PhD Position F/M: IOT-ML: Secure Machine Learning on IOT Traces for Daily Activity Discovery

Supervision : Luc Bouganim (Luc.Bouganim_at_inria.fr), Philippe Pucheral (Philippe.Pucheral_at_inria.fr)
Applications can also be sent by email (CV, motivation letter).

Context.
The PETRUS team (Inria/UVSQ), in association with the Hippocad company (a subsidiary of La Poste group) and the Yvelines District, is currently deploying secure home boxes for 10,000 patients. These boxes, based on the team's research results (DBMS embedded in secure hardware), include a personal medical-social database to improve care coordination for dependent people at home. Medical and social workers interact with these secure boxes via a smartphone application. Our objective is to enhance these boxes with the ability to communicate with IoT sensors measuring e.g., luminosity, movement, and temperature to improve patient monitoring. The sensors' raw data will be analyzed by Machine Learning (ML) techniques to identify the patient's activities and thus, detect risk situations like depression or illness. These raw data are however very intrusive. The originality of our approach is to process these raw data inside each box, within the hardware security element, in order to externalize only the relevant information: alerts, aggregated values, and patient dashboards.

Challenges.
Data Mining (e.g., [2]) and ML models (e.g., [4]) have already been proposed in the Daily Activity Recognition and Discovery field. The accuracy of these models highly depends on the presence and on the quality of a labeled training dataset. In our context (see above), obtaining a reliable labeled dataset from elderly people is definitely unrealistic. However, datasets labeled by healthy individuals dedicated to daily activity discovery (e.g., [1]) already exist and could be used as a baseline. Our expectation is to use online/semi-supervised learning approaches [3, 4] to derive an *elderly daily activity* model from a *healthy daily activity* model and dynamically adapt its behavior to each patient. The challenge is then twofold: (1) define the best strategy to build such a *customizable elderly daily activity* model and (2) make its execution compatible with the highly constrained RAM and computing resources of a secure microcontroller (i.e., a secure box).

Methodology and expected results
- State of the art of relevant ML strategies, supervised learning, semi-supervised learning, analysis of a real dataset and of a synthetic dataset. Evaluation of preliminary strategies to discover daily activities
- Capitalize on previous results to answer the scientific challenge described above. Besides these scientific challenges, an experimental work is expected and should lead to a real Proof of Concept that can be demonstrated in the medical/social field.

Applicant's expected background.
- ML algorithm knowledge
- DBMS algorithm knowledge
- Basic knowledge in data security, secure hardware, embedded programming is a plus (but not mandatory)
- Programming language: Python, C or Rust will be appreciated

Benefits package
- Subsidized meals, Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Social, cultural and sports events and activities
- Access to vocational training

Remuneration
- 1st and 2nd year monthly gross salary : 2.051 euros
- 3rd year monthly gross salary : 2.158 euros

References.
[1] Alemdar, Hande & Ertan, Halil & Incel, Ozlem & Ersoy, Cem. (2013). ARAS Human Activity Datasets in Multiple Homes with Multiple Residents. 232-235. pdf: https://tinyurl.com/DADref1
[2] Jérémie Saives, Clément Pianon, Gregory Faraut. Activity Discovery and Detection of Behavioural Deviations of an Inhabitant from Binary Sensors. IEEE Transactions on Automation Science and Engineering, Institute of Electrical and Electronics Engineers, 2015, 12 (4), pp.1211 - 1224. pdf: https://tinyurl.com/DADref2
[3] van Engelen, J.E., Hoos, H.H. A survey on semi-supervised learning. Mach Learn 109, 373–440 (2020). pdf: https://tinyurl.com/DADref3
[4] Diane Cook and Narayanan Krishnan. Activity learning: Discovering, recognizing, and predicting human behavior from sensor data. Wiley publishing. Feb. 2015