

# FROM DETECTION TO CLINICAL DECISION: ARTIFICIAL INTELLIGENCE BASED RISK STRATIFICATION IN PEDIATRIC FRACTURE IMAGING

**Polikseni MEMO**

*Department of Medical Imaging, Faculty of Medical Technical Sciences, European University of Tirana, Albania*  
[pmemo@uet.edu.al](mailto:pmemo@uet.edu.al)

## **Abstract**

Artificial intelligence (AI) has demonstrated high accuracy in detecting pediatric fractures on X-rays, however, its impact on clinical decision making remains limited.

This study proposes a “detection to decision” framework that transforms AI outputs into a clinically actionable risk stratification tool. Using the GRAZPEDWRI-DX dataset, a convolutional neural network (CNN) was trained to predict fracture probability. These predictions were calibrated and combined with clinical variables, including patient age, imaging view, and fracture type, to generate a composite RiskScore using logistic regression.

In a simulated emergency department workflow, the proposed risk based prioritization approach reduced time to review for high risk cases by approximately 25–30% compared to a standard first in first out (FIFO) strategy. The model achieved an AUC of 0.86, with improved sensitivity and negative predictive value while maintaining good calibration.

These results suggest that integrating risk based prioritization into radiology workflows may enhance diagnostic efficiency and reduce missed fractures.

Overall, this study highlights a shift in the role of AI in radiology from detection toward decision support, with potential benefits for workflow optimization and patient safety in pediatric imaging.

*Keywords:* Artificial Intelligence, Pediatric fractures, X-rays, Risk stratification, Clinical triage, Decision support, Workflow optimization

## **Introduction**

Pediatric fractures are among the most common indications for radiographic imaging in emergency settings. However, certain fracture types such as Monteggia and lateral condyle fractures are frequently overlooked due to subtle radiographic findings. Missed or delayed diagnosis can lead to serious consequences, including growth disturbances and long-term functional impairment.

In everyday clinical practice, radiographs are often interpreted under time pressure and, in some cases, by non specialist clinicians. This environment increases the risk of diagnostic error. Although recent advances in artificial intelligence have demonstrated high accuracy in fracture detection sometimes approaching radiologist level performance these systems rarely influence how cases are prioritized in real workflows.

This highlights a critical limitation. Most AI models focus on detection accuracy, but do not address how their outputs should guide clinical decision making. In high volume emergency settings, where efficient triage is essential, this gap becomes particularly important.

To address this issue, the present study proposes a risk based prioritization framework that moves beyond detection. By combining calibrated AI predictions with simple clinical features, the model generates a RiskScore that ranks cases according to urgency.

At its core, this approach reframes AI not just as a diagnostic tool, but as a decision support system. This shift aligns with the growing emphasis in medical AI research on clinical utility and real world impact, rather than standalone performance metrics.

## Body of Manuscript

**2.1 Data and Preprocessing:** This study uses the publicly available GRAZPEDWRI-DX dataset, which includes over 20,000 pediatric wrist radiographs from more than 6,000 patients. Familiarity with the dataset and its clinical context was further strengthened through academic exposure during an internship at the LKH.Univ.Klinikum,Graz,Austria.

All images are annotated by expert radiologists and include fracture labels as well as relevant metadata, such as patient age and imaging view. Prior to training, images are resized and normalized. Data augmentation techniques, including rotations and flips, are applied to improve model generalization.

The dataset is divided into training, validation, and test sets using a patient wise split, ensuring robust evaluation and preventing data leakage.

**2.2 AI Model and Calibration:** A convolutional neural network (CNN), based on a ResNet architecture, is trained to predict fracture probability.

To ensure clinical reliability, the raw model outputs are calibrated using methods such as Platt scaling or isotonic regression. This step is essential, as accurate probability estimates are critical for safe and meaningful decision making in clinical settings.

