Interpretable Weak Labelling of Time Series through Motif Discovery and Structured Temporal Representations

Christian Luna, José María Luna, and Sebastián Ventura

Department of Computing and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Córdoba, 14071 Córdoba, Spain
{c.luna,jmluna,sventura}@uco.es

PhD supervisors: José María Luna and Sebastián Ventura.
PhD first enrolment: November 2024.
Expected PhD defence: 2027/2028 academic year.

Abstract. Time series are increasingly collected in many scientific and industrial settings, but using them effectively still depends, in many cases, on having meaningful labels. Obtaining these labels manually is costly, requires expert knowledge, and becomes particularly challenging when the aim is to annotate specific points or temporal segments rather than a complete series. This doctoral research focuses on automatic and weak labelling of time series, with an emphasis on representations that remain interpretable. The main idea is that recurrent temporal patterns, or motifs, can serve as a bridge between raw signals and higher-level descriptions, yielding symbolic labels, rules, and structured representations useful for both human analysis and machine learning.
The thesis combines motif discovery, adaptive symbolic representation, grammatical inference, and reproducible experimentation. The work developed so far includes a motif-based labelling method based on variable-length motifs and evolutionary label composition, a grammar-based approach for error-bounded symbolic labelling, and an ongoing review and benchmark of time series motif discovery methods. Future work will focus on using the generated labels for interpretable learning, validating the framework in medical time series and predictive maintenance, and extending the approach towards online labelling.

Keywords: Time series · Weak supervision · Weak labelling · Motif discovery · Symbolic representations · Grammar inference · Interpretability

1 Background and Motivation

Time series are one of the most common data modalities in real-world artificial intelligence applications. They appear in medical monitoring, industrial systems, predictive maintenance, environmental sensing, finance, mobility, and energy management, among many other domains where observations are collected

continuously over time. In these scenarios, data are not only large and heterogeneous, but also sequential, noisy, and often difficult to interpret directly [5, 28]. This makes time series analysis a central problem for modern machine learning, especially when the final goal is not only to obtain accurate predictions, but also to understand the temporal behaviours that support them.

A major limitation in many time series applications is the lack of high-quality labels. Manual annotation is expensive, time-consuming, and usually requires domain experts [4, 6, 8]. This problem becomes even more difficult when labels are needed at point or segment level, since the expert must decide not only which behaviour is present, but also where it starts and ends [2, 18]. As a result, many real datasets remain weakly labelled, partially labelled, or completely unlabelled, limiting the development of reliable supervised models [29]. Although weak supervision and weak labelling have emerged as practical alternatives to reduce annotation effort [17, 29], many existing time-series approaches still rely on predefined segments, domain-specific heuristics, or labels assigned to complete sequences rather than to meaningful temporal regions [2, 8, 24].

A promising way to address this limitation is to exploit the series's internal structure. Time series often contain recurrent patterns, commonly known as motifs, which may correspond to states, events, transitions, or characteristic behaviours of the underlying system [21, 13]. Motifs provide a natural bridge between raw continuous signals and symbolic representations, since they reveal repeated subsequences that can make temporal behaviour easier to describe and inspect [12, 10]. However, using motifs for labelling is not straightforward. Many motif discovery methods require sensitive parameters such as window length, similarity thresholds, alphabet size, or the expected number of motifs [16, 26]. They often assume fixed-length patterns, exact or near-exact repetitions, and offline processing [7, 15]. Moreover, even when relevant motifs are discovered, transforming them into a coherent, compact, and interpretable sequence of labels remains a challenging and only partially addressed problem [11, 19].

This situation calls for methods that can derive structured temporal representations directly from raw time series, with limited dependence on manual annotation. In this thesis, the goal is to reduce the burden placed on domain experts without losing the temporal information that makes the data meaningful. The labels generated in this way are therefore not conceived merely as a preprocessing artefact. Instead, they are viewed as an intermediate representation between the signal and the learning task, one that can be examined by humans and exploited by machine learning models. This intermediate level is particularly useful when annotations are scarce, costly, or unavailable at the point or segment level.

