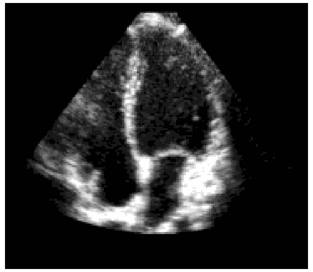
MICCAI 2026

MultiTab 2026

Beyond Echocardiography: Toward Clinical Diagnosis Through Multimodal Fusion

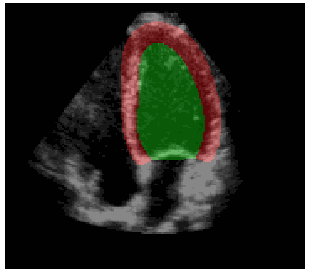
Prof. Olivier Bernard
CREATIS Lab., University of Lyon, France, IUF member

Echocardiography



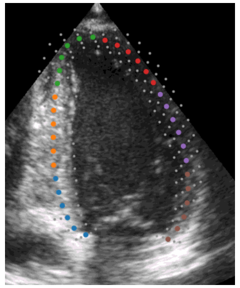
Unique imaging modality

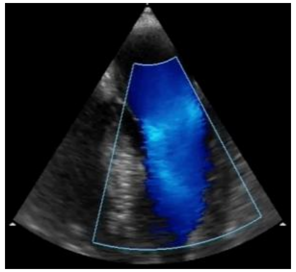
- Accessibility
- Non-invasive / no-radiation
- Real-time imaging
- Bedside applicability



Multimodality

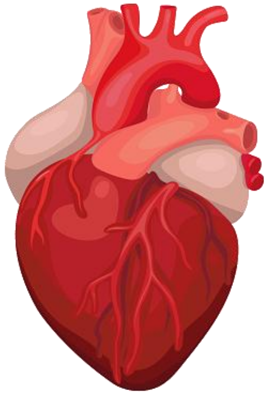
- B-mode sequences
- Doppler sequences
- ECG
- Report

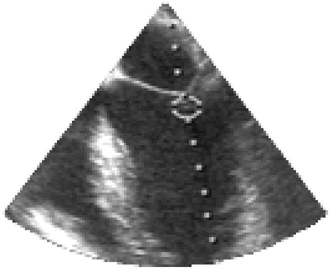




Annotations

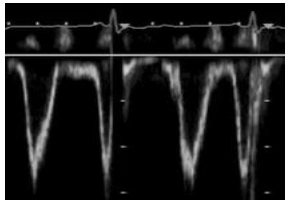
- Interobserver variability





Scarcity

- Annotations
- Sequence availability



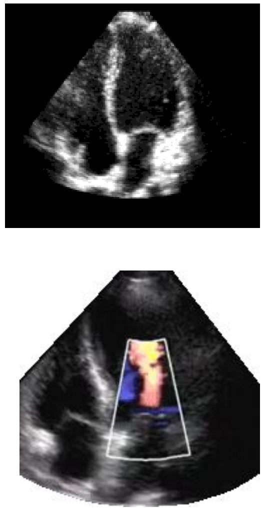
Domain shifts

- Multicentre
- Multivendor
- Acquisition settings

Echocardiography - Intense research in multimodal integration

[Holste et al., medRxiv*, 2024]* *[Christensen et al., Nature Medicine, 2024]* *[Vukadinovic et al., Nature, 2025]*

[Kim et al., IEEE TMI, 2025] *[Heidari et al., arXiv, 2026]* *[Qin et al., arXiv, 2026]*



Ultrasound images

Multimodal integration → Automatic diagnosis

Reports

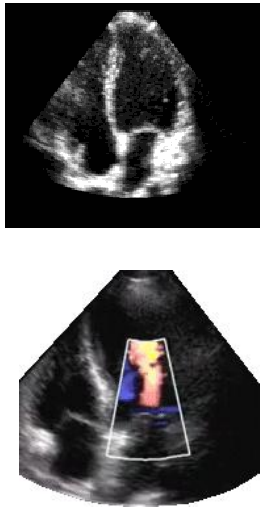
Lab tests

EHR (tabular data)

"This is a patient with hypertrophy because […]"

Echocardiography - Report generation ▸ EchoPrime

[Vukadinovic et al., Nature, 2025]



Ultrasound images

Multimodal integration

Reports
(during training)

Generation of a report
at inference

“This is a patient with hypertrophy because […]”

Echocardiography - Report generation ▸ EchoPrime

Videos / report alignment

Dataset

- 109,000 patients
- 275,000 exams
- 12,1 M videos

Model

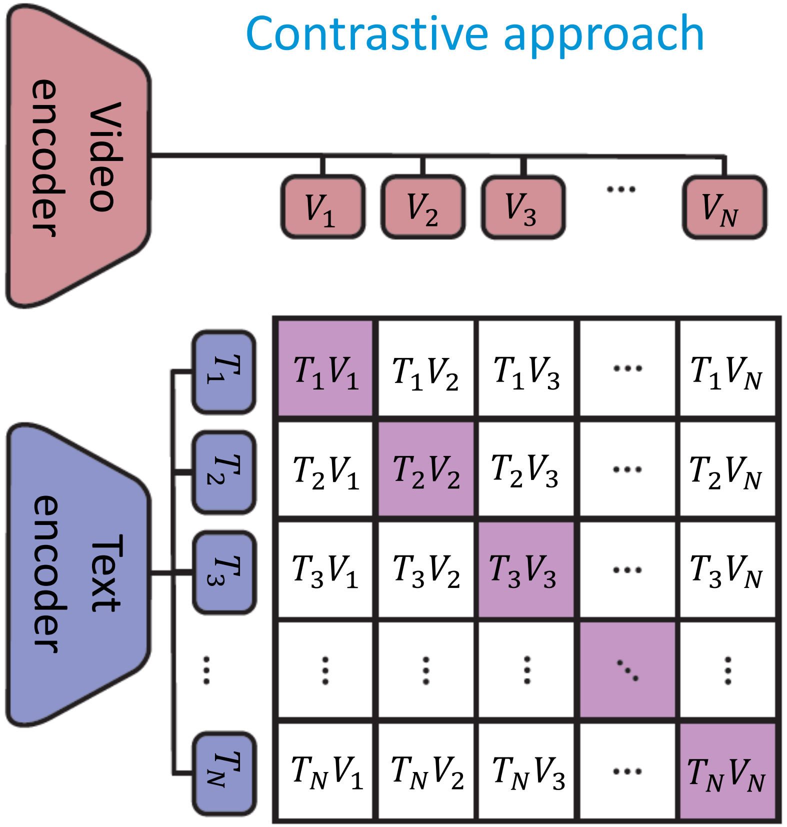
- Video encoder
- Text encoder
- View classification
- Cross-modal retrieval

Left ventricle:
LV ejection fraction 60%
Mid diastolic dysfunction

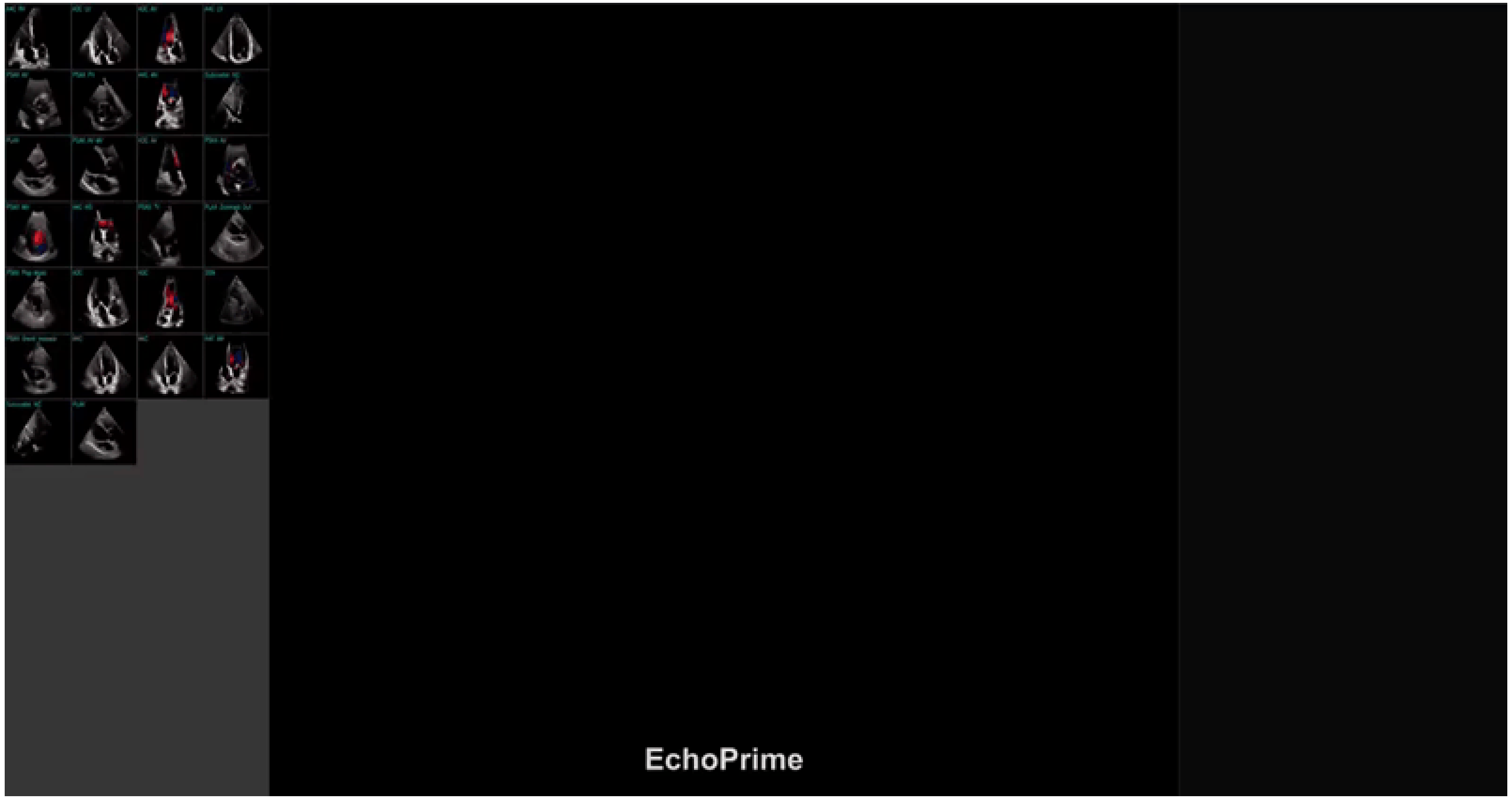
Aortic valve:
A bioprosthetic valve is present

