



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

FRAC-MAS: A Safe and Explainable Multi-Agent System for Fracture Diagnosis

Hardik Iyer^{1*}, Tirath Bhathawala², Mihir Panchal³, Ying-Jung Chen⁴, Kiran Bhowmick¹, Pankaj Sonawane¹, and Meera Narvekar¹

¹ Dwarkadas J Sanghvi College of Engineering, Mumbai, India

² University of Amsterdam, Amsterdam, Netherlands

³ National University of Singapore, Singapore

⁴ Georgia Institute of Technology, Atlanta, USA

hardikiyer17@gmail.com, tirath.bhathawala@student.uva.nl,
mihir@comp.nus.edu.sg, yingjungcd@gmail.com, kiran.bhowmick@djsce.ac.in,
pankaj.sonawane@djsce.ac.in, meera.narvekar@djsce.ac.in

Abstract. Fracture detection and its clinical interpretability see notable improvements when deep vision models are integrated with agentic AI architectures. While deep learning models achieve high diagnostic performance, their black-box nature limits clinical adoption. We propose FRAC-MAS, an agentic AI system for automated, explainable, and safe bone fracture detection. The framework combines a stacked ensemble of four vision models with conformal prediction to produce statistically grounded differential diagnoses, while a multi-agent workflow performs independent verification, retrieves clinical guidelines, and generates patient-friendly reports. A pipeline-depth ablation study confirms that our multi-agent critic triages 86.6% of cases into a high-confidence auto-confirmed cohort while escalating uncertain cases, outperforming a single-agent baseline. Patient preference studies against Llama, MedGemma, and Gemini further demonstrate significantly more comprehensible clinical reports. These results suggest that integrating multi-agent critics with conformal guarantees enables safer radiology triage while preserving clinician oversight. More broadly, FRAC-MAS demonstrates how cooperative agentic architectures can serve as auditable, human-in-the-loop decision support systems for safety-critical healthcare. Our code is available at <https://github.com/hardik1712/FRAC-MAS>, and the website is available at <https://frac-mas.vercel.app>.

Keywords: Agentic AI · Multi-Agent Systems · Medical Imaging · Explainable AI · Clinical Decision-Making

1 Introduction

Musculoskeletal injuries, especially bone fractures, represent a substantial proportion of cases in emergency and orthopedic departments [26]. Interpreting plain radiographs is essential for accurate fracture diagnosis and management,

yet it remains challenging due to subtle fracture morphologies, inter-observer variability, and the high volume of cases seen in clinical practice [18]. Undetected fractures frequently progress to malunion, avascular necrosis, and permanent disability, while missed diagnoses drive avoidable surgeries, prolonged hospitalizations, and significant economic losses [5].

Studies report that about 2–3% of all orthopedic injuries are missed at first presentation in trauma and emergency settings, and plain radiographs may not indicate the presence of fractures in about 10–20% of some skull and skeletal injuries [7]. Therefore, a non-trivial proportion of fractures remains under-detected at first contact [13]. Explainable fracture-detection systems are increasingly being used to highlight the specific image regions that drive a fracture prediction [24]. This allows clinicians to verify whether the model is focusing on plausible features such as cortical discontinuity or trabecular collapse rather than irrelevant artifacts, improving trust and supporting safer adoption of AI-assisted diagnosis [11].

Addressing these gaps, we propose FRAC-MAS, a multi-agent AI system for automated bone fracture detection with clinical explainability. The system emphasizes human-AI interaction by validating outputs with practicing orthopedic surgeons, since human oversight is critical in clinical AI pipelines to mitigate automation bias, ensure alignment with clinical values, and improve decision accuracy where AI alone cannot reliably handle ethical ambiguity or edge cases [16]. Our multi-agent architecture bridges the gap between black-box deep learning predictions and clinical trust by decomposing the diagnostic pipeline into specialized, interpretable components, each independently verifiable and collectively designed for real-world deployment.

