SYNTHEMA + ERN-EuroBloodNet

Joint Training Programme on Synthetic Data Generation in SCD and AML

Funded by the European Union

Data standardisation & interoperability and synthetic data in PHEMS

Max Salmi, VEIL.AI

Why is PHEMS needed?

Urgent need for safe data sharing

1 **Small numbers of complex pediatric and rare diseases patients per country**

- Limited ability to conduct research, develop new therapies, and improve patient outcomes

2 **Limited real-world data to evaluate newer expensive therapies**

- Health technology assessment (HTA) bodies lack data and processes to evaluate treatment outcomes

3 **Need for more engagement with Medtech industry & innovation ecosystems**

- Large data sets are needed to leverage benefits of artificial intelligence and machine learning and support innovation

4 **Many countries only have one or two children's hospitals**

- Shared data are necessary for benchmarking to improve system performance and efficiency

phems

Building a Pediatric Health Data Space (PHDS)

Objectives

- **Accelerate research in pediatric and rare diseases** by increasing access to health data while protecting patient privacy
- Increase **efficiency and quality of healthcare delivery** through benchmarking
- **Improve patient outcomes** by leveraging the power of federated health data analysis, machine learning, and synthetic data

The **Pediatric Health Data Space (PHDS)** will consist of **technical infrastructure** and **governance frameworks**, empowering institutions to **collaborate without relinquishing control** over their data.

phems

How do we do it?

Collaboration | **Technology** | Use Cases | Governance

Data standardization

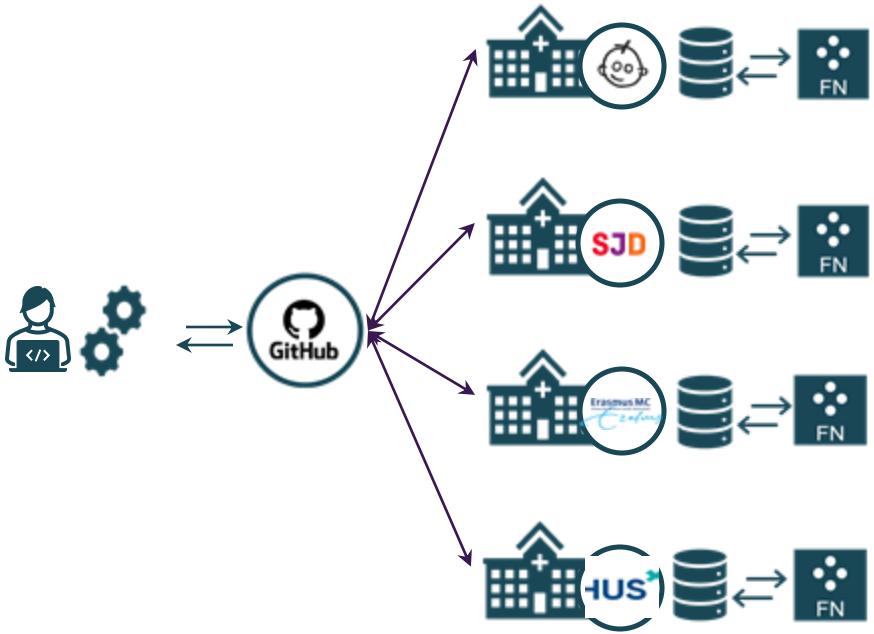
- OMOP CDM
- Prerequisite for Federated network: Transform hospital EHR data to OMOP

Data interoperability

- Federated network
  - Cross-border data collaboration
  - Combine small data sources
- Federated Analysis
- Federated Learning

Data privacy protection

- Data anonymization and synthetic data generation to complement Federated network



phems

Funded by the European Union

UK Research and Innovation

How do we do it?

Collaboration | **Technology** | Use Cases | Governance

Data anonymization and synthetic data

An essential technological component of the PHDS that:

- **Provides additional privacy safeguard** before Federated Analysis/Learning processes
- **Accelerate data access** with anonymous data
- **Ensures data anonymity** evaluation before granting access to results

Original (raw data)
"Hospital use"

John – 35 y/o male
Original data

Anonymous synthetic data
"Enabling new use cases"

Male – 35 y/o
Individual level data based on real data

Software application is deployed in data controllers' (hospital) environment resulting in anonymized GDPR compliant OMOP data

veil.ai
ENABLING USE OF SENSITIVE DATA

www.veil.ai

phems

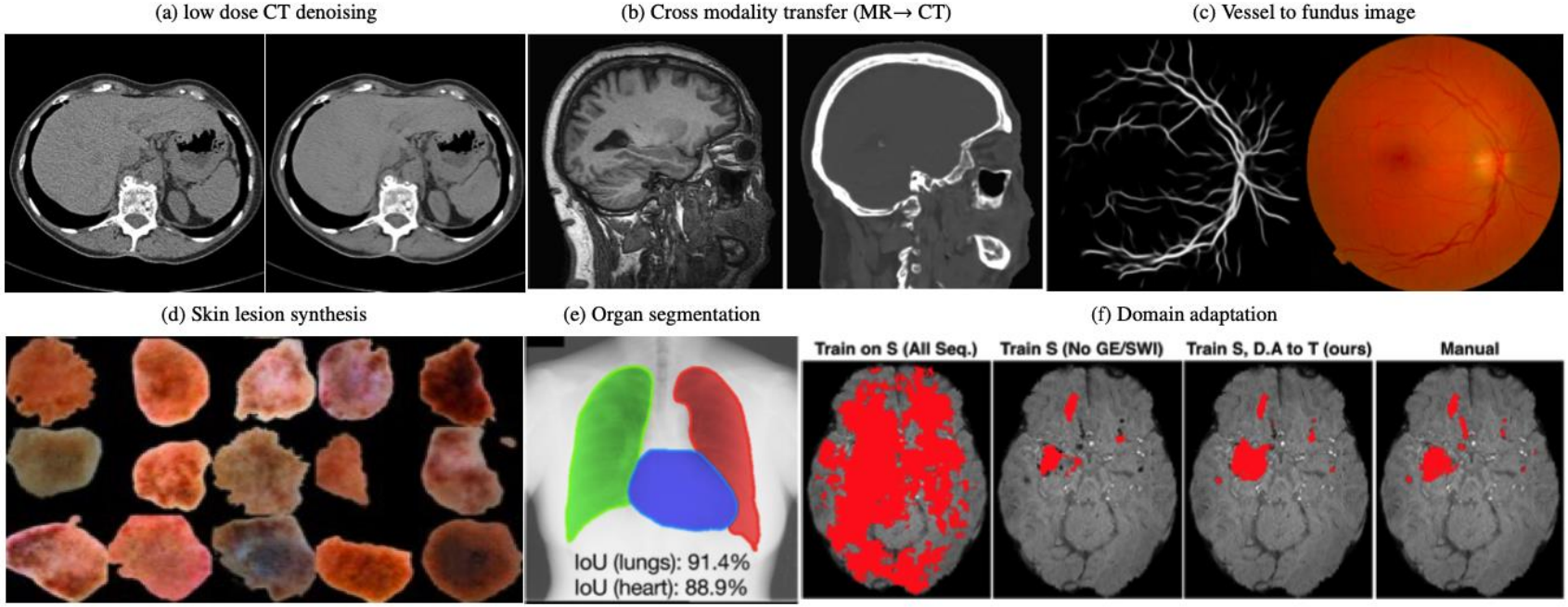
Funded by the European Union
UK Research and Innovation

PHASE IV AI

Rafael Redondo

Medical Image Synthesis

A wide range of applications



Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical image analysis, 58, 101552.

Medical Image Synthesis

What for?

