Athlete monitoring in professional road cycling using similarity search on time series data*

Arie-Willem de Leeuw[1], Tobias Oberkofler[2], Mathieu Heijboer[3], and Arno Knobbe[2]

[1] Department of Computer Science, University of Antwerp - imec, Antwerp, Belgium
arie-willem.deleeuw@uantwerpen.be
[2] Leiden Institute of Advanced Computer Science (LIACS), Leiden, the Netherlands
[3] Team Jumbo-Visma, 's-Hertogenbosch, the Netherlands

Abstract. In sports, athlete monitoring is important for preventing injuries and optimizing performance. The multitude of relevant factors during the exercise sessions, such as weather conditions, makes proper individual athlete monitoring labour intensive. In this work, we develop an automated approach for athlete monitoring in professional road cycling that takes into account the terrain on which the ride is executed by finding segments with similar elevation profiles. In our approach, the matching is focused on the shapes of the segments. We use 2.5 years of data of a single rider of Team Jumbo-Visma and assess the performance of our approach by determining the quality of the best matches for a selection of 700 distinct segments, consisting of the most representative shapes for the elevation profiles. We demonstrate that the execution time is within seconds and more than ten times faster than exhaustive search. Therefore, our method enables real-time deployment in large scale applications with potentially many requests from multiple users. Moreover, we show that on average our approach has similar accuracy when considering the correlation to a target segment and approximately only has a twice as large mean squared error when compared to exhaustive search. Finally, we discuss a practical example to demonstrate how our approach can be used for athlete performance monitoring.

Keywords: Sports Analytics · Data Mining · Time Series Data · Road Cycling

1 Introduction

Over the last decades, technical developments have led to new opportunities for detailed athlete monitoring in sports. Here, the main focus is on tracking the fatigue from exercising and the corresponding recovery. In particular, the aim is finding the right balance between training stress and recovery on an athlete-specific level [13]. Monitoring this balance has two main benefits for sports practitioners. First, if training programs contain insufficient recovery, this can result

* This work is part of the Nationale SportInnovatorprijs 2020 *WielerFitheid*, which is financed by ZonMW.

in severe injuries. Therefore, detailed monitoring could signal early symptoms of potentially severe injuries that might occur in the future [18]. Second, the performance of athletes can be enhanced by tracking the adaptions of the body after completing a training session. Hence, coaches keep an eye on their athletes on a daily basis to optimize performance and prevent injuries [7,11].

For a coach, athlete monitoring is a complicated task as there is a multitude of different factors that need to be considered, such as the characteristics of training sessions or the wellness of an athlete [29]. Hereby, it is crucial to interpret the findings by taking into account the right contextual information. For example, an elevated heart rate during a training session could be explained by a higher body temperature due to environmental factors [15]. Additionally, there is a potential risk of missing possible relevant information in the often vast amounts of collected data. As coaches are usually responsible for a group of athletes who all need an individualized analysis, athlete monitoring is labour intensive and one of the most important daily occupations of a coach. Hence, there is a need for efficient and easy-to-implement athlete monitoring methods that can assist coaches in retrieving the most valuable information [21].

Road cycling is a prime example of a sport with many opportunities for developing these automated approaches in athlete monitoring [26]. Bikes of cyclists are typically equipped with multiple sensors and therefore detailed information of the bike rides is available. This collection of ample sensor data opens up many avenues to apply machine learning techniques in elite cycling [12,17,30]. In this work, we will consider an application of machine learning techniques that uses sensor data in road cycling to develop an approach for comparing the performance of a cyclist in different training sessions. In particular, for a given part of a bike ride, we automatically find other bike rides on a similar terrain and compare the physiological characteristics, such as the relationship between heart rate and produced power. By monitoring possible changes in the physiological characteristics of the rider between both bike rides taking place on distinct dates, our method can assist in signaling the physical development of cyclists.

The remainder of this article is structured as follows. First, we review some related work. Hereafter, we discuss the materials that are considered in this work and elaborate on the modeling approach that we have developed. Subsequently, we present the results of experiments on the performance of our approach and give an example of a typical outcome. Finally, we discuss our results and end with a conclusion.