This shift toward modular, specialized AI pipelines aligns with recent literature investigating targeted deep learning applications for fracture detection. Joonho Oh et al. [15] demonstrated potential leaps in streamlining training processes and augmenting diagnostic precision in fracture detection, particularly focusing on X-ray images of the wrist bone using attention modules [25]. RAD-DINO has been established as a strong self-supervised vision backbone for medical imaging and has been systematically benchmarked across multiple radiology tasks, including musculoskeletal X-ray classification [19]. Furthermore, hybrid CNN-transformer and multi-stream architectures are necessary to capture both fine-grained local features and global contextual structure across heterogeneous radiographic views [9,23]. However, validating language model outputs generated from these visual features remains an open challenge; studies like Chung et al. [3] highlight the need for verification and LLM-as-a-judge protocols to validate documentation generated by clinical LLMs. Agentic extensions of these pipelines to real-world clinical deployment further validate retrieval-grounded, orchestrated architectures for trustworthy diagnostic AI [14].

Post-hoc interpretability methods are now integral to clinically deployable medical imaging AI, with Grad-CAM serving as the dominant visualization paradigm [21]. Successfully applied to pelvic fracture detection [11], these developments establish that spatial explainability is essential for regulatory compli-

ance and building clinician trust. Multi-agent systems (MAS) have demonstrated measurable superiority over single-agent baselines across clinical benchmarks. For example, Ying-Jung Chen et al. [2] showed MAS achieves 59% mortality prediction accuracy versus 56% for single-agent systems, with mean length-of-stay error reduced from 5.82 to 4.37 days [2,12]. These results collectively motivate agent-orchestrated architectures wherein specialized roles including retrieval, reasoning, and critique are decomposed across coordinated agents to achieve diagnostic performance unattainable by monolithic models.

Briefly, the major contributions of this study are as follows. First, we design a multi-agent clinical decision support workflow, where a Patient Interaction Agent orchestrates Knowledge, Critic, and Educational agents to ground evidence, perform blind verification with human escalation, and generate patient-facing explanations with Grad-CAM overlays. Second, we introduce an ensemble with conformal prediction that resolves commonly confused fracture categories via stacking and converts outputs into differential diagnoses with explicit empirical coverage safety. Finally, we rigorously validate the agent layer via a novel pipeline-depth ablation study, proving that our multi-agent architecture successfully triages uncertain cases, and establish its patient-communication superiority via a preference survey against leading LLMs (Llama, MedGemma, Gemini).

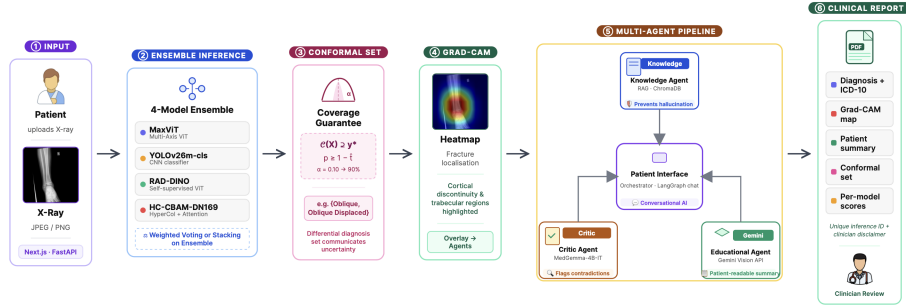


Fig. 1. End-to-end system architecture: integrating ensemble inference, Grad-CAM localization, multi-agent reasoning, and conformal prediction for verifiable clinical report generation.

2 Methodology

2.1 Dataset and Models Ensemble

The system is trained on an augmented version of the HBF MID dataset [17]. We utilized this dataset because it categorizes radiographs by precise fracture morphology rather than mere anatomical location. We balanced and augmented the training set to 1,952 images across 8 categories, while maintaining unaugmented validation (106 images) and test (112 images) splits.

Our system employs an ensemble of 4 models to capture diverse fracture appearances: MaxViT [22], YOLOv26m-cls [20], RAD-DINO, and a custom HyperColumn-CBAM DenseNet-169 [8,10]. This combines multi-axis attention, self-supervised radiological domain features, and low-level edge detail preservation. Predictions are combined through a two-pass weighted soft voting scheme or a stacking meta-learner (logistic regression) to optimally disambiguate morphologically similar categories like Oblique and Transverse fractures.

2.2 Conformal Prediction

To complement the ensemble output with a rigorous measure of uncertainty, the system wraps the ensemble output in a split conformal prediction set with a distribution-free empirical coverage guarantee [1], calibrated on the validation set. Given a miscoverage level α (e.g., $\alpha = 0.10$ for 90% coverage), every class whose probability exceeds a calibrated threshold enters the prediction set. This maps naturally to a differential diagnosis, communicating empirical uncertainty directly to downstream agents.

