Retrieval Over Training:
Similarity search-based Model Selection for Time Series Anomaly Detection

Christos Panourgias, Roberto Stanzione*, Adrien Petralia, Themis Palpanas, Paul Boniol*
** Equal contribution*

NeurIPS 2026









No one-size-fits-all

- No single anomaly detector **universally outperforms others**
- We need:
  - in-depth **domain knowledge**
  - **understanding** of detectors
- For this reason, an emerging research direction focuses on **automated solutions**

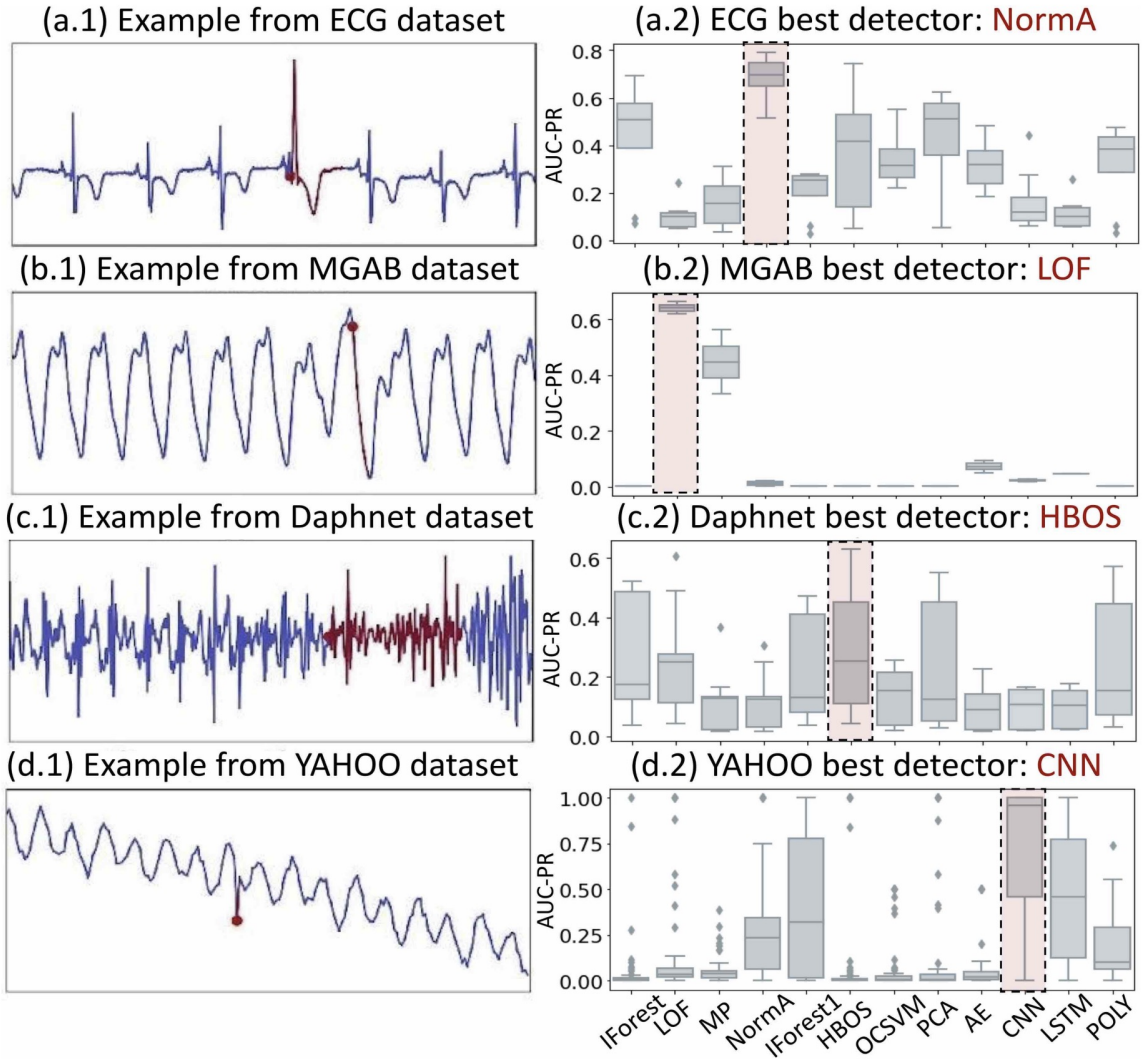


Image from: Sylligardos, E., Boniol, P., Paparrizos, J., Trahanias, P., & Palpanas, T. (2023). Choose wisely: An extensive evaluation of model selection for anomaly detection in time series. *Proceedings of the VLDB Endowment*, *16*(11), 3418-3432.

Model Selection: Current Limitations

L1 *Existing automatic solutions for time-series anomaly detection involve a trade-off between robustness, flexibility, and computational efficiency*

L2 *Model selection addresses this issue by choosing a single detector, but most approaches rely on trained models and are tied to a fixed detector pool.*

L3 *Model selectors may generalize poorly under domain shift*

Can model selection be solved without training a selector at all?

