An Unsupervised Big Data Visualization-Based Scheme for Anomalous Sound Detection of Facilities

Sana Talmoudi (s.talmoudi@srd.mech.tohoku.ac.jp)
Tohoku University https://orcid.org/0000-0002-1321-0344
Tetsuya Kanada
D'isum Inc.
Yasuhisa Hirata
Tohoku University: Tohoku Daigaku

Research Article

Keywords: unsupervised predictive scheme, vibration/audio data, full spectrum, data visualization, generic, scalable, 0-training, white box

Posted Date: January 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-139087/v1

RESEARCH

An Unsupervised Big Data Visualization-Based Scheme for Anomalous Sound Detection of Facilities

Sana Talmoudi[1,2*], Tetsuya Kanada[2] and Yasuhisa Hirata[1]

Abstract

This paper is dedicated to validating the scalability and generalization of our previously proposed "machinery failure predictive scheme". Our aim is to have a generic core technology to provide a solution applicable in industry that is low-cost and low-intrusive.

Background: In our previous works, we proposed an unsupervised predictive scheme combining the use of full spectrum of vibration/audio data and data visualization techniques. We then proposed a real time data tracker (RTDT) and we applied our proposal on vibration data of bearings. In this paper, we are applying our predictive scheme on a facility (composite system) rather than a specific mechanical component (singular system). We chose to apply our proposal on the MIMII dataset as it was used in task 2 of the DCASE 2020 challenge for the detection of anomalous sounds given normal data only.

Methodology: We adopted two approaches: (1) the same scheme used in our application on bearing vibration data and (2) with a slightly modified approach where we apply a high pass filter (HPF) on the audio data to reduce the effect of the background noise. To effectively evaluate the accuracy of our scheme in detecting and recognizing anomalous sounds, we are comparing our results to the performance of the baseline system proposed by the organizers of the challenge as well as the results from the 40 participating teams. For the evaluation, we used the same metrics used in the challenge: the area under the receiver operation characteristic (ROC) curve (AUC) and the partial AUC (pAUC).

Results: We obtained satisfactory values of AUC and pAUC compared to the related works. We also outperformed the baseline system in 13 out of 16 machines in terms of AUC and 15 out of 16 machines in terms of pAUC.

Merits: Compared to the current related works, our "machinery failure predictive scheme" is featured by 0-training, and no complex preprocessing or de-noising techniques. Furthermore, our solution based on our scheme is provided as a white box, users will have the following merits: (1) our solution can be in operation right after a normal data set is obtained usually in a few days and (2) our solution can be built into conventional operation and maintenance systems without advanced background in artificial intelligence or data science.

Keywords: unsupervised predictive scheme; vibration/audio data; full spectrum; data visualization; generic; scalable; 0-training; white box

Introduction

A fault that we overlook and fail to detect can cost a life. Maintenance techniques have been developed since the birth of the first machines, to detect and recognize a faulty behavior and take the right action to avoid a catastrophe. Today's most common direction in maintenance is predictive maintenance, becoming one of the pillars of Industry 4.0 and permitting us to be few steps ahead of the fault. Condition Based Monitoring (CBM) is the most known predictive technique, that can be today translated in an IoT-based predictive solution.

For anomaly detection, recognition and prediction, literature is rich of approaches combining the use of

*Correspondence: s.talmoudi@srd.mech.tohoku.ac.jp
[1]Department of Robotics, Graduate Faculty of Engineering, Tohoku University, Aoba, Sendai, Japan
Full list of author information is available at the end of the article

vibration or/and audio as sensor data and the use of Artificial Intelligence (AI) based methods for data processing and decision making. On vibration data, we can cite as example, Windau and Itti [1] using Support Vector Machine (SVM) and Neural Network (NN), Kankar et al. [2] using Artificial Neural Network (ANN) and SVM, Zhang et al. [3] using ANN and Galloway et al. [4] using Auto Encoder (AE) and SVM. On audio data, we can cite as example, Koizumi et al. [5] using Gaussian Mixture Model (GMM) and Variational Auto Encoder (VAE), Kawaguchi et Endo [6] using Long-Short-Term Memory (LSTM) autoencoder, Oh and Yun [7] and Koizumi et al. [8] using AE.

