



Learning Compact Representations from Unlabeled ICU Time Series

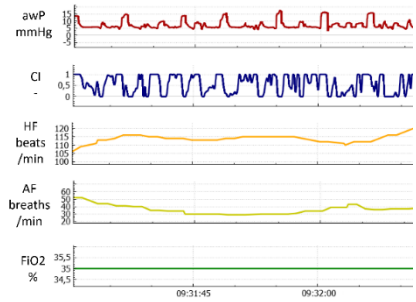
Motivation

- ▶ ICU monitoring produces large amounts of unlabeled high-frequency time series data
- ▶ Dense clinical labels are rare, time-consuming, and often event-specific
- ▶ Large parts of the data remain unused for data-driven analysis

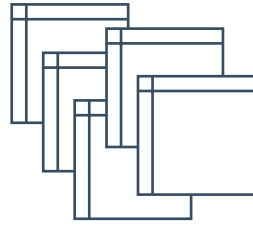
Motivation

- ▶ ICU monitoring produces large amounts of unlabeled high-frequency time series data
- ▶ Dense clinical labels are rare, time-consuming, and often event-specific
- ▶ Large parts of the data remain unused for data-driven analysis
- ▶ Unsupervised learning solution: Clustering

Motivation

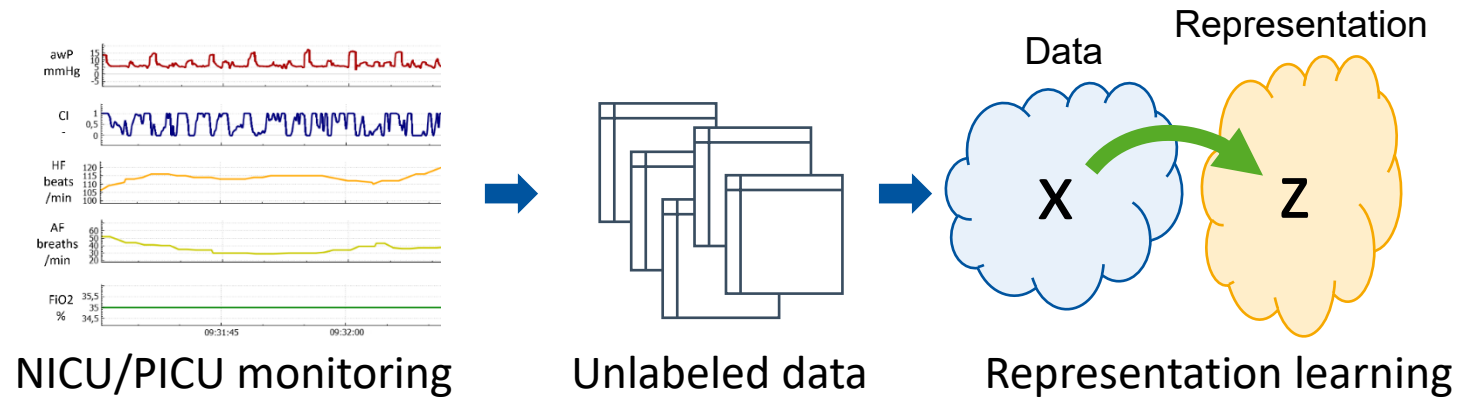


NICU/PICU monitoring

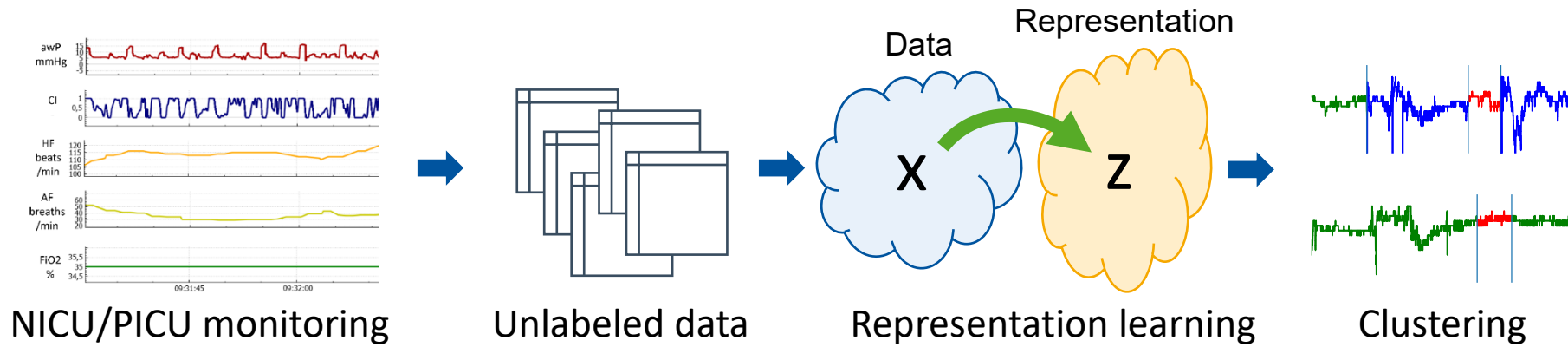


Unlabeled data

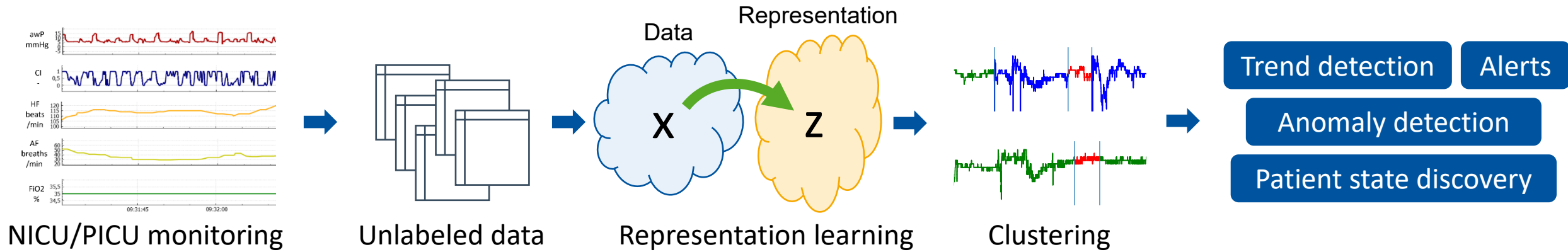
Motivation



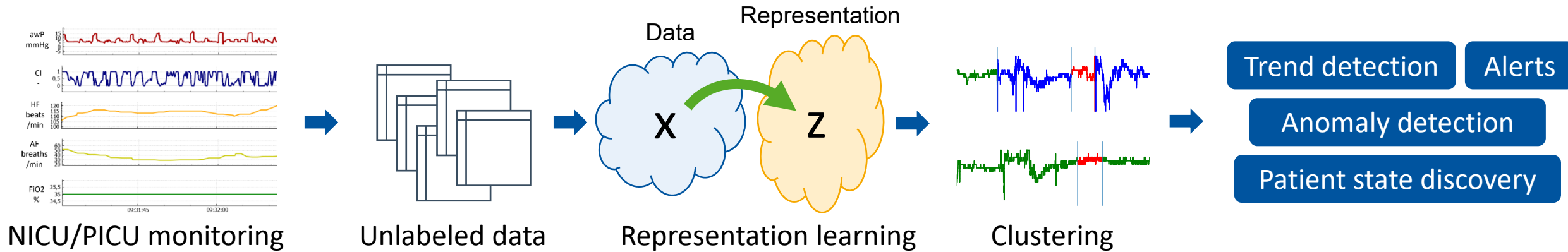
Motivation



Motivation



Motivation



- ▶ **Goal: Discover recurring patterns by learning compact and relevant representations**

- ▶ Data heterogeneity (Many different measurement sources and data types)
 - Different and asynchronous sampling intervals
 - Data with different physiological dynamics
 - Missing data
 - Patient-specific baselines, interventions and anomalies

- ▶ Data heterogeneity (Many different measurement sources and data types)
 - Different and asynchronous sampling intervals
 - Data with different physiological dynamics
 - Missing data
 - Patient-specific baselines, interventions and anomalies
- ▶ Computational requirements
 - High-frequency signals quickly lead to large input windows
 - Example: 10s = 1250 values of airway flow data

Evaluation methodology

Evaluation type	Examples	Limitations
Cluster geometry	Silhouette score, Davies-Bouldin score, Calinski-Harabasz score	- Assumes geometric (distance-metric based) separation - Does not evaluate clinical usefulness
Robustness	Comparison across patients/window sizes/sampling rate	Does not evaluate clinical usefulness
Cluster characteristics	Mean, variance, catch22, prototypes, histograms	- Time-consuming inspection - Computationally heavy - Does not evaluate clinical usefulness
Expert evaluation	Manual review	Expensive, time-consuming
Labeled tasks	Evaluation on labeled dataset	Requires labels

