓	✓

(Continues)

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | (Continued)

Expanded information set			
Absence of apophyseal joint ankylosis	✓	✓	✓
Differential diagnosis			
DISH	5	5	5
Degenerative spinal disease	4	4	3
Ankylosing spondylitis	0	0	0
Osteoarthrosis	5	5	4
Refined differential diagnosis (likelihood)			
DISH	Highly likely		
Osteoarthrosis (OA)	Highly likely (coexisting)		
Degenerative spinal disease (DSD)	Moderately likely		
Ankylosing spondylitis (AS)	Ruled out		

prompt in which all relevant information is incorporated into the text, for example, "I am analysing a palaeopathological case ...." The prompt should include contextual information, the observed pathological features, relevant metadata, and the analytical objectives and conclude with a clearly formulated question, such as "Based on this information, what are the most likely differential diagnoses? Please consider both infectious and noninfectious aetiologies and explain how each diagnosis aligns or conflicts with the observed features." However, when dealing with a specific case, for instance, a spinal injury, it is advisable to avoid requesting unnecessary extensions of the analysis (e.g., to infectious and noninfectious diseases) if the etiology is already reasonably well understood. The chatbot should not be encouraged to extend its reasoning beyond the defined scope of the case.

For ChatGPT, the ideal approach is for the chatbot to take on the role of a paleopathologist: "You are a specialist in palaeopathology. Given the following observations on skeletal remains, provide a differential diagnosis including possible pathological conditions ...." As with Copilot, all relevant contextual information must be included explicitly, without assuming that the AI possesses prior knowledge. Specific, targeted questions should be asked, relating solely to the lesion or injury under investigation.

For Gemini, it is first necessary to request the generation of a suitable prompt; the system will then produce one in markdown format: "Please perform a differential diagnosis for the following paleopathological case. I will provide details on the skeletal remains, observed lesions, and the archaeological context. **Individual:** [e.g., Age at death (e.g., mature adult, subadult, infant), Sex (if determinable), Skeletal element(s) affected (e.g., femur, tibia, cranium, vertebrae)] ...." Although the output style can be adjusted, the markdown format was found to yield the most effective results. Copying and pasting the prompt while preserving Gemini's formatting optimizes the model's response.

For all three chatbots, a detailed description of the uploaded images is essential, as AI systems cannot reliably recognize dry bone or distinguish between anatomical structures. It is therefore crucial to specify what has been uploaded, which anatomical region it represents, which type of image it is, and which orientation it has (e.g., frontal or lateral view). For Copilot and ChatGPT, images may be pasted directly, whereas Gemini performs best with single images uploaded individually, preferably in PNG or JPEG/JPG format (see Figures S1.1 and S1.2 for spinal defect images submitted to chatbots and Figures S2.1, S2.2, and S2.3 for patellar bone lesion images submitted to chatbots).

Using our differential diagnosis as a reference, we instructed the chatbots to remain neutral, to disagree when appropriate, and to reconsider pathologies they had initially excluded: "Given that you can see that slightly oval subperiosteal and cortical hole, a lesion measuring less than 1 cm, a well-defined lytic lesion, and a reactive osteosclerotic border, why didn't you also consider osteoid osteoma?" to evaluate how it would respond after a suggestion.

3.2 | Differential Diagnosis

Table 3 includes the differential diagnosis with reasons performed by the three chatbots for the case of spinal pathology.

Without radiographic imaging, Copilot identified DISH as the most likely explanation but considered several alternatives. After X-ray examination, DISH was confirmed, with AS and metabolic disorders excluded based on preserved disc spaces, lack of sacroiliac involvement, and absence of systemic sclerosis.

Gemini also regarded DISH as the most likely diagnosis from the outset but placed greater emphasis on the co-occurrence of generalized OA. Whereas Copilot considered OA a secondary condition, Gemini interpreted it as a central component of the individual's morbidity profile. ChatGPT likewise identified DISH as the most probable pre-imaging diagnosis but placed special weight on morphological criteria. Following imaging, ChatGPT confirmed DISH as highly likely by applying established radiological standards (Resnick and Niwayama 1976),

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Differential diagnosis form of the skeletal individual with a patella exhibiting a lesion and consensus among the anthropologists.

Initial information set				
Specimen ζ20				
Species	*Homo sapiens*			
Ethnicity/provenance	Ferrara (N-E Italy)			
Sex	Male (Schutkowski 1993; Rogers 1999, 2009)			
Age at death	15 ± 3 (Ubelaker and Grant 1989)			
Stature	168.72 ± 3.2 cm (Ruff et al. 2012)			
Body mass	55.84 ± 7.8 kg (Ruff 2007)			
Initial evaluation				
Datation	XV century			
Burial circumstances	Single soil grave			
State of preservation	Good			
Taphonomic observations	Partially fragmented			
Ecology				
Environment	Urban			
Nutrition	—			
Original climate	—			
Social context	Middle-low classes			
Archaeological context	Medieval cemetery of the St. Nicolò Church (Ferrara, Italy)			
Expanded information set				
Problem list				
		A1	**A2**	**A3**
Lytic lesion located on the left lateral superior margin of the lateral articular surface of the left patella		✓	✓	✓
The lesion is rounded		✓	✓	✓
The lesion presents regular borders with an internal cavity		✓	✓	✓
Empty cavity is observed inside the lesion		✓	✓	✓
Other lesions				
Lytic lesions are present in the left mastoid process		✓	✓	✗
Pitting observed on the endocranial surface		✓	✗	✗
Porotic hyperostosis, with some pores showing signs of healing		✗	✓	✓
Mild cavitation is observed in the vertebral bodies		✓	✓	✓
Imaging plain and CT				
Subperiosteal and cortical slightly oval hole		✓	✓	✓
Lesion measuring < 1 cm		✓	✓	✓
Well-defined lytic lesion		✓	✗	✓

(Continues)

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | (Continued)

Expanded information set			
Reactive osteosclerosis rim	✓	✓	✓
Differential diagnosis			
Osteomyelitis hematogenous/chronic/Brodie's abscess	2	3	3
Osteoarticular tuberculosis	0	✗	1
Giant cell tumor	2	1	1
Aneurysmal bone cyst	0	0	0
Chondroblastoma	4	5	4
Brown tumor of hyperparathyroidism	2	1	✗
Osteoblastoma	4	3	3
Osteosarcoma	1	1	0
Osteoid osteoma	5	5	5
Dorsal defect of the patella	✗	5	5
Refined differential diagnosis (likelihood)			
Osteoid osteoma (OO)	Highly likely		
Dorsal defect of the patella (DDP)	Highly likely		
Chondroblastoma (CBL)	Moderately likely		
Osteoblastoma (OB)	Less likely		
Osteomyelitis hematogenous/chronic/Brodie's abscess (O/BA)	Less likely		
Giant cell tumor (GCT)	Very unlikely		
Brown tumor of hyperparathyroidism (BTH)	Very unlikely		
Osteosarcoma (OS)	Very unlikely		
Aneurysmal bone cyst (ABC)	Ruled out		
Osteoarticular tuberculosis (O TB)	Ruled out		

including flowing ossification across at least four contiguous vertebrae, preserved disc spaces, and absence of sacroiliac or apophyseal joint fusion.