Accordingly, this doctoral research studies automatic and weak labelling of time series from a structural and interpretable perspective. Its central hypothesis is that recurrent temporal patterns can be transformed into symbolic labels, rules, and higher-level representations that help describe the behaviour of a time series while remaining useful for downstream learning. This perspective connects motif discovery, weak labelling, symbolic representation, grammatical inference,

explainable artificial intelligence, and time series mining, with the broader aim of making temporal data more interpretable and more usable in data-driven systems.

2 Objectives

The main objective of this thesis is to develop automatic weak-labelling methods for time series based on motif discovery, symbolic representation, and grammatical inference, and to exploit the generated labels as an intermediate, interpretable layer for downstream machine learning models. The thesis is therefore not limited to producing annotations, but aims to define a path from raw temporal signals to symbolic labels, structured temporal representations, and interpretable learning models. This general objective is divided into the following specific objectives:

O1. To systematize time series motif discovery. This objective aims to analyse the evolution of time series motif discovery methods, identifying their assumptions, methodological families, limitations, and practical trade-offs. It includes the study of exact and approximate algorithms, fixed- and variable-length motifs, raw and symbolic representations, multivariate settings, streaming approaches, and encoding-based methods [9, 21]. The expected outcomes are a structured taxonomy, a unified implementation framework, and a reproducible benchmark that provide the methodological basis for positioning the proposed labelling methods within the state of the art.

O2. To develop symbolic labelling methods based on motifs and grammatical structures. This objective develops automatic labelling methods that turn raw time series into symbolic sequences by identifying recurrent temporal patterns and organising them into meaningful temporal structures. In the motif-based line, variable-length motifs will be combined to build coherent label maps over the series. In the grammar-based line, repeated symbolic fragments will be used to infer reusable rules able to describe regular behaviours and temporal dependencies [11, 19]. The aim is not only to generate shorter symbolic descriptions of the data, but to obtain representations that preserve the relevant temporal structure and remain interpretable after the labelling process. For this reason, the methods will be designed to depend as little as possible on manual annotation, fixed window sizes, exact repetitions, or predefined segment boundaries.

O3. To exploit motif-based labels for interpretable learning. This objective examines how the labels, motifs, and grammatical rules generated by the proposed methods can serve as input to downstream learning models. In particular, the thesis will explore classifiers and predictive models built from symbolic label sequences, motif occurrences, and inferred rules [25, 27]. The aim is to assess whether these representations preserve relevant predictive information while making the decision process easier to inspect and understand, and to analyse the trade-off between predictive performance, model simplicity, and interpretability.

O4. To validate the proposed framework across realistic domains and tasks. This objective analyses the behaviour of the proposed methods in settings with different levels of control and complexity. Synthetic datasets will be used to test specific properties when the relevant temporal structures are known in advance, including reconstruction quality, compression, robustness, and temporal coverage. The evaluation will then be extended to real-world time series, where the methods must deal with noisier data, greater variability, incomplete annotations, and behaviours that vary across application domains. The analysis will also consider computational cost, interpretability, and the usefulness of the generated labels in downstream tasks such as classification, anomaly detection, or semantic segmentation. Particular attention will be given to domains where expert annotation is expensive or limited, including medical time series, industrial monitoring, predictive maintenance, and other sensor-based scenarios [6, 8].

O5. To extend the framework towards online and adaptive scenarios. This objective analyses the possible extension of the proposed methods to data streams and online learning settings. In many real applications, time series arrive continuously, and the underlying data distribution may evolve due to changing operating conditions, evolving temporal dynamics, or individual variability [23]. Therefore, the thesis will study how motifs, labels, and symbolic rules could be updated incrementally as new observations arrive, without recomputing the complete representation from scratch [15]. The aim is to explore the feasibility of real-time or near-real-time labelling while maintaining interpretability and computational feasibility.

3 Methodology and Work Plan

The methodology of this thesis combines theoretical analysis, algorithmic design, reproducible experimentation, and application-oriented validation. The work is organised as a progressive research programme in which each stage supports the following one: first, the field of time series motif discovery is analysed and systematised; then, automatic symbolic labelling methods are designed; afterwards, the generated labels and rules are exploited for interpretable learning; finally, the framework is validated in realistic domains and explored towards online and adaptive scenarios.