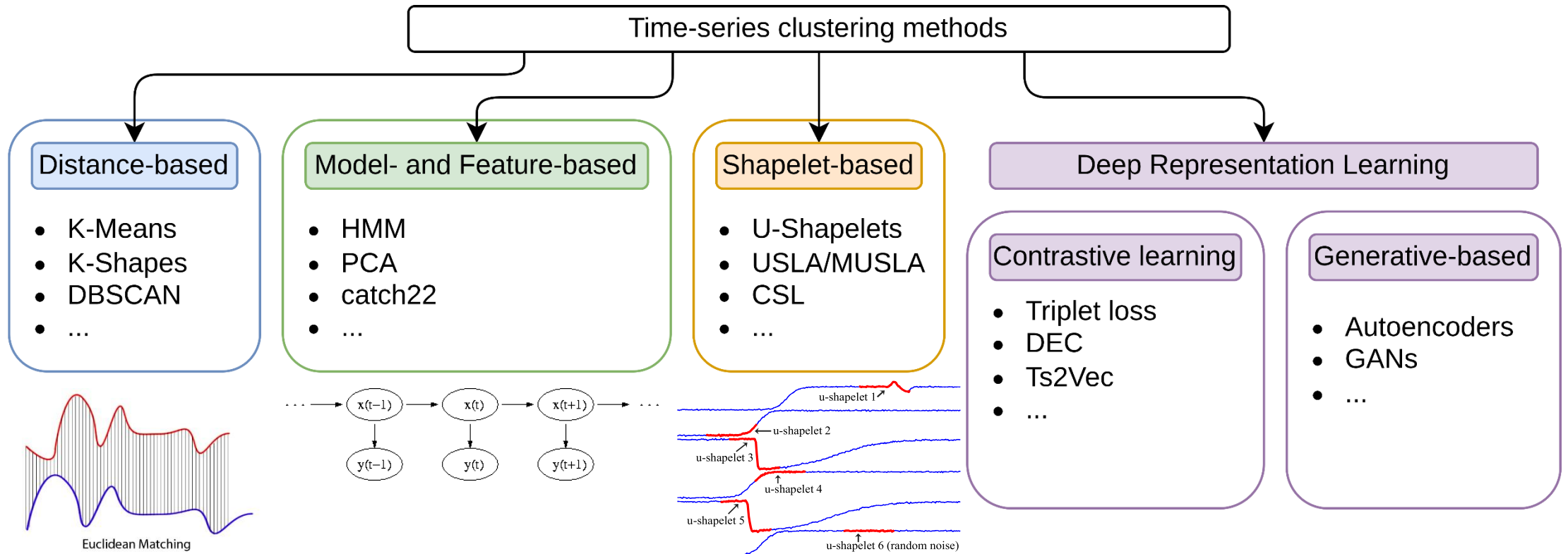


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26

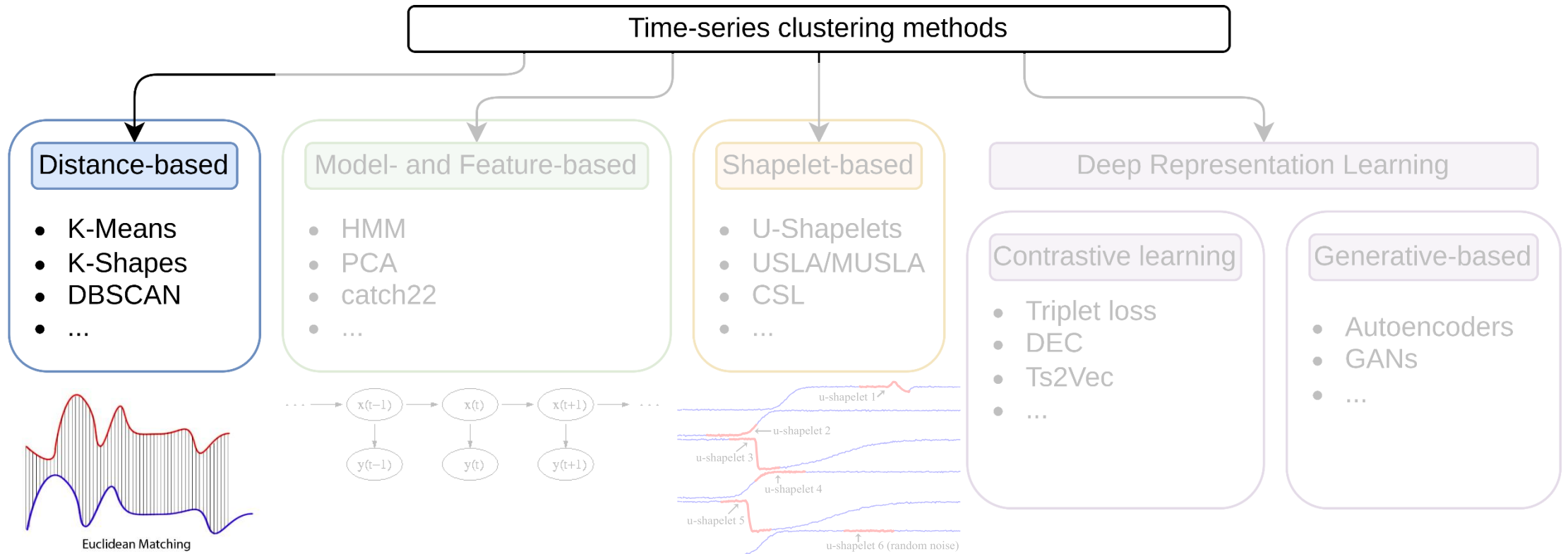


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26

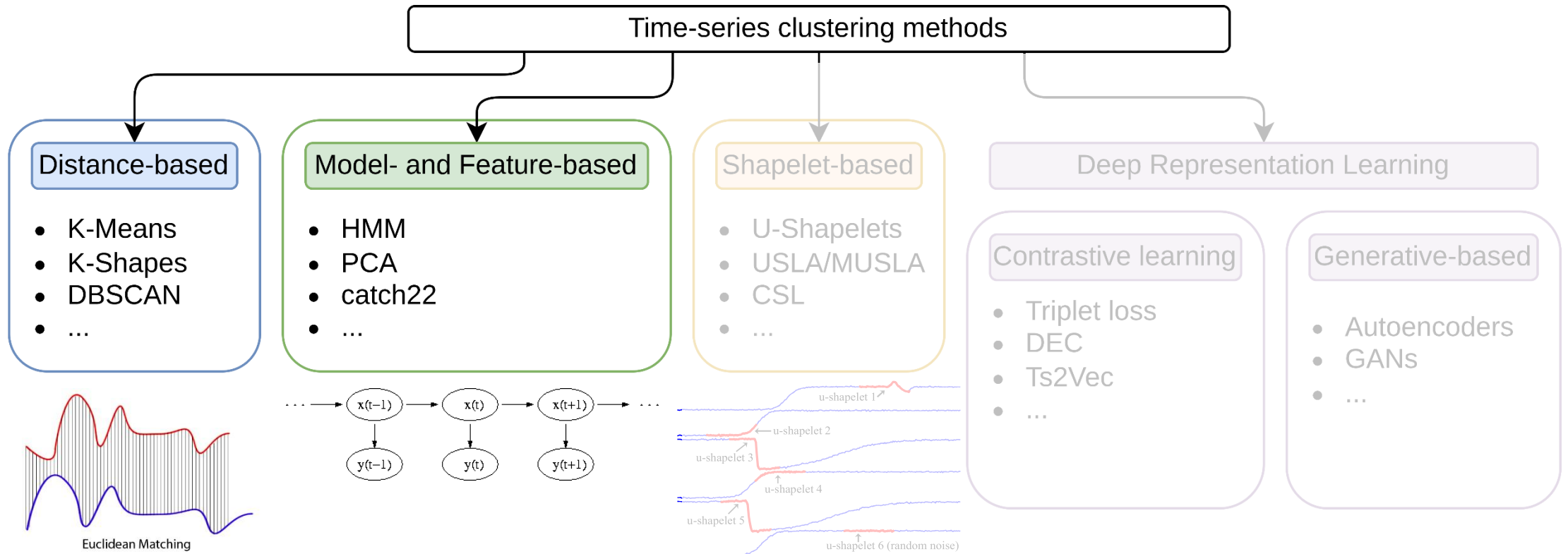


