



Funded by
the European Union

Project: 101226456 — MLCARE — HORIZON-MSCA-2024-DN-01



DC 10 MSCA Doctoral Network MLCARE (Machine Learning Computational Advancements for peRsonalized mEdicine)

Individual Research Project title

Enhancing Health Studies: Effective Use of Continuous Health Indicators from Electronic Health Records

Primary research lines

Foundation Models, OMOP, electronic health records, measurements

Lead beneficiary (place of employment)

University of Tartu, Institute of Computer Science

Narva mnt 18, 51009, Tartu, Estonia

Read more about the university: <https://ut.ee/en/about-us>

Secondments at:

1. **Max-Planck Gesellschaft zur Förderung der Wissenschaften e.V. (Germany):** incorporating continuous intensive care unit variables in Foundation Models for sepsis prediction (3 months, tentatively in Dec 2027 - Feb 2028).
2. Tentatively: **Siemens Healthineers AG (Germany):** prompting for continuous time measurements in FMs (3 months, tentatively in Aug 2028 - Oct 2028).

Supervisors

Dr. Asad Munir (University of Tartu) (main supervisor)

Dr. Sulev Reisberg (University of Tartu)

Secondment mentors

Dr. Bernadette Stolz (Max-Planck Gesellschaft zur Förderung der Wissenschaften e.V)

Dr. Tobias Heimann (Siemens Healthineers AG).



Project tasks and objectives

Turn underused health measurement data (continuous, sparse, variable-quality) into robust numerical and temporal embeddings that can be integrated into Foundation Models to improve patient-level trajectory and outcome prediction.

Subtopic 1. Representing clinical time-series measurements as embeddings

Develop and benchmark embeddings of clinical measurements that can later be integrated into a Foundation Model. Build a reproducible pipeline to transform health measurement data into several baseline representations. For example,

- scalar tokens (value + unit + discretized timestamp),
- continuous learned embeddings, and
- interval tokens (min/max/last/trend per time window).

Evaluate these embeddings for information loss and predictive utility on baseline tasks such as short-term event prediction or next-measurement prediction. **Demonstrate that learned embeddings contain clinically meaningful information** by comparing their performance to naïve representations (e.g., last value) using simple machine learning models.

Subtopic 2. Modelling temporal dynamics of patient measurements

Learn how clinical measurements evolve over time, even when irregular, missing, or variable in quality. Train temporal models to encode measurement sequences into time-sensitive embeddings. For example,

- point-process models to predict time to next measurement,
- time-aware transformers to handle variable gaps, or
- adaptive sampling to emphasize clinically informative measurements.

Evaluate whether temporal embeddings improve predictive performance on downstream tasks (e.g., 30-day hospitalization risk). Preliminary evaluation can be performed using simple predictive models or frozen Foundation Model embeddings, but full fine-tuning of a Foundation Model is reserved for Subtopic 3.

Subtopic 3. Integrating embeddings into a Foundation Model for outcome prediction

Start with a pretrained clinical Foundation Model and fine-tune it on all patient sequences, incorporating static (subtopic 1) and temporal (subtopic 2) embeddings. This step teaches the Foundation Model to interpret the new embeddings within its internal patient representations. Select 3 clinically meaningful outcomes and predict each outcome using the integrated embeddings, optionally with lightweight task-specific output layers if needed.



Perform retrospective validation and compare predictive performance across three scenarios:

- baseline using raw health data without embeddings,
- learned embeddings without temporal modelling, and
- Foundation Model with learned embeddings and temporal modelling.

Optionally, explore different prompting strategies to assess interpretability or zero-shot predictive ability. **Demonstrate that the added embeddings consistently improve predictive performance.**

As the field of Artificial Intelligence is developing rapidly, specific tasks, methods, and objectives in this project may be refined or adjusted during the course of the PhD to reflect new scientific insights, emerging methodologies, or available datasets.