However, it remains challenging for the AI-based techniques to be applied for machinery failure prediction despite achieving outstanding results in other fields of application such as speech recognition and computer vision. Supervised methods suffer from the lack of training data and the absence of real failure data since the machines are initially designed for a long lifetime. Unsupervised methods on the other hand achieved good results using normal data only for training the classifiers and models to recognize and detect abnormal data. But these methods, relying on a specific set of extracted features, can rarely be generic and scalable and has to be trained to specific machines and sensor data to achieve highly accurate results. Moreover, there is still a big gap between industry and AI despite the popularity of these approaches. It is usually required to have data science background to be able to deploy, understand and effectively interpret the obtained results. In [9], Jardine et al. mentioned that there are the two main difficulties with neural networks: the difficulty to obtain physical explanation of the trained model and the difficulty of the training process. For some factories, it is still considered intrusive and costly to hire data scientists and to implement these solutions in the first place.

Therefore, we are aiming for a scheme that is generic and scalable to provide industry with a solution that is low-cost and low-intrusive. In other words, a solution that doesn't need to be re-designed or re-developed for each machine/sensor type and that can be used with the existing monitoring system already installed in the factory/plant. We are also aiming to provide a solution that can be fully comprehended even without an advanced background in data science.

In this context, we proposed a novel approach for machinery failure diagnosis and prediction featured by 0-training and by quick start after installation. In our previous work [10] we proposed the backbone of our analysis scheme. We use the full spectrum data, without any specific feature extraction, combined with visualization techniques such as toorPIA [11] and t-SNE [12]. The visualization technique maps the highly dimensional data vectors (spectra) in a 2D map according to the machine behavior. The philosophy of our proposal is to ensure the simplicity of the process with less possibility to overlook any important information and to provide a human-friendly and human-understandable output. Additionally, by having 0-training, the solution would adapt to any type of machine without the need to re-develop or re-design it.

In [10], we were able to validate the ability of our analysis scheme to recognize abnormal data from normal data using sound data from a miniature toy motor where we mimicked failure states. However, our main goal is not a solution that only detects and recognizes anomalous data from normal data but further a real time data tracker that can track and visualize the signs of degradation to anticipate the anomalies. Therefore, in [13], we extended our proposal to be applied as a predictive tool and we proposed the Real Time Data Tracker (RTDT). The RTDT consists of first, generating a reference map (RM) from the early acquired data (the reference data), supposedly to be the healthy (normal) data. To generate the RM, we use the same analysis scheme as proposed in [10]. Then, as we monitor the target machine/component, we acquire new data and using a geometry-based algorithm, we add these data (the monitoring data), into the RM in real time. By checking the movement of the newly added data on the RM, we track and asses the machine condition. In [13], we were able to validate the RTDT on a run-to-failure test vibration dataset on bearings, provided by the Intelligent Maintenance Systems (IMS), University of Cincinnati [14, 15] .

In this paper we are not proposing a novel analysis scheme but instead verifying the scalability and generalization of our machine failure predictive scheme. Thus, in this work we are extending further our field of application to a composite system rather than a singular system as in [13]. We chose to apply our predictive scheme on valves, pumps, fans and slide-rails as composite systems from the MIMII dataset [16]. Each machine type in the MIMII dataset has different operational mode [17] rather than rotating at a constant speed as the bearings used in [13]. This made the abnormal data more challenging to be detected and recognized from the normal data compared to our work in [13]. Additionally, in the MIMII dataset, the monitoring data is audio data recorded by a contactless microphone and mixed with a background noise rather than vibration data recorded by an accelerometer directly attached to the shaft where the bearings were mounted. This made the sensor signal more complex to find the key information than what we used in [13].

Despite these changes in the target machines and the monitoring data, we are keeping the core technology without any fundamental changes to test (1) how scalable our scheme is by validating on composite system and (2) how generic our scheme is by validating on several machine types (valves, pumps, fans and slide-rails) and on different sensor data (audio data rather than vibration data). We are presenting our results from two adopted approaches:

- **Approach 1**: the use of the scheme exactly as in [13].
- **Approach 2**: the use of a High Pass Filter (HPF) on the input audio data to improve the signal and reduce the effect of the background noise.

