21 Juin 2023

# 9 ÈME JOURNÉE RÉGIONALE DES DOCTORANTS EN AUTOMATIQUE

# Anomaly Detection for IoT multivariate time series data with statistical and machine learning techniques


Thi Thuy Van Nguyen[a,b], Kim Phuc Tran[a], Cédric Heuchenne[b]

[a]University of Lille, ENSAIT, GEMTEX - Laboratoire de Génie et Matériaux Textiles, F-59000 Lille, France

[b]HEC Management School - University of Liège, Belgium


## Context

- Good decisions are critical for business success in today competitive global economy.
- IoT applications aid in decision-making and automating daily tasks for convenience
- Embedded AI integrates AI into electronic systems, such as home automation systems, smart wearables, and autonomous vehicles
- Anomaly detection enables quick identification of anomalies or unexpected patterns for effective decision-making.
- Control charts (in SPC): well-established and reliable method, easy construction and user-friendly usage.
- Deep learning (DL) have capability to automatically learn and adapt to complex patterns and variations in data

→ uncover anomalies that may be overlooked by traditional statistical methods,

→ provide a higher level of accuracy compared to traditional statistical methods.

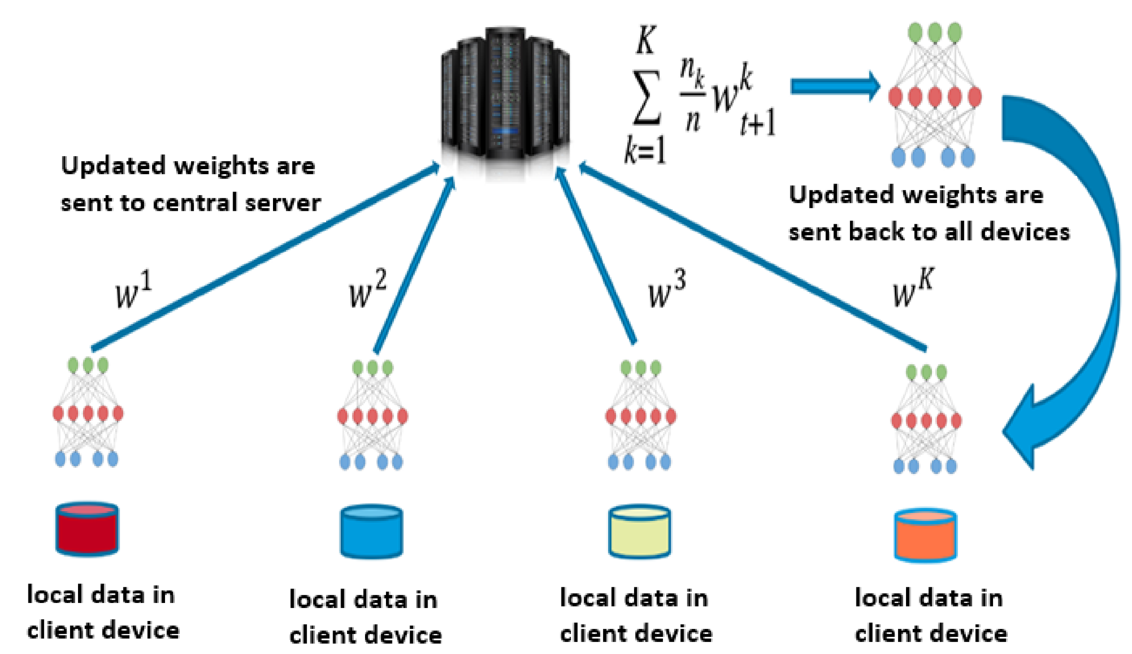
## Problems

- **Control charts** rely on data distribution assumptions & may struggle with complex data → **Deep Learning** can handle these struggles, however:
- Require significant computational resources
- Large amount of industrial data cannot be easily collected from a single silo.
- Training DL models normally requires a large amount of data and it can lead to privacy concerns.
- DL models are like black-box, hard to interpret

## Objective

Develop new Explainable Anomaly Detection (EAD) algorithms which combine embedded AI with IoT technology in a federated setting for IoT Multivariate time series data using statistic and DL techniques

- Design an SVDD control chart based on EMWMA technique to monitor CoDa – a type of data which is often encountered in real life
- Use federated setting to solve issues such as data availability and privacy concerns
- Design a framework for anomaly detection in a federated setting: A robust graph transformer network combining with the OCSVM/SVDD for multivariate time series anomaly detection
- Design of interpretable/explainable-AI based module to help explain the predictions of proposed DL model.




