

Latent Space Unsupervised Semantic Segmentation

Knut J. Strommen¹, Jim Tørresen¹, Ulysse Côté-Allard¹

Abstract—The development of compact and energy-efficient wearable sensors has led to an increase in the availability of biosignals. To analyze these continuously recorded, and often multidimensional, time series at scale, being able to conduct meaningful unsupervised data segmentation is an auspicious target. A common way to achieve this is to identify change-points within the time series as the segmentation basis. However, traditional change-point detection algorithms often come with drawbacks, limiting their real-world applicability. Notably, they generally rely on the complete time series to be available and thus cannot be used for real-time applications. Another common limitation is that they poorly (or cannot) handle the segmentation of multidimensional time series. Consequently, the main contribution of this work is to propose a novel unsupervised segmentation algorithm for multidimensional time series named Latent Space Unsupervised Semantic Segmentation (LS-USS), which was designed to work easily with both online and batch data. When comparing LS-USS against other state-of-the-art change-point detection algorithms on a variety of real-world datasets, in both the offline and real-time setting, LS-USS systematically achieves on par or better performances.

Index Terms—Multi-dimensional Time Series, Semantic Segmentation, Unsupervised Learning

I. INTRODUCTION

The physiological processes occurring in the human body generate a plethora of biosignals (e.g. motion, muscle activity, biopotential) that can provide otherwise inaccessible insights into a person’s health and activity, and may even be leveraged for the development of human-computer interfaces [1]. Through the rise of the Internet of Things and the development of more compact and energy efficient sensors, wearable technologies can now provide continuous, non-intrusive and multimodal monitoring of these biosignals in real-time. However, the sheer amount of data generated by these devices make the processing and analysis of this information challenging and time-consuming to perform. As an example, a single 9-axis inertial measurement unit cadenced at a low sampling rate of 60Hz generates around 2 million data-points every hour. Having to manually annotate these unlabelled data streams can thus rapidly become impractical. Another commonly occurring type of data are weakly labeled time series which include imprecise or inexact labels of when a change actually occurred. For both unlabeled and weakly labeled time series, the most useful information often lies in the precise location where a transition in the time series take place. Thus to make use of this data at scale, being able to identify these critical points in the time series in an unsupervised manner can become highly beneficial. Unsupervised change-point detection (CPD)

algorithms attempt to identify these abrupt change in the data generating process [2].

Unfortunately, CDPs algorithms also tend to suffer from limitations that reduce their suitability for real-world applications. Notably, they tend to make assumptions and require prior knowledge of the data which in practice implicitly or explicitly restrict them to a specific domain (as opposed to being domain agnostic) [2–4]. Another common limitation is that many CDPs algorithms are defined only for offline applications (i.e. they require having access to the full time series before performing the segmentation) [2, 5]. Further, for many real-world applications fast change-point detection are necessary to execute time-sensitive actions. Thus, algorithms that cannot handle online streaming data are ill-adapted for this type of reality, as they require a complete rerun every time a new data point is added. This problem is compounded by the fact that time series are often recorded over a long time and with a high sampling rate. Thus, algorithms with high computational complexity or poor scalability are also inappropriate for these types of real-world applications. Therefore, developing an unsupervised semantic segmentation algorithm that is hyperparameter lite, domain agnostic, scalable and works on online streaming data is an auspicious target.

One state-of-the-art algorithm, which meets many of these requirements is: Fast Low-cost Unipotent Semantic Segmentation [3] (FLUSS). FLUSS is a domain agnostic, scalable CDP algorithm, that can work with both online and offline data. The main hypothesis behind the development of FLUSS is that subsequences (small snippets of the time series data extracted around each time step in the time series) in similar segments are more similar than subsequences occurring after a changepoint. FLUSS handles multidimensional data by implicitly calculating the likelihood of a change-point at each time step for each channel independently before taking the average likelihood over all the channels. However, some of the considered dimensions might not contain information that is helpful for the segmentation, or they might be heavily correlated. Thus, when taking the average over all the channels, these “not so useful” channels will dilute the information from the meaningful channels. The authors acknowledge this problem and recommend that when dealing with high dimensional data one should do a search or manually find the most useful subset of channels and remove the rest.

As time series data is often sampled from multiple sources and sensors, especially when working with wearable devices, it would be helpful to have an algorithm that can automatically extract the most useful information from high-dimensional times series data during segmentation. Consequently, the main contribution of this work is the introduction of Latent Space Unsupervised Semantic Segmentation (LS-USS). LS-USS is

¹Knut J. Strommen, Jim Tørresen and Ulysse Côté-Allard are with the University of Oslo, Oslo, Norway. Corresponding author: Knut J. Strommen (knut.jstr@ifi.uio.no)

an unsupervised semantic segmentation algorithm that is hyperparameter lite, domain agnostic and capable of working with both streamed online and offline multidimensional data. LS-USS is based on FLUSS, but instead of using the similarity between subsequences, it finds the similarity between lower-dimensional encodings of the multidimensional subsequences obtained via an autoencoder. The idea is that the dimensionality reduction learned by the autoencoder will reduce redundant and correlated information between channels, which in turn will improve time series segmentation.

This work is organized as follows; an overview of the related work is given in Section II. Section III introduces the notions necessary for the description of the proposed algorithm in Section IV. The experiments are then presented in Section V. Finally, the results and their associated discussion are covered in Section VI and VII respectively.

II. RELATED WORK

CPD algorithms can be divided into two categories depending on their reliance on labeled data: Supervised and Unsupervised CPD. Supervised CPD usually entails extracting subsequences over a sliding window where each subsequence is assigned a class or label. Most of the classifiers used in other machine learning tasks are also leveraged for supervised CPD such as naïve Bayes [6], Support Vector Machines (SVMs) [6], Gaussian Mixture Models (GMMs) [6], Decision Trees[6], Hidden Markov Models [7], and Neural Networks [8]. When labels are available, the supervised methods are often the preferred solution. Unfortunately, labeling time series data can be prohibitively expensive (in terms of time, cost and/or human labor) making it an impractical solution for a wide variety of domains. Another challenge with supervised CPD is that they generally require training samples from all possible states (classes) of the signals beforehand and thus cannot easily adapt to novel behavior in the time series. Therefore, unsupervised CPD remains of great interest in many practical applications.

Most of the existing unsupervised CPD make statistical assumptions on the data (e.g. stationarity, independent and identically distributed) [9–15], and/or require extensive tuning of model parameters as they rely on predefined parametric models [15–17]. As such, the type of algorithms can be cumbersome to apply on new domains and less robust to changes in the data over time. In contrast, LS-USS was designed to work using only minimal assumptions on the data and without requiring domain knowledge.

Another common limitation of existing CPD algorithms is that they require being used on batched data and are thus ill-suited for real-time applications [2, 15, 16, 18] such as detecting change-points in a patient’s vital signs [19] or continuously monitoring the wear and tear of industrial robots [20]. In contrast, LS-USS can be efficiently updated every time a new data point is added to the time series allowing it to run in real-time.

A popular form of unsupervised CPD algorithms relies on clustering subsequences of the time series [21–23]. These methods cluster individual subsequences extracted from running a sliding window over the time series. The idea being

that if two temporally close subsequences belong to different clusters, a change-point most likely exist between the two. However, Keogh et al. [24] have shown that using subsequence clustering essentially produces cluster centers which tend towards a sinusoidal signal with random phases that average to the mean of the time series for any dataset used. In other words, the change-points detected are essentially random. To address this issue, Keogh et al. [24] proposes to instead cluster the time series data points using *motifs*. In the context of data mining, motifs can be seen as fingerprints for time series data, as all non-random time series data produced by the same process are bound to contain some reoccurring patterns. Thus, the difference between clustering motifs and subsequence clustering is that in the former case clusters are made up of short individual time series, while in the latter case the clusters are derived from subsequences extracted from a sliding window. In other words, clustering motifs means that the cluster-centers do not represent averages over all the data but averages over motifs (similar patterns in the data). Several methods have been proposed to extract motifs in time series data [18, 25–27]. In particular, *matrix profile* [27] is a recent approach to efficient motif discovery. As matrix profiles are a core aspect of LS-USS, they will be presented in more details in section III-A.

Many of the current states-of-the-art CPD algorithms [16, 28–30], including FLUSS [3], were primarily designed for one-dimensional data. Unfortunately, this also means that they often cannot efficiently be applied to multidimensional data [2] as they poorly (if at all) take into account that multidimensional time series often have varying levels of correlated information across dimensions which can increase the complexity of finding change-points accurately. To alleviate this issue, CPD algorithms have been specifically designed for multidimensional data [31–36]. In [31] principal component analysis (PCA) is first used to obtain a one-dimensional signal (using the principal component) before performing the segmentation. Kim et al. [32] aims to segment out driving patterns from sensors on driving vehicles by leveraging word2vec [37] to make an encoded representation of time series data consisting of both categorical and numerical information in varying scales. Autoencoders [38] have also been employed in the domain of anomaly detection with great success [34–36]. The most common approach is to first train the autoencoder on time series data which does not contain any anomalies (or as few as possible). At inference time, the reconstruction error is then used as a measure of how likely the current data contain an anomaly. The idea being that if the autoencoder cannot reconstruct the current signal well enough, it is most likely because it differs from what was seen during training. While this approach successfully finds change-points that are anomalies, it is not straightforward to adapt it for segmentation as the different segments are a natural part of the data. In other words, an autoencoder trained on specific segments will also have a low reconstruction error whenever the data is part of one of the trained segment. Nevertheless, as shown in [33], unsupervised segmentation can still be performed using an autoencoder while achieving state-of-the-art results by calculating a distance between consecutive window in the latent

space. This method referred to as Latent Feature Maximal Distance (LFMD), is illustrated in Figure 1. The central idea behind LFMD, using a learned latent space of an autoencoder as a way to efficiently characterize multidimensional data for CDP is also a core concept in LS-USS. As such, LFMD will be used in this work to better contextualize the performance of the proposed algorithm.