2 Related Work

In this work, we are dealing with time series data. This type of data is omnipresent in multiple domains covering climate studies as well as finance and medicine research. Therefore, there is a large variety in time series data analyses [9,10,14], such as forecasting, classification or regression settings. Here, we consider the task of finding the part of a time series that is similar to a given segment of distinct time series. So, we are in the research area of time series similarity or matching of time series [16].

There are many approaches for addressing matching of time series, ranging from naive brute force methods to statistical analysis and deep representation learning. Most approaches rely on dimensionality reduction of the time series data using techniques such as Discrete Fourier Transform [2], Discrete Wavelet Transform [6] or t-SNE [20]. More recently, the UMAP algorithm [5] is also applied to map single and multi-attribute biomedical time series data, into a lower-dimensional feature space [4]. After the dimensionality reduction, the similarity of time series can be assessed by comparing their key characteristics.

The time series can also be matched by calculating a distance measure between entire time series [8]. Here, the most common approach is a point-by-point comparison of the absolute distance by using the Euclidean distance [31]. Alternatively, the similarity can be determined by using Dynamic Time Warping [25], or by solely focusing on the shape of the time series [3]. After obtaining the distance between a collection of time series, there are two main options for determining the similarity between time series [16]. First, given a time series T, we can explore a database to retrieve all other time series that are within a predefined threshold distance of T. Second, clustering approaches can be applied to find the groups of similar time series. For example, k-mediods clustering with the Dynamic Time Warping distance can be used [22]. An overview of the various approaches of time series clustering can be found in Refs. [1,19].

3 Materials

In this section, we will describe the materials that are used in the study and elaborate on the preprocession that we have applied.

3.1 Materials

We consider 2.5 years of training and competition data of an elite cyclist of Team Jumbo-Visma. During the rides, many attributes are collected by using sensors and a bike computer. We have physiological attributes, such as the produced power and heart rate, and also environmental information, including the location and altitude of the terrain. In total, the rider completed almost 800 sessions. The data of each session is a time series, where the information is collected with a resolution of 1 Hz. Ignoring sessions with malfunction of the bike computer, such as a session shorter than 1 kilometer, we find that the average length and duration of a session is 95.6 ± 56.9 kilometer and 179 ± 95.6 minutes (mean $\pm$ std), respectively.

3.2 Data preprocessing

Before using the data in our modeling approaches, we first apply some preprocessing. In this step, we developed a pipeline to remove outliers, inconsistent

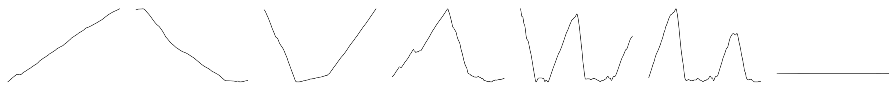

Fig. 1. The seven main distinct shape types that can be encountered when investigating the elevation profile of road cycling rides.

data points and missing values. Most importantly, we applied Gaussian smoothing on the altitude variable to overcome the step-wise increase of the altitude values in the raw data. After exploring different values of the standard deviation of the Gaussian kernel, we set $\sigma = 3$ to remove the discontinuous behavior and in the meantime preserve most fluctuations. Moreover, segments with less than 60 seconds of consecutive missing values, are filled by applying spline interpolation. Finally, we down-sampled the original one Hertz data to a 15 second sampling rate. Although we hereby remove some details, the precision is sufficient to retrieve accurate information for our road cycling application.

4 Methodology

The goal of our work is to retrieve core information between comparable ride segments in different recorded sessions of a given rider. More specifically, we consider the following challenge

Given: A segment S defined as a specific part of an entire workout that is of arbitrary length, selected from a collection of time series, and of the form $\{(d_s, h_s), \ldots, (d_f, h_f)\}$. Here, d_j is the covered distance, h_j is the corresponding altitude, and the indices s and f correspond to the first and last point of the segment, respectively.

Goal: Find the segments that are conditioned on similar terrain, i.e., have a similar elevation profile compared to S.