Pericardium:
Small pericardium effusion

CLIP-based method
Contrastive approach



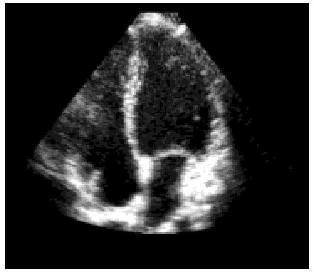
Echocardiography - Report generation ▸ EchoPrime



End of the story ? – Limitations of such approaches

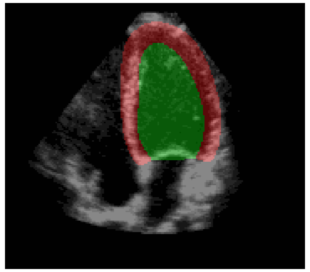
- Lack of visual evidence
- Uncertainty not considered
- Limited explainability
- No interpretability

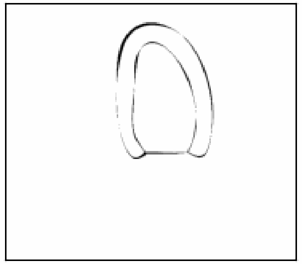
B-mode only



“This patient has an ejection fraction of 60%, trust me”

All we need is transparent and trustworthy AI models



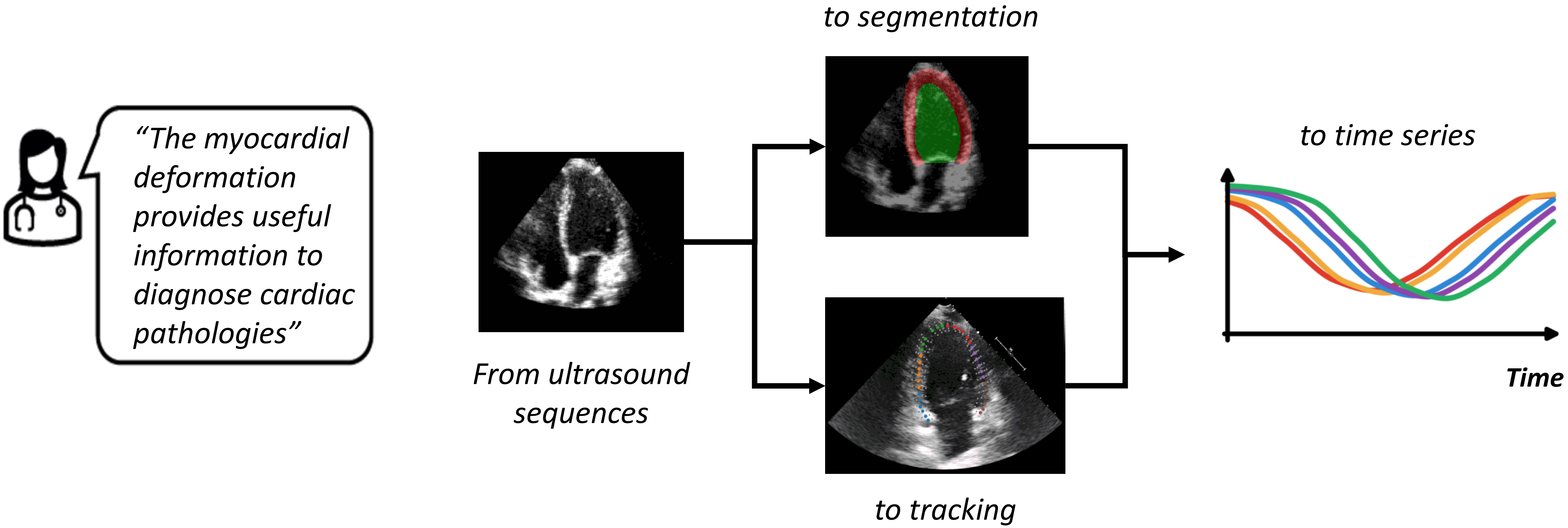


“This patient has an ejection fraction of 60%, trust me”

B-mode with segmentation & uncertainty

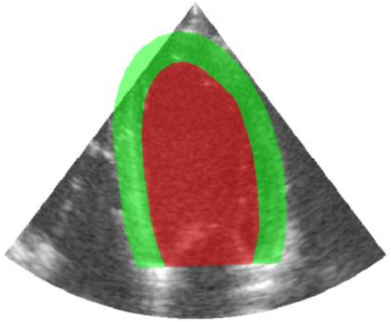
Approach considered – Maximizing the use of clinical knowledge

- Reduce the need to train from huge datasets and models
- Reinforce explainability et interpretability



On going work – Automatic extraction of time series from echocardiography

▸ Currently focusing on the left ventricle (LV) and the myocardium



Segmentation of cardiac structures

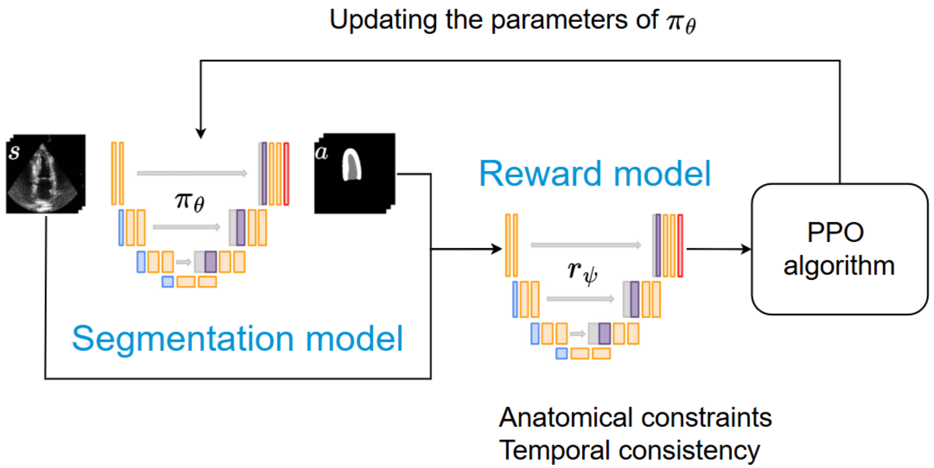
- Robustness: domain adaptation
- Trustworthy: uncertainty estimation
- Time series:
  - Surface of the LV
  - LV length along its main axis
  - Regional myocardial thickness

Tracking of cardiac structures

- Robustness: realistic synthetic cohorts
- Time series:
  - Global LV strain
  - Regional LV strain

Echocardiography - Segmentation ▸ domain adaptation

▸ Reinforcement learning

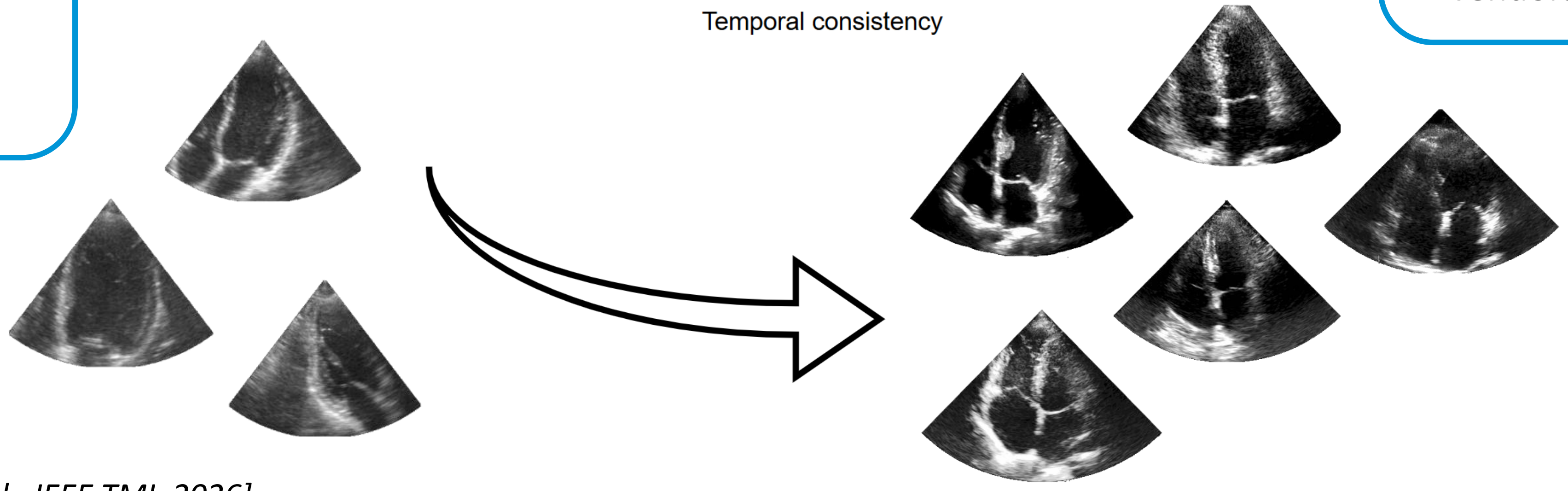


CAMUS dataset

- 500 patients
- Annotations
- Mono center / vendor

Private dataset

- 25,000 patients
- No annotation
- Multi centers / vendors



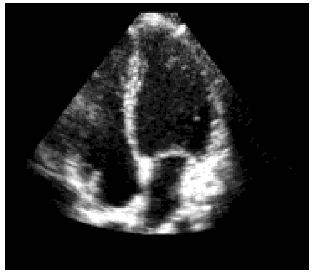
[A. Judge et al., IEEE TMI, 2026]

