MEDIMP: 3D Medical Images with clinical Prompts from limited tabular data for renal transplantation

L. Milecki[1], V. Kalogeiton[2], S. Bodard[3,4], D. Anglicheau[3,4], J.-M. Correas[3,4], M.-O. Timsit[3,4], M. Vakalopoulou[1]

[1] MICS, CentraleSupelec, Paris-Saclay University, Inria Saclay, France
[2] LIX, École Polytechnique/CNRS, Institut Polytechnique de Paris, France
[3] HEGP, Necker Hospital, Paris University, France
[4] UFR Médecine, Paris-Cité University, France

MIDL Nashville 2023

CONTEXT

- **Renal transplantation:** best solution for end-stage renal disease[1]; **But risk of transplant chronic dysfunction**[2]
- Graft monitoring is **multi-disciplinary** (nephrology, urology, radiology)
- **Goal: learning relevant representations of renal transplants MRI data coupled with clinicobiological information**

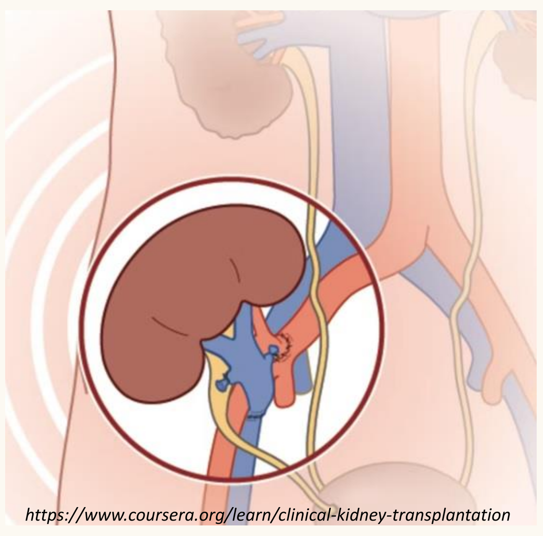
MOTIVATIONS

- Recent advances in NLP **(LLMs[3,4]): textual data perfect candidate** for weakly-supervised tasks
- **Multi-modal pretraining using multiview contrastive learning** for jointly training an image and text encoder: CLIP[5] for natural images, ConVIRT[6] for medical images and radiology reports.

