Francesco Vitale

Anomaly Detection by Process Mining and Machine Learning for Industrial Cyber-Physical Systems

Tutor: Prof. Nicola Mazzocca

Cycle: XXXVII Year: Third

Candidate's information

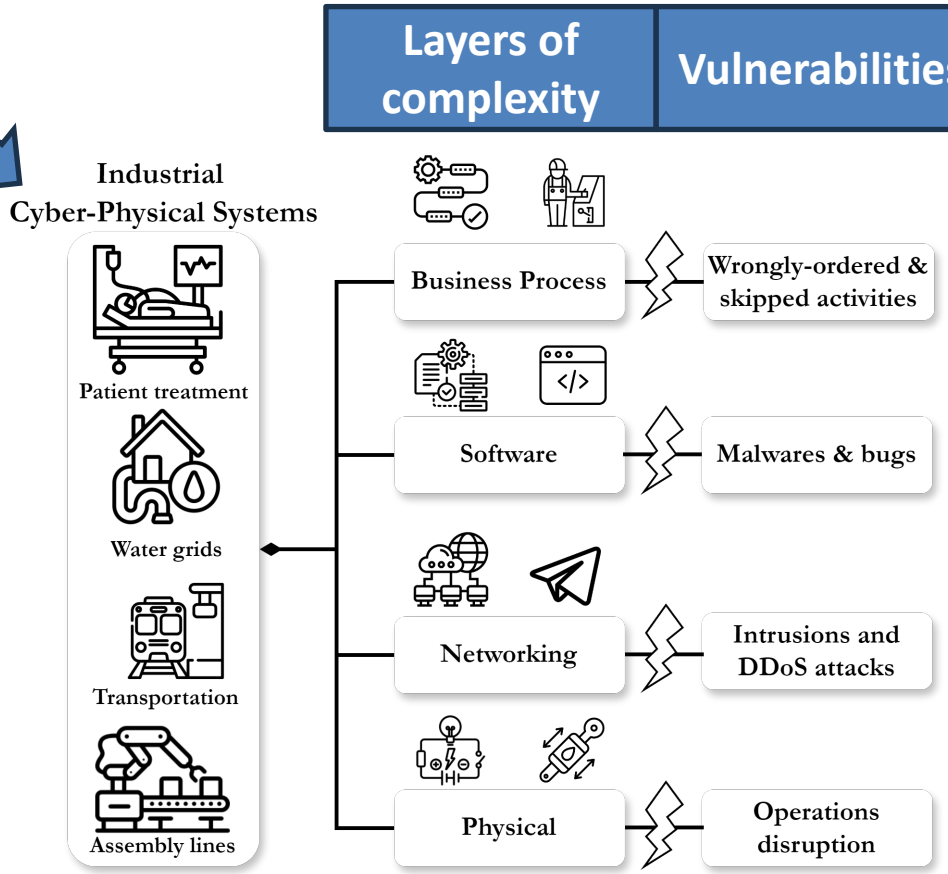
- MSc degree: **Computer Engineering @ UNINA, 2021**
- Research group: **SECLab**
- PhD start date – end date: **01/11/21 – 31/10/24**
- Scholarship type: **Company-funded scholarship**
- Partner company: **Hitachi Rail STS**
- Periods abroad: **Visiting at RWTH Aachen University, 01/02/23 – 31/07/23**

Summary of study activities

- **First year**
  - Courses on computer systems modeling, modern technologies, and data analysis
  - Literature reviews, proof-of-concept evaluations, prototype development for the funding company
  - Supervision of MSc students' theses
  - Scientific writing
- **Second year**
  - Courses on data analysis and English
  - Literature reviews, proof-of-concept evaluations, development of foundational work for my thesis while abroad
  - Scientific writing
- **Third year**
  - Course on scientific writing
  - Literature reviews, refinement of the work done abroad, and development of new research directions
  - Supervision of MSc students' theses
  - Scientific writing