Echocardiography - Segmentation ▸ uncertainty estimation

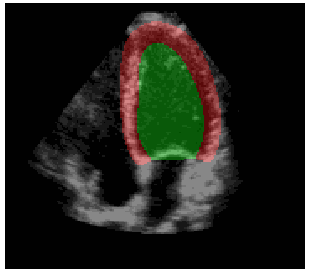
- Reinforcement learning

Successful case →

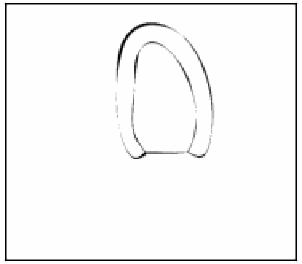
Input B-mode



Segmentation model π_θ

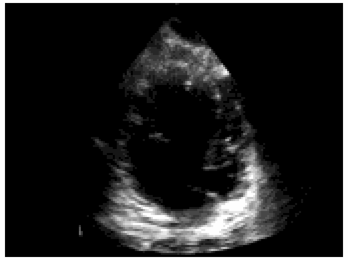


Reward / Uncertainty model r_ψ



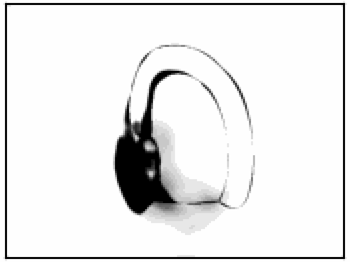
Detected failure case with correction →

Input B-mode



Segmentation model π_θ

Reward / Uncertainty model r_ψ



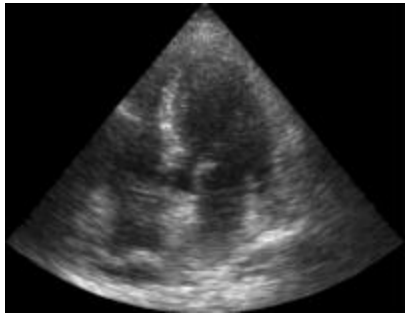
[A. Judge et al., MICCAI, 2024]

Echocardiography - Image quantification ▸ robust strain estimation

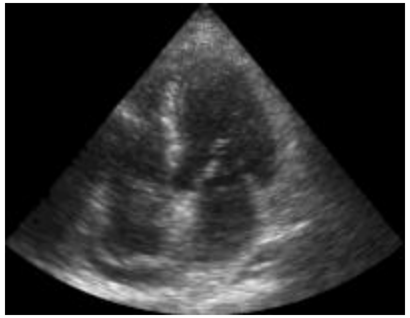
▸ Simulation pipeline for realistic synthetic cohort

Inference: real B-mode

Real B-mode sequence

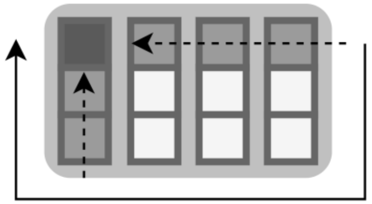


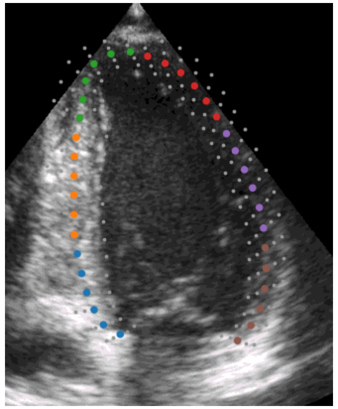
Synthetic sequence using SIMUS physical simulator

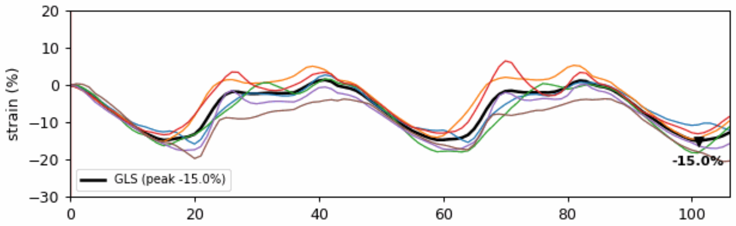


TAS-Net

Tracking trajectories
Transformer: cross-track / time attention







TAS-1K dataset

1478 synthetic sequences with motion reference

[T. Judge et al., arxiv, 2026]

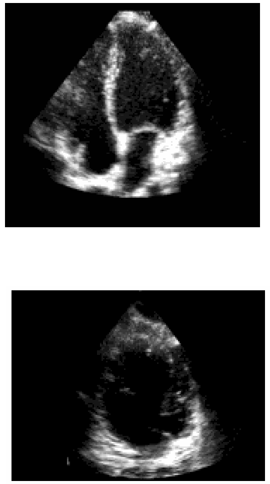
[T. Judge et al., STACOM, 2026]

On going work – Fusion of echocardiographic time-series with patient data

On a better characterization of patients with arterial hypertension (HT)

CARDINAL private dataset

- 239 patients
- Mono center / vendor
- US sequences + EHR
- 3 discrete outputs
  - White coat HT (#21)
  - Controlled (#140)
  - Uncontrolled (#78)

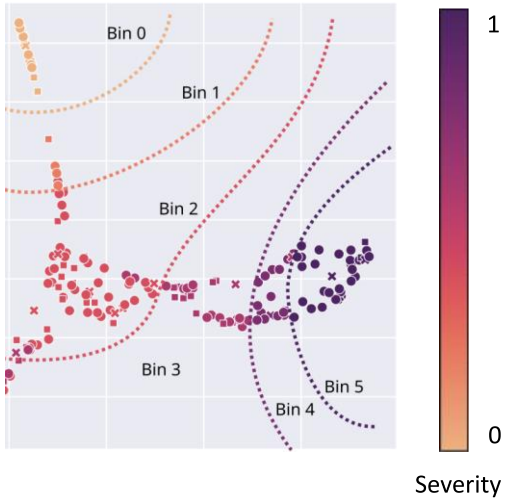


US sequences
(time-series)

Multimodal integration

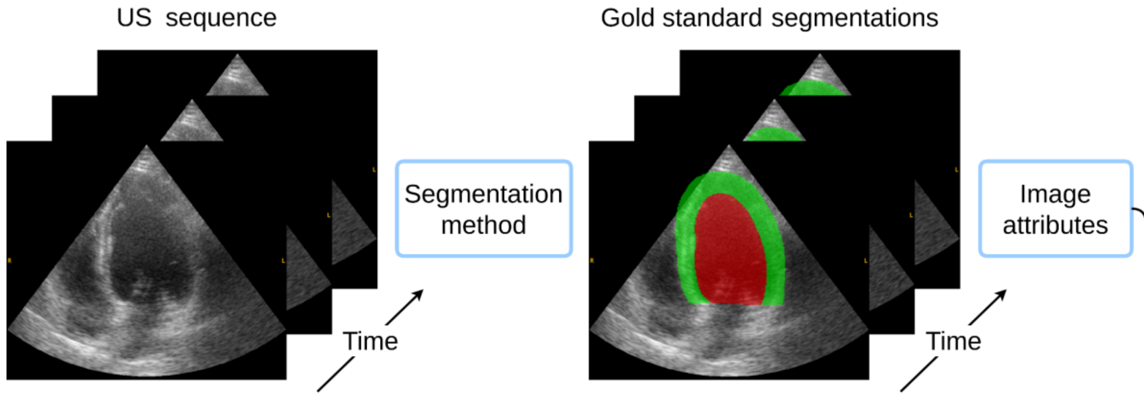
HT stratification

EHR
(tabular data)



[N. Painchaud et al., IEEE TUFFC, 2025]

HT stratification – General overview

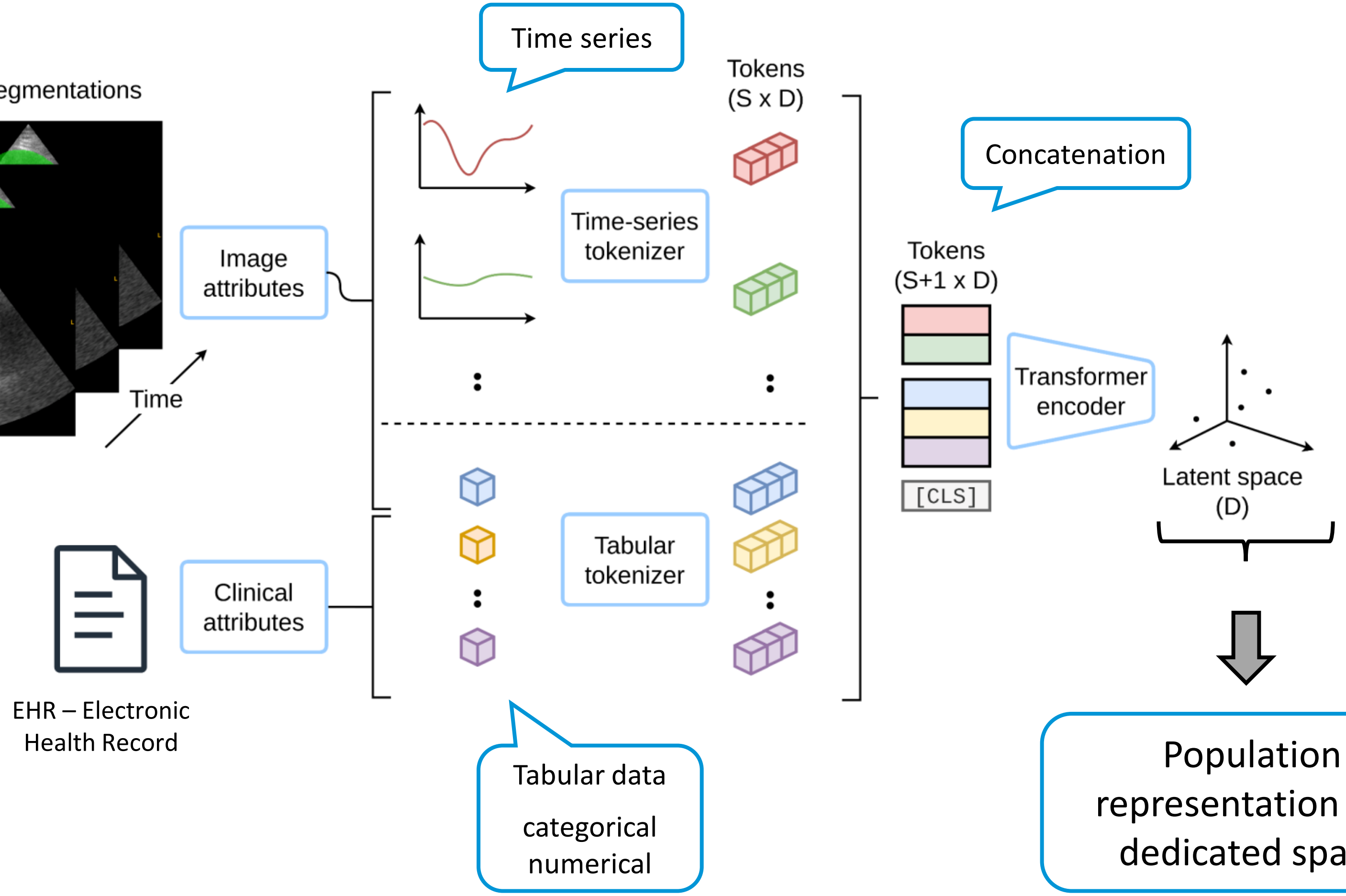


Time series (#14)

