Towards frugal unsupervised detection of subtle abnormalities in medical imaging

Geoffroy Oudoumanessah[1,2,3], Carole Lartizien[2], Michel Dojat[1], Florence Forbes[3]
1. Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut des Neurosciences 2. Univ. Lyon, CNRS, Inserm, INSA Lyon, UCBL, CREATIS, UMR5220, U1294, F-69621 3. Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK
Contact: geoffroy.oudoumanessah@inria.fr

Green Unsupervised Anomaly Detection (UAD)

- Deep learning has raised environmental concerns
- Green AI solutions [1] have not permeated medical imaging domain yet
- We propose a **frugal approach** of UAD in medical imaging:
  - Likelihood approach with **online learning**
- Application to the detection of **subtle abnormalities** in *de novo* Parkinsonian patients
- **PPMI** dataset [2] that contains **controls (HC)** and patients (PD)
- Assess the evolution of the disease through the **Hoehn and Yahr (HY)** score at stages 1 and 2

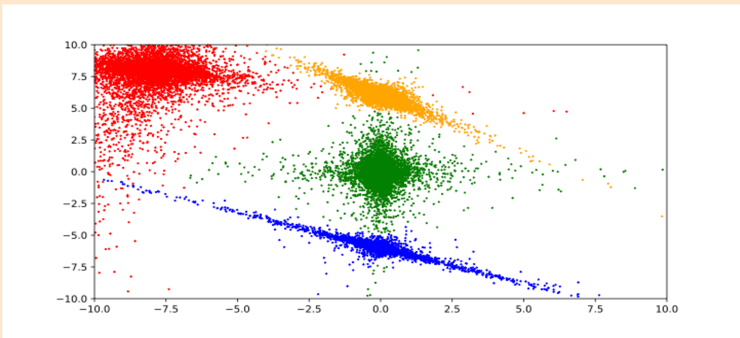
UAD Steps with mixture models

Reference Model

Trained on the set of healthy voxels

$$\mathbb{Y}_H = \{\mathbf{y}_v, v \in \mathbb{V}_H\}$$

Mixture model of multiple scale t-distributions $\mathcal{MST}$ [X]

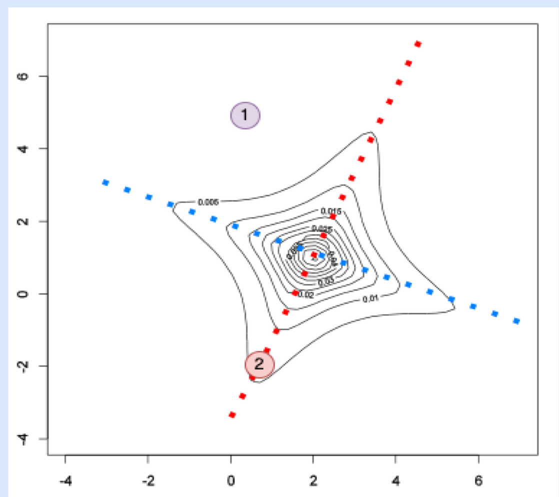


$$f_H(\mathbf{y}_v; \mathbf{\Theta}_H) = \sum_{k=1}^{K} \pi_k f(\mathbf{y}_v; \boldsymbol{\theta}_k)$$

Proximity measure

Quantify the proximity of an observation towards the reference model

$$r(\boldsymbol{y}_v, \mathbf{\Theta}_H) \propto \max_{m=1:M} \left(\frac{\nu_m}{2} + \frac{[\boldsymbol{D}^T(\mathbf{y}_v - \boldsymbol{\mu})]^2_m}{2A_m} \right)^{-1}$$



Decision Rule

Choose an appropriate False Positive Rate and compute the empirical quantile

$$\mathbb{P}\left[r(\mathbf{y}_v, \mathbf{\Theta}_H) < \tau_\alpha\right] = \alpha$$

Decision rule:

$$\begin{cases} \mathbf{y_v} \text{ is an anomaly} & \text{if } r(\mathbf{y}_v, \mathbf{\Theta}_H) < \tau_\alpha \\ \mathbf{y_v} \text{ is not an anomaly} & \text{otherwise} \end{cases}$$