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26

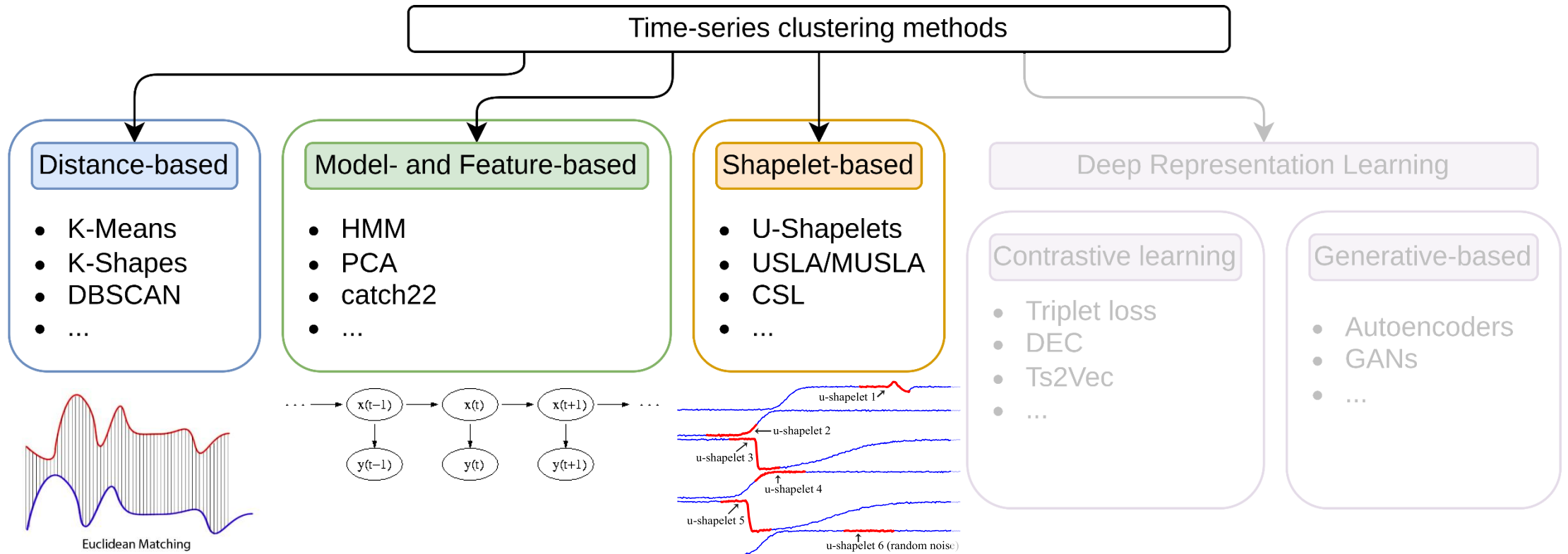


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26

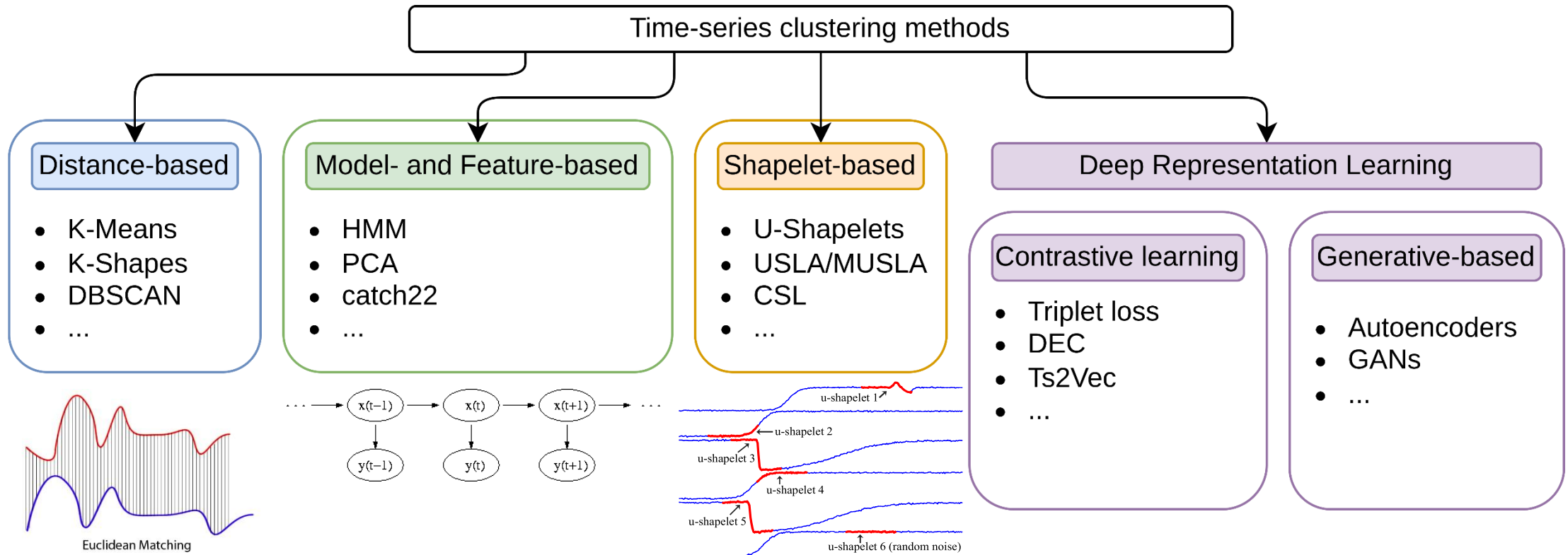
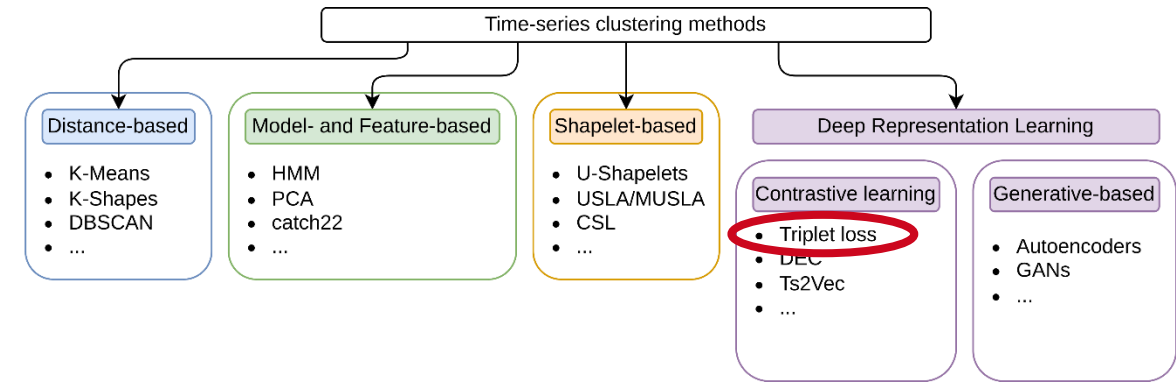


