

An Evaluation of Classification Methods for 3D Printing Time-Series Data^{*}

Vivek Mahato^{*} Muhannad Ahmed Obeidi^{**}
Dermot Brabazon^{**} Pádraig Cunningham^{*}

^{*} *I-Form SFI Research Centre for Advanced Manufacturing, University
College Dublin, Ireland (e-mail: author@i-form.ie).*

^{**} *I-Form SFI Research Centre for Advanced Manufacturing, Dublin
City University, Ireland (e-mail: author@i-form.ie).*

Abstract: Additive Manufacturing presents a great application area for Machine Learning because of the vast volume of data generated and the potential to mine this data to control outcomes. In this paper we present preliminary work on classifying infrared time-series data representing melt-pool temperature in a metal 3D printing process. Our ultimate objective is to use this data to predict process outcomes (e.g. hardness, porosity, surface roughness). In the work presented here we simply show that there is a signal in this data that can be used for the classification of different components and stages of the AM process. In line with other Machine Learning research on time-series classification we use k -Nearest Neighbour classifiers. The results we present suggests that Dynamic Time Warping is an effective distance measure compared with alternatives for 3D printing data of this type.

Keywords: Classification, Process Control, Time-Series.

1. INTRODUCTION

It is generally accepted that one of the big challenges with Additive Manufacturing (AM) is that the processing parameters are difficult to tune (Qi et al., 2019). On the positive side, AM produces vast amounts of data that can be mined to address this problem. A strategy for using Supervised Machine Learning in AM is shown in Fig. 1. The objective is to analyse data gathered through sensors during the process and analyse that data to predict process outcomes. In the context of AM, these outcomes might relate to product quality, for instance hardness, surface roughness or porosity.

It is in the nature of AM that the data coming from the sensors is likely to be in the form of a time-series. The time-series data might be aggregated into feature vector representations so that standard Machine Learning (ML) methods can be applied; however one of our objectives is to work with the raw time-series data coming from the process.

Our overall objective is to use pyrometers to measure melt-pool temperature and mine this data to predict process outcomes. The hypothesis is that anomalies such as pores will have characteristic signatures in the temperature time-series. Our first step in this direction is to find out how much signal there is in the pyrometer data.

^{*} This publication has resulted from research supported in part by a grant from Science Foundation Ireland (SFI) under Grant Number 16/RC/3872 and is co-funded under the European Regional Development Fund.

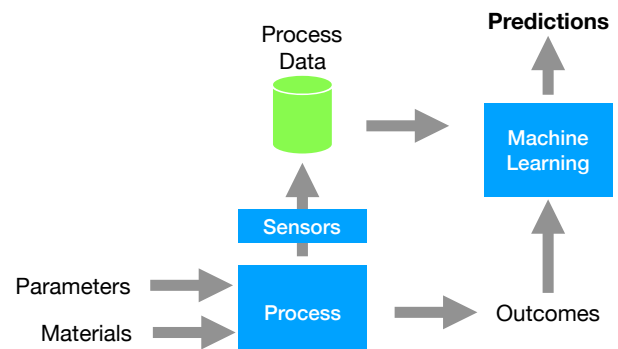


Fig. 1. Supervised Machine Learning in AM.

We have data from the printing of 5mm sample cubes printed in batches of 27. We consider two different ML tasks on this data:

- Can we distinguish different cubes based on the temperature time-series?
- Can we distinguish different layers in the print process?

The details about the data and the specifics of these classification tasks are presented in the next section. The time-series classification methods we evaluate are described in section 3. The results of the evaluation are presented in section 4 and our next steps are outlined in section 5.

2. CONTEXT

The analysis presented here uses data from the Aconity MINI 3D printer.¹ This is a powder-bed fusion (PBF) printer using selective laser melting (SLM). While AM using PBF has great potential, the process is very sensitive to input parameters and there is a lot that can go wrong during a build. Common problems are voids, pores and lack of fusion due to spatter or under or over melting in the meltpool (Özel et al., 2018).

