

Cross-Continental Clinical Representation: Validating CLMBR-T-Base for Length of Stay Prediction in an Italian Research Hospital

Giovanna Nicora¹, Francesca Negri, Valentina Tibollo², Alberto Malovini², Arianna Dagliati¹, Lucia Sacchi¹, Federico Sottotetti², Paola Baiardi², Riccardo Bellazzi^{1,2}

¹ Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

² IRCCS ICS Maugeri Hospital, Pavia, Italy

Intro and Objective

- Clinical Foundation Models promise generalist medical AI^{1,2} — **but do models trained on one healthcare system transfer to another?**
- Our goal is to evaluate the cross-continental applicability of CLMBR-T-Base, a FM model pretrained on Stanford Medicine data, by applying it to a distinct Italian cohort to predict hospital Length of Stay (LoS)^{3,4}.

Data and methods

Data overview and preprocessing

- OMOP EHR dataset from the Italian research Hospital IRCCS Maugeri in Pavia (data from 2012 to 2025)
- OMOP instances (day 1 of hospitalization) converted to the MEDS (Medical Event Data Standard) schema
- 9M hospital concepts mapped via 3-level hierarchy to CLMBR-T-Base's 26,249-token vocabulary.

Clinical data representation

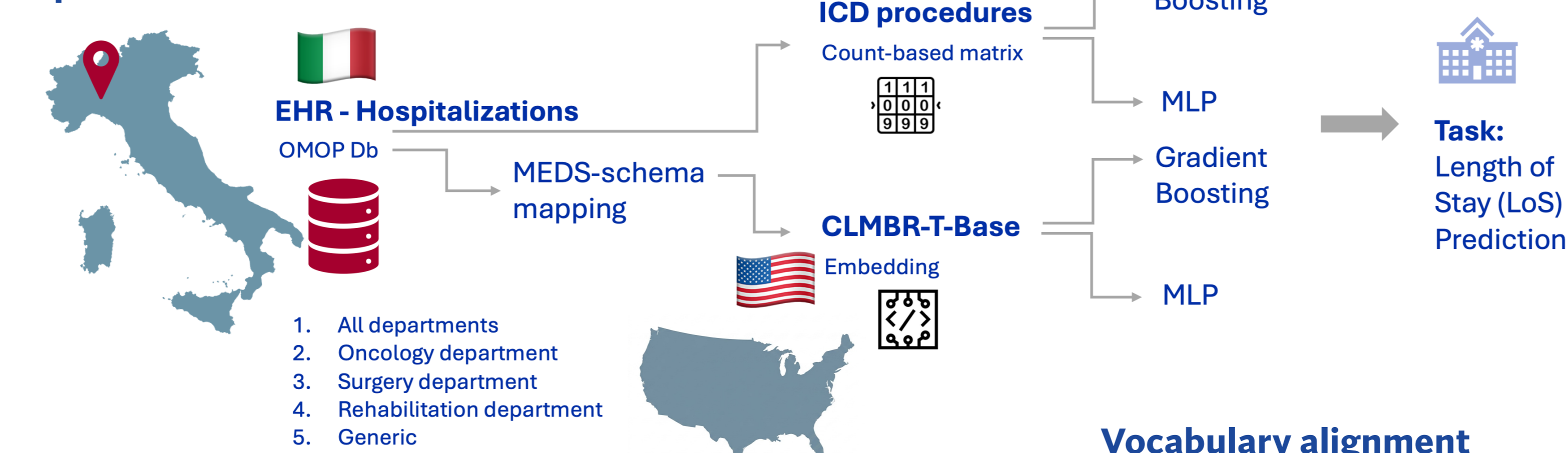
- FM Embeddings:** CLMBR-T-Base on day-1 MEDS data
- Count-based ICD codes:** ICD-10 code counts

Model development

- Model types:**
 - XGBoost & MLP on embeddings vs. counts.
 - Baseline Regression with covariates : age, gender, department and # of previous hospitalizations in a 6-month window
 - Target:** log(LoS).
- Evaluation strategy:** patient-level splitting (80-10-10)
 - Performance (R2) evaluated on complete hospital model and models trained across 4 clinical macro-categories (Surgical, Rehabilitation, Oncology, Generic).
 - Embedding cohesion validated using latent space analysis (HDBSCAN clustering).