The reference outputs indicate notable inconsistencies across systems. Gemini cites plausible authors but includes incorrect titles and one publication that could not be verified. ChatGPT presents references with generally appropriate authorship but includes incorrect DOIs and wrong publication years. Copilot generates the most extensive set of references, though its reliability is mixed: Several are accurate and complete, whereas others contain mismatched DOIs or authorship inconsistencies (Appendix S2.1).

The matrix (Figure 2) summarizes the likelihood scores (0 = *ruled out*; 5 = *highly likely*) assigned by each chatbot and the human expert to potential diagnoses associated with the observed vertebral alterations. A strong consensus is evident for DISH and OA, both rated as highly likely (score = 5) by all evaluators. DSD received slightly lower but still consistent scores, whereas the other conditions were unanimously ruled out.

Table 4 includes the differential diagnosis with reasons performed by the three chatbots for the patellar lesion.

Copilot initially approached the lesion within a neoplastic and proliferative framework, identifying *CBL* as the most probable diagnosis while considering also *LCH* and O/BA as less likely alternatives. Following CT examination, it continued to favor chondroblastoma emphasizing its occasional small size and circumscribed margins in subadult cases. After the query, DDP was explicitly ruled out, and OO was considered very unlikely.

Gemini concluded that the size and location of the lesion are pathognomonic for DDP and highlighted signs of systemic anemia, cribra orbitalia, and porotic hyperostosis as indicators of broader physiological stress. Post-imaging assessment confirmed the developmental interpretation, excluding neoplastic and infectious processes on the basis of the lesion's smooth, stable borders and absence of periosteal reaction.

ChatGPT initially favored an infectious interpretation (O/BA) as the leading diagnosis. However, CT imaging led to a significant revision: The sharply demarcated, smooth-walled cavity with

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Differential diagnosis with reason performed by the three chatbots for the case of spinal pathology.

AI chatbots	Copilot	Gemini	ChatGPT
Differential diagnoses (pre-imaging)	1. DISH (highly likely). 2. OA (likely, coexisting). 3. AS (unlikely). 4. F/HA (very unlikely)	1. DISH (highly likely). 2. OA (moderate–high likelihood, coexisting). 3. SA (less likely). 4. Severe spondylosis deformans (less likely)	1. DISH (highly likely). 2. OA (concomitant). 3. AS (unlikely). 4. Forestier's disease (= DISH). 5. Infectious/metabolic disorders (remote)
Differential diagnoses (post-imaging)	1. DISH (highly likely). 2. OA (coexisting). 3. ASs (ruled out). 4. F/HA (ruled out)	1. DISH (confirmed). 2. Generalized OA (confirmed comorbidity). 3. SA (ruled out)	1. DISH (highly likely). 2. OA (coexisting). 3. AS (excluded). 4. TS (ruled out). 5. F/HA (ruled out)
Radiographic features	Flowing anterolateral ossification T5–T11; preserved disc spaces; absence of vertebral fusion; osteophytes	Lateral radiograph: Thick, undulating, continuous anterior ossification (T5–T11); preserved intervertebral disc height	Radiography: Contiguous ossification across ≥ 4 vertebrae; right-sided predominance; disc height preserved; intact cortical bone
Consistent characteristics	Flowing "candle-wax" ossification T5–T11; preserved disc spaces; extraspinal enthesopathies; femoropatellar and cervical OA; macroscopic spinal changes; extraspinal enthesopathies; age; body mass	Age > 50, female sex; thoracic hyperostosis (T5–T11); peripheral OA (hips, knees, clavicles, iliac crest); preserved disc height; spinal morphology typical of DISH; radiographic morphology diagnostic for DISH; severe peripheral OA and enthesopathies	Flowing anterolateral ossification (T5–T11); preserved disc spaces; elderly female; systemic enthesophytes (clavicle, scapula, iliac crest); radiographic flowing ossification; exclusion of sacroiliac/zygapophyseal fusion; systemic enthesopathy
Contrasting characteristics	Absence of sacroiliac fusion; no vertebral squaring; no bamboo spine; absence of sacroiliac involvement or vertebral fusion	Severe eburnation of peripheral joints is more pronounced than expected for secondary OA in DISH, indicating primary OA comorbidity; degree of peripheral OA is greater than expected in secondary OA alone	No sacroiliac fusion; no inflammatory erosions; ossification pattern horizontal (not vertical syndesmophytes); absence of inflammatory features, lytic lesions, or systemic skeletal disease
Inclusion/exclusion criteria	Included: Age, spinal ossification, extraspinal enthesopathies. Excluded: Inflammatory joint disease, systemic metabolic disorders.	Included: Flowing "candle-wax" morphology, preserved disc spaces, enthesopathies. Excluded: Thin syndesmophytes, severe disc space narrowing.	Included: ≥ 4 contiguous vertebrae ossified, preserved disc height, systemic enthesopathies. Excluded: Sacroiliac/apophyseal fusion, inflammatory or infectious pathology.
Historical context	Rural parish cemetery, late 19th to early 20th century; lifestyle and diet potential contributing factors	Late 19th to early 20th century, rural north-west Italy; context consistent with age-related and metabolic disease	Forestier and Rotes-Querol (1950) first described DISH; consistent with later paleopathological recognition; parallels with published European cases
Taphonomic considerations	Well preserved; postmortem fractures and bone loss not obscuring pathology	Postmortem fragmentation at vertebral margins, but flowing hyperostosis intact	Minimal postmortem alteration; ossification firmly attached, antemortem in origin
Paleopathological comparisons	DISH documented in older adults in postmedieval Europe, frequently coexisting with OA (Rogers and Waldron 2001; Verlaan et al. 2007)	Radiographic pattern directly comparable to canonical DISH cases (Resnick criteria) (Resnick and Niwayama 1976; Rogers and Waldron 2001)	Consistent with archaeological reports of DISH in elderly individuals (e.g., Roger and Waldron 1995; Pinhasi and Mays 2008)

Abbreviations: AS: ankylosing spondylitis; DISH: diffuse idiopathic skeletal hyperostosis; DSD: degenerative spinal disease; F/HA: fluorosis/hypervitaminosis A; OA: osteoarthritis; SA: spondyloarthropathy; TS: tuberculous spondylitis.