## Methodology

**SVDD control chart based on MEWMA technique for monitoring CoDa**

- SVDD is an one-class algorithm: detects **abnormal** observations by **modeling the normal ones**

$$\min_{R,\mathbf{a}} R^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{subject to} \quad \|\mathbf{x}_i - \mathbf{a}\| \le R^2 + \xi_i, \quad i = 1,\dots,n; \ \ \xi_i \ge 0$$

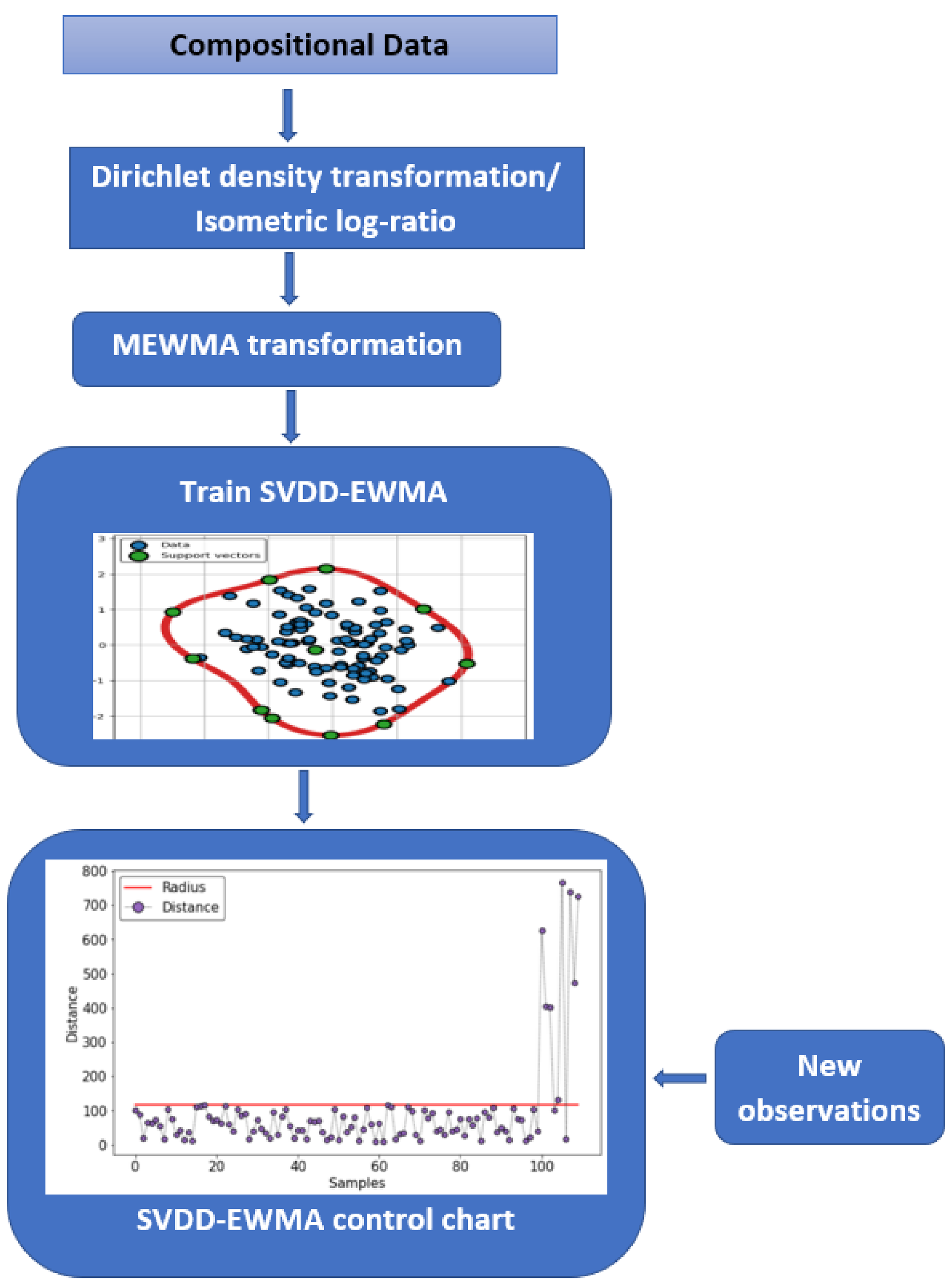
- CoDa is special data: Its components are strictly positive and sum to a constant → can not be treated as normal data.

