

# Canonical Polyadic Decomposition and Deep Learning for Machine Fault Detection

Frusque Gaëtan<sup>1</sup>, Michau Gabriel<sup>2</sup>, and Fink Olga<sup>3</sup>

<sup>1,2,3</sup> *ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland*  
*frusque.gaetan@outlook.fr*

## ABSTRACT

Acoustic monitoring for machine fault detection is a recent and expanding research path that has already provided promising results for industries. However, it is impossible to collect enough data to learn all types of faults from a machine. Thus, new algorithms, trained using data from healthy conditions only, were developed to perform unsupervised anomaly detection. A key issue in the development of these algorithms is the noise in the signals, as it impacts the anomaly detection performance. In this work, we propose a powerful data-driven and quasi non-parametric denoising strategy for spectral data based on a tensor decomposition: the Non-negative Canonical Polyadic (CP) decomposition. This method is particularly adapted for machine emitting stationary sound. We demonstrate in a case study, the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) baseline, how the use of our denoising strategy leads to a sensible improvement of the unsupervised anomaly detection. Such approaches are capable to make sound-based monitoring of industrial processes more reliable.

## 1. INTRODUCTION

Early machine fault detection is a highly relevant industrial problem and is one of the pillars of the industry 4.0 to lower the cost of maintenance (Stojanovic & Dinic, 2016). Sound is one type of signals that can be recorded to monitor the health state of the machine. These signals are currently a topic of interest because acoustic monitoring offers a great benefit for the industries since it is cheap, non-invasive, non-disruptive and easy to install also as a refurbishment to existing assets. However, the discrimination between sounds from healthy and abnormal machines remain a challenging task due to following reasons: (1) failures are typically rare, (2) failures recorded are not necessarily representative of all possible failure types, (3) the noise levels and the presence of artefacts can be high (Purohit et al., 2019), (Chao, Kulkarni, Goebel, & Fink, 2020),

Gaëtan Frusque et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(4) different operational modes make it difficult to characterize normal from abnormal states of the machine.

To tackle these problems, unsupervised anomaly detection strategies were proposed (Purohit et al., 2019), (Michau, Thomas, & Fink, 2017), (Tagawa, Tadokoro, & Yairi, 2015). The idea is to use healthy data only for the training of the anomaly detector. These approaches are designed to identify anomalies without any knowledge on the characteristics of the faulty system conditions. This method allows taking into account the imbalance between the two classes (healthy and abnormal case) as well as the lack of recordings of all possible faults.

Deep learning based methods have recently shown remarkable performances for unsupervised detection problem (and more generally for audio signal processing (Purwins et al., 2019)). These methods usually are a 1-class classifier (Michau, Hu, Palmé, & Fink, 2020) or an autoencoder with a decision on the residual to perform detection (Purohit et al., 2019). In (Michau et al., 2020), a stacked extreme machine learning procedure is used to perform simultaneously unsupervised feature learning and feature aggregation in a single indicator for fault detection. This method can learn the relevant features from real condition monitoring data with higher accuracy compared to manually selected features. More recently (Michau & Fink, 2021), a 1-class classifier was combined with an adversarial domain discriminator to transfer the developed models and the operational experience between machines subject to changing operating conditions, in a setup where the target machine has only limited operational experience.

The condition monitoring signals are often recorded at a high sampling rate, and the number of samples is often too high for applying end-to-end deep learning methods on the raw signals directly. Feature extraction is then often applied for high frequency condition monitoring data. One possible way to extract informative features is to transform the signals in a different domain. Transformation in the frequency domain, via a spectrogram for example (Purohit et al., 2019), has been shown to be particularly suitable. However, more specific transformations using wavelet analysis or dictionary learning have been proposed (Whitaker, Suresha, Liu, Clifford, & An-

derson, 2017), (Chen, Wei, & Zhang, 2020). In addition to reducing the dimension of the signal, a good signal transform highlights the relevant motifs characterising the dataset and facilitates the learning of the neural network.

However, these transformations do not necessarily eliminate the noise and artefacts present in the dataset. A signal corrupted by several sources of noise will result in a poor anomaly detection performance by increasing the overfitting effect or by blocking the neural network training in irrelevant local minima.

In this work, we focus on the importance of denoising the high frequency data before using a deep learning method for a subsequent anomaly detection task. The core idea of our proposed strategy is to reduce the complexity of the dataset, thereby, improving the classification performance of the neural network. To improve the quality of the spectral representation, we take advantage of the data structure, consisting of discontinuous recordings of the machine at several instants and focus on eliminating recordings-specific artefacts as well as noise. We propose a powerful, data-driven and quasi non-parametric denoising strategy using the non-negative Canonical Polyadic (nnCP) decomposition (Lim & Comon, 2009). This procedure is particularly suited when the sound of the machine is stationary. We compare this method to another denoising strategy using a more standard decomposition: the Non-Negative Matrix (NMF) factorisation (Lee & Seung, 2001).

This article is organised as follows: we first introduce, in section 2, the proposed denoising strategies using the NMF and nnCP decomposition. These methods are then applied to a case study, the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) baseline, which is presented in section 3. Comparison and evaluation of the different methods using the MIMII dataset are presented in section 4. Finally, results and perspectives are discussed in section 5.