The thesis follows an iterative workflow rather than a single linear process. Each stage begins with a review of the relevant literature in order to identify the assumptions, limitations, and open problems that motivate the work. From there, a methodological hypothesis is defined and developed into an algorithmic proposal. The proposed method is then implemented within a common software framework, enabling experiments to be reproduced and compared under consistent conditions. Its behaviour is evaluated through both quantitative and qualitative criteria, including reconstruction fidelity, symbolic compactness, temporal coverage, computational cost, and usefulness in downstream tasks. The results of this evaluation are then used to refine the method, analyse its limitations, and compare it with representative baselines.

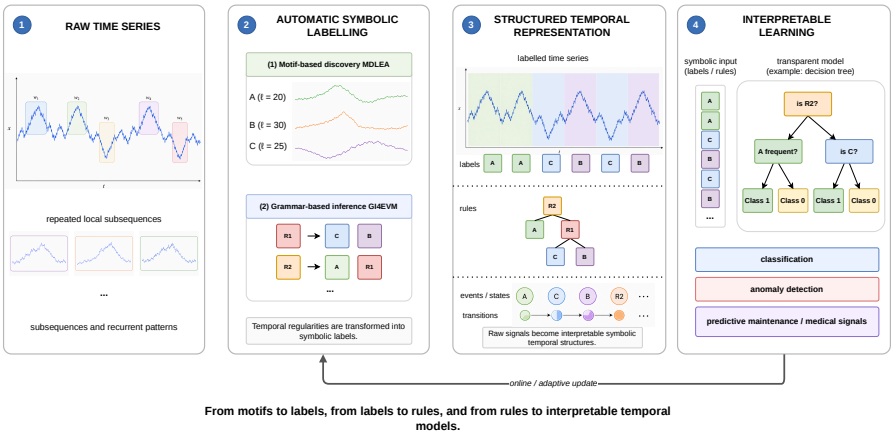


Fig. 1. Overview of the proposed doctoral framework, from raw time series to symbolic labelling, structured temporal representations, and interpretable downstream learning.

Figure 1 summarises the overall methodological workflow of the thesis, linking motif-based and grammar-based labelling with structured temporal representations and downstream interpretable learning.

The work plan is structured according to the objectives defined in Section 2. For **O1**, the thesis provides a theoretical and experimental review of time-series motif discovery. This includes the analysis of exact and approximate methods, similarity-, frequency-, and encoding-based approaches, raw and symbolic representations, fixed- and variable-length motifs, multivariate extensions, and online settings [9, 14, 21]. In parallel, representative algorithms are being implemented or integrated under a common Python interface, allowing their behaviour to be compared under homogeneous experimental conditions [22]. The expected result is a structured taxonomy, a unified implementation framework, and a reproducible benchmark that provide the methodological basis for the rest of the thesis.

For **O2**, the thesis designs automatic labelling methods that transform raw time series into symbolic label sequences. Initial work has resulted in two complementary methodological lines. The first one, MDLEA, combines error-bounded variable-length motif discovery with an evolutionary procedure for label composition. Its goal is to build coherent label maps that balance reconstruction fidelity, temporal coverage, and reduction of unlabelled gaps. The second one, GI4EVM, combines adaptive error-bounded discretisation with noise-tolerant grammatical inference to obtain compact symbolic sequences and reusable rules while reducing sensitivity to fixed windows and alphabet choices, as well as to exact pattern matching. The remaining work in this objective focuses on refining these methods, improving motif composition mechanisms, consolidating grammar-based labelling, and comparing the proposals against representative discretisation, motif discovery, and grammar-based baselines [12, 19].

For **O3**, the generated labels, motif occurrences, and grammatical rules will be reused in downstream learning tasks. The aim is to move beyond labelling as a final output and treat symbolic representations as an intermediate layer between raw time series and predictive models. Existing classifiers and rule-based models will be adapted to use label sequences, motif-based features, and rule-based representations, and then compared with models trained directly on raw or conventional time-series representations [1, 27]. The evaluation will consider predictive performance, model complexity, qualitative interpretability, and the extent to which model decisions can be traced back to discovered temporal behaviours.

For **O4**, the proposed framework will be evaluated in both controlled and application-oriented scenarios. Synthetic datasets will be used to analyse properties such as noise tolerance, temporal coverage, compression, and reconstruction fidelity. Real-world datasets will be used to assess practical usefulness in domains where expert annotation is costly or limited. Particular attention will be paid to medical time series and predictive maintenance, while additional sensor-based domains may be considered when they provide a suitable evaluation setting [3, 8]. The validation will also assess the usefulness of the generated labels for selected downstream tasks, primarily classification and, where appropriate, exploratory analyses in anomaly detection and semantic segmentation.