2.3 Multi-Agent System

To make the prediction pipeline auditable end-to-end, we route every case through a four-agent workflow (see Fig. 1). By decomposing the diagnostic process into specialized roles, the system prevents monolithic LLMs from hallucinating medical advice and forces all outputs to be explicitly grounded and independently verified.

Patient Interface Agent. The Patient Interface Agent acts as the entry point and primary orchestrator. It manages the conversational memory and evaluates the patient’s context (e.g., age, reported pain levels). It utilizes a deterministic LangGraph state machine to route interactions: simple administrative queries are handled directly, while clinical assessments of uploaded radiographs invoke the downstream pipeline. This architecture ensures that diagnostic responses are generated strictly through the verified pipeline rather than through unbounded, parametric LLM generation.

Knowledge Agent. The Knowledge Agent is responsible for anchoring the narrative to verified clinical sources, acting as an advanced Retrieval-Augmented Generation (RAG) module. It queries a locally hosted, curated ChromaDB vector store that contains structured clinical records for all 8 fracture classes. These records include standardized ICD-10 codes, expected recovery timelines, severity ratings, and official treatment guidelines from established orthopedic bodies (e.g., AO/OTA fracture classifications). For each radiograph case, the agent uses sentence-transformer embeddings to perform a dense semantic search. The retrieved top-k guidelines are then supplied as the absolute, non-negotiable context to a fast language model (*gemini-2.5-flash-lite* [4]), which synthesizes a focused clinical background that strictly avoids hallucinating non-retrieved treatments.

Critic Agent. Operating as the core safety mechanism, the Critic Agent prevents automation bias by performing a blind-first triage. Upon receiving an image, it produces an independent zero-shot diagnostic assessment *before* seeing the ensemble’s provisional label, preventing anchoring bias. It then evaluates the ensemble’s output and issues a verdict $\in \{\text{‘yes’}, \text{‘no’}, \text{‘uncertain’}\}$. A rigorous hybrid consensus rule forces human escalation under three conditions: (1) if the Critic explicitly rejects the ensemble’s diagnosis, (2) if the Critic’s independent top prediction differs from the ensemble with a confidence greater than a preset safety threshold (default 0.6), or (3) if the conformal prediction set contains multiple classes, indicating intrinsic model ambiguity. This ensures the clinician remains the final arbiter for all difficult cases.

Educational Agent. Finally, the Educational Agent is tasked with patient communication. It receives the clinically verified diagnosis, the retrieved context from the Knowledge Agent, and the Grad-CAM [21] heatmap overlays generated by the vision ensemble. Using this rich context, it prompts a highly capable reasoning model (*gemini-2.5-pro* [6]) to generate a lay-friendly summary. The agent explicitly highlights the anatomical regions activated in the Grad-CAM heatmap to explain *why* the diagnosis was made, translating complex radiographic findings into accessible language while providing clear next-steps guidance (e.g., immediate emergency room visits for displaced fractures versus scheduled orthopedic follow-ups for non-displaced injuries).

3 Results and Analysis

3.1 Pipeline Agent Ablation

To quantify the marginal contribution of each architectural layer, we evaluated five incrementally richer arms of FRAC-MAS on the held-out test set (seed 42, $\alpha=0.10$, stacking meta-learner enabled). As shown in Table 1, accuracy is invariant across pipeline depth because downstream components do not alter the raw argmax prediction. The single-agent baseline escalates every test case due to raw confidence thresholds, providing no practical automation benefit. In contrast, the multi-agent Critic selectively auto-confirms a large majority of cases, reducing unnecessary escalations significantly. The confirmed cohort achieves a strong safety margin over the raw ensemble baseline, demonstrating that the Critic successfully filters erroneous predictions. Finally, the full pipeline achieves a complete task-completion rate for verifiable, retrieval-grounded generation.

3.2 External Dataset Evaluation

To assess generalizability beyond HBFMID, we evaluated binary fracture detection on two external datasets that the models never encountered during training. As detailed in Table 2, the stacking meta-learner robustly identifies unseen fracture types on the 10-class Roboflow dataset, demonstrating strong morphological transferability. Conversely, evaluating on FracAtlas provides a stress test for

Table 1. Pipeline-depth ablation on the 112-sample test set ($\alpha=0.10$, stacking). *Conf.* *Acc.* is accuracy on cases not flagged for human escalation. [†]The Critic uses local MedGemma-4B-IT.

