Scalable AI Model Monitoring for Medical Imaging:

An Open-Source Framework for Drift-Aware Medical AI Deployment

Stephanie Mehltretter, Kristina John, Justin MacPherson

AKROS AG

Objective of the Study

This study proposes a scalable and reproducible framework for continuous monitoring of machine learning (ML) models in medical imaging. The objective is to enable systematic post-deployment validation under real-world conditions, ensuring model robustness, reliability, and traceability in clinical environments. In contrast to prior work focusing primarily on model development, this work addresses in addition the **operational phase of medical AI systems, with a focus on monitoring and lifecycle management and a continuous validation of model outputs.**

Motivation & Background

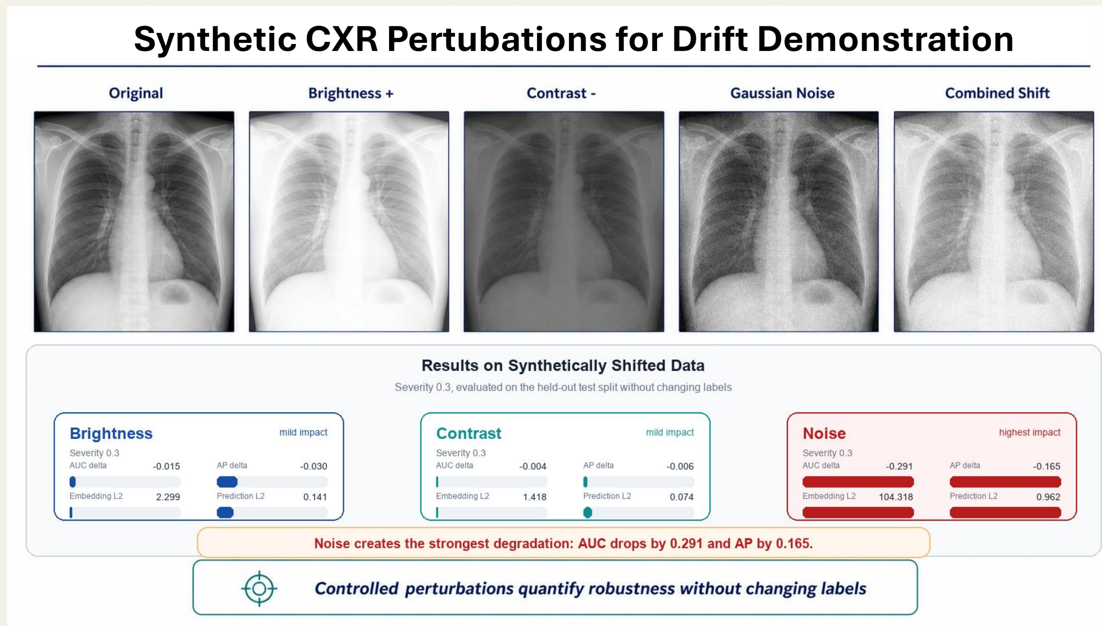
Although AI systems have demonstrated strong performance in medical imaging tasks, their clinical adoption remains limited due to poor generalization across institutions. **A central challenge is domain shift, caused by variations in imaging protocols, hardware, and patient populations, which can significantly degrade model performance under real-world conditions** [1,2]. Both **data drift** (changes in input distributions) and **concept drift** (changes in the relationship between inputs and outputs) pose **critical risks in safety-sensitive domains** such as healthcare [3]. Undetected drift may lead to systematically biased or unreliable predictions, raising concerns regarding clinical safety and regulatory compliance. Despite increasing awareness of these challenges, practical and reproducible frameworks for monitoring deployed models remain underdeveloped, particularly for high-dimensional medical imaging data.