To effectively evaluate the accuracy of detection our scheme, we are comparing our results to the related works on the MIMII dataset. The MIMII dataset was used in task 2 from the IEEE AASP Challenge on Detection and Classification for Acoustic Scenes and Events, DCASE 2020. This task is known as the Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring. The task rule was to use the normal data only as training data and do not include any abnormal data instance in the training phase. Two metrics were used for ranking the participating teams: the area under receiver operating characteristic (ROC) curve AUC and the partial AUC, pAUC [18]. We then calculated the AUC and pAUC from our results to quantitively evaluate our results and position our scheme among the existing works.

Our achievements in this paper are as follow: (1) from the viewpoint of the solution generalization, our contribution is that not only our method can be used on both complex and singular systems, but it can be also used with different types of machines and with different types of sensor data. We used the same anomaly score formula and same analysis method for all the 4 machine types and all the machine IDs. (2) As a comparison to the results of task 2 of DECASE 2020 in terms of accuracy of detection of anomalous sound data, we obtained satisfactory values of AUC and pAUC compared to the related works. We outperformed the baseline system in 13 out of 16 machines in terms of AUC and 15 out of 16 machines in terms of pAUC. Additionally, it should be mentioned that contrarily to the current related works, our scheme is 0-training, no complex preprocessing nor de-noising techniques. This makes our solution reproducible and generic. The solution is a white box, simple to understand and use by operators, even with the absence of data science background. This permits more effective interpretation of the results leading to a better prediction, less false alerts and more effective maintenance.

Moreover, our solution is originally designed not only for detection of anomalous events but also to track the degradation before reaching the anomalous stage. We were able to prove this specific feature in our work on the IMS dataset [13] since it was a run-to-failure test data. However, in the MIMII dataset, the anomalies were deliberately caused in the machines and therefore, we only applied our scheme to detect anomalous sound data from normal data rather than a real-time data tracker.

The rest of the paper will start by brief recapitulation of the core technology and methodology of our predictive scheme. We are then explaining how we used the MIMII dataset to validate our scheme. Next, we are detailing our analysis steps while in the section that follows, we are showing our results. In the next section we are presenting the evaluation of the obtained results. In a later section, we are discussing further the added value of the work presented in this paper compared with our previous works. We finally conclude our work and present our future works.

Methodology

In [10] we presented the backbone of our solution for machinery failure detection and prediction. We then proposed the RTDT as a real time data tracker in [13]. Our proposal is an unsupervised predictive method consisting of two main steps. (1) Generation of the reference map RM from the early acquired data (reference data) and supposed to be the healthy (normal) data. Then (2) adding the newly acquired data (the monitoring data) in real time into the RM for tracking and assessment of the machine condition. **Fig.1** illustrates the process flow of the proposed predictive scheme.

In the next sections, we will use the following notations:

- We refer to a vector U from the origin in the multi-dimensional space by $\boldsymbol{U}$, where $\boldsymbol{U} \in \Re^n$
- We refer to a vector U from the origin in the two-dimensional plane by $\boldsymbol{u}$, where $\boldsymbol{u} \in \Re^2$
- We refer to the vector formed by a point A and a point B in the multi-dimensional space by $\overrightarrow{AB}$
- We refer to the vector formed by a point A and a point B in the two-dimensional plane by $\overrightarrow{ab}$
- We refer to the inner product between a vector U and a vector V by $\langle \boldsymbol{U}, \boldsymbol{V} \rangle$

Generation of the RM

Preprocessing phase: on the reference data, we convert the raw data into reference data vectors.

We start by dividing the time-domain data into segments. Next, on each segment we apply a window function, the Hanning window, and we then apply the Fast Fourier Transform (FFT). We only consider the components of the positive frequencies. The obtained spectrum from each data segment is a multidimensional

data vector $\boldsymbol{V}$. In certain cases where the sampling rate is high, we smoothen the spectrum to reduce the white noise effect.

In our scheme, we do not extract any specific feature but instead we consider the whole frequency contents as a high dimensional input vector to minimize the possibility of overlooking any important information within the spectrum [13].

The preprocessing phase results in transforming the raw reference data to a set of high dimensional reference data vectors $\{\boldsymbol{R_i}\}_1^{N_r}$ where $\boldsymbol{R_i} \in \Re^n$, N_r is the number of reference data vectors and n is the number of dimensions, i.e. frequency components of the spectrum.