Configuration	Acc. (%)	Conf. Acc. (%)	Escalation Rate (%)	Set Size	Coverage (%)	Know/Educ (%)
Single-agent baseline	91.1	—	100.0	—	—	—
Ensemble only	91.1	—	0.0	—	—	—
+Conformal ($\alpha=0.10$)	91.1	—	0.0	1.08	92.0	—
+Conformal + Critic	91.1	93.8	13.4	1.08	92.0	—
Full pipeline	91.1	93.8	13.4	1.08	92.0	100.0

out-of-distribution (OOD) performance. The optimal threshold yielded a modest accuracy (AUC 0.652), confirming that features learned from wrist/hand X-rays do not transfer robustly to full-body radiographs (e.g., spine, pelvis). Crucially, the multi-agent Critic is designed precisely to intercept such OOD uncertainties, flagging anomalous inputs via low ensemble confidence and conformal ambiguity for human escalation, thus maintaining system safety despite backbone limitations.

Table 2. Binary fracture detection on the Roboflow dataset (140 images, all positive). Det. Rate = recall; $\bar{s}$ = mean fracture score.

Configuration	Det. Rate	F1	$\bar{s}$
MaxViT	0.950	0.974	0.947
HC-CBAM-DenseNet169	0.900	0.947	0.903
RAD-DINO	0.914	0.955	0.901
YOLOv26m-cls	0.921	0.959	0.927
Ensemble (Wt. Avg)	0.929	0.963	0.920
Ensemble (Stacking)	0.964	0.982	0.916

3.3 Conformal Prediction

Conformal prediction converts the hardest classes into useful differentials. As illustrated in Table 3, the procedure achieves strong overall empirical coverage. The largest coverage gains concentrate among morphologically similar and highly error-prone categories (Oblique, Transverse). Even when the argmax prediction is incorrect, the conformal set often includes the true label, converting a hard error into a clinically actionable differential diagnosis.

3.4 Grad-CAM Analysis

Different vision backbones attend to complementary evidence, explaining the ensemble’s overall gain. As quantified in Table 4, MaxViT consistently produces sparse attention maps that isolate the fracture line or cortical discontinuity. In

Table 3. Per-class ensemble accuracy and conformal coverage on the test set. Coverage failures concentrate among morphologically similar types.

Class	N	Acc. (%)	Coverage (%)	
			$\alpha=0.05$	$\alpha=0.10$
Comminuted (COM)	17	100.0	100.0	100.0
Greenstick (GRN)	13	100.0	100.0	100.0
Healthy (HLT)	10	100.0	100.0	100.0
Oblique (OBL)	17	70.6	76.5	82.4
Oblique Displaced (OBL-D)	9	100.0	100.0	100.0
Spiral (SPR)	12	100.0	100.0	100.0
Transverse (TRV)	17	70.6	70.6	70.6
Trans. Displaced (TRV-D)	17	88.2	88.2	94.1
Overall	112	89.3	90.2	92.0

Table 4. Grad-CAM attention statistics. Active fraction is the proportion of pixels exceeding 0.3 activation. MaxViT attends to 2%-21% of an image while the HyperColumn model covers 53%-67%.

Fracture	Active (%)		Mean Act.	
	MaxViT	HC	MaxViT	HC
Comminuted	2.4	59.8	0.04	0.55
Oblique Disp.	7.4	53.0	0.23	0.52
Spiral	7.4	66.8	0.14	0.61
Transverse	20.6	58.5	0.27	0.57

contrast, the HyperColumn-CBAM model distributes attention broadly, capturing the surrounding structural context. This complementary diagnostic strategy allows the downstream Educational Agent to ground its explanations on both localized and global features.

3.5 Human Validation

Clinician Reader Study. Three orthopedic clinicians provided a blind diagnosis based solely on the radiograph before reviewing the system-generated output. As summarized in Table 5, the agent demonstrated strong alignment with clinical standards and substantial inter-rater reliability. This confirms that the multi-agent pipeline effectively bridges raw classification outputs to highly accurate and comprehensible explanations for human review.

Patient Preference Survey. To evaluate utility from the patient’s perspective, we conducted a robust preference survey on 35 distinct radiographic images, uniformly spanning all 7 fracture types and the healthy baseline. Human evaluators blindly reviewed reports generated by FRAC-MAS alongside outputs

Table 5. Clinician ratings of the system’s educational outputs (1–5 scale) and inter-rater agreement across three orthopedic raters.