Results

Pipeline

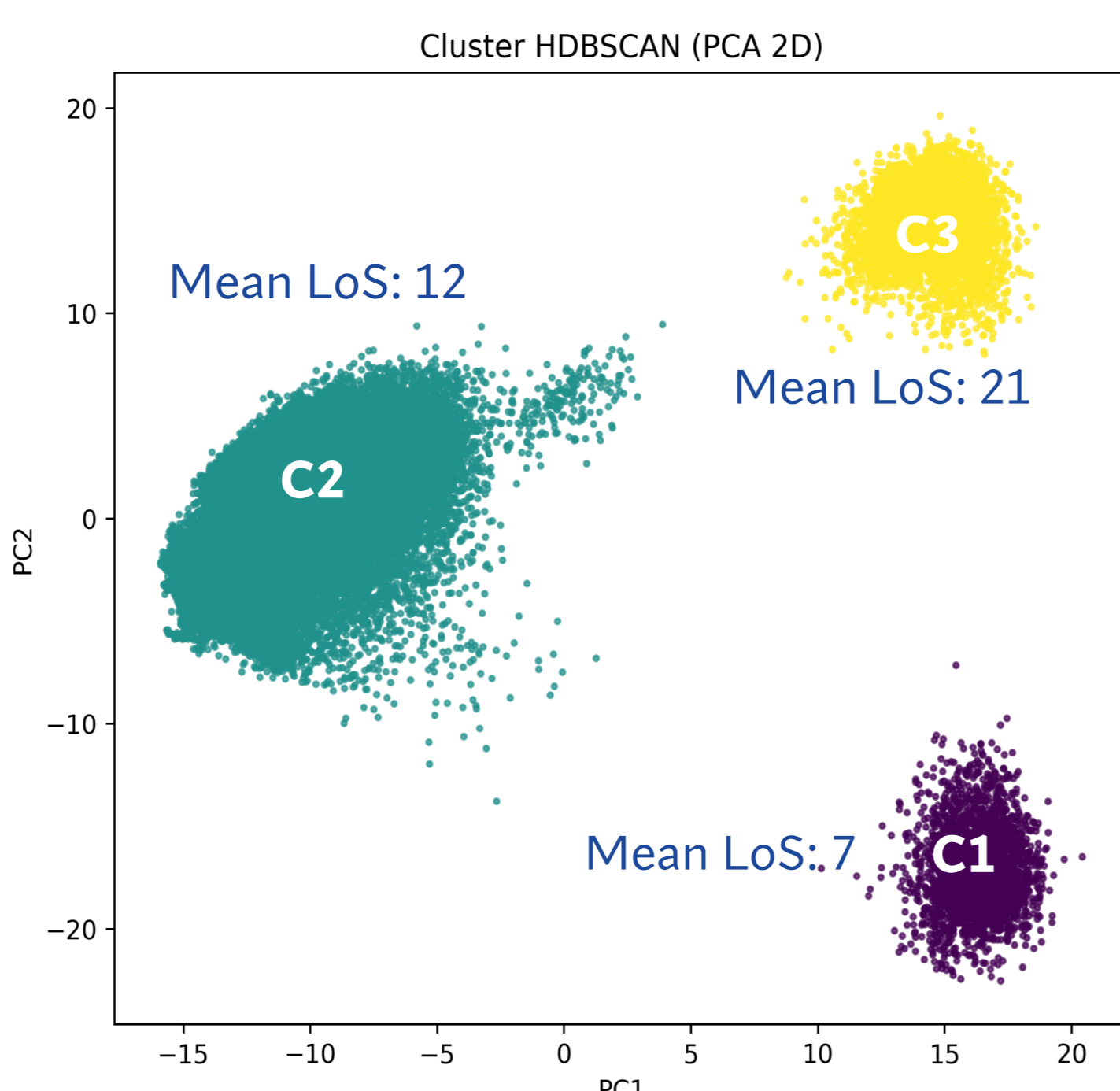


Data overview and preprocessing

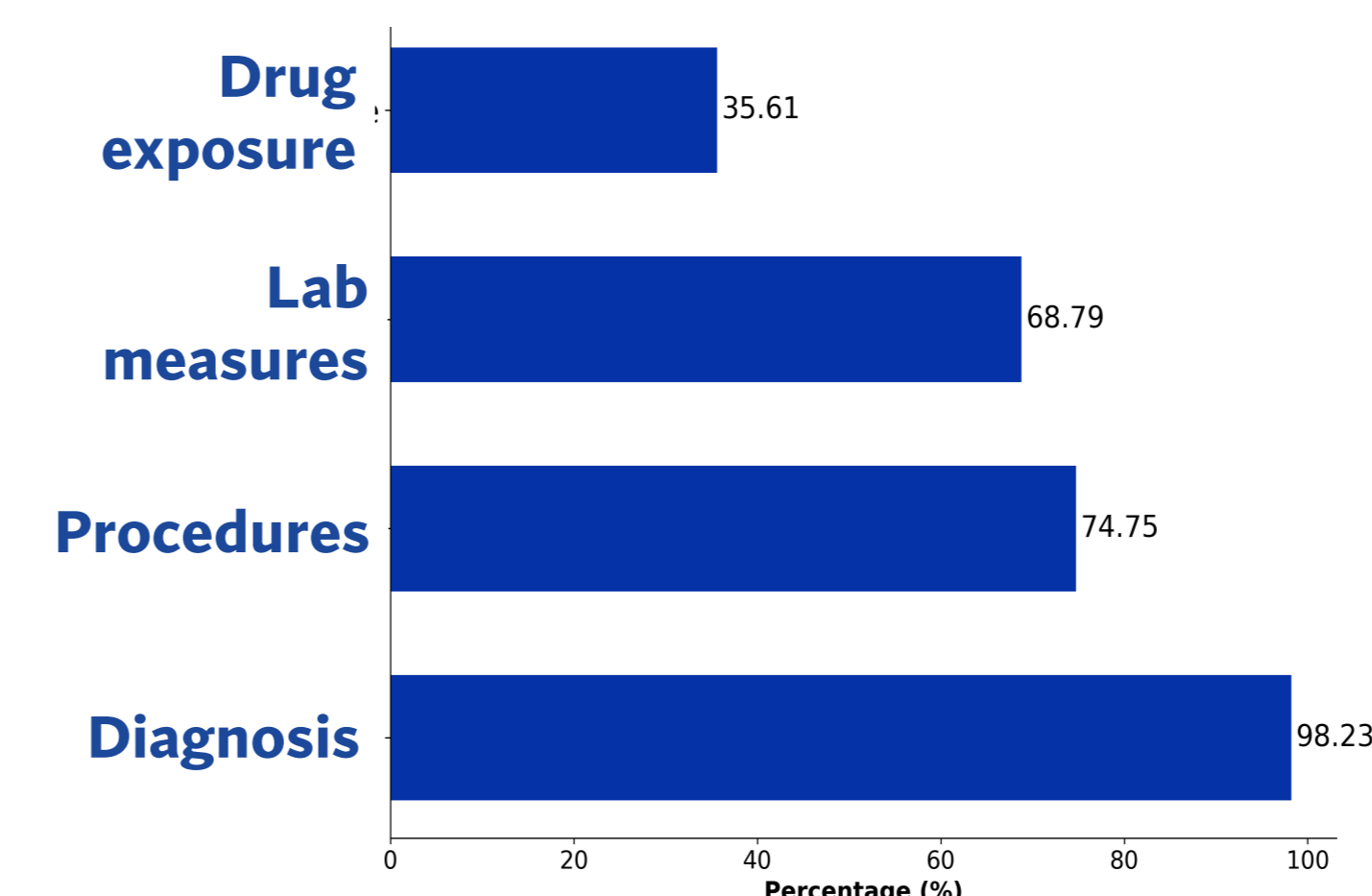
department	admissions	patients	median LoS
All	81,371	52,585	11
Oncology	9,232	8,413	6
Surgery	7,740	4,264	22
Rehabilitation	32,947	24,077	7
Generic	3,452	21,704	2