For **O5**, the thesis will study possible extensions towards online and adaptive scenarios. This stage addresses the limitations of offline labelling in settings where data arrive continuously or where temporal behaviour changes over time. The work will analyse online and streaming motif discovery strategies [15], mechanisms to update labels and rules without recomputing the full representation from scratch, and the computational feasibility of real-time or near-real-time monitoring scenarios.

At this stage, the thesis has progressed beyond the initial definition of the problem. The main methodological directions are already in place, two algorithmic proposals have been developed, and the broader review and experimental framework are being completed. The work that remains is mainly concerned with strengthening these contributions: extending the empirical validation, analysing how the generated labels can be used in interpretable learning models, developing selected application-oriented case studies, and studying possible online extensions.

The main milestones planned for the remainder of the thesis are:

- **Review, taxonomy, and benchmark (mid-2026):** completion and refinement of the comprehensive review, taxonomy, and benchmark of time-series motif discovery methods.
- **Grammar-based labelling framework (mid-2026):** further refinement of the proposed motif- and grammar-based labelling methods, including extended experimental validation against representative baselines.
- **Predoctoral research stay (September–November 2026):** research stay at INESC TEC (Porto, Portugal) to discuss and extend the proposed framework towards online and adaptive learning scenarios.

– **Interpretable downstream learning (late-2026 – early-2027):** investigation of how the generated labels, motifs, and grammatical rules can support interpretable predictive models.
– **Application-oriented validation (2027):** validation of the proposed framework in selected application domains, including medical time series, predictive maintenance, and other sensor-based scenarios.
– **Doctoral dissertation (late-2027):** integration of the methodological, experimental, and application-oriented contributions into the final doctoral thesis.
– **Thesis defence (2027/2028 academic year):** expected completion and public defence of the doctoral dissertation.

This research stay is expected to support the final stages of the thesis by providing an international environment for discussion, external feedback, and potential collaboration, especially in relation to online and adaptive extensions of the proposed framework.

4 Preliminary Results

Several preliminary results have already been obtained during the development of this thesis, including methodological contributions, publications, benchmarking activities, software infrastructure, and application-oriented extensions. Together, they show the progressive consolidation of the research line, from motif discovery to weak labelling and interpretable temporal representations.

– **Motif-based weak labelling.** The first methodological contribution is a motif-based weak-labelling approach that combines error-bounded variable-length motif discovery with evolutionary label composition to build coherent label maps over time series. This work has already led to a peer-reviewed publication:
  • C. Luna, J. M. Luna, and S. Ventura. *Motif Discovery and Labelling via Evolutionary Algorithm in Time Series*. In SADASC 2026: Sixth International Conference on Smart Applications and Data Analysis for Smart Cyber-Physical Systems, Marrakech, Morocco, 2026. Best Paper Award.
– **Grammar-based symbolic labelling.** The second contribution extends this line by combining adaptive error-bounded symbolic representation with grammatical inference, aiming to obtain compact symbolic sequences and reusable rules with reduced dependence on fixed windows and exact repetitions. This work is currently under journal review:
  • C. Luna, J. M. Luna, and S. Ventura. *Automatic Time Series Labelling via Grammatical Inference with Error-Bounded Variable-Length Motifs*. Submitted to IEEE Transactions on Cybernetics; under review after second-round revision.

– **Review, taxonomy, and benchmark.** In parallel, the thesis is developing a comprehensive review and benchmark of time-series motif discovery methods, including a structured taxonomy, representative algorithms, and a unified experimental protocol. This work provides the basis for positioning the proposed labelling methods within the state of the art and selecting representative baselines:
  - C. Luna, J. M. Luna, and S. Ventura. *Discovering Time Series Motifs: A Comprehensive Review and Benchmark of 25 Years of Algorithms.* Manuscript in preparation.
– **Application-oriented validation.** The proposed methods are also being explored in realistic domains where temporal annotations are costly or incomplete, including medical time series, predictive maintenance, and other sensor-based scenarios. The aim is to assess whether the generated labels can support signal inspection, weak annotation, and interpretable downstream learning.
– **Reproducible software framework.** A Python framework for time-series motif discovery is being developed to integrate representative algorithms under a common interface. Rather than an isolated software output, it is conceived as a research instrument for reproducible experimentation, systematic comparison, and empirical validation of the proposed methods.