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26

Deep Representation Learning

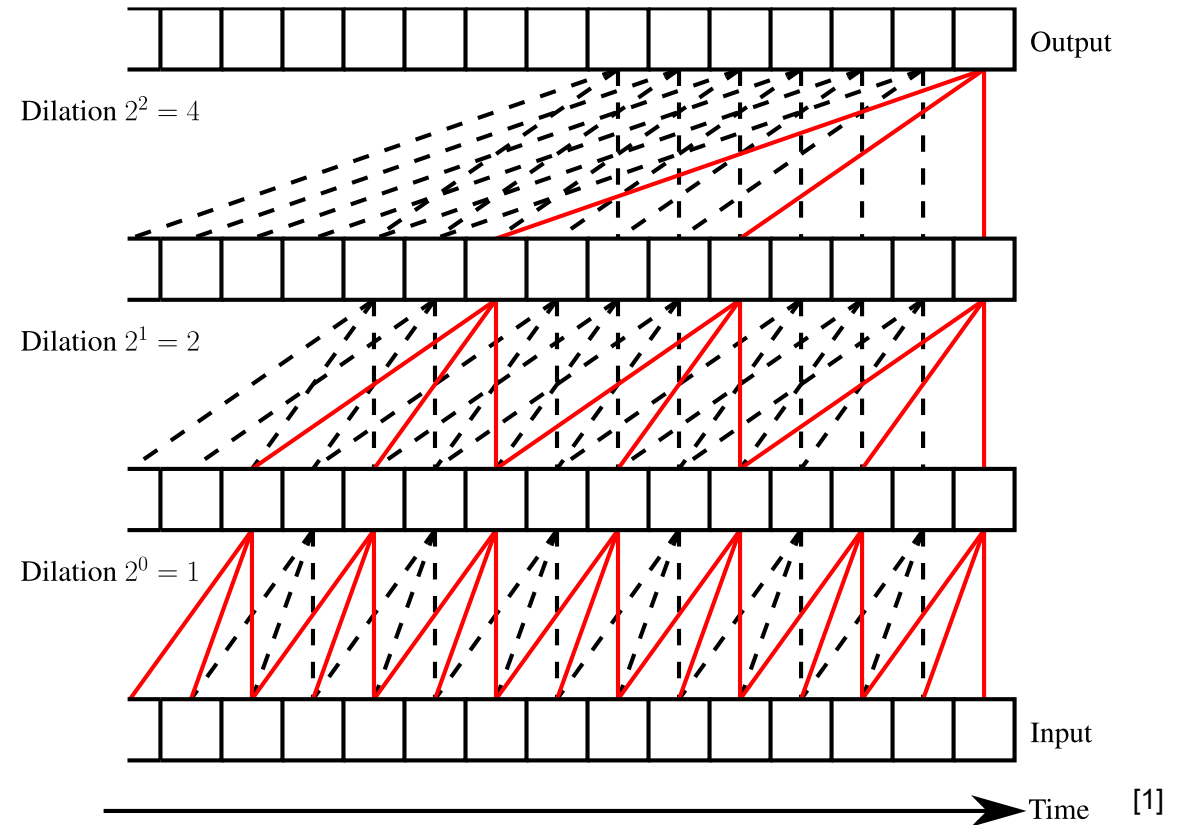
► USRL^[1]:



[1] Jean-Yves Franceschi et al. (2019). Unsupervised Scalable Representation Learning for Multivariate Time Series. *Thirty-third Conference on Neural Information Processing Systems*.

Deep Representation Learning

- ▶ USRL^[1]:
 - Convolutional encoder

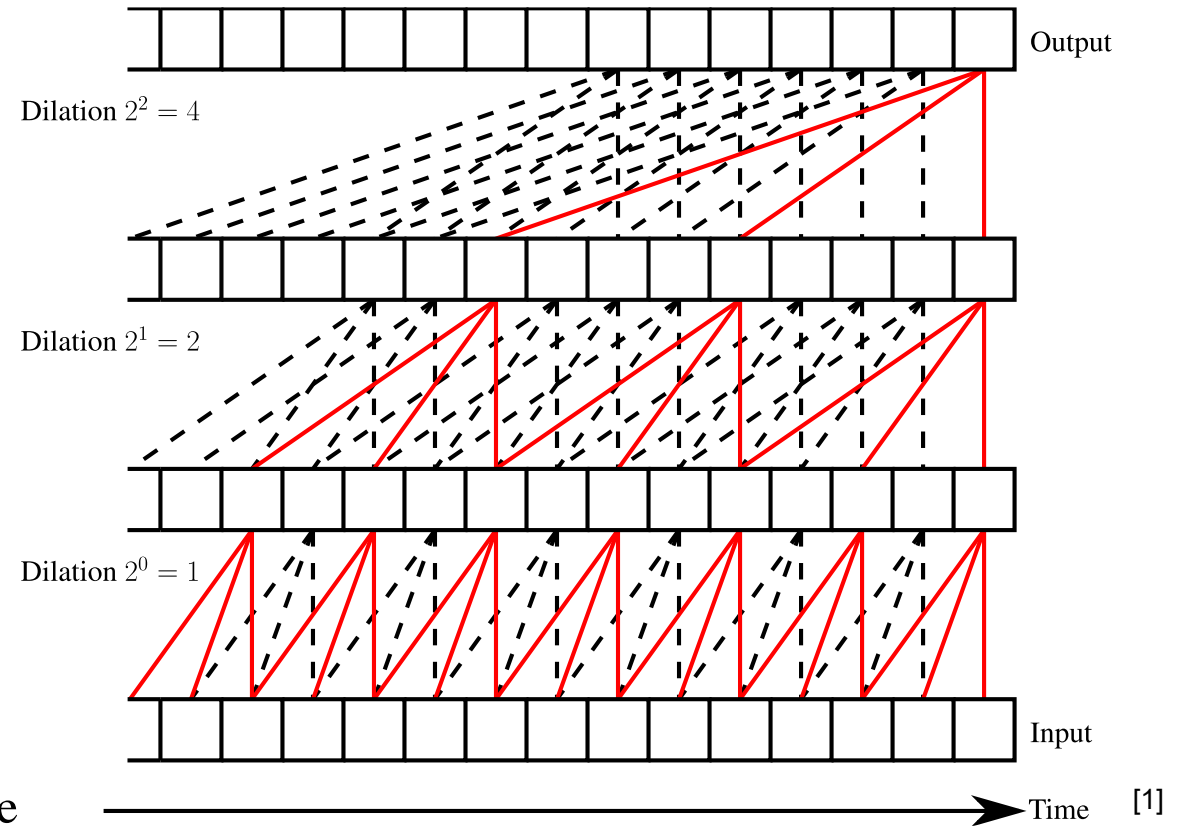
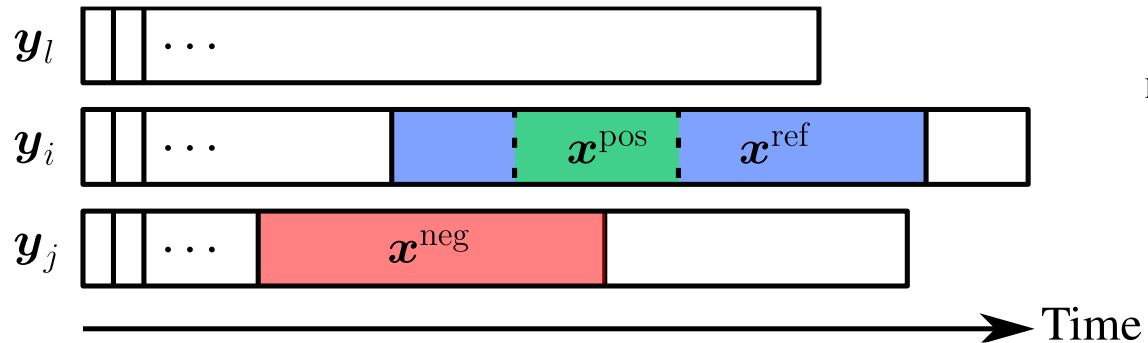


[1] Jean-Yves Franceschi et al. (2019). Unsupervised Scalable Representation Learning for Multivariate Time Series. *Thirty-third Conference on Neural Information Processing Systems*.

