Evaluating Fracture Healing with Artificial Intelligence

Using Transfer Learning to Predict RUST Scores from Radiographs. A Major Extremity Trauma Research Consortium Project.

Shen Zhou Hong <shong@jhu.edu>

Johns Hopkins Bloomberg School of Public Health
METRC Major Extremity Trauma and Rehabilitation Consortium
Goldsmiths University of London
University of London

Introduction and Table of Contents

- Table of Contents:

Background Information

Using AI to Evaluate Fracture Healing

- We want to develop an AI that can infer RUST scores from radiographs.
- RUST: Radiographic Union Score for Tibial Fractures. Also used in other long bone fractures.
- RUST measures the progression of fracture healing via callus formation and bridging.

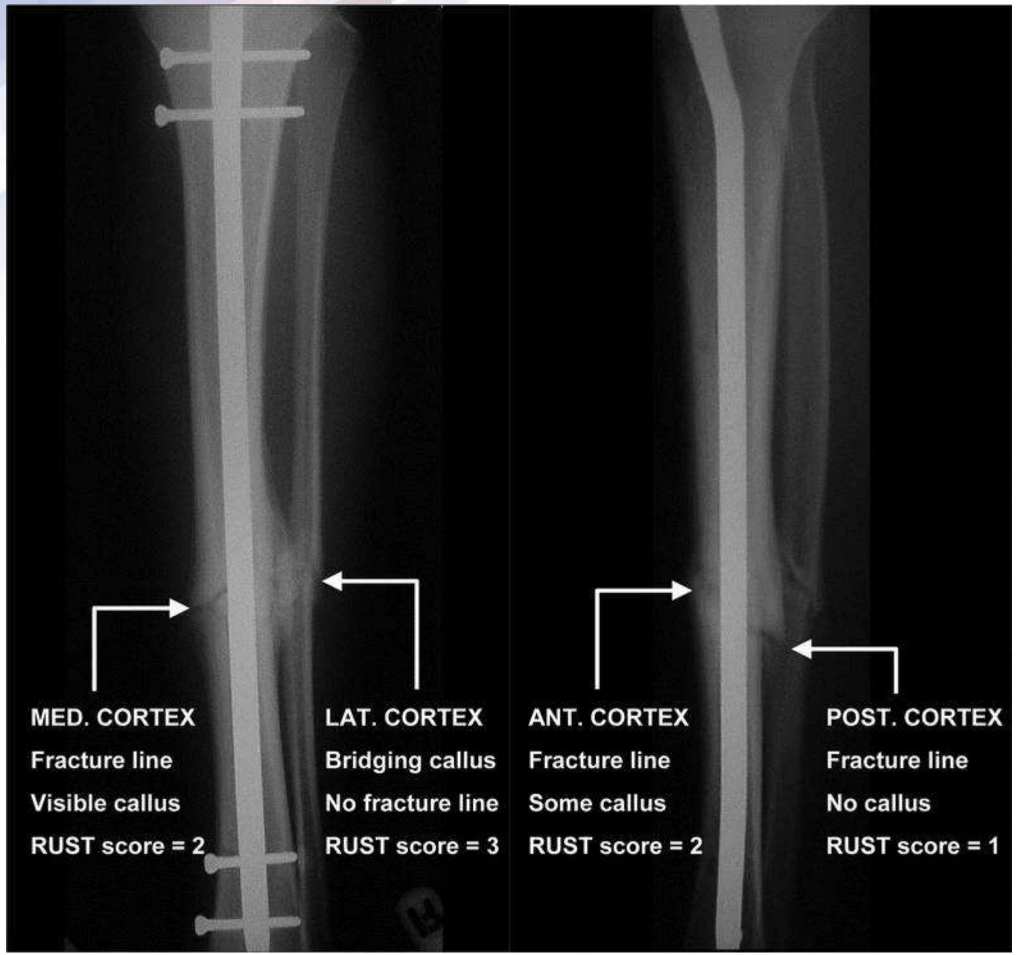

Image source: D. B. Whelan, M. Bhandari, D. Stephen, et al. 2010

Why is this project useful?