## 2. METHOD

We consider a typical dataset used as the input feature of an 1-class classifier or an autoencoder for fault detection; which is the time-frequency representation of each audio recording from a machine. This dataset can be conveniently written as a three-dimensional cube of data, called tensor and denoted by  $\mathcal{X} \in \mathbb{R}^{F \times T \times N}$ , where  $F$  is the number of frequencies considered,  $T$  is the number of frames and  $N$  is the number of recordings. We present in this section two strategies using the NMF and the CP to denoise and remove artefacts in  $\mathcal{X}$ .

### 2.1. Non-Negative Matrix Factorisation (NMF)

NMF is a well-known matrix decomposition that can be seen as an instance of soft clustering (Aggarwal & Reddy, 2013). It can be used for feature extraction and was recently applied on acoustic signals for event detection (Komatsu, Toizumi,

Kondo, & Senda, 2016). Since this method has to be applied on a matrix and not a tensor, we consider the matrix representation of the tensor  $\mathcal{X}$ , noted by  $\mathbf{X}_{MAT} \in \mathbb{R}^{F \times TN}$  and defined as :

$$\mathbf{X}_{MAT} = [\mathbf{X}_{:,1}, \mathbf{X}_{:,2}, \dots, \mathbf{X}_{:,N}]. \quad (1)$$

Notice that  $\mathbf{X}_{MAT}$  is the concatenation of the time-frequency representations of each recording. Applying the NMF on  $\mathbf{X}_{MAT}$  aims to find two matrices of lower dimensions  $\mathbf{A} \in \mathbb{R}^{F \times K}$  and  $\mathbf{B} \in \mathbb{R}^{TN \times K}$  which are the solution of the following optimisation problem :

$$\underset{\mathbf{A} > 0, \mathbf{B} > 0}{\operatorname{argmin}} \|\mathbf{X}_{MAT} - \mathbf{A}\mathbf{B}^t\|_F^2 \quad (2)$$

Where  $\|\bullet\|_F$  is the Frobenius norm.  $\mathbf{A}$  and  $\mathbf{B}$  are constrained to be non-negative matrices (i.e. all their elements have to be higher or equal to zeros). The only hyperparameter here is  $K$ , the number of columns (or components) in each matrix.

The columns of the matrix  $\mathbf{A}$  can be seen as the main spectral motifs characterising the dataset. Each motif activates through time according to the corresponding activation profile which are the columns of the matrix  $\mathbf{B}$ . The non-negativity constraint is important here since it will induce sparsity in the obtained motifs which contribute to obtaining denoised components (Frusque, Jung, Borgnat, & Gonçalves, 2020). Also, since spectral data are non-negative, it will help to obtain interpretable components (Donoho & Stodden, 2004), (Frusque, Borgnat, Gonçalves, & Jung, 2020).

Thus, the idea of denoising with NMF, is to reconstruct the input feature  $\mathbf{X}_{MAT}$  using the denoised motifs in the matrix  $\mathbf{A}$ . The denoised version of  $\mathbf{X}_{MAT}$  using NMF is noted as  $\hat{\mathbf{X}}_{MAT}$  and corresponds to :

$$\hat{\mathbf{X}}_{MAT} = \mathbf{A}\mathbf{B}^t. \quad (3)$$

Please note that by performing the inverse transformation of Eq. 1, the denoised data from the matrix  $\hat{\mathbf{X}}_{MAT}$  can also be represented in the form of a tensor, noted as  $\hat{\mathcal{X}}$ .

There are many algorithms to solve the optimisation problem stated in Eq. 2. The major difficulty here is that Eq. 2 is non-convex. Multiplicative update (Lee & Seung, 2001) or alternating least square (Cichocki & Zdunek, 2007) are state of the art approaches and are widely used algorithms to find a solution to the problem Eq. 2.

### 2.2. Non-Negative Canonical Polyadic Decomposition (nnCP)

We propose to take advantage of the characteristic 3-dimensional structure of the tensor  $\mathcal{X}$ . Tensor decompositions like canonical polyadic decomposition are known to have good pattern extraction properties. A canonical polyadic decomposition was recently used for fault detection in vibration data (Hu, Wang, & Bai, 2019). The nnCP can then be seen as an extension of

the NMF for data represented as a tensor. Before introducing the nnCP, it is important to note that the product between two matrices can be written as the sum of rank-1 matrices such as :

$$\mathbf{AB}^t = \sum_{k=1}^K \mathbf{a}_{:k} \times \mathbf{b}_{:k} \quad (4)$$

With  $\times$  being the vector product (let  $\mathbf{u} \in \mathbb{R}^F$  and  $\mathbf{v} \in \mathbb{R}^T$ , the tensor product  $\mathbf{u} \times \mathbf{v} = \mathbf{M} \in \mathbb{R}^{L \times T}$  is such that  $m_{ft} = u_f v_t$  (Comon, 2014)). Then, applying a nnCP decomposition on  $\mathcal{X}$  aims to approximate this tensor with three matrices  $\mathbf{A} \in \mathbb{R}^{F \times K}$ ,  $\mathbf{B} \in \mathbb{R}^{T \times K}$  and  $\mathbf{C} \in \mathbb{R}^{N \times K}$  maximising the following optimisation problem :

$$\underset{\mathbf{A}>0, \mathbf{B}>0, \mathbf{C}>0}{\operatorname{argmin}} \quad \|\mathcal{X} - \sum_{k=1}^K \mathbf{a}_{:k} \times \mathbf{b}_{:k} \times \mathbf{c}_{:k}\|_F^2, \quad (5)$$

The motivation behind the non-negativity constraint is the same as for the NMF. Moreover, it warrants the existence of a global solution for this decomposition (Qi, Comon, & Lim, 2016). An illustration of the nnCP is provided in Figure 1.