Length-of-stay prediction task and latent space cohesion

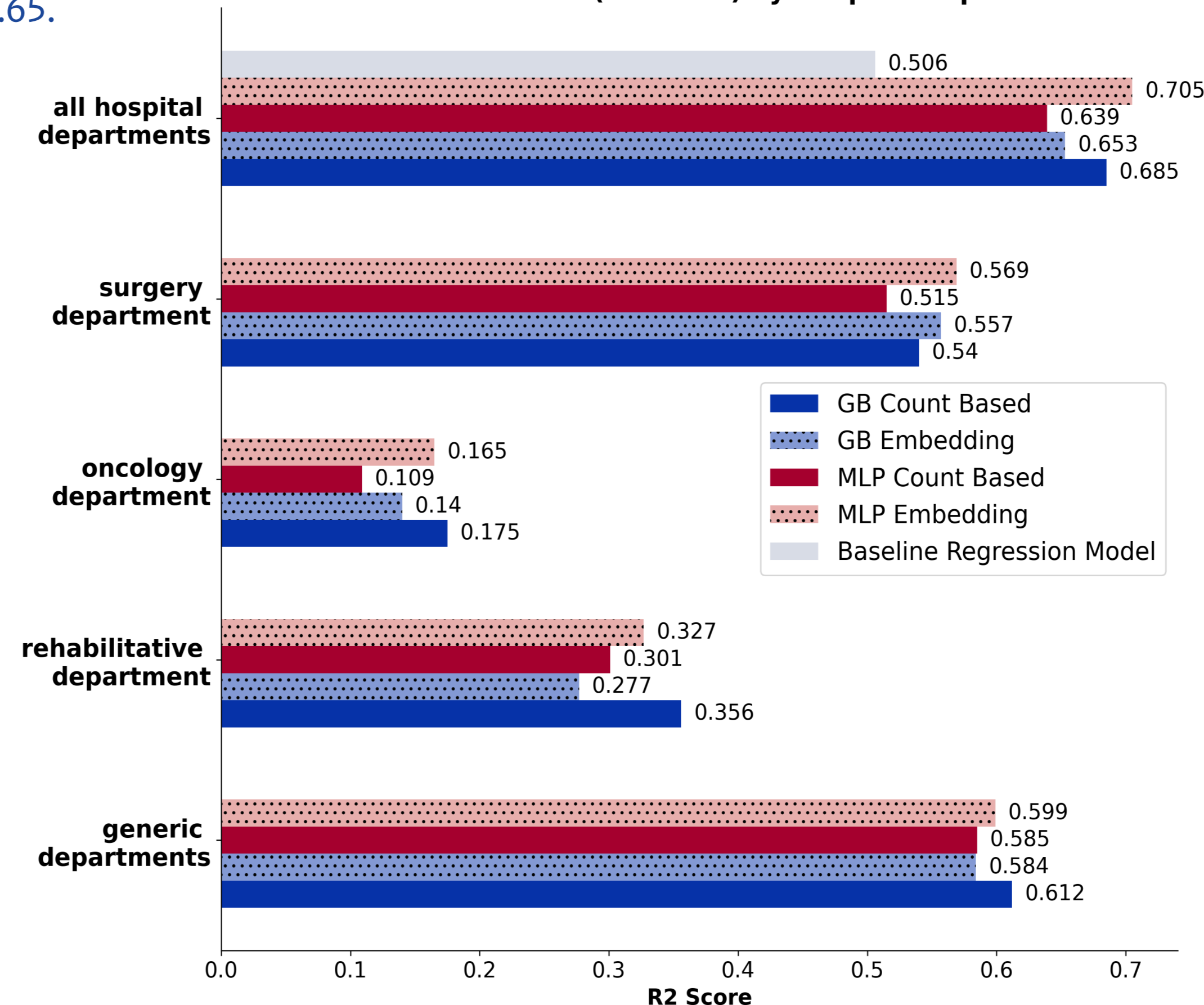
- $R^2=0.705$ (MLP on embeddings, best overall)
- Pre-pandemic training preserves SOTA, $R^2=0.65$.
- HDBSCAN: 3 clinically coherent clusters (mean LoS 7 / 12 / 21)



Vocabulary alignment



Model Performance (R2 Score) by Hospital Department



Conclusion

- A Foundation Model pre-trained on U.S. patient data effectively generalizes to an Italian healthcare context**, achieving SOTA performance in predicting hospital Length of Stay
- Oncology has lowest predictability, probably due to high information concentration**, where 10 codes cover 99.2% of observations, and the absence of critical data (disease stages) in current OMOP mapping
- A key contribution is the validation of data standardization as a bridge for cross-continental AI.** The use of the OMOP and the MEDS schema was essential to harmonize disparate clinical lexicons.
- The model's temporal robustness** when tested on the post-2020 cohort, indicates that FM-derived representations are stable to major organizational shifts, such as those induced by the COVID-19 pandemic.
- Pre-trained clinical FMs can be successfully deployed in new geographical settings** with minimal local infrastructure, potentially democratizing access to advanced predictive tools for hospitals with limited computational resources

Future works

- Local fine-tuning for bridge vocabulary gaps
- Integration of multimodal clinical data
- New predictive tasks (e.g. mortality)
- Embedding interpretability

References

- Hou, W., et al. Improving Clinical Foundation Models with Multi-modal Learning and Domain Adaptation for Chronic Disease Prediction. IEEE J Biomed Health Inform PP, (2025).
- Bommasani, R. et al. On the Opportunities and Risks of Foundation Models. Preprint at <https://doi.org/10.48550/arXiv.2108.07258> (2022).
- Wornow, M., et al. EHRSHOT: An EHR Benchmark for Few-Shot Evaluation of Foundation Models. Advances in Neural Information Processing Systems 36, 67125–67137 (2023).
- Steinberg, E. et al. Language models are an effective representation learning technique for electronic health record data. Journal of Biomedical Informatics 113, 103637 (2021).

Contact us

giovanna.nicora@unipv.it