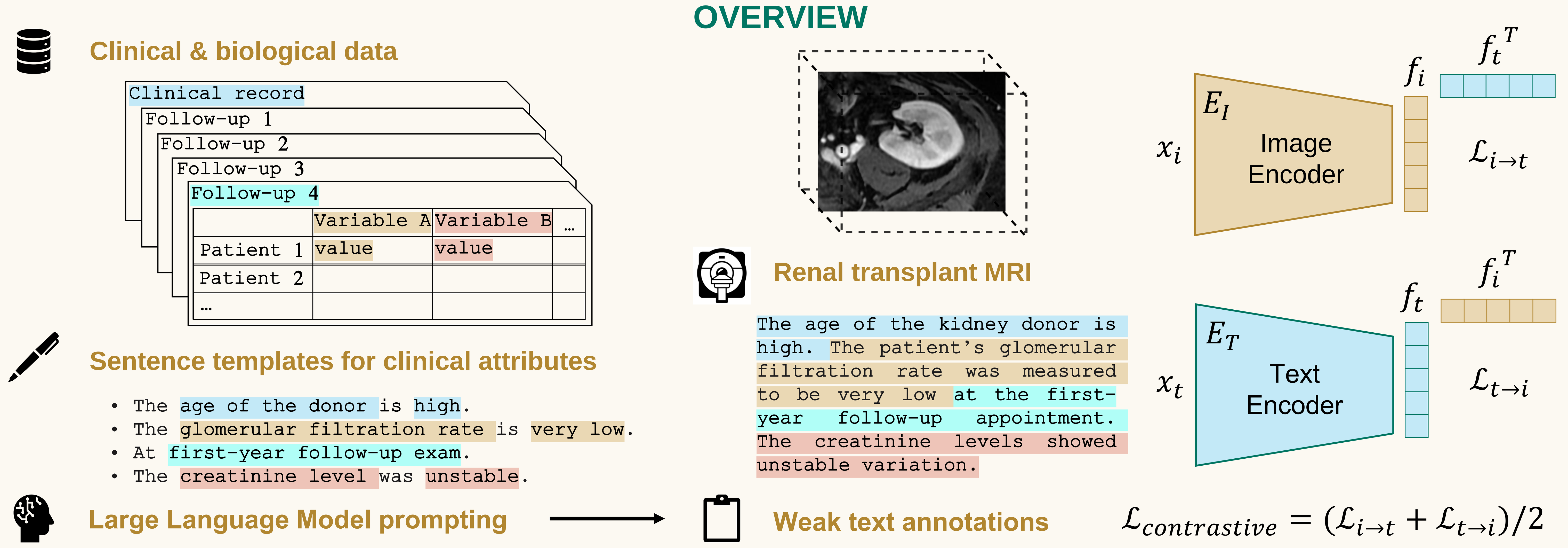
CONTRIBUTIONS

- Semi-automatic **medical prompt generation from tabular data** to leverage the expressive representation of textual data from NLP models
- Extending existing works on combining text with imaging data by integrating **limited 3D medical imaging and textual data and fine-tuning strategies**

CODE



OVERVIEW



1. Medical clinical prompts generation

2. Representation learning setup

MATERIALS & METHODS

➡ **Dataset: 105 patients** subject to kidney transplantation having **DCE MRI series at 4 follow-up exams** (Train/Validation: 72/5, Test: 28).

1. **Medical clinical prompts generation**

- Categorize continuous variables into **text labels**
- Produce **one template sentence** per variable
- Use of **LLMs to produce textual data augmentations**

2. Representation learning setup

- Image encoder: **3D ResNet50** initialized with CLIP[5] weights
- Text encoder: **BERT** initialized with **Bio+Clinical BERT**[7]
- **Fine tune only the last layer** of the text encoder transformer
- **Multiview contrastive learning** using $\mathcal{L}_{contrastive} = (\mathcal{L}_{i\to t} + \mathcal{L}_{t\to i})/2$

$$\text{with } \mathcal{L}_{i\to t} = \sum_{b=1}^{B} -\log \frac{e^{\cos(f_{ib}, f_{tb})/\tau}}{\sum_{k=1}^{B} e^{\cos(f_{ib}, f_{tk})/\tau}}$$

- Downstream task from the image encoder

➡ **Our visual features encode the clinicobiological information incorporated from the medical prompts in the multimodal learning**

