Advanced clustering technique for automatic labelling of welding signals

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 138 (2026) 1213–1217

www.elsevier.com/locate/procedia

18th CIRP Conference on Intelligent Computation in Manufacturing Engineering

Advanced clustering technique for automatic labelling of welding signals

Mario Vozza[a,b*], Tommaso Forni[a,b], Fabio Le Piane[a,c], Alessandro Petrella[a], Giulio Mattera[d], Emily Yap[e], Luigi Nele[d], Francesco Mercuri[a]

[a] DAIMON Lab, CNR-ISMN, Bologna, Italy
[b] Department of Control and Computer Engineering, Polytechnic University of Turin, Turin, Italy
[c] Department of Computer Science and Engineering, University of Bologna, Bologna, Italy
[d] Department of Chemical, Materials and Production Engineering, University of Naples Federico II, Naples, Italy
[e] Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong NSW, Australia

* Corresponding author. *E-mail address: mario.vozza@polito.it*

Abstract

In the field of additive manufacturing, effective monitoring is indispensable for ensuring the quality of produced components. This task is usually solved using supervised learning approaches, whom need to deal with manually labelled datasets. In this research we leverage advanced artificial intelligence techniques, such as data-driven features extraction, dimensionality reduction and clustering algorithms, welding current and welding voltage during a Wire Arc Additive Manufacturing deposition have been labelled automatically. Features from both time and frequency domain have been automatically extracted, while Uniform Manifold Approximation and Projection (UMAP) algorithm is employed to reduce the number of features from 889 to 10 using a non-linear projection. Finally, Hierarchical clustering has been employed to generate labels.

Peer-review under responsibility of the scientific committee of the 18th CIRP Conference on Intelligent Computation in Manufacturing Engineering (CIRP ICME '24)

Keywords: welding; clustering; UMAP, features extraction

1. Introduction

Wire Arc Additive Manufacturing (WAAM) [1], [2] is an Additive Manufacturing technology based on the principle of welding which consist in depositing in a layer-by-layer fashion welding seams with the aim to build mid-sized metal components. Nowadays several industrial applications can be developed using Artificial Intelligence techniques like Machine Learning (ML) for feedback control [3], [4], [5], process monitoring [6], [7], [8] and parameters optimisation [9]. However, since most of ML techniques are supervised, labelling technique plays a crucial role in the algorithm performance. This task is usually solved by human, who analysing the output data, the welding audio, the surface appearance and conducting metallurgical inspection of the specimens, associated labels to data. Clustering techniques are algorithms of unsupervised learning family which can associate similar labels to data which exhibits a similar characteristic, employing distance based [10] or density-based methods [11]. In this work a methodology to perform advanced clustering of welding current and welding voltage signals of a WAAM process is presented. Once data has been collected and segmented, a data-driven approach for features extraction and dimensionality reduction is presented. Using the reduced features space, a Hierarchical clustering technique is employed to associate labels to each welding data.

2. Materials and methods

2.1. Experimental setup and data collection

The experimental setup, shown in Figure 1, consists of a collaborative robot UR10e and a Fronius TPS500i welding machine used in constant voltage mode. In the experiment, 4 walls composed of 12 layers of 65 mm length each were printed using the parameters reported in Table 1. During deposition,

2212-8271 © 2025 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 18th CIRP Conference on Intelligent Computation in Manufacturing Engineering (CIRP ICME '24)
10.1016/j.procir.2026.01.209

the Contact-To-Workpiece Distance (CTWD) was maintained constant from Wall 2 to the Wall 4 at the distance of 20 mm, giving the possibility to explore the effect of an uncontrolled CTWD on the final deposited layer during the building process of the wall 1. This generated defective parts which contain porosity and humping. The Gas Flow Rate was fixed at the value of 20 L/min, while the inter-pass temperature was maintained fixed to the value of 150°C.

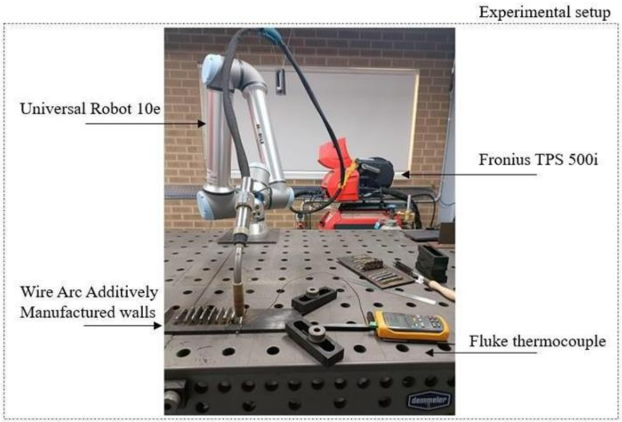

Figure 1: The experimental setup utilised in this study. A Cobot is used as motion platform to deposit walls using WAAM technology and a thermocouple is used to maintain the interpass temperature to the value of 150°C.