- LV surface
- LV main axis length
- Global & local strain
- Local myo. thickness

EHR (#28 tab. data)

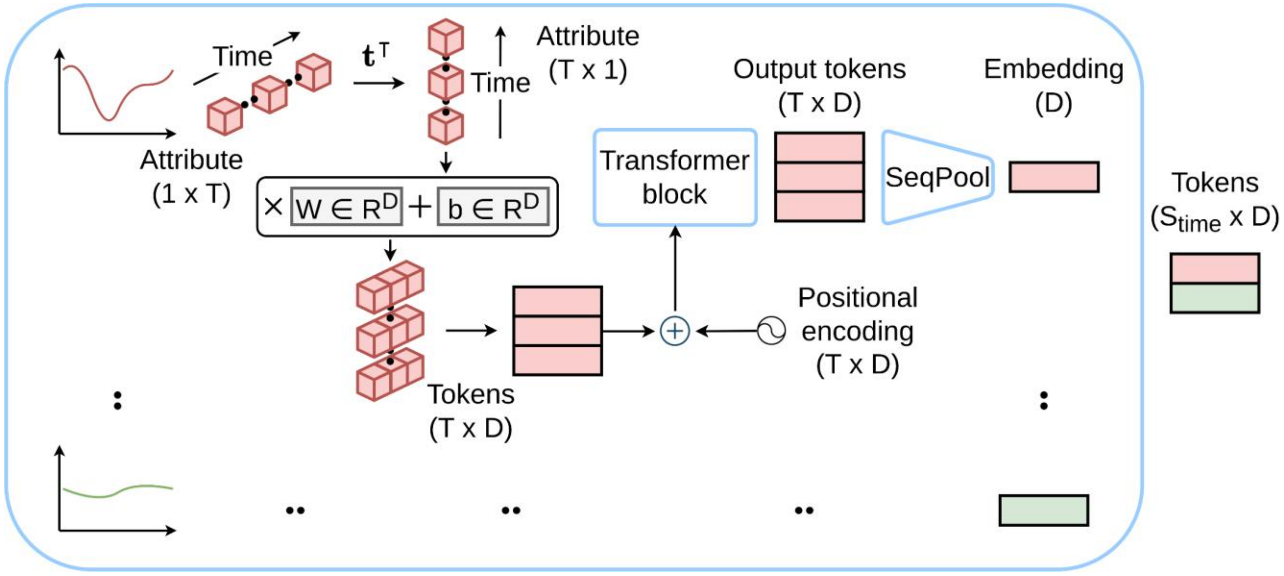
- Demographic info.
- Medical history
- Holter monitor
- Biological exam
- Treatments



Population representation in a dedicated space

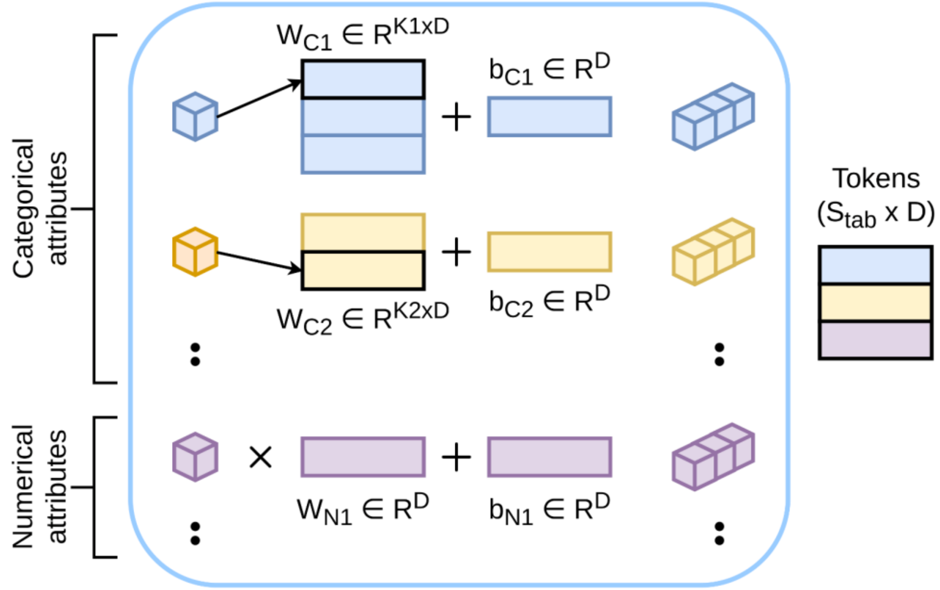
HT stratification – Tokenizers

Time series tokenizer



[Pellegrini et al., MEDIA, 2023]

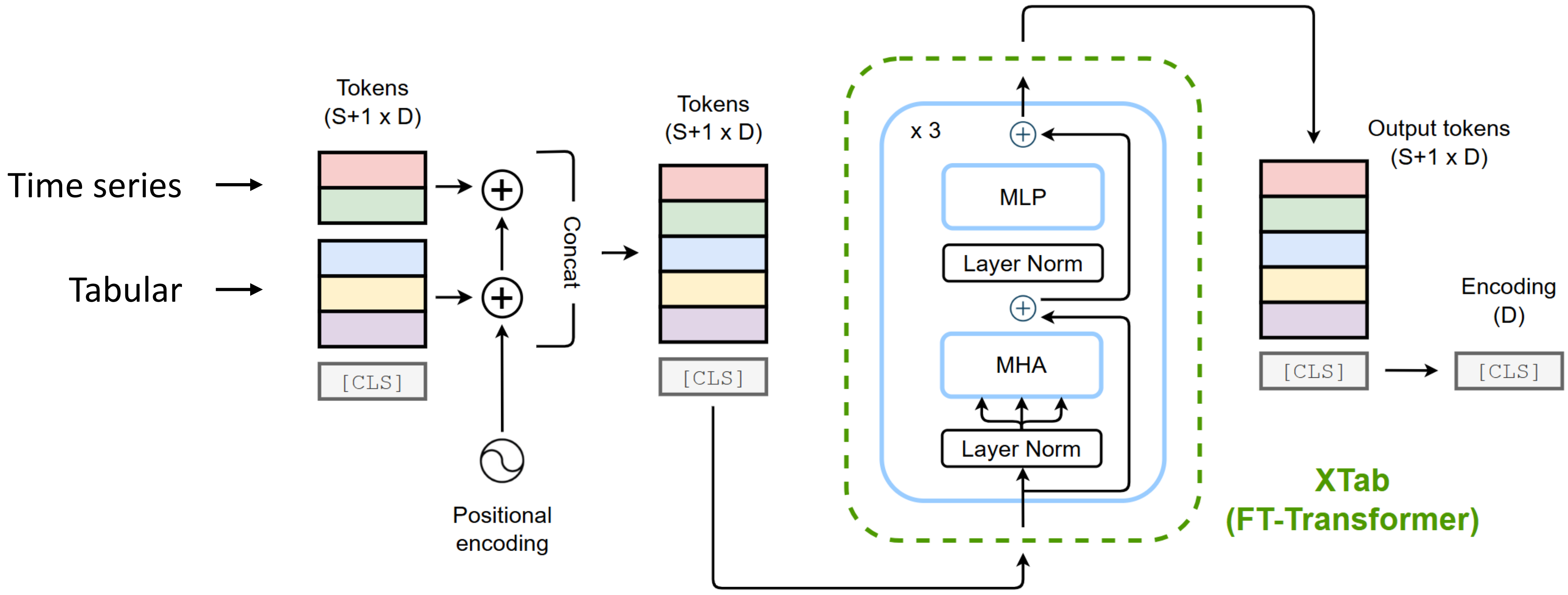
Tabular tokenizer



[Gorishniy et al., NeurIPS, 2021]