Similarity quantification: on the reference data vectors, we quantify the similarity between the data vectors. We use the distance in the multidimensional space as shown by (1) to quantify the similarity between a data vector $\boldsymbol{R_i} \in \Re^n$ and a data vector $\boldsymbol{R_j} \in \Re^n$.

$$d_{ij} = \sqrt{\langle \boldsymbol{R_i}, \boldsymbol{R_i} \rangle + \langle \boldsymbol{R_j}, \boldsymbol{R_j} \rangle - 2\langle \boldsymbol{R_i}, \boldsymbol{R_j} \rangle} \tag{1}$$

Where:

- $\boldsymbol{R_i}, \boldsymbol{R_j} \in \Re^n$, $\boldsymbol{R_i} = (r_{i1}, r_{i2}, \cdots, r_{in})$, $\boldsymbol{R_j} = (r_{j1}, r_{j2}, \cdots, r_{jn})$ $\forall i, j \in [1, N_r]$
- $\langle \boldsymbol{R_i}, \boldsymbol{R_j} \rangle$ is the inner product of $\boldsymbol{R_i}$ and $\boldsymbol{R_j}$

From the distance values between all the reference data vectors, we form a similarity matrix S. S is an $n \times N$ matrix and can be simplified to an upper triangular matrix given the following characteristics: (a) $d_{ij} = d_{ji}$ and (b) $d_{ii} = 0$. The similarity matrix S has the format shown by (2).

$$S = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ 0 & 0 & d_{13} & \cdots & d_{2n} \\ 0 & 0 & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{2}$$

Data visualization: based on the obtained similarity matrix S, the high dimensional reference data vectors are mapped in a two-dimensional plane using a dimension reduction technique such as toorPIA [11] and t-SNE [12]. Dimensional reduction techniques tend to translate the similarity seen in the multidimensional space to clusters in the two-dimensional plane while preserving as much of the significant structure of the high-dimensional data as possible in the low-dimensional map [12]. The differences among the existing methods relies in the definition of the significant structure to preserve. In [13], we discuss further the differences between the dimensional reduction techniques and the reason for choosing toorPIA and t-SNE for this field of application.

Therefore, for each $\boldsymbol{R_i} \in \Re^n$ we then assign a $\boldsymbol{r_i} \in \Re^2$ where $\boldsymbol{r_i} = (r_{ix}, r_{iy})$. We then obtain $\{\boldsymbol{r_i}\}_1^{N_r}$, the set of the N_r reference data vectors in the two-dimensional plane forming the RM.

The RM, formed by the early acquired data, would consist of the normal (safe) zone.

Adding the new data into the RM

Preprocessing phase: on each newly acquired data (test data), we start by generating our data vectors from the raw data following the same exact preprocessing phase as in the generation of the RM.

Data plotting: as we mentioned in [13], there are many approaches that can be implemented to plot the new data into the RM. We are employing a geometry-based algorithm as follow:

- We consider G, the center of gravity of the RM, as origin for our plotting algorithm.
- For each X of the data vectors generated by the preprocessing phase from the newly acquired data (test data):
  1. We define $\boldsymbol{Z} \in \Re^n$ as the closest reference data to $\boldsymbol{X} \in \Re^N$ by distance in the multidimensional space as defined by 1.
  2. We define $\boldsymbol{Y} \in \Re^n$ as the closest reference data to $\boldsymbol{X} \in \Re^n$ by angle from G, i.e. forming the smallest angle, θ, with X from G.
  3. Using the angle, θ obtained from Y and the reduction ratio r obtained from Z, we obtain $\boldsymbol{x} \in \Re^2$, $x = (x_x, x_y)$. We then place the newly acquired data X into the RM as shown in **Fig.2**.

Decision making

If the machine is still healthy, the newly acquired data will fall into the safe zone. However, if the machine is degrading, the data points move out of the normal (safe) zone. This permits to track the machine condition by visualization. Additionally, we are quantitively tracking the machine condition by calculating a warning factor, ρ, from the two-dimensional position of the newly plotted data. From $\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g} \in \Re^2$ the 2D position of each newly acquired data X, the 2D position of Z, the closest reference data to $\boldsymbol{X} \in \Re^n$ by distance in the multidimensional space and the 2D position of G, the gravity center of the RM, we define $\rho(X)$ as given by (3).

$$\rho(X) = \frac{\|\overrightarrow{gx}\|}{\|\overrightarrow{gz}\|} \tag{3}$$

The warning factor $\rho(X)$ has the following characteristic: $\rho(X) = 1 + \epsilon$ where:

- ϵ would take a large positive value when the data vector $\boldsymbol{X}$ corresponds to anomalous data.
- ϵ would take a small or negative value, leading to a value of ρ almost equal to 1 when the data vector $\boldsymbol{X}$ corresponds to a normal data.