N	Wire Feed Speed [m/min]	Welding speed [m/min]	Welding voltage [V]
1	4.4	350	18.3
2	4.4	350	18.3
3	4.4	450	18.3
4	4.4	550	18.3

Table 1: Design of Experiments

The welding signals of welding current and welding voltage have been acquired using a sample rate of 5kHz using a NI-6009 device. As shown in Figure 2, the welding data collected has been segmented into 5000 length signals, aiming to labelling every second of deposition.

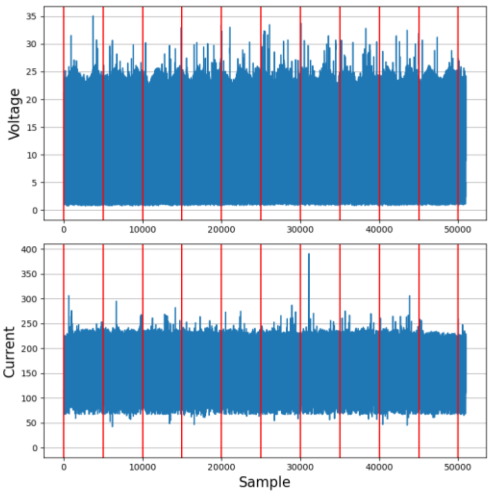

Figure 2: Collected and segmented welding signals.

2.2. Features extraction

The *TSFresh* library [12] has been used for automatic feature extraction from the two time-series of welding voltage and welding current. TSFresh extracts features in both the time domain and the frequency domain, such as mean, variance in the time domain energy and entropy calculated from the Fourier Transform of the input signal. In particular, the frequency domain features extracted from welding current and voltage signals has been demonstrated to be particularly beneficial for WAAM application. [13] Once the features were extracted, they were cleaned from missing values and the resultant features have been scaled with respect to the minimum and maximum values of each feature, bringing each column of the dataset into the range 0-1. After that, we performed another round of feature selection based on the data variance; this decision has been made because only the features that vary among the various samples are useful for differentiation. For this reason, only those with a variance greater than 0.02 were considered, resulting in a reduction from 1575 features to 889.

2.3. UMAP

Uniform Manifold Approximation and Projection (UMAP) [14] is a powerful technique used for dimensionality reduction, particularly in the field of machine learning and data visualization. Compared to traditional methods such as Principal Component Analysis (PCA) or t-distributed Stochastic Neighbor Embedding (t-SNE), UMAP offers several advantages including faster computation and better preservation of both local and global structure of the data. The primary logic behind this method involves constructing a high-dimensional representation of the data in the form of a graph. This graph captures the relationships between data points, emphasizing both local and global structures.
By optimizing a low-dimensional representation of this graph, UMAP aims to preserve the underlying structure of the data in fewer dimensions. Based on the previous considerations, in this study we decided to utilize UMAP on the previously filtered data to project the features onto a lower-dimensional space. The utility of this methodology lies in the fact that our dataset comprised over 1500 features. As demonstrated in numerous studies [15], [16] high-dimensional feature spaces often require many samples for effective clustering. By projecting with UMAP into a low-dimensional space, we avoid the pitfalls of high-dimensional data while preserving the topological structure of the original dataset, preserving most of the information it contained; moreover, by reducing the number of features, we also limit the impact of eventual noise present in the original data. Additionally, it allows for the acceleration of clustering computation, making the overall process more efficient. Therefore, the projection of our dataset was performed using a 10-component space. To ensure the preservation of a comprehensive representation of the features, a neighbourhood size of 20 was chosen, and Euclidean distance

was employed as the metric. It is worth noting that a lower neighbourhood parameter tends to provide a more local representation of the dataset, but there is a risk of capturing noise. Conversely, increasing this parameter leans towards a more global representation of the dataset. For the experiment reproducibility we choose a random seed of 42.