As mentioned previously in the related work section, there are several options for addressing this task. For practical usage, our approach should be sufficiently fast while minimizing the risks of missing relevant matches. Here, we need to meet this requirement for all different types of elevation profiles. After inspection of all time series data in our database, we find that an elevation profile is typically equal to one of the seven different kind of shapes shown in Fig. 1.

In this work, we have applied two different approaches that are accurate and fast enough for all distinct elevation profiles. Before we will elaborate on these two procedures in more detail, we first discuss an additional step that both approaches have in common.

4.1 Selection of potential matches

For our methods to be sufficiently fast, it is unfeasible to take a naive approach and perform a comparison to all other segments of the same length. Note that this is also unnecessary as the characteristics of most of these segments will be quite different. Therefore, we determine general properties of the selected segment S and preselect potential matches by imposing conditions on these characteristics.

First, we determine the extreme points in altitude of segment S. Next, we define a minimum and maximum allowed altitude that is 10% lower or higher, respectively. Hereafter, we consider all segments of the same length as S and select the ones for which at all time points, the altitude is in between the minimum and maximum allowed altitude. Subsequently, we filter out matches with a Pearson's correlation of less than 0.7 with the query segment.

4.2 Taylor-made approach

After selecting the potential matches, we can apply our approaches for finding similar segments to S. In our first method, we use a similarity measure to determine the similarity between the segment S and the potential matches.

In principle, there are multiple similarity measures that can be applied. For optimizing the usability for sport practitioners in this specific use case, an agreement in shape of the segments is most important. Moreover, in the selection of potential matches, we already ensured that the altitude values are in the same range by restricting the difference between altitude in both segments and enforcing a minimal correlation between them. Therefore, we have chosen the *peak alignment* as our similarity measure, which closely corresponds to human judgment of mountainous terrain in cycling. With the peak alignment, we match the identified peaks between two series in sequential order and compute the sum of weighted horizontal differences between the summits. The differences are weighted by the summits relative altitude that is defined as the absolute altitude of the summit divided by the sessions maximum altitude. Although here we focus on the objective of peak alignment, the approach can be easily extended to allow for different (combinations) of evaluation metrics.

We use the scipy signal package[4] to identify peaks. If there are no clear peaks in the original segment, such as sprints or simple climbs, the Pearson correlation coefficient is used as similarity score. Finally, we sort all matches from most to least similar and remove overlapping segments to ensure variety in the results.

4.3 Dimensionality reduction approach

In our second approach, we utilize techniques of dimensionality reduction to project the time series into low dimensional space and apply a k-nearest neighbour (KNN) classification to identify the most similar segments for any given segment S. In principle, dimensional reduction can be quite time-consuming, as there is a need for extensive preprocessing and the buffering for non-linear reduction methods requires a huge memory. From a practical perspective, it is often not feasible to use this approach if dimensionality reduction has to be

[4] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html`

performed for each new query. However, in the case of fixed sized windows, the computational complexity reduces to a $\mathcal{O}(1)$ lookup operation at execution time. In this case, the dimensionality reduction only has to be performed once and the results can be stored and retrieved on demand. We experimented with different forms of dimension reduction such as UMAP, t-SNE and Principal Component Analysis (PCA). For efficiency reasons, we here opt for PCA.

Before performing the dimensionality reduction, we add the slope of the terrain as additional feature to the original time series data. Hereafter, we apply normalization of the features and learn an n-dimensional representation of segments with a given length k. We project all sub-parts of our data collection with length k into the low-dimensional space and determine the nearest neighbours with the Euclidean distance measure. The value of n is determined by using a representative sample of different distance altitude segments for a given k, and evaluating the average top-5 mean-squared-error as well as the correlation to the target segment.

Although less efficient when not restricted to segments of fixed length, this approach also has an advantage compared to the first method. If we apply an n-dimensional reduction of segments for $n < 4$, we can visualise the characteristics of all segments of a given length k that are present in our data collection. Thereby, we can explore all different type of segments and also find potential clusters of similar fragments. Hence, the user can visually explore the landscape of all available segments and find the similar segments without specifying one specific segment S in advance.

5 Results

In this section, we will present the assessment of the performance of our modeling approaches and illustrate how our approach can be used for athlete monitoring.