Therefore, if $\rho(X)$ takes a large value, the machine condition is abnormal. Else if $\rho(X)$ takes a small value (almost 1), the machine is normal. In terms of tracking the machine condition, if $\rho(X)$ is increasing, the machine is degrading and will have anomaly. Else if $\rho(X)$ is maintained around 1, the machine is maintaining a healthy condition.

Experimental material

In this paper, we chose to apply our proposed predictive scheme on the MIMII dataset [17] but as it was used in the task 2 of the DCASE 2020 challenge [19]. This task is known as the Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring. The aim of the task is to have solutions that are not limited to simple classification but instead are able to have a "prompt detection" of degradation by observing the machine sounds. Hence, the task rule was to use the normal data only as training data and do not include any abnormal data instance in the training phase. The task used two datasets: the MIMII dataset [17] and the ToyADMOS dataset [20]. However, in this work, we chose to focus on the MIMII dataset, since it consists of the sound data of real machines with background environment conditions. By using this dataset, we can also have a wide range of related works results to compare our results to and to effectively verify the validity of our proposal.

Data characteristics

The MIMII dataset in task 2 of the DECASE 2020 challenge consists of audio data that had been recorded from 4 types of machines: valve, pump, fan and slide-rail. Each machine type had 4 monitored machines. Each of the machine types had different operational modes [17]. All data were provided as 10-sec-long segments and down-sampled to 16 kHz [19]. Originally, the data were acquired on an 8-channel microphone [17] but for the sake of the simplicity in the task, only channel 1 data were considered [19]. Additionally, in the set considered in the task 2 in the DECASE 2020 challenge, the audio data consist of both, the monitored machine sound as well as the background noise [19]. Further details on the full MIMII dataset can be found in [17] and on the dataset as used in the task 2 of DECASE 2020 challenge in [19].

Data structure

The MIMII dataset as used in task 2 of DECASE 2020 challenge was presented in the form of 3 sets:

- Development dataset:
  - Training set: as training data, only normal data, labeled, intended to be used to train classifiers and models for learning-based and training-based methods.
  - Test set: as test data, normal and anomaly data, labeled, intended to be used to test the trained classifiers and models. This set is not allowed to be used in the training phase by rule of the organizers.
- Evaluation dataset: as test data, normal and anomaly data, not labeled.
- Additional training dataset: as training data, can be used for training however this set was open later in the challenge.

In the current work, we are focusing on the development dataset and the data is used as described in **Table 1** and as follow, for each target machine:

- The data from the training set are used to make the reference map, RM. As we detailed in the previous section, no training is performed despite the name of the set. Therefore, we refer to this data as the reference data.
- The data from the test set is used as test data and is to be plotted into the RM. We refer to this data as the test data.

Moreover, in the development dataset, we have the data of the 4 types of machines: valve, pump, fan and slide-rail (slider). And for every machine type, we have 4 machine IDs: 00, 02, 04 and 06. Then, in total, we are testing our predictive scheme on 16 machines. Therefore, we can verify ability to recognize the anomalous sounds by having normal data only as reference and over different machine types. Consequently, we can verify the scalability and generalization of our scheme.

Analysis

We are analyzing each machine type and each machine ID. Therefore, the following analysis process has been conducted over the 16 analysis cases.

Preprocessing phase (vectors generation)

On both, the reference data and the test data from the development set, we generate the data vectors as follow:

1 We consider each file (audio clip) as a segment. Therefore, a segment was 10-sec-long and contained 220500 sample audio data as shown in **Fig.3.a**.

2 We add dummy data to reach a segment length equal to $2^{18} (= 262144)$ i.e. we add 41644 dummy data, each equal to 0.