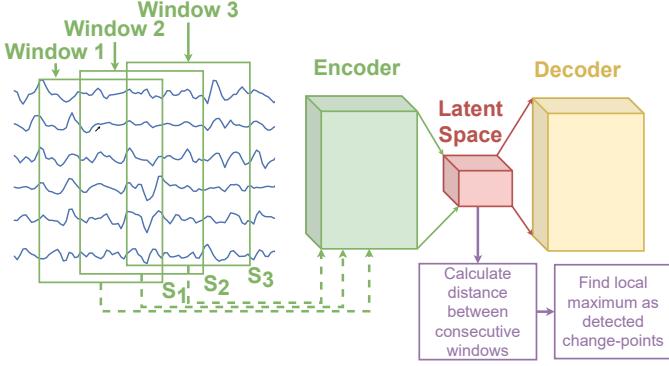


Fig. 1. The diagram of the LFMD algorithm presented in [33]. The distance of the latent features are calculated between each adjacent window. The distance's local maxima are selected as the change-points to be returned by the algorithm.⁵

III. PRELIMINARIES

The following section presents an overview of the fundamental building blocks used by LS-USS.

A. Matrix Profiles

One of the core concepts behind LS-USS is the matrix profile [27], a data structure for time series that facilitates change-point detection and motif discovery. The matrix profile represents the distance between each motif and their most similarly associated motif (excluding themselves). To calculate the matrix profile, one first has to find the set of all subsequence \mathbf{A} from the time series \mathbf{T} by utilizing a sliding window of length $m \in \mathbb{N}$ using a step size of 1 to extract all the possible subsequences of \mathbf{T} . After this a distance matrix is calculated by computing the z-normalized Euclidian distance between every subsequence in \mathbf{A} with every other subsequence in \mathbf{A} . Figure 2 shows what the resulting distance matrix looks like for the time series GunPoint from the UCR time series archive [39]. Each row in the distance matrix, referred to as the distance profile \mathbf{D} depicts the distance from the subsequence at the current row index to every other subsequence in the time series. The distance profile \mathbf{D} is calculated efficiently in $O(n \log(n))$, using a technique referred to as Mueen's algorithm for similarity search (MASS) [40].

The matrix profile is then obtained by taking the Euclidian distances between all the subsequences in \mathbf{A} with the nearest non-trivial neighbor in \mathbf{A} itself. In other words, the matrix profile is created by extracting the smallest value from each row in the distance matrix, as this will be the distance to the nearest neighboring subsequence. Note, however, that the most similar subsequence to a given subsequence will always be the subsequence itself (as they are identical). Similarly,

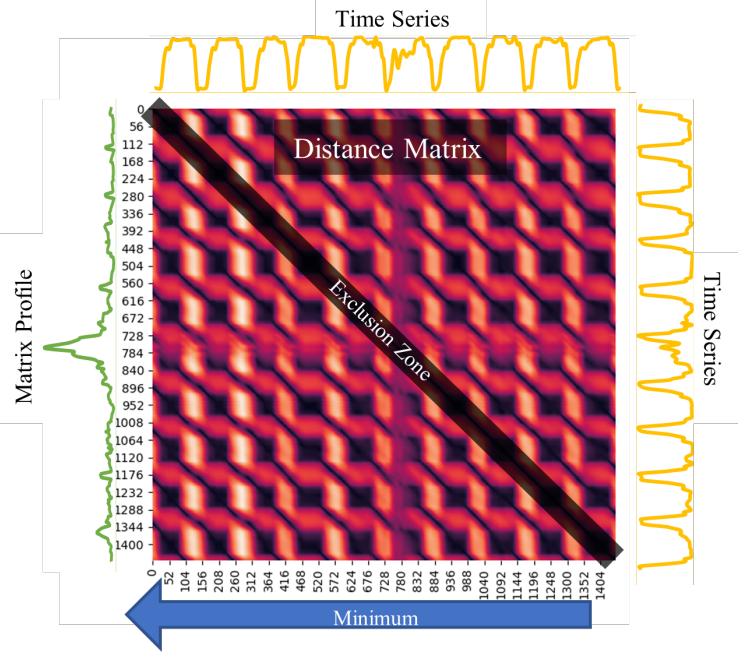


Fig. 2. Plot depicting a time series and its corresponding distance matrix and matrix profile. The time series presented is an excerpt from the dataset NoGunGun from the UCR time series archive.

subsequences that are extracted close in time will also be nearly identical. To avoid these trivial matches, an exclusion zone around each index is created by setting the values in these areas of \mathbf{A} to infinity. As seen in Figure 2, this exclusion zone will be along the diagonal of the distance matrix.

The matrix profile can be used to facilitate the discovery of motifs and discord in the data. In the areas with relatively low values, the subsequences in the original time series must have (at least one) relatively similar subsequence elsewhere in the data. These regions are reoccurring patterns which are classified as motifs and always come in pairs. In contrast, for areas with relatively high values, the subsequence in the original time series must be a unique shape since it does not match any other subsequence. These areas of discords can be considered anomalies in the data. Figure 3 shows an example of how motifs and discords can easily be identified using a matrix profile.

When calculating the matrix profile \vec{p}_A , one can also extract the index of the nearest neighbor for each row in the distance matrix to make the matrix profile index \vec{i}_A , which is defined as the vector containing the indices of the nearest non-trivial neighbor in \mathbf{A} for every subsequence in \mathbf{A} . The matrix profile index can thus be seen as a time series containing pointers to the nearest neighbors for each subsequence and will serve as the basis for the segmentation algorithm in LS-USS. The matrix profile can be efficiently calculated in $O(n^2 \log(n))$ using the Scalable time series Anytime Matrix Profile (STAMP) [27] outlined in Algorithm 1.

B. Fast Low-cost Unipotent Semantic Segmentation

Fast Low-cost Unipotent Semantic Segmentation (FLUSS) [3] is a segmentation algorithm that builds

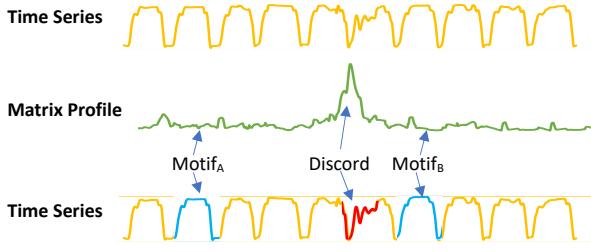


Fig. 3. The distance matrix is a matrix constructed from all the distance profiles, which shows how similar each subsequence in the times series is to every other subsequence in the same time series. The matrix profile is derived by taking the minimum distance of each row. The original time series is shown in yellow and the matrix profile in green.

Algorithm 1 Scalable time series Anytime Matrix Profile (STAMP)

Input:

\vec{t} – time series

m - subsequence length

Output:

\vec{p}_A, \vec{i}_A – The incrementally updated matrix profile and the associated matrix profile index

```

1: procedure STAMP( $\vec{t}, m$ )
2:    $\vec{p}_A = \text{inf}'s, \vec{i}_A = \text{zeros}$ 
3:    $A \leftarrow \text{the all-subsequence set from } \vec{t}$ 
4:   for  $\text{idx}=0 : \text{length}(\vec{t})-m$  do
5:      $D[\text{idx}] = \text{MASS}(A[\text{idx}], TA)$   $\triangleright$  MASS calculates
       the distance profile for the current subsequence
6:      $\vec{p}_A[\text{idx}], \vec{i}_A[\text{idx}] = \text{ElementWiseMin}(D[\text{idx}])$  Save
       the minimum values in the distance profile and the corresponding index
return  $\vec{p}_A, \vec{i}_A$ 

```

upon the matrix profile. This algorithm utilizes the matrix profile index (\vec{i}_A), to segment time series data. The intuition behind FLUSS can be illustrated through this simple example: In time series, data gathered from a person wearing an activity sensor and performing the activities walking and running, one would expect that most of the walking subsequences would point to other walking subsequences, and most running subsequences would point to other running sequences. In other words, for a given index in the time series \vec{t} , the number of arcs crossing "over" that index would be small if it is in an area where the time series is changing, whereas a high arc-count would be expected in areas with a clear homogeneous pattern. If these arc-crossings are counted for each index, the end result is the Arc Curve (AC).

The AC of a given time series \vec{t} of length n will itself be a time series of length n containing only non-negative values. The value at the i^{th} index in the AC specifies the number of arcs spatially crossing over location i in the original time series \vec{t} [3].