PBF operates in an inert gas environment and the flow of this gas can have an impact on cooling, particularly for high gas flow rates or when a turbulent flow occurs. Excessive amounts of powder may be used within the process if it is not well controlled. There is a strong incentive to recycle powder but potential problems due to recycled powder quality can prevent this (Gorji et al., 2020). Degradation of the powder feedstock due to repeated recycling may negatively impact on the mechanical properties of components built using this powder.

Our objective is to be able to identify these problems as they arise during the build process. In this paper we assess the potential to do this by analysing pyrometer data that tracks the temperature of the melt pool.

2.1 3D Printing Data

The temperature data is recorded by means of two pyrometers from KLEIBER Infrared GmbH. The two pyrometers detect the heat emission light in the range of 1500 to 1700nm via two detectors. The measured light which is reflected from the meltpool area, is split into two paths by means of optical filters and transmitted to the pyrometers via optical fibre cables. In the Aconity 3D printer used in the work, the scanner and the pyrometers are configured to cover x and y values (for each layer) in the range of 0 to 32768 bit covering an area of 400×400 mm. This results in a calibration value of 81.92 bit/mm. Taking in account the pyrometer frequency of 100 kHz, this will produce a response of one single measurement of the meltpool temperature in every $10\mu\text{s}$; e.g. one measurement in each ten microns in the x and y directions based on a scanning speed of 1000 mm/s.

The data analysed here comes from a build of 27×5 mm blocks built with a layer thickness of $20\mu\text{m}$ resulting in 250 layers in the build (see Fig. 2). The temperature data for each layer (all 27 blocks) contains approximately 700,000 data points. The laser scan for each block comprises the perimeter scan shown in Fig. 2 and a raster scan to fill in the interior of the block. A sample temperature time-series for a perimeter scan for Block 0 is shown in Fig. 3. These values are the emissivity measured in (mV) which are directly proportional to the melting temperature.

Because these temperature readings are evidently noisy, we also consider a smoothed version of the time-series. The same time-series passed through a Butterworth filter is shown in Fig. 4.²

¹ <https://aconity3d.com/products/aconitymini/>

² Data will be available for download through the Time Series Classification data repository <https://timeseriesclassification.com>.

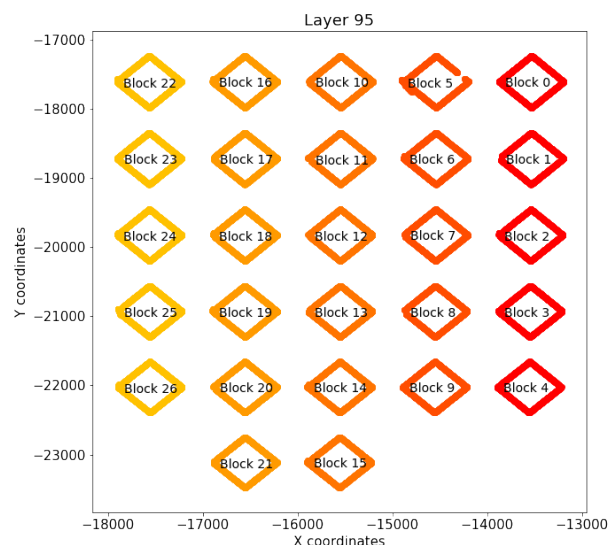


Fig. 2. Perimeter scans from a sample layer of the 27 block build.