This decomposition leads to remove several sources of noise and artefacts that are specific to the particular recordings. Indeed, a component of the matrix  $\mathbf{C}$ , noted  $\mathbf{c}_{:k}$ , regroups recordings that share a similar spectral motif  $\mathbf{a}_{:k}$ . This spectral motif has to activate synchronously through time for every recording of the group (coinciding with the temporal activation  $\mathbf{b}_{:k}$ ). Thus, only time-coherent motifs that are redundant in several recordings will be extracted by this decomposition. Moreover, the nnCP is particularly beneficial if the sound emitted by the machine is stationary. In that case, the spectral motifs characterising the signal will be associated with constant activation profiles present in several recordings.

Finally, denoising with nnCP corresponds to reconstructing the input features  $\mathcal{X}$  using the denoised motifs  $\mathbf{A}$ . The denoised version of  $\mathcal{X}$  using nnCP is :

$$\hat{\mathcal{X}} = \sum_{k=1}^K \mathbf{a}_{:k} \times \mathbf{b}_{:k} \times \mathbf{c}_{:k}. \quad (6)$$

Several algorithms have been proposed to solve the optimisation problem Eq. 5 which is also non-convex. In this work, we used the Hierarchical Alternating Least Square (HALS) algorithm (Cichocki & Phan, 2009) as it is easy to implement and showed a good performance compared to other algorithms.

### 3. CASE STUDY

We apply our proposed denoising procedure to an acoustic monitoring case study, which is the benchmark from the MIMII dataset (Purohit et al., 2019). This dataset consists of four different machine operating sounds recorded in real factory environments. The four machines are a pump, a fan, a slider and a valve. The benchmark consists of an autoencoder-

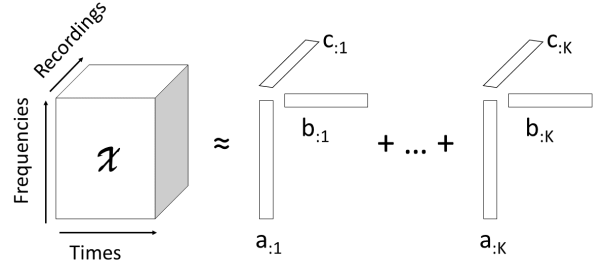


Figure 1. Illustration of the canonical polyadic decomposition.

based anomaly detector with the log-Mel spectrogram of the sound signal as input features. In the following, we provide more details about this benchmark case study.

#### 3.1. Time-frequency representation

The sound of each machine are recorded via an 8-channel microphone array. Only the signal from the first channel is considered. Each recording corresponds to a 10-second sound segment sampled at 16 kHz. To reduce the dimension of the input feature, a time-frequency representation of each recording is performed via a log-Mel spectrogram (McFee et al., 2015). Compared to a classic spectrogram, a Mel-spectrogram simulates the sound perception of humans. Indeed, humans do not perceive frequencies on a linear scale and are better at detecting differences in lower frequencies than higher ones. Then, in a Mel-spectrogram, the frequencies are rescaled (via the Mel scale) such that equal distances in pitch sounds similarly distant to the listener. This transformation is relevant if the difference between normal emission and anomaly can be easily perceived by a human with trained ears.

The application of the log-Mel spectrogram on the  $N$  recordings from one machine yields the following dataset denoted  $\mathcal{X} \in \mathbb{R}^{F \times T \times N}$ , with  $F$  and  $T$  respectively the number of Mel-frequencies and frames used for the log-Mel spectrogram computation of each sound recording. We use the same hyperparameters to compute the Mel-spectrogram as in (Purohit et al., 2019): we set  $F = 64$  and  $T = 314$ .

#### 3.2. Anomaly detector

The anomaly detector is based on an autoencoder. Considering the input feature  $\mathbf{x} = \mathbf{x}_{:m} \forall t \in 1, \dots, T$  and  $\forall n \in 1, \dots, N$  which is the log-Mel-frequencies of one frame from one recording, the loss function associated with the autoencoder is given by:

$$\mathcal{L}(\mathbf{x}; \theta) = \|\mathbf{x} - D_{\theta_d}(E_{\theta_e}(\mathbf{x}))\|_F^2, \quad (7)$$

where  $\theta = (\theta_e, \theta_d)$ ,  $E_{\theta_e}(\bullet)$  and  $D_{\theta_d}(\bullet)$  are the encoder and the decoder networks, with parameters  $\theta_e$  and  $\theta_d$  respectively. The encoder network  $E_{\theta_e}(\bullet)$  comprises two dense layers of 64

nodes plus one of 8 nodes. The decoder is composed of two dense layers of 64 nodes. The activation function used is the Rectified Linear Units (ReLUs).