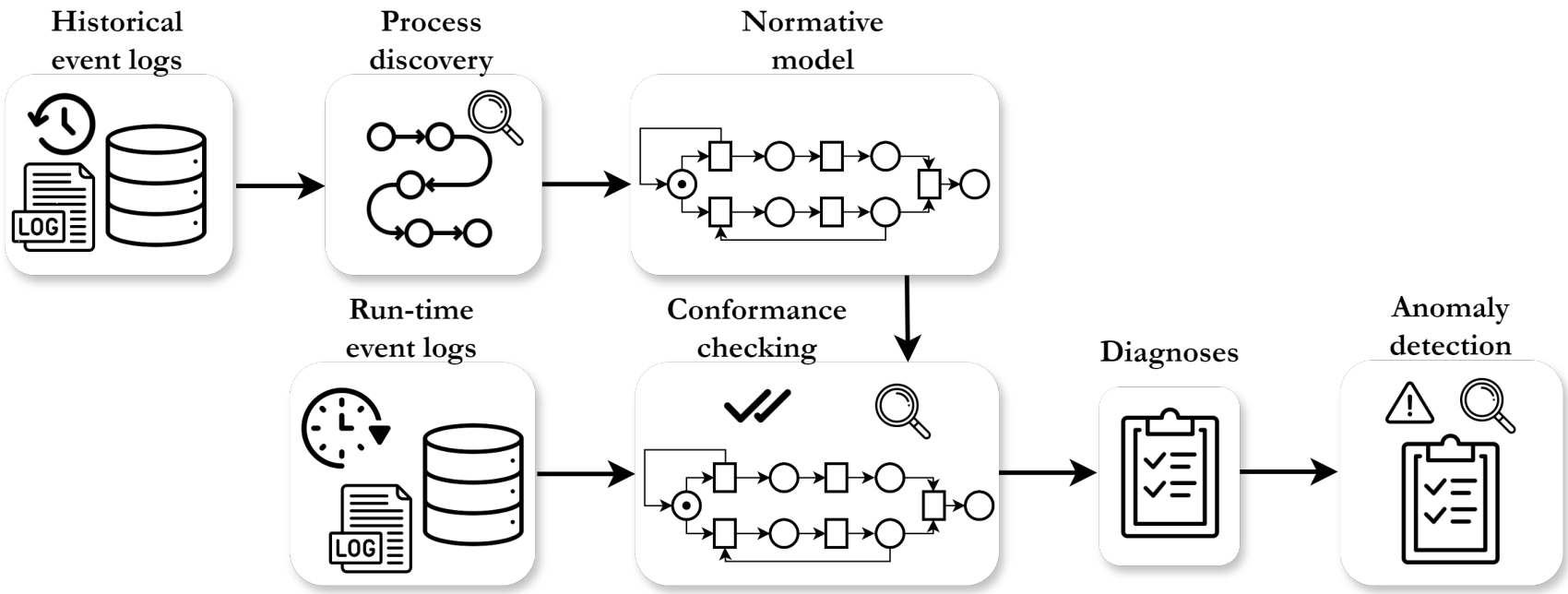
Research area(s)

- My research focuses on:
  1. **Industrial Cyber-Physical Systems (ICPSs)**
     - Several layers of complexity and vulnerabilities
  2. **Anomaly detection**
     - Several data and anomaly types
  3. **Process Mining (PM)**
     - Process discovery
     - Conformance checking
  4. **Machine Learning (ML)**
     - Clustering, dimensionality reduction



Research problem

- **Application of PM for anomaly detection in ICPSs**



Research products

[J1]	P. Singh, M. S. Azari, **F. Vitale**, F. Flammini, N. Mazzocca, M. Caporuscio, and J. Thornadtsson, *Using log analytics and process mining to enable self-healing in the Internet of Things* **Environment Systems and Decisions,** Vol. 42, pp. 234 - 250, 2022, https://doi.org/10.1007/s10669-022-09859-x.
[J2]	A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, and **F. Vitale,** *Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept* **IEEE Transactions on Industrial Informatics,** Vol. 19, pp. 11553 - 11563, 2023, https://doi.org/10.1109/TII.2023.3246983.
[J3]	S. Guarino, F. Vitale, F. Flammini, L. Faramondi, N. Mazzocca, and R. Setola, *A Two-Level Fusion Framework for Cyber-Physical Anomaly Detection* **IEEE Transactions on Industrial Cyber-Physical Systems**, Vol. 2, pp. 1 - 13, 2023, https://doi.org/10.1109/TICPS.2023.3336608.
[J4]	M. Cinque, L. De Simone, N. Mazzocca, D. Ottaviano, and **F. Vitale,** *Evaluating virtualization for fog monitoring of real-time applications in mixed-criticality systems* **Real-Time Systems** Vol. 59, pp. 534 - 567, 2023, https://doi.org/10.1007/s11241-023-09410-4.
[J5]	F. Vitale, F. De Vita, N. Mazzocca, and D. Bruneo, *A Process Mining-based unsupervised Anomaly Detection technique for the Industrial Internet of Things,* **Internet of Things,** vol. 24, p. 100993, 2023, https://doi.org/10.1016/j.iot.2023.100993.

