

Research internship, Master 2 or final year of engineering school, 5 months

## Improving mortality prediction from ECGs

Laboratoire d'Informatique et des Systèmes (LIS), Aix-Marseille Université

**Supervisors:** Paul Chauchat, Luca Thiebaud, Mustapha Ouladsine (LIS), Stéphane Delliaux (C2VN)

**Location:** Campus Saint-Jérôme, 52 av Escadrille Normandie Niemen, Marseille

**Duration:** 5 months, flexible start date, first half of 2027

**Funding:** Institut Laënnec (AMU), standard internship stipend

### Context

The electrocardiogram (ECG) is a diagnostic tool commonly used in medical practice. It is made of 12 leads (electrodes), each producing a time series. A growing number of studies investigate its prognostic value, in particular the possibility of using machine learning to associate it with a short- or medium-term mortality risk [1, 2]. Thanks to large databases, several foundation models for the ECG, trained on millions of signals, have been proposed. Their results are promising, but still imperfect [3].

Recent work by the team has shown that useful and well-generalising representations can be learned from *a thousand times fewer* ECGs than existing foundation models, using biometric pre-training: recognising the same patient from two different recordings [4, 5]. These representations lead to state-of-the-art performance in one-year mortality prediction. One factor has a strong impact on the results: how the 12 leads of the ECG are combined. The leads are highly correlated, partly because of the acquisition setup (electrode placement), which encourages overfitting. Current methods regularise strongly by processing the leads independently before fusing the results.

### Internship objectives

The internship extends this work along two axes:

1. **Lead fusion:** propose and evaluate new fusion methods, for instance taking into account the spatial organisation of the leads through graph neural networks [6, 7] and/or attention mechanisms [8, 9].
2. **Survival analysis:** study mortality beyond binary classification, using survival analysis methods (e.g. DeepSurv, DeepHit [10, 11]), in order to assess the benefit of the proposed methods in this setting.

Depending on the skills and interests of the selected candidate, one of the two axes may be emphasised. The project relies on the public cohorts CODE-15% [12] and MIMIC-IV-ECG [13, 14]. The internship will be carried out in coordination with the ongoing PhD thesis in the team on the same topic, whose work it directly extends [4, 5]. Depending on the results, application to medical data from Hôpital Nord in Marseille may be considered, subject to administrative approval. The results may lead to a scientific publication.

## Candidate profile

- Master 2 or final-year engineering student, with a strong background in **statistics and/or machine learning**, and a keen interest in **biomedical applications**;
- Good programming skills in **Python** (ideally with experience in PyTorch);
- Prior knowledge of the ECG, or experience in signal processing or survival analysis, is a plus.
- Critical thinking about the use gen AI (pros and cons)

## How to apply

Send an e-mail to `paul.chauchat@lis-lab.fr` and `luca.thiebaud@lis-lab.fr`:

- Describe your interest in the topic and how your profile fits it directly in the email;
- Attach your CV and your academic transcripts.

Application deadline: 30 November. Applications will be reviewed on a rolling basis.

## References

- [1] S. Raghunath et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nature Medicine*, 26:886–891, 2020. <https://doi.org/10.1038/s41591-020-0870-z>
- [2] P. Lukyanenko et al. Deep survival analysis from adult and pediatric electrocardiograms: a multi-center benchmark study. *BioData Mining*, 19(1), 2025. <https://doi.org/10.1186/s13040-025-00510-4>
- [3] M.A. Al-Masud, J.M. Lopez Alcaraz, N. Strodthoff. Benchmarking ECG FMs: A Reality Check Across Clinical Tasks. *ICLR*, 2026. <https://doi.org/10.48550/arXiv.2509.25095>
- [4] L. Thiebaud, P. Chauchat, M. Ouladsine, S. Delliaux. Learning Cardiac Features: ECG Biometrics Across Time and Exercise. *AIH*, 2026, p. 94–107. <https://doi.org/10.48550/arXiv.2609.21962>
- [5] L. Thiebaud, P. Chauchat, M. Ouladsine, S. Delliaux. BiomECG: Learning Transferable ECG Representations from Small Cohorts. Preprint, 2026. <https://hal.science/hal-05778973>
- [6] A. Müller, M. Scheibl, T. Uhe, W.-R. Schäbitz, B. Wrede. Application of Graph Neural Networks on ECG Data: A Systematic Literature Review. *IEEE J. Biomed. Health Inform.*, 30(7):6255–6269, 2026. <https://doi.org/10.1109/JBHI.2026.3651842>
- [7] Y. Qiang, X. Dong, X. Liu, Y. Yang, Y. Fang, J. Dou. Conv-RGNN: An efficient Convolutional Residual Graph Neural Network for ECG classification. *Comput. Methods Programs Biomed.*, 257:108406, 2024. <https://doi.org/10.1016/j.cmpb.2024.108406>
- [8] L. Fu, B. Lu, B. Nie, Z. Peng, H. Liu, X. Pi. Hybrid Network with Attention Mechanism for Detection and Location of Myocardial Infarction Based on 12-Lead Electrocardiogram Signals. *Sensors*, 20(4):1020, 2020. <https://doi.org/10.3390/s20041020>
- [9] Y. Li, G. Wang, Z. Xia, W. Yang, L. Sun. A Dual-Scale Lead-Separated Transformer for ECG Classification. *45th Annual Int. Conf. of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, p. 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340556>
- [10] J.L. Katzman et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.*, 18:24, 2018. <https://doi.org/10.1186/s12874-018-0482-1>
- [11] C. Lee, W. Zame, J. Yoon, M. van der Schaar. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. *AAAI*, 32(1), 2018. <https://doi.org/10.1609/aaai.v32i1.11842>
- [12] A.H. Ribeiro et al. CODE-15%: a large scale annotated dataset of 12-lead ECGs. Zenodo, 2021. <https://doi.org/10.5281/zenodo.4916206>
- [13] B. Gow et al. MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset. PhysioNet, 2023. <https://doi.org/10.13026/4nqg-sb35>
- [14] A.E.W. Johnson et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 2023. <https://doi.org/10.1038/s41597-022-01899-x>