RAMSAD: a different paradigm

- We propose **RAMSAD** (*Retrieval-Augmented Model Selection for Anomaly Detection*)

- The idea is simple: **similar time series** are likely to favor the **same anomaly detector**.

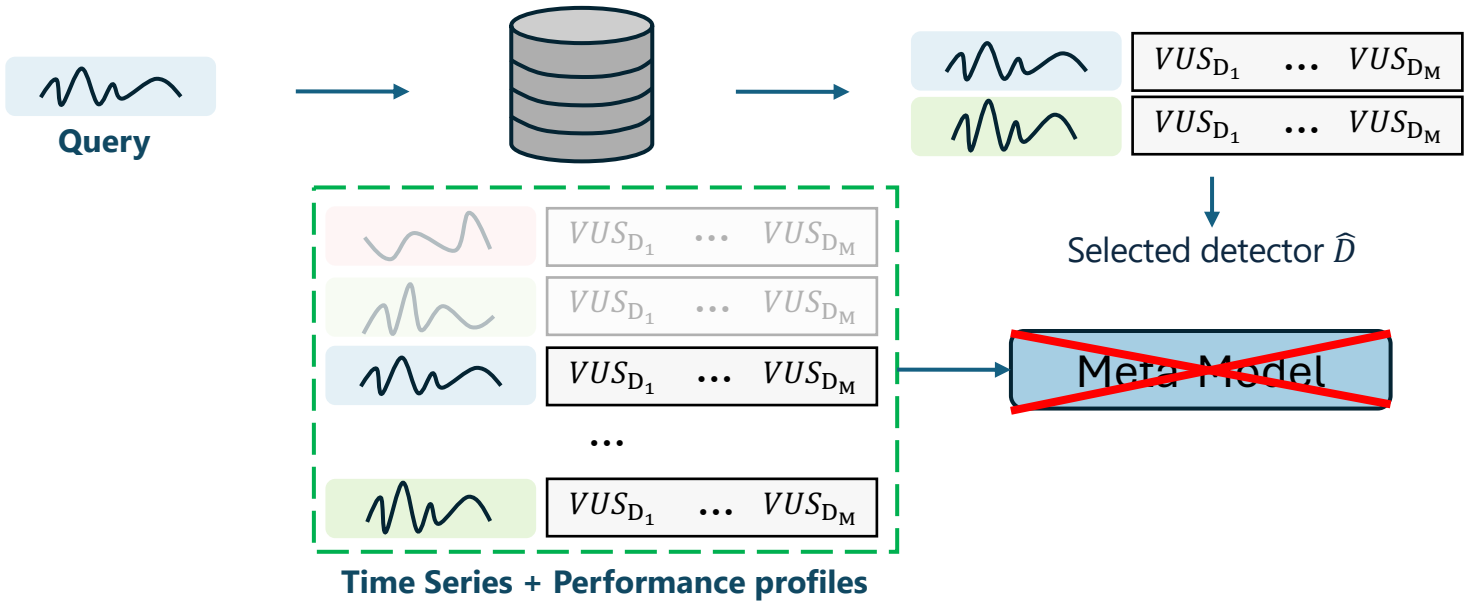
- We define a **search space** by leveraging:
  - **Previous knowledge** with **TSFM** models
  - **Encoding** additional **semantic knowledge**

- We **turn model selection into a similarity-search problem**, avoiding a trained selector



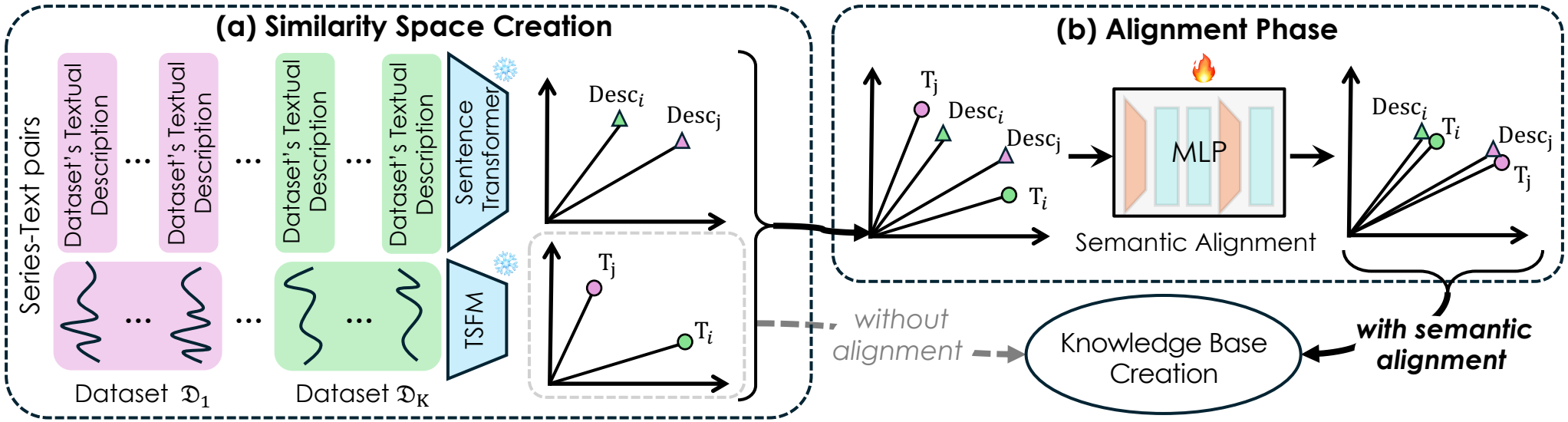
Panourgias C., Stanzione R., Petralia A., Palpanas T. & Boniol P. (2026).
Retrieval over training: similarity search-based model section for time series anomaly detection.
NeurIPS 2026.