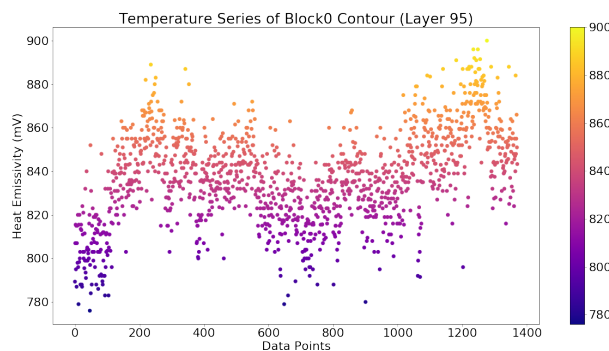


Fig. 3. The temperature time-series of a typical Block 0 perimeter.

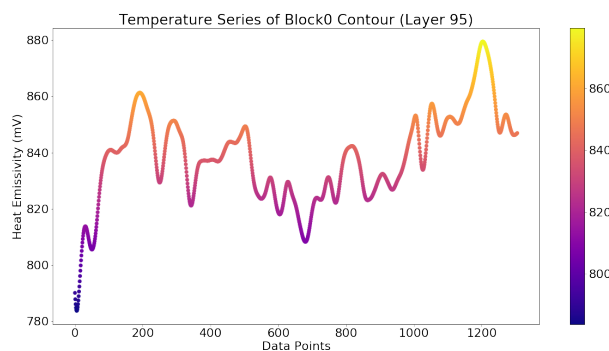


Fig. 4. The time-series in Fig. 3 after passing through a Butterworth filter.

2.2 The Classification Tasks

The Aconity printer uses Argon gas to maintain an inert environment. The gas flows right to left across the build plate as shown in Fig. 2. For the first classification tasks we consider the perimeter scan of a block as a temperature time-series. We consider:

- Block 0 versus Block 22 (easy),
- Block 0 versus Block 1 (hard),
- Block 1 versus Block 22 (easy).

Each block is represented by 250 samples, one for each layer. The parameter tuning is done by 6-fold cross-validation over the bottom 212 layers, and then the model is evaluated over the top 38 layers that have been held back from model selection and training.

The second classification task is to distinguish between the top 10 and bottom 10 layers. With 27 blocks there are 270 samples in each class. The evaluation is done by holding back data from 8 blocks for testing.

3. MACHINE LEARNING FOR TIME-SERIES

We are dealing with time-series data of specific characteristics. When it comes to time-series classification k -Nearest Neighbor (k -NN) has a special status, which may be because some popular ML methods such as Decision Trees and Support Vector Machines will not work with time-series data. There is a presence of several similarity/distance methods that developed to applied with k -NN to temporal data, which makes it more efficient. We examine three such methods in our evaluation:

- Dynamic Time Warping (DTW),
- Symbolic Aggregate approxIMATION (SAX),
- Symbolic Fourier Approximation (SFA).

These methods are described in the subsections that follow, and then we present the results.

3.1 Dynamic Time Warping

Euclidean distance is a popular metric for assessing similarity between feature value representations. Euclidean distance will work well with time-series data if the time-series are well aligned (see Figure 5(a)). However, a small misalignment will result in a large Euclidean distance (Figure 5(b)). Nevertheless, Euclidean distance is included as a baseline in our experiments. As the name suggests, DTW attempts to address this misalignment by allowing more flexible mapping in the time dimension ((Figure 5(c)). The DTW distance is defined as follows:

$$DTW(\mathbf{q}, \mathbf{x}) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} d(q_i, x_j)^2} \quad (1)$$

where $\pi = [\pi_1, \dots, \pi_l, \dots, \pi_L]$ is the optimum path (mapping) having the following properties:

- $m = |\mathbf{q}|, n = |\mathbf{x}|$
- $\pi_1 = (1, 1), \pi_L = (m, n)$
- $\pi_{l+1} - \pi_l \in \{(1, 0), (0, 1), (1, 1)\}$

A cost matrix is constructed by DTW, where each cell (i, j) contains the distance between q_i and x_j . The overall distance is the sum of distances taken by the shortest path through the grid. The extent of the deviation from the main diagonal reflects the *warping*. The computational complexity of DTW is $O(n, m)$ because it entails a search through the matrix. This complexity is effectively $O(n^2)$ in the length of the time-series – so DTW is computationally expensive. To improve the performance of DTW, and reduce its time and memory complexity, we utilise the Sakoe-Chiba (1978) global constraint in the model.