Overall, these results show that the thesis already has concrete methodological contributions and a reproducible basis. The remaining work will strengthen the empirical validation and extend the framework towards interpretable learning, application-oriented studies, and online scenarios.

5 Relevance of the Proposed Research

The relevance of this thesis lies in addressing a central limitation of many time-series learning scenarios: the difficulty of obtaining meaningful labels at the temporal scale at which relevant behaviours occur. In many real applications, raw time series are abundant, but expert annotations are scarce, costly, or unavailable [4, 8]. This is particularly problematic when the objective is not only to classify complete sequences, but to identify the temporal regions, events, states, or transitions that explain the behaviour of the system [2, 10]. This thesis uses automatic labelling methods based on motifs, symbolic representations, and grammatical inference to transform raw time series into intermediate representations that are easier to interpret and more useful for learning tasks.

From a scientific perspective, the thesis contributes to time-series mining, weakly supervised learning, motif discovery, and explainable artificial intelligence. Motif discovery has traditionally been studied primarily as a pattern-mining task, where the objective is to find repeated subsequences [13, 21]. This thesis adopts a different perspective by treating motifs as building blocks for automatic labelling and interpretable temporal representation. In this sense, the

contribution is not limited to detecting repeated patterns but aims to transform motifs into labels, labels into rules, and rules into representations that can support later learning tasks.

A second contribution is methodological. Many existing symbolic and motif-based approaches still depend, to some extent, on design choices such as fixed window sizes, discretisation parameters, or assumptions about how repetitions should appear in the data [7, 19]. These choices can be difficult to set in real applications, where the same underlying behaviour may reappear with changes in amplitude, duration, alignment, or local shape [21]. The proposed approaches address this limitation through variable-length motif discovery, error-bounded grouping, adaptive symbolic discretisation, and noise-tolerant grammatical inference. At the same time, the thesis contributes to reproducibility through a taxonomy, a unified implementation framework, and a benchmark of time-series motif discovery methods, helping clarify the assumptions, trade-offs, and practical behaviour of existing algorithms [9, 22].

From an applied perspective, the thesis is especially relevant in domains where time series are routinely collected, but detailed annotations are difficult to obtain. In medical time series, automatically generated symbolic labels could help describe recurrent physiological behaviours and point experts towards abnormal or ambiguous regions that require closer inspection [8]. In predictive maintenance and industrial monitoring, motif-based labels may help identify operational states, repeated machine behaviours, degradation patterns, or unusual transitions [20]. The same problem also occurs in other sensor-based domains, such as environmental monitoring, energy systems, and finance, where long, continuous signals need to be summarised into representations that can be interpreted and used for decision-making.

Finally, the proposed framework may help develop more interpretable downstream models. Instead of training predictive models directly on raw time series, the thesis explores the use of symbolic label sequences, motif occurrences, and grammatical rules as intermediate features or decision units [25, 27]. This can make model behaviour easier to inspect, since predictions can be related to discovered temporal behaviours rather than only to opaque numerical representations. The possible extension towards online and adaptive scenarios further increases the practical relevance of the work, especially in monitoring and decision-support applications where data arrive continuously, and temporal behaviours may evolve over time [15].

6 Conclusions

This doctoral research addresses automatic time-series labelling from a representation-centred, interpretability-oriented perspective. Its main premise is that labelling should not be understood only as the generation of annotations, but also as a way of transforming raw temporal signals into structured, reusable, and interpretable representations.

The work completed so far has established the basis of the thesis through two complementary methodological directions: motif-based labelling with evolutionary composition, and grammar-based labelling with adaptive symbolic representation. These contributions are supported by a broader theoretical and experimental effort aimed at systematising time-series motif discovery through a review, taxonomy, a unified implementation framework, and a benchmark.