RAMSAD: a different paradigm



- The **knowledge base** contains time series and the performance of detectors
- Given a new query series, we retrieve its **top-k** nearest neighbors.
- According to the performanes on the **k** series, we choose a detector
- Because performance is stored per series and per detector, **the detector pool and the database can change without retraining anything**

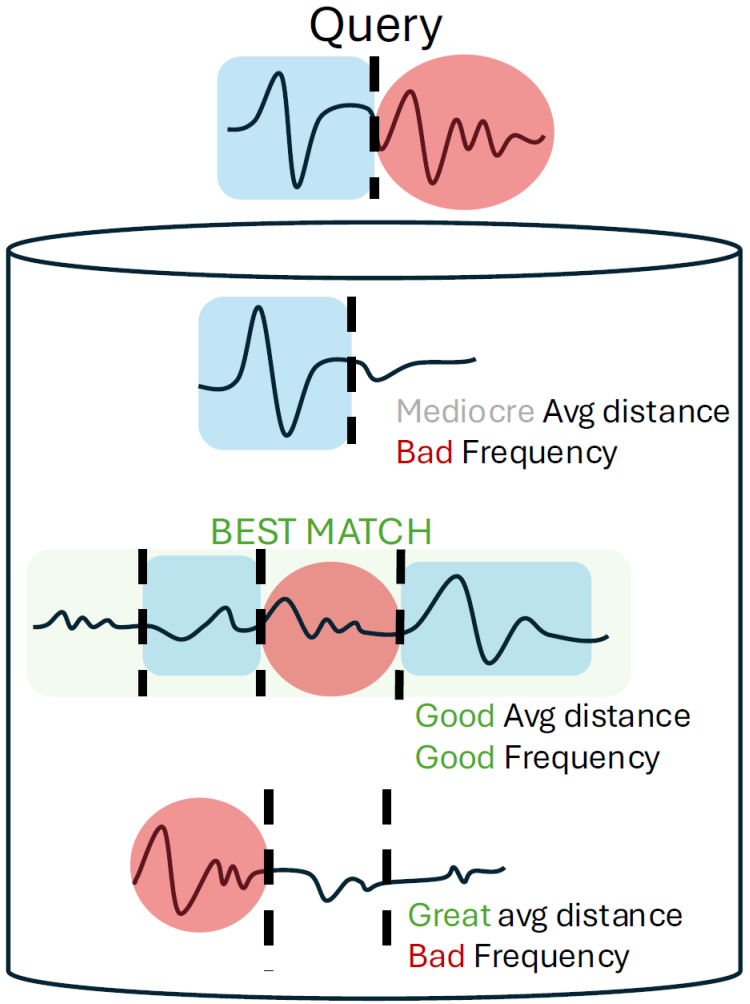
RAMSAD: Alignment



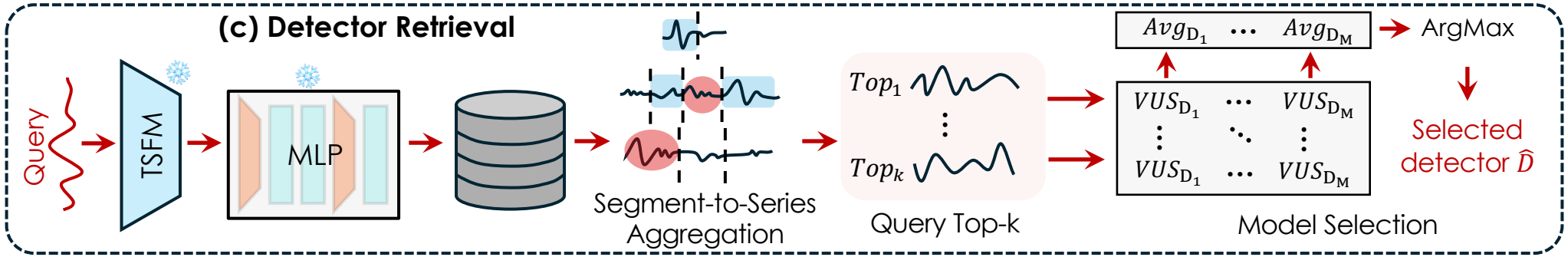
RAMSAD: Retrieval

- Different query segments may retrieve **segments from the same series**, so we **group** them segments by their source