Alternative UAD methods and their costs

- Reconstruction error (RE) [3]:
  - Patch based method
  - Inefficient on subtle anomalies
- FastFlow (FF) [4]:
  - Patch based method
  - Model the latent distribution
  - Faster than RE

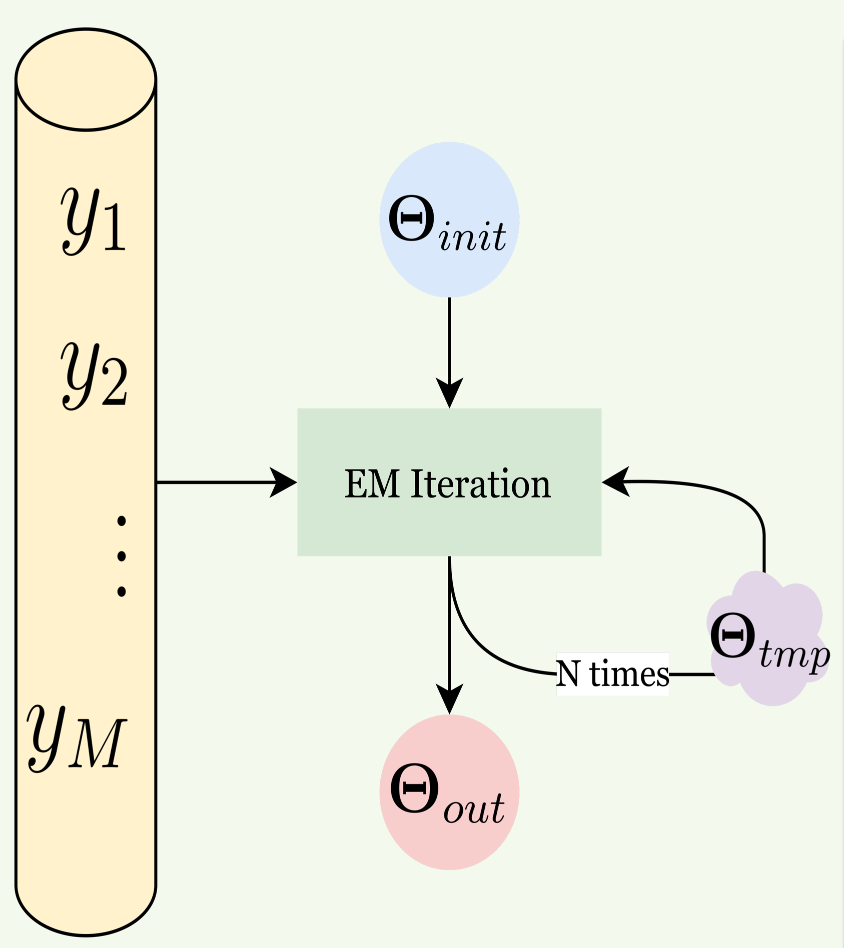
Both methods rely on an **auto-encodeur** architectures:

- Imply a lot of parameters to estimate
- Highly energy-consuming for training and inference

Method	Backend	Training Time	Training Consumption	Training DRAM peak	Inference Time	Inference Consumption	Inference DRAM peak	Gmean	Parameters
Online Mixtures (ours)									
OGMM	CPU	**50s**	**85 kJ**	**494 MB**	**17min**	**23 kJ**	**92 MB**	0.65	140
OMMST	CPU	1min20	153 kJ	958 MB	18min	32 kJ	96 MB	**0.67**	**128**
Lightweight AE									
RE	GPU	1h26	5040 kJ	26 GB	3h30	8350 kJ	22 GB	0.61	5266
FF	GPU	4h	6854 kJ	27 GB	3h53	13158 kJ	27 GB	0.55	1520
Resnet-18									
RE	GPU	17h40	53213 kJ	26 GB	59h	108593 kJ	28 GB	0.64	23730218
FF	GPU	4h10	7234 kJ	28 GB	19h45	18481 kJ	28GB	0.61	1520