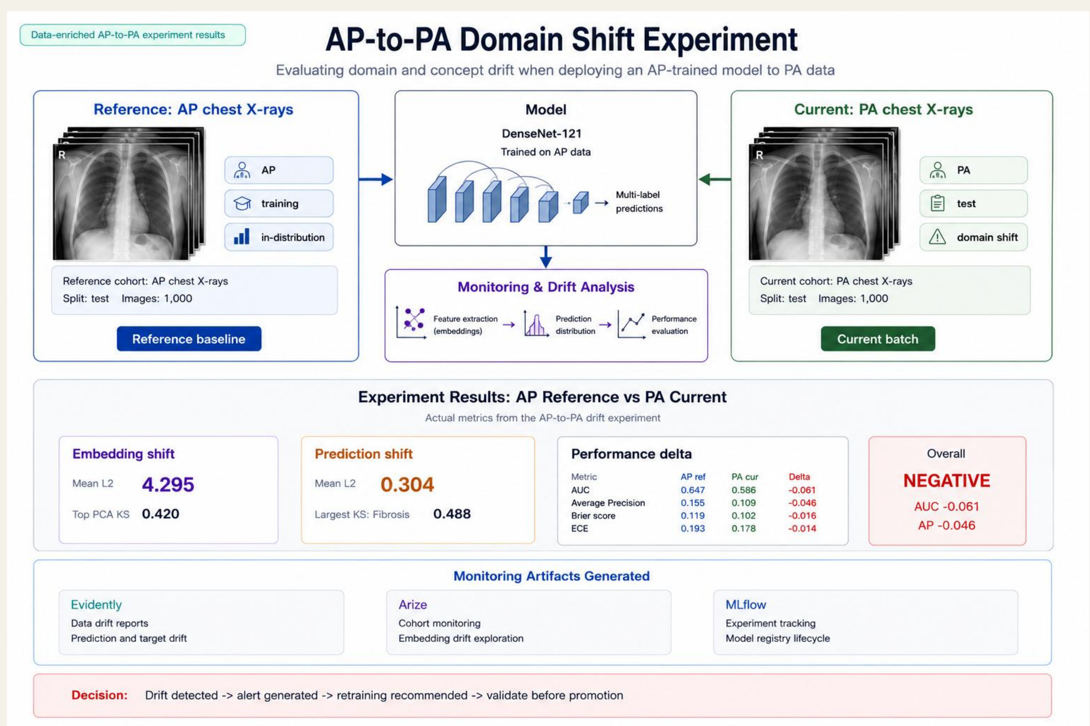
Short Description of Methods

We propose a **modular, open-source MLOps framework that integrates experiment tracking, deployment, and continuous monitoring. The architecture is centered on MLflow, Evidently and Arize while combining multiple monitoring layers: Data Monitoring, Model Monitoring, Embedding-based Monitoring, and Observability Layer**. The framework is evaluated using the publicly available **NIH ChestX-ray14 dataset** [4]. Controlled distribution shifts are introduced via intensity transformations (brightness & contrast), noise injection, and domain-based splits to simulate realistic deployment scenarios. The proposed architecture is **informed by practical experience** in building scalable data platforms and production-grade analytics systems in industrial environments.

Results

The proposed framework detects distribution shifts and links them to downstream performance degradation. **Embedding-based monitoring** proves particularly effective for identifying subtle changes in high-dimensional image data. Consistent with prior findings, models trained on a single dataset exhibit measurable performance degradation under shifted conditions [2,5]. The integration of monitoring into the **MLOps pipeline enables automated alerting, continuous evaluation, and reproducible analysis** across deployment scenarios and can trigger re-training.