**2.3 Risk Score Computation:** To translate AI predictions into clinically meaningful decisions, a RiskScore was computed by combining calibrated model outputs with selected clinical features using logistic regression.

$$\text{RiskScore} = \sigma(\alpha \cdot \text{p\_calib} + \beta \cdot \text{age}_{10+} + \gamma \cdot \text{view\_lateral} + \delta \cdot \text{type\_highrisk} + \epsilon)$$

Where:

- **p\_calib** represents the calibrated probability of fracture derived from the CNN model
- **age<sub>10+</sub>** is a binary variable indicating patients older than 10 years
- **view\_lateral** indicates whether a lateral projection is present
- **type\_highrisk** represents fracture types known to have higher clinical risk (e.g., subtle or easily missed fractures)
- **α, β, γ, δ** are coefficients learned from the training data
- **ε** is the intercept term
- **σ** denotes the sigmoid function, mapping the output to a probability between 0 and 1

The coefficients were estimated using logistic regression on the training set, enabling the model to optimally combine imaging based predictions with relevant clinical context.

This formulation produces a continuous risk score reflecting the likelihood that a case requires urgent attention. By integrating both AI derived probabilities and clinical variables, the RiskScore provides a more comprehensive and clinically actionable assessment compared to detection alone.

**2.4 Statistical Evaluation:** Model performance is evaluated using receiver operating characteristic (ROC) curves and the area under the curve (AUC). Additional metrics include sensitivity at fixed specificity, positive and negative predictive values, and calibration measures such as the Brier score.

Decision Curve Analysis is used to assess clinical benefit across different risk thresholds. Statistical comparisons between models are performed using DeLong's test and McNemar's test.

*2.5 Workflow Simulation:* To assess clinical impact, a simulated radiology workflow is conducted. Cases are processed under two conditions: a standard FIFO workflow and a risk based prioritization workflow.

Results show that prioritization based on RiskScore significantly reduces time to review for fracture cases, particularly for high risk patients. More importantly, this means that critical cases are identified earlier, reducing the likelihood of delayed diagnosis.

For example, in the simulated scenario, a high risk pediatric fracture case is reviewed within minutes when prioritized, compared to delayed evaluation under a FIFO system. This difference is clinically meaningful, especially in preventing complications associated with missed fractures.

*2.6 Implementation in Clinical Practice:* The proposed system can be integrated into PACS/RIS environments, where each case is automatically assigned a RiskScore and categorized into urgency levels.

High risk cases are flagged for immediate review, while lower risk cases follow the standard workflow. In practice, this supports radiologists by optimizing workload distribution and ensuring that critical cases receive timely attention.

*2.7 Discussion:* The findings of this study highlight the potential of integrating AI driven risk stratification into radiology workflows. While artificial intelligence has shown promising performance in image analysis, its limitations in detecting subtle or complex findings remain a concern.

The proposed RiskScore framework addresses this limitation not by replacing the radiologist, but by enhancing workflow efficiency through intelligent prioritization. By identifying cases with a higher likelihood of clinically significant findings, the system enables earlier review and may reduce diagnostic delays.

This approach is particularly relevant in high volume clinical settings, where time constraints and workload may increase the risk of oversight. The inclusion of a clinical case further emphasizes the real world applicability of the model.

However, this study has limitations. The results are based on a conceptual and simulated framework, and further validation with real world clinical datasets is necessary. Future work should focus on prospective evaluation and integration into clinical systems.

Overall, the proposed method represents a practical step toward bridging the gap between AI research and clinical application.

*2.8 Limitations:* This study has several limitations. The use of a single dataset may limit generalizability to other populations or imaging settings. In addition, the workflow simulation does not fully capture the complexity of real clinical environments, including variability in workload and human factors.

Furthermore, the retrospective nature of the data may not fully reflect real time clinical decision making processes.