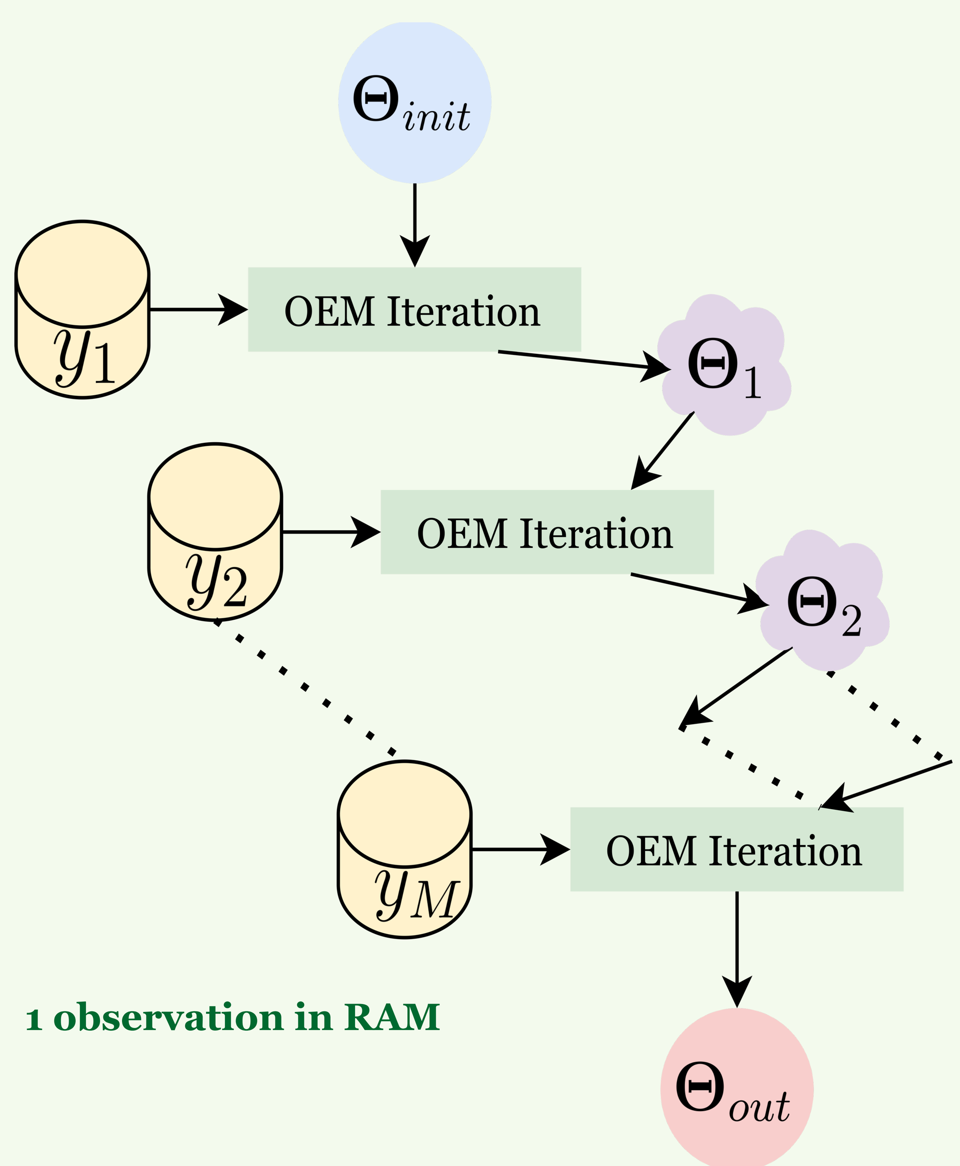
Energy & Memory efficient learning of the reference model: Online EM (OEM) [6,7]

Batch EM



M observations in RAM

Online EM



1 observation in RAM

$$f_{\mathcal{MMST}}(\mathbf{y}; \mathbf{\Theta}) \propto \exp(\boldsymbol{\phi}(\mathbf{\Theta})\mathbf{s}(\mathbf{y}) - \mathbf{\Psi}(\mathbf{\Theta}))$$

EM iteration

$$\begin{cases} L(\mathbf{\Theta}_{n-1}) & = \sum_{i=1}^{M} \mathbb{E}_{\mathbf{\Theta}_{n-1}}[\log f_{\mathcal{MMST}}(\boldsymbol{Y}_i; \mathbf{\Theta}_{n-1})] & \text{E-step} \\ \mathbf{\Theta}_n & = \text{argmax } L(\mathbf{\Theta}_{n-1}) & \text{M-step} \end{cases}$$