3.2 Symbolic Aggregate approximation

In the past few decades, there was much research around developing symbolic representations of time-series data. The idea is to harness the power of text processing algorithms to solve time-series tasks. A summary of such methods is provided by Lin et al. (2003).

Symbolic Aggregate Approximation (SAX) is one such algorithm that converts a time-series into a series of symbols and obtains dimensionality and numerosity reduction (i.e. more compact representation) of the original time-series. These transformations present a distance measure that is lower bounding on corresponding measures on the original series (Lin et al., 2003).

Piecewise Aggregate Approximation SAX employs Piecewise Aggregate Approximation (PAA) for dimensionality reduction of the original time-series. PAA achieves this by slicing the time-series into bins of equal sizes. The series can then be represented by the mean values in these bins (the PAA coefficients).

Consider a time series S of length n . PAA reduces the series S from size n to m , where $m \leq n$. It achieves this reduction by transforming S into a vector $\bar{S} = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_m)$, where each of s_i is computed as follows:

$$\bar{s}_i = \frac{m}{n} \sum_{j=\frac{n}{m}(i-1)+1}^{\frac{n}{m}i} s_j \quad (2)$$

With PAA there are two exceptional situations worth noting (Keogh and Pazzani, 2000):

- $m = n$, the transformation is similar to the original time-series.
- $m = 1$, the transformation is the mean of the original time-series.

Due to the difficulty in comparing two time-series of different scales, SAX normalizes the original series so that the mean is zero and standard deviation is one, before passing to PAA for transformation (Lin et al., 2003; Keogh and Kasetty, 2002).

SAX passes the PAA transformed series through another discretisation procedure that converts them into symbols. SAX achieves this conversion by discretising the levels into a bins of approximately equal size. These discretised levels typically follow a Gaussian distribution, so these bins get bigger further from the mean. The discretised bins are separated by breakpoints that forms a sorted list $B = \beta_1, \dots, \beta_{a-1}$, in a way that the area under a $N(0, 1)$ Gaussian curve from β_i to $\beta_{i+1} = \frac{1}{a}$, where β_0 and β_a are $-\infty$ and ∞ respectively (Lin et al., 2003).

SAX is provided with a pool of symbols $S = (s_1, s_2, \dots, s_m)$, where m is the size of the pool. After computing all the breakpoints, SAX performs the symbolic transformation as follows. First, the algorithm transforms the original series into a series of PAA coefficients. We take the smallest breakpoint β_1 first and map all the PAA coefficients that are less than β_1 to the symbol s_1 . Then we take β_2 (the second smallest) breakpoint, and all the PAA coefficients that fall between β_1 and β_2 gets mapped to s_2 , and so