In order to learn the distribution of the normal signals, the autoencoder is trained on sound from healthy machines only, referring to this training dataset as  $\mathcal{X}^{train}$ . We are looking for the best parameters  $\theta^*$  which minimise the root mean square error between input features and their reconstructions :

$$\operatorname{argmin}_{\theta} \sum_{\mathbf{x} \in \mathcal{X}^{train}} \mathcal{L}(\mathbf{x}; \theta), \quad (8)$$

The trained model will have a high reconstruction error when applied to an abnormal machine signal because it will be outside the distribution of the training samples. The separation between normal and abnormal sound can be made by thresholding the reconstruction error. More explicitly, considering a recording  $\mathbf{X}$ , we use the following decision function  $\mathcal{F}(\mathbf{X}, \theta^*)$  to determine if the recording is from a normal or abnormal machine :

$$\mathcal{F}(\mathbf{X}, \theta^*) = \begin{cases} \text{Normal if } \sum_{\mathbf{x} \in \mathbf{X}} \mathcal{L}(\mathbf{x}; \theta^*) < \phi, \\ \text{Anormal if } \sum_{\mathbf{x} \in \mathbf{X}} \mathcal{L}(\mathbf{x}; \theta^*) \geq \phi. \end{cases} \quad (9)$$

The selection of threshold  $\phi$  is a compromise between detecting the anomalies and limiting the false detection. This can be done by using a validation dataset containing abnormal machine sound or by analysing the distribution of the reconstruction error of the training dataset.

### 3.3. Insertion of the denoising strategy

To improve the autoencoder performance, we propose to use the denoised signal  $\hat{\mathbf{X}}$  derived by NMF or nnCP as input to the autoencoder. An overview of the proposed denoising strategy is provided Figure 2. We use the same autoencoder as in the baseline to make sure that the observed improvement is due to the effect of denoising applied on the input. The only hyperparameter to consider for each method is the number of components  $K$ . If  $K$  is too low, the reconstruction is done with few spectral motifs. In this case, we will have good denoising properties and artefact removal but a crucial spectral motif can be missed. On the contrary, if  $K$  is too high, the noise and artefacts present in the dataset will start to be reconstructed. However, several values of  $K$  leads to relevant results. An Elbow criterion can be used to help with the selection of this hyperparameter.

Finally, two additional remarks have to be discussed:

- The log-Mel spectrogram contain only non-positive elements whereas NMF and nnCP have the non-negativity constraint. Then, the decomposition Eq. 2 and Eq. 5 are performed on the absolute value of  $\mathcal{X}$ .
- The MIMII baseline used a data-augmentation strategy,

where the input feature is not directly  $\mathbf{x}$  (or  $\hat{\mathbf{x}}$ ), but the concatenation of five successive frames from the same recording. We use the same procedure here.

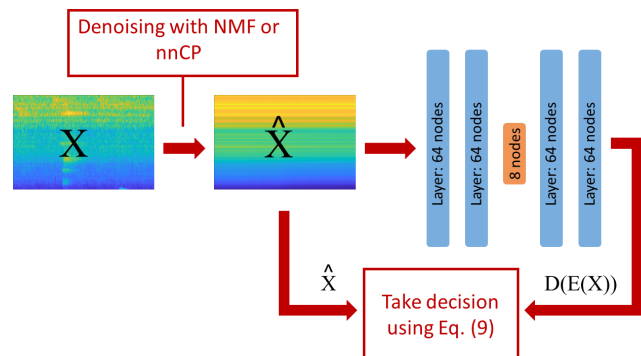


Figure 2. Illustration of the detection method, the input feature is first denoised using the NMF or nnCP strategy, a decision is made by comparing the output of the autoencoder with the denoised input using Eq. 9

## 4. RESULTS

We consider the openly available dataset MIMII to validate our proposed approach. We use all machines available and divide them into two groups. When the signal of interest is stationary: containing the fan and pump; and when the signal of interest is non-stationary: containing the valve and the slider. For example, in the case of the valve, the acoustic emissions are impulsive and sparse. This distinction is important because we illustrate in the results why the nnCP is particularly adapted for stationary signal, but not necessary for non-stationary ones. For each machine, we generated training and validation datasets in the same manner as in the paper (Purohit et al., 2019). The training dataset  $\mathcal{X}^{training}$  comprises only normal sound. The validation dataset  $\mathcal{X}^{valid}$  contains as many normal recordings as abnormal ones. The objective is to detect only the abnormal sound from the validation dataset after having trained the network with the training dataset.

### 4.1. Validation

The autoencoder presented in Section 3.2 is trained with the Adam optimization technique (Kingma & Ba, 2014) for 50 epochs. We refer to the parameters obtained by solving the problem in Eq. (8) using the  $\mathcal{X}^{training}$  dataset (baseline case) as  $\theta^*$ . Similarly, we refer to the parameters obtained by solving the problem in Eq. (8) using the denoised  $\hat{\mathcal{X}}^{training}$  dataset with the NMF and the nnCP decomposition respectively as  $\theta_{NMF}^*$  and  $\theta_{nnCP}^*$ .