Deep Representation Learning

► USRL^[1]:

- Convolutional encoder
- Triplet loss

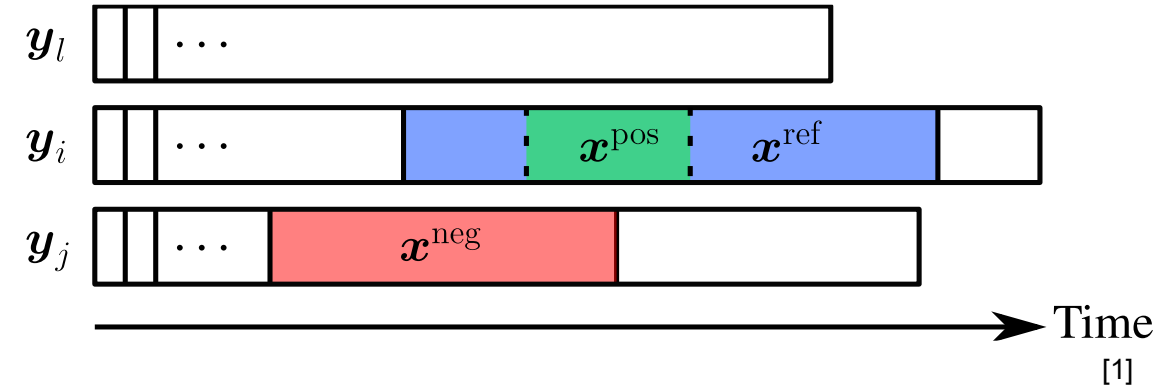


[1] Jean-Yves Franceschi et al. (2019). Unsupervised Scalable Representation Learning for Multivariate Time Series. *Thirty-third Conference on Neural Information Processing Systems*.

Deep Representation Learning

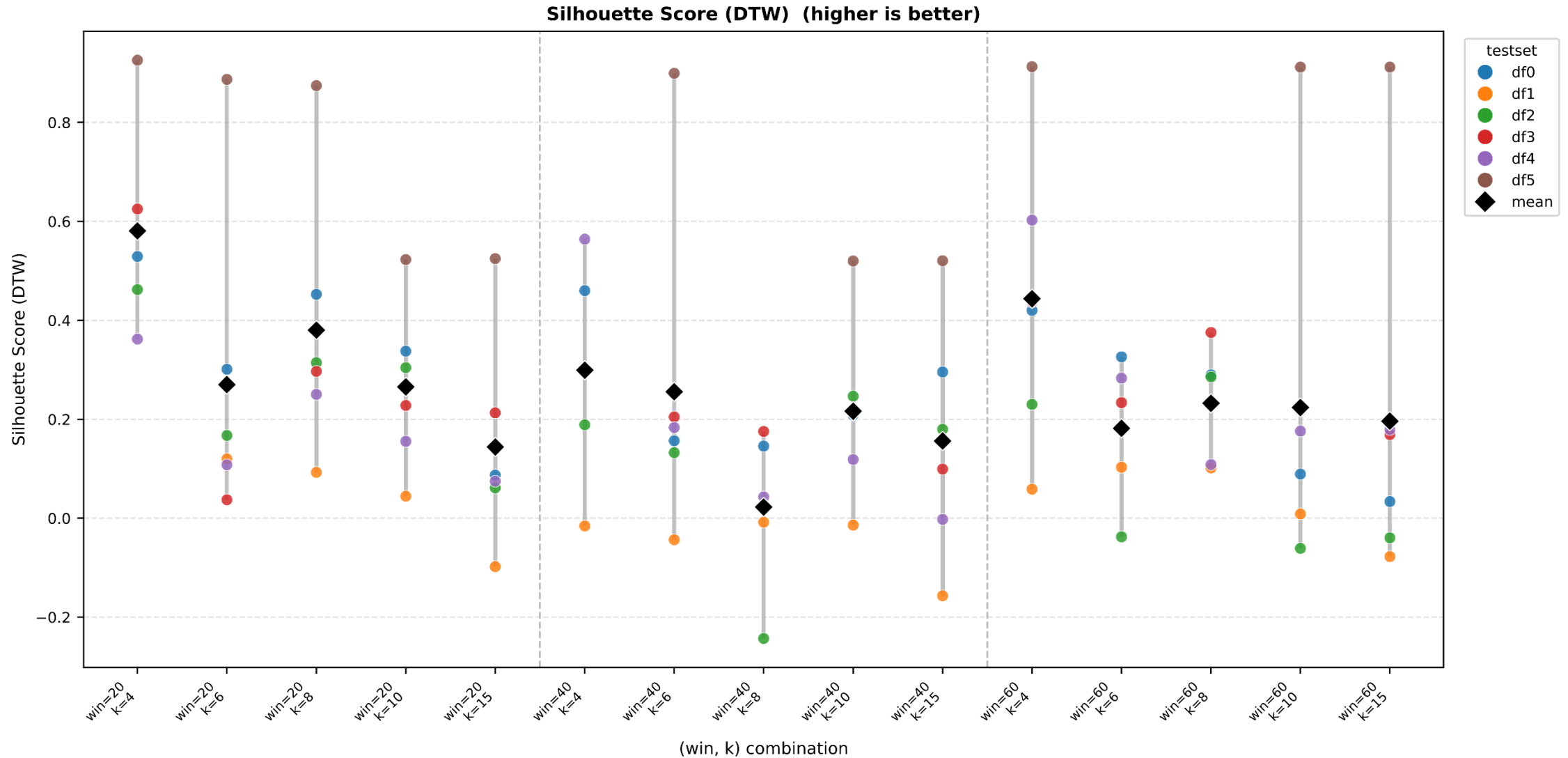
► USRL^[1]:

- Convolutional encoder
- Triplet loss



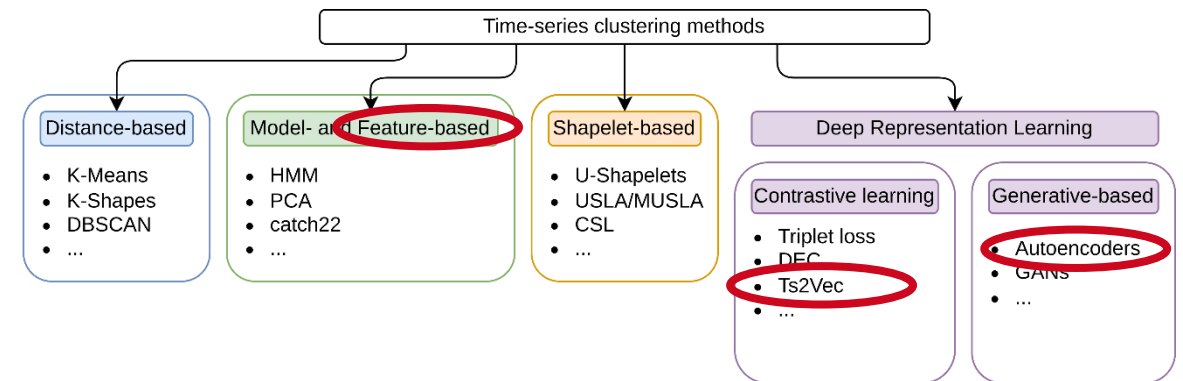
[1] Jean-Yves Franceschi et al. (2019). Unsupervised Scalable Representation Learning for Multivariate Time Series. *Thirty-third Conference on Neural Information Processing Systems*.