Figure 4 shows that the AC is close to zero at the transition between the segments. However, by definition, the arc count will naturally be lower closer to the edges until it become

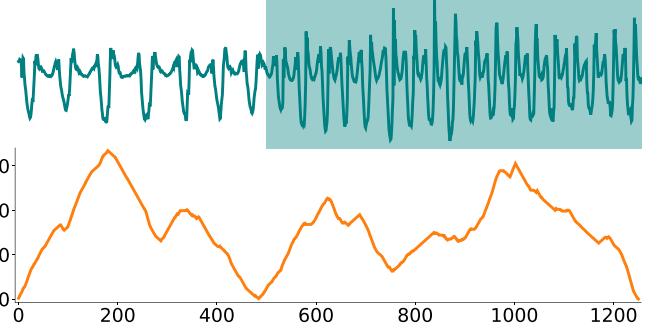


Fig. 4. The top plots show a time series recorded from a gyroscope placed on a person's arm [41]. During the recording, the person is initially walking but then starts jogging at around timestep 500. The bottom plot shows the corresponding arc curves. Importantly, the arc counts are low around timestep 500, corresponding to the time series true change-point. However, the arc counts also decrease towards the beginning and end of the time series, which corresponds to a construction artifact that has to be corrected for.

zero as no arcs can cross the borders of the time series. To compensate for this, the authors in [3] divide the AC with an inverted parabola with a height equal to half the length of the time series. This parabola is called the idealized arc curve (IAC) and is what the AC would look like for a time series with no structure, where all arc curves would just point to random locations. The empirical and theoretical IAC is depicted in the top plot in Figure 5. Dividing the AC with the IAC will normalize the time series between 0 and 1 and solve the edge effects. The resulting vector is known as the Corrected Arc Curve (CAC). A min function is also applied to ensure that the CAC is between 0 and 1, even in the unlikely event that $AC > IAC$.

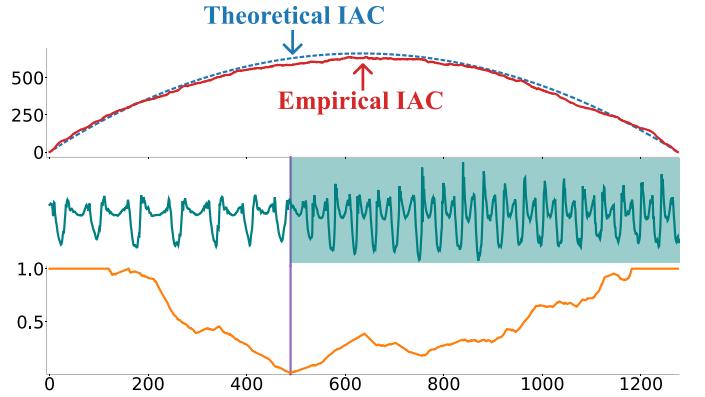


Fig. 5. The top plot shows the empirical vs theoretical idealized arc curve (IAC). The bottom plot shows the corrected arc curve (CAC) computed on the walking-jogging time series from [41]. As shown by the purple vertical line, the minimum value on the CAC can be used to identify this time series' change-point.

1) Regime Extraction Algorithm: Low values in the CAC are used to identify potential change-points. In [3], change-points are then selected using the Regime Extraction Algorithm (REA). REA search for the k lowest "valley" points in the CAC. However, if the point at time t is the lowest point on the CAC, the points at time $t-1$ and $t+1$ are likely the second and third-lowest point. To avoid all the change-points ending up in one "valley", an exclusion zone is set around

the already detected “valley” points, the width of which is an hyperparameter. Note that REA only works when the number of change-points is known beforehand. Unfortunately such a requirement can be hard/impossible to fulfill, in practice, for a wide variety of applications. The pseudo-code for REA is outlined in Algorithm 2.

Algorithm 2 Region Extraction Algorithm (REA)

Input:

CAC - a Corrected Arc Curve

numRegimes - number of regime changes

NW - Subsequence size

Output:

locRegimes - the location of the change-points

```

1: procedure REA(CAC,numRegimes, NW)
2:   locRegimes = empty array of length numRegimes
3:   for idx=0 : numRegimes do
4:     locRegimes[i] = indexOf(min(CAC))
5:     Set exclusion zone of  $5 \times NW$  around locRegimes[i]
            $\triangleright$  To prevent matches to "self"
   return locRegimes

```

2) *Fast Low-cost Online Semantic Segmentation*: Adding a new point to the CAC takes only $O(n\log(n))$. However, removing the oldest point takes $O(n^2)$ as every subsequences could point to the ejected point, thus requiring the whole matrix profile to be updated. As such, while FLUSS can easily be run on offline datasets, it does not scale well in the context of streaming data. To solve this issue, the authors in [3] propose an online version of FLUSS, referred to as Fast Low-cost Online Semantic Segmentation (FLOSS). FLOSS addresses the online streaming issue by explicitly forcing each arc to only point towards a newer data point. Consequently, as no arc can point to an older point, ejecting the oldest point in the currently considered time series can now be done in $O(1)$. Thus, maintaining the one-directional CAC_{1D} can be done in $O(n\log(n))$. It should be noted however that in the case of the CAC_{1D}, the IAC will now be skewed to the right as the rightmost part of the CAC_{1D} will have a higher chance of arc crossings (see Figure 6).

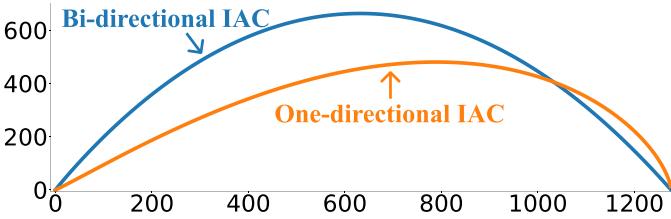


Fig. 6. The bi-directional idealized arc curve (IAC) vs. the one-directional IAC.

FLOSS in conjunction with the one-directional CAC thus enables online streaming using the matrix profile. However, it should be expected that FLOSS will perform worse than FLUSS due to the directional constraints imposed on the arcs.

3) *Locality - Temporal Constraint*: In some time series datasets, the same activities or events can arise multiple times

in a disjointed manner (e.g. a time series representing the electrocardiogram of a person laying down and standing up multiple times in succession). In these cases, the CAC is not a good indicator of a regime change as the arcs for a given event-type (e.g. laying down) would have practically the same likelihood of pointing to subsequences in the segment they are occurring in than they would have in pointing to subsequences from other segments of the same event-type. This challenge can be addressed by using a temporal constraint (TC) that limits how far in time an arc can point, thus forcing each index in the CAC to only consider a local area around itself. While this adds a hyperparameter (length of TC) to the algorithm, it also reduces the computational time of FLUSS. If considered, this hyperparameter is suggested in [3] to be set to circa maximum expected segment length, thus allowing to inject some domain knowledge to the algorithm if available. Furthermore, as pointed out in [3], this hyperparameter is not very sensitive and thus only requires a rough idea of the order of the temporal scale at which the change-points occurs.

C. Autoencoder

Autoencoders [38] are a type of neural network which aim to learn a mapping from a high-dimensional observation to a lower-dimensional feature space (latent space) in an unsupervised manner. Importantly, this mapping is learned with the aim of being able to reconstruct as closely as possible the original observation from the lower-dimensional feature space.

In other words, define \mathcal{X} as the input space and consider a probability measure p on \mathcal{X} . The goal of an autoencoder is then to learn two functions ψ and ϕ , such that:

$$\psi : \mathcal{X} \rightarrow \mathcal{Z} \quad (1)$$

$$\phi : \mathcal{Z} \rightarrow \mathcal{X} \quad (2)$$

$$\psi, \phi = \arg \min_{\psi, \phi} \mathbb{E}_{x \sim p} \left[\| x - (\phi \circ \psi)(x) \| \right] \quad (3)$$

Where \mathcal{Z} is the latent feature space that serves as an information bottleneck to the original feature space.

The input of the autoencoder is the time series partitioned into windows of length w . Note that these subsequences can be defined with an overlap. The detailed autoencoders’ architectures employed in this work will be presented in Section IV.

IV. METHOD

LS-USS proposes to perform unsupervised segmentation by leveraging an autoencoder to learn a meaningful latent feature space from which the corrected arc curve can be computed from. The following details the proposed algorithms and contributions of this work.

A. Autoencoder Implementation

For comparison’s sake, the fully connected network’s architecture used by LFMD [33] is employed in this work. Further, a simple convolutional network is also considered to evaluate if modeling the input’s temporal characteristics can help find a better latent representation than the fully connected version.

1) *Fully Connected Model*: Following the autoencoder implementation [33], the fully connected model reproduced uses two hidden layers for both the encoder and decoder, transposed weights for the decoder, and the sigmoid function as its activation function. For the decode encoder, each hidden layer is half the size of the previous layer and the opposite for the decoder layer (the initial size of which will depend on the input/subsequence size). Adam [42] is employed to optimize the network's weights using the mean square error as the loss function. Finally, the dimension of the latent space is defined so that the ratio between the feature representation and the input $\frac{\dim(x)}{\dim(z)}$ is 0.1. Where $x \in \mathcal{X}$ and $z \in \mathcal{Z}$.