Research products

[J6]	**F. Vitale**, S. Guarino, F. Flammini, L. Faramondi, N. Mazzocca, and R. Setola, *Process Mining for Digital Twin Development of Industrial Cyber-Physical Systems,* **IEEE Transactions on Industrial Informatics,** 2024, https://doi.org/10.1109/TII.2024.3465600.
[J7]	**F. Vitale**, F. Flammini, M. Caporuscio, and N. Mazzocca, *Combining Process Mining and Unsupervised Machine Learning for Monitoring Resilient Computer Systems,* Awaiting first decision, submitted to IEEE Transactions on Dependable and Secure Computing.
[J8]	S. Guarino, **F. Vitale**, E. Del Prete, L. Faramondi, N. Mazzocca, and R. Setola, *d-TV-DBN: A Hierarchical and Distributed Architecture for Scalable Cyber-Physical Anomaly Detection with Bayesian Networks*, Awaiting first decision, submitted to IEEE Transactions on Emerging Topics in Computing.
[J9]	**F. Vitale**, M. Pegoraro, W. M. P. van der Aalst, and N. Mazzocca, *Control-flow anomaly detection by process mining-based feature extraction and dimensionality reduction*, Under review, submitted to Knowledge-based Systems.
[J10]	**F. Vitale**, N. Dall'Ora, S. Gaiardelli, E. Fraccaroli, N. Mazzocca, F. Fummi, *Process Mining-Driven Fault Diagnosis and Simulation for Cyber-Physical Systems,* In progress

PhD thesis overview

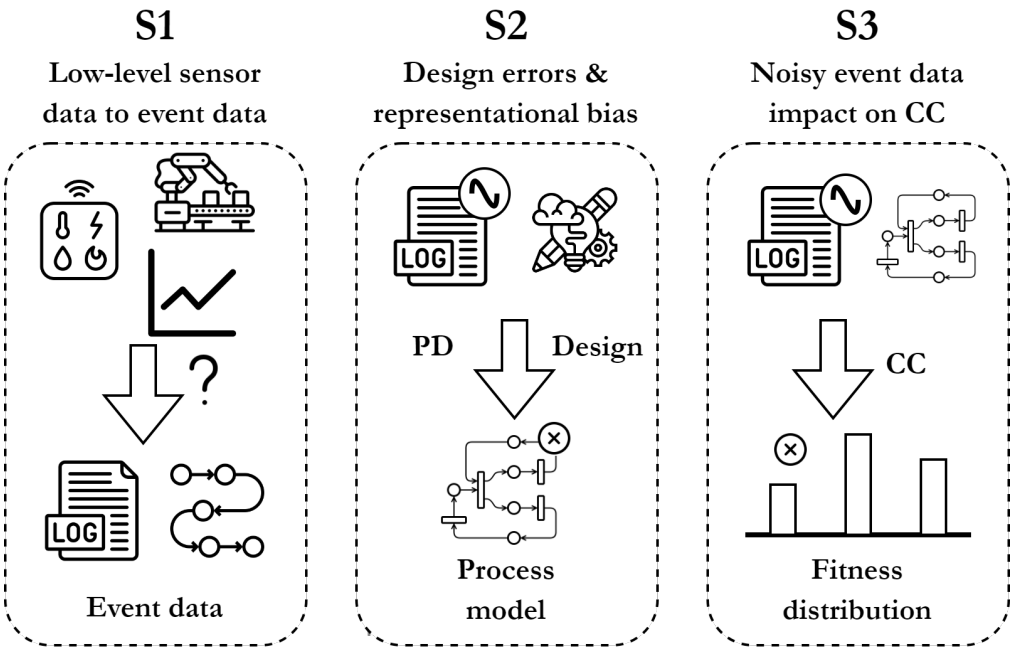
- **Problem statement**
  - **How can PM be used for anomaly detection in ICPSs?**
  - Existing ML solutions are process-agnostic and lack explainability
  - PM for anomaly detection can help to provide process-based descriptions of the ICPS and improve the explainability of the results
- **Objective**
  - **Systematic development of flexible, explainable and effective PM-based anomaly detection for ICPSs**