- The AI model can automate the analysis of archived study data.
- RUST scores have good biomechanical correlation, can serve as guidance for rehab.
- Project is novel: prior work in AI-radiography mainly focused on fracture detection.

Prior Research in AI-Radiography

- MURA: Anomaly detection. 40.5k radiographs, achieved AUROC of 0.929
- Lindsey et al: Anomaly detection & localization. 31.0k radiographs, achieved AUROC of 0.967
- Kim et MacKinnon: 1.3k radiographs, achieved AUROC of 0.954

Why has nobody done this before?

- Most individual institutions do not have access to datasets large enough to perform training.
- Most large radiography datasets do not have specialized, *evidence-based, adjudicated* labels.
- Most AI research organizations do not have access to medical research organizations.

METRC is uniquely positioned to conduct this research.

Design, Methodology, & Endpoints

AI Models versus Statistical Models

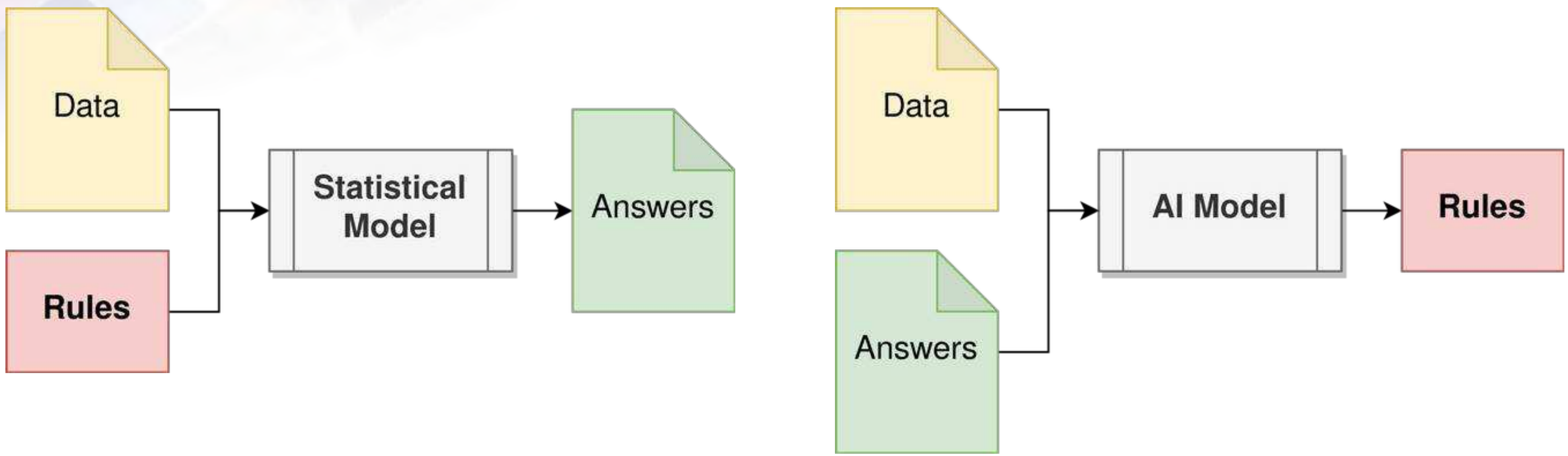

Image source: METRC 2022

METRC Radiography Datasets

- RetroDEFECT: 741 radiographs
- OUTLET: 707 radiographs
- PAIN: 370 radiographs
- PACS: 195 radiographs
- Total: ~2,013 radiographs
- Our study architecture is data-constrained.