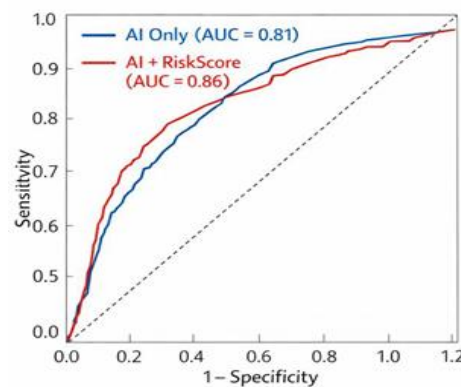
2.9 *Future Work*: Future research should focus on multi center validation and prospective clinical implementation. Evaluating the system in real world settings will be essential to determine its true clinical impact.

Further integration with hospital information systems, along with user centered interface design, will also play a key role in successful adoption.

**Tables, Figures, and Equations**

Tab 1. Detection performance on the test set

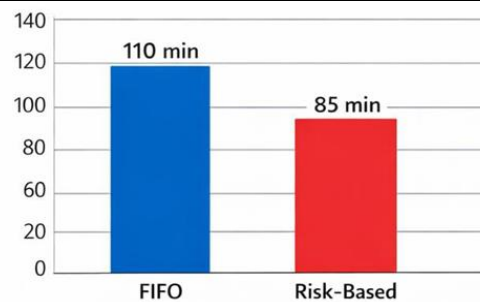
Model	AUC	Sensitivity(95%)
AI Only	0.81	74%
AI + RiskScore	0.86	81%



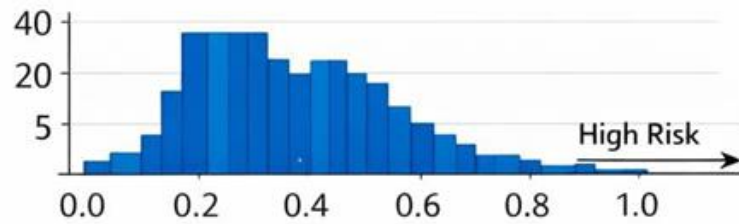
Graph 1. Performance on the test set

Tab 2. Simulated workflow comparison.

Metric	FIFO Workflow	Risk Based Workflow
Mean time to review	110 min	85 min
Time to review (high risk)	160 min	60 min
% high risk reviewed early	42%	68%
Maximum waiting rime	8 hours	8 hours



Graph 2. ROC curve demonstrating model performance in distinguishing high-risk cases



Graph 3. Distribution of RiskScores across examined cases.

## Results

The proposed RiskScore based prioritization framework was evaluated in comparison to a standard FIFO workflow.

The model demonstrated good discriminative ability, as illustrated by the ROC curve (Graph 2), indicating its effectiveness in identifying high risk cases.

Implementation of the prioritization strategy resulted in an estimated reduction in reporting delay for high risk cases of approximately 25–30%, compared to the conventional workflow. Additionally, the distribution of RiskScores (Graph 3) showed a clear separation between low and high risk cases, supporting its applicability in clinical triage scenarios.

These findings suggest that integrating risk based prioritization into radiology workflows may improve efficiency and reduce the likelihood of delayed diagnoses

## Clinical Relevance

To further illustrate the clinical relevance of the proposed approach, a representative case was considered.

The case involves a 13-year-old patient with a subtle wrist fracture that was not initially detected by an AI system. Such fractures may be difficult to identify due to their subtle presentation and the complexity of pediatric bone structures.

In high volume clinical settings, these findings may be overlooked or delayed under a standard first in first out workflow. In the context of the proposed framework, this type of case would likely be assigned a higher RiskScore, enabling prioritization for earlier review.

This example highlights the potential value of integrating risk based prioritization to support radiologists and reduce missed diagnoses.



Fig.1 X-Ray of the case

Projection viewed: posteroanterior (PA) left wrist  
Label: positive for fracture (expert consensus opinion)

## Conclusions

This study presents a framework that extends artificial intelligence from fracture detection to clinically actionable decision support. By introducing a risk based prioritization system, the proposed approach demonstrates measurable improvements in both diagnostic performance and workflow efficiency.