Deep Representation Learning

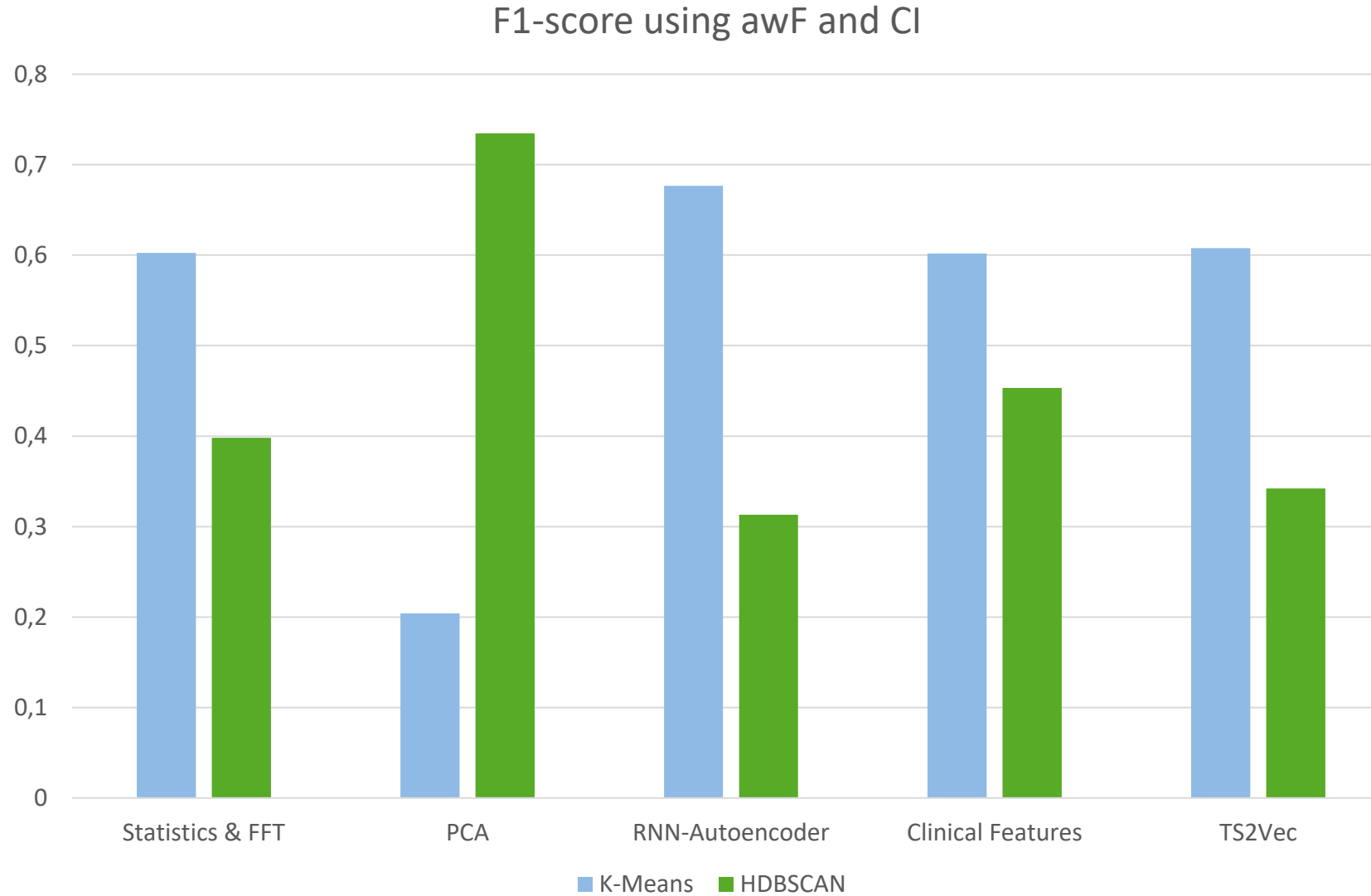


Labeled task evaluation

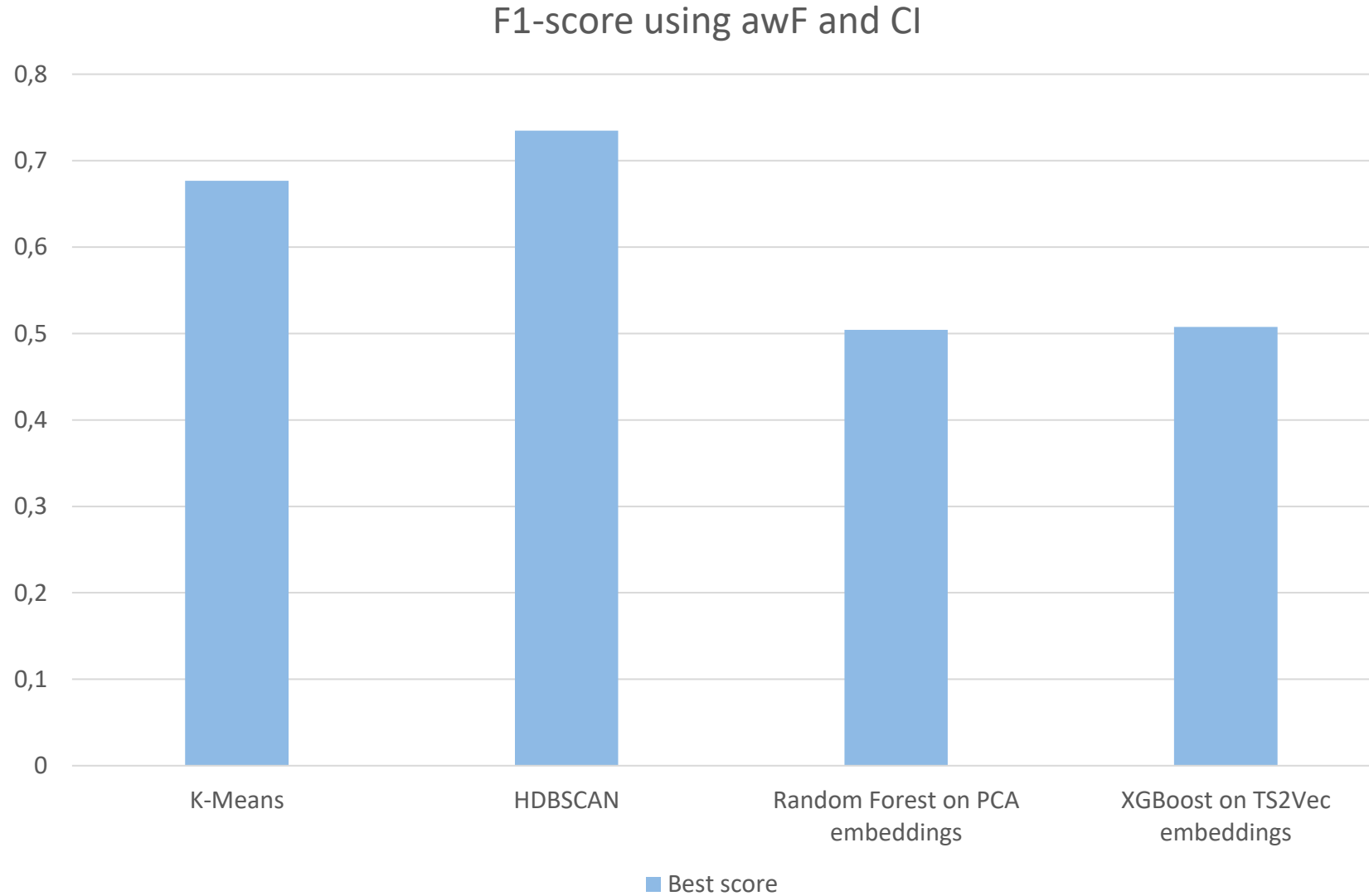
- ▶ Dataset with 11 patients (~2700 hours)
- ▶ Multivariate complication use-case
- ▶ 3% of labeled data (positive and negative)



Labeled task evaluation



Labeled task evaluation



Next steps

- ▶ Evaluate USRL and end-to-end supervised methods in labeled use-case
- ▶ Shapelet-based methods:
 - Option to pre-define shapelets
 - Better explainability
 - Candidate: CSL ^[2]

[2] Liang, Z. et al. (2024). Timecs: Unsupervised contrastive learning of general shapelets for explorable time series analysis. *Proceedings of the VLDB Endowment*.