For each recording of the validation dataset  $\mathbf{X}^{valid} = \mathbf{X}_{:,n}^{valid}$  which can be either normal or abnormal, we compare the outputs of the three setups, referred to as 'Baseline', 'NMF' and 'nnCP', obtained by applying the decision function in

Eq. 9 with the following parameters:

$$\bullet \text{ 'Baseline' : } \mathcal{F}(\mathbf{X}^{valid}, \theta^*) \quad (10)$$

$$\bullet \text{ 'NMF' : } \mathcal{F}(\hat{\mathbf{X}}^{valid}, \theta_{NMF}^*) \quad (11)$$

$$\bullet \text{ 'nnCP' : } \mathcal{F}(\hat{\mathbf{X}}^{valid}, \theta_{nnCP}^*) \quad (12)$$

Please note that for the case 'NMF' and 'nnCP' the validation dataset  $\mathcal{X}^{valid}$  is denoised using the NMF or the nnCP decomposition respectively. The number of components  $K$  used is the same as for the denoising of the training dataset. NMF and nnCP are applied on all datasets independently.

The comparison between the decision obtained from all recordings of the validation dataset to the ground truth is made via area under the receiver operating characteristic curve (AUC). Each experiment is repeated 5 times with different weight initialisations for the neural network.

#### 4.2. Stationary case

We consider machines under stationary operation in the MIMII dataset: the fan and the pump. Figure 3 shows the impact of the denoising for three recordings of the Fan machine id 02. To illustrate the denoising performances, we choose empirically  $K = 10$  for the NMF and  $K = 20$  for the nnCP. The first recording doesn't present any particularities, we can see that even with a higher number of components  $K$  than for the NMF, the nnCP seems to remove more noise. The second recording contains different spectral motifs compared to the first one. In this case, the nnCP can retrieve every spectral motif, which is not the case for the NMF (a low-frequency motif is missing). Finally, the third recording presents an artefact delimited in time. This artefact is eliminated with the nnCP. Indeed, every event that is not present in a consequent number of recordings in the same period will be removed (if  $K$  is not too high). This demonstrates the advantage of nnCP to denoise stationary processes as it will remove one-off and temporally inconsistent events.

The Table 1 shows the mean of the AUC for the different fan and pump datasets available for the three applied approaches 'Baseline', 'NMF' and 'nnCP'. We consider dataset with a medium signal to noise ratio (SNR=0dB) and a low SNR (SNR=-6dB). For the 'NMF' and 'nnCP' approaches, we apply grid search  $K = [5, 10, 20]$  and only retain for each machine the best mean a posteriori based on the anomaly detection performance. We can observe that the best result is often obtained using  $K = 20$  for the nnCP. More refined results could be found by optimizing  $K$  based on a validation dataset. In every case, the nnCP provides a sensible improvement compared the baseline and the NMF method (except for the pump id 06 with SNR=-6dB). We have an average improvement of the AUC between the nnCP approach and the baseline of 0.13 for a SNR of 0 dB and 0,9 for a SNR of -6 dB. These results demonstrate the relevance of the nnCP denoising to lower the

Table 1. Mean of AUC for the 5 experiments performed with each stationary machine and with the 'baseline', 'NMF' and 'nnCP' procedures.

Machine	SNR	Baseline	NMF	nnCP
Pump-00	0 dB	0.62	0.67	<b>0.86</b>
	-6 dB	0.59	0.69	<b>0.74</b>
Pump-02	0 dB	0.49	0.72	<b>0.88</b>
	-6 dB	0.53	0.65	<b>0.73</b>
Pump-04	0 dB	0.94	0.97	<b>0.98</b>
	-6 dB	0.91	<b>0.93</b>	<b>0.93</b>
Pump-06	0 dB	0.81	0.83	<b>0.87</b>
	-6 dB	<b>0.65</b>	0.50	0.63
Fan-00	0 dB	0.63	0.57	<b>0.70</b>
	-6 dB	0.56	0.51	<b>0.61</b>
Fan-02	0 dB	0.87	0.87	<b>0.96</b>
	-6 dB	0.69	0.64	<b>0.77</b>
Fan-04	0 dB	0.77	0.72	<b>0.90</b>
	-6 dB	0.59	0.50	<b>0.68</b>
Fan-06	0 dB	0.99	<b>1.00</b>	<b>1.00</b>
	-6 dB	0.88	0.90	<b>0.97</b>
<b>Average</b>	0 dB	0.77	0.80	<b>0.90</b>
	-6 dB	0.67	0.66	<b>0.76</b>

complexity of the dataset for stationary machines.

Figure 4 provides three ROC curves obtained the three approaches It illustrates, for these example, that the nnCP denoising decreases at the same time the false-positive rate and the false-negative rate. We obtain similar improvements also for other machines.

#### 4.3. Non-stationary case

We consider now the non-stationary machines. We show the impact of the denoising in Figure 5 for three recordings of the machine valve id 00. As illustrated by the recording 1 and 2, the valve acoustic emission is characterised by sparse and periodic events. However, from one recording to another, the spike does not activate at the same time periods. Thus, the NMF can retrieve these spikes but not the nnCP. For the nnCP, the spikes are smoothed to account for several recordings that have spike activation approximatively at the same time. However, the quality of spike reconstruction is drastically impaired. For the last recording (corresponding to an abnormal sound), the spike does not activate periodically anymore, they are all removed by the nnCP method because this one-off configuration of the spikes is considered as an artefact.