Dimension	Mean Score	Score ≥ 4 (%)	Fleiss’ κ
Technical Accuracy	4.14	75.0	0.72
Comprehensibility	3.95	68.2	0.65

from established large language model baselines (Llama, MedGemma, Gemini) that were supplied with the exact same diagnostic inputs. Evaluators ranked the responses based on clarity, empathy, and perceived clinical helpfulness. Our proposed model significantly outperformed the generic baselines, achieving a top win rate of 55.4% and establishing the best average rank of 1.77. A non-parametric repeated-measures Friedman test ($\chi^2(3) = 135.4371, p < 0.001$) and subsequent post-hoc pairwise Wilcoxon signed-rank tests unequivocally validated this statistically superior performance across the dataset. Ultimately, the structured, multi-agent orchestration, which explicitly blends Grad-CAM localization with RAG-retrieved clinical severity, creates a definitive advantage in generating concise, comprehensible, and patient-preferred outputs.

4 Limitations and Conclusion

We presented a modular multi-agent framework that combines an ensemble of diverse vision backbones with Grad-CAM explainability, retrieval-anchored clinical context, a VLM-based critic, and conformal prediction to produce human-verifiable orthopedic diagnoses. Our experiments demonstrate strong single-model and ensemble performance, achieving a 96.4% stacking detection rate on an anatomically matched external set, while conformal calibration yields meaningful coverage gains (92.0% empirical coverage at $\alpha = 0.10$) for morphologically ambiguous subtypes where differentials are clinically valuable. Crucially, our pipeline-depth ablation demonstrated that our multi-agent critic successfully triages cases, offering a +2.7 percentage point safety margin on confirmed cases while reducing escalations $7.5\times$ relative to a monolithic agent baseline. Furthermore, human validation via both clinician reader studies and patient preference surveys proved our system’s outputs are significantly preferred over single-model baselines like Llama and Gemini, confirming high clinical accuracy (mean 4.14/5) and comprehensibility. While our evaluation surfaces practical limitations requiring future work, namely finite-sample constraints due to the small calibration pool, domain-shift brittleness on FracAtlas, and potential VLM confirmation bias, addressing these through multi-anatomy training and explicit domain adaptation provides a clear developmental roadmap. Ultimately, FRAC-MAS demonstrates a viable path toward auditable, patient-friendly AI assistance for fracture diagnosis while preserving the clinician as the final arbiter through intelligent triage.

References

1. Angelopoulos, A.N., Bates, S.: A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *Foundations and Trends in Machine Learning* **16**(4), 494–591 (2023). <https://doi.org/10.1561/2200000101> 4
2. Chen, Y.J., Albarqawi, A., Chen, C.S.: Enhancing clinical decision-making: Integrating multi-agent systems with ethical ai governance. In: 2025 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–7. IEEE (2025) 3
3. Chung, P., Swaminathan, A., Goodell, A.J., Kim, Y., Momsen Reincke, S., Han, L., Deverett, B., Sadeghi, M.A., Ariss, A.B., Ghanem, M., et al.: Verifying facts in patient care documents generated by large language models using electronic health records. *NEJM AI* **3**(1), AIdbp2500418 (2025) 2
4. Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al.: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:2507.06261 (2025) 4
5. Flores, M.J., Brown, K.E., O’Marr, J.M., Adejuyigbe, B., Rodarte, P., Gomez-Alvarado, F., Nwachuku, K., Urva, M., Shearer, D.: The economic impact of infection and/or nonunion on long-bone shaft fractures: a systematic review. *OTA International* **7**(3), e337 (2024) 2
6. Gemini Team, Google: Gemini: A family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2023), <https://arxiv.org/abs/2312.11805> 5
7. Gupta, S., Kapoor, P.S., Ghosh, P., Cr, V.: Patterns of musculoskeletal injuries diagnosed through x-ray and mri in trauma patients in varanasi, uttar pradesh. *Journal of Heart Valve Disease* **29**, 68–75 (2024) 2
8. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 447–456 (2015). <https://doi.org/10.1109/CVPR.2015.7298642> 4
9. Hassan, A., Munib, N., Batool, A., Noor, H., et al.: Fracture detection in x-rays using custom convolutional neural network (cnn) and transfer learning models. arXiv preprint arXiv:2509.06228 (2025) 2
10. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2261–2269 (2017). <https://doi.org/10.1109/CVPR.2017.243> 4
11. Kassem, M.A., Naguib, S.M., Hamza, H.M., Fouda, M.M., Saleh, M.K., Hosny, K.M.: Explainable transfer learning-based deep learning model for pelvis fracture detection. *International Journal of Intelligent Systems* **2023**(1), 3281998 (2023) 2
12. Kim, Y., Park, C., Jeong, H., Chan, Y.S., Xu, X., McDuff, D., Lee, H., Ghassemi, M., Breazeal, C., Park, H.W.: Mdagents: An adaptive collaboration of llms for medical decision-making. *Advances in Neural Information Processing Systems* **37**, 79410–79452 (2024) 3
13. Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., et al.: Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences* **115**(45), 11591–11596 (2018) 2
14. Oettl, F.C., Pruneski, J., Zsidai, B., Yu, Y., Cong, T., Feldt, R., Winkler, P.W., Hirschmann, M.T., Samuelsson, K., Group, E.A.I.W.: Artificial intelligence agents