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

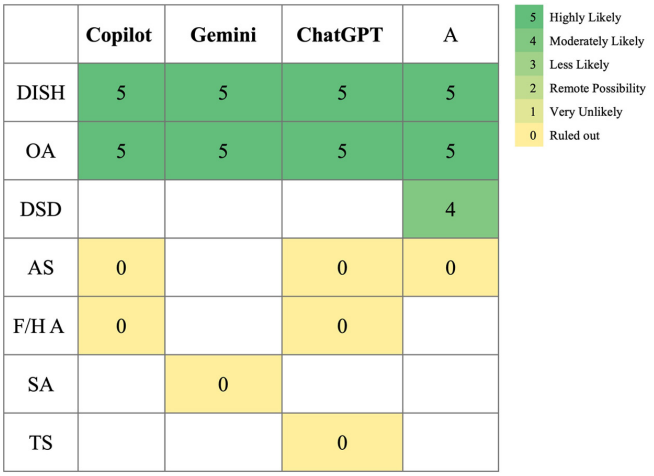


FIGURE 2 | Matrix of the likelihood of each pathology showing consensus among the three chatbots and the anthropologists (A) for the spinal lesion. AS: ankylosing spondylitis; DISH: diffuse idiopathic skeletal hyperostosis; DSD: degenerative spinal disease; F/HA: fluorosis/hypervitaminosis A; OA: osteoarthritis; SA: spondyloarthropathy; TS: tuberculous spondylitis. [Colour figure can be viewed at wileyonlinelibrary.com]

minimal reactive sclerosis was found to be inconsistent with infection. Consequently, the lesion was reclassified as a DDP. Post query, OO was deemed moderately likely.

With respect to bibliographic references, all three systems produced numerous citations encompassing paleopathological, medical, and online sources. ChatGPT generated the most comprehensive set, but as observed in the previous case, multiple inaccuracies were detected, including wrong publication years, missing authors, and incomplete or unverifiable references. Gemini exhibited different inconsistencies, often conflating authors and titles from unrelated works. Copilot, although occasionally accurate in matching authors and publication years, also contained minor errors in the title (Appendix S2.2).

The matrix (Figure 3) illustrates the level of diagnostic agreement for each hypothesized condition, with a score ranging from 0 (*excluded*) to 5 (*highly probable*). Overall, the highest agreement is observed for DDP. There is also moderate consensus for OO and CBL, whereas other conditions, such as LCH or OS, received low or inconsistent probability scores.

4 | Discussions and Conclusions

In this study, we evaluated the capability of three commercial AI software (Gemini, ChatGPT, and Copilot) to perform differential diagnosis in paleopathology. We tested strategies for submitting data to the chatbots, assessed their diagnostic outputs, and compared these with anthropologists' interpretations, analyzing the reasons for any discrepancies. Two lesion types were used, and interobserver agreement among three anthropologists was tested to ensure a reliable comparison with the software.

Despite minor inconsistencies, complete concordance was achieved for the diagnosis of DISH, which all evaluators recognized as highly consistent with the observed features (100% agreement). Divergence in the list of recorded problems is not unexpected in paleopathological interobserver studies, reflecting differences in both lesion recognition and diagnostic weighting (Anderson et al. 2021; Grootscholten et al. 2008). In contrast, the patellar lesion presented a substantially greater diagnostic challenge. Its rarity in archaeological contexts, the absence of soft tissues, and the impossibility of observing disease progression all contribute to the inherent difficulty of distinguishing benign bone tumors from tumor-like conditions in paleopathology, a problem further aggravated by the osteological paradox (Siek 2013).

Nevertheless, broad agreement was also achieved for the patellar lesion, which was interpreted as highly consistent with both OO and DDP, both considered benign, chronic, and nonaggressive processes. Consensus favored OO as highly likely, whereas DDP was judged highly likely by two of the three anthropologists. However, the absence of a visible nidus and lack of periosteal thickening prevent the lesion from being considered diagnostic of OO (Hashemi et al. 2011; Zeng et al. 2022).

The interobserver reliability achieved here demonstrates the value of integrating macroscopic and CT data to enhance diagnostic precision and reduce subjective bias (Coqueugniot et al. 2020).

GenAI technologies rely fundamentally on user-provided texts, and the quality and relevance of their outputs depend heavily on the specificity and clarity of the input prompts (Jarrahi et al. 2023; Robertson et al. 2024; Liu et al. 2023; Zamfirescu-Pereira et al. 2023). This concept is well documented within medical literature, which extensively examines prompt engineering to optimize AI performance in clinical contexts (Maaz et al. 2025; Patil et al. 2024; Stadler et al. 2024). In our observations, consensus among different chatbots is not always achieved, and the structure of effective prompts varies across platforms: Copilot requires a narrative description of lesions, whereas ChatGPT responds more effectively when instructed to adopt a paleopathologist's perspective, and Gemini frequently suggests Markdown-formatted prompts. Across all three systems, text-based interaction remains necessary. Moreover, each platform demonstrates difficulty interpreting images, making a detailed description of the bone, projection, and imaging type essential.

We also instructed the chatbots to remain neutral, as LLMs are known to exhibit flattery, uncritically accepting users' incorrect statements as facts. Sycophancy is widespread in state-of-the-art language models, including GPT-5, and increases with the difficulty of the problem (Naddaf 2025). In our test, ChatGPT revised its initial assessment after we asked it to consider OO, including it as "moderately likely," whereas the others included OO following the query but considered it as "very unlikely."

In assessing the spinal lesion, all three chatbots concluded that the spinal changes were typical of DISH with coexisting OA. Nevertheless, their diagnostic reasoning, emphasis, and interpretative frameworks differed. Copilot adopted a cautious, stepwise differential exclusion approach with confirmation through

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Differential diagnosis with reason performed by the three chatbots for the case of patellar lesion.