5.1 Modeling performance

To obtain realistic estimates of the performance of our approaches under real-world conditions, we present the results of experiments on a sample of 700 manually selected segments of about 1 hour length equally representing the 7 distinct shape types displayed in Fig. 1. The experiments are executed on a machine with 32 x 8-core Intel(R) Xeon(R) CPU E5-2630-v3 central processing units with a combined RAM of 440 GB using parallel creation and comparison of the sliding segment-windows.

Before we can compare our methods, we first need to find the optimal number of components in our dimensionality reduction approach. After investigating the performance on a log-2 scale, we find that the optimal number of components is around 2^4. However, for more than 2^2 components the performance, defined as the average correlation and mean squared error, only marginally increases. We find that representations with 2- or 3-dimensions already result in compelling clusters which can be used for visualization.

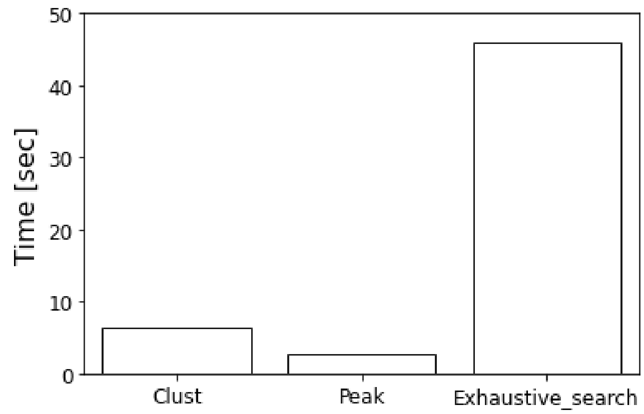

Fig. 2. Computation time for finding the three best matches for 100 segments of each of the 7 distinct shape types of Fig. 1. We display the average computation time for our taylor made approach (peak), dimensional reduction (clust) approach and exhaustive search.

We assess the performance of our method by investigating the computation time and accuracy of our different approaches. In Fig.2, we show the average time necessary for finding the three best matches. We observe that exhaustive search has the largest computational time. This method takes more than ten times longer than our taylor-made approach, which on average needs under three seconds to obtain the three best matches for a given segment S. The overall quality of the matches is displayed in Fig.3. We observe that the overall quality of retrieved matches is very high with an average Pearson's correlation around 0.85 and mean squared error smaller than 120m altitude for all methods. The Pearson's correlation of all methods is comparable, while the exhaustive search

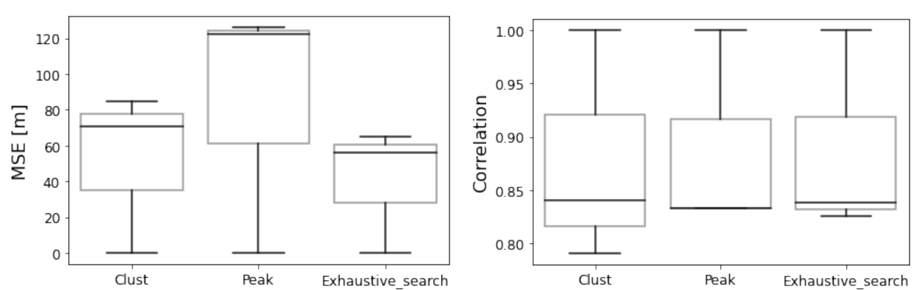

Fig. 3. Comparison of the accuracy of our taylor made approach (peak), dimensional reduction (clust) approach and exhaustive search. We show the distribution of the mean squared error (left) and Pearson's correlation (right) for experiments on a total 700 segments where all 7 different shape types of Fig. 1 are equally represented.