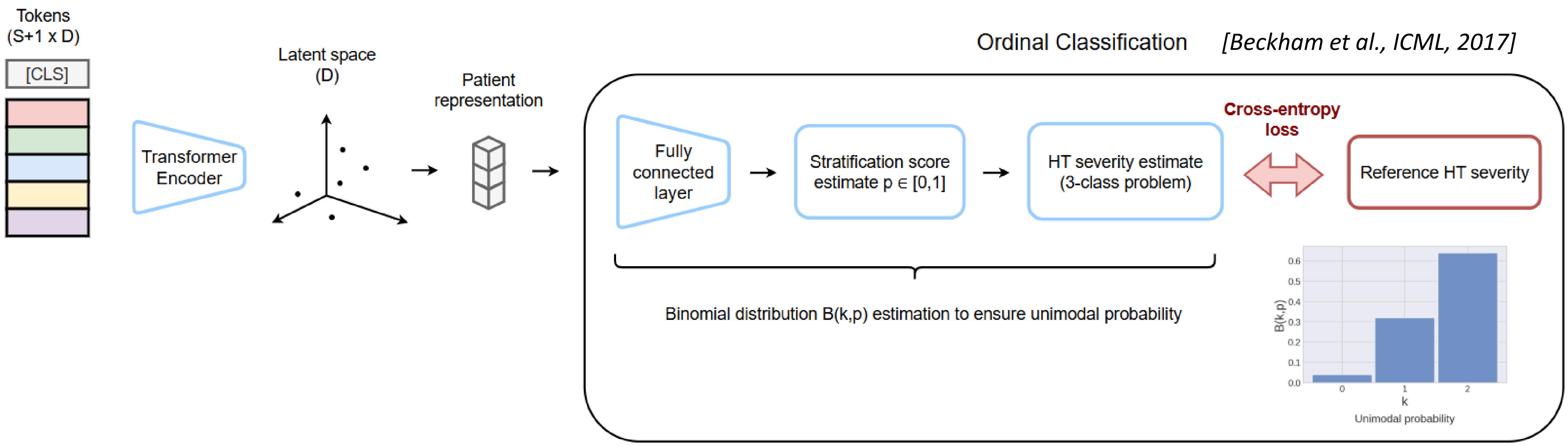
HT stratification – Transformer encoder

- XTab tabular foundation model built on FT-Transformer backbone

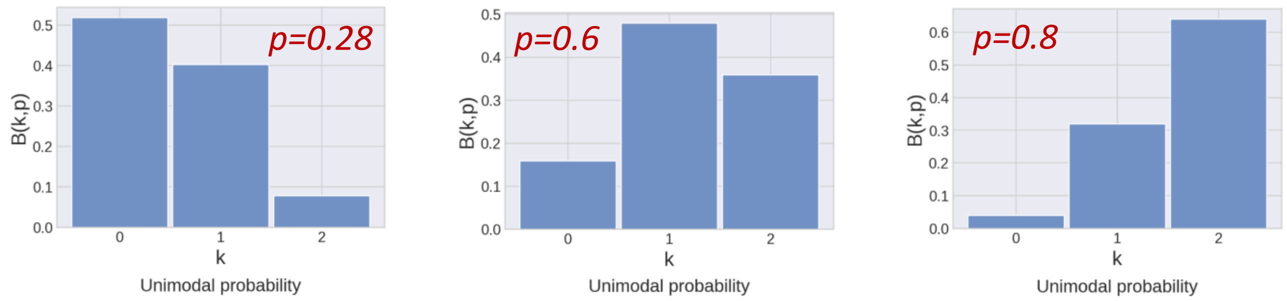


[Zhu et al., ICML, 2023]

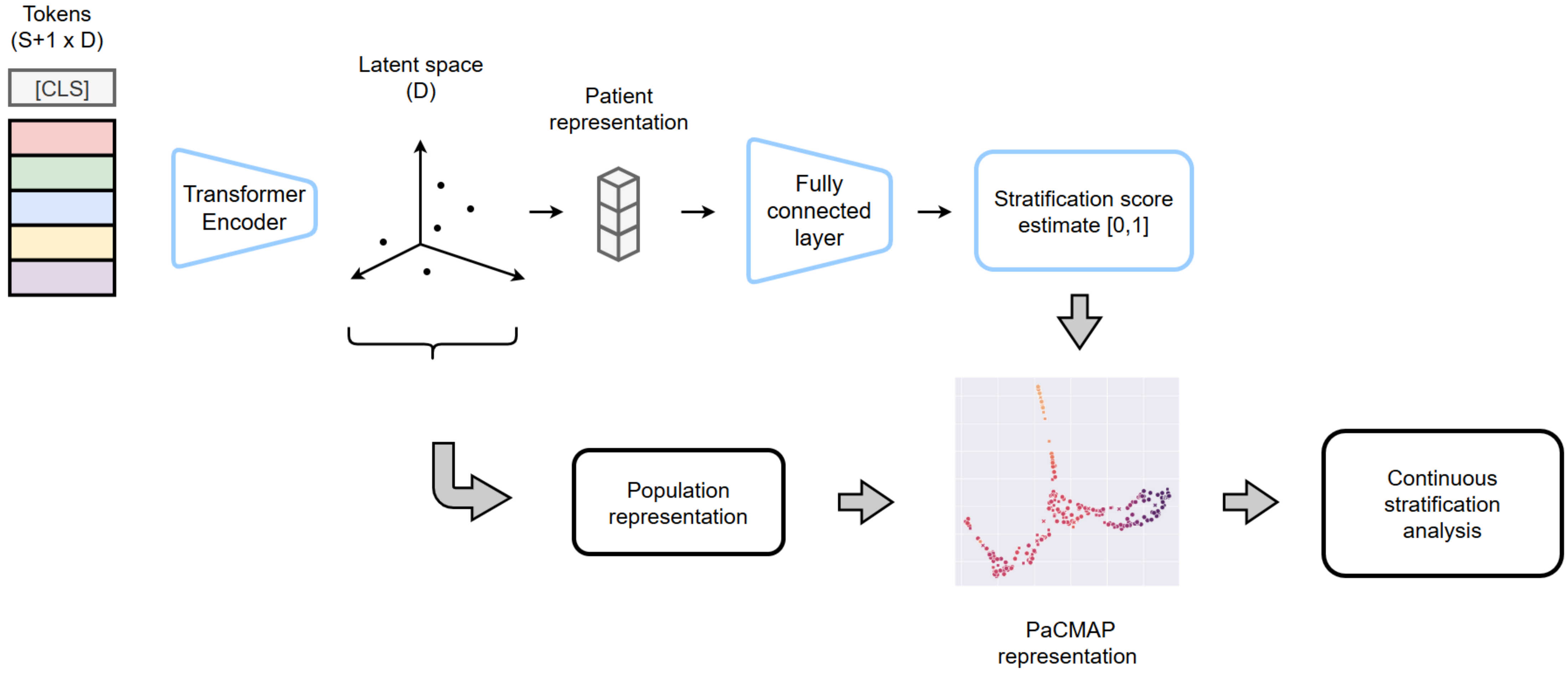
HT stratification – Ordinal classification



Unimodal probability distribution



HT stratification – Inference



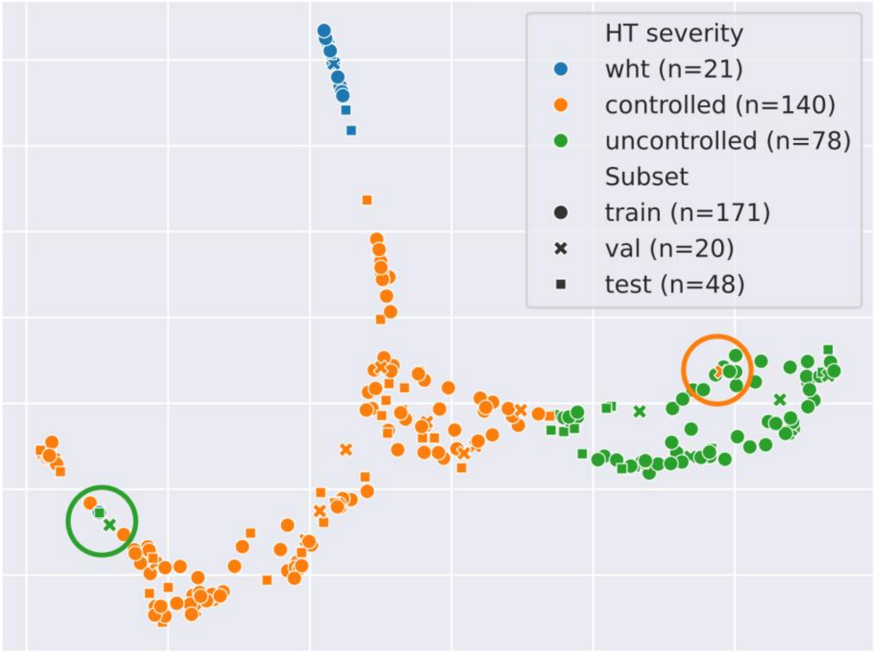
HT stratification – Classification performance

- 5-fold cross validation

XTab	tabular+time-series	# param.
Accuracy (%)	**96.8 ± 3.1**	1.1 M

Best classification accuracy using the 28 clinical descriptors + 14 time series descriptors

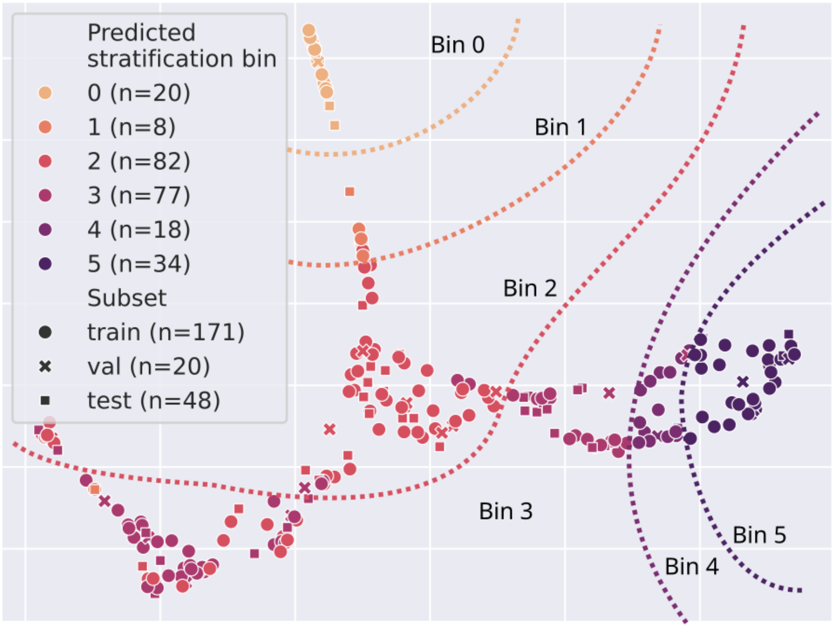
Holter monitor (24 hours) is key for high classification accuracy