→ introduce a new transformation method: **Dirichlet density transformation**: decrease the dimension of CoDa, remove constant constraint

- Incorporate MEWMA technique: Transform data using MEWMA formula

$$\mathbf{w}_i = r(\mathbf{x}_i - \boldsymbol{\mu}_0) + (1-r)\mathbf{w}_{i-1},\ i = 1, 2, \dots$$

→ leverage all available observed data, both current and historical, to assess the performance of the process




## Results

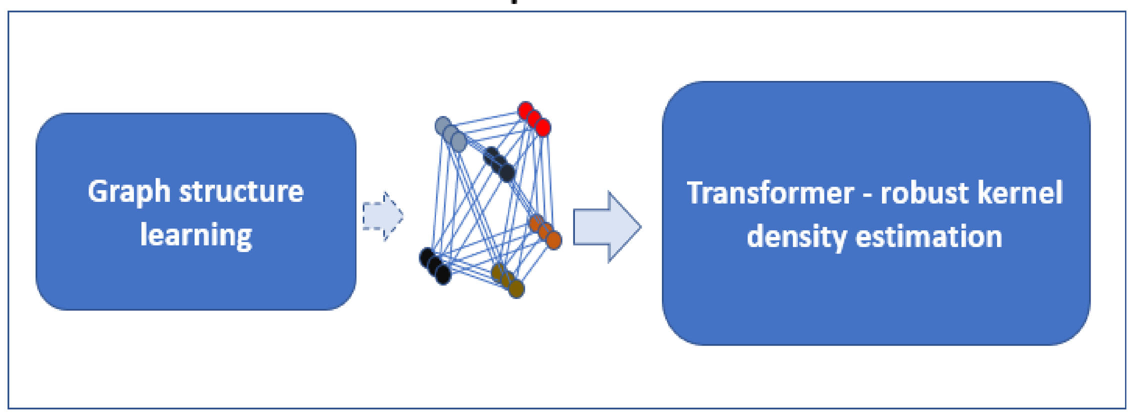
- Dirichlet density transformation: eliminating inherent constraints and reducing data dimensionality in CoDa
- SVDD-EWMA control charts perform well without requirement of data distribution
- Two SVDD-EWMA control charts outperform the classical MEWMA-CoDa for monitoring CoDa in term of Average Run Length