The Table 2 shows the mean of the AUC for the different valve and slider datasets available for the three approaches 'Baseline', 'NMF' and 'nnCP'. We only consider datasets with

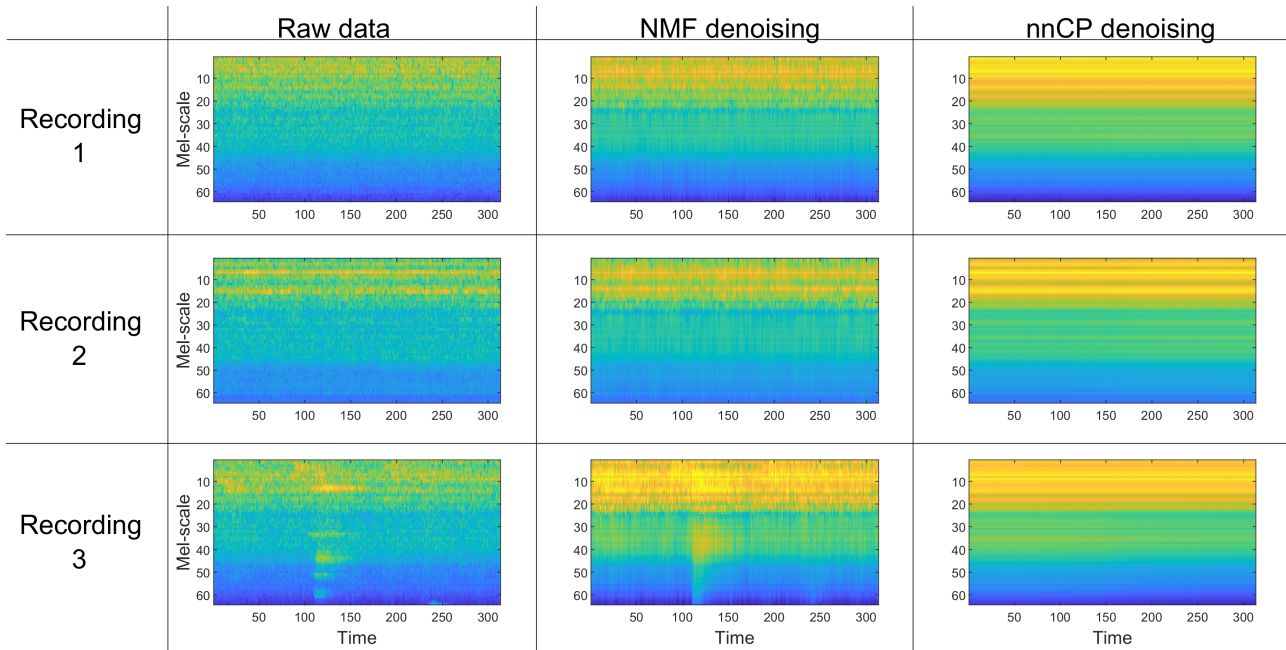


Figure 3. Table of figures where rows are three different recordings of the machine fan id 02. The columns show the influence of the denoising strategies with NMF and nnCP case for each recording.

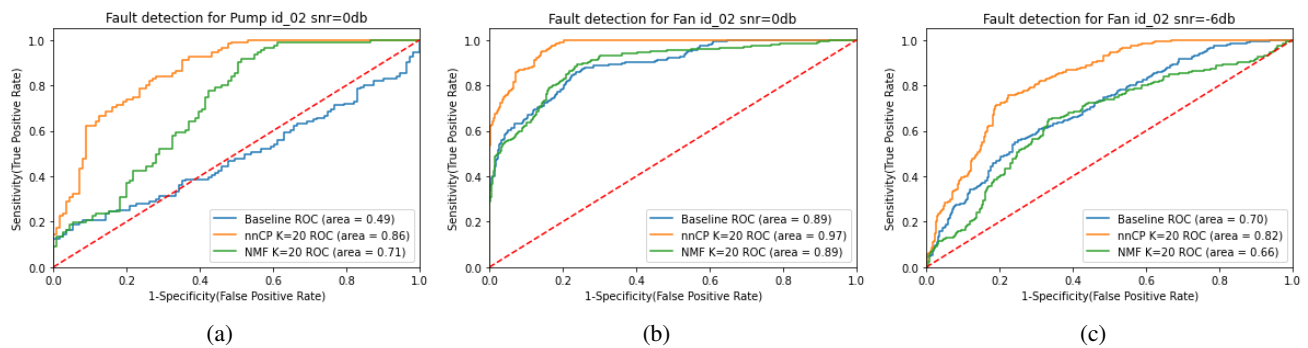


Figure 4. ROC curves for the 'baseline', 'NMF' and 'nnCP' denoising procedures for (a) the pump id 00 with 0dB (b) the fan id 02 with 0 dB and (c) the fan id 02 with 6 dB



Table 2. Mean of AUC for the 5 experiments performed with each non-stationary machine and with the 'baseline', 'NMF' and 'nnCP' procedures.