Population representation: coloration according to HT severity

HT stratification – Phenogroups analysis

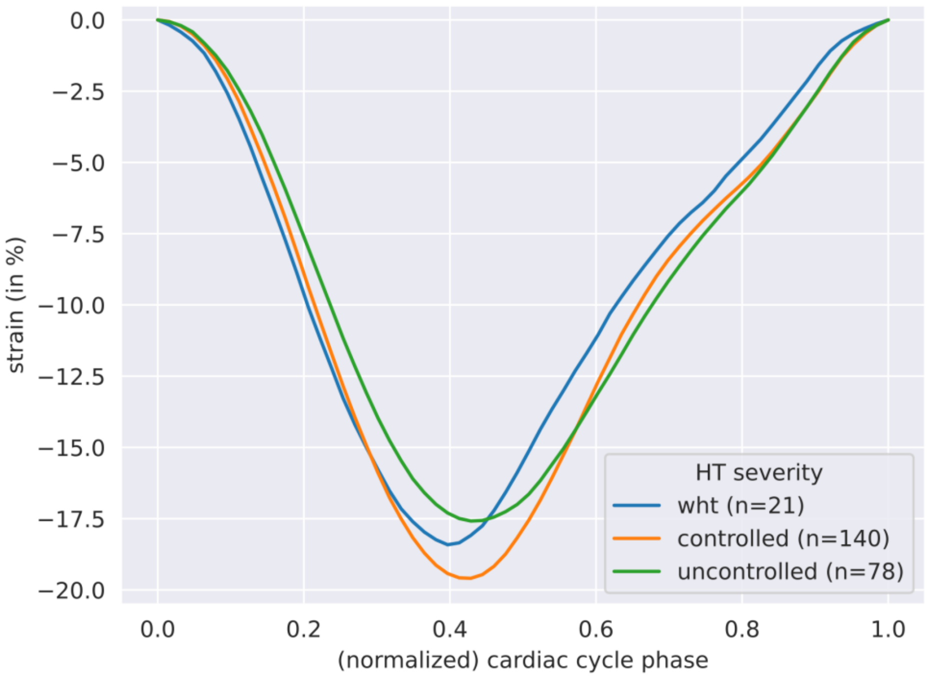
- Discretization of the continuous stratification score to define ordinal phenogroups



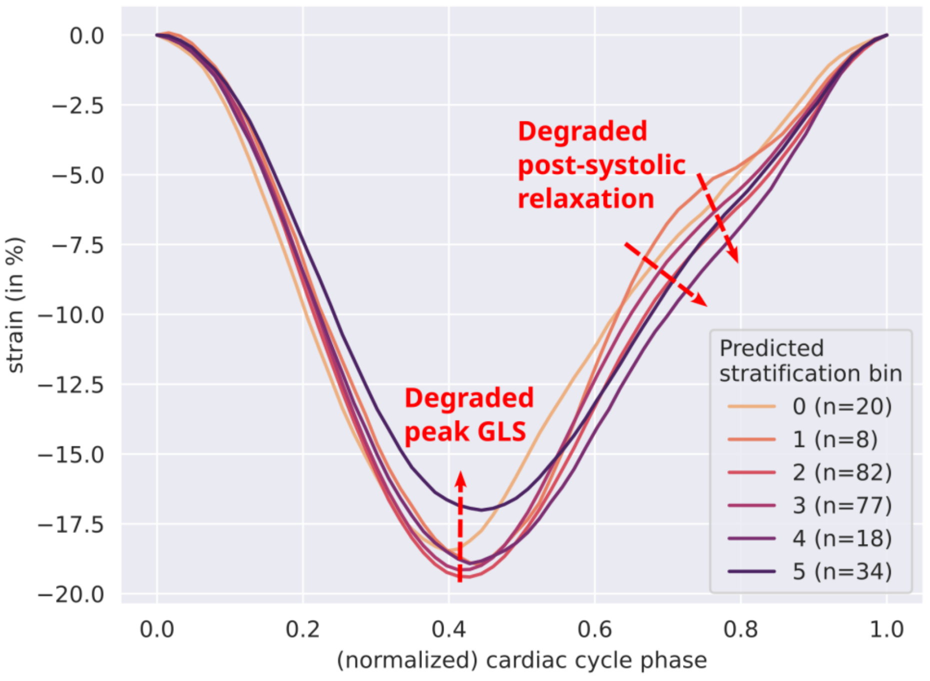
Discretization with arbitrary 6 bins

HT stratification – Phenogroups analysis

- Post-systolic relaxation: a potential marker to assess the progression of hypertension



Patient groups according to the 3-class severity score



Ordinal phenogroups

HT stratification – Conclusions

- ✓ Pilot study on the integration of time series with patient data
- ✓ Derivation of a potential marker for the assessment of HT
- ✓ Shows stability in the results despite the use of a small dataset

- ✗ Evaluation on a mono centric / mono vendor small cohort (238 patients)
- ✗ No time series from hemodynamic imaging (e.g. Doppler)
- ✗ Simple multimodal fusion scheme (concatenation)
- ✗ Explainability and robustness remain key challenge
- ✗ Clinical adaptation remains essential

On going work – Fusion of echocardiographic time-series with patient data

Asymmetric fusion of tabular and echo. data for cardiac hypertension diagnosis

- A 3-category HT classification problem
- Without the use of the Holter monitor (24 hours)

Cardinal internal dataset

Time series (#14)
- LV surface
- LV main axis length
- Global & local strain
- Local myo. thickness

EHR (#13 tab. data)
- Echo. Measures (#10)
- Biological exam (#1)
- Demographic info (#1)
- Treatment (#1)

3 classes
- White coat HT
- Controlled
- Uncontrolled

Input data

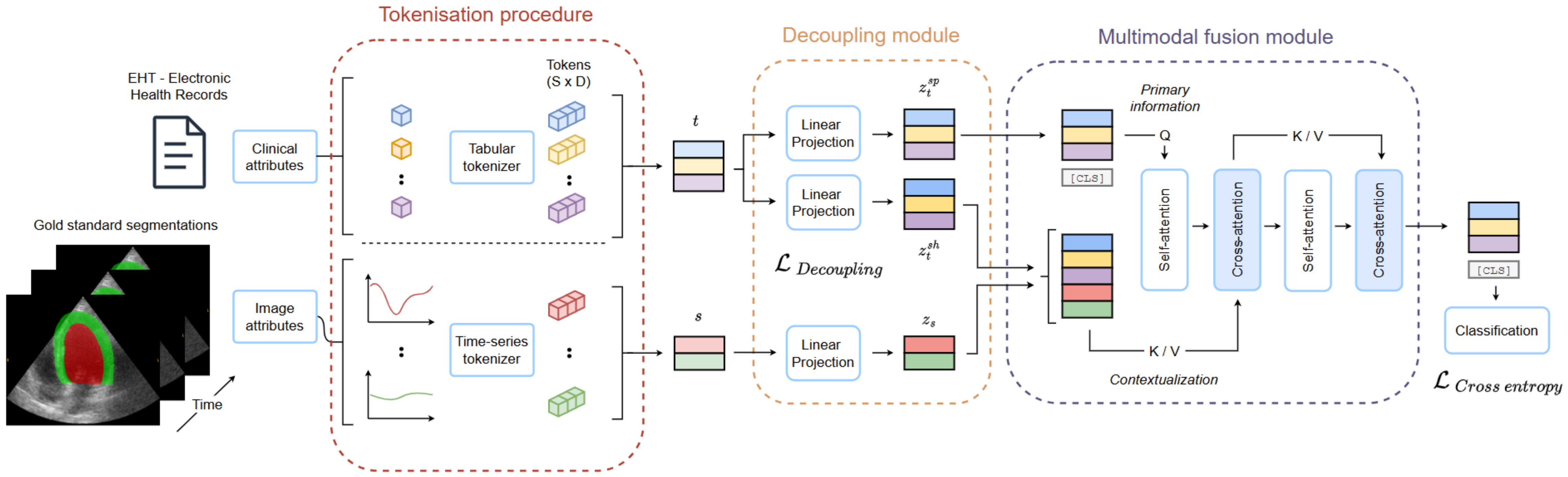
Output data

Classification performance

Model	ROC AUC	# parameters
Unimodal models		
XGBoost	87.4	N/A
FT-T Tab	85.8 ± 4.8	863K
FT-T TS	52.2 ± 2.3	1.0M
Multimodal fusion models		
XGBoost	79.7	N/A
MLP	81.8 ± 1.3	391K
IRENE	86.7 ± 2.8	102.9M
FT-T	88.9 ± 1.4	1.1M

- Unbalanced information contribution

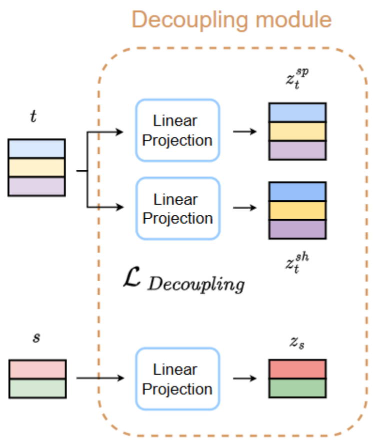
HT classification – General overview

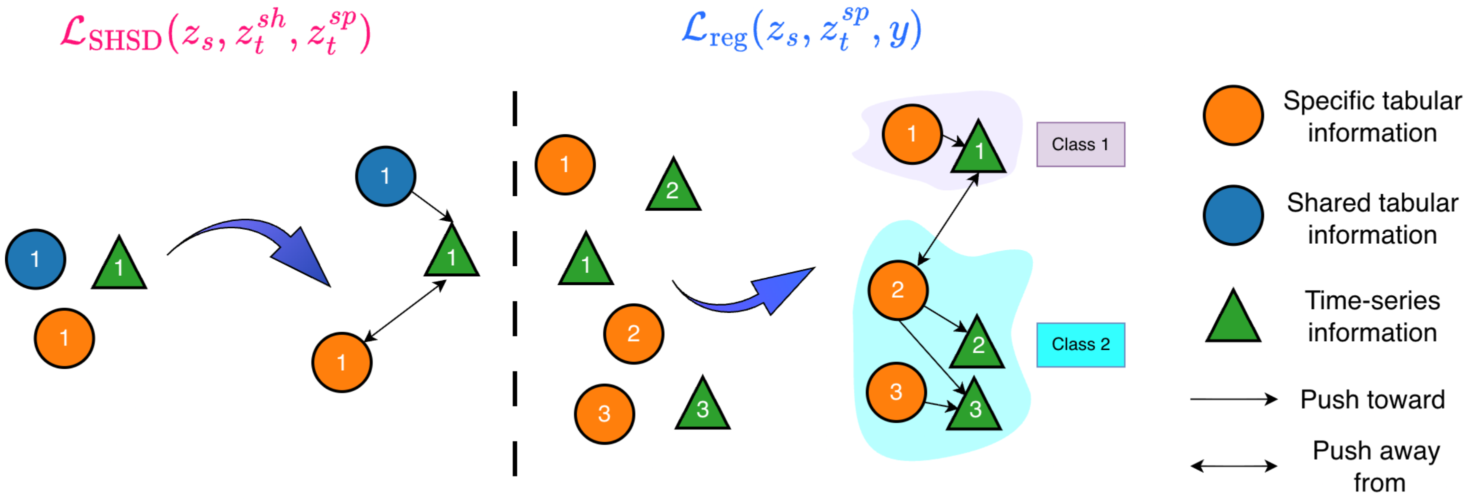


[J. Stym-Popper et al., MIDL, 2025]