Specifically, the integration of the RiskScore improved AUC from **0.81 to 0.86** and significantly reduced time to review for high risk cases. These findings suggest that combining AI predictions with clinical context can enhance the prioritization of urgent cases and reduce the likelihood of missed diagnoses.

Beyond performance metrics, this work emphasizes a critical shift in the role of AI in radiology from standalone detection tools to systems that actively guide clinical decision making. Such approaches have the potential to improve patient safety, optimize workflow, and support radiologists in high demand clinical environments.

Future validation in multi center and real world settings will be essential to confirm the generalizability and clinical impact of the proposed framework.

## Nomenclature

RiskScore – Prioritization score

p\_calib – Calibrated probability

CNN – Convolutional Neural Network

AUC – Area Under Curve

PACS/RIS – Radiology systems

## Appendix

Detailed implementation steps, model parameters, and simulation methodology.

## References

- [1] Nagendran M., Chen Y., Lovejoy C., Gordon A., Komorowski M., Harvey H., et al., 2020. Artificial Intelligence versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies in Medical Imaging. *BMJ: British Medical Journal*, Vol. 368, pp. 1–14.
- [2] Goldhahn J., Daumer M., Holzinger A., 2018. Could Artificial Intelligence Make Doctors Obsolete? *BMJ: British Medical Journal*, Vol. 363.
- [3] Horgan D., Karlsen R., Røyrvik E.C., et al., 2019. Artificial Intelligence: Power for Civilisation – and for Better Healthcare. *Public Health Genomics*, Vol. 22, No. 5/6, pp. 145–161.
- [4] Bonifazi F., De Stefano L., De Falco F., et al., 2020. Machine Learning Systems Applied to Health Data and System. *European Journal of Health Law*, Vol. 27, No. 3, pp. 242–258.
- [5] Larasati R. and DeLiddo A., 2020. Building a Trustworthy Explainable AI in Healthcare. *Human Computer Interaction and Emerging Technologies: Adjunct Proceedings from the INTERACT 2019 Workshops*, pp. 209–214.
- [6] World Health Organization, 2021. Artificial Intelligence Is Changing the Health Sector. WHO Consultation Towards the Development of Guidance on Ethics and Governance of Artificial Intelligence for Health: Meeting Report, Geneva, Switzerland, 2–4 October 2019, pp. 3–7.
- [7] Uthoff J.C., 1879. The Diagnosis And Treatment Of Colles's Fracture Of The Radius. *The British Medical Journal*, Vol. 2, No. 980, pp. 574–574.
- [8] Donaldson L.J., Reckless J.P., Scholes S., et al., 2008. The Epidemiology of Fractures in England. *Journal of Epidemiology and Community Health*, Vol. 62, No. 2, pp. 174–180.

- [9] Shelmerdine S.C., White R.D., Liu H., et al., 2022. Artificial Intelligence for Radiological Paediatric Fracture Assessment: A Systematic Review. *Insights Imaging*, Vol. 13, pp. 94.
- [10] Kutbi M., 2024. Artificial Intelligence-Based Applications for Bone Fracture Detection Using Medical Images: A Systematic Review. *Diagnostics*, Vol. 14, No. 17, pp. 1879.
- [11] Marletta D.A., Nanni M., Giuca G., et al., 2024. The Role of Artificial Intelligence in Predicting and Managing Pediatric Fracture Overgrowth: A Comprehensive Review. *Applied Sciences*, Vol. 14, pp. 11652.
- [12] Oppenheimer J., Lüken S., Hamm B., Niehues S.M., 2023. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. *Life*, Vol. 13, pp. 223.
- [13] Sharma S., 2023. Artificial Intelligence for Fracture Diagnosis in Orthopedic X-rays: Current Developments and Future Potential. *SICOT Journal*, Vol. 9, pp. 21.
- [14] Nagy E., Janisch M., Hrzić F., Sorantin E., Tschauer S., 2022. A Pediatric Wrist Trauma X-ray Dataset (GRAZPEDWRI-DX) for Machine Learning. *Scientific Data*, Vol. 9, Article 327.