AI chatbots	Copilot	Gemini	ChatGPT
Differential diagnoses (pre-imaging)	1. CBL (highly likely). 2. LCH (moderately likely). 3. O/BA (moderately likely). 4. ABCt (less likely). 5. GC tumor (remote possibility). 6. OO (very unlikely). 7. DDP (ruled out)	1. DDP (highly likely). 2. CA (systemic) (highly likely); CBL, SCGABC (moderately likely). 3. Intraosseous ganglion cyst, O/BA (less likely). 4. LCH/eosinophilic granuloma (remote possibility). 5. OO (very unlikely)	1. O/BA (highly likely). 2. Multifocal O TB (moderately likely). 3. Benign bone cyst/developmental defect (less likely). 4. CBL/OO (remote possibility). 5. Malignant neoplasm or trauma (very unlikely/ruled out)
Differential diagnoses (post-CT imaging and after query)	1. CBL (highly likely). 2. LCH (moderately likely). 3. O/BA. 4. ABC (less likely). 5. GCT (remote possibility). 6. OO (very unlikely). 7. DDPt (ruled out)	1. Highly likely. DDP (systemic): C A (iron deficiency/thalassemia). 2. Less likely: SCG. 3. Very unlikely: OO, O/BA. 4. Ruled out: CBL, ABC, LCH	1. DDP (highly likely). 2. OO (moderately likely). 3. Brodie's abscess (less likely). 4. OCD (remote possibility). 5. O TB (very unlikely/ruled out)
Radiographic features	Slightly oval, well-circumscribed lytic lesion (~9.86 × 8.79 mm). Located on the lateral articular surface of the left patella. Internal smooth-walled cavity. Mild osteosclerotic border. No periosteal reaction or cortical expansion	Size: ≈1 cm (9.86 mm × 9.79 mm). Location: Pathognomonic on the superolateral articular (dorsal/posterior) margin. Borders: Sharp, well-defined, nonsclerotic ("punched-out"). Internal cavity: Lytic (empty) and communicates with the joint surface.	CT: Sharply demarcated subchondral lytic cavity (~9.9 × 7.9 mm); thin osteosclerotic rim; smooth internal contour; no periosteal reaction; absence of central calcified nidus; cortex continuous; articular surface penetration minimal.
Consistent characteristics	Epiphyseal-equivalent location Juvenile age (15 ± 3 years). Reactive border. Good preservation. Systemic signs (mastoid and orbital porosity) Internal cavity without aggressive destruction	DDP: Size, precise location, sharp nonsclerotic borders, and occurrence in an adolescent male. Anemia: Porotic hyperostosis and cribra orbitalia are classic indicators.	– Solitary, round/oval defect < 1 cm—Subchondral location in adolescent. – Thin rim of reactive bone (slow, chronic, or developmental process) – No evidence of aggressive destruction.
Contrasting characteristics	No nidus (excludes OO). No bilateral symmetry (excludes dorsal defect). No periosteal reaction or cloaca (limits osteomyelitis). No expansion or fluid levels (limits ABC). No calcifications or cortical thinning (limits GCT)	Neoplasms (CBL/OO): Lack of large size or the characteristic sclerotic rim. Infection (osteomyelitis): Lack of a dense, reactive sclerotic border. DDP vs. systemic: The DDP is a localized developmental defect and not related to the generalized health markers (anemia, lytic mastoid), indicating comorbidity.	DDP: Fixed superolateral facet position, thin uniform rim, no nidus. OO: Nidus with dense central calcification and marked peripheral sclerosis (absent). Brodie's abscess: Irregular margin, sclerotic shell thick, possible sinus (absent). TB: poorly defined, "moth-eaten" lysis (absent).
Inclusion/exclusion criteria	Included: Age, lesion location, morphology, systemic signs Excluded: Developmental anomalies, aggressive neoplasms, inflammatory joint disease	Inclusion (DDP): The superolateral articular margin location and the nonsclerotic, sharp lytic defect are the key inclusion criteria. Exclusion (OO/infection): The complete absence of a sclerotic rim is the definitive exclusion criterion for osteoid osteoma and chronic, encapsulated osteomyelitis.	Include DDP: Typical site + geometry + smooth sclerotic rim. Include OO: Small round lytic lesion < 1 cm + rim sclerosis. Exclude Brodie's abscess: No cloaca, uniform rim. Exclude TB: No cortical destruction, no multifocal pattern in the patella. Exclude neoplasms (chondroblastoma, giant cell): Size and border morphology are inconsistent.

(Continues)

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | (Continued)

AI chatbots	Copilot	Gemini	ChatGPT
Historical context		Chronology and location: XV Century CE, Urban Medieval Cemetery in Ferrara, N-E Italy. Significance: High prevalence of chronic infectious disease and poor sanitation (explaining the lytic mastoid and vertebral changes). Geographic context: Thalassemia (hemolytic anemia) and severe iron deficiency anemia were historically endemic to many Italian coastal and urban areas, accounting for the porotic hyperostosis.	Urban medieval environment with endemic infectious and anemic stress; chronic otitis (mastoid lesion) common; skeletal TB present but rare in juveniles; DDP and minor subchondral anomalies underreported historically. Nutritional stress could contribute to bone remodeling anomalies.
Taphonomic considerations	No root etching or erosion. Lesion formed in vivo. Well-preserved skeletal remains. No postmortem damage affecting lesion interpretation	Preservation: Good, but with some postmortem damage. Impact: The excellent preservation and use of CT allowed for accurate assessment of the lesion margins and internal structure, minimizing the effect of postmortem damage on the differential diagnosis of the patellar lesion.	Inferior patellar pole fractured postmortem (irregular edge, open trabeculae). Lesion margins display vital sclerosis → antemortem formation. Cavity infilled by sediment-burial diagenesis, not pathology.
Paleopathological comparisons	CBL documented in juvenile remains (Ortner 2019) LCH with multifocal lesions (Aufderheide et al. 1998) Patellar CBL with secondary ABC (Zheng et al. 2021) Dorsal defect as nonpathological variant (Resnick and Niwayama 1988) Rare OO cases in the patella (Greco et al. 2001)	The case is a textbook example of comorbidity, where a localized developmental variant (DDP) exists concurrently with widespread indicators of chronic environmental stress/disease (anemia/infection), a pattern typical of high-density medieval urban populations.	Teenage DDP case (Gerrie et al. 2016); 7 identical DDP lesions (Goergen et al. 1979); pictorial review of DDP (Masroori et al. 2024); no prior paleopathological patellar OO cases reported.