10.14778/3685800.3685907

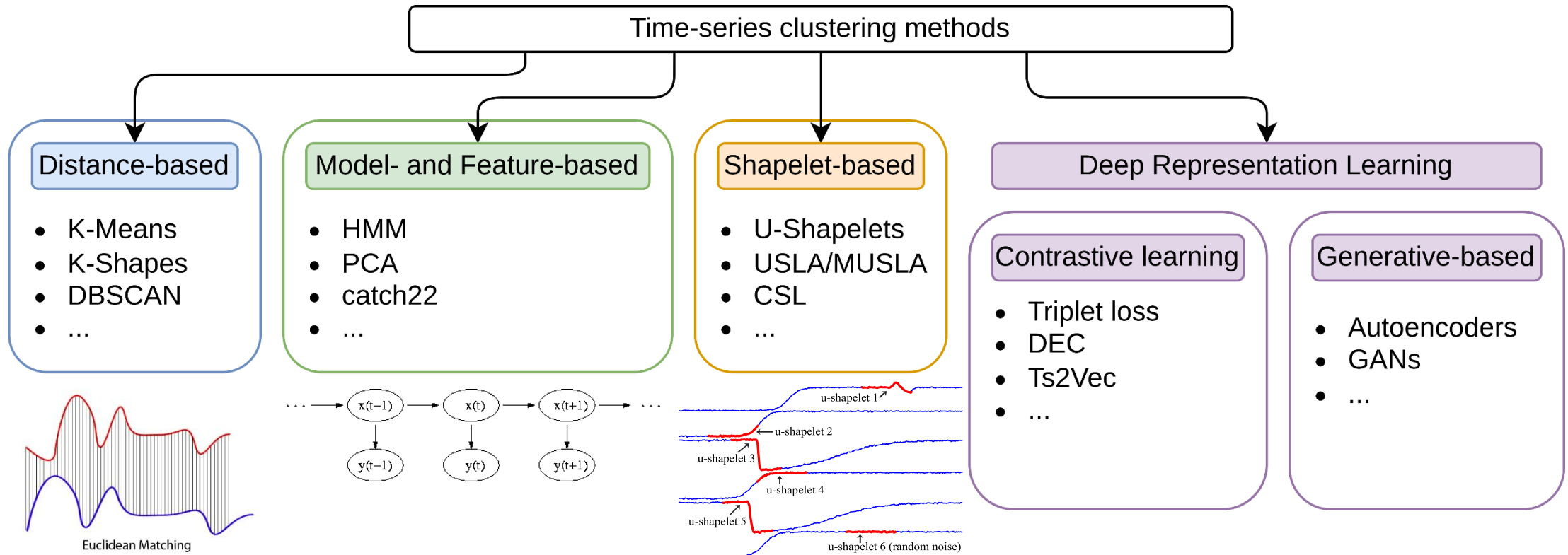


Diagram based on: Paparrizos, J., Yang, F., & Li, H. (2024). *Bridging the gap: A decade review of time-series clustering methods*. arXiv. DOI: 10.48550/ARXIV.2412.20582 and Paparrizos et al. (2025). Time-series clustering: A comprehensive study of data mining, machine learning, and deep learning methods. *Proceedings of the VLDB Endowment*, 18(11), 4380–4395. DOI: 10.14778/3749646.3749700

Image sources: Wikimedia commons, Zakaria, J. et al. (2012). Clustering time series using unsupervised-shapelets. *2012 IEEE 12th International Conference on Data Mining*, 785–794. DOI: 10.1109/ICDM.2012.26