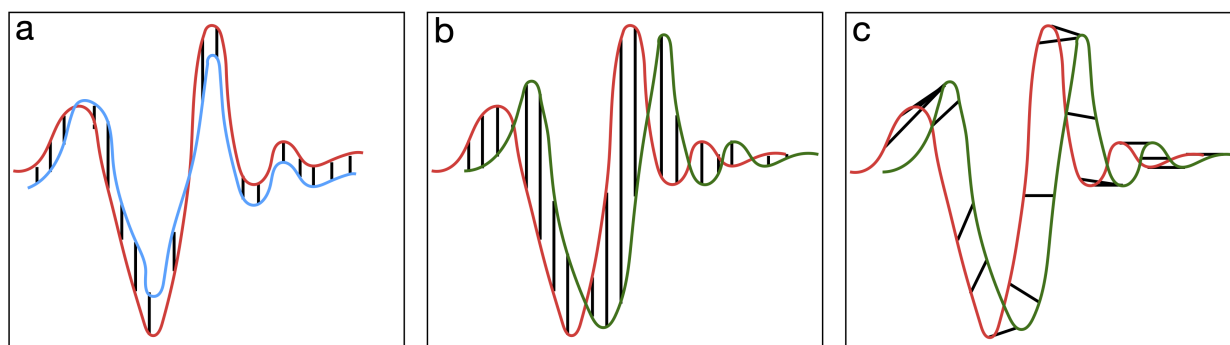


Fig. 5. Dynamic Time Warping; (a) Two similar time-series. (b) Two similar time-series displaced in time – Euclidean distance is large. (c) An example DTW mapping for the two time-series in (b).

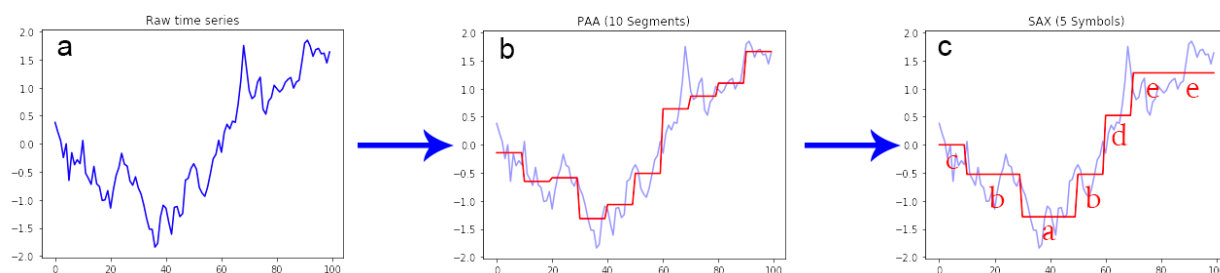


Fig. 6. Symbolic Aggregate Approximation; The raw time-series in (a) will be represented by the sequence **cbabdee** in (c) Mahato et al. (2018).

on until all the PAA coefficients get mapped to its corresponding symbol.

The SAX algorithm includes a module that uses a sliding window technique with an adjustable size. The idea is to extract all the symbols inside the window and concatenate them to create a SAX word. The sliding window then shifts to its right, and extract the symbols in this new window to create another SAX word, and this goes on until the sliding window hits the last frame to create the last word. This collection of words also known as a “bag-of-words”, represents the original time-series.

After the transformation of all the time-series data in our dataset, we can easily calculate the distance between any two given time-series by using any string metric on their symbolic representation. Levenshtein distance (Yujian and Bo, 2007) is one such example of a string metric that is popular.

3.3 Symbolic Fourier Approximation

SFA (Schäfer and Höggqvist, 2012) is another example of an algorithm built on the idea of dimensionality reduction by symbolic transformation. SAX tries to keep the data in the time-domain, whereas SFA transforms the data to bring it in the frequency domain because, in the frequency domain, each dimension has approximate information of the entire series. One can also enhance the overall quality of the estimate by increasing the dimensions. Working with the time-domain requires deciding the length of approximation in advance, and a prefix of this length factors just a subset of the time-series (Schäfer and Höggqvist, 2012).

Discrete Fourier Transform SFA employs Discrete Fourier Transformation (DFT) as its dimensionality reduction

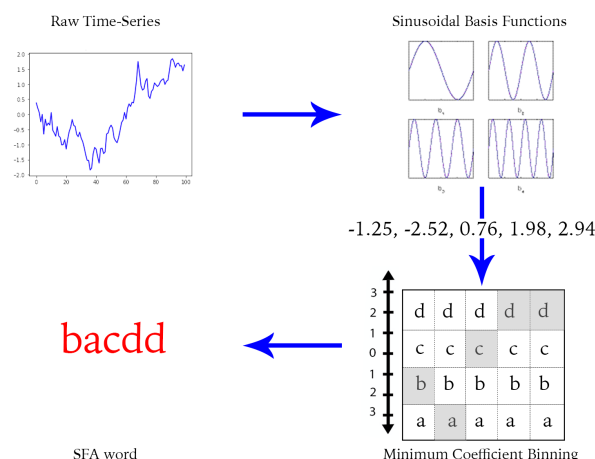


Fig. 7. Symbolic Fourier Approximation; The raw time-series will be represented by the sequence **bacdd** (Mahato et al., 2018).