Machine	Baseline	NMF	nnCP
Valve-00	0.62	<b>0.69</b>	0.47
Valve-02	<b>0.59</b>	<b>0.59</b>	0.62
Valve-04	<b>0.65</b>	0.61	0.48
Valve-06	0.67	<b>0.76</b>	0.56
Slider-00	0.98	<b>1.00</b>	0.93
Slider-02	0.83	0.86	<b>0.90</b>
Slider-04	0.81	<b>0.82</b>	0.70
Slider-06	0.55	<b>0.62</b>	0.45
Average	0.71	<b>0.74</b>	0.64

medium SNR (SNR=0dB). These results demonstrate that the nnCP decomposition is not adapted for these kinds of non-stationary signals. However, most of the time, the NMF leads to an improvement of the baseline

### 5. CONCLUSION AND PERSPECTIVES

In this work, we focus on the importance of pre-processing the data before using a deep learning method. We propose two decompositions, the non-negative matrix factorisation (NMF) and the non-negative canonical polyadic decomposition (nnCP) as data-driven, quasi non-parametric modules for spectral data denoising prior to using a neural network. We show on an acoustic emission monitoring case study, the MIMII baseline, that nnCP reduces sensibly the complexity of the dataset, and considerably improves the MIMII baseline in the case of machines emitting stationary sound. However, the nnCP seems unsuitable when the sound is non-stationary.

Two main exciting research paths could be pursued from this work.

Concerning the limited applicability of the nnCP to non-stationary cases, a first approach would be to centre the acoustic emission of the machine for each recording. However, this pre-processing can be time-consuming. Another approach could be to explore more deeply the use of the NMF decomposition since it showed interesting results compared to the baseline for the non-stationary cases. Finally, an extension of the nnCP for the non-stationary case could be proposed by imposing the time mode of the decomposition to be shift-invariant (Tang, Chen, & Dong, 2014), (Kuang et al., 2015).

One of the main limitations of the method is the fact that after the training, a new recording must be denoised before applying the decision rule. Indeed, the network was trained on the denoised dataset, with a limited amount of noise and artefacts. Then, a large number of new recordings have to be regrouped to be effectively denoised with the nnCP decomposition. To

overcome this problem, an online version of the nnCP, where its components are updated each time a new recording is considered, could be proposed (Mairal, Bach, Ponce, & Sapiro, 2010).

### ACKNOWLEDGMENT

This study was supported by the Swiss Innovation Agency (Innosuisse) under grant number: 47231.1 IP-ENG.

### NOMENCLATURE

In order to facilitate the reading of the article, lower-case letter are used for scalars ( $x$ ), vector are in bold ( $\mathbf{x}$ ), matrix are in bold capital ( $\mathbf{X}$ ) and tensor are in bold calligraphic letter ( $\mathcal{X}$ ).  $f, t, n$  and  $k$  are indices and  $F, T, N$  and  $K$  denote their index upper bounds.

$\mathcal{X}$	Multi-recording time-frequency representation
$\mathbf{X}$	Time-frequency rep. of one recording
$\mathbf{x}$	Frequencies for one frame
$\mathbf{A}$	Spectral components
$\mathbf{B}$	Temporal components
$\mathbf{C}$	Recording components
$f$	frequency index
$t$	time (frame) index
$n$	recording index
$k$	component index
$\theta$	Parameters of the neural network
$E(\bullet)$	Encoder network
$D(\bullet)$	Decoder network
$\mathcal{L}(\bullet)$	Loss function
$\mathcal{F}(\bullet)$	Decision function

### REFERENCES

- Aggarwal, C. C., & Reddy, C. K. (2013). *Data clustering: algorithms and applications*. CRC press.
- Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2020). Fusing physics-based and deep learning models for prognostics. *arXiv preprint arXiv:2003.00732*.
- Chen, Y., Wei, S., & Zhang, Y. (2020). Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. *Medical & Biological Engineering & Computing*, 58(9), 2039–2047.
- Cichocki, A., & Phan, A.-H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, com-*