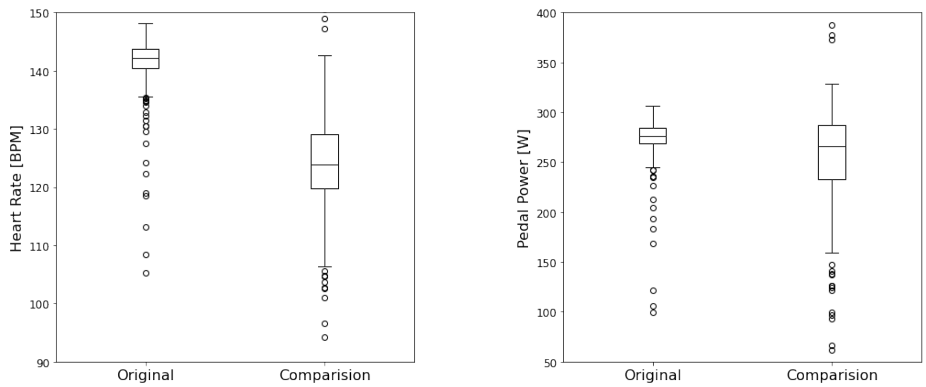

Fig. 4. The heart rate (left) and pedal power (right) values for the selected segment and the best match. We observe there is quite a difference in heart rate values although the produced power in both cases in similar.

method has a mean squared error that is approximately two times smaller than our taylor-made approach. Note that we combined the results for all distinct shapes of the elevation profiles, but we obtain similar results if we consider the distinct shapes separately.

5.2 Athlete monitoring

As an illustration of the usefulness of our approach, we consider of one session the El Teide climb in Spain with starting point in Chio. This case is of particular interest for the coaches as this is a popular training location for cyclists. For this example, we find that the best match is a different climb on Tenerife. Moreover, the top-5 matches for the given query segment all have a Pearson's correlation coefficient above 0.99 and a mean-squared error difference smaller than 50 meters.

To obtain interesting insights for athlete monitoring, we analyze the exercise intensity via the produced pedal power and relate this to the heaviness that is experienced by the rider by means of the heart rate. In Fig. 4, we compare the heart rate and produced pedal power of the original segment and the best match. Although for the most similar segment, there are some more large values, overall the produced pedal power in both cases is quite similar and differs by only 16 W. On the other hand, the heart rate values are much different and on average, we find that the heart rate is almost 18 bpm, or roughly 15%, higher. This found difference in heart rate at similar exercise intensities, points to coach to having a closer look at both sessions and consider some contextual information.

For example, it is worthwhile to compare the temperature in both cases. We observe that during the ride on the original segment the average temperature is almost 8 degrees Celsius lower than during the compared session. This might indicate that the temperature difference could be the explanation for the observed dissimilarity in heart rate. We can further investigate this claim by comparing the second best match to the original segment. In this case, we have a comparable difference in temperature as found before. Moreover, compared to the original segment, the average heart rate is 8 bpm lower and on average the cyclist produced 28 W less power. Hence, there is a larger difference in produced pedal power, but the heart rate is more similar. Therefore, this suggests that the difference in heart rate as shown in Fig. 4 can not only be caused by the change in temperature. In a similar fashion a coach could consider other contextual information, such as the run-up to the segment in both rides. Hereby, it is possible to study whether the observed difference in heart rate at similar produced pedal power was a consequence of a change in fitness of the rider or there was another explanation.

6 Discussion

We have presented three approaches for finding similar elevation profiles in professional road cycling. Although exhaustive search is most accurate, there is only a relatively small gain compared to our other two approaches, especially in terms of the correlation metric. On the other hand, exhaustive search often takes over a minute to find results and our taylor-made method retrieves comparable results in under three seconds. This demonstrates that this approach is suitable for any real-time deployment if the application needs to be scaled up to a service with multiple simultaneous requests.

Thereafter, we illustrated the usefulness of our approach by considering a practical example. Although the races were executed on similar terrain and the exercise intensities were comparable, we observed that there were significant differences in heart rate that could not be explained by only looking at the temperature difference in both sessions. This elevated heart rate at the same exercise intensity could indicate a decrease in performance [32] or a reduction in training volume [23,24] at the original segment. However, before drawing these conclusions, it is important that the coach also takes into account all contextual information. For example, the exercise intensity and exercise duration before starting the segment could have been different. Therefore, it would be worthwhile to extend our approach by including some restrictions, such as enforcing similar physiological starting conditions. For instance, the energy used until the starting point of a segment could by calculated by determining the work done up until starting the segment.