Expected Results

1. Development of innovative techniques to incorporate sparse and variable-quality continuous health measurements into Foundation Models effectively.
2. Predictive methods over electronic health records that incorporate continuous health indicators.
3. Publication of at least three papers in international, peer-reviewed, high-impact journals or conferences.
4. First-author contribution in at least two of these publications.
5. Contribution to the following deliverables of the MLCARE network:
 - 5.1. D1.2 “Techniques for training Foundation Models in clinical data” (due month M21, Sept 2027) – Content: Multi-modal self-supervision methods to train Foundation Models over electronic health records. Tokenization techniques for continuous health measurements. Strategies for medical reasoning and training. Self-supervision methods in mobile monitorization data.
 - 5.2. D1.6 “Benchmarking clinical Foundation Models” (due M45, Aug 2029) – Content: Interpretable prompting mechanisms to avoid costly fine-tuning in multiple clinical downstream tasks. Benchmarks for privacy-preserving. Predictive methods over electronic health records with continuous health indicators. A benchmark for sequential decision making.
6. Leading the production of deliverable:
 - 6.1. D1.5 “Clinical Foundation Models” (due M33, Sep 2028) – Content: Clinical validation methods in the hospital dataset. Identifying evidence sources and outcomes elements across biomedical contexts. Studies to relate specific continuous health measures to patient outcomes. A series of finetuned multi-modal Foundation Models for medical reasoning.

As the field of Artificial Intelligence is developing rapidly, the expected results may slightly change during the studies, accordingly.



Eligibility

Applicants must:

- Be - at the date of recruitment - a doctoral candidate (i.e., not already in possession of a doctoral degree).
- Be - at the date of recruitment - formally admitted to a PhD programme leading to the award of a degree in at least one EU Member State or Horizon Europe associated country. For that purpose, candidates must meet the national requirements for doctoral enrolment in the host country. Proof of admission must be provided prior to the start of the contract.
- Not have resided or carried out their main activity (work, studies, etc.) in the country of the recruiting beneficiary for more than 12 months in the 36 months immediately before the recruitment date — unless as part of a compulsory national service or a procedure for obtaining refugee status under the Geneva Convention.
- Be working exclusively for the action.

Evaluated criteria

- Study records, including Bachelor and Master
- Previous work & research experience
- Letter of motivation
- Letters of recommendation
- Positive attitude, previous mobility experience, good communication skills
- English proficiency and other languages

Preferred starting date

September 2026

Additional requirements for the studies

During their studies, the doctoral student must comply with all regulations of both the MSCA Doctoral Network and the University of Tartu. The University's doctoral study regulations are available at: <https://ut.ee/en/doctoral-studies>

Please note that while the MSCA Doctoral Network programme has a duration of 3 years, doctoral studies at the University of Tartu may extend up to a maximum of 4 years.



Approximate gross salary

A full employment contract with the Doctoral Candidate will be signed with the following approximate salary for the duration of MSCA Doctoral Network programme (3 years).

The living allowance (gross salary) is **€2,780 per month**. In addition, the financial package includes monthly allowances of **€531** for mobility and **€370** for family (if applicable), in accordance with the MSCA Doctoral Network financial regulations. These amounts are subject to mandatory deductions under national law, such as social security contributions and income tax. Please check the applicable deductions in the beneficiary's country. Based on current laws and coefficients in Estonia, the **approximate net salary** is **€2,490 per month** (without family) or **€2,768 per month** (with family).

While the MSCA Doctoral Network programme lasts 3 years, doctoral studies at the University of Tartu may continue for up to 4 years. The salary in the 4th year is less than in the previous years and it does not include the monthly allowances for mobility nor family. The exact gross salary in the 4th year depends on the minimum salary of the doctoral students in University Tartu at that time which currently is €2,300 per month.

Application

Applications will include:

1. Detailed CV
2. Motivation letter
3. Bachelor and Master diplomas and transcripts along with the translations into English
4. Proof of English proficiency (<https://ut.ee/en/sisu/phd-language-requirements>)
5. Copy of passport
6. At least 2 letters of recommendation

There are two **mandatory steps** for applicants:

1. Fill out the **central application form**. This form lets the consortium know whether you intend to apply for multiple positions within the Network — but please complete it even if you are applying for only one position: <https://mlcare.webs.tsc.uc3m.es/open-positions/>
2. Apply for DC10 at: <https://estonia.dreamapply.com/courses/course/1496>