## Developing RGT

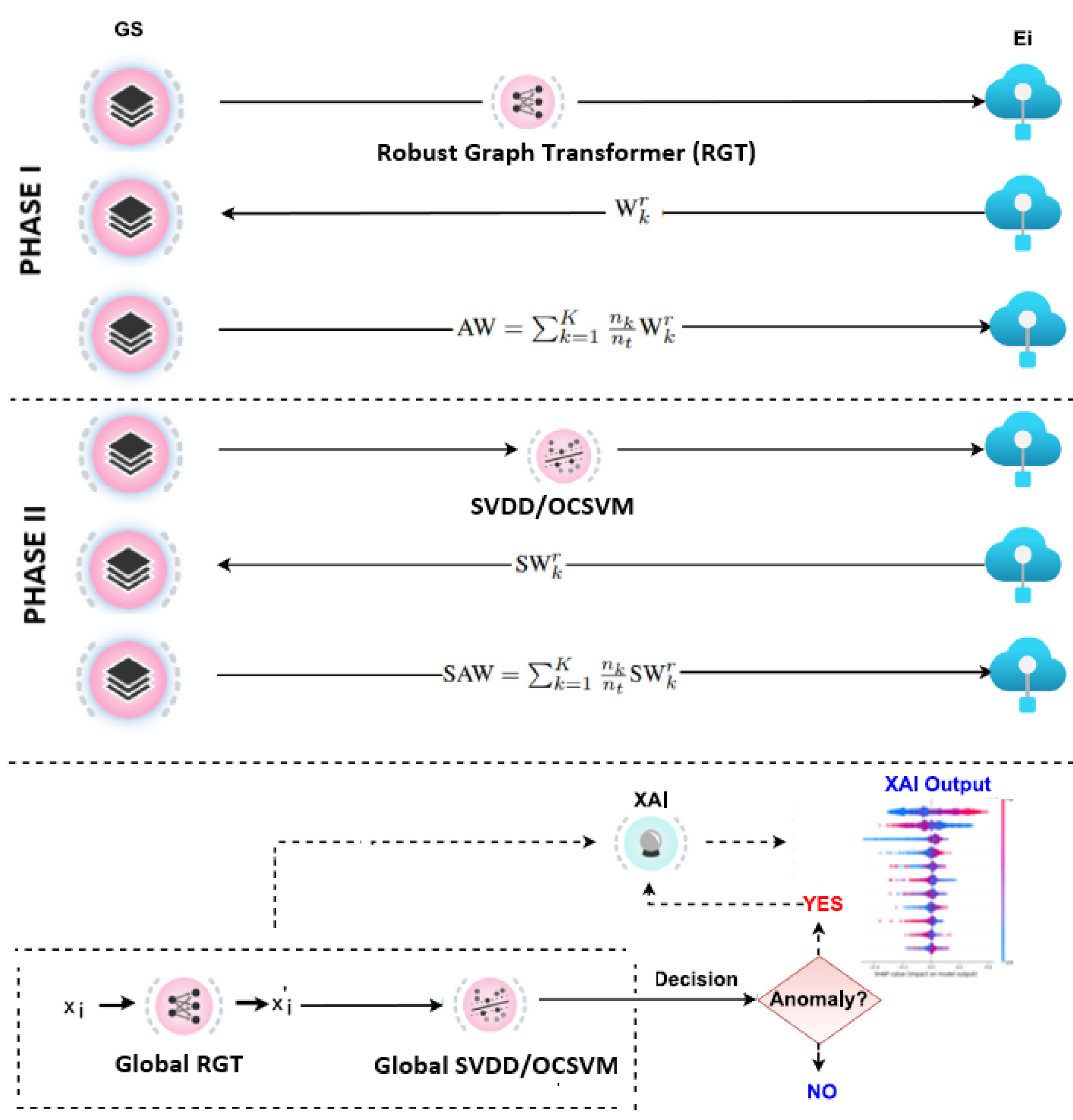
Propose developing a **robust graph transformer** (RGT) for AD in a federated setting.

- Federated learning: handle data privacy
- RGT: extract important information from the data
- OCSVM/SVDD: detect abnormal objects from the inputs
- XAI: enhances transparency, and trustworthiness

**Robust Graph Transformer**




**Proposed framework**




## Conclusion

- Control charts are easy to implement and user-friendly usage, but struggle data assumption and complex data
- Proposed SVDD-EWMA control charts overcomes data distribution assumption, outperform traditional MEWMA-CoDa
- Future direction: a robust graph transformer in a federated architecture together with XAI is proposed to detect anomalies in complex data

## Références


T.T. Van Nguyen, C. Heuchenne, and K. P. Tran. "Machine learning for compositional data analysis in Support of the Decision Making Process." *Machine Learning and Probabilistic Graphical Models for Decision Support Systems*. CRC Press, 2021. 184-215.