Model Architecture

- *Transfer learning*: train an AI model on a large, general-purpose dataset. Then *fine-tune* on the smaller, task-specific dataset.
- Data Augmentation & K-Fold Validation
- We will use ImageNet and MURA as base models for transfer-learning.

Anatomy of a Deep Neural Network

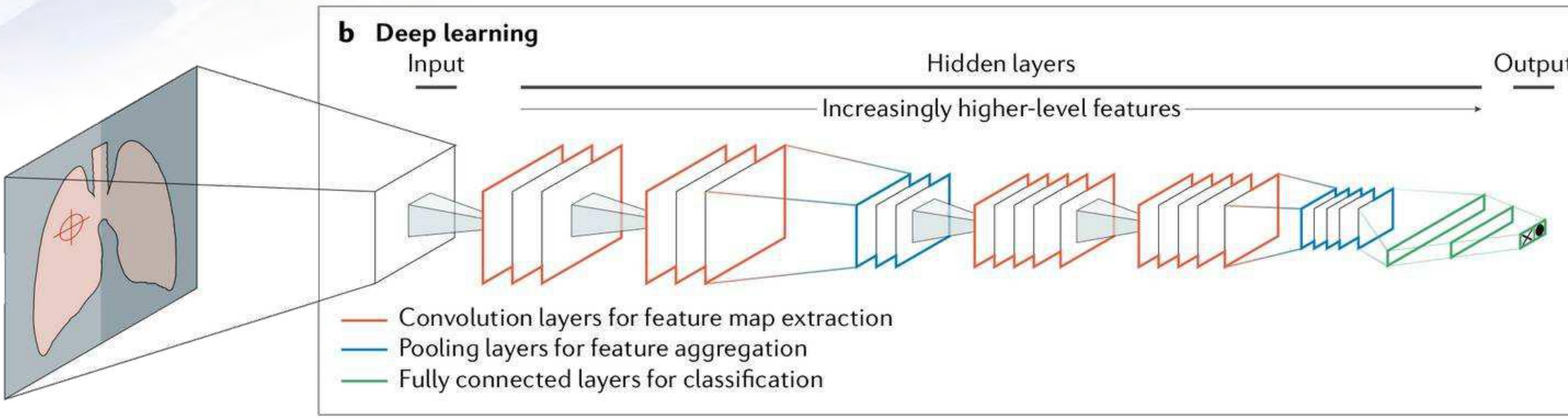

Image source: Hosny, Parmar, Quackenbush et al. 2018

Endpoints

- Model performance measured via AUROC: Area Under Curve (of) Receiver Operating Characteristic.
- Endpoint 1: Model AUROC > 0.50
- Endpoint 2: Model AUROC > "naive" ConvNet
- Endpoint 3: Model AUROC > 0.75

Progress, Roadmap, & Challenges

Roadmap and Current Progress

- 2022-11-15: Complete background research
- 2022-12-01: Study and DNN architecture design
- 2022-12-12: Initial exploration of METRC datasets
- 2023-01-02: Preprocessing of METRC radiographs. (**We are here)**
- 2023-01-31: Initial model development
- 2023-02-27: Fine-tuning, hyperparameter search.
- 2023-03-31: **Deadline: Impl. & Analysis**
- 2023-05-02: **Deadline: Poster Presentation at UoL Goldsmiths**
- 2023-05-12: **Deadline: Submission to UoL Goldsmiths**

Challenges

- Dataset validation and "cleaning."
- Combating model overfitting.
- Programmatically parsing data from REDCap.

REDCap Libraries and Tools

- REDCap Branch Parser: https://github.com/metrc/redcap-branch-parser
- REDCap Schema: https://github.com/metrc/redcapschema

Future Pathways

- Heat-maps and "Explainable AI"
- Further assessment of model performance with human orthopedic specialists/surgeons
- Collect more RUST data. Train a more robust model, conduct serious evaluations of using AI as a diagnostic tool.

The End. Thank you for your attention!