- in orthopaedics: Concepts, capabilities and the road ahead. *Knee Surgery, Sports Traumatology, Arthroscopy* **33**(12), 4475–4483 (2025) [2](#)
15. Oh, J., Hwang, S., Lee, J.: Enhancing x-ray-based wrist fracture diagnosis using hypercolumn-convolutional block attention module. *Diagnostics* **13**(18), 2927 (2023) [2](#)
 16. Olawade, D.B., Plabon, S.B., Ojo, A., Ogunbona, M.A., Makanjuola, B.D., Olasilola, R.: Human in the loop artificial intelligence in healthcare: applications, outcomes, and implementation challenges. *International Journal of Medical Informatics* p. 106362 (2026) [2](#)
 17. Parvin, S., Rahman, A.: A real-time human bone fracture detection and classification from multi-modal images using deep learning technique. *Applied Intelligence* **54**(19), 9269–9285 (2024). <https://doi.org/10.1007/s10489-024-05588-7> [3](#)
 18. Pinto, A., Berritto, D., Russo, A., Riccitiello, F., Caruso, M., Belfiore, M.P., Papapietro, V.R., Carotti, M., Pinto, F., Giovagnoni, A., et al.: Traumatic fractures in adults: missed diagnosis on plain radiographs in the emergency department. *Acta Bio Medica: Atenei Parmensis* **89**(Suppl 1), 111 (2018) [2](#)
 19. Pérez-García, F., Sharma, H., Bond-Taylor, S., Bouzid, K., Salvatelli, V., Ilse, M., Bannur, S., Castro, D.C., Schwaighofer, A., Lungren, M.P., Wetscherek, M.T., Codella, N., Hyland, S.L., Alvarez-Valle, J., Oktay, O.: Exploring scalable medical image encoders beyond text supervision. *Nature Machine Intelligence* (2025). <https://doi.org/10.1038/s42256-024-00965-w> [2](#)
 20. Sapkota, R., Cheppally, R.H., Sharda, A., Karkee, M.: YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection. *arXiv preprint arXiv:2509.25164* (2025), <https://arxiv.org/abs/2509.25164> [4](#)
 21. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. pp. 618–626 (2017). <https://doi.org/10.1109/ICCV.2017.74> [2](#), [5](#)
 22. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.: MaxViT: Multi-axis vision transformer. In: *Computer Vision – ECCV 2022. Lecture Notes in Computer Science*, vol. 13684, pp. 459–479. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20053-3_27 [4](#)
 23. Uddin, A., Chowdhury, R., Saha, A., Hosen, M.H., Roy, P.S., Uddin, M.N.: Mobilevit and yolov8: Improving bone fracture detection and classification through deep learning. In: *2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)*. pp. 212–217. IEEE (2024) [2](#)
 24. Van der Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A.: Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical image analysis* **79**, 102470 (2022) [2](#)
 25. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*. pp. 3–19 (2018) [2](#)
 26. Yan, J., Li, F., Zhou, J., Ding, Y., Qin, Q., Jin, C.: The global burden of fractures and its underlying etiologies: results from and further analysis of the global burden of disease study 2021. *Archives of osteoporosis* **20**(1), 111 (2025) [1](#)