In addition to the application of our method for athlete monitoring, we have another useful utilization for sport practitioners in road cycling. Instead of finding the most similar segment in historical data, we can also use the elevation profile of a future race. In this case, we can select parts that are expected to be important for the race outcome and use our approach to find similar segments in our data collection. Hereby, we can determine specific areas in popular training destinations that have a similar elevation profile. This can be an asset in the build-up for important races as this allows riders to experience the terrain of the race, without the additional need for travelling to the specific location.

There are multiple opportunities for future research. As mentioned before, for athlete monitoring it can be important to extend our approach and include some restrictions on the matches. Moreover, it is interesting to investigate multiple riders as this also allows for comparison of the physiological characteristics between different riders. Finally, we can also explore alternatives for some choices that we have made in our approach. For instance, we have preselected candidate segments based on the extreme values of the altitude of the target segment and the Pearson's correlation. These specific choices are most appropriate for segments that include high altitudes and sufficient elevation differences. On the other hand, a preselection based on absolute errors between the elevation profiles might be more reliable for segments that contain little elevation differences. While the segments with large altitude difference are the most important for coaches as these are typically very demanding for cyclists, it also might be insightful to accurately compare other type of segments. Therefore, an extension of our approach could be a more flexible procedure for the preselection of potential matches that is based on the characteristics of the target segment. Finally, we could also apply different approaches, such as hierarchical clustering [27], or clustering based on Dynamic Time Warping Barycenter Averaging [28].

7 Conclusion

In this work, we have developed methods for athlete monitoring in professional road cycling. We obtained insights about the physical abilities and fatigue of a professional road cyclist by finding similar elevation profiles of bike rides. Our main approach uses multi-stage filtering and peak alignment to assess this similarity, which is most in line with the human perception if segments are alike. We have shown that this approach is sufficiently accurate and fast to allow for real-time application. In addition to comparing the physiological characteristics of a rider between segments with similar elevation profiles occurring on different dates, our approach can also be used to prepare for future races by identifying areas in training locations with similar terrain as a given future race. Concluding, we have constructed a valuable tool for sport practitioners in professional road cycling that can be used for efficient and effective athlete monitoring to support performance optimization.

References

1. Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y.: Time-series clustering–a decade review. Information Systems **53**, 16–38 (2015)
2. Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. In: International conference on foundations of data organization and algorithms. pp. 69–84. Springer (1993)

3. Agrawal, R., Lin, K.I., Sawhney, H.S., Shim, K.: Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In: Proceedings of the 21th International Conference on Very Large Data Bases. p. 490–501. VLDB '95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995)
4. Ali, M., Jones, M.W., Xie, X., Williams, M.: Timecluster: dimension reduction applied to temporal data for visual analytics. The Visual Computer **35**(6), 1013–1026 (2019)
5. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W., Ng, L.G., Ginhoux, F., Newell, E.W.: Dimensionality reduction for visualizing single-cell data using umap. Nature biotechnology **37**(1), 38–44 (2019)
6. pong Chan, K., Fu, A.W.C.: Efficient time series matching by wavelets. Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337) pp. 126–133 (1999)
7. Coutts, A., Cormack, S.: Monitoring the training response. High-performance training for sports pp. 71–84 (2014)
8. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. Proceedings of the VLDB Endowment **1**(2), 1542–1552 (2008)
9. Esling, P., Agon, C.: Time-series data mining. ACM Comput. Surv. **45**(1) (dec 2012). https://doi.org/10.1145/2379776.2379788, `https://doi.org/10.1145/2379776.2379788`
10. Fakhrazari, A., Vakilzadian, H.: A survey on time series data mining. In: 2017 IEEE International Conference on Electro Information Technology, EIT 2017. pp. 476–481. IEEE International Conference on Electro Information Technology, IEEE Computer Society (Sep 2017). https://doi.org/10.1109/EIT.2017.8053409, 2017 IEEE International Conference on Electro Information Technology, EIT 2017 ; Conference date: 14-05-2017 Through 17-05-2017
11. Halson, S.L.: Monitoring training load to understand fatigue in athletes. Sports medicine **44**(2), 139–147 (2014)
12. Hilmkil, A., Ivarsson, O., Johansson, M., Kuylenstierna, D., van Erp, T.: Towards machine learning on data from professional cyclists. arXiv preprint arXiv:1808.00198 (2018)
13. Kellmann, M.: Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring. Scandinavian journal of medicine & science in sports **20**, 95–102 (2010)
14. Keogh, E., Kasetty, S.: Kasetty, s.: On the need for time series data mining benchmarks: A survey and empirical demonstration. data mining and knowledge discovery 7(4), 349-371. Data Mining and Knowledge Discovery **7**, 349–371 (01 2003). https://doi.org/10.1023/A:1024988512476
15. Kirschen, G.W., Singer, D.D., Thode Jr, H.C., Singer, A.J.: Relationship between body temperature and heart rate in adults and children: A local and national study. The American Journal of Emergency Medicine **38**(5), 929–933 (2020)
16. Kontaki, M., Papadopoulos, A., Manolopoulos, Y.: Similarity Search in Time Series Databases (01 2006). https://doi.org/10.4018/9781591405603.ch106
17. de Leeuw, A.W., Heijboer, M., Hofmijster, M., van der Zwaard, S., Knobbe, A.: Time series regression in professional road cycling. In: Appice, A., Tsoumakas, G., Manolopoulos, Y., Matwin, S. (eds.) Discovery Science. pp. 689–703. Springer International Publishing, Cham (2020)
18. de Leeuw, A.W., van der Zwaard, S., van Baar, R., Knobbe, A.: Personalized machine learning approach to injury monitoring in elite volleyball players. European journal of sport science pp. 1–10 (2021)

