Article

Title

Firstname Lastname [1,†,‡], Firstname Lastname [2,‡] and Firstname Lastname [2,*]

[1] Affiliation 1; e-mail@e-mail.com
[2] Affiliation 2; e-mail@e-mail.com
[*] Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials) +xx-xxxx-xxx-xxxx (F.L.)
[†] Current address: Affiliation.
[‡] These authors contributed equally to this work.

Abstract: A single paragraph of about 200 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: (1) Background: place the question addressed in a broad context and highlight the purpose of the study; (2) Methods: describe briefly the main methods or treatments applied; (3) Results: summarize the article's main findings; (4) Conclusions: indicate the main conclusions or interpretations. The abstract should be an objective representation of the article, it must not contain results which are not presented and substantiated in the main text and should not exaggerate the main conclusions.

Keywords: keyword 1; keyword 2; keyword 3 (List three to ten pertinent keywords specific to the article; yet reasonably common within the subject discipline.)

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Related Works

The K-means algorithm is used in [1,2] to label the degradation states of bearings. In [1], the labelled timeseries are than used to train a CNN recognition model. Traditional statistical features (TSF) and MFCC are used to form the hyperspace in which to cluster the data. When evaluating the state of the system, the raw data are fede to CNN model withowt the need to extract features. The authors validated this algorithm on the IMS bearing dataset [3]. In [2], the TSF are used toghether with the Shannon's entropy as features while performing the clustering. The authors then converted the timeseries into images and used a CNN (Alexnet) to classify the degradation states. This method was validated on the IMS and the CWRU [4] datasets.

A method to efficiently initialize the clusters centers in the K-means algorithm is proposed in [5]. The authors merged the Ant Colony Optimization (ACO) algorithm with the K-means. This new ACO-K-means method has been validated applying a three-layer Wavelet Packet Decomposition (WPD) on the CWRU dataset. The dataset has been aughmented using sliding window method to prove the applicability of the method to large datasets.

In [6], the authors propose a method to quantify the health of bearing. The time-domain features are extracted from the vibration signal and reduced, applying a cross-correlation filter to remove the redundant features. The K-means algorithm is then used to select only the most relevant features (optimising for obtaining the most dense and separated clusters). The SOM algorithm is then used to compute a health indicators for the bearing. The authors validated this method on the IMS dataset.

In [7], the authors propose a subset based deep auto-encoder model to automatically learn discriminative features from datasets. This approach has been validated on the CWRU, IMS and SPCP bearing vibration datasets.

A case-study on the IMS dataset is proposed in [8]. The anthors leverage the knowledge the Fault frequencies of the bearings to provide labels to the data. The features considered initially are TSF and RSGWPT coefficients. The dimensionality of the feature space is reduced by applying PCA. Lastly, K-means, SVM and agglomerative clustering are used to perform anomaly detection and to identify the failure modes.

A powerful approach to data streams classification is proposed in [9]. It is not a one-class approach as it manages more than one "normal" class as well as the emerging of novel classes during the classification task.

A metric for quantifying how novel a record is in a dataset has been developed by [10]. This metric is called Local Outlier Factor (LOF) and depends on the position of the point and the density of known points around it. This concept has been extended to consider also the clustered structure of the data in [11], that define the metric cluster-based LOF (CBLOF).

A very simple way of quantifying the normality of data using a distance metric is proposed in [12]. The authors propose to compute a novelty score that is the distance of the record from the nearest cluster, normalized by the standard deviation of the distance of the known points in the cluster. This method has been tested on vibration data collected on a jet engine.

Another distance-based method is proposed in [13]. The method profides a hard classification of the data as normal, extension or unknown, based on the radius of the closest cluster and the position of the point to be evaluates. If the point is outside the decision boundary but whithin a tolerance, than it is classified as extension and the learned model is updated. If the point is outside the tolerance, than it is classified as unknown. Unknown points are kept in a buffer and used to update the model when new classes emerge whithin the data.

A computer vision method to detect anomalies in mechanical systems is proposed in [14]. This has the advantage of evaluating vibrations in multiple points of interest without phisical contact with the observed component.