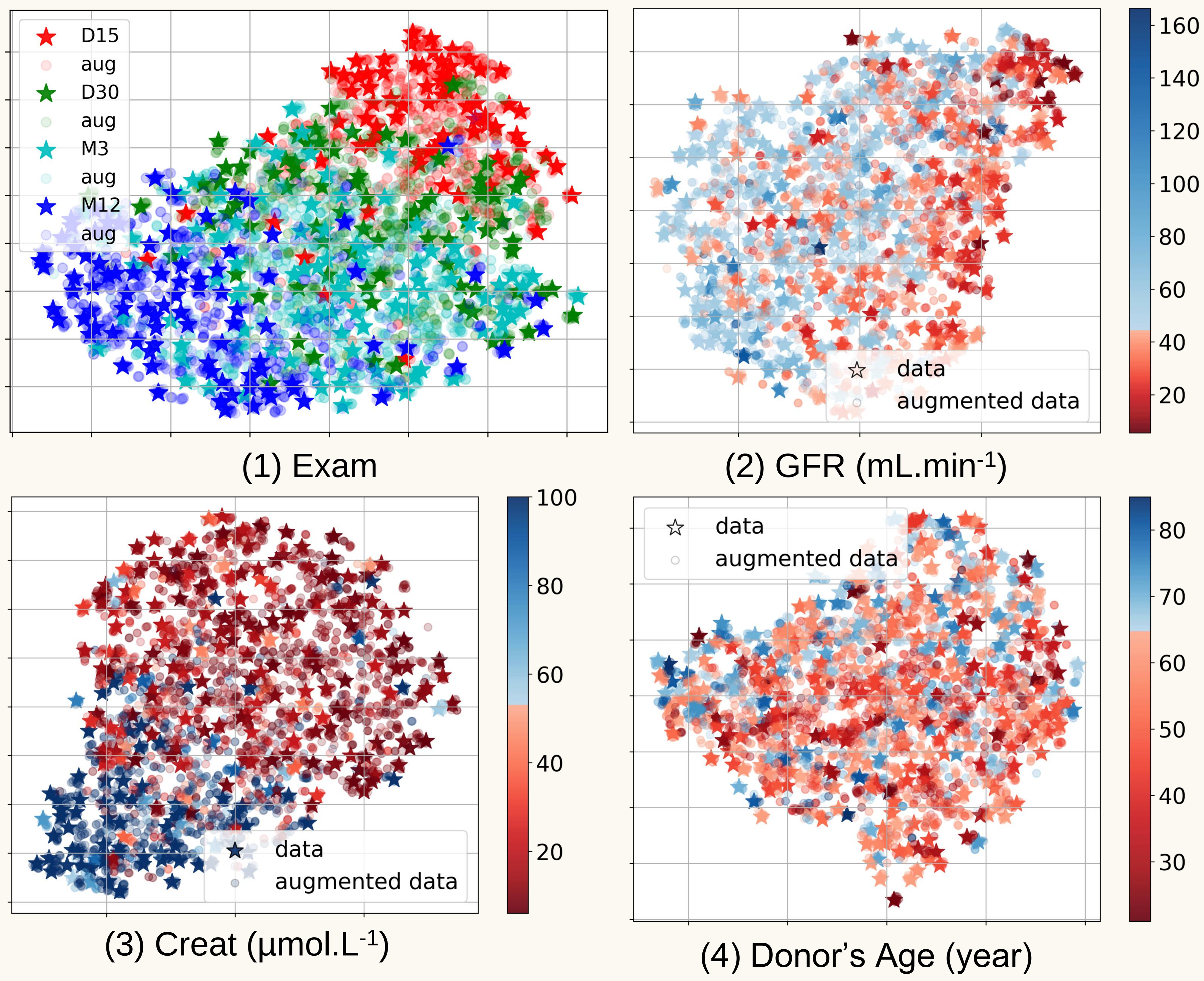


Figure 1: t-SNE visualizations of the features of the last layer of MEDIMP image encoder using the DCE MRI exams.

RESULTS

Table 1: **Comparison of MEDIMP vs SOTA**. Evaluation at 2,3,4 years post-transplantation and report the mean. Ablations in weak annotations from either the comparison CosEmbLoss pretaining or our proposed generated textual data.

Method	Weak annotations				2 years		3 years		4 years		Mean	
	GFR	Exam	Creat	D.A.	AUC	F1	AUC	F1	AUC	F1	AUC	F1
CLIP weights					62.6	73.7	52.5	78.1	51.3	54.6	55.5	68.8
CosEmbLoss[8]	✓				76.2	86.2	77.8	70.6	67.0	77.3	73.6	78.1
CosEmbLoss++	✓			✓	84.4	88.9	82.5	**86.4**	73.9	85.7	80.3	87.0
CosEmbLoss++	✓		✓	✓	78.2	87.0	75.0	83.3	74.8	87.0	76.0	85.8
CosEmbLoss++	✓	✓	✓	✓	75.5	85.7	62.0	69.8	63.5	80.9	67.0	78.8
MEDIMP	✓				56.5	83.3	51.9	79.1	49.6	**90.2**	52.6	84.2
MEDIMP	✓	✓		✓	76.9	73.2	**86.3**	85.7	74.8	**90.2**	79.3	83.0
MEDIMP	✓	✓	✓		72.8	86.4	71.9	81.0	71.3	71.8	72.0	79.7
MEDIMP	✓	✓	✓	✓	**85.0**	**89.4**	84.4	83.7	**75.7**	**90.2**	**81.7**	**87.8**

CONCLUSION

We introduced an approach to learn **powerful manifolds of renal transplant DCE MRI** data toward transplant function forecasting, providing:

- an elegant way to **incorporate clinical or biological information** into the learning process of feature extraction of medical imaging data;
- presented **representation learning strategy** enabling us to outperform the state of the art in the challenging task of creatinine prediction;
- promising results advocating the **use textual data from emerging LLMs** to assist in training robust medical imaging models.

Acknowledgments & References:
Work performed using HPC resources from the "Mésocentre" (CentraleSupelec , ́Ecole Normale Supérieure Paris-Saclay) supported by CNRS, Région Ile-de-France, and from IDRIS (2022-AD011013541).

[1] Suthanthiran *et al*. Renal transplantation. 1994.
[2] Hariharan *et al*. Long-term survival after kidney transplantation. 2021.
[3] T. Brown *et al*. Language models are few-shot learners. NeurIPS. 2020.
[4] OpenAI. Chatgpt: Optimizing language models for dialogue, 2022.
[5] A. Radford *et al*., Learning transferable visual models from natural language supervision. 2021.
[6] Y. Zhang *et al*., Contrastive learning of medical visual representations from paired images and text. 2020.
[7] Alsentzer *et al*. Publicly available clinical BERT embeddings. 2019.
[8] Milecki *et al*. Contrastive masked transformers for forecasting renal transplant function. 2022.