3 We apply Hanning window.
4 We perform the FFT: the original spectrum size is 131072 and the original frequency resolution of the spectrum is equal to 0.08 Hz as shown in **Fig.3.b**.
5 The original frequency resolution of the spectrum being high, we smoothen the spectrum by averaging every 256 consecutive frequency components. This also reduces the white noise effect. We obtain a spectrum of size down to 512 with an augmented frequency resolution of 19.58 Hz as shown in **Fig.3.c**.
6 We crop the spectrum to 8 kHZ since the original signal has been down sampled to 16 kHz and the original spectrum was found equal to 0 after 8 kHz. For each data vector, the number of dimensions became equal to 372.

The preprocessing phase resulted then in two sets of multidimensional data vectors:

1 $\{\boldsymbol{R_i}\}_1^{N_r}$ where $\boldsymbol{R_i} \in \Re^{372}$, N_r is the number of reference data vectors.
2 $\{\boldsymbol{X_i}\}_1^{N_x}$ where $\boldsymbol{X_i} \in \Re^{372}$, N_x is the number of test data vectors.

We used the $\{\boldsymbol{R_i}\}_1^{N_r}$ to generate the RM the we added each of the $\{\boldsymbol{X_i}\}_1^{N_x}$ into the RM. We evaluated the anomaly score of each X_i as described in the Methodology section.

The parameters used in the preprocessing phase such as the segment length, the spectrum size and the frequency resolution, were kept the same among all the machine types and IDs to verify how general our scheme can be.

RM generation

On the reference data vectors we had the following steps:

Similarity quantification: to define the similarity between two data vectors, we evaluated the distance in the multi-dimensional space as given by (1). In this work, we are considering the data vectors in 2 approaches:

- **Approach 1, *FFR_approach*** (**F**ull **F**requency **R**ange approach): we used the full spectrum range, $0 \sim 8$ kHz. We then had 372 dimensions, and each audio clip was considered as a data vector $\boldsymbol{R_i} \in \Re^{372}$.
- **Approach 2, *HPF_approach*** (**H**igh **P**ass **F**ilter approach): we used a High Pass Filter (HPF) to reduce the effect of the background noise. The covered frequency range was 1.5 kHz $\sim$ 8 kHz. We then had 302 dimensions and each audio clip was considered as a data vector $\boldsymbol{R_i} \in \Re^{302}$.

The structure of the adopted multi-dimensional space for the RM of each of the 16 cases, in both approaches is summarized by **Table 2**.

For each of the 16 cases, we generated the similarity matrix S as given by (2).

Dimension reduction and visualization: we performed the dimension reduction from $\Re^n$ to $\Re^2$. In the current work, we mapped the high-dimensional data vectors in a two-dimensional plane using toorPIA as a dimensional reduction technique. We then obtained the RM. The position of the high-dimensional data vectors in the RM was defined based on the multidimensional distance between them given in the similarity matrix S.

By the end of this step, we generated 16 RMs from the dedicated reference set as described in **Table 1** and using the exact same steps and same analysis parameters.

As we discussed in [13], t-SNE is an alternative dimension reduction technique. However, in the current work we chose toorPIA to maintain the core analysis of the predictive scheme similar to our work in [13].

Adding the test data into the RM

On each of the 16 RMs, we added the corresponding test set as described in **Table 1** . For the 16 analysis cases, we used the same geometry-based algorithm as described in the Methodology section.

Anomaly score calculation

To quantitively describe the state of a test sound data X, an anomaly score is calculated for X. By definition, an anomaly score takes a large value when the input signal seems to be anomalous, and vice versa [19].

In the MIMII dataset, the anomalies were deliberately caused in the machines. Therefore, we are only applying our scheme to detect and recognize anomalous sound data from normal data rather than real-time tracking of the machines condition. However, the previously defined warning factor can be regarded as an anomaly score. The same formula given by (3) can be used to define the anomaly score $\rho(X)$ for each test data X. If $\rho(X)$ takes a large value, the machine condition is abnormal. Else if $\rho(X)$ takes a small value (almost 1), the machine is normal.

For every test data vector X, we then evaluated the anomaly score $\rho(X)$ using the formula in (3). The same anomaly score formula was used for all the 16 analysis cases.

Results

The RMs structure

The obtained RMs for all the 16 analysis cases using the $FFR\_approach$ are shown in **Fig.4**. The RMs for all the machines presented somewhat a similar distribution. The reference (normal) data distributed into dense subclusters surrounded by scattered data. We

then inspected the frequency contents of these areas on the RM to understand the reason of such distribution.