Abbreviations: ABC: aneurysmal bone cyst; CA: chronic anemia; CBL: chondroblastoma; DDP: dorsal defect of the patella; GCT: giant cell tumor; LCH: Langerhans cell histiocytosis; O TB: osteoarticular tuberculosis; O/BA: subacute/chronic osteomyelitis (Brodie's abscess); OCD: osteochondritis dissecans; OO: osteoid osteoma; SCG: subchondral cyst/geode.

radiographic corroboration; Gemini emphasized comorbidity and systemic burden, integrating spinal and extraspinal manifestations into a single clinical–paleopathological picture; and ChatGPT relied upon a pattern-recognition and criteria-based approach, foregrounding radiography and established diagnostic frameworks.

For the patellar lesion, their differential diagnoses again reflected distinct interpretive orientations. Copilot's interpretation was guided by the lesion's epiphyseal-equivalent position and its sharply circumscribed lytic appearance, both features sometimes seen in benign bone tumors. Gemini took a developmental and systemic interpretive stance from the outset. ChatGPT's interpretation reflected attention to the lesion's lytic nature and the contextual prevalence of infection in the medieval population. After CT evaluation and comparative review, Gemini and ChatGPT identified the lesion as highly consistent with DDP. This finding is supported by the combination of its size (< 1 cm), subchondral superolateral position, sharply defined borders, and smooth internal cavity, features regarded as pathognomonic of DDP in both clinical and paleopathological contexts. Copilot did not include DDP, also after query, and includes OO, but as very unlikely. ChatGPT includes OO as moderately likely after the query, and Gemini as very unlikely.

A comparative analysis with the anthropologist yielded several insights: (1) the anthropologist exercised greater caution when assigning specific pathologies; (2) consensus was reached in certain cases, though significant disagreement persisted in others; (3) the chatbots proposed additional diagnostic possibilities not initially considered, some of which displayed vague but noteworthy correspondence with observed lesions, thereby opening valuable avenues for further inquiry; and (4) diagnosing patellar lesions is challenging even for anthropologists due to the absence of soft tissues, and AI software appears to encounter similar difficulties.

GenAI tools, when critically employed by experienced anthropologists, may therefore offer valuable preliminary support in structuring a differential diagnosis, helping to generate possible interpretations without replacing specialist expertise.

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

	Copilot	Gemini	ChatGPT	A
DDP	0	5	5	5
CA		5		
SCG		3		
OO	1	1	4	5
O/BA	4	1	3	3
LCH	4	0		
GCT	2			3
OCD			2	
O TB			1	0
BTH				1
OB				3
OS				1
ABC	3	0		0
CBL	5	0		4

5 Highly Likely
4 Moderately Likely
3 Less Likely
2 Remote Possibility
1 Very Unlikely
0 Ruled out

FIGURE 3 | Matrix based on the likelihood of pathology showing consensus among the three chatbots and the anthropologists (A) for the patellar lesion. ABC: aneurysmal bone cyst; BTH: brown tumor of hyperparathyroidism; CA: chronic anemia; CBL: chondroblastoma; DDP: dorsal defect of the patella; GCT: giant cell tumor; LCH: Langerhans cell histiocytosis; O TB: osteoarticular tuberculosis; O/BA: subacute/chronic osteomyelitis (Brodie's abscess); OB: osteoblastoma; OCD: osteochondritis dissecans; OO: osteoid osteoma; OS: osteosarcoma; SCG: subchondral cyst/geode. [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | Limitations of the Study

This study has some limitations. First, only two cases were analyzed, limiting the generalizability of the findings; the present work should, therefore, be considered a preliminary exploration of how commercially available AI may support human anthropologists and how information is processed within these systems. Nevertheless, we have chosen two examples of different complexity to highlight the strengths and limits of the method.

Another limitation concerns the reliability of AI-generated references. Although some well-known sources were reproduced correctly, many citations contained errors in authorship or publication year, and some were entirely fabricated. We carefully checked every single citation and emphasized that LLM outputs should be cross-checked against reliable scientific databases such as PubMed, Scopus, or Web of Science (Alkaissi and McFarlane 2023; Sarode et al. 2024). Scopus AI, introduced in August 2023, reduces the risk of "hallucinations" or false information generated by AI, as it draws exclusively from the verified knowledge within the Scopus database (Scopus AI 2024). Furthermore, AI outputs depend on their training data, which can introduce bias. Current medical chatbots lack the expertise to interpret skeletal pathology, taphonomy, and archaeological context. Therefore, expanding AI in paleopathology is essential to develop specialized datasets and improve reliability. A dedicated AI tool trained on paleopathological data could enhance diagnostic accuracy and support research on ancient diseases.

Finally, ethical concerns emerge when AI-derived suggestions are misinterpreted as equivalent to expert clinical judgment. Balancing the efficiency of AI with the irreplaceable value of professional expertise remains a major challenge. Nevertheless, AI's capacity to synthesize large volumes of literature rapidly can support researchers in avoiding low-probability or unproductive investigative directions (Dave et al. 2023).

Author Contributions

Jessica Mongillo: conceptualization, methodology, data curation, investigation, formal analysis, writing – original draft, writing – review and editing. **Giulia Vescovo:** data curation, investigation, writing – review and editing. **Giulia Manzo:** data curation, investigation, formal analysis, writing – review and editing. **Nicoletta Zedda:** methodology, data curation, investigation, formal analysis, writing – review and editing, supervision. **Barbara Bramanti:** methodology, data curation, investigation, writing – review and editing, supervision.

Acknowledgments

Gratitude is extended to Pietro Marro for his valuable input on the papers argument. Open access publishing facilitated by Universita degli Studi di Ferrara, as part of the Wiley - CRUI-CARE agreement.

Funding

Open Access funding wa's enabled by the University of Ferrara CRUI agreement.

Ethics Statement

This study was conducted in accordance with relevant ethical guidelines, and all human remains were handled respectfully and responsibly.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data examined in this study are reported in the article and in the Supporting Information.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

The authors used commercial artificial intelligence tools (ChatGPT, Copilot, and Gemini) to test these tools' ability to support the differential diagnosis of paleopathologies in human skeletal remains. After using these tools, the authors clearly distinguished the artificial intelligence output and reviewed the content as necessary, taking full responsibility for the content of the published article.

References

Acsádi, G., and J. Nemeskéri. 1970. *History of Human Life Span and Mortality*. Akadémiai Kiadó.