Abbreviations

The following abbreviations are used in this manuscript:

IMS	Intelligent Maintenance Systems
MFCC	Mel-frequency cepstral coefficients
TSF	Traditional Statistical Features
CWRU	Case Western Reserve University
SOM	Self-Organizing Maps
SPCP	self-priming centrifugal pump
SVM	Support Vector Machine
RSGWPT	Redundant Second Generation Wavelet Packet Transform
LOF	Local Outlier Factor
CBLOF	Cluster-Based Local Outlier Factor
NM	Novelty Metric
WPD	Wavelet Packet Decomposition

References

1. Zhou, Q.; Shen, H.; Zhao, J.; Liu, X.; Xiong, X. Degradation State Recognition of Rolling Bearing Based on K-Means and CNN Algorithm. *Shock and Vibration* **2019**, *2019*, 8471732, [https://onlinelibrary.wiley.com/doi/pdf/10.1155/2019/8471732]. https://doi.org/https://doi.org/10.1155/2019/8471732.
2. Pinedo-Sánchez, L.A.; Mercado-Ravell, D.A.; Carballo-Monsivais, C.A. Vibration analysis in bearings for failure prevention using CNN. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **2020**, *42*, 628. https://doi.org/10.1007/s40430-020-02711-w.
3. Lee, J.; Qiu, H.; Yu, G.; Lin, J.; Services, R.T. Bearing Data Set. IMS, University of Cincinnati, NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007.
4. University, C.W.R. Bearing Data Center: Seeded Fault Test Data. http://data-acquisition.case.edu/. Accessed: 2024-06-05.
5. Wan, L.; Zhang, G.; Li, H.; Li, C. A Novel Bearing Fault Diagnosis Method Using Spark-Based Parallel ACO-K-Means Clustering Algorithm. *IEEE Access* **2021**, *9*, 28753–28768. https://doi.org/10.1109/ACCESS.2021.3059221.

6. Chalouli, M.; Berrached, N.e.; Denai, M. Intelligent Health Monitoring of Machine Bearings Based on Feature Extraction. *Journal of Failure Analysis and Prevention* **2017**, *17*, 1053–1066. https://doi.org/10.1007/s11668-017-0343-y.
7. Zhang, Y.; Li, X.; Gao, L.; Li, P. A new subset based deep feature learning method for intelligent fault diagnosis of bearing. *Expert Systems with Applications* **2018**, *110*, 125–142. https://doi.org/https://doi.org/10.1016/j.eswa.2018.05.032.
8. Gattino, C.; Ottonello, E.; Baggetta, M.; Razzoli, R.; Stecki, J.; Berselli, G. Application of AI failure identification techniques in condition monitoring using wavelet analysis. *The International Journal of Advanced Manufacturing Technology* **2023**, *125*, 4013–4026. https://doi.org/10.1007/s00170-022-10549-w.
9. Masud, M.; Gao, J.; Khan, L.; Han, J.; Thuraisingham, B.M. Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints. *IEEE Transactions on Knowledge and Data Engineering* **2011**, *23*, 859–874. https://doi.org/10.1109/TKDE.2010.61.
10. Breunig, M.; Kröger, P.; Ng, R.; Sander, J. LOF: Identifying Density-Based Local Outliers. 06 2000, Vol. 29, pp. 93–104. https://doi.org/10.1145/342009.335388.
11. He, Z.; Xu, X.; Deng, S. Discovering cluster-based local outliers. *Pattern Recognition Letters* **2003**, *24*, 1641–1650. https://doi.org/https://doi.org/10.1016/S0167-8655(03)00003-5.
12. Clifton, D.A.; Bannister, P.R.; Tarassenko, L. Learning Shape for Jet Engine Novelty Detection. In Proceedings of the Advances in Neural Networks - ISNN 2006; Wang, J.; Yi, Z.; Zurada, J.M.; Lu, B.L.; Yin, H., Eds., Berlin, Heidelberg, 2006; pp. 828–835.
13. Garcia, K.D.; Poel, M.; Kok, J.N.; de Carvalho, A.C.P.L.F. Online Clustering for Novelty Detection and Concept Drift in Data Streams. In Proceedings of the Progress in Artificial Intelligence; Moura Oliveira, P.; Novais, P.; Reis, L.P., Eds., Cham, 2019; pp. 448–459.
14. Spytek, J.; Machynia, A.; Dziedziech, K.; Dworakowski, Z.; Holak, K. Novelty detection approach for the monitoring of structural vibrations using vision-based mean frequency maps. *Mechanical Systems and Signal Processing* **2023**, *185*, 109823. https://doi.org/https://doi.org/10.1016/j.ymssp.2022.109823.