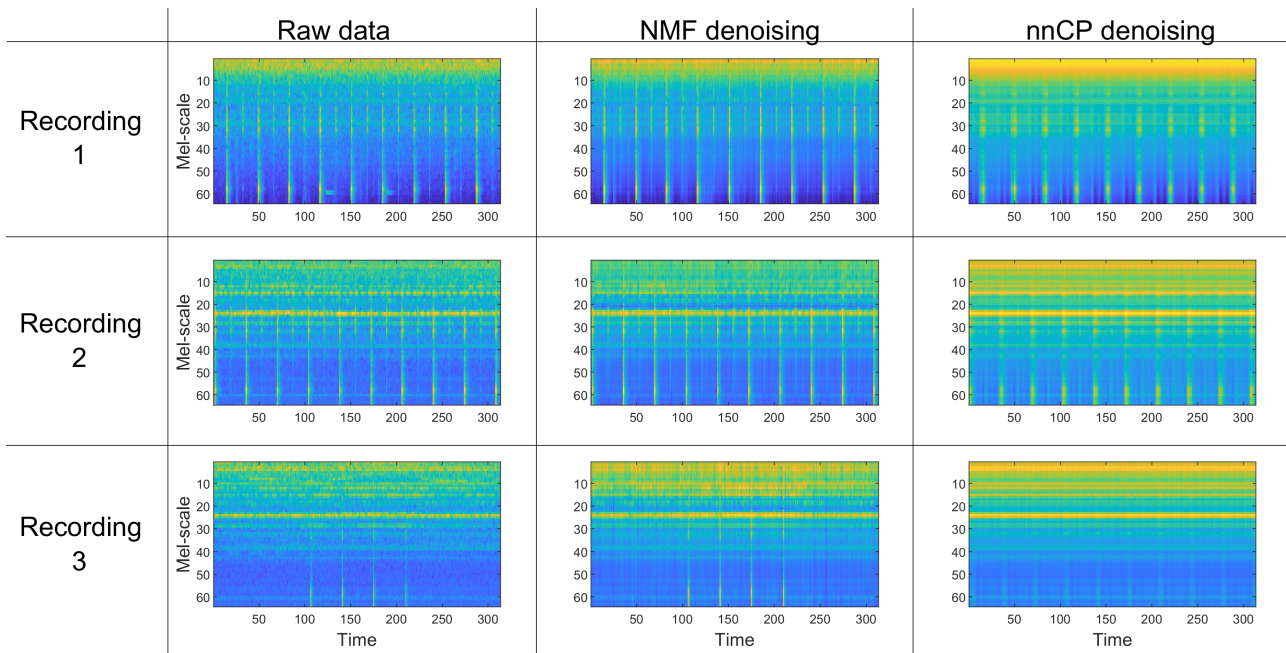


Figure 5. Table of figures where rows are three different recordings of the machine valve id 00. The columns show the influence of the denoising strategies with NMF and nnCP for each recording.

*munications and computer sciences*, 92(3), 708–721.

Cichocki, A., & Zdunek, R. (2007). Regularized alternating least squares algorithms for non-negative matrix/tensor factorization. In *International symposium on neural networks* (pp. 793–802).

Comon, P. (2014). Tensors: a brief introduction. *IEEE Signal Processing Magazine*, 31(3), 44–53.

Donoho, D. L., & Stodden, V. C. (2004). When does non-negative matrix factorization give a correct decomposition into parts?

Frusque, G., Borgnat, P., Gonçalves, P., & Jung, J. (2020). Semi-automatic extraction of functional dynamic networks describing patient’s epileptic seizures. *Frontiers in neurology*, 11.

Frusque, G., Jung, J., Borgnat, P., & Gonçalves, P. (2020). Multiplex network inference with sparse tensor decomposition for functional connectivity. *IEEE transactions on Signal and Information Processing over Networks*, 6, 316–328.

Hu, C., Wang, Y., & Bai, T. (2019). A tensor-based approach for identification of multi-channel bearing compound faults. *IEEE Access*, 7, 38213–38223.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Komatsu, T., Toizumi, T., Kondo, R., & Senda, Y. (2016). Acoustic event detection method using semi-supervised non-negative matrix factorization with a mixture of local dictionaries. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016)* (pp. 45–49).

Kuang, L.-D., Lin, Q.-H., Gong, X.-F., Cong, F., Sui, J., & Calhoun, V. D. (2015). Multi-subject fmri analysis via combined independent component analysis and shift-invariant canonical polyadic decomposition. *Journal of neuroscience methods*, 256, 127–140.

Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems* (pp. 556–562).

Lim, L.-H., & Comon, P. (2009). Nonnegative approximations of nonnegative tensors. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(7-8), 432–441.

Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(1).

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference* (Vol. 8, pp. 18–25).

Michau, G., & Fink, O. (2021). Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems*, 216, 106816.

Michau, G., Hu, Y., Palmé, T., & Fink, O. (2020). Feature learning for fault detection in high-dimensional condition monitoring signals. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 234(1), 104–115.

Michau, G., Thomas, P., & Fink, O. (2017). Deep feature



- learning network for fault detection and isolation. In *Phm 2017, st. petersburg, usa, 2-5 october 2017* (pp. 108–118).
- Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). Mimi dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. *arXiv preprint arXiv:1909.09347*.
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2), 206–219.
- Qi, Y., Comon, P., & Lim, L.-H. (2016). Uniqueness of nonnegative tensor approximations. *IEEE Transactions on Information Theory*, 62(4), 2170–2183.
- Stojanovic, L., & Dinic, A. (2016). Big-data-driven anomaly detection in industry (4.0): An approach and a case study. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1647–1652).
- Tagawa, T., Tadokoro, Y., & Yairi, T. (2015). Structured denoising autoencoder for fault detection and analysis. In *Asian conference on machine learning* (pp. 96–111).
- Tang, H., Chen, J., & Dong, G. (2014). Sparse representation based latent components analysis for machinery weak fault detection. *Mechanical Systems and Signal Processing*, 46(2), 373–388.
- Whitaker, B. M., Suresha, P. B., Liu, C., Clifford, G. D., & Anderson, D. V. (2017). Combining sparse coding and time-domain features for heart sound classification. *Physiological measurement*, 38(8), pp 1701–1713.