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Alduhailan, H. W., M. A. Alshamari, and H. A. Wahsheh. 2025. "A Comprehensive Comparison and Evaluation of AI-Powered Healthcare Mobile Applications' Usability." *Healthcare* 13: 1829. https://doi.org/10.3390/healthcare13151829.

Alkaissi, H., and S. I. McFarlane. 2023. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing." *Cureus* 15: e35179. https://doi.org/10.7759/cureus.35179.

Alyas, F., S. L. James, A. M. Davies, and A. Saifuddin. 2007. "The Role of MR Imaging in the Diagnostic Characterisation of Appendicular Bone Tumours and Tumour-Like Conditions." *European Radiology* 17: 2675–2686.

Anderson, A. S., M. L. Sutherland, L. O'Donnell, et al. 2021. "Do Computed Tomography Findings Agree With Traditional Osteological Examination? The Case of Porous Cranial Lesions." *International Journal of Paleopathology* 33: 209–219. https://doi.org/10.1016/j.ijpp.2021.04.008.

Angel, J. L. 1981. "History and Development of Paleopathology." *American Journal of Physical Anthropology* 56: 509–515. https://doi.org/10.1002/ajpa.1330560425.

Appleby, J. 2022. "Differential Diagnosis and Rigor in Paleopathology." In *The Routledge Handbook of Paleopathology*, 43–63. Routledge.

Appleby, J., R. Thomas, and J. Buikstra. 2015. "Increasing Confidence in Paleopathological Diagnosis—Application of the Istanbul Terminological Framework." *International Journal of Paleopathology* 8: 19–21. https://doi.org/10.1016/j.ijpp.2014.07.003.

Atif, M., O. H. Hasan, U. Ashraf, and M. Mustafa. 2018. "Benign Tumours and Tumour-Like Lesions of Bone." *Journal of the Pakistan Medical Association* 68: 1502.

Aufderheide, A. C., C. Rodríguez-Martín, and O. Langsjoen. 1998. *The Cambridge Encyclopedia of Human Paleopathology*. Cambridge University Press.

Bayala, Y. L. T., W. J. S. Zabsonré/Tiendrebeogo, D. D. Ouedraogo, et al. 2025. "Performance of the Large Language Models in African Rheumatology: A Diagnostic Test Accuracy Study of ChatGPT-4, Gemini, Copilot, and Claude Artificial Intelligence." *BMC Rheumatology* 9: 54. https://doi.org/10.1186/s41927-025-00512-z.

Brothwell, D. R. 2012. "Tumours: Problems of Differential Diagnosis in Paleopathology." In *A Companion to Paleopathology*, edited by A. L. Grauer, 420–433. Wiley-Blackwell.

Buikstra, J. E. 2010. "Paleopathology: A Contemporary Perspective." In *A Companion to Biological Anthropology*, 395–411. Wiley-Blackwell.

Buikstra, J. E., and D. H. Ubelaker. 1994. *Standards for Data Collection from Human Skeletal Remains*. Arkansas Archaeological Survey.

Coqueugniot, H., B. Dutailly, and O. Dutour. 2020. "The Third Dimension in Palaeopathology: How Can Three-Dimensional Imaging by Computed Tomography Bring an Added Value to Retrospective Diagnosis?" *International Journal of Osteoarchaeology* 30: 538–550. https://doi.org/10.1002/oa.2881.

Currie, G., C. Singh, T. Nelson, C. Nabasenja, Y. Al-Hayek, and K. Spuur. 2023. "ChatGPT in Medical Imaging Higher Education." *Radiography (London)* 29: 792–799. https://doi.org/10.1016/j.radi.2023.05.011.

Dave, T., S. A. Athaluri, and S. Singh. 2023. "ChatGPT in Medicine: An Overview of Its Applications, Advantages, Limitations, Future Prospects, and Ethical Considerations." *Frontiers in Artificial Intelligence* 6: 1169595. https://doi.org/10.3389/frai.2023.1169595.

De Cremer, D., and D. Narayanan. 2023. "How AI Tools Can—and Cannot—Help Organizations Become More Ethical." *Frontiers in Artificial Intelligence* 6: 1093712. https://doi.org/10.3389/frai.2023.1093712.

Ezra, D., R. Mader, A. Yosef, et al. 2025. "Prevalence and Characteristics of Osseous Bridging Between Vertebral Bodies in the Cervical Spine: A Skeletal Study." *Clinical Anatomy* 38: 75–82.

Fang, Y. T., Q. Lan, T. Xie, Y. F. Liu, S. Y. Mei, and B. F. Zhu. 2020. "New Opportunities and Challenges for Forensic Medicine in the Era of Artificial Intelligence Technology." *Fa Yi Xue Za Zhi* 36: 77–85.

Forestier, J., and J. Rotes-Querol. 1950. "Senile Ankylosing Hyperostosis of the Spine." *Annals of the Rheumatic Diseases* 9: 321–330. https://doi.org/10.1136/ard.9.4.321.

Gerrie, B. J., C. L. McCrum, D. M. Knapik, and M. J. Salata. 2016. "Dorsal Defect of the Patella in a Teenage Male Football Player: Case Report and Differential Diagnosis." *Case Reports in Orthopedics* 4, no. 9: 4389549. https://doi.org/10.1155/2016/4389549.

Goergen, T. G., D. Resnick, G. Greenway, and S. L. Saltzstein. 1979. "Dorsal Defect of the Patella (DDP): A Characteristic Radiographic Lesion." *Radiology* 130: 333–336.

Grauer, A. L. 2012. *A Companion to Paleopathology*. Wiley-Blackwell.

Greco, F., F. C. Tamburrelli, G. Ciabattoni, and C. A. Logroscino. 2001. "Osteoid Osteoma of the Patella: A Case Report and Review of the Literature." *Journal of Orthopaedic Surgery* 9: 59–62.

Grootscholten, C., I. M. Bajema, S. Florquin, et al. 2008. "Interobserver Agreement of Scoring of Histopathological Characteristics and Classification of Lupus Nephritis." *Nephrology, Dialysis, Transplantation* 23: 223–230. https://doi.org/10.1093/ndt/gfm555.

Hakim, D. N., T. Pelly, M. Kulendran, and J. A. Caris. 2015. "Benign Tumours of the Bone: A Review." *Journal of Bone Oncology* 4: 37–41. https://doi.org/10.1016/j.jbo.2015.02.001.