Privacy compliant data

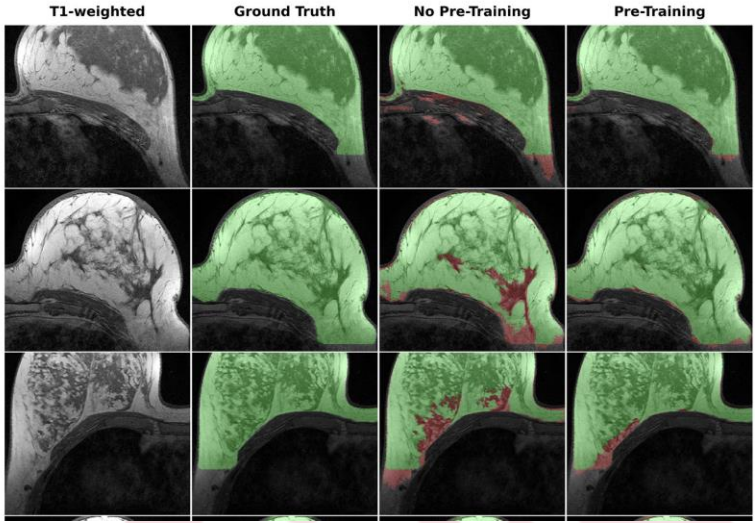
Differential privacy and synthetic data as anonymized data [Ziller et al. 2021, Ficek et al. 2021]

Downstream tasks: data scarcity, faster acquisition

As data augmentation for fine-tuning detection models [Ferreira et al. 2024, Khader et al. 2023] or scan time reduction [Moschetto et al. 2025]

As tools for disease forecasting

Prediction of disease evolution, e.g. cancerous nodule growth in lung CT scans [Tang et al. 2026, Puglisi et al. 2025]



Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., & Kaissis, G. (2021). Medical imaging deep learning with differential privacy. *Scientific Reports*, *11*(1), 13524.
Ficek, J., Wang, W., Chen, H., Dagne, G., & Daley, E. (2021). Differential privacy in health research: A scoping review. *Journal of the American Medical Informatics Association*, *28*(10), 2269-2276.
Ferreira, A., Li, J., Pomykala, K. L., Kleesiek, J., Alves, V., & Egger, J. (2024). GAN-based generation of realistic 3D volumetric data: A systematic review and taxonomy. Medical Image Analysis, 103100.
Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., ... & Truhn, D. (2023). Denoising diffusion probabilistic models for 3D medical image generation. Scientific Reports, 13(1), 7303.
Moschetto, A., et al. (2025, September). Benchmarking GANs, Diffusion Models, and Flow Matching for T1w-to-T2w MRI Translation. In International Conference on Image Analysis and Processing (pp. 429-440). Cham: Springer Nature Switzerland.
Tang, X., et al(2026). NGP-Net: a Lightweight Growth Prediction Network for Pulmonary Nodules. IEEE Transactions on Medical Imaging.
Litrico, M., et al (2025). Temporally-aware diffusion model for brain progression modelling with bidirectional temporal regularisation. CMIG 102688.

Why Generative Models?

Capabilities and Challenges

Faithfully approximate complex real data distributions

- High fidelity (realistic)
- Plausible instances, not copies

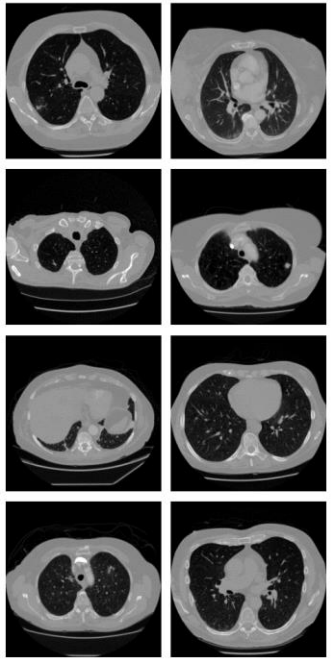
Improved reliability

- Address data scarcity in certain domains (multi-domain)
- Adaptable to different clinical contexts (conditional models)
- Access to prediction probability
- Visually interpretable predictions (visual feedback)

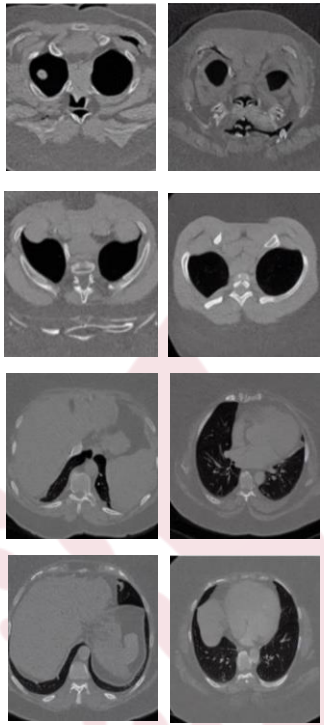
Challenges: hallucination

- Bias: imbalanced datasets induce bias on the synthesis
- AI safety: as decision-support tools, always require supervision
- Fairness: how to measure balanced and realistic generation?

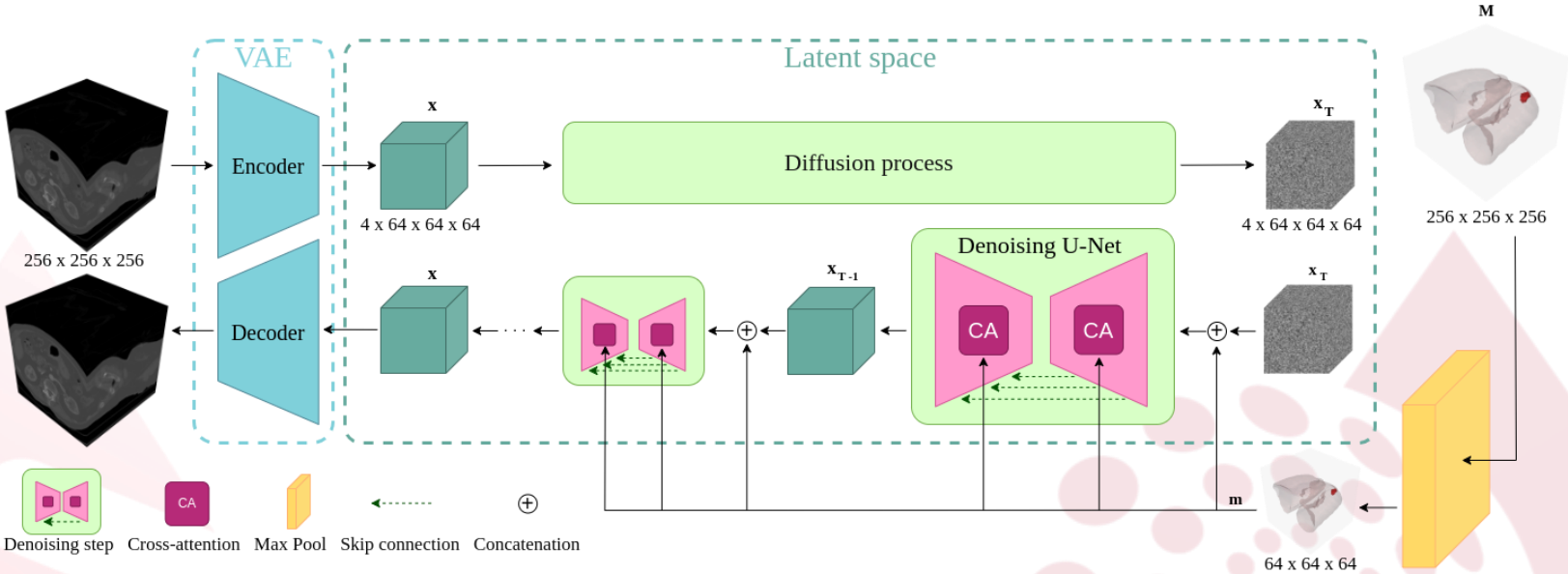
Real CTs



Fake CTs
(animal pre-trained)



LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance

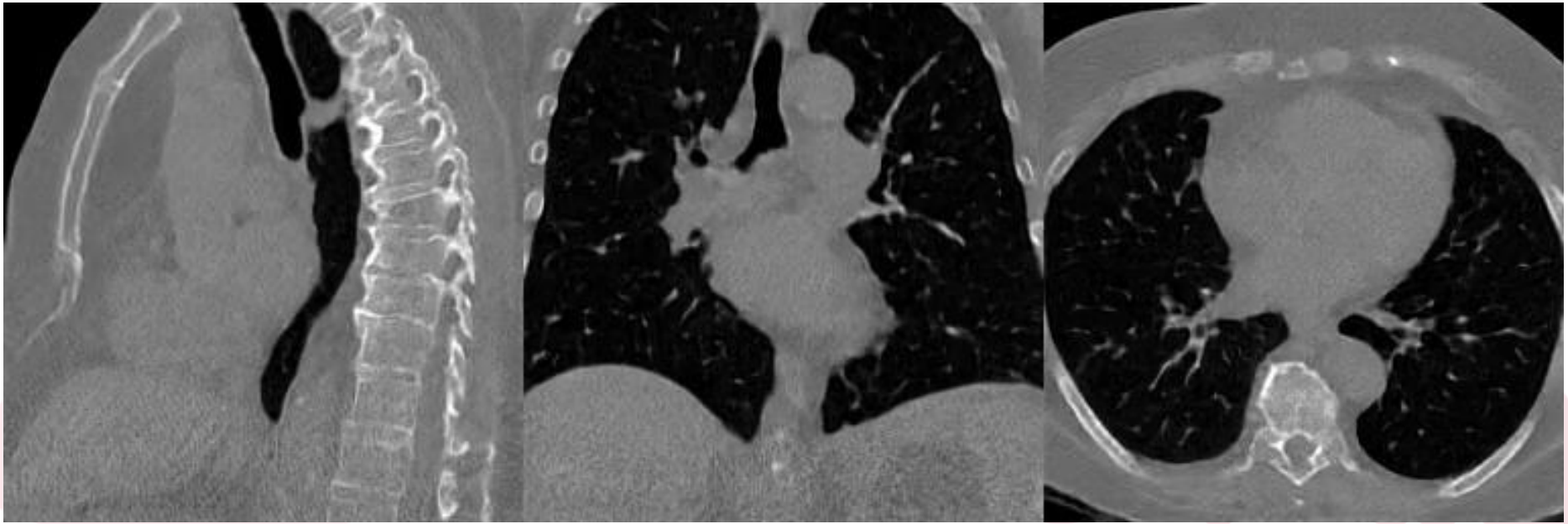


A. Oliveras, R. Marí, R. Redondo, O. Guardià-Olivella, A. Tost, B. Nagarajan, C. Migliorelli, V. Ribas, P. Radeva. “LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance.” Medical Imaging meets EurIPS: MedEurIPS, 2025.

LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance

Synthetic 3D Volume

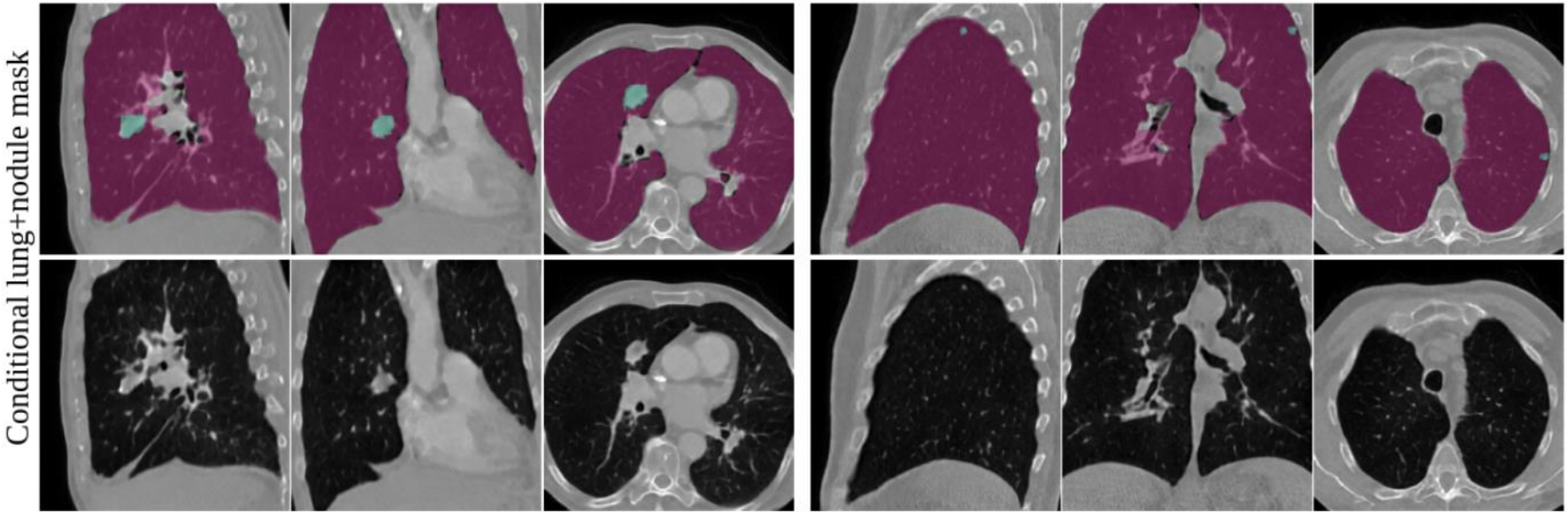
256^3 resolution



Sagital *Coronal* *Transverse*

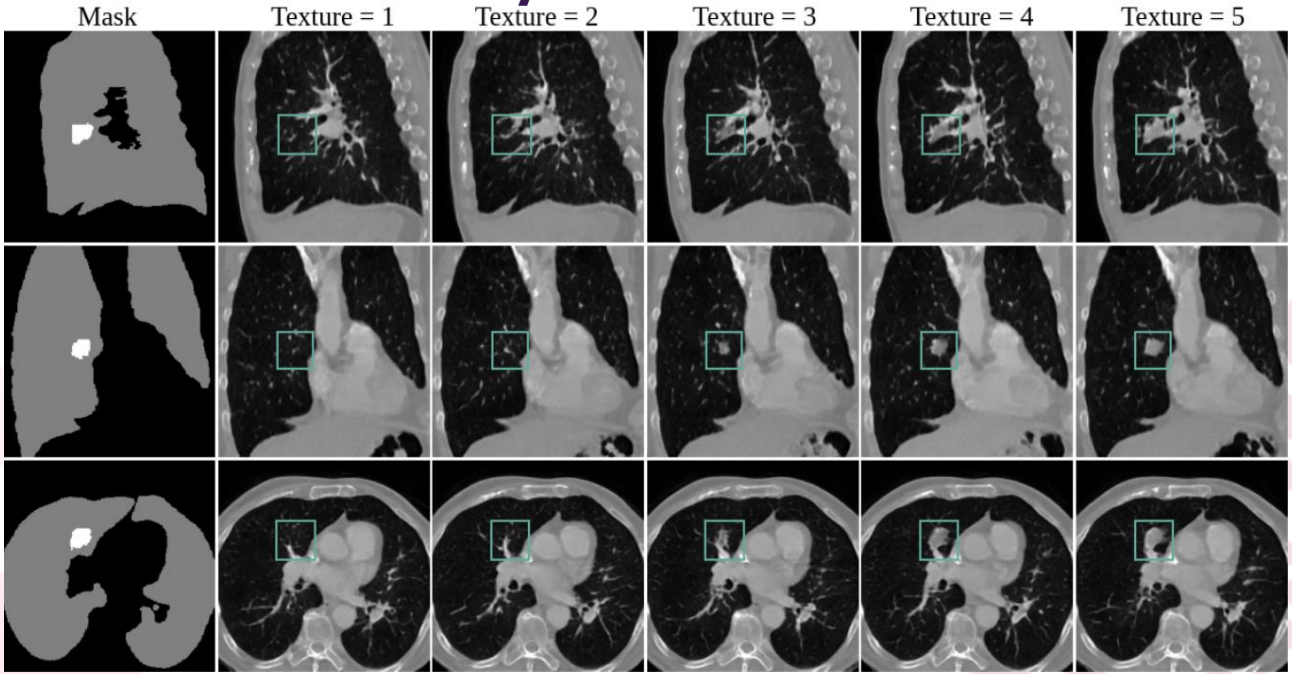
LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance

Anatomy-guided Lung+Nodule 3D Conditional Synthesis



LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance

Nodule Texture Conditional Synthesis



Trustworthiness: reliability of synthetic data

Synthetic Image Quality Metrics

OBJECTIVE
Direct Metrics

Statistical distribution models

- **Fidelity:** Fréchet Inception Distance (FID) or Fréchet Radiomic Distance (FRD).
- **Variety:** Similarity metrics, e.g. MS-SSIM.
- **Authenticity**: reals vs. copies ratio based on similarity metrics.