2) *Convolutional Model*: The architecture employed for the convolutional model is presented in Table I. The model is trained in the same way as the fully connected model using mean square error as loss function and Adam as optimizer. The feature size, which is now defined as the ratio ($\frac{\dim(z)}{NC \times NW}$), is set to $\frac{1}{6}$.

B. Latent Space Matrix Profile

As previously stated, LS-USS leverages an autoencoder as a way to learn a meaningful representation of a multidimensional time series. Because the time series is first segmented into windows with each window going through the autoencoder, the resulting representation of the time series in the latent space is not a continuous time series. Unfortunately, this also means that the matrix profile cannot be computed from this latent space using the STAMP algorithm as presented in Section III-A as doing so is only possible if a sliding window can be applied to the time series itself. Thus, this work introduces the Latent Space Matrix Profile (LSMP) which, as the name implies, is the matrix profile computed directly on the latent representation. The LSMP is defined by the same logic used for calculating the regular matrix profile (see Section III-A). The difference being the use of latent representation instead of subsequences. By exploiting the fact that each distance calculation can be done independently, it is possible to implement the computation of the Euclidean distance between pairs of vectors to run in parallel (e.g. GPU). The pseudo-code to compute the lsmp is presented in Algorithm 3.

Similarly to FLUSS, performing time series segmentation based on the LSMP and CAC is only meaningful when the same segment-type is not repeating over the time series (see Section III-B3). Therefore, a TC is also applied when computing the CAC from the LSMP. Adding this TC also comes with the added benefit that only the distances between the latent features located inside the temporal constraint need to be calculated. The top of figure 7-A depicts the collapse of the temporally constrained latent distance matrix into the LSMP.

A memory problem arise for long time series as it is necessary to save 2^*TC data points per time step to make the temporally constrained distance profile, which can rapidly become intractable. To address this memory issue, instead of considering the full distance matrix (constrained by TC), this work instead propose to use an overlapping window on the

Algorithm 3 Latent Space Matrix Profile (LSMP)

Input:

T_A – time series

m - subsequence length

Output:

P_F, I_F – The incrementally updated latent space matrix profile and the associated latent space matrix profile index

```

1: procedure LSMP( $T_A, m$ )
2:    $P_A = inf$ 's,  $I_A = zeros$ 
3:    $F \leftarrow$  the latent all-subsequence set from  $T_A$ 
4:   for  $idx=0 : length(T_A)-m$  do
5:      $D[idx] = CDIST(A[idx], TA)$   $\triangleright$  CDIST calculates
       the distance profile for the current subsequence.
6:      $P_F[idx], I_F[idx] = ElementWiseMin(D[idx])$  Save
       the minimum values in the distance profile and the corre-
       sponding index
return  $P_F, I_F$ 
```

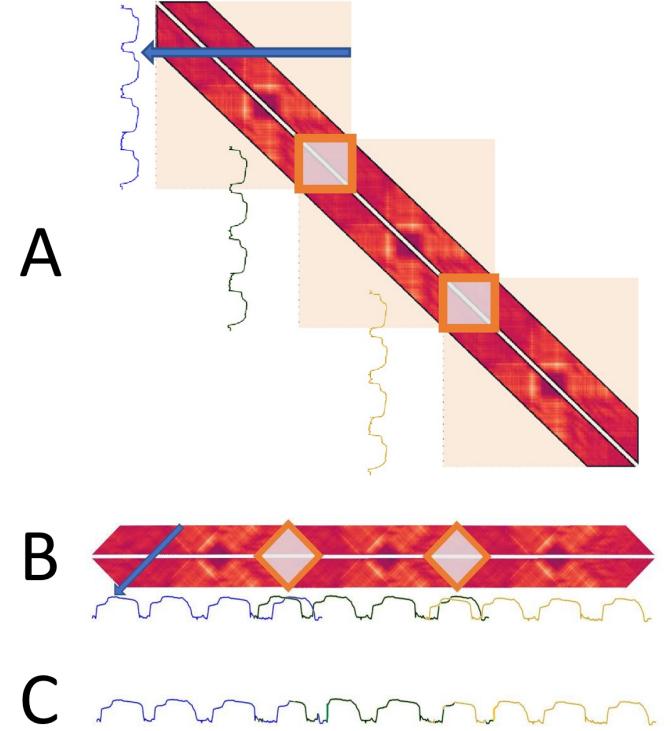


Fig. 7. The top part of figure A (blue arrow) depicts the batched collapse of the temporally constrained latent distance matrix into the LSMP. Figure A as a whole shows the temporally constrained matrix profile of a time series of length 10400. Figure B shows the three batched matrix profiles calculated where they overlap with each other. Figure C shows the full matrix profile after taking the minimum value in the overlapping area.

Layer	Type	Input Length	Output Length	Input Width	Output Width	Kernel Size	Stride	Padding	Activation
Hidden Layer 1 Encoder	Convolution	NW	NW/2	NC	2xNC	3	2	1	ReLU
Hidden Layer 2 Encoder	Convolution	NW/2	NW/4	2xNC	4xNC	3	2	1	ReLU
Reshape Encoder	Reshaping	NW/4	NW	4xNC	1	-	-	-	-
Hidden Layer 3 Encoder	Fully Connected	NW	NW/2	1	1	-	-	-	ReLU
Hidden Layer 4 Encoder	Fully Connected	NW/2	NW/4	1	1	-	-	-	ReLU
Hidden Layer 5 Encoder	Fully Connected	NW/4	NW/6	1	1	-	-	-	ReLU
Hidden Layer 1 Decoder	Transposed Hidden Layer 5 Encoder	NW/6	NW/4	1	1	-	-	-	ReLU
Hidden Layer 1 Decoder	Transposed Hidden Layer 4 Encoder	NW/4	NW/2	1	1	-	-	-	ReLU
Hidden Layer 2 Decoder	Transposed Hidden Layer 3 Encoder	NW/2	NW	1	1	-	-	-	ReLU
Reshape Decoder	Reshaping	NW	NW/4	1	4xNC	-	-	-	-
Hidden Layer 1 Encoder	Transposed Convolution	NW/4	NW/2	4xNC	2xNC	3	1	1	ReLU
Hidden Layer 1 Encoder	Transposed Convolution	NW/2	NW	2xNC	NC	3	2	1	Linear

TABLE I

OVERVIEW OF THE ARCHITECTURE USE FOR THE CONVOLUTIONAL NETWORK VERSION OF THE AUTOENCODER. NC IS THE ORIGINAL NUMBER OF CHANNELS FROM THE INPUT DATA, WHILE NW REPRESENTS THE SUB-SEQUENCE LENGTH.

time series where a distance matrix will be computed for each window. A matrix profile for each distance matrix is then computed. Finally, the different matrix profiles are merged together by taking the minimum value (and the corresponding index) over the overlapping area of each matrix profile. This operation is referred to as the batched collapse of the matrix profile, a depiction of which is shown in Figure 7 (the pseudo-code is also provided in Algorithm 4).

Algorithm 4 Batched Collapse algorithm

```

1: procedure BATCHEDCOLLAPSE( $t_{lim}$ )
2:    $D \leftarrow$  The distance matrix, calculated from timestamp
   0 to  $t_{lim}$ 
3:    $P \leftarrow$  The matrix profile obtained by collapsing  $D$ 
4:    $t_{lim\_curr} \leftarrow t_{lim}$ 
5:   while  $t_{lim} < \text{len}(T)$  do
6:      $t_{lim\_prev} \leftarrow t_{lim\_curr}$ 
7:      $t_{lim\_curr} \leftarrow t_{lim\_curr} + t_{lim}$ 
8:      $D_{new} \leftarrow$  Calculate distance matrix from  $t_{lim\_prev}$ -
        $TC \times 2 - 1$  to  $t_{lim\_curr}$ 
9:     Collapse  $D_{new}$  to get the matrix profile  $P_{new}$ 
10:     $P_{merged} \leftarrow$  Merge  $P$  with  $P_{new}$   $\triangleright$ 
      keep the minimum value between  $P$  and  $P_{new}$  where the
      two vectors overlap in time
11:     $P \leftarrow P_{merged}$ 
return  $P$ 

```

As seen in the Batched collapse algorithm, collapsing the matrix profile before processing the entire time series comes at the cost of recalculating the last ($TC \times 2 - 1$) time steps of the matrix profile. This recalculation is necessary as the first time steps in the new matrix profile can point back to the old ones (see Figure 7).

The “batch collapse” algorithm can also be used to make LS-USS ϵ real-time (i.e. an algorithm that requires at least ϵ data points to detect a change-point. A completely online algorithm would then be 1 real-time) by accumulating a batch of data before adding it to the end of the matrix profile. Doing this entails a trade-off between batch size and calculation time as every time a batch is added, $TC \times 2 - 1$ time step must be recalculated.

In section III-B2, FLOSS was made to work online by forcing arcs to only be able to point forwards in time. This trick can also be borrowed to update the LSMP in real-time. If one only looks for the closest distance forward in time, no subsequence can have a nearest-neighbor in the previous LSMP. As such, the requirement of recalculating $TC \times 2 - 1$ time steps at each new batch also disappears (see Figure 8). As with FLOSS, it should be expected that using only forward-pointing arcs will have a negative impact on the algorithm’s performance.