PhD thesis motivation

S1: The application of PM to low-level sensor data is challenging

S2: The existing PM-based solutions are affected by low-quality process models

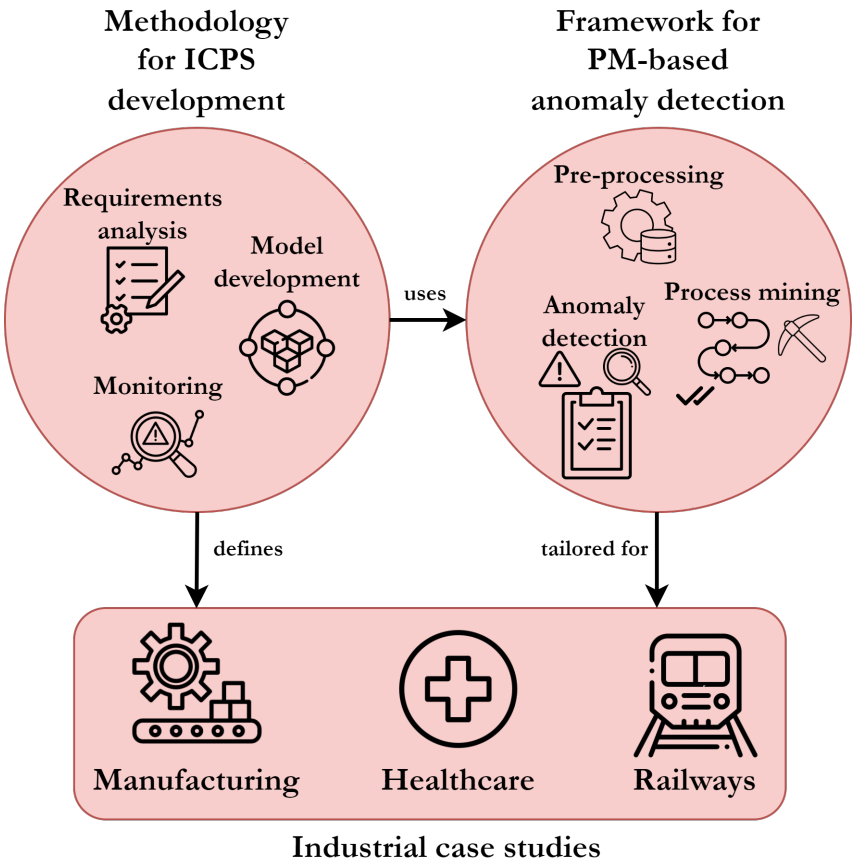
S3: Noisy data affect the results of PM-based solutions too



PhD thesis contributions

Overview

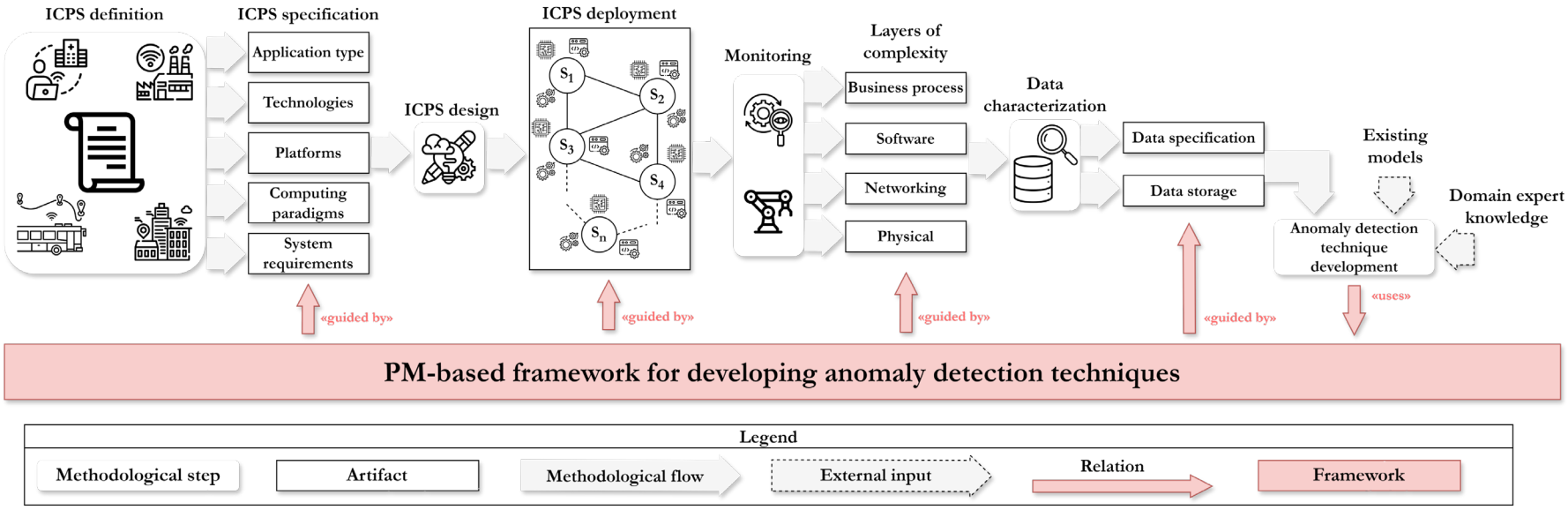
- **Methodology for ICPS development**
- **Framework for PM-based anomaly detection**
- **Validation against several industrial case studies**