Hashemi, J., M. Gharahdaghi, E. Ansaripour, F. Jedi, and S. Hashemi. 2011. "Radiological Features of Osteoid Osteoma: Pictorial Review." *Iranian Journal of Radiology* 8: 182–189. https://doi.org/10.5812/kmp.iranjradiol.17351065.3392.

Hosseini, M., C. A. Gao, D. M. Liebovitz, et al. 2023. "An Exploratory Survey About Using ChatGPT in Education, Healthcare, and Research." *PLoS ONE* 18: e0292216. https://doi.org/10.1371/journal.pone.0292216.

Jaffe, H. L. 1935. "Osteoid-Osteoma: A Benign Osteoblastic Tumour Composed of Osteoid and Atypical Bone." *Archives of Surgery* 31: 709–728. https://doi.org/10.1001/archsurg.1935.01180170034003.

Jarrahi, M. H., D. Askay, A. Eshraghi, and P. Smith. 2023. "Artificial Intelligence and Knowledge Management: A Partnership Between Human and AI." *Business Horizons* 66: 87–99.

Klaus, H. D. 2017. "Paleopathological Rigor and Differential Diagnosis: Case Studies Involving Terminology, Description, and Diagnostic Frameworks for Scurvy in Skeletal Remains." *International Journal of Paleopathology* 19: 96–110.

Kotěrová, A., D. Navega, M. Štepanovský, Z. Buk, J. Brůžek, and E. Cunha. 2018. "Age Estimation of Adult Human Remains From Hip Bones Using Advanced Methods." *Forensic Science International* 287: 163–175. https://doi.org/10.1016/j.forsciint.2018.03.047.

Kuperus, J. S., F. A. A. Mohamed Hoesein, P. A. de Jong, and J. J. Verlaan. 2020. "Diffuse Idiopathic Skeletal Hyperostosis: Etiology and Clinical Relevance." *Best Practice & Research: Clinical Rheumatology* 34: 101527. https://doi.org/10.1016/j.berh.2020.101527.

Lawler, D. F. 2017. "Differential Diagnosis in Archaeology." *International Journal of Paleopathology* 19: 119–123.

Lichtenstein, L. 1965. *Bone Tumors*, 104–120. C.V. Mosby Company.

Liu, P., W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig. 2023. "Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing." *ACM Computing Surveys* 55: 1–35.

Lorenzo, J., M. Horowitz, Y. Choi, H. Takayanagi, and G. Schett, eds. 2015. *Osteoimmunology: Interactions of the Immune and Skeletal Systems*. Academic Press.

Lovejoy, C. O., R. S. Meindl, T. R. Pryzbeck, and R. P. Mensforth. 1985. "Chronological Metamorphosis of the Auricular Surface of the Ilium:

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A New Method for the Determination of Adult Skeletal Age at Death." *American Journal of Physical Anthropology* 68, no. 1: 15–28. https://doi.org/10.1002/ajpa.1330680103.

Maaz, S., J. C. Palaganas, G. Palaganas, and M. Bajwa. 2025. "A Guide to Prompt Design: Foundations and Applications for Healthcare Simulationists." *Frontiers in Medicine* 11: 1504532. https://doi.org/10.3389/fmed.2024.1504532.

Masroori, Z., S. Haseli, E. Abbaspour, et al. 2024. "Patellar Non-Traumatic Pathologies: A Pictorial Review of Radiologic Findings." *Diagnostics (Basel)* 14: 2828. https://doi.org/10.3390/diagnostics14242828.

Meskó, B. 2023. "Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial." *Journal of Medical Internet Research* 25: e50638. https://doi.org/10.2196/50638.

Naddaf, M. 2025. "AI Chatbots Are Sycophants—Researchers Say It's Harming Science." *Nature* (Advance online publication) 647: 13–14. https://doi.org/10.1038/d41586-025-03390-0.

Nakhaeizadeh, S., I. E. Dror, and R. M. Morgan. 2014. "Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias." *Science & Justice* 54: 208–214. https://doi.org/10.1016/j.scijus.2013.11.003.

OpenAI. 2025. *Introducing GPT-5*. OpenAI Accessed October 28, 2025. https://openai.com/index/introducing-gpt-5/.

Ortner, D. J. 2003. *Identification of Pathological Conditions in Human Skeletal Remains*. Second ed. Academic Press.

Ortner, D. J. 2012. "Differential Diagnosis and Issues in Disease Classification." In *A Companion to Paleopathology*, edited by A. L. Grauer, 250–267. Wiley-Blackwell.

Ortner, D. J. 2019. In *Ortner's Identification of Pathological Conditions in Human Skeletal Remains*, edited by J. E. Buikstra, 3rd ed. Academic Press, an imprint of Elsevier.

Patil, R., T. F. Heston, and V. Bhuse. 2024. "Prompt Engineering in Healthcare." *Electronics* 13: 2961. https://doi.org/10.3390/electronics13152961.

Pinhasi, R., and S. Mays, eds. 2008. *Advances in Human Paleopathology*. Wiley.

Ragsdale, B. D., and L. M. Lehmer. 2012. "A Knowledge of Bone at the Cellular (Histological) Level is Essential to Paleopathology." In *A Companion to Paleopathology*, edited by A. L. Grauer, 227–249. Wiley-Blackwell.

Raisch, S., and S. Krakowski. 2021. "Artificial Intelligence and Management: The Automation–Augmentation Paradox." *Academy of Management Review* 46: 192–210.

Resnick, D., and G. Niwayama. 1976. "Radiographic and Pathologic Features of Spinal Involvement in Diffuse Idiopathic Skeletal Hyperostosis (DISH)." *Radiology* 119: 559–568. https://doi.org/10.1148/119.3.559.

Resnick, D., and G. Niwayama. 1988. *Diagnosis of Bone and Joint Disorders*. 2nd ed., Vols. 1–6. W.B. Saunders Co.

Resnick, D., S. R. Shaul, and J. M. Robins. 1975. "Diffuse Idiopathic Skeletal Hyperostosis (DISH): Forestier's Disease With Extraspinal Manifestations." *Radiology* 115: 513–524. https://doi.org/10.1148/15.3.513.

Robertson, J., C. Ferreira, E. Botha, and K. Oosthuizen. 2024. "Game Changers: A Generative AI Prompt Protocol to Enhance Human-AI Knowledge Co-Construction." *Business Horizons* 67: 499–510.

Roger, J., and T. Waldron. 1995. *A Field Guide to Joint Disease in Archaeology*. Wiley.