C. Latent Space Unsupervised Semantic Segmentation (LS-USS)

LS-USS uses an autoencoder to encode the multidimensional all subsequences set \mathbf{A} into the one-dimensional latent all subsequence set \mathbf{F} . \mathbf{F} is then used to compute the LSMP \vec{P}_F and the corresponding LSMP index vector \vec{I}_F as described in the previous section. The indices \vec{I}_F are then used to make the CAC, which is the graph that contains the number of arc crossings at each time step. Like in FLUSS, when few arcs are crossings over a particular data point, this indicates a high likelihood of a change-point at that time step, and a high number of arc crossings indicates a low likelihood for a change-point.

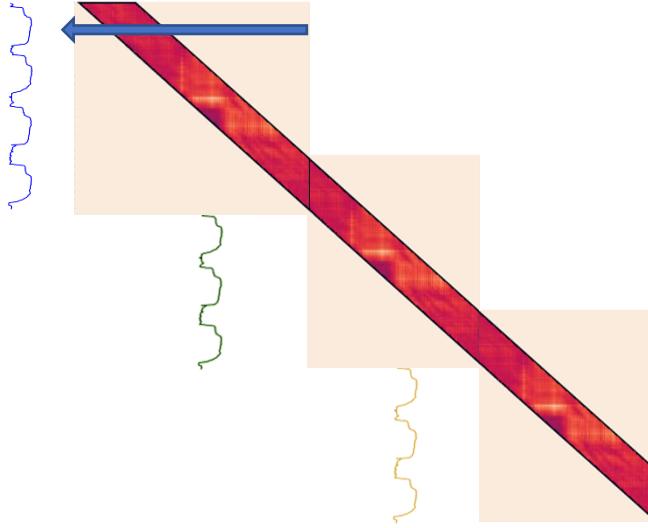


Fig. 8. Online version of LSMP on a time series of length 10400. No recalculation is needed when updating the distance matrix

LS-USS online is similar to LS-USS, except that it uses the version of the LSMP that works online by only considering right-pointing arcs. Doing this makes it possible to update \vec{P}_F without recomputing the distance matrix for the last $(TC \times 2 - 1)$ time steps. As mentioned, the regular LSMP can also be ϵ real-time by accumulating a batch of data before adding it to the end of the \vec{P}_F , making the regular LS-USS ϵ real-time. An overview of the relation between LS-USS, LS-USS online, FLUSS, FLOSS and LFMD is shown in Figure 9.

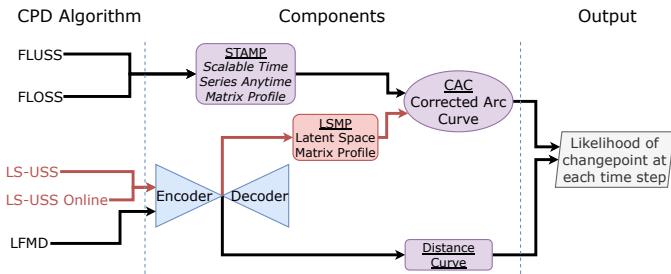


Fig. 9. Overview of the LS-USS and LS-USS online algorithms compared to the other considered algorithms.

D. Local Regime Extraction Algorithm

REA was presented as an algorithm for extracting segments based on the k-lowest "valley" points of the CAC. For most long time series data the change-points are distributed relatively evenly in time. In these cases, when extracting change-points, local minimums are often more meaningful than the global minimums. The REA algorithm will always pick the global k-lowest "valley" points on the CAC instead of locating where the CAC is at its lowest compared to the local area around it. To address this issue, this work introduces the Local Regime Extraction Algorithm (LREA). The method scales each point in the CAC to zero mean and unit variance based

on the local mean and standard deviation calculated over a rolling window (as shown in Equation 4). After the scaling, the same procedure used in the REA algorithm extracts the lowest k "valley points" from the scaled CAC.

$$CAC_{\text{scaled}} = \frac{CAC - \mu_{\text{rolling}}}{\sigma_{\text{rolling}}} \quad (4)$$

Note that by increasing the window size, the extraction of threshold will be increasingly global and thus tend towards REA.

E. Local Threshold Extraction Algorithm

For comparisons between the offline CPD algorithms in this work, the algorithms REA and LREA are used. However, these algorithms require information about the number of segments, which is often unavailable in practice. To address this, this work introduces the Local Threshold Extraction Algorithm (LTEA). As the name suggests, this algorithm works by scaling the CAC before doing threshold-based change-point extraction.

The same CAC scaling used in the LREA algorithm is first applied to LTEA. Note that in the case where the algorithm is applied on an online (streaming) time series, the rolling statistics can naturally be computed only on prior data (which is the main difference with offline dataset).

In LTEA, the scaled CAC-values over a given threshold are disregarded by being set to a value of one. When the scaled CAC is standardized to zero mean and unit variance, a good threshold value was found to be around minus one (one standard deviation).

After thresholding the CAC, the local minimum obtained are identified as the change-point location. An exclusion zone like the one used in LREA and REA is also applied to avoid trivial matches caused by valleys close in time. The pseudo-code to compute LTEA is presented in Algorithm 5 and an example of the application of LTEA is shown in Figure 10.

V. EXPERIMENTS

A. Datasets

To benchmark the performance of LS-USS (and LS-USS Online) against FLUSS, FLOSS and LFMD, two biosignals-based dataset are considered.

1) *UCI Human Activity Recognition*: The UCI Human Activity Recognition Using Smartphones Dataset [43] contains 30 volunteers between the ages of 19 and 48 who perform six activities wearing a smartphone on the waist. The activities are walking, walking up stairs, walking down stairs, sitting, standing, and laying down. The collected data came from the 3-axial accelerometer and 3-axial gyroscope located in a Samsung Galaxy S2 and were sampled at 50Hz. The raw acceleration data signals have three main components: body movement, gravity, and noise. The sensor data was filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz for noise removal. A low pass Butterworth filter, with a corner frequency of 0.3 Hz, is used to separate the body movement and gravity signals. The reason for choosing to use a 0.3 Hz cut-off

Algorithm 5 Local Threshold Extracting Algorithm (LTEA)

Input:

*CAC – a Corrected Arc Curve or Distance Curve
localWindowSize – The size of the window used to normalize the CAC
threshold – Values of the CAC_{scaled} above this threshold will be set to 1. A good default value is to set this to -1.*

Output:

locRegimes – the location of the change-points

```

1: procedure LTEA(CAC, localWindowSize, threshold=-1.)
2:   CACscaled  $\leftarrow$  empty array with same length as the CAC
3:   for i=0:length(CACscaled) do
4:     localWindow  $\leftarrow$  CAC[idx-(localWindowSize) : idx+(localWindowSize)]
5:      $\mu_{rolling} \leftarrow mean(localWindow)$ 
6:      $\sigma_{rolling} \leftarrow std(localWindow)$ 
7:     CACscaled[i]  $\leftarrow (CAC[i] - \mu_{rolling}) / \sigma_{rolling}$ 
8:     if CACscaled[i] > threshold then
9:       CACscaled[i]  $\leftarrow 1$ 
      valleys  $\leftarrow$  List of sequences in CACscaled, where a sequence correspond to all the consecutive points with a value different than 1 in the CACscaled.
10:    for for valley in valleys do
11:      locRegimes[i]  $\leftarrow indexOf(min(valley))$ 
return locRegimes

```

frequency is that gravitational forces are assumed only to have low-frequency components. This work will use the data from the gyroscope, accelerometer, and the filtered body movement from the accelerometer. As all these signals are sampled in all three spatial dimensions, the dataset contains nine channels in total. Nine subjects constitute the training set, five subjects are used as a validation set, while the remaining 16 subjects are picked for the test set. The labeled change-point detection is available in this dataset with the goal of detecting when the participant starts a new activity. This dataset will be referred to as the UCI dataset.

2) *EMG-Based Long-Term 3DC Dataset*: The EMG-Based Long-Term 3DC Dataset [44] contains data from 20 able-bodied participants performing eleven hand gestures while recording their forearm's muscle activity over a period of 21 days (the recording sessions took place every \sim 7 days). The goal for the unsupervised segmentation is to detect when a participant transitions towards a new gesture. This is of particular interest for myoelectric-based control as being able to detect such transitions accurately in real-time would improve the performance of previously proposed self-learning classifiers [45]. The armband used when recording this dataset is the 3DC Armband [46]. The 3DC is a ten-channel, dry electrode 3D printed EMG band with a sampling rate of 1000 Hz per channel. While the armband also features a 9-axis Magnetic, Angular Rate, and Gravity (MARG) sensor, only the 10 EMG channels are considered in this work. The dataset is divided into training sessions and evaluation sessions. In the training session, the participants were asked to hold each of the 11 gestures for 5 seconds. The transitions between gesture