- For each candidate series we compute:
  - the **average distance of its matched segments**
  - **its frequency** (fraction of query segments for which that series appears among the neighbors)
  - The final score is the **average distance divided by the frequency**, so lower is better



RAMSAD: Retrieval

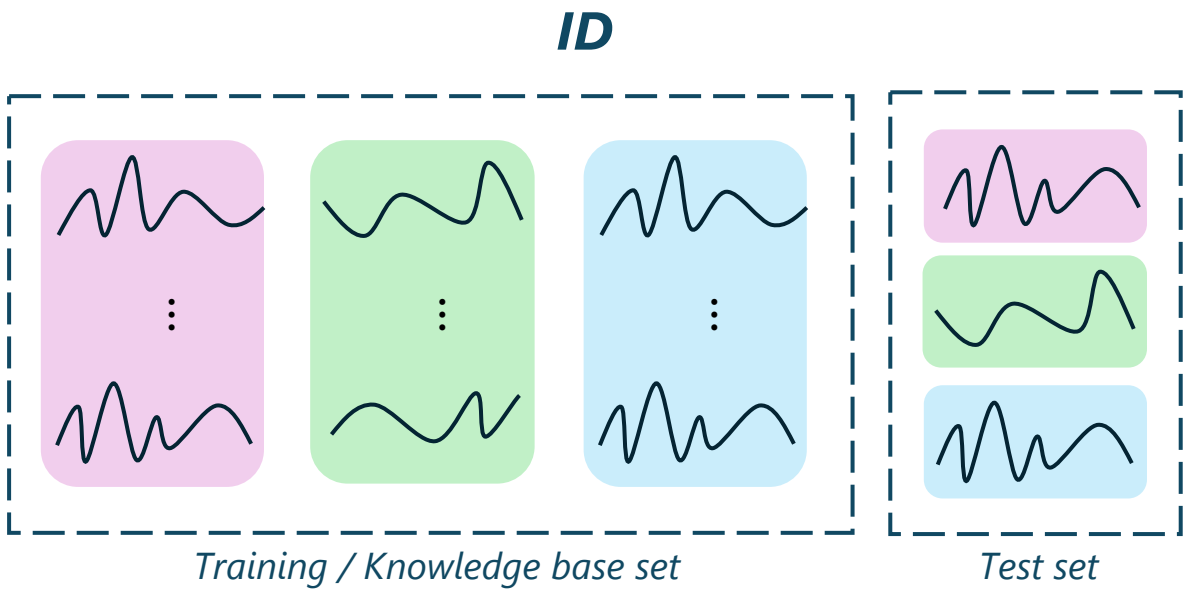


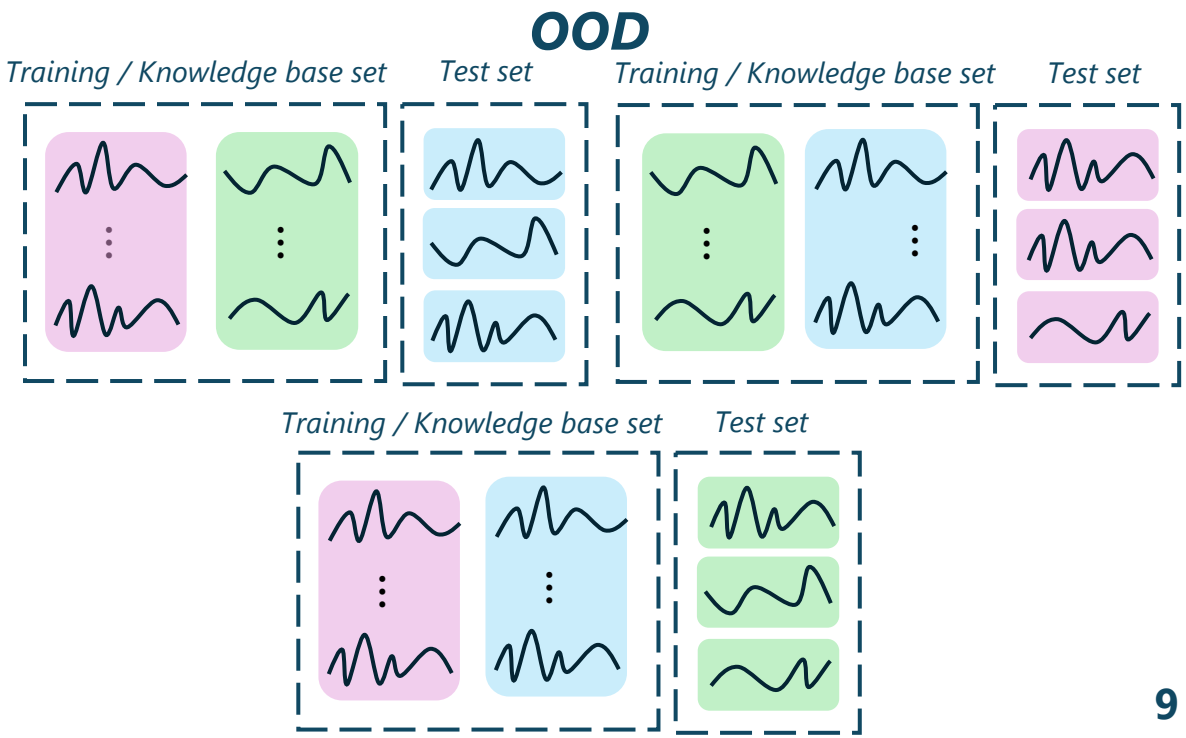
- For each candidate detector, we look up its **VUS-PR** on every series in the neighborhood and **average it**
- The detector with the highest average is selected:
  - to build a small ensemble, we take the top **N** detectors
- Result: a recommendation obtained without training any selector, without labels for the query, and in about 20 ms