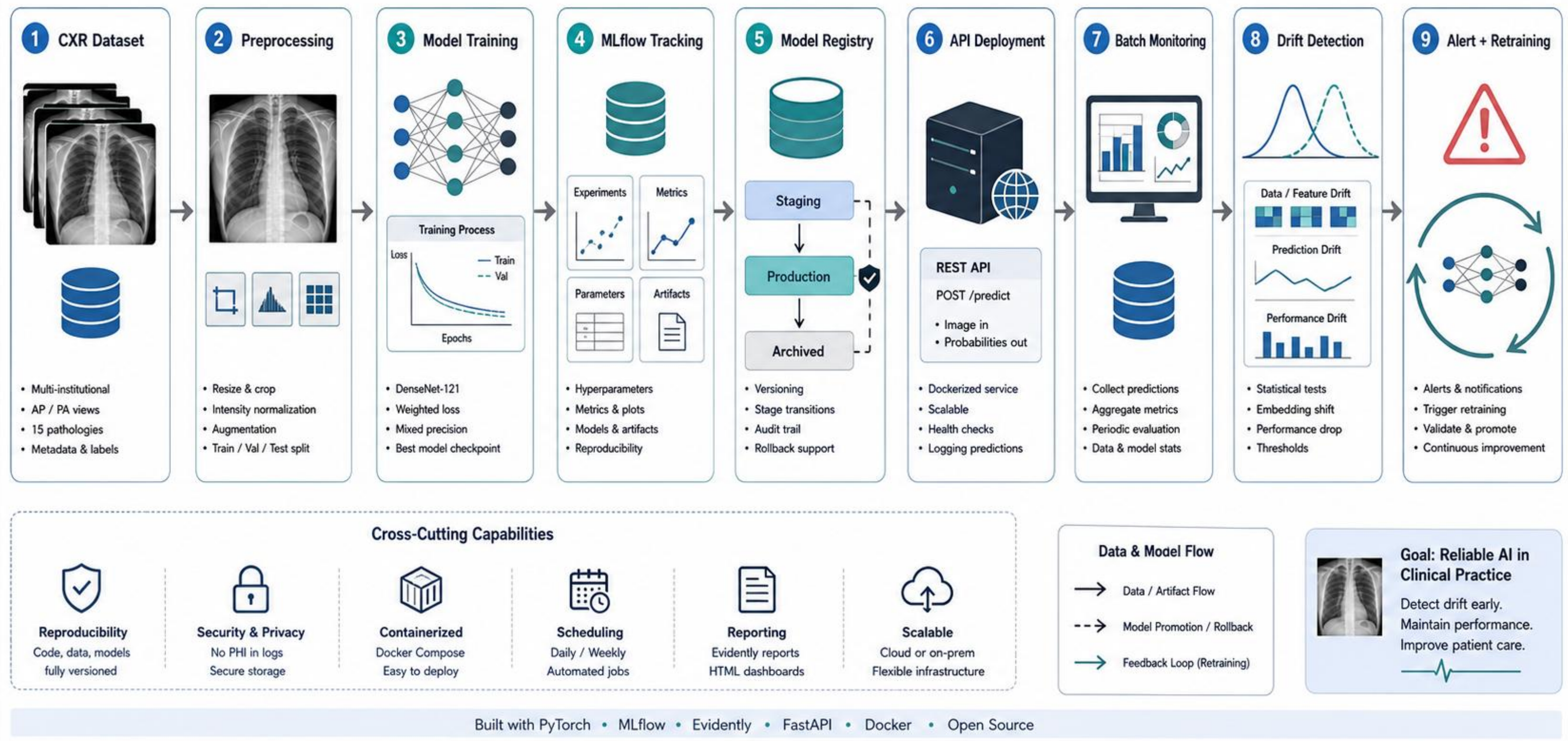
List of References

[1] Kelly, C. J., et al. "Key challenges for delivering clinical impact with artificial intelligence." BMC Medicine, 17, 195 (2019).
[2] Zech, J. R., et al. "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs." PLoS Medicine, 15(11), e1002683 (2018).
[3] Gama, J., et al. "A survey on concept drift adaptation." ACM Computing Surveys, 46(4), 44 (2014).
[4] Wang, X., et al. "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly supervised classification and localization." Proceedings of the IEEE CVPR (2017).
[5] Recht, B., et al. "Do ImageNet classifiers generalize to ImageNet?" Proceedings of the ICML (2019).