✓ Objective, reproducible.

⊘ Data-biased, not medical domain-tailored, or threshold-dependent.

[Alaa et al.]

OBJECTIVE
Indirect Metrics

Utility assessment

- **Data augmentation:** downstream tasks, e.g. Lung Cancer detection: *detection* (precision, recall), *localization* (Jaccard Index, Dice score), *geometry* (volumetric similarity, relative volume difference).
- **Performance similarity:** ML methods should perform equally on synthetic data.

✓ Utility assessment in specific clinical applications.

⊘ Data and model biased, low interpretability.

[Xu et al., Khader et al., Chen et al., Onishi et al. , Toda et al.]

SUBJECTIVE
Metrics

Visual Turing Test

Human assessment of synthetic samples.

- Confusion matrices **subjective score opinion 1-4**, e.g. across image quality, slice consistency, or anatomic correctness.

✓ Tailored to clinical purposes

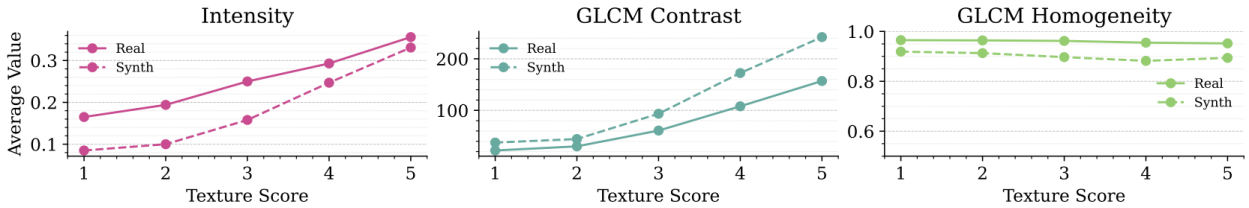
⊘ Time-consuming.

[Khader et al., Wang et al.]..

LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance

Fidelity & Variety Assessment

	Method	FID↓ (LIDC)	FID↓ (NLST)	MS-SSIM↓	Mem GB↓
Unconditional	PatchDDM [2]	317.534	376.397	0.390	19.61
	WDM [9]	15.236	32.662	**0.270**	**7.27**
	LAND Unconditional	5.062	4.759	0.294	7.38
Conditional	LAND nodule mask	4.518	5.821	0.297	7.52
	LAND lung+nodule mask	**4.475**	**3.373**	0.290	7.52
	LAND lung+nodule+texture mask	4.603	3.865	0.289	7.52

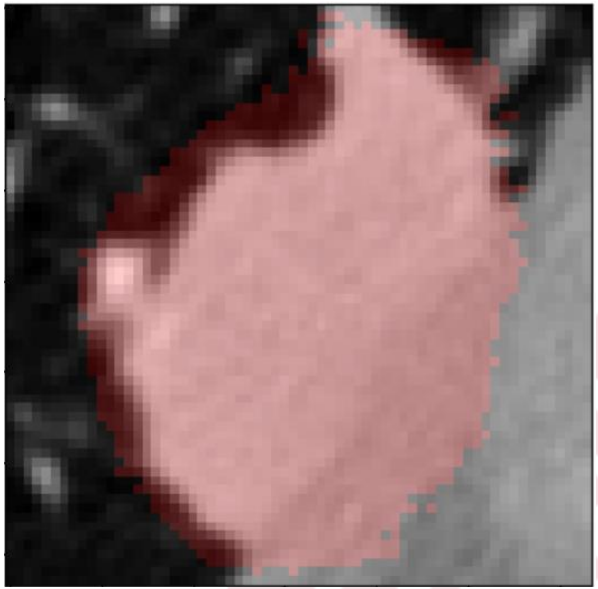


Intensity and statistical GLCM texture descriptors to assess nodule synthesis.

Downstream Task: Lung Nodule Segmentation

Indirect objective utility assessment

Idea: To test whether synthetic CTs used as data augmentation can improve performance in specific clinical tasks.



Downstream Task: Lung Nodule Segmentation

Indirect objective utility assessment

Dataset: LIDC dataset (1012 studies). Resolution: Variable + LAND synthetic dataset (880 studies). Resolution: 1mm.

Model: Monai's 3D UNet implementation.

Evaluation metrics: Dice and Jaccard.

Output type: 3D binary mask of the nodule.

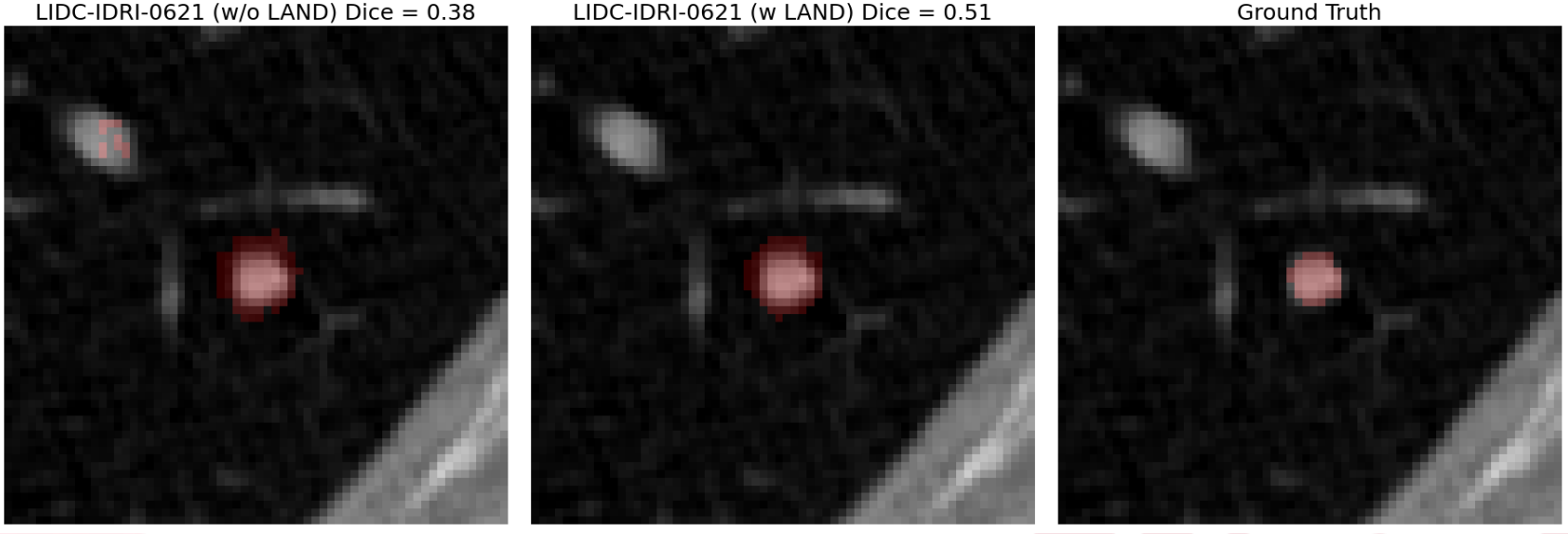
Two different experimental setups:

 # Experiment 1
Training: 50% LIDC
Validation: 20% LIDC
Test: 30% LIDC

 # Experiment 2
Training: 50% LIDC + **LAND**
Validation: 20% LIDC
Test: 30% LIDC

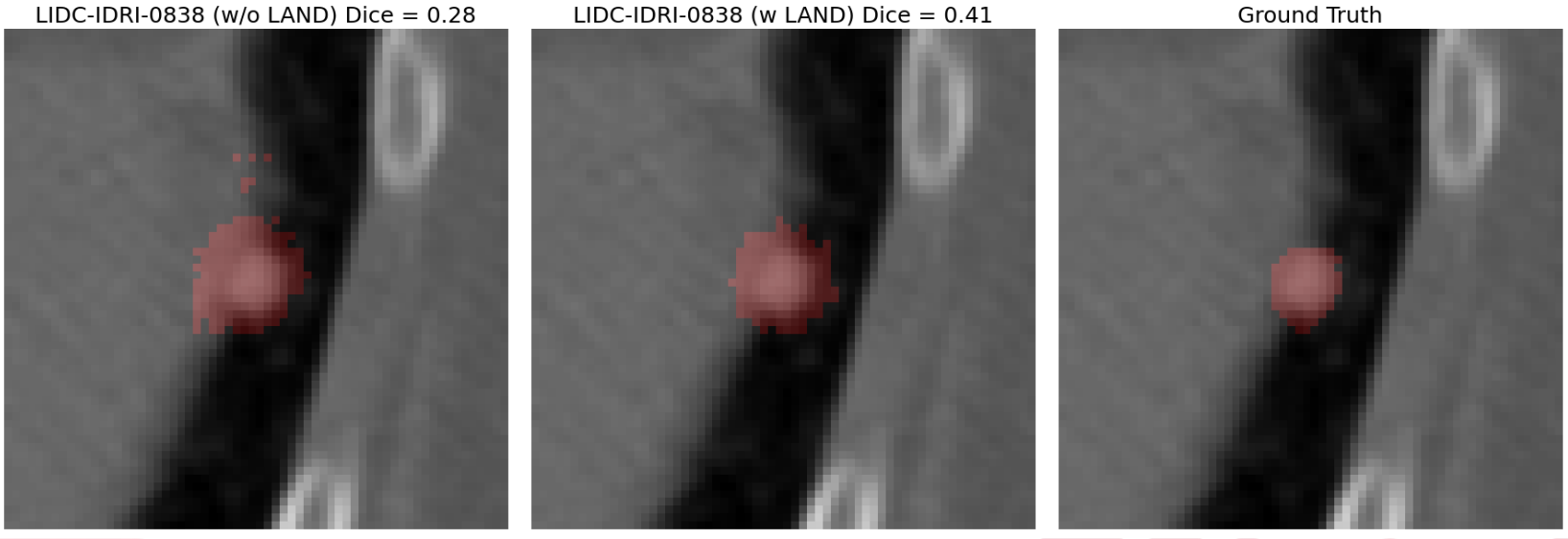
Downstream Task: Lung Nodule Segmentation

Indirect objective utility assessment



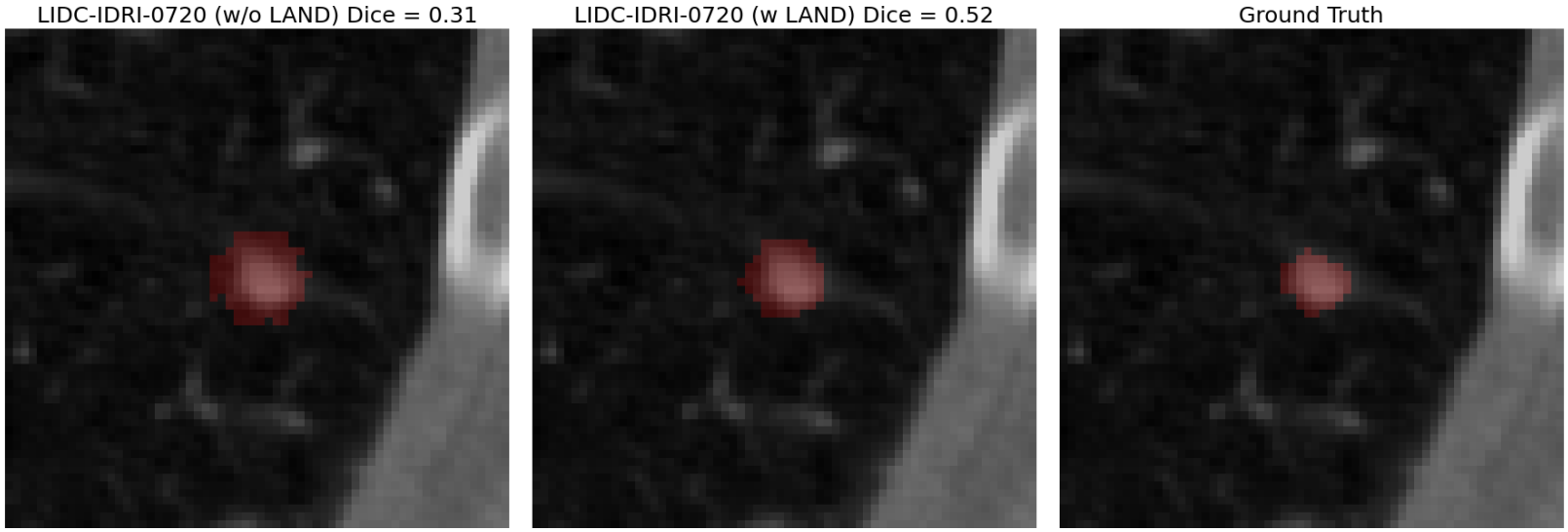
Downstream Task: Lung Nodule Segmentation

Indirect objective utility assessment



Downstream Task: Lung Nodule Segmentation

Indirect objective utility assessment



Medical Image Synthesis

Conclusions and Open Challenges

- **Improved image realism**, fewer artifacts and hallucinations thanks to conditional architectures; however, **eliminating clinically relevant inconsistencies** remains an open challenge.
- **Assessing fidelity for clinical use** is still unresolved: current image quality metrics are not fully aligned with diagnostic relevance or clinical decision-making.
- Synthetic data could **balance underrepresented and rare cases** in real datasets, though ensuring their **clinical validity** is critical.
- **Privacy-related similarity metrics** remain limited, often relying on subjective or **arbitrary thresholds**, making robust privacy guarantees difficult to standardize.
- Bridging technical advances with **regulatory frameworks** (AI Act, GDPR) is a key challenge, requiring translation of concepts like fidelity, variability, and privacy into measurable, auditable criteria.

Introduction to FLUTE & synthetic data in healthcare

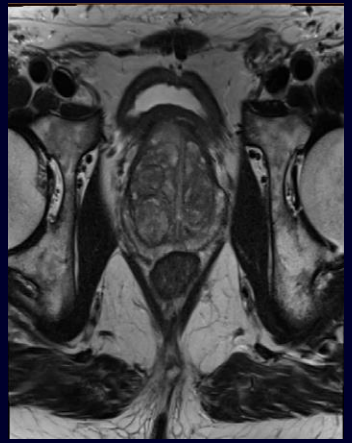
Jan Ramon, INRIA, Lille, France

FLUTE main idea

- Federated (cross border) machine learning:
  - Sufficient training data for high performance
  - Geographical robustness
- Challenges:
  - Privacy-preserving and regulatory compliant
    - No data leaves secured data owner premises in form accesible without data owner collaboration
  - Standardization
    - Standardize/normalize data of different owners

Prostate cancer use case

- Prostate cancer diagnosis:
  - Symptoms? → blood test
  - high PSA? → MRI image
  - clinically significant lesions? → biopsy
- FLUTE objectives:
  - Given: 7 clinical variables + MRI image (= 3D image)
  - Predict: biopsy needed? (and where are lesions to consider?)
  - Generalize: use developed model in other hospitals



Synthetic data

Synthetic data - Introduction

- Why?
  - Train humans (while hiding personal data)
  - Data augmentation: add synthetic data to training set hoping to improve learning process
- What?
  - Synthetic tabular data
  - Synthetic image data
  - Synthetic multi-modal data

Synthetic data - generation

- **UMAP**:
  - repeatedly remove features and impute new values
  - Quite instance-based → to make privacy-preserving federated versión, expensive multi-party computation needed.
- GAN, VAE, **Diffusion models**:
  - Build model to generate data from randomness source
  - Requires lots of training data
- Fusion into multi-modal synthetic data

Synthetic data - validation

- Validation by distribution
  - Is the distribution of properties similar to the distribution of the same properties for the real data?
- Validation by human experts (ongoing)
  - Limited number of synthetic instances can be evaluated
- Data augmentation result: does synthetic data improve prediction performance (ongoing)?

