Doctoral Dissertation
Doctoral Program in Bioengineering and medical-surgical sciences (38^{th}cycle)

Methods and tools for Augmented Humans in Medicine

By

Francesca Giada Antonaci

******

Supervisor(s):
Prof. Sandro Moos, Supervisor
Dr. Elena Carlotta Olivetti, Co-Supervisor

Doctoral Examination Committee:

Politecnico di Torino
2026

Abstract

While clinical Artificial Intelligence (AI) systems are increasingly integrated into patient care, their deployment often precedes a comprehensive understanding of their boundary conditions and structural failure modes. This thesis addresses this vulnerability by introducing epistemic humility as an architectural design principle: clinical AI systems should make the limits of their knowledge visible and available for clinical review before their outputs influence care. Methodologically, this principle is examined across three points of the AI lifecycle: clinically validated data augmentation before training, explainable AI with clinician-controlled automation during use, and image reconstruction for auditing after deployment. Proximal femur radiography, echocardiography, and chest radiography are treated as clinical case studies through which these methodological contributions are developed.

To address the first stage of this lifecycle, data augmentation is examined through the lens of clinical governance in proximal femur fracture classification. Because augmentation alters the training distribution, selecting transformations solely for statistical diversity can generate images that detach from clinical reality. To address this, each candidate transformation was reviewed alongside an orthopaedic team prior to training, leading to the removal of three augmentations deemed clinically implausible. A YOLOv8 model trained on this clinically curated dataset achieved an accuracy of 0.90 and an F1-score of 0.85, demonstrating that clinical oversight at the data-pipeline stage effectively prevents unrealistic variations from compromising model behaviour.

Moving from data preparation to deployment, the second contribution addresses explainable AI and clinician-controlled automation for Left Ventricular Ejection Fraction (LVEF) estimation from echocardiographic video. The system named CardioSmartAssist combines MultiResUNet segmentation, area-frame curve construction, phase identification, volume computation, and multi-cycle aggregation

within a workflow co-designed with cardiologists through UML modelling. By exposing the intermediate steps of this pipeline, the system ensures that measurements remain accessible for inspection and correctable before entering the clinical record. When evaluated, MultiResUNet achieved a Dice coefficient of 0.9328, while multi-cycle aggregation reduced mean absolute LVEF error to 5.09 percentage points. Furthermore, in an initial feasibility study on 27 patients, the system achieved a mean absolute difference of 10 percentage points against clinical measurements, falling below the documented inter-observer variability up to 13.9 percentage points for manual echocardiographic assessment.

Finally, for the post-deployment phase, the third contribution introduces image reconstruction as an auditing strategy for deployed chest radiograph classifiers. Autoencoders trained on the MIMIC-CXR distribution generated architecture-independent reconstruction scores to signal potential classifier unreliability on anomalous or out-of-distribution cases. Across five pathologies, review of the top 5% of cases by reconstruction score recovered up to 13.8% of all false negatives, with false negatives accounting for 19–34% of flagged errors in the most stringent review subset. The limited overlap with classifier confidence, below 10% across all tested conditions, shows that the reconstruction signal captures cases largely missed by confidence-based review.

Across these three case studies, these contributions translate epistemic humility into practical design choices across the AI lifecycle, shaping the data the model learns from, the form in which automated measurements enter clinical review, and the signals used to identify failures after deployment. Through this approach, this thesis aims to demonstrate that AI enhances diagnostic capabilities when its outputs remain structurally inspectable and accountable to responsible clinical judgement.