Contact & Link to Github Repository

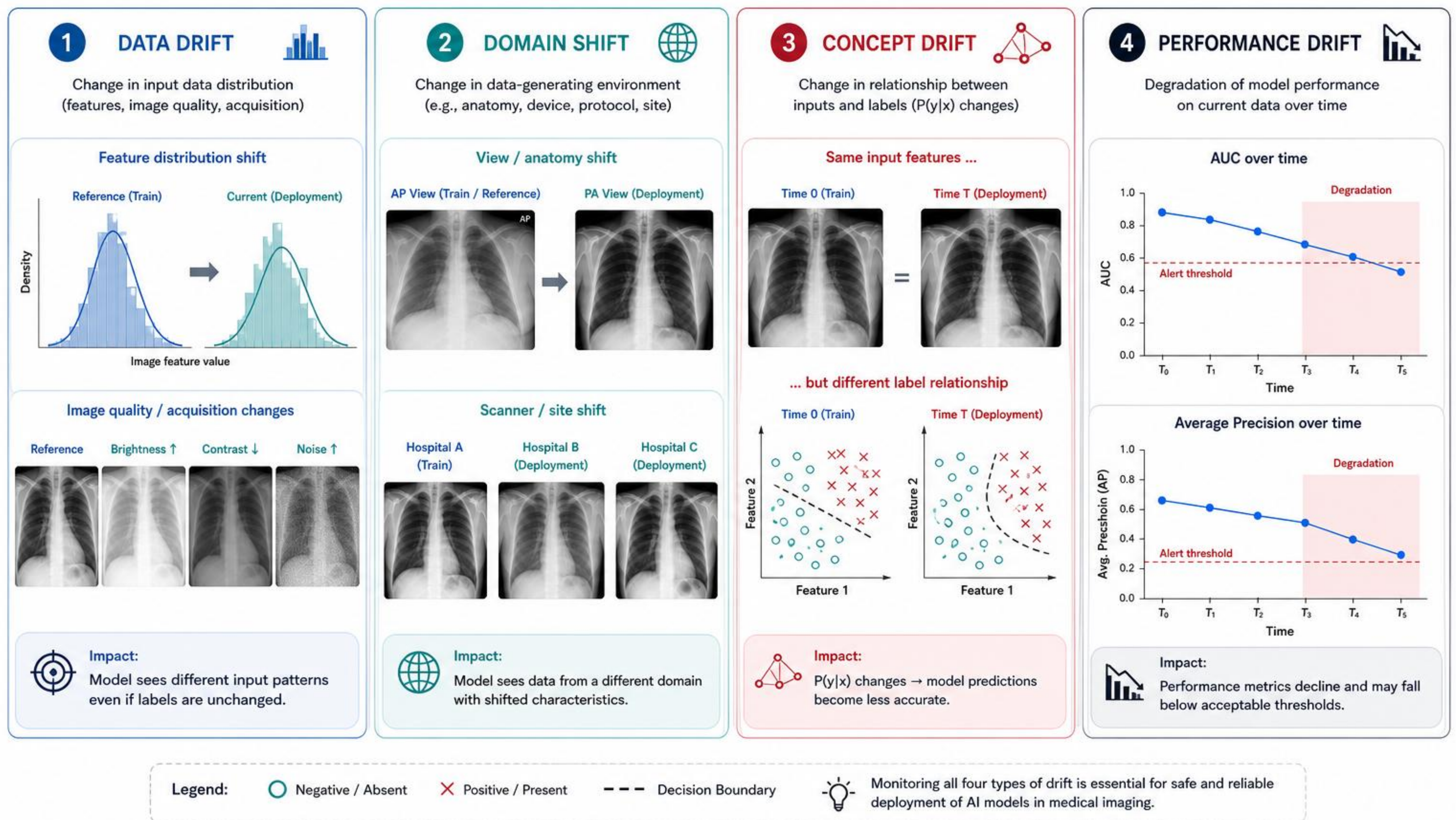
AKROS | Schweizer IT Dienstleister | Massgeschneiderte Lösungen
https://www.akros.ch
stephanie.mehltretter@akros.ch,
kristina.john@akros.ch,
justin.macpherson@akros.ch

Open-Source MLOps Framework for Medical Imaging



Drift Types in Medical Imaging

Understanding how data, environment, labels, and performance change over time



Model Lifecycle: Monitor, Retrain, Promote, Roll Back

A closed-loop MLOps framework for reliable medical AI in production