Rogers, J., and T. Waldron. 2001. "DISH and the Monastic Way of Life." *International Journal of Osteoarchaeology* 11: 357–365.

Rogers, T. L. 1999. "A Visual Method of Determining the Sex of Skeletal Remains Using the Distal Humerus." *Journal of Forensic Science* 44: 57–60.

Rogers, T. L. 2009. "Sex Determination of Adolescent Skeletons Using the Distal Humerus." *American Journal of Physical Anthropology* 140: 143–148. https://doi.org/10.1002/ajpa.21060.

Ruff, C. 2007. "Body Size Prediction From Juvenile Skeletal Remains." *American Journal of Physical Anthropology* 133: 698–716. https://doi.org/10.1002/ajpa.20568.

Ruff, C. B., B. M. Holt, M. Niskanen, et al. 2012. "Stature and Body Mass Estimation From Skeletal Remains in the European Holocene." *American Journal of Physical Anthropology* 148: 601–617. https://doi.org/10.1002/ajpa.22087.

Sarode, G., S. Singh, and S. C. Sarode. 2024. "Cautionary Note: Chatbots Generate Fictitious References on the Topics Related to Oral Pathology." *Oral Oncology Reports* 9: 100168. https://doi.org/10.1016/j.ajog.2023.04.004.

Schutkowski, H. 1993. "Sex Determination of Infant and Juvenile Skeletons: I. Morphognostic Features." *American Journal of Physical Anthropology* 90: 199–205. https://doi.org/10.1002/ajpa.1330900206.

Scopus AI. 2024. *Trusted Content. Powered by Responsible AI*. Elsevier Accessed January 8, 2024. https://www.elsevier.com/en-in/products/scopus/scopus-ai.

Siek, T. 2013. "The Osteological Paradox and Issues of Interpretation in Paleopathology." *Vis-à-Vis: Explorations Anthropology* 12: 1–20.

Siek, T. 2019. "A Systematic Review of Palaeoepidemiological Investigations in Neoplastic Disease." *American Journal of Physical Anthropology* 168: 228 [Unpublished review].

Singh, J., S. L. James, H. M. Kroon, K. Woertler, S. E. Anderson, and A. M. Davies. 2009. "Tumour and Tumour-Like Lesions of the Patella—A Multicentre Experience." *European Radiology* 19: 701–712.

Stadler, M., A. Horrer, and M. R. Fischer. 2024. "Crafting Medical MCQs With Generative AI: A How-To Guide on Leveraging ChatGPT." *GMS Journal for Medical Education* 41: Doc20. https://doi.org/10.3205/zma001675.

Steinbock, R. 1976. *Paleopathological Diagnosis and Interpretation: Bone Disease in Ancient Human Populations*. Charles Thomas Publisher.

Takita, H., S. Walston, H. Tatekawa, et al. 2025. "A Systematic Review and Meta-Analysis of Diagnostic Performance Comparison Between Generative AI and Physicians." *npj Digital Medicine* 8: 175.

Ubelaker, D. H. 1982. "The Development of American Paleopathology." In *A History of American Physical Anthropology, 1930–1980*, edited by F. Spencer, 1–28. Academic Press.

Ubelaker, D. H., and L. G. Grant. 1989. "Human Skeletal Remains: Preservation or Reburial?" *American Journal of Physical Anthropology* 32, no. S10: 249–287.

Uddin, M., S. U. Arfeen, F. Alanazi, S. Hussain, T. Mazhar, and M. Arafatur Rahman. 2026. "A Critical Analysis of Generative AI: Challenges, Opportunities, and Future Research Directions." *Archives of Computational Methods in Engineering* 33, no. 2: 1763–1793.

Van Holsbeeck, M., B. Vandamme, G. Marchal, M. Martens, J. Victor, and A. L. Baert. 1987. "Dorsal Defect of the Patella: Concept of its Origin and Relationship With Bipartite and Multipartite Patella." *Skeletal Radiology* 16: 304–311.

Verlaan, J. J., F. C. Oner, and G. J. Maat. 2007. "Diffuse Idiopathic Skeletal Hyperostosis in Ancient Clergymen." *European Spine Journal: Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 16: 1129–1135. https://doi.org/10.1007/s00586-007-0342-x.

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Workshop of European Anthropologists. 1980. "Recommendations for Age and Sex Diagnoses of Skeletons." *Journal of Human Evolution* 9: 517–549.

Younis, H. A., T. A. E. Eisa, M. Nasser, et al. 2024. "A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges." *Diagnostics* 14: 109. https://doi.org/10.3390/diagnostics14010109.

Zamfirescu-Pereira, J. D., R. Y. Wong, B. Hartmann, and Q. Yang. 2023. "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts." In *Proc. 2023 CHI Conference on Human Factors in Computing System*, 1–21. Association for Computing Machinery (ACM).

Zeng, H., H. He, X. Tong, Z. Wang, R. Luo, and Q. Liu. 2022. "Osteoid Osteoma of the Proximal Femur: Pitfalls in Diagnosis and Performance of Open Surgical Resection." *Frontiers in Surgery* 9: 922317. https://doi.org/10.3389/fsurg.2022.922317.

Zheng, J., N. Niu, J. Shi, et al. 2021. "Chondroblastoma of the Patella With Secondary Aneurysmal Bone Cyst, an Easily Misdiagnosed Bone Tumor: A Case Report With Literature Review." *BMC Musculoskeletal Disorders* 22: 381.

Zuckerman, M. K., B. L. Turner, and G. J. Armelagos. 2011. "Evolutionary Thought in Paleopathology and the Rise of the Biocultural Approach." In *A Companion to Paleopathology*, edited by A. L. Grauer, 34–57. Wiley-Blackwell.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:1** Approach used for differential diagnoses of DISH. **Table S1:2** Approach used for differential diagnosis of patellar lesion (Osteoid Osteoma and Dorsal Defect of Patella-DDP). **Appendix S1:** Optimal prompts for AI chatbots used in this study. **Appendix S2:1** References and sitography used by each chatbot to make a differential diagnosis of DISH. **Appendix S2:2** References and sitography used by each chatbot to make a differential diagnosis of patellar lesion. **Figure S1:1** Three perspectives of spinal defect. **Figure S1:2** RX of spinal defect. **Figure S2:1** Dorsal and ventral view of patellar bone. **Figure S2:2** CT scan of the patella with measurements. **Figure S2:3** Three view of CT scan of patellar bone.

10991212, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oa.70129 by CochraneItalia, Wiley Online Library on [15/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License