Synthetic data - Privacy

- Synthetic data is not necessarily a solution for privacy
  - Synthetic data generation models can contain sensitive information
- Differential privacy
  - Slightly more noisy due to hiding
  - Relatively straightforward for optizable models (GAN, VAE, diffusion …)
  - Dedicated methods for more instance-based methods (UMAP …)

Alternative: foundational models

- Similar purpose as data augmentation:
  - Gather patterns / information from one dataset and use it to improve training on other (typically more specialized) data
- What?
  - NN predicting pixels from rest of image (Idea as in UMAP: understand object if we can reconstruct missing pixel/values)
- Advantages/disadvantages:
  - Allows for transferring patterns without constraint of being represented as data instances
  - Good experimental performance
  - Less interpretable / can’t represent as data instance

Conclusions

- Synthetic data can help to (a) get a feeling of how data looks like in a privacy-friendly way, (b) improve performance through data augmentation
- Challenge: large amount of training data (and time) needed, especially data is limited in medical settings

THANK YOU

FLUTE project

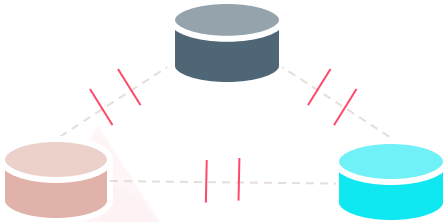
Jan.Ramon@inria.fr

AISym4MED Project
Synthetic data generation and evaluation

Inês Sousa, Fraunhofer Portugal AICOS

Motivation

Healthcare data is the basis of AI in medicine

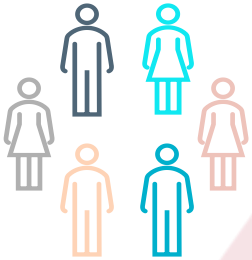


1

- Highly sensitive
- Private
- Scattered or Siloed

2

- Incomplete
- Imbalanced
- Unstructured



3

- Biased
- Underrepresenting minority populations and rare diseases

Needs

Healthcare Data

1

- Privacy-preservation
- Harmonisation
- Standardisation
- Improved quality and completeness

AI-Based Solutions

2

Access to better data:

- Large amounts
- High-quality
- Representative

Real-world validation

Continuous auditing

Improved Care

3

- Reduction of bias
- Increased efficiency
- Knowledge sharing
- Consistency
- Better representation of rare pathologies

Consortium

Strong network of partners

Interdisciplinary

Different backgrounds

Complementary knowledge, skills and resources.

15 partners

8 countries

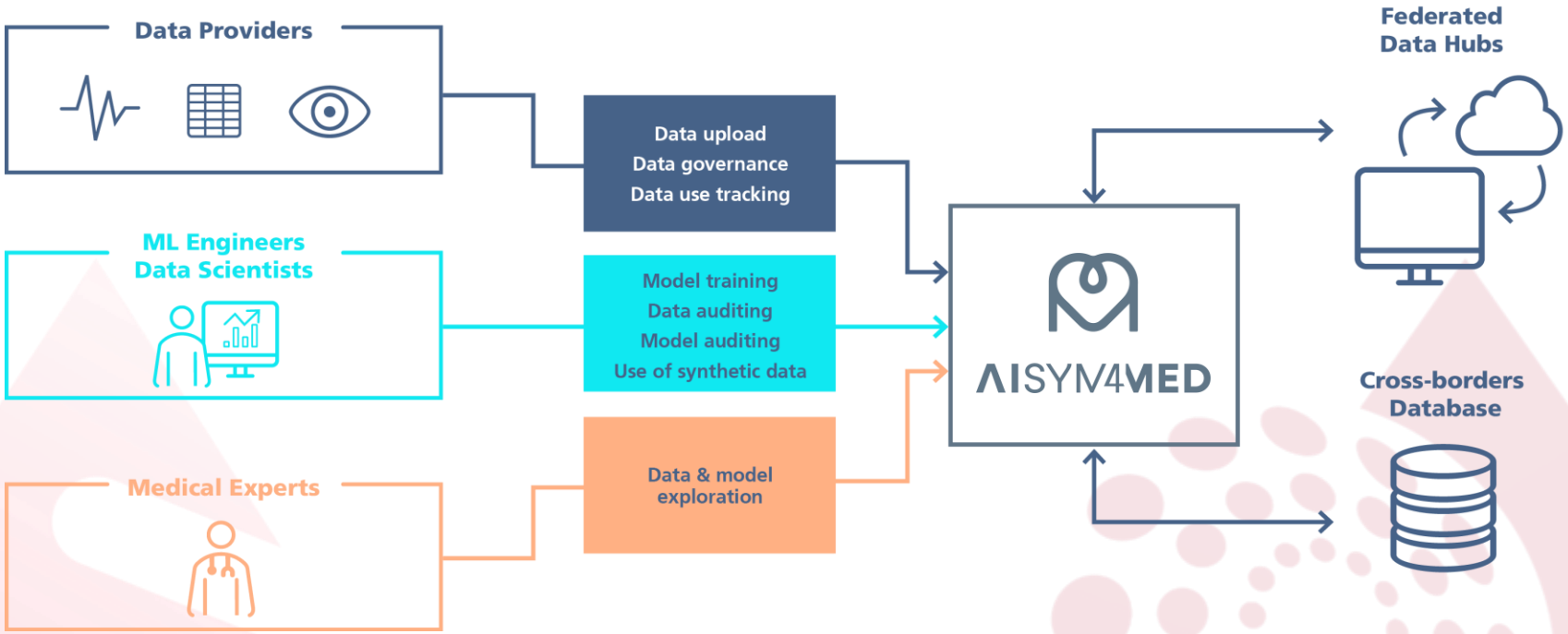
4 years

+6M€
Total budget

eBN | EuroBloodNet

SYNTHEMA

Concept

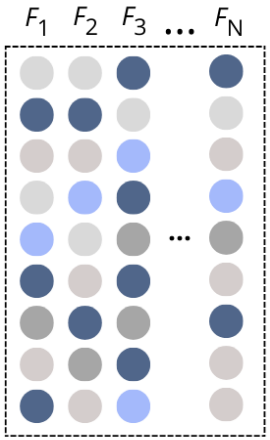

Data Providers
ML Engineers
Data Scientists
Medical Experts
Data upload
Data governance
Data use tracking
Model training
Data auditing
Model auditing
Use of synthetic data
Data & model
exploration
AISYM4MED
Federated
Data Hubs
Cross-borders
Database

Chronic heart and respiratory diseases monitoring

Objectives

Create an ML model to detect decompensation states of patients

Data:



24 Clinical Observations

13 sensor measures
7 patient reports
4 clinical reports

254 Training Samples

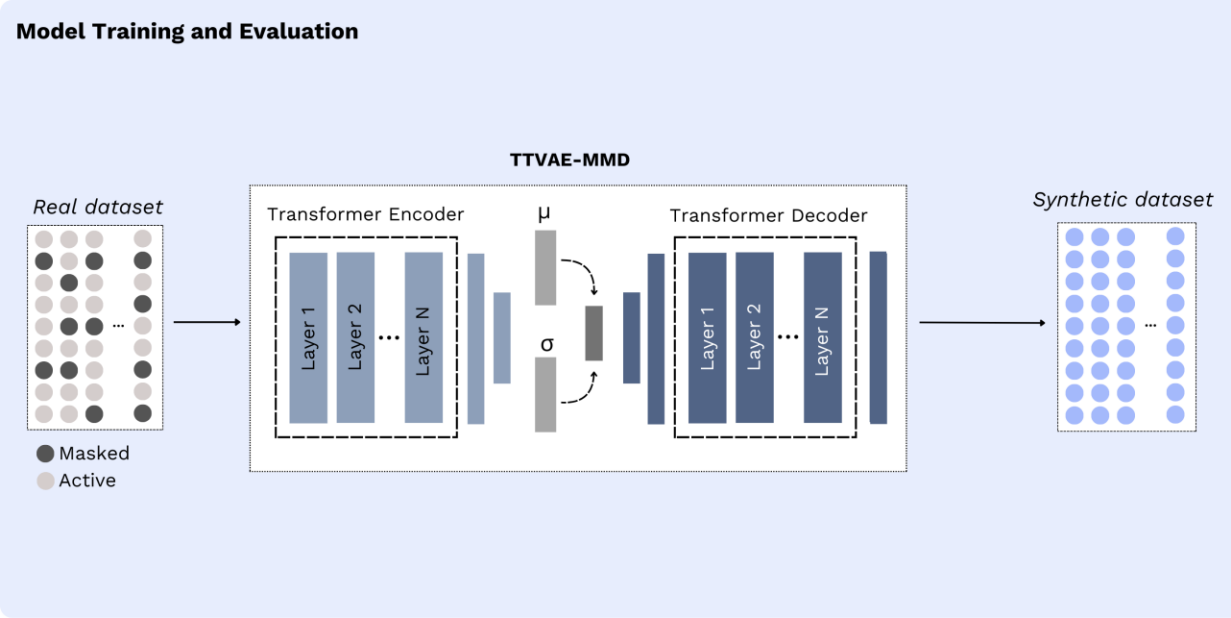
147 Compensated
107 Decompensated

Why synthetic data?