Cost: M expectation computations

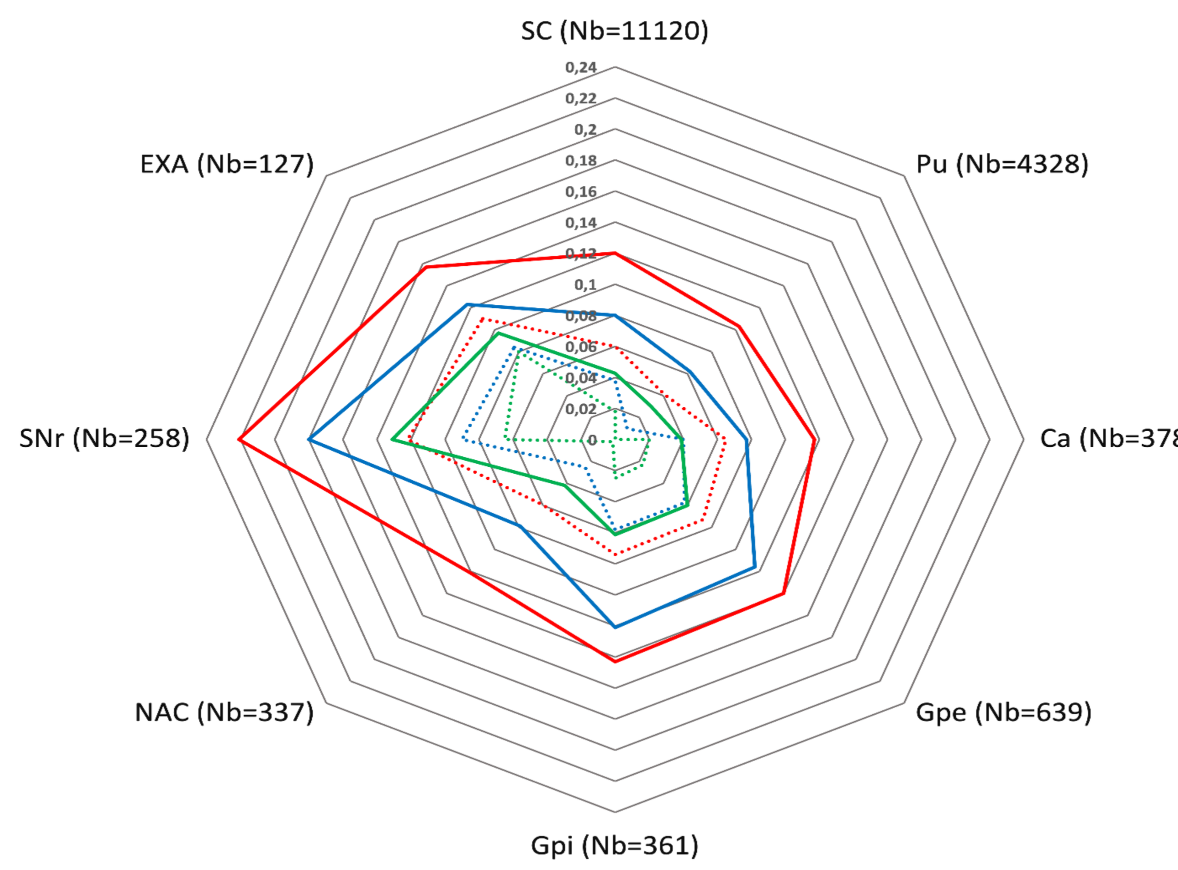
OEM iteration

$$\begin{cases} S_n & = (1-\gamma_n)S_{n-1} + \gamma_n \mathbb{E}_{\mathbf{\Theta}_{n-1}}[\boldsymbol{s}(\boldsymbol{Y}_n)] & \text{E-step} \\ \mathbf{\Theta}_n & = \text{argmax } \boldsymbol{\phi}(\mathbf{\Theta}_{n-1})S_n + \mathbf{\Psi}(\mathbf{\Theta}_{n-1}) & \text{M-step} \end{cases}$$

