

PRODIGY AI SOLUTIONS

CohortForge AI

Clinical & CT Evidence Studio

Generative modelling for balanced, complete
and harmonised oncology research cohorts

MVP / Laboratory validation

Research demonstrator | Laboratory validation



A research workflow for incomplete cancer cohorts

For data stewards, imaging scientists and AI development teams

01

Incomplete records

Generate mixed clinical records and reconstruct masked fields while retaining observed values.

Output: synthetic or completed tables

02

Underrepresented groups

Set cohort composition targets and report achieved counts, shortages and distribution changes.

Output: cohort evaluation reports

03

Incomplete CT data

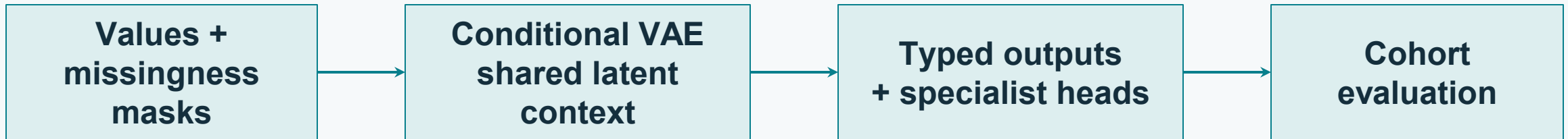
Reconstruct withheld CT slices and compare image errors with transparent interpolation baselines.

Output: images and error evidence

Specialist models with a shared evidence pipeline

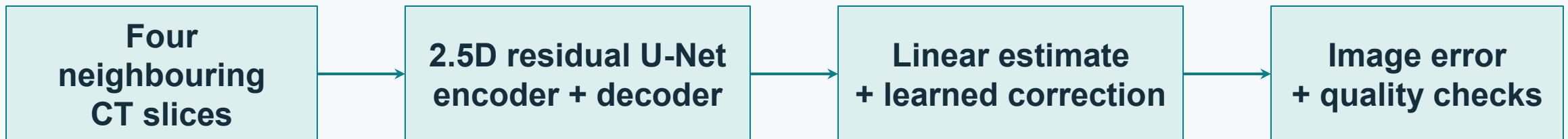
Clinical records and CT images use different model architectures

CLINICAL MODEL



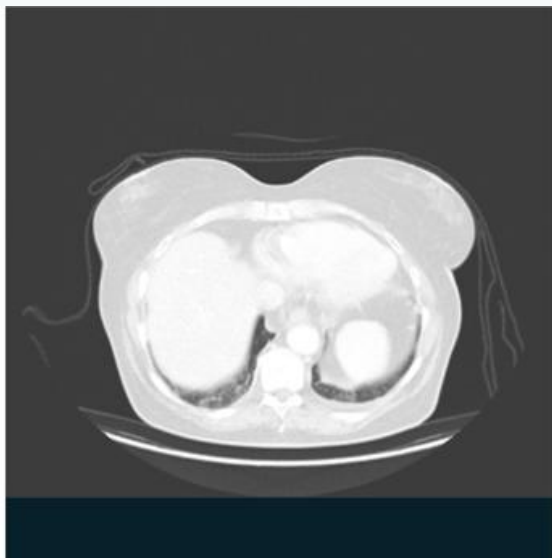
Gender, stage and histotype condition generation; specialist age-quantile and PD-L1 Beta heads model selected fields.

CT MODEL

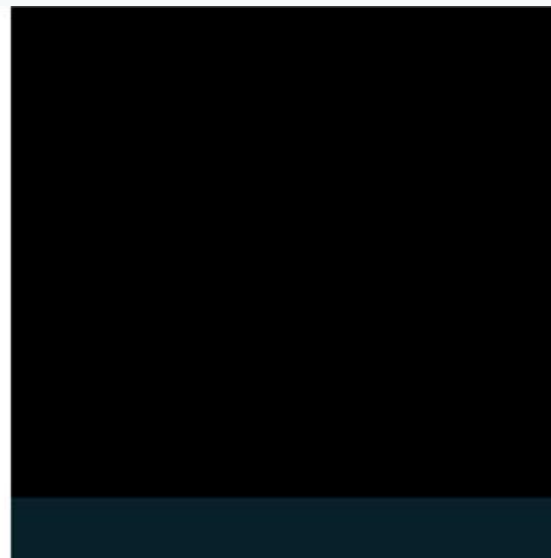


CT reconstruction in the MVP

A withheld centre slice makes reconstruction error measurable



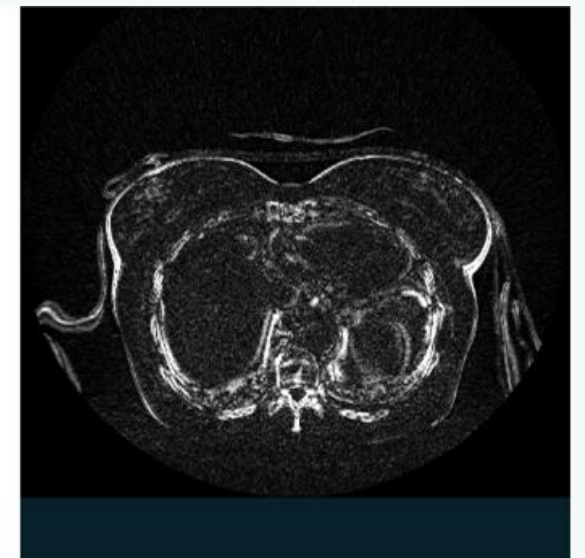
Original



Masked input



Reconstruction



Error map

Illustrative reconstruction workflow — not a clinical result. Quantitative laboratory results follow.

CT learning curve in laboratory testing

Median masked-slice MAE reduction relative to linear interpolation



22.45%

**median MAE reduction
vs linear interpolation at**

16 training subjects
6 / 6 evaluation subjects
improved

15 fits · 7,830 slice predictions
22 subjects in one collection

Internal one-collection feasibility study; not clinical or multicentre validation. CPU-only study

Laboratory evidence and qualification gates

Clinical component results

88 / 90

PD-L1 paired blocks improved
Median change in KS: -0.0534

81 / 90

Age paired blocks improved
Median change in KS: -0.0422

Lower KS = closer marginal distribution

Artificial clinical data; isolated comparisons.

Integrated system

**720 evaluations completed
with zero technical failures**

Age component qualified in integration.
PD-L1 component remains at STOP.

**Full generator: qualification remains
open**

Fixed criteria retain failures and blocked runs.
Human review governs model promotion.

Team and technical execution

Alessandro Ludovici

Technical lead · ML, imaging & MLOps

20+ years in enterprise software development

TensorFlow-certified

Generative modelling and reproducible experiments

10,000

**GPU-hour allocation
previously managed
and utilised**

Alena Baida

Project, governance & commercial lead

14+ years in product growth, go-to-market and strategy

Hands-on experience with
CINECA Leonardo A100
workload orchestration

The next validation step

Research scope

Fresh subject panels and multicentre CT data

Clinical fidelity and subgroup evaluation

GPU/HPC profiling of larger workloads

Verbis Graph evaluation on HPC

Discussion topic: workload, benchmark and success criteria

Meeting outcome

Agree a bounded validation use case

Dataset and access

Baseline and metrics

Expert review

Compute and deliverables

MVP walkthrough

med.prodigy-ai-solutions.com