- Data collection is laborious
- Limited dataset size
- Risk of bias and overfitting of ML model
- Improve generalization of ML model

Chronic heart and respiratory diseases monitoring

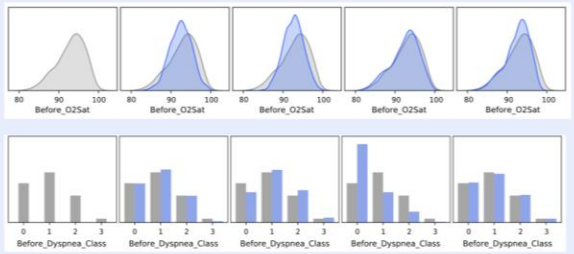
Synthetic Data Generation: TTVAE-MMD



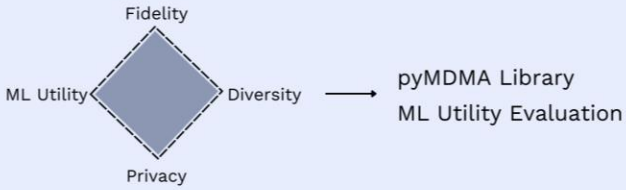
Chronic heart and respiratory diseases monitoring

Synthetic Data Generation: TTVAE-MMD

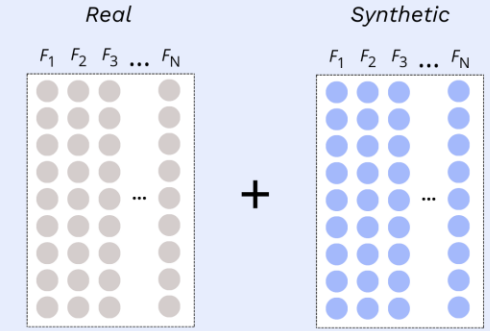
Qualitative Evaluation



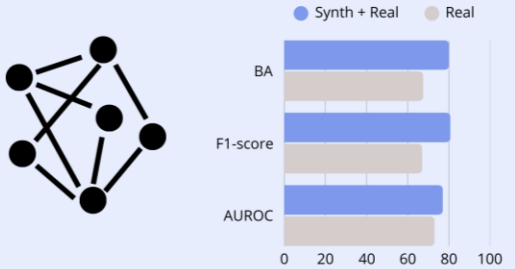
Quantitative Evaluation



ML Utility Evaluation

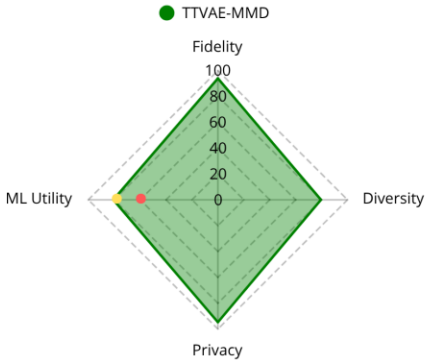


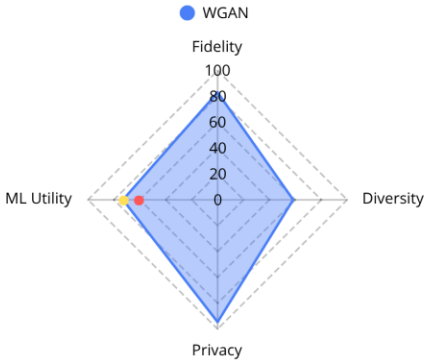
Classification Task

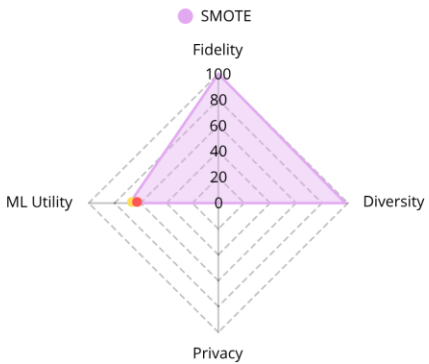


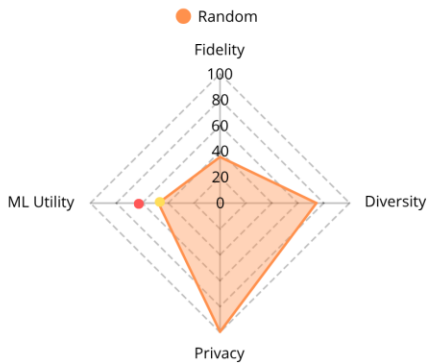
Chronic heart and respiratory diseases monitoring

Synthetic Data Generation: Quantitative Evaluation









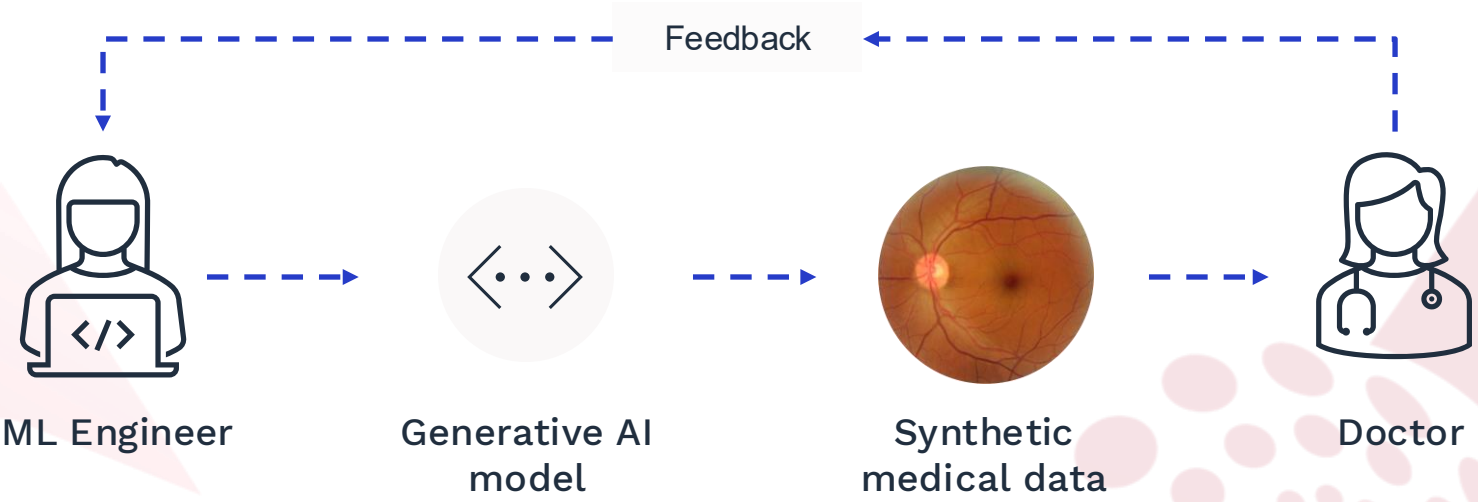
Baseline (Train Real)