HT classification – Information decoupling

- $\mathcal{L}_{Decoupling} = \mathcal{L}_{SHSD}(z_s, z_t^{sh}, z_t^{sp}) + \mathcal{L}_{reg}(z_s, z_t^{sh}, y)$

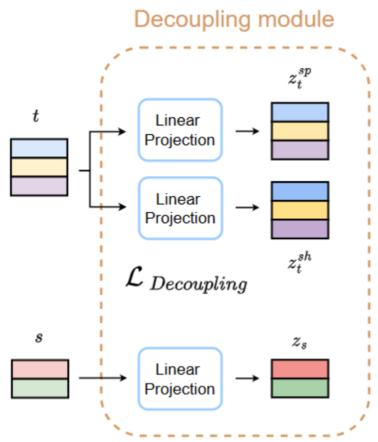




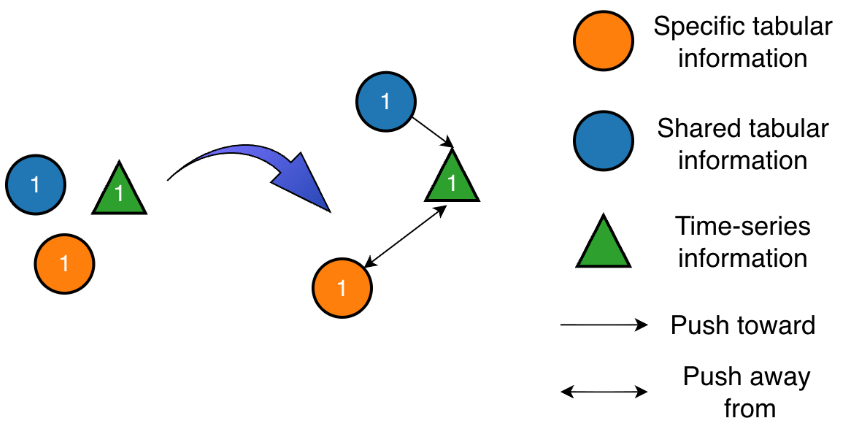
- z_s time series token
- z_t^{sh} share tabular token
- z_t^{sp} specific tabular token
- y patient label

HT classification – Information decoupling

- $\mathcal{L}_{Decoupling} = \mathcal{L}_{SHSD}(z_s, z_t^{sh}, z_t^{sp}) + \mathcal{L}_{reg}(z_s, z_t^{sh}, y)$



$\mathcal{L}_{\mathrm{SHSD}}(z_s, z_t^{sh}, z_t^{sp})$



Contrastive loss

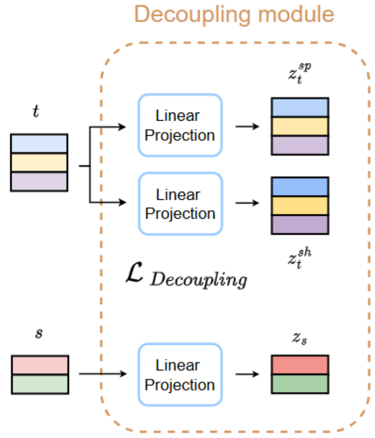
$$\mathcal{L}_{\mathrm{SHSD}}(\boldsymbol{z}_s, \boldsymbol{z}_t^{sh}, \boldsymbol{z}_t^{sp}) = \frac{1}{2N} \sum_{i=1}^{N} \left(l_i^{s,t} + l_i^{t,s} \right)$$

$$l_i^{s,t} = -\log \left(\frac{\exp\{\mathrm{sim}(\boldsymbol{z}_{s_i}, \boldsymbol{z}_{t_i}^{sh})/\tau\}}{\sum_{k=1}^{N} \exp\{\mathrm{sim}(\boldsymbol{z}_{s_i}, \boldsymbol{z}_{t_k}^{sp})/\tau\}} \right)$$

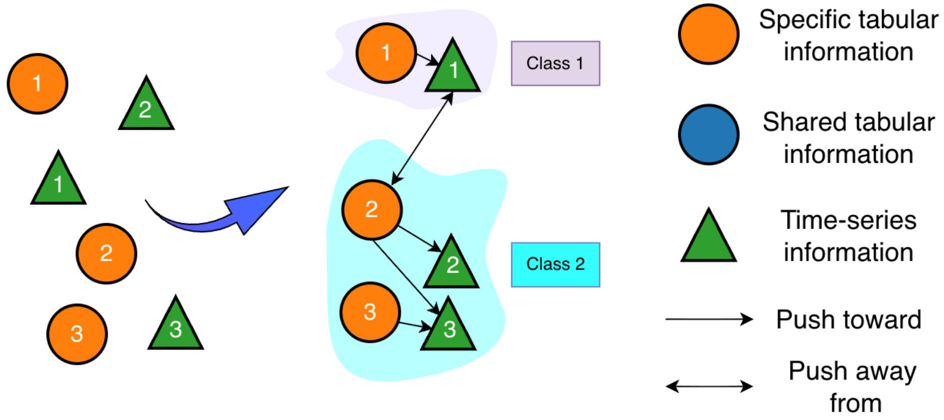
- z_s time series token
- z_t^{sh} share tabular token
- z_t^{sp} specific tabular token
- y patient label

HT classification – Information decoupling

- $\mathcal{L}_{Decoupling} = \mathcal{L}_{SHSD}(z_s, z_t^{sh}, z_t^{sp}) + \mathcal{L}_{reg}(z_s, z_t^{sh}, y)$



$\mathcal{L}_{\text{reg}}(z_s, z_t^{sp}, y)$



Regularization loss

$$\mathcal{L}_{\text{reg}}(\boldsymbol{z}_s, \boldsymbol{z}_t^{sp}, y) = \frac{1}{2N}\sum_{i=1}^{N}\left(r_i^{s,t} + r_i^{t,s}\right)$$

$$r_i^{t,s} = -\frac{1}{S_i}\sum_{j=1}^{N} \mathbb{1}\{y_j = y_i\} \log\left(\frac{\exp\{\text{sim}(\boldsymbol{z}_{t_i}^{sp}, \boldsymbol{z}_{s_j})/\tau\}}{\sum_{k=1}^{N}\exp\{\text{sim}(\boldsymbol{z}_{t_i}^{sp}, \boldsymbol{z}_{s_k})/\tau\}}\right)$$

- z_s time series token
- z_t^{sh} share tabular token
- z_t^{sp} specific tabular token
- y patient label

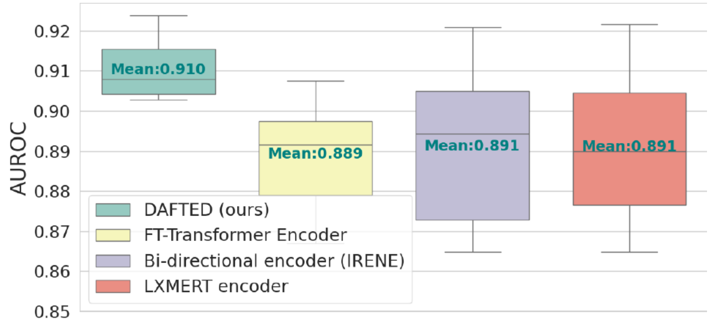
HT classification – Classification performance

- 10 training runs with different seeds

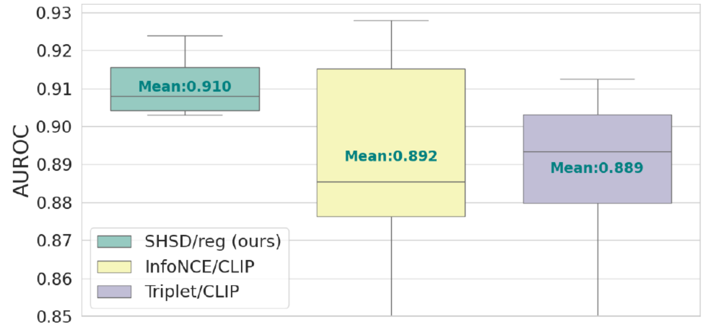
Model	ROC AUC	# parameters
Unimodal models		
XGBoost	87.4	N/A
FT-T Tab	85.8 ± 4.8	863K
FT-T TS	52.2 ± 2.3	1.0M
Multimodal fusion models		
XGBoost	79.7	N/A
MLP	81.8 ± 1.3	391K
IRENE	86.7 ± 2.8	102.9M
FT-T	88.9 ± 1.4	1.1M
DAFTED (ours)	$\mathbf{91.0 \pm 0.7}$	3.0M