Experimental Setting

- **In Distribution (ID):** For a given domain, we have time series both in the training/knowledge base set and in the test set
- **Out-of-Distribution (OOD):** For each domain, we consider time series from other domains as training/knowledge base set

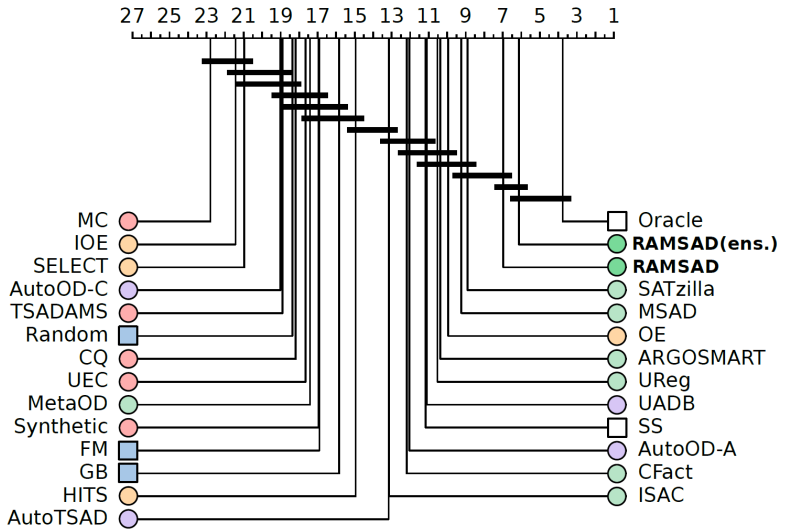




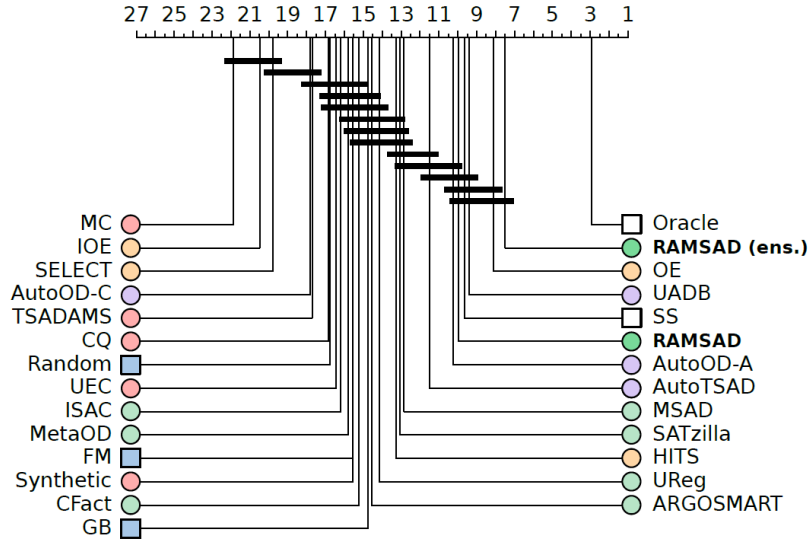


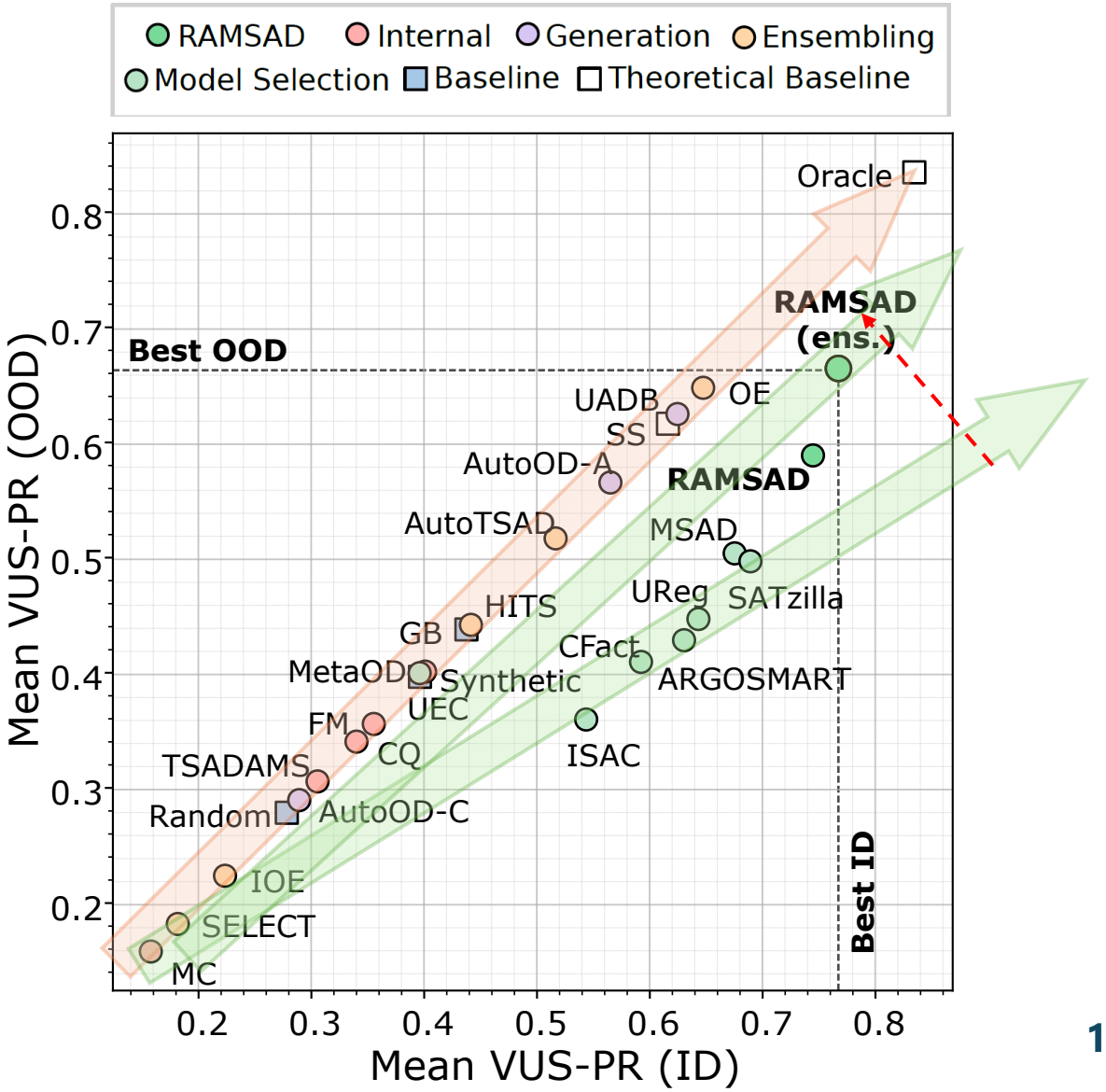
Results

ID

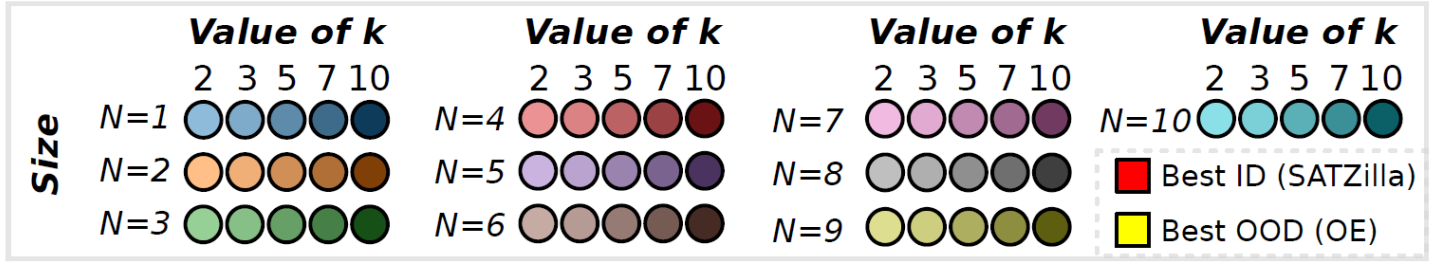


OOD