19. Liao, T.W.: Clustering of time series data—a survey. Pattern recognition **38**(11), 1857–1874 (2005)
20. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
21. McGuigan, H., Hassmén, P., Rosic, N., Stevens, C.: Monitoring of training in high-performance athletes: What do practitioners do? Journal of Science in Sport and Exercise **5**, 121–129 (03 2021). https://doi.org/10.36905/jses.2021.02.05
22. Meert, W., Hendrickx, K., Craenendonck, T.V.: wannesm/dtaidistance v2. 0.0. Zenodo (2020)
23. Mujika, I., Padilla, S.: Detraining: loss of training-induced physiological and performance adaptations. part i: short term insufficient training stimulus. Sports medicine **30**(2), 79–87 (2000). https://doi.org/10.2165/00007256-200030020-00002
24. Mujika, I., Padilla, S.: Detraining: loss of training-induced physiological and performance adaptations. part ii: Long term insufficient training stimulus. Sports medicine **30**(3), 145–54 (2000). https://doi.org/10.2165/00007256-200030030-00001
25. Müller, M.: Dynamic Time Warping, pp. 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
26. Municio, E., Daneels, G., De Brouwer, M., Ongenae, F., De Turck, F., Braem, B., Famaey, J., Latré, S.: Continuous athlete monitoring in challenging cycling environments using iot technologies. IEEE Internet of Things Journal **6**(6), 10875–10887 (2019)
27. Nielsen, F.: Hierarchical Clustering, pp. 195–211 (02 2016)
28. Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern recognition **44**(3), 678–693 (2011)
29. Saw, A.E., Main, L.C., Gastin, P.B.: Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. British Journal of Sports Medicine **50**(5), 281–291 (2016). https://doi.org/10.1136/bjsports-2015-094758, `https://bjsm.bmj.com/content/50/5/281`
30. Silacci, A., Taiar, R., Caon, M.: Towards an ai-based tailored training planning for road cyclists: A case study. Applied Sciences **11**(1) (2021). https://doi.org/10.3390/app11010313, `https://www.mdpi.com/2076-3417/11/1/313`
31. Soon, L.K., Lee, S.H.: An empirical study of similarity search in stock data. In: Proceedings of the 2nd International Workshop on Integrating Artificial Intelligence and Data Mining - Volume 84. p. 31–38. AIDM '07, Australian Computer Society, Inc., AUS (2007)
32. Thorpe, R., Atkinson, G., Drust, B., Gregson, W.: Monitoring fatigue status in elite team-sport athletes: Implications for practice. International journal of sports physiology and performance **12**(Suppl 2), S227–S234 (2017). https://doi.org/10.1123/ijspp.2016-0434