The next stages of the thesis will build on the results obtained so far. The work will focus on strengthening the proposed methods, extending their empirical evaluation, and studying how the generated labels, motifs, and grammatical rules can be used in interpretable learning tasks. The thesis will also examine how these ideas could be adapted to online and adaptive settings, where new observations arrive over time, and the symbolic representation may need to be updated progressively. According to the current work plan, the thesis is expected to be completed during the 2027/2028 academic year.

In this way, the thesis aims to provide not only new methods but also software resources and evaluation tools for interpretable time-series analysis. Its main contribution lies in bringing together motif discovery, symbolic labelling, grammatical inference, and downstream learning within a common framework. This connection is intended to support representations of temporal data that are easier to inspect, easier to reuse, and more useful for building understandable machine learning models.

Acknowledgements

This research was supported in part by the PID2023-148396NB-I00 project of Spanish Ministry of Science and Innovation and the European Regional Development Fund, by the ProyExcel-0069 project of the Andalusian University, Research and Innovation Department, and by a predoctoral grant FPU23/02732 from the Spanish Ministry of Universities.

References

1. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data mining and knowledge discovery **31**(3), 606–660 (2017). https://doi.org/10.1007/s10618-016-0483-9
2. Blalock, D.W., Guttag, J.V.: Extract: Strong examples from weakly-labeled sensor data. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). pp. 799–804. IEEE (2016). https://doi.org/10.1109/icdm.2016.0093, backup Publisher: IEEE
3. Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A., Keogh, E.: The UCR Time Series Classification Archive. IEEE/CAA Journal of Automatica Sinica **6**(6), 1293–1305 (2019). https://doi.org/10.1109/jas.2019.1911747
4. Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X.: Label-efficient time series representation learning: A review. IEEE Transactions on Artificial Intelligence **5**(12), 6027–6042 (2024). https://doi.org/10.1109/tai.2024.3430236

5. Esling, P., Agon, C.: Time-series data mining. ACM Computing Surveys (CSUR) **45**(1), 1–34 (2012). https://doi.org/10.1145/2379776.2379788
6. Fredriksson, T., Mattos, D.I., Bosch, J., Olsson, H.H.: Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In: International conference on product-focused software process improvement. pp. 202–216. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-64148-1_13, backup Publisher: Springer
7. Gao, Y., Lin, J.: HIME: discovering variable-length motifs in large-scale time series. Knowledge and Information Systems **61**(1), 513–542 (Oct 2019). https://doi.org/10.1007/s10115-018-1279-6, http://link.springer.com/10.1007/s10115-018-1279-6
8. Goswami, M., Boecking, B., Dubrawski, A.: Weak supervision for affordable modeling of electrocardiogram data. In: AMIA Annual Symposium Proceedings. vol. 2021, p. 536. American Medical Informatics Association (2021)
9. Guerrini, V., Germain, T., Truong, C., Oudre, L., Boniol, P.: Time Series Motif Discovery: A Comprehensive Evaluation. Proceedings of the VLDB Endowment (PVLDB) **18**(7), 2226–2239 (2025). https://doi.org/10.14778/3734839.3734857
10. Jain, S., Hallac, D., Sosic, R., Leskovec, J.: MASA: Motif-Aware State Assignment in Noisy Time Series Data (Jun 2019). https://doi.org/10.48550/arXiv.1809.01819, http://arxiv.org/abs/1809.01819, arXiv:1809.01819 [cs]
11. Li, Y., Lin, J.: Approximate variable-length time series motif discovery using grammar inference. In: Proceedings of the Tenth International Workshop on Multimedia Data Mining. pp. 1–9. ACM, Washington D.C. (Jul 2010). https://doi.org/10.1145/1814245.1814255, https://dl.acm.org/doi/10.1145/1814245.1814255
12. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. Data Mining and knowledge discovery **15**(2), 107–144 (2007). https://doi.org/10.1007/s10618-007-0064-z
13. Lonardi, J., Patel, P.: Finding motifs in time series. In: Proc. of the 2nd workshop on temporal data mining. pp. 53–68. -, - (2002)
14. Mueen, A.: Time series motif discovery: dimensions and applications. WIREs Data Mining and Knowledge Discovery **4**(2), 152–159 (Mar 2014). https://doi.org/10.1002/widm.1119, https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1119
15. Mueen, A., Keogh, E.: Online discovery and maintenance of time series motifs. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1089–1098. ACM, Washington DC USA (Jul 2010). https://doi.org/10.1145/1835804.1835941, https://dl.acm.org/doi/10.1145/1835804.1835941
16. Nunthanid, P., Niennattrakul, V., Ratanamahatana, C.A.: Parameter-free motif discovery for time series data. In: 2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. pp. 1–4. IEEE, Phetchaburi, Thailand (May 2012). https://doi.org/10.1109/ECTICon.2012.6254126, http://ieeexplore.ieee.org/document/6254126/, backup Publisher: IEEE
17. Ratner, A., Bach, S.H., Ehrenberg, H., Fries, J., Wu, S., Ré, C.: Snorkel: rapid training data creation with weak supervision. The VLDB Journal **29**(2), 709–730 (2020). https://doi.org/10.1007/s00778-019-00552-1
18. Schaekermann, M., Law, E., Larson, K., Lim, A.: Expert Disagreement in Sequential Labeling: A Case Study on Adjudication in Medical Time Series Analysis. In: SAD/CrowdBias@ HCOMP. pp. 55–66 (2018)