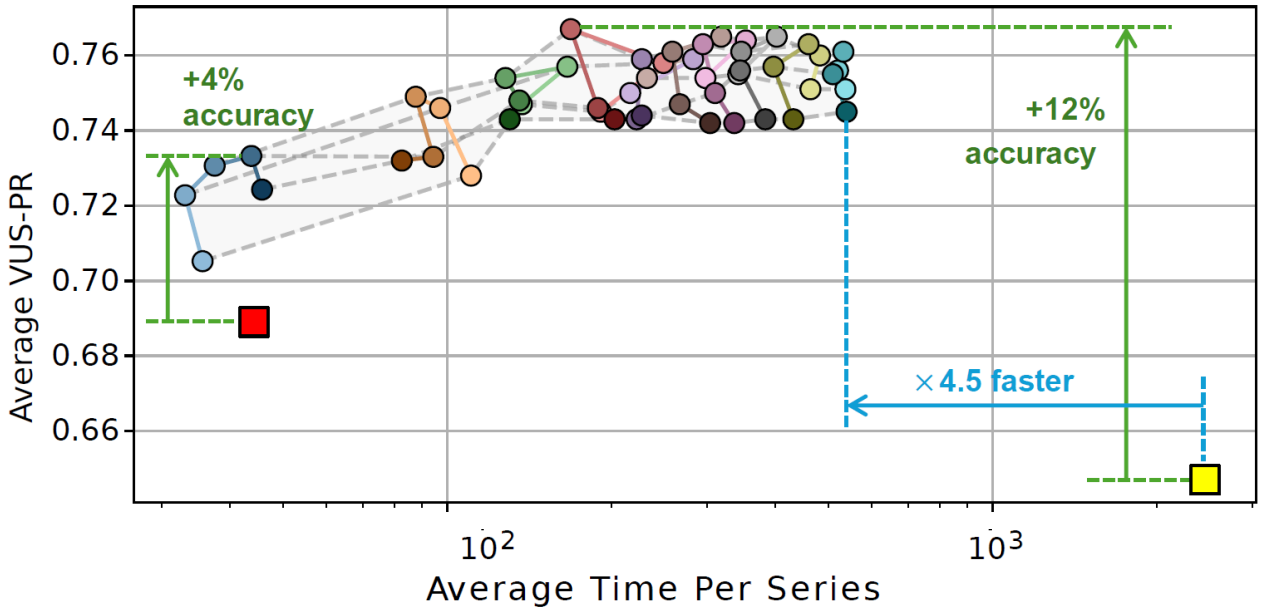


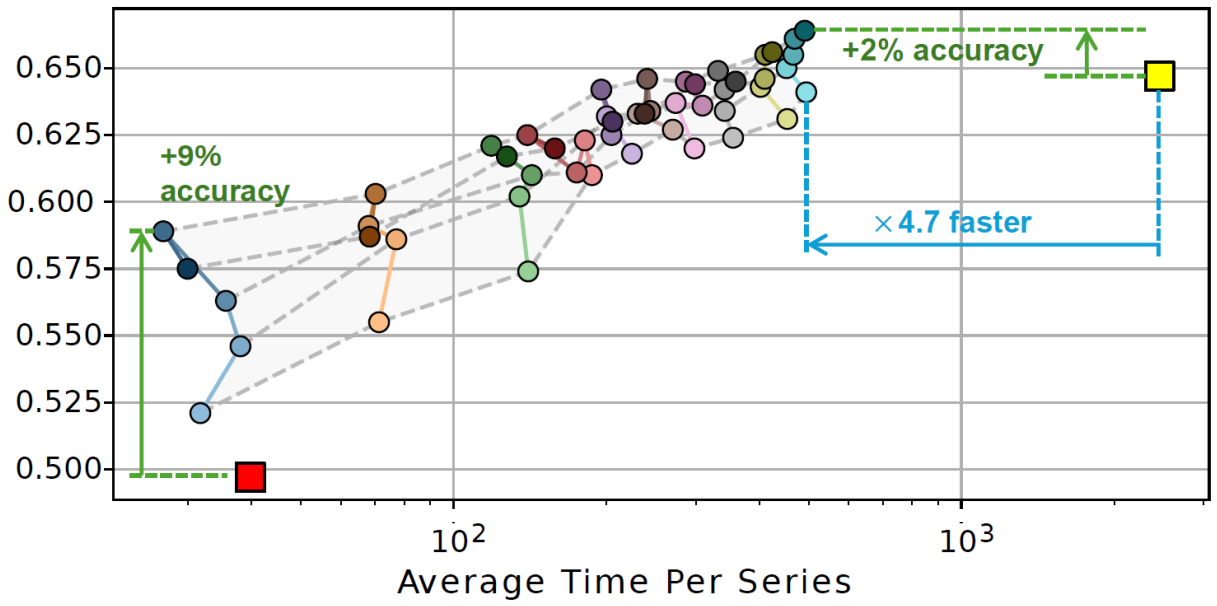
Effects of k and N



k=number of retrieved **examples**

N=number of **models** chosen according to the k examples



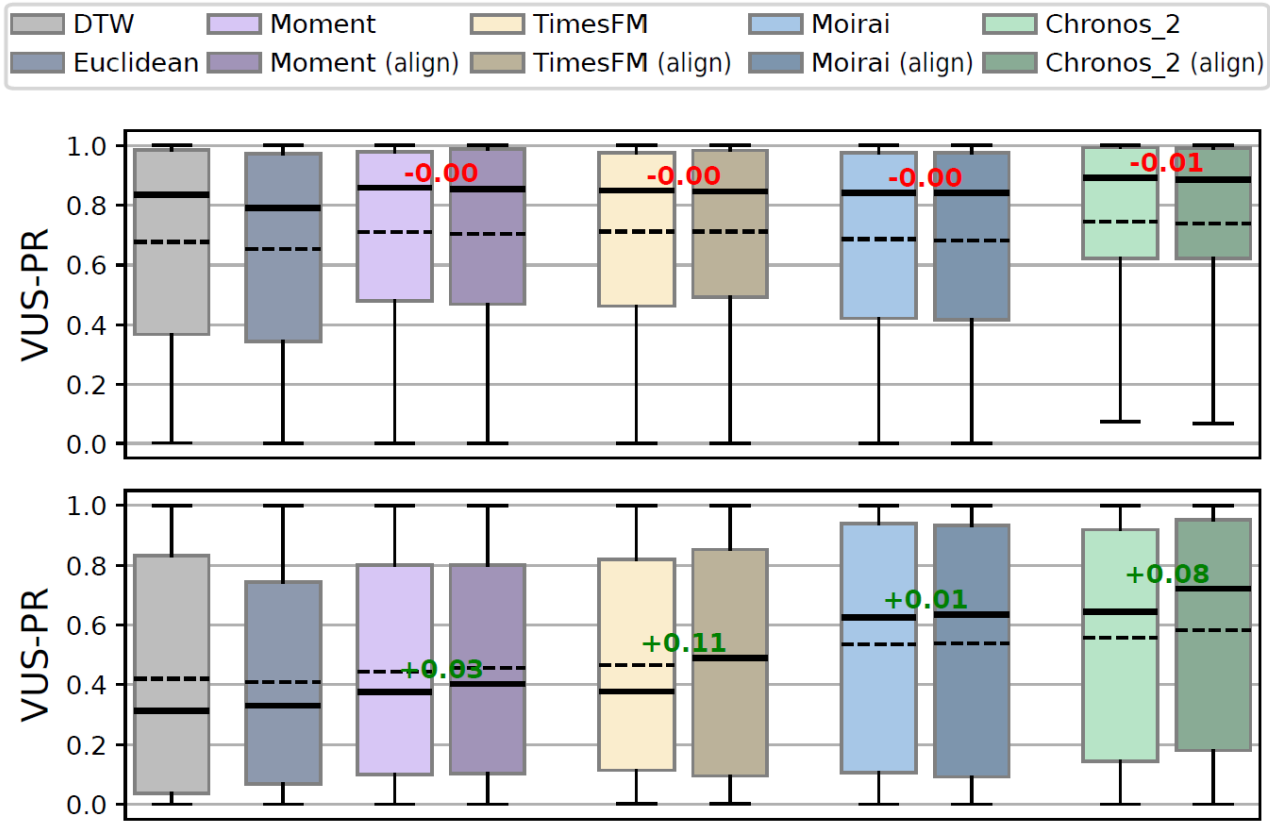


Comparison of similarity spaces

- We compare **three retrieval spaces**:
  - **Original space** (DTW, Euclidean)
  - **TSFM embedding**
  - **Semantically aligned** embeddings.

- **In Distribution:** Aligned-TSFM are similar to the plain TSFM space

- **Out of Distribution:** Aligned-TSFM improve on plain TSFM.
  - In OOD, temporal patterns alone may be insufficient, and incorporating semantic info leads to improvements



Conclusions

- **Retrieval instead of training:** RAMSAD picks a detector by looking at similar past series, so there is no selector to train.
- **Best model selector:** it beats all model-selection baselines.
- **Strong ensembles at lower cost:** near-oracle performance, beating the best ensemble while running only 1/3 of the detectors, 4.7× faster.
- **Robust out-of-distribution:** semantically aligned TSFM spaces give the best retrieval, especially OOD.

THANK YOU!

Paper

Repo

Retrieval Over Training: *Similarity search-based Model Selection for Time Series Anomaly Detection*

Christos Panourgias, Roberto Stanzione*, Adrien Petralia, Themis Palpanas & Paul Boniol*

** Equal contribution*

panourgiaschris@gmail.com