Train Real + Synthetic

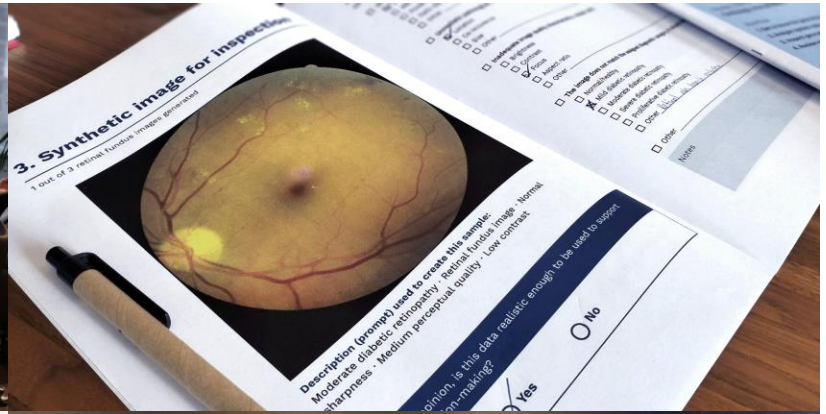
Augmentation factor = 5x

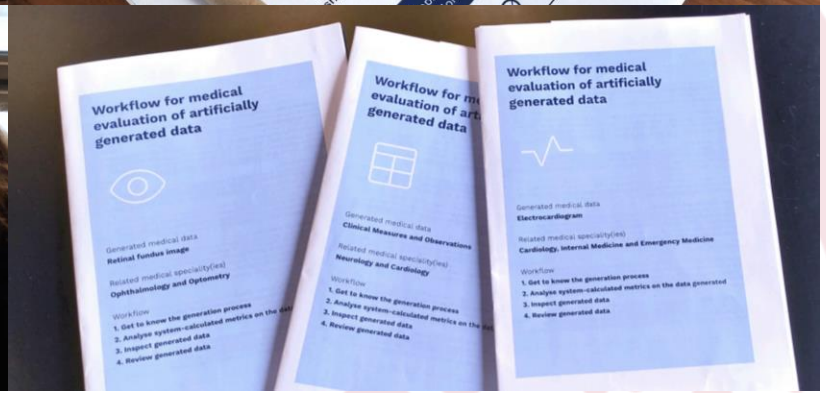
Doctor-in-the-loop

Human Expert evaluation of Synthetic Data



Participatory Design Process





EBN | EuroBloodNet

SYNTHEMA

AISYM4MED

Pages

Home

Help

Logout

Medical data representativeness

How representative is the medical data used for the generation purposes?

- Extremely representative
- Quite representative
- Somewhat representative
- Slightly representative
- Not representative

Comments (optional):

Add comments here...

Submit

Decompensated HF or COPD exacerbation prediction dataset

DOCTOR3

Understand the generation | Analyse metrics | Review synthetic data | Summarize review

Step 1 of 4

Get to know the generation base: Generation purpose, Medical data used, Generative model used.

Consider that limitations in the medical dataset used for generation may be reflected in the generated data, resulting in underrepresentation, bias and shortcomings cases.

Generation purpose

Purpose
The aim is to augment the limited clinical tabular data on decompensated and compensated states in patients with heart failure or chronic obstructive pulmonary disease, based on repeated physiological measurements and clinical observations. Ultimately, the goal is to improve the performance and generalisation of models for predicting cardiac decompensation and pulmonary exacerbation.

Real medical data used

Real dataset information

Real dataset statistical summary

Generative model used

Generative model information

Title	Transformer-based Tabular Variational Autoencoder with Maximum Mean Discrepancy regularization and input masking (TTVAE-MMD)
Summary	The TTVAE-MMD generative model combines the power of self-attention with a MMD regularizer and input masking to encourage both faithful reconstruction and diverse generalization. The architecture was adapted to include input masking. Instead of the traditional Kullback-Leibler divergence term, it was adopted an MMD loss due to its variance stability and posterior collapse mitigation that would, in the end, lead to more informative latent representations and improved generative quality
Generation options available	• Random sampling from the learned latent space. Multiple seeds can produce diverse records. • Conditioning is being integrated but not yet available.
Limitations	• The model may generate clinically implausible records (e.g., distance walked < 0 m, oxygen saturation > 100%). • Privacy risks were mitigated but cannot be fully eliminated. Not guaranteed to generalize to unseen populations.
Biases	• Reflects the biases present in the original dataset, including demographic and measurement biases, which may influence generated outputs.

AISYM4MED

Pages

Home

Help

Logout

Medical data representativeness

How representative is the medical data used for the generation purposes?

- Extremely representative
- Quite representative
- Somewhat representative
- Slightly representative
- Not representative

Comments (optional):

Add comments here...

Submit

Filter by review status

All | Pending | Realistic | Not realistic

Selected: 0 • Pending: 324 • Realistic: 4 • Not realistic: 1

Select one or more rows to add your review.

#	Diagnosis	Disease phase	Before - DBP	Before - SBP	Before - HR	During - HR	After - HR	Before - Ox	During - Ox	Review
1	2-COPD	1-Decompe...	67	122	79	94	72	95	97	Not realistic
2	1-Heart failu...	1-Decompe...	64	106	82	83	83	98	95	Realistic
3	1-Heart failu...	1-Decompe...	80	127	88	88	79	95	97	Realistic
4	3-Both	1-Decompe...	48	123	74	77	82	94	90	Realistic
5	2-COPD	0-Compens...	73	137	73	79	78	96	101	Pending
6	1-Heart failu...	1-Decompe...	83	96	92	110	95	91	91	Realistic
7	1-Heart failu...	1-Decompe...	66	118	88	94	86	93	94	Pending
8	1-Heart failu...	1-Decompe...	83	107	68	90	72	90	99	Pending
9	2-COPD	1-Decompe...	64	94	92	108	89	93	90	Pending
10	1-Heart failu...	0-Compens...	54	124	65	86	64	96	99	Pending
11	2-COPD	1-Decompe...	77	117	94	96	85	92	85	Pending
12	1-Heart failu	1-Decompe	60	89	77	85	77	90	88	Pending

'Not realistic' data feedback

Data reviewed as not realistic and the feedback provided

Filter by column

Choose options

Selected rows	Column	Issue(s)	Comment	Timestamp
1	Before - DBP	Unnatural or erratic data fluctuations, Unrealistic variable distribution	test	10 Apr 2026, 10:20:02
1	Disease phase	Unnatural or erratic data fluctuations, Unrealistic variable distribution	test	10 Apr 2026, 10:20:02

Save comment changes

AISYM4MED

Pages

Home

Help

Logout

Medical data representativeness

How representative is the medical data used for the generation purposes?

- Extremely representative
- Quite representative
- Somewhat representative
- Slightly representative
- Not representative

Comments (optional):

Add comments here...

Submit

Filter by review status

All | Pending | Realistic | Not realistic

Selected: 1 • Pending: 324 • Realistic: 4 • Not realistic: 1

Selected Row IDs: [7]

Realistic

Feedback on not realistic

Issue type(s)

Choose options

- Select all
- Biologically implausible values
- Unnatural or erratic data fluctuations
- Lacks internal coherence or consistency
- Unrealistic or impossible intervals
- Unrealistic variable distribution
- Inconsistent formatting/units

Save feedback

#	Diagnosis	Disease phase	Before - DBP	Before - SBP	Before - HR				
1	2-COPD	1-Decompe...	67	122	79				
2	1-Heart failu...	1-Decompe...	64	106	82				
3	1-Heart failu...	1-Decompe...	80	127	88				
4	3-Both	1-Decompe...	48	123	74				
5	2-COPD	0-Compens...	73	137	73				
6	1-Heart failu...	1-Decompe...	83	96	92				
7	1-Heart failu...	1-Decompe...	66	118	88	94	86	93	94	Pending
8	1-Heart failu...	1-Decompe...	83	107	68	90	72	90	99	Pending
9	2-COPD	1-Decompe...	64	94	92	108	89	93	90	Pending
10	1-Heart failu...	0-Compens...	54	124	65	86	64	96	99	Pending
11	2-COPD	1-Decompe...	77	117	94	96	85	92	85	Pending
12	1-Heart failu	1-Decompe	60	89	77	85	77	90	88	Pending

'Not realistic' data feedback

Data reviewed as not realistic and the feedback provided

Filter by column

Choose options

Selected rows	Column	Issue(s)	Comment	Timestamp
1	Before - DBP	Unnatural or erratic data fluctuations, Unrealistic variable distribution	test	10 Apr 2026, 10:20:02
1	Disease phase	Unnatural or erratic data fluctuations, Unrealistic variable distribution	test	10 Apr 2026, 10:20:02

Save comment changes

Thanks!

Any questions?

Keep in touch!

eurobloodnet.eu /ERNEuroBloodNet @ERNEuroBloodNet @erneurobloodnet.bsky.social

synthema.eu /synthema @SYNTHEMA_EU @synthemaeu.bsky.social

Funded by
the European Union