PhD thesis contributions

Methodology

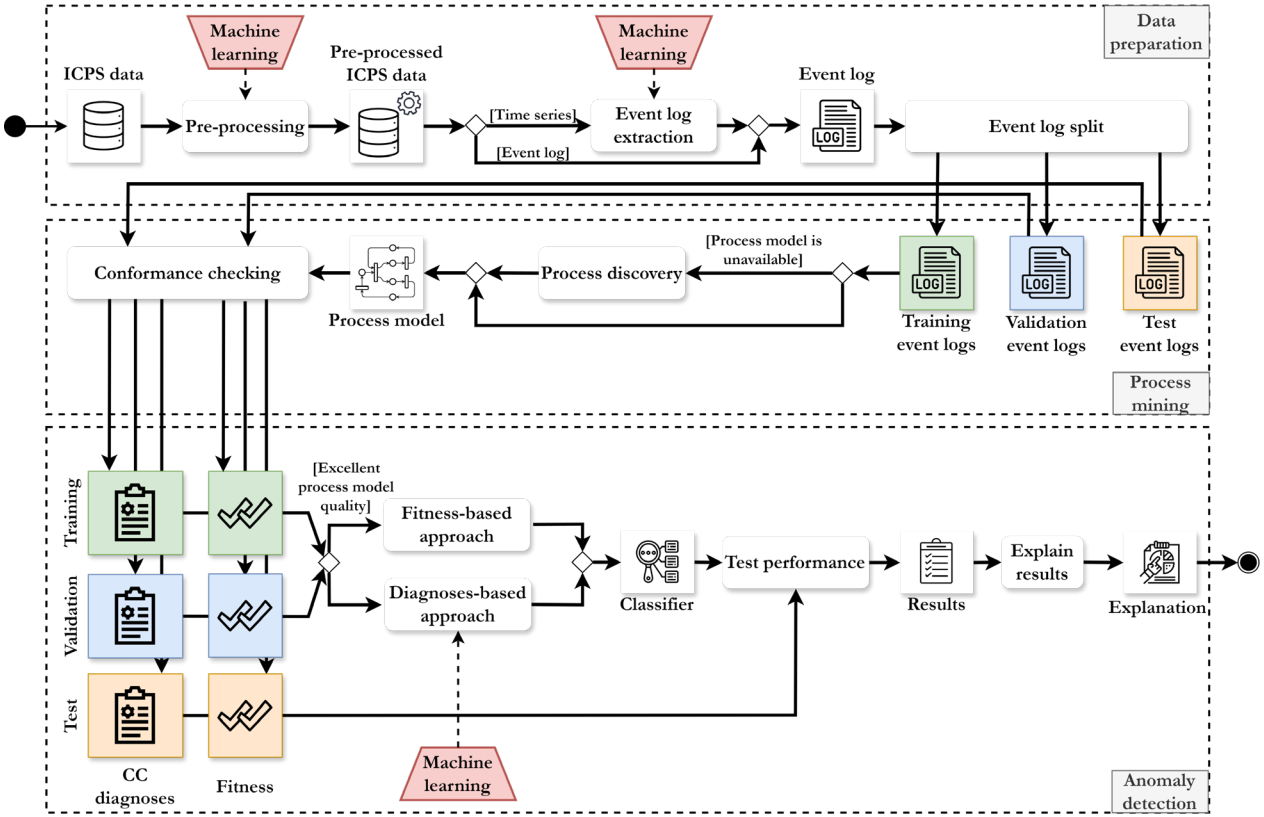
- **The methodology is a procedure to characterize the reference ICPS and identify the type of data and anomalies to detect**



PhD thesis contributions

Framework

- **The framework is a flexible set of steps and tools to integrate PM with ML based on the characterization of the reference ICPS**



PhD thesis contributions

Validation and results

- **Pre-processing with ML enables unsupervised characterization of the reference ICPS and allows PM-based time series anomaly detection (S1)**
- **ML improves the unsupervised characterization, leading to better PM-based time series anomaly detection effectiveness (S3)**
- **PM-based feature extraction combined with ML improves anomaly detection effectiveness while maintaining the explainable nature of PM (S2 and S3)**

Thanks for your attention

Any questions?