technique to focus the data in the frequency domain, rather than PAA used in SAX. DFT and the continuous Fourier Transform for signals are alike, and are known only at N instants by sample times T , which is a finite series of data.

Let $S(t)$ be the continuous signal which is the source of the data. Let N samples be denoted $s[0], s[1], \dots, s[N-1]$. The Fourier transformation of the original signal, $S(t)$, would be:

$$F(\omega_k) \triangleq \sum_{n=0}^{N-1} s(t_n) e^{-j\omega_k t_n}, k = 0, 1, 2, \dots, N-1 \quad (3)$$

To determine the signal’s frequency content at $S[k]$, DFT examines a time-domain signal at $s(n)$ by comparing it against sinusoidal basis functions through correlation. The first few basis functions describe the gradual changing regions, while the later basis functions describe the rapid changes like gaps and noise. Therefore, using only the first few basis functions, one can have a good approximation of the entire series (Schäfer and Högvist, 2012).

The SFA model uses DFT approximation as part of pre-processing by transforming the original time-series into a series of DFT coefficients. SFA then applies Multiple Coefficient Binning (MCB) method for computing multiple discretisations of the coefficients series (see Figure 7). MCB engages in the mapping of the DFT coefficients to their respective symbols and then concatenates them to form an SFA word, transforming the time-series into a symbolic representation.

SFA also uses a sliding window in the same manner as SAX and the output is also a “bag-of-words” representation of the time-series.

4. EVALUATION

In the evaluation we consider the two classification tasks described in section 2.2. For convenience we refer to these as ‘Up Wind versus Down Wind’ and ‘High versus Low’. For the tasks we consider two versions of the data, the raw data and the filtered data as shown in Fig. 4. In addition to the three specialised time-series methods presented in section 3 we also consider Euclidean distance and the Mean of the time-series as baselines. The Mean is included to show that very simple aggregate statistics are not sufficient.

4.1 Up Wind versus Down Wind

The results on the first set of tasks are shown in Tables 1 and 2. It is clear that the Mean and Euclidean baselines do not perform well and DTW appears to be the overall winner.

Table 1. Up Wind versus Down Wind (Raw)

Model	0 vs 22	0 vs 1	1 vs 22	0 vs 1 vs 22
Mean	55.26	50.0	51.32	33.33
Euclidean	77.63	64.47	64.47	55.26
DTW	86.84	89.47	94.74	80.70
SAX	80.26	61.84	65.79	57.02
SFA	82.89	56.58	67.11	52.63

Table 2. Up Wind versus Down (Filtered)

Model	0 vs 22	0 vs 1	1 vs 22	0 vs 1 vs 22
Mean	59.21	52.63	51.32	33.33
Euclidean	73.68	61.84	67.11	56.14
DTW	88.16	65.79	86.84	64.04
SAX	82.89	57.89	84.21	59.65
SFA	71.05	57.89	76.32	45.61

4.2 High versus Low

The same evaluation is repeated for the High versus Low task. In this case there is only one task so the results are presented in a single table – Table 3. Again, DTW is the

Table 3. High versus Low

Model	Raw	Filtered
Mean	71.25	74.38
Euclidean	67.5	68.75
DTW	89.38	90.63
SAX	59.38	59.38
SFA	43.75	54.38

clear winner. It is interesting to note that, whereas Mean was no better than random guessing in the first exercise, it has some classification power here.