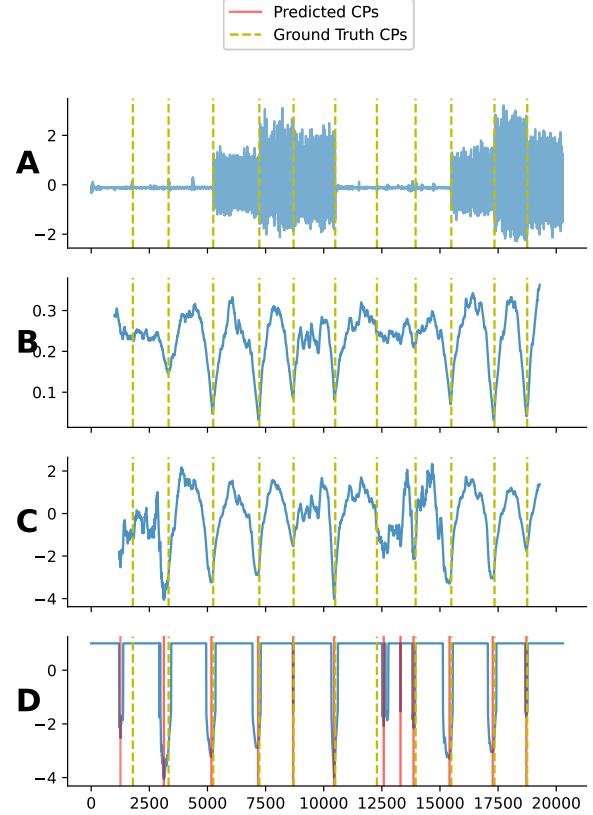


Fig. 10. Plot A shows one channel from subject 4 in the UCI dataset. Plot B shows the regular CAC for subject 4. Plot C shows the scaled CAC. Plot D shows the thresholded CAC and the predicted change-points when using LTEA on this CAC. Note that as LTEA is not provided with the number of change-points in a given time series, it can in reality find more or less change-points than are actually present. In this case, an additional change-point compared to the ground truth was detected at index ca. 13500.

were not recorded, yielding a discontinuous time series. This procedure was repeated four times by each participant on every recording day. For the evaluation session, the participants were randomly asked to hold a total of 42 gestures. The requested gestures were selected at random every 5 seconds (the time series associated with an evaluation session thus lasted 210 seconds). Each participant recorded a minimum of 6 evaluation sessions (twice per recording session). Importantly, for the evaluation session, the transition between each gesture was recorded, yielding a continuous time series.

For this work, two datasets were derived from the Long-Term 3DC Dataset: The EMG Artificial Dataset and the EMG Dataset.

The EMG Artificial Dataset was created from the original training sessions by concatenating these gesture recordings together to form a continuous time series for each participant. The artificial training set consists of training sessions recorded from ten participants. The validation set consists of 15 time series made by gestures from four participants, while the test sets includes 30 time series from eight participants.

The EMG Dataset was created from the evaluation sessions. The evaluation runs of six participants is used as the training

set, the evaluation runs of five participants is used as the validation set, and the evaluation runs for the remaining nine participants is used as the test set. This dataset will be referred to as the EMG Dataset.

B. Hyperparameter Selection

The segmentation algorithms are divided into two categories: online and offline.

Even if they are not fully online, LS-USS, FLUSS and LFMD are also included in the online category as they can be made to work ε real-time when using a temporal constraint. The hyperparameter optimization is done using grid search by leveraging the previously described training and validation sets for the three considered datasets. The step size, used in LFMD, decides the amount of overlap between subsequences. For the segmentation algorithms that uses an autoencoder (LS-USS and LFMD), the fully connected and convolutional model is included in the hyperparameter search. For the offline algorithms the extraction algorithms REA and LREA are included in the hyperparameter search. Four data scaling techniques are applied to data for comparing the segmentation algorithms on the different datasets. These methods use the implementation provided in scikit-learn [47] and are widely used for machine learning and data mining tasks. The scaling is done independently over all channels, and the statistics used for scaling the data are computed using only the training data. The choice of scaling method is also incorporated into the hyperparameter search.

The mean distance between the change-points from the training set is calculated for use as the local window size for scaling the CAC when using the LREA and LTEA extractors. The training set is also utilized for training the autoencoders. As the autoencoders require both training and validation sets, 80% of the training set is used as training examples, while the remaining 20% is used to validate the autoencoder. The validation portion of the dataset is used to find the optimal hyperparameters of the segmentation algorithms. After the best performing model configurations on the validation set are found, the test set is used to make the final comparisons between the different segmentation algorithms.

For a detailed presentation of the considered hyperparameters see Table IV and Table V in the Appendix. Furthermore, Table VI and Table VII in the appendix shows the hyperparameters selected for the offline and online evaluation respectively.

C. Evaluation metrics

The performance of the offline algorithms are evaluated based on the ScoreRegimes from Table 3 in [3]. The definition of the ScoreRegimes is as follows:

$$\text{ScoreRegimes} = \frac{\sum_{i=1}^{N_{GT}} |CP_{pred} - CP_{actual}|}{N_{GT} * n} \quad (5)$$

Where N_{GT} is the number of ground truth change-points, and n is the length of the time series.

Note that the ScoreRegimes requires knowing the number of ground truth change-points and are thus ill-adapted to evaluate online CPD algorithms as they might identify a different

	<u>UCI</u>		
	LS-USS	FLUSS	LFMD
Mean	0.00687	0.00931	0.01580
Friedman Rank	1.33	2.07	2.60
H0 (Adjusted p-value)	-	0 (0.04461)	0 (0.00105)
<u>EMG Artificial</u>			
	LS-USS	FLUSS	LFMD
Mean	0.00214	0.00718	0.01839
Friedman Rank	1.03	1.97	3.00
H0 (Adjusted p-value)	-	0 (0.00030)	0 (<0.00001)
<u>EMG</u>			
	LS-USS	FLUSS	LFMD
Mean	0.00452	0.00593	0.00715
Friedman Rank	1.00	2.00	3.50
H0 (Adjusted p-value)	-	0 (0.01431)	0 (<0.00001)

TABLE II
COMPARISON BETWEEN THE CONSIDERED CPD ALGORITHMS IN THE OFFLINE SETTING.

number of change-points compared to the ground truth. Thus, the online algorithms are evaluated using the Prediction Loss mean absolute error (MAE), which is a slight modification of equation 12 in [33]. The Prediction Loss MAE used in this work is defined as follows:

$$\text{Prediction loss MAE} = \left| 1 - \frac{N_{pred}}{N_{GT}} \right| \times MAE \quad (6)$$

Where N_{pred} is the number of predicted change-points and N_{GT} is the number of ground truth change-points. Thus, instead of relying on pre-defined change-points, the Prediction loss MAE weights the mean absolute error with the prediction ratio. Importantly, as the metrics used for the offline and online case are not related, it is also not possible to make comparisons between the online and offline algorithms' performance.

As suggested in [48], a two-step statistical procedure is applied to compare LS-USS against the relevant CPD algorithms. First, Friedman's test ranks the algorithms amongst each other. Then, Holm's post-hoc test is applied using the best ranked method as a comparison basis. The null hypothesis of the post hoc test is that the performance of the two models is the same. The null hypothesis is rejected when $p < 0.05$.

VI. RESULTS

The comparisons in this section are made on three datasets in both the offline and online setting: UCI, EMG artificial, and EMG.

A. Offline

In this subsection, for each dataset and participant, the full time series was available when performing unsupervised segmentation and the number of change-points in each time series was also known. Table VI-A presents the results obtained in the offline setting.

B. Online

In this subsection, data was fed to the CPD algorithms as if they were acquired in real-time. Further, the total number of change-points in a time series was not given. Table VI-B presents the results obtained in the online setting.

VII. DISCUSSION

1) *Observations from the hyperparameter search:* Table VI in the appendix shows that LREA is chosen seven out of nine times for the three algorithms that are being compared on the three different datasets. This indicates that scaling the CAC based on the local statistics is useful when dealing with longer time series data. Another interesting finding from the hyperparameter selection is that the autoencoder selected by the hyperparameter search for the LS-USS models (both offline and online) varies between the datasets; some use the less complex, fully connected model, while others use the convolutional model. As alluded to in [33], using a smaller model with only two hidden layers might be beneficial for feature representation as it could lead to a more general way to represent the data. The drawback of simpler autoencoder models is that one risks losing some of the information needed for segment differentiation. Note that an in-depth analysis of the impact of the autoencoder's architecture was outside the scope of the current work and will be the focus of future works.

2) *Offline:* As shown in Table VI-A, in the offline setting LS-USS is shown to systematically and significantly outperform the other segmentation algorithms for all three considered datasets. This indicates that there is a clear advantage in terms of overall performance of the CDP algorithm in learning a representation which can consider the multidimensional representation of the data simultaneously compared to using FLUSS directly. Further, the fact that LFMD was the worst performing CDP algorithm illustrates the usefulness of leveraging the latent space matrix profile to segment the time serie.

3) *Online:* The highest-ranked algorithm on the UCI dataset is LS-USS, but the difference with the other algorithms is not significant. ε -real time LS-USS is also the best-ranked algorithm on both the EMG and EMG Artificial datasets and the difference in performance between it and the other algorithms is also significant (except for LS-USS online on the EMG Artificial dataset and LS-USS online and FLOSS on the EMG dataset). Overall, the ε -real time LS-USS LTEA model seems to be the most consistent performing online segmentation algorithm across all the datasets.

A. Limitations

The segmentation outcomes for each of the algorithms are dependent on the hyperparameter search. Therefore, it would be beneficial to do a more in-depth hyperparameter search that includes sampling the hyperparameters from predefined distributions instead of a simple grid search. Further, a more in-depth evaluation of the hyperparameter selection sensitivity will be conducted in future works.