2.4. Hierarchical clustering

Hierarchical clustering is a widely used technique in the field of data analysis and machine learning for grouping similar data points into clusters based on their distance or similarity. Unlike partitioning methods like k-means, hierarchical clustering can be used for both specifying number of clusters or not. The fundamental idea behind hierarchical clustering is to build a hierarchy of clusters, where clusters at the same level of the hierarchy are more like each other than clusters at various levels. This hierarchy is often represented as a dendrogram, a tree-like structure that illustrates the arrangement of the clusters [17]. In this work we use agglomerative hierarchical clustering, with the distance threshold. For the selection of the maximum distance, two criteria were used to help assess the quality of the formed clusters. The silhouette score that measure how similar an instance is to its cluster compared to other clusters; higher silhouette score indicates that instances have been assigned to appropriate clusters [18]. The Davies-Bouldin Index (DBI) that measures the dispersion between clusters. A lower value indicates denser and more compact clusters [19]. For this reason, therefore, considering also the image below, a distance equal to 15 was chosen. The evaluation metric for the clusters, instead, was selected as the Euclidean metric.

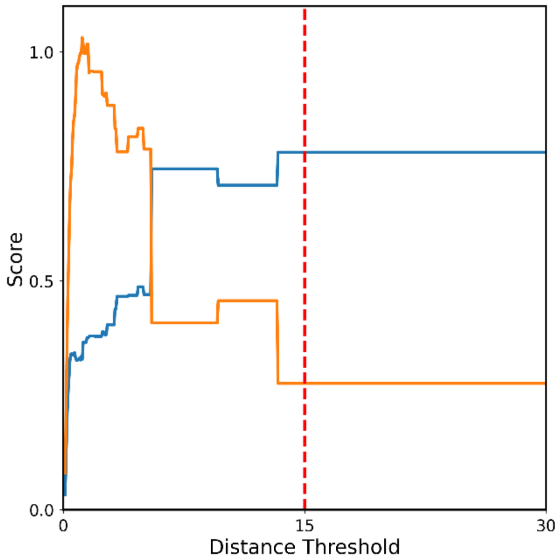

Figure 3: Comparison of Silhouette Score (blue line) and Davies-Bouldin Index values (orange line) for the optimal clustering configuration. The selected distance (red dotted line) represents the point where the Silhouette Score is at its maximum and the Davies-Bouldin Index is at its minimum, indicating good cluster cohesion and significant separation between clusters.

3. Results and discussion

At the end of the clustering process, 4 different clusters have been found, as illustrated in the Figure 3.

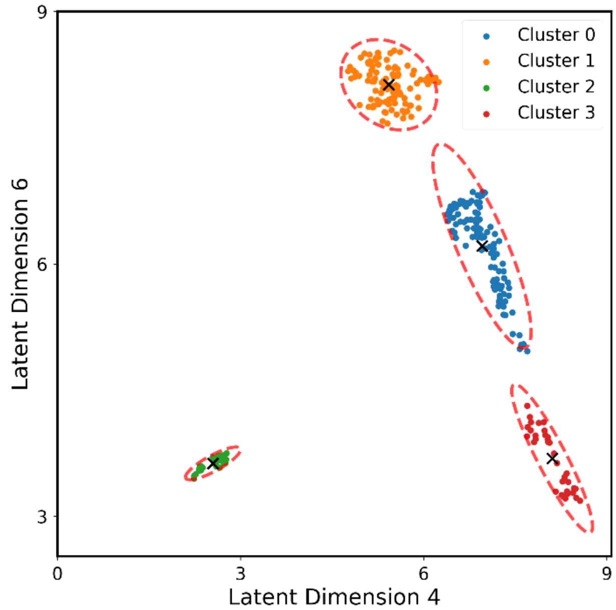

Figure 4: Cluster visualisation in the latent dimension 4 and 6 obtained employing UMAP algorithm. Each black 'x' denotes the centroid of a cluster, while red dotted lines depict the confidence ellipses for each cluster.

Cluster 0 and Cluster 1 contain samples related to normal deposition conditions. As shown in Figures 4, the Cluster 1 signals are associated with a deposition in which the droplet is transferred in the melting pool with a constant frequency, indicating a stable deposition process. In Figure 5 is shown the typical signal of the cluster 0, in which is possible to appreciate a less stable process, associated with changing in CTWD, but that can still be considered a normal deposition. However, if a new sample is associated with this cluster, a warning for the operator should be generated.

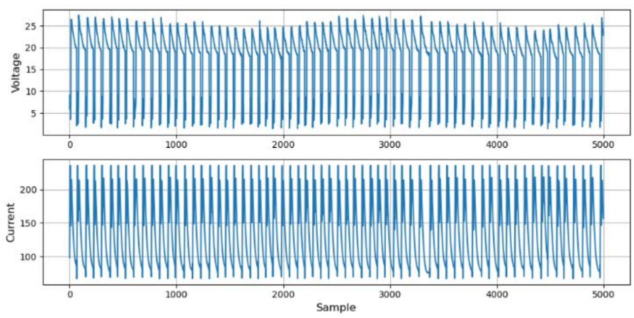

Figure 5: Cluster 1 contain depositions associated to stable deposition process.