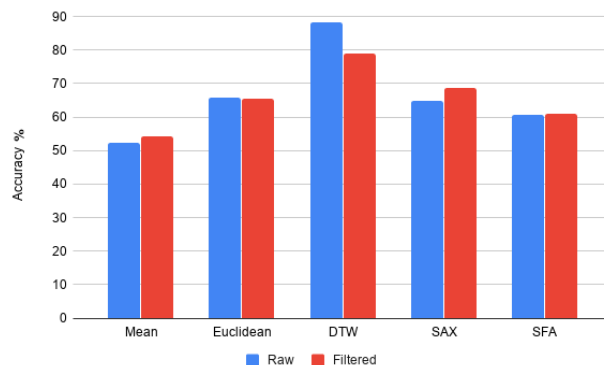


Fig. 8. Average model accuracy across all tasks.

4.3 Discussion

Fig. 8 shows the average model accuracy across all tasks. There are a few clear conclusions that can be drawn:

- It is pretty clear that Euclidean distance is not a competitive distance measure on this data. This is because of the variable length of the time-series and the associated problems of alignment.
- Of the three methods that are specifically conceived to deal with time-series, it seems clear that DTW beats both SAX and SFA. This is probably due to the global nature of DTW that considers the whole time-series whereas SAX and SFA work well when specific signatures in sub-regions of the time-series are important (Mahato et al., 2018).
- It is interesting to note that DTW works better with the raw data whereas SAX and SFA perform best with the data that has passed through the Butterworth filter.

5. CONCLUSION & FUTURE WORK

This evaluation shows that the temperature time-series data does capture information about the AM process. It also shows that ML methods specialised for time-series analysis are required to get the most from the data. As a next step we are carrying out CT scans on the blocks to identify pores and then see if these pores have a characteristic signature in the temperature time-series.

ACKNOWLEDGEMENTS

This publication has resulted from research supported in part by a grant from Science Foundation Ireland (SFI) under Grant Number 16/RC/3872 and is co-funded under the European Regional Development Fund.

REFERENCES

- Gorji, N.E., Saxena, P., Corfield, M., Clare, A., Rueff, J.P., Bogan, J., González, P.G., Snelgrove, M., Hughes, G., O'Connor, R., et al. (2020). A new method for assessing the utility of powder bed fusion (pbf) feedstock. *Materials Characterization*, 161, 110167.
- Keogh, E. and Kasetty, S. (2002). On the need for time series data mining benchmarks. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 02*. doi: 10.1145/775060.775062.
- Keogh, E.J. and Pazzani, M.J. (2000). Scaling up dynamic time warping for datamining applications. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 00*. doi: 10.1145/347090.347153.
- Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery - DMKD 03*. doi: 10.1145/882082.882086.
- Mahato, V., O'Reilly, M., and Cunningham, P. (2018). A comparison of k -NN methods for time series classification and regression. In R. Brennan, J. Beel, R. Byrne, J. Debattista, and A.C. Junior (eds.), *Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, 2018.*, volume 2259 of *CEUR Workshop Proceedings*, 102–113. CEUR-WS.org. URL <http://ceur-ws.org/Vol-2259/aics.11.pdf>.
- Özel, T., Shaurya, A., Altay, A., and Yang, L. (2018). Process monitoring of meltpool and spatter for temporal-spatial modeling of laser powder bed fusion process. *Procedia CIRP*, 74, 102–106.
- Qi, X., Chen, G., Li, Y., Cheng, X., and Li, C. (2019). Applying neural-network-based machine learning to additive manufacturing: Current applications, challenges, and future perspectives. *Engineering*.
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. doi:10.1109/TASSP.1978.1163055.
- Schäfer, P. and Höggqvist, M. (2012). SFA: a symbolic fourier approximation and index for similarity search in high dimensional datasets. In *Proceedings of the 15th International Conference on Extending Database Technology*, 516–527. ACM.
- Yujian, L. and Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6), 1091–1095.