An additional limitation of LS-USS compared to FLUSS is that despite the improvement observed in terms of performance, it does come at the cost of an increase in the hyperparameters due to the use of an autoencoder.

The experiments conducted in this work have mainly been on multidimensional data where the channels are from similar sensors, similar sampling rates, and numerical data only. Therefore, how the algorithm performs when considering time series data containing channels from a broader range of sources with both numerical and categorical variables remains to be evaluated. Furthermore, doing an in-depth analysis on how the number of channels affects the performance of LS-USS will also be considered in future works.

VIII. CONCLUSION

The main contribution of this work is the segmentation algorithms LS-USS and LS-USS online. Through extensive testing conducted on both artificial and real-world datasets from various domains and sensors, it was found that LS-USS generally delivers on par or better segmentation scores compared to other state-of-the-art algorithms such as FLUSS/FLOSS and LFMD. The LS-USS algorithms have many desirable properties. They can be implemented both online and in an “anytime” fashion. They are domain agnostic (beyond knowing the order of time scale considered for change-points) and do not need extensive tuning of hyperparameters to achieve state-of-the-art performance. Further, they do not make any statistical assumptions about the input data. The LSMP component used in LS-USS also shows that it is possible to calculate a temporarily constrained matrix profile from feature vectors by exploiting highly parallelized hardware. The same methods used for constructing the LSMP can be used on any vector-based Representation Learning algorithms [49], which represents an excellent potential direction for future research.

The extraction algorithms LREA and LTEA presented in this work also show potential. LREA usually outperformed the more global REA algorithm, which shows that scaling the CAC based on the local statistics is highly useful, especially for long time series data. The online extraction algorithm LTEA uses the same CAC scaling as LREA but uses a threshold to identify the change-points instead of extracting the n lowest “valley” points. Consequently, in contrast to REA and LREA, LTEA does not need any information on the number of change-points to extract, making it applicable to a wider array of real-world segmentation problems.

REFERENCES

- [1] E. Campbell, A. Phinyomark, and E. Scheme, “Current trends and confounding factors in myoelectric control: Limb position and contraction intensity,” *Sensors*, vol. 20, no. 6, p. 1613, 2020.
- [2] S. Aminikhahgahi and D. J. Cook, “A survey of methods for time series change point detection,” *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.
- [3] S. Gharghabi, Y. Ding, C.-C. M. Yeh, K. Kamgar, L. Ulanova, and E. Keogh, “Matrix profile viii: domain agnostic online semantic segmentation at superhuman

	UCI				
	LS-USS	LS-USS Online	FLUSS	FLOSS	LFMD
Mean	45.54	114.82	47.27	87.36	51.99
Friedman Rank	2.50	4.20	2.56	3.12	2.60
H0 (Adjusted p-value)	-	0 (0.01294)	1	1	1
	EMG Artificial				
	LS-USS	LS-USS Online	FLUSS	FLOSS	LFMD
Mean	6.64	16.10	26.73	86.44	97.54
Friedman Rank	1.82	2.28	2.80	3.88	4.22
H0 (Adjusted p-value)	-	1	0 (0.03202)	0 (<0.00001)	0 (<0.00001)
	EMG				
	LS-USS	LS-USS Online	FLUSS	FLOSS	LFMD
Mean	121.97	164.68	251.41	163.71	388.82
Friedman Rank	1.50	2.50	3.70	2.55	4.75
H0 (Adjusted p-value)	-	1	0 (0.00003)	1	0 (<0.00001)

TABLE III
COMPARISON BETWEEN THE CONSIDERED CPD ALGORITHMS IN THE ONLINE SETTING.

- performance levels,” in *2017 IEEE international conference on data mining (ICDM)*. IEEE, 2017, pp. 117–126.
- [4] J. F.-S. Lin, M. Karg, and D. Kulić, “Movement primitive segmentation for human motion modeling: A framework for analysis,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 325–339, 2016.
- [5] C. Truong, L. Oudre, and N. Vayatis, “Selective review of offline change point detection methods,” *Signal Processing*, vol. 167, p. 107299, 2020.
- [6] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, “Using mobile phones to determine transportation modes,” *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 13:1–13:27, Mar. 2010. [Online]. Available: <https://doi.org/10.1145/1689239.1689243>
- [7] I. Cleland, M. Han, C. Nugent, H. Lee, S. McClean, S. Zhang, and S. Lee, “Evaluation of Prompted Annotation of Activity Data Recorded from a Smart Phone,” *Sensors*, vol. 14, no. 9, pp. 15 861–15 879, Sep. 2014, number: 9 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/1424-8220/14/9/15861>
- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [9] L. I. Kuncheva, “Change Detection in Streaming Multivariate Data Using Likelihood Detectors,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 5, pp. 1175–1180, May 2013, conference Name: IEEE Transactions on Knowledge and Data Engineering.
- [10] R. P. Adams and D. J. C. MacKay, “Bayesian Online Changepoint Detection,” *arXiv:0710.3742 [stat]*, Oct. 2007, arXiv: 0710.3742. [Online]. Available: <http://arxiv.org/abs/0710.3742>
- [11] R. Malladi, G. P. Kalamangalam, and B. Aazhang, “Online Bayesian change point detection algorithms for segmentation of epileptic activity,” in *2013 Asilomar Conference on Signals, Systems and Computers*, Nov. 2013, pp. 1833–1837, iSSN: 1058-6393.
- [12] Z. Harchaoui, E. Moulines, and F. Bach, “Kernel Change-point Analysis,” in *Advances in Neural Information Processing Systems*, vol. 21. Curran Associates, Inc., 2008. [Online]. Available: <https://proceedings.neurips.cc/paper/2008/hash/08b255a5d42b89b0585260b6f2360bdd-Abstract.html>
- [13] P. R. Rosenbaum, “An exact distribution-free test comparing two multivariate distributions based on adjacency,” *Journal of the Royal Statistical Society B*, p. 2005.
- [14] J. H. Friedman and L. C. Rafsky, “Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests,” *The Annals of Statistics*, pp. 697–717, 1979.
- [15] N. Itoh and J. Kurths, “Change-Point Detection of Climate Time Series by Nonparametric Method,” p. 4, 2010.
- [16] Y. Kawahara, T. Yairi, and K. Machida, “Change-Point Detection in Time-Series Data Based on Subspace Identification,” in *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, Oct. 2007, pp. 559–564, iSSN: 2374-8486.
- [17] K. Yamanishi and J.-i. Takeuchi, “A unifying framework for detecting outliers and change points from non-stationary time series data,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD ’02. New York, NY, USA: Association for Computing Machinery, Jul. 2002, pp. 676–681. [Online]. Available: <https://doi.org/10.1145/775047.775148>
- [18] T. Rakthanmanon, E. J. Keogh, S. Lonardi, and S. Evans, “Time series epenthesis: Clustering time series streams requires ignoring some data,” p. 10.
- [19] P. Yang, G. Dumont, and J. M. Ansermino, “Adaptive change detection in heart rate trend monitoring in anesthetized children,” *IEEE transactions on bio-medical engineering*, vol. 53, no. 11, pp. 2211–2219, Nov. 2006.
- [20] C. Nentwich and G. Reinhart, “A Combined Anomaly and Trend Detection System for Industrial Robot Gear Condition Monitoring,” *Applied Sciences*, vol. 11, p. 10403, Nov. 2021.
- [21] T. Yairi, Y. Kato, and K. Hori, “Fault detection by mining association rules from housekeeping data.”
- [22] T.-C. Fu, F.-L. Chung, V. Ng, and R. Luk, “Pattern discovery from stock time series using self-organizing maps.”
- [23] C.-S. Li, P. S. Yu, and V. Castelli, “MALM: a