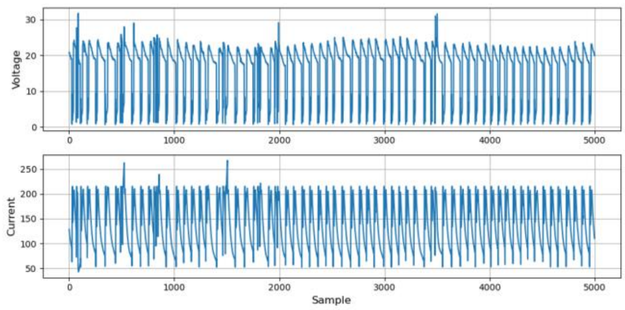

Figure 6: Cluster 0 contain depositions associated to good deposition process which is moving to instability.

For what concern the Clusters 2 and 3, they are related to unstable deposition conditions. Cluster 2 is associated to outliers related to arc ignition phenomena, which happen at the beginning of the deposition, or to undesired short circuit which results into defects on the final deposited layer, as shown in Figures 6 and Figure 7 for respectively the signals time series and the layer appearance.

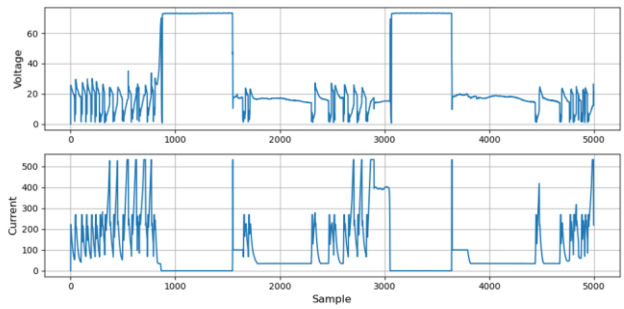

Figure 7: Welding signals associated to unstable deposition process which led into porosity in the deposited layers typical of Cluster 2.

Figure 8: Layer appearance for unstable deposition and undesired short circuit.

Finally, the Cluster 3, shown in Figure 8, is associated to uncontrolled deposition in which the frequency of the droplet release in the melting pool is different with respect to the planned one. In this case the layer can present an undesired layer geometry, since the welding current and the welding voltage are far from the planned one, and problems like spatter on the adjacent layers which can compromise the final proprieties of the printed component.

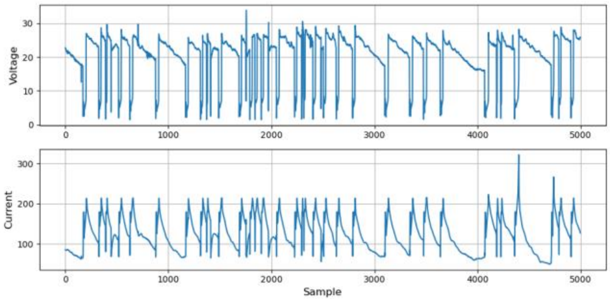

Figure 9: Cluster 3 contain depositions associated to unstable deposition process.

4. Conclusion

In this research work, a clustering approach based on UMAP and Agglomerative Hierarchical Clustering (AHC) was developed with the aim to automatically label layers segments in Wire Arc Additive Manufacturing process. The developed methodology was studied through an experimental campaign performed using normal deposition conditions and data coming from an unstable process due to uncontrolled CTWD. An automatic features extraction method has been employed to extract features from both time and frequency domain and UMAP algorithm has been employed to reduce the number of features from 889 to 10 trough a non-linear projection. Finally, the AHC method has been employed and an optimal threshold distance has been computed based on the Silhouette score and Davies-Bouldin Index. The results shown the presence of 4 clusters in the data, and observing the signals and layers appearance, a meaning to each cluster has been associated. Using the proposed methodology the dataset can be automatically labelled and the samples can be used to train supervised learning classifier. Furthermore, the centroids of the obtained clusters can be used in a distance-based classification system in which new samples coming from the WAAM process can be labelled as the closest centroid label. The presented methodology can be extended to different materials, anomalies and defects and deployed in IoT networks to reach zero-defects manufacturing goals.