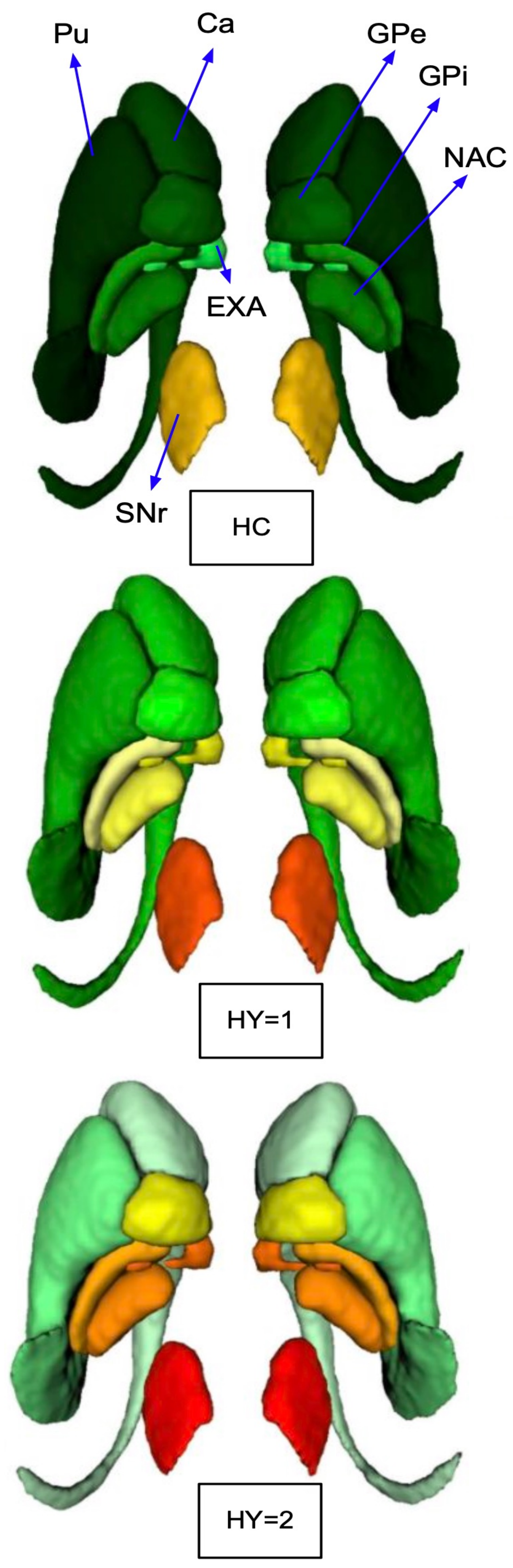
Cost: 1 expectation computation

Parkinson disease application

- Increasing number of abnormalities.
- The **Substancia Nigra Reticula** (SNr) greatly impacted
- Classification of controls and patients
  Gmean of **0.67** for *MMST* and **0.65** for *GMM*



Percentage of anomalies per structure
controls, HY=1, HY=2, MMST, GMM



Conclusion

- Design of an **energy and memory efficient method for UAD in medical imaging**
- We aim at enhancing our cost-effective method to effectively handle both **high-dimensional** and **spatial** information for UAD

1. Schwartz, R., Dodge, J., Smith, N., Etzioni, O.: Green AI. Commun. ACM 63(12), 54–63 (2020)
2. Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C., Caspell-Garcia, C., Simuni, T., Jennings, D., M. Tanner, C., Q. Trojanowski, J., Shaw, L., Seibyl, J., Schuff, N., Singleton, A., Kieburtz, K., W. Toga, A., Mollenhauer, B., Galasko, D., M. Chahine, L.: The Parkinson's progression markers initiative - establishing a PD biomarker cohort. Ann. Clinical Translational Neurology p. 1460–1477 (2018)
3. **Pinon, N., Oudoumanessah, G., Trombetta, R., Dojat, M., Forbes, F., Lartizien, C.: Brain subtle anomaly detection based on auto-encoders latent space analysis**
4. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Un- supervised anomaly detection and localization via 2D normalizing flows. ArXiv abs/2111.07677 (2021)
5. Forbes, F., Wraith, D.: A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering. Statistics and Computing 24(6), 971–984 (2014)
6. McLachlan, G.J., Krishnan, T.: The EM algorithm and extensions. John Wiley & Sons (2007)
7. Cappé, O., Moulines, E.: On-line Expectation-Maximization algorithm for latent data models. Journal of the Royal Statistical Society B 71, 593–613 (2009)