Classification performance

- Ablation study

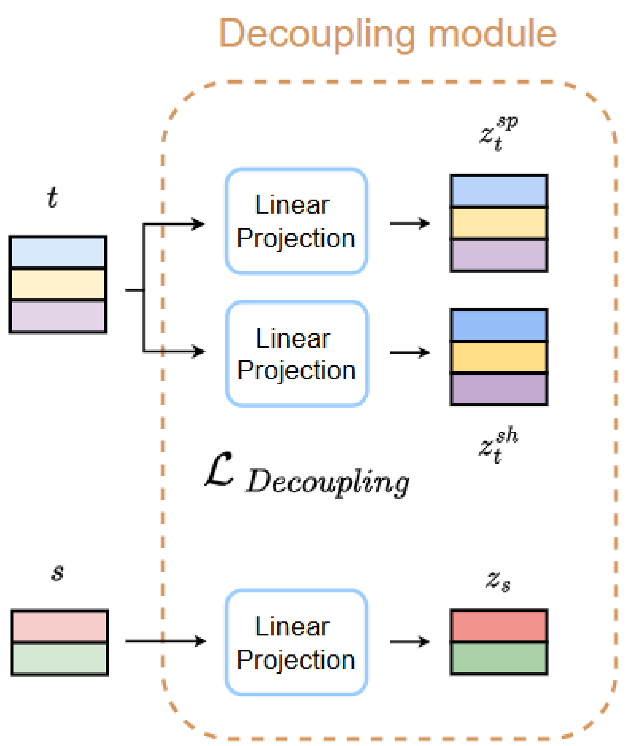


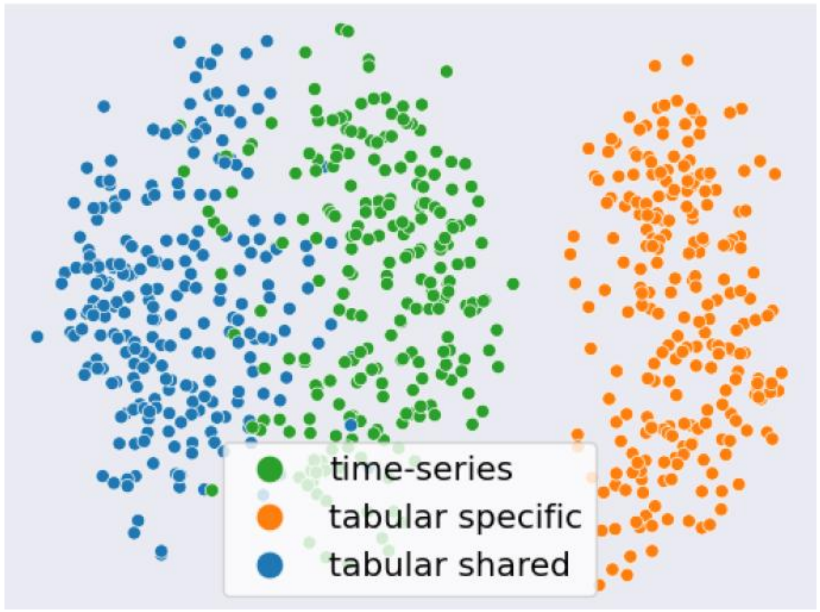
Relevance of fusion scheme



Relevance of the decoupling module

HT classification – Latent space structuration





PaCMAP representation of latent vector before fusion with decoupling $(z_s, z_t^{sh}, z_t^{sp})$

HT classification – A first step towards explainability

- L_2 distance between shared and specific embeddings for each tabular features

Abbreviation	Description	Source
sbp_tte	Systolic Blood Pressure (SBP) during TTE	Echo. measure
lvm_ind	Left Ventricular Mass (LVM) indexed to BSA	Echo. measure
pp_tte	Pulse Pressure (PP) during TTE	Echo. measure
la_volume	Left Atrial (LA) volume indexed to body surface area (BSA)	Echo. measure
lateral_e_prime	Lateral mitral annular velocity (e')	Echo. measure
septal_e_prime	Septal mitral annular velocity (e')	Echo. measure
e_e_prime_ratio	Ratio of E velocity over e': E/e'	Echo. measure
a_velocity	A-wave (active blood flow caused by atrial contraction) velocity	Echo. measure
diastolic_dysfunction	1 point per parameter of diastolic dysfunction: *dilated_la*, *reduced_e_prime*, *d_dysfunction_ratio*, *ph_vmax_tr*	Echo. measure
pw_d	Left ventricular Posterior Wall (PW) thickness at end-Diastole (D)	Echo. measure
gfr	Glomerular Filtration Rate (GFR) indexed to standard body surface area	Biological exam
age	Age	Demographic info
ddd	Defined Daily Dose (DDD) of blood pressure medication	Treatment

13 tabular features

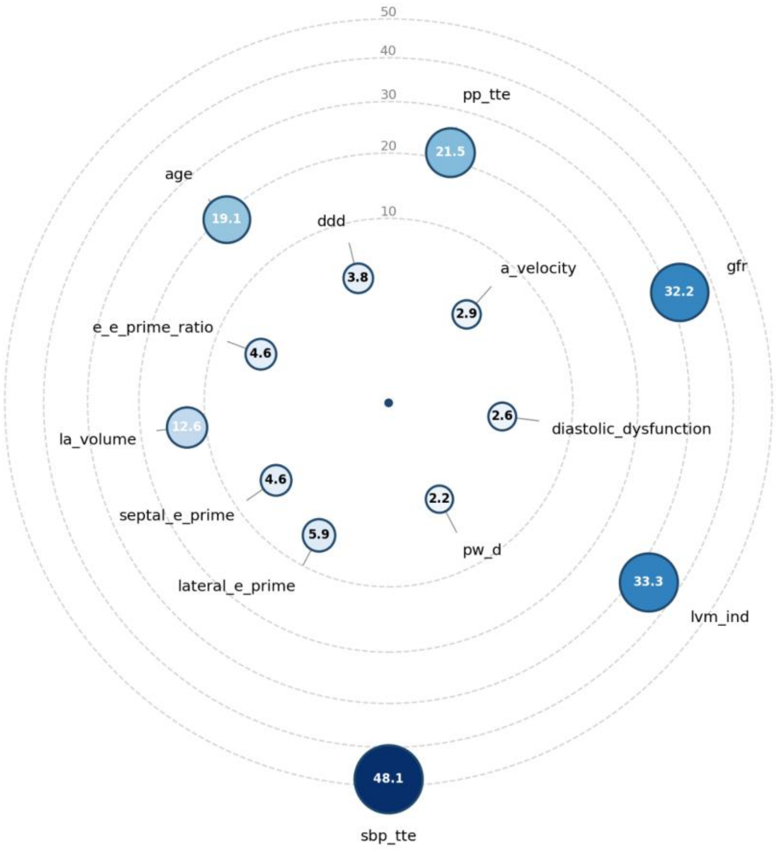
Hypothesis

$dist(z_s, z_t^{sh})$ is low
$\Rightarrow$ Tabular info is highly specific

$dist(z_s, z_t^{sh})$ is high
$\Rightarrow$ Tabular info includes time series info

HT classification – A first step towards explainability

- L_2 distance between shared and specific embeddings for each tabular features



$dist(z_s, z_t^{sh})$ is low

Features related:
- Local blood flow (a_velocity)
- Local tissue velocities (e_e_prime_ratio)

$dist(z_s, z_t^{sh})$ is high

Features related:
- cardiac pressure (sbp-tte, pp_tte)
- cardiac structures (lvm_ind, la_volume)
- age

Conclusions

- ✓ Pilot study on the hierarchical integration of time series with patient data
- ✓ Information decoupling appears to be a promising direction

- ✗ Evaluation on a mono centric / mono vendor small cohort (238 patients)
- ✗ No time series from hemodynamic imaging (e.g. Doppler)
- ✗ Explainability and robustness remain key challenge
- ✗ Clinical adaptation remains essential

Perspectives

► On the construction of more solid multimodal datasets in echocardiography

ORCHID private dataset

- 1500 patients
- two centers / two vendors
- US sequences (A2C, A3C, A4C) + EHR
- Prediction of the etiology of cardiac diseases
  - Amyloidosis
  - Hypertension
  - Ischemic
  - Non ischemic

Improved time-series extraction

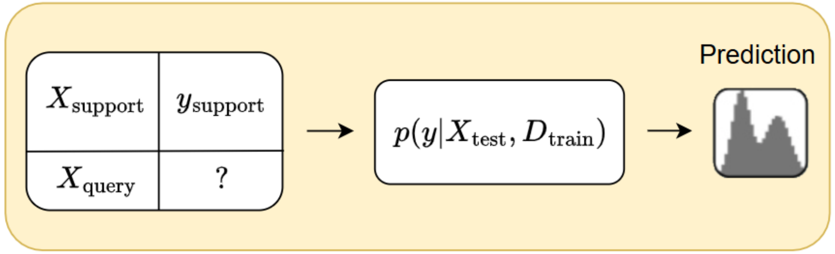
- *RL4Seg3D-v2* model for large scale cardiac segmentation
- *TAS-1K* – 1,478 simulated echocardiographic videos with reference myocardial motion
- *TAS-39K* – 39,000 simulated echocardiographic videos with reference myocardial motion
- *TAS-Net-u* model for accurate and robust global and regional strain estimation with uncertainty estimates

Perspectives

- On the integration of more robust tabular / time series representations for small/medium databases
- Foundation models built from in-context learning

Tabular data

- TabPFN models
- Learning from synthetic data
- Learning of inductive bias
  - Causality
- Generalization on real tabular data



Time series (TS)

- TSPFN / TimEE models
- Learning from synthetic data
- Continued pre-training on real data
- Learning of inductive bias
  - Autoregressive process with exogeneous inputs
- Generalization on TS extracted from real data

[J. Stym-Popper et al., STACOM, 2026]

Many perspectives

- On the computation of blood flow / local tissue velocity time series
- On the computation of uncertainty in motion / deformation estimation
- On the integration of uncertainty in diagnostic prediction
- On the generation of intrinsically explainable models

Master offers

Coming soon

PhD offers

Coming soon

Postdoc offer

Coming soon

Help to get permanent researcher position in France

Many thanks to – my collaborators (*and friends*)


Nicolas Duchateau


Pierre-Marc Jodoin


Christian Desrosiers


Nicolas Thome


Damien Garcia


Clément Rambour


Federica Granese


Pierre-Yves Courand


Eric Saloux

Many thanks to – my PhD students / postdocs


Arnaud Judge


Thierry Judge


Jérémie Stym-Popper


Celia Goujat


Nathan Painchaud


Hang Jung Ling


Khuram Faraz

Many thanks to – funding agencies / company that support this research

anr agence nationale de la recherche

cnrs

La Région Auvergne-Rhône-Alpes

IUF institut universitaire de France

iCARDIO.ai In a Heartbeat.