Acknowledgement

The INVITALIA Project on NEMESI and POR Campania FESR 2014/2020 Project TOP are gratefully acknowledged for their support to this research work. We acknowledge the Facility for Intelligent Fabrication laboratory of University of Wollongong and their research group for providing equipment and support throughout this research. Furthermore, we would like to thank the DESIGN-IT research project (protocol number 69, Public Notice Innovation Agreements DM 31/12/2021 – I Sportello), funded by the Ministry of Enterprises and Made in Italy under Mission 4 (Education and research), Component 2 (From research to enterprise), and the COST action CA22143 - European Materials Informatics Network (EuMINe), supported by COST (European Cooperation in Science and Technology).

References

[1] Z. Pan, D. Ding, B. Wu, D. Cuiuri, H. Li, and J. Norrish, "Arc Welding Processes for Additive Manufacturing: A Review," 2018, pp. 3–24. doi: 10.1007/978-981-10-5355-9_1.

[2] J. Norrish, J. Polden, and I. Richardson, "A review of wire arc additive manufacturing: development, principles, process physics, implementation and current status," *J Phys D Appl Phys*, vol. 54, no. 47, p. 473001, Nov. 2021, doi: 10.1088/1361-6463/ac1e4a.

[3] G. Mattera, A. Caggiano, and L. Nele, "Optimal data-driven control of manufacturing processes using reinforcement learning: an application to Wire Arc Additive Manufacturing," *J Intell Manuf*, 2024.

[4] Y. Wang *et al.*, "Coordinated monitoring and control method of deposited layer width and reinforcement in WAAM process," *J Manuf Process*, vol. 71, pp. 306–316, 2021, doi: https://doi.org/10.1016/j.jmapro.2021.09.033.

[5] H. Mu, J. Polden, Y. Li, F. He, C. Xia, and Z. Pan, "Layer-by-layer model-based adaptive control for wire arc additive manufacturing of thin-wall structures," *J Intell Manuf*, vol. 33, no. 4, pp. 1165–1180, Apr. 2022, doi: 10.1007/s10845-022-01920-5.

[6] G. Mattera, L. Nele, and D. Paolella, "Monitoring and control the Wire Arc Additive Manufacturing process using artificial intelligence techniques: a review," *J Intell Manuf*, 2023, doi: 10.1007/s10845-023-02085-5.

[7] J. Y. I. Alcaraz, W. Foqué, A. Sharma, and T. Tjahjowidodo, "Indirect porosity detection and root-cause identification in WAAM," *J Intell Manuf*, Apr. 2023, doi: 10.1007/s10845-023-02128-x.

[8] Y. Li *et al.*, "A defect detection system for wire arc additive manufacturing using incremental learning," *J Ind Inf Integr*, vol. 27, p. 100291, May 2022, doi: 10.1016/j.jii.2021.100291.

[9] G. Mattera, A. Caggiano, and L. Nele, "Reinforcement learning as data-driven optimization technique for GMAW process," *Welding in the World*, 2023, doi: 10.1007/s40194-023-01641-0.

[10] R. Reisch, T. Hauser, B. Lutz, M. Pantano, T. Kamps, and A. Knoll, "Distance-Based Multivariate Anomaly Detection in Wire Arc Additive Manufacturing," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Dec. 2020, pp. 659–664. doi: 10.1109/ICMLA51294.2020.00109.

[11] G. Mattera, J. Polden, A. Caggiano, P. Commins, L. Nele, and Z. Pan, "Anomaly Detection of Wire Arc Additively Manufactured Parts via Surface Tension Transfer through Unsupervised Machine Learning Techniques," in *17th CIRP Conference on Intelligent Computation in Manufacturing Engineering*, Naples: Procedia CIRP, 2023.

[12] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018, doi: 10.1016/j.neucom.2018.03.067.

[13] G. Mattera, J. Polden, and L. Nele, "A Time-Frequency domain features extraction approach enhanced by computer vision for Wire Arc Additive Manufacturing monitoring using Fourier and Wavelet transform," *Journal of Advanced Manufacturing Systems*, 2024.

[14] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.03426

[15] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans Neural Netw*, vol. 13, no. 3, pp. 780–784, May 2002, doi: 10.1109/TNN.2002.1000150.

[16] M. Mittal, L. M. Goyal, D. J. Hemanth, and J. K. Sethi, "Clustering approaches for high-dimensional databases: A review," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 3, May 2019, doi: 10.1002/widm.1300.

[17] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, Jan. 2012, doi: 10.1002/widm.53.

[18] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J Comput Appl Math*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[19] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans Pattern Anal Mach Intell*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.