- framework for mining sequence database at multiple abstraction levels,” in *Proceedings of the seventh international conference on Information and knowledge management*, ser. CIKM ’98. Association for Computing Machinery, pp. 267–272. [Online]. Available: <https://doi.org/10.1145/288627.288666>
- [24] E. Keogh and J. Lin, “Clustering of time series subsequences is meaningless: Implications for previous and future research,” p. 20.
- [25] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover, “Exact discovery of time series motifs,” vol. 2009, pp. 473–484.
- [26] T. L. Bailey, “STREME: accurate and versatile sequence motif discovery,” vol. 37, no. 18, pp. 2834–2840. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btab203>
- [27] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, “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)*. Ieee, 2016, pp. 1317–1322.
- [28] C. Alippi, G. Boracchi, D. Carrera, and M. Roveri, “Change detection in multivariate datastreams: Likelihood and detectability loss,” p. 7.
- [29] T. Gu, S. Chen, X. Tao, and J. Lu, “An unsupervised approach to activity recognition and segmentation based on object-use fingerprints,” vol. 69, no. 6, pp. 533–544. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169023X10000133>
- [30] Y. Kawahara and M. Sugiyama, “Sequential changepoint detection based on direct density-ratio estimation,” vol. 5, no. 2, pp. 114–127. [Online]. Available: <https://doi.org/10.1002/sam.10124>
- [31] A. Qahtan, S. Wang, B. Alharbi, and X. Zhang, “A PCA-Based Change Detection Framework for Multidimensional Data Streams,” p. 10.
- [32] H. Kim, H. K. Kim, M. Kim, J. Park, S. Cho, K. B. Im, and C. R. Ryu, “Representation learning for unsupervised heterogeneous multivariate time series segmentation and its application,” *Computers & Industrial Engineering*, vol. 130, pp. 272–281, Apr. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360835219301159>
- [33] W.-H. Lee, J. Ortiz, B. Ko, and R. Lee, “Time series segmentation through automatic feature learning,” *arXiv preprint arXiv:1801.05394*, 2018.
- [34] M. Sakurada and T. Yairi, “Anomaly detection using autoencoders with nonlinear dimensionality reduction,” in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, ser. MLSDA’14. Association for Computing Machinery, 2014, pp. 4–11. [Online]. Available: <https://doi.org/10.1145/2689746.2689747>
- [35] C. Zhou and R. C. Paffenroth, “Anomaly detection with robust deep autoencoders,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’17. Association for Computing Machinery, 2017, pp. 665–674. [Online]. Available: <https://doi.org/10.1145/3097983.3098052>
- [36] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, “A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 1409–1416.
- [37] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [38] M. Tschannen, O. Bachem, and M. Lucic, “Recent advances in autoencoder-based representation learning,” *arXiv preprint arXiv:1812.05069*, 2018.
- [39] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, “The ucr time series archive,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293–1305, 2019.
- [40] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, M. B. Westover, Q. Zhu, J. Zakaria, and E. Keogh, *Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping*, Aug. 2012, vol. 2012, journal Abbreviation: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Publication Title: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [41] O. Banos, M. A. Toth, M. Damas, H. Pomares, and I. Rojas, “Dealing with the effects of sensor displacement in wearable activity recognition,” *Sensors*, vol. 14, no. 6, pp. 9995–10 023, 2014.
- [42] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [43] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in *Esann*, vol. 3, 2013, p. 6.
- [44] U. Côté-Allard, G. Gagnon-Turcotte, A. Phinyomark, K. Glette, E. Scheme, F. Laviolette, and B. Gosselin, “A transferable adaptive domain adversarial neural network for virtual reality augmented emg-based gesture recognition,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 546–555, 2021.
- [45] U. Côté-Allard, G. Gagnon-Turcotte, A. Phinyomark, K. Glette, E. J. Scheme, F. Laviolette, and B. Gosselin, “Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition,” *IEEE Access*, vol. 8, pp. 177 941–177 955, 2020.
- [46] U. Côté-Allard, G. Gagnon-Turcotte, F. Laviolette, and B. Gosselin, “A low-cost, wireless, 3-d-printed custom armband for semg hand gesture recognition,” *Sensors*, vol. 19, no. 12, p. 2811, 2019.
- [47] scikit-learn: machine learning in python — scikit-learn 0.23.2 documentation. [Online]. Available: <https://scikit-learn.org/stable/>

- //scikit-learn.org/stable/
- [48] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
 - [49] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

APPENDIX A

HYPERPARAMETERS SEARCH AND SELECTION

Table IV and Table V shows the hyperparameters considered for the offline algorithm and online algorithms respectively.

Offline Models	Data Scalers	NW		TC		Step-Size	Autoencoder	Extraction Algorithm
LS-USS	All Datasets	All datasets		EMG Datasets	UCI	All datasets	All Datasets	All Datasets
	No Scaler	50	300	1000	800	-	Fully Connected	REA
	Standard Scaler	100	400	1500	1200	-	Convolutional	LREA
	Robust Scaler	150	500	2000	1600	-	-	-
FLUSS	Min Max Scaler	200	-	-	-	-	-	-
	No Scaler	50	300	1000	800	-	-	REA
	Standard Scaler	100	400	1500	1200	-	-	LREA
	Robust Scaler	150	500	2000	1600	-	-	-
LFMD	Min Max Scaler	200	-	-	-	-	-	-
	No Scaler	50	300	1000	800	25	200	REA
	Standard Scaler	100	400	1500	1200	50	250	LREA
	Robust Scaler	150	500	2000	1600	100	350	-
	Min Max Scaler	200	-	-	-	150	500	-

TABLE IV

HYPERPARAMETERS CONSIDERED FOR THE CPD ALGORITHMS IN THE OFFLINE SETTING. NW IS THE SUB-SEQUENCE LENGTH. TC IS THE TEMPORAL CONSTRAINT.

Online Models	Data Scalers	NW		TC		Step-size	Autoencoder	NW	Extraction Algorithm
LS-USS (ϵ -real time)	All Datasets	All datasets		EMG datasets	UCI	All datasets	All Datasets	All datasets	All Datasets
	No Scaler	50	300	1000	800	-	Fully Connected	-0.5	-2.0
	Standard Scaler	100	400	1500	1200	-	Convolutional	-1.0	-2.5
	Robust Scaler	150	500	2000	1600	-	-	-1.5	-3.0
LS-USS - Online	Min Max Scaler	200	-	-	-	-	-	-	-
	No Scaler	50	300	1000	800	-	Fully Connected	-0.5	-2.0
	Standard Scaler	100	400	1500	1200	-	Convolutional	-1.0	-2.5
	Robust Scaler	150	500	2000	1600	-	-	-1.5	-3.0
FLUSS (ϵ -real time)	Min Max Scaler	200	-	-	-	-	-	-	-
	No Scaler	50	300	1000	800	-	-	-0.5	-2.0
	Standard Scaler	100	400	1500	1200	-	-	-1.0	-2.5
	Robust Scaler	150	500	2000	1600	-	-	-1.5	-3.0
FLOSS	Min Max Scaler	200	-	-	-	-	-	-	-
	No Scaler	50	300	1000	800	-	-	-0.5	-2.0
	Standard Scaler	100	400	1500	1200	-	-	-1.0	-2.5
	Robust Scaler	150	500	2000	1600	-	-	-1.5	-3.0
LFMD	Min Max Scaler	200	-	-	-	-	-	-	-
	No Scaler	50	300	1000	800	25	200	-0.5	-2.0
	Standard Scaler	100	400	1500	1200	50	250	Fully Connected	-1.0
	Robust Scaler	150	500	2000	1600	100	350	Convolutional	-1.5
	Min Max Scaler	200	-	-	-	150	500	-	-3.0

TABLE V

HYPERPARAMETERS CONSIDERED FOR THE CPD ALGORITHMS IN THE ONLINE SETTING. NW IS THE SUB-SEQUENCE LENGTH. TC IS THE TEMPORAL CONSTRAINT. THE THRESHOLD PARAMETER IS THE THRESHOLD USED IN THE LTEA EXTRACTION ALGORITHM.

Tables VI and VII shows the hyperparameters selected for the offline and online algorithms after doing hyperparameter search. The selected parameters is the ones used for the comparisons between the different models in Section VI and VII.

<u>UCI</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	LREA	Robust Scaler	50	-	250	Fully Connected
LS-USS	REA	Min Max Scaler	50	800	1	Fully Connected
FLUSS	LREA	Robust Scaler	50	1200	1	-
<u>EMG Artificial</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	REA	Standard Scaler	200	-	500	Conv. Model
LS-USS	LREA	Standard Scaler	50	1500	1	Conv. Model
FLUSS	LREA	Standard Scaler	50	1500	1	-
<u>EMG</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	LREA	Standard Scaler	50	-	100	Fully Connected
LS-USS	LREA	Robust Scaler	50	2000	1	Conv. Model
FLUSS	LREA	Standard Scaler	50	1500	1	-

TABLE VI

HYPERPARAMETERS SELECTED FOR THE COMPARED CPD ALGORITHMS IN THE OFFLINE SETTING. NW IS THE SUB-SEQUENCE LENGTH. TC IS THE TEMPORAL CONSTRAINT.

<u>UCI</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	LTEA	Min Max Scaler	100	-	100	Fully Connected
FLUSS	LTEA	Robust Scaler	50	1600	1	-
FLOSS	LTEA	Robust Scaler	150	1600	1	-
LS-USS	LTEA	Min Max Scaler	150	800	1	Fully Connected
LS-USS Online	LTEA	Robust Scaler	50	1200	1	Fully Connected
<u>EMG Artificial</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	LTEA	Standard Scaler	300	-	100	Fully Connected
FLUSS	LTEA	Standard Scaler	50	1500	1	-
FLOSS	LTEA	Standard Scaler	100	2000	1	-
LS-USS	LTEA	Standard Scaler	50	2000	1	Conv. Model
LS-USS Online	LTEA	Standard Scaler	50	1500	1	Conv. Model
<u>EMG</u>						
CPD Algorithm	Extraction Algorithm	Data Scaler	NW	TC	Step-size	Autoencoder
LFMD	LTEA	Standard Scaler	500	-	500	Fully Connected
FLUSS	LTEA	Standard Scaler	50	1000	1	-
FLOSS	LTEA	Standard Scaler	200	1000	1	-
LS-USS	LTEA	Standard Scaler	300	1000	1	Fully Connected
LS-USS Online	LTEA	Standard Scaler	200	1500	1	Fully Connected

TABLE VII

HYPERPARAMETERS SELECTED FOR THE COMPARED CPD ALGORITHMS IN THE ONLINE SETTING. NW IS THE SUB-SEQUENCE LENGTH. TC IS THE TEMPORAL CONSTRAINT. THE THRESHOLD PARAMETER IS THE THRESHOLD USED IN THE LTEA EXTRACTION ALGORITHM.