19. Senin, P., Lin, J., Wang, X., Oates, T., Gandhi, S., Boedihardjo, A.P., Chen, C., Frankenstein, S.: GrammarViz 3.0: Interactive Discovery of Variable-Length Time Series Patterns. ACM Transactions on Knowledge Discovery from Data **12**(1), 1–28 (Feb 2018). https://doi.org/10.1145/3051126, https://dl.acm.org/doi/10.1145/3051126
20. Torkamani, S., Lohweg, V.: Identification of Multi-Scale Motifs. In: ProcEEDings 24. Workshop comPutational intElligEncE. p. 277 (2014)
21. Torkamani, S., Lohweg, V.: Survey on time series motif discovery. WIREs Data Mining and Knowledge Discovery **7**(2), e1199 (Mar 2017). https://doi.org/10.1002/widm.1199, https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1199
22. Van Wesenbeeck, D., Yurtman, A., Meert, W., Blockeel, H.: Quantitative evaluation of motif sets in time series. Data Mining and Knowledge Discovery **40**(1), 5 (2026). https://doi.org/10.1007/s10618-025-01169-5
23. Wang, K., Lu, J., Liu, A., Zhang, G.: TS-DM: A time segmentation-based data stream learning method for concept drift adaptation. IEEE Transactions on Cybernetics **54**(10), 6000–6011 (2024). https://doi.org/10.1109/tcyb.2024.3429459
24. Wilson, G., Doppa, J.R., Cook, D.J.: Multi-source deep domain adaptation with weak supervision for time-series sensor data. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1768–1778. ACM (2020). https://doi.org/10.1145/3394486.3403228
25. Yan, R., Ma, T., Fokoue, A., Chang, M., Julius, A.: Neuro-symbolic models for interpretable time series classification using temporal logic description. In: 2022 IEEE International Conference on Data Mining (ICDM). pp. 618–627. IEEE (2022). https://doi.org/10.1109/icdm54844.2022.00072, backup Publisher: IEEE
26. Yeh, C.C.M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H.A., Silva, D.F., Mueen, A., Keogh, E.: Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). pp. 1317–1322. IEEE, Barcelona, Spain (Dec 2016). https://doi.org/10.1109/ICDM.2016.0179, http://ieeexplore.ieee.org/document/7837992/, backup Publisher: Ieee
27. Yin, M.S., Tangsripairoj, S., Pupacdi, B.: Variable Length Motif-Based Time Series Classification. In: Recent Advances in Information and Communication Technology: Proceedings of the 10th International Conference on Computing and Information Technology (IC2IT2014). pp. 73–82. Springer International Publishing (2014). https://doi.org/10.1007/978-3-319-06538-0_8, backup Publisher: Springer
28. Zhang, K., Wen, Q., Zhang, C., Cai, R., Jin, M., Liu, Y., Zhang, J.Y., Liang, Y., Pang, G., Song, D., et al.: Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. IEEE transactions on pattern analysis and machine intelligence **46**(10), 6775–6794 (2024). https://doi.org/10.1109/tpami.2024.3387317
29. Zhou, Z.H.: A brief introduction to weakly supervised learning. National science review **5**(1), 44–53 (2018). https://doi.org/10.1093/nsr/nwx106