

A Radiology M&M-Style Root-Cause Analysis Framework for Artificial Intelligence Errors: Development of a Six-Domain Taxonomy

Colin Wu¹, Ravi Patel¹, Youyou Cheng¹, Sina Ashraf¹, Justin Kuyn², Edward Pettyjohn³, Zhuangwei Kang⁴, Sarah Pettyjohn⁵

¹University of the Incarnate Word School of Osteopathic Medicine ² University of Colorado Boulder ³ University of Illinois College of Medicine Peoria (UICOMP) ⁴ Independent Researcher ⁵ Rutgers Robert Wood Johnson Medical School

Introduction

Artificial intelligence (AI) tools are increasingly integrated into radiology workflows, yet errors from these systems remain poorly understood and rarely analyzed. Traditional quality and safety mechanisms such as morbidity and mortality (M&M) conferences, root-cause analyses (RCA), incident reporting, and RADPEER were designed for human and procedural errors, not algorithmic failures. Radiology, therefore, lacks a systematic method to identify, classify, and learn from AI errors. This project developed a clinically usable, M&M-style root-cause framework tailored to AI failures, enabling radiologists to categorize errors across data, model, workflow, and human–AI interaction domains and convert algorithmic mistakes into structured learning opportunities

Methods

We conducted a conceptual analysis integrating:

1. AI failure literature (data drift, label noise, generalizability, miscalibration);
2. Human-factors research (automation bias, anchoring);
3. Radiology QI frameworks (peer learning, RADPEER, RCA, FMEA); and
4. Real-world failures in breast, lung, and liver AI tools.

These informed a proposed six-domain taxonomy and conceptual workflow for AI-focused M&M discussions.

Results

Six primary root-cause domains captured most clinically relevant AI error types:

1. Data Failures: image quality, motion, contrast timing, vendor variability, data drift.
2. Label Failures: noisy ground truth, interobserver variability, outdated labels.
3. Model Failures: overfitting, miscalibration, brittle thresholds.
4. Integration Failures: PACS/VNA issues, routing errors, latency.
5. Workflow Failures: timing misuse, preprocessing errors, disrupted reading flow.
6. Human–AI Interaction Failures: automation bias, overreliance, misread AI output.

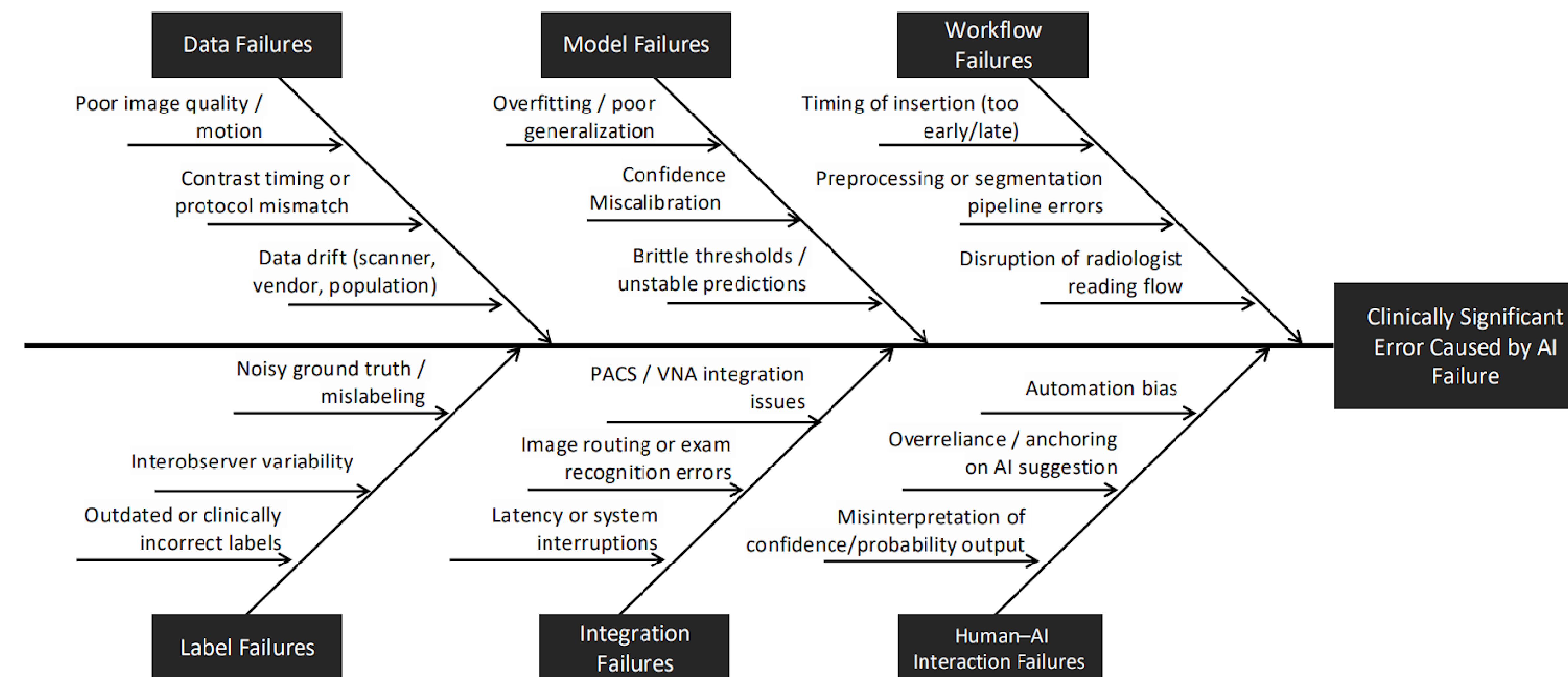


Figure 1. Fishbone (Ishikawa) diagram illustrating six primary root-cause domains of clinically significant radiology AI errors: data failures, label failures, model failures, integration failures, workflow failures, and human–AI interaction failures. Each domain contains representative sub-causes that contribute to downstream diagnostic or workflow errors attributable to AI malfunction or misuse. This taxonomy supports M&M-style structured analysis of AI-related adverse events in clinical imaging.

Conclusion

AI failures influence patient care, yet radiology lacks the structured error-analysis systems long applied to human performance. This M&M-style framework provides the first radiology-specific taxonomy for AI failure classification and a scalable model for peer learning and quality improvement. Extending M&M culture to algorithmic events can foster safer, more transparent AI integration and turn mistakes into system-level learning opportunities.

References

- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56.
- Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.
- Oakden-Rayner L. Exploring large-scale public medical image datasets. *Acad Radiol.* 2020;27(1):106–112.
- Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *Eur Radiol.* 2020;30(6):3576–3584.
- Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. *Radiology.* 2019;293(2):436–440.
- Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit Med.* 2020;3:41.
- Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak.* 2020;20(1):310.
- Graber ML, Kissam S, Payne VL, et al. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf.* 2012;21(7):535–557.