In **Fig.5**, we are presenting the frequency contents of the subclusters as well as the scattered points for valve 00, pump 00, fan 00 and slider 00 as examples of each machine type. From the spectrum of the segments of each area on the RMs, we can see that:

- The core area had smooth spectrum (**Fig.5.v-1**, **Fig.5.v-2**, **Fig.5.p-1**, **Fig.5.p-2**, **Fig.5.p-3**, **Fig.5.f-1**, **Fig.5.f-2**, **Fig.5.f-3**, **Fig.5.s-1**, **Fig.5.s-2** and **Fig.5.s-3**).
- The fringe area formed by:
  - Spectra with peaks at around 1 kHz (**Fig.5.v-3**, **Fig.5.v-4**, **Fig.5.v-5**, **Fig.5.v-6**, **Fig.5.p-4**, **Fig.5.p-5**, **Fig.5.p-6**, **Fig.5.f-4**, **Fig.5.f-5**, **Fig.5.f-6**, **Fig.5.f-7**, **Fig.5.f-8**, **Fig.5.f-9**, **Fig.5.f-10**, **Fig.5.s-4**, **Fig.5.s-5**, **Fig.5.s-6**, **Fig.5.s-7** and **Fig.5.s-8**).
  - Spectra with high amplitude over all the frequency components (**Fig.5.v-7**, **Fig.5.p-7**, **Fig.5.f-11** and **Fig.5.s-9**).

Since the data was recorded with a background noise (on purpose), these peaks can be related to the background noise effect. Moreover, the structure of the RM is the same among the different types of machines. Therefore, we cannot explain the subclusters by the different operational mode of the machine at the normal status. Instead, it can be explained by the background noise since for all the 4 types of machine, the sound data was recorded with the same background noise.

These findings inspired us to apply an HPF on the audio data, limiting the low frequencies to 1.5 kHz. We can then reduce the effect of the background noise on the results. The obtained RMs for all the 16 analysis cases using the *HPF_approach* are shown in **Fig.6**. The maps became more homogeneous forming one dense area with few data vectors scattered around and having less complex structure. The background noise effect, causing the normal data to be heterogenous, was then decreased.

Decision making on the test data

We added the test data into the RMs generated by both approaches, to compare their performances. From the position of the newly plotted test data into the RM, we calculated the anomaly score for each test data. The distribution of the values of the anomaly scores, with both *FFR_approach* and *HPF_approach*, is given in **Fig.7**, **Fig.8**, **Fig.9** and **Fig.10**, for all the 4 valve IDs, all the 4 pump IDs, all the 4 fan IDs and all the 4 slide rail IDs, respectively.

For some machines, *HPF_approach* was better than the *FFR_approach* in differentiating between the anomaly scores of the normal data and the abnormal data. In these cases, the high frequency region contained valuable information to distinguish the anomalies, for example the valve case in **Fig.7**. But for some machines, only the low frequency range was significant such as the fan case in **Fig.9**. Therefore, by filtering the signal, we lost key information to detect anomalies and the difference between the anomaly scores of the abnormal data and normal data was decreased.

Evaluation

Background on the used metrics

To evaluate the performance of our predictive scheme, we used the two metrics: the AUC and the pAUC. The AUC refers to the area under the receiver operation characteristic (ROC) curve (AUC) and pAUC refers to the partial AUC (pAUC). The ROC graph has been known to be used to evaluate the performance of a classifier. A classifier, mapping instance data to prediction classes, produces true positives and false positives. True positive corresponds to when the instance is positive and predicted to be positive. False positive means an instance that is negative but was predicted to be positive. Accordingly, the true positive rate (TPR) means the hit rate while the false positive rate (FPR) means the false alerts rate. TPR and FPR are given by (4) and (5), respectively. As a matter of fact, the ROC space has its x-axis as the FPR and its y-axis as TPR. Therefore, the ROC graph represents the tradeoff between the hit rate and the false alerts [18].

$$TPR = \frac{\text{positive instance predicted as poistive}}{\text{total number of positives}} \quad (4)$$

$$FPR = \frac{\text{negative instance predicted as poistive}}{\text{total number of negatives}} \quad (5)$$

A discrete classifier has as its output only as one class label (normal or abnormal) and is presented by a point in the ROC space having a unique (FPR, TPR) pair as shown in **Fig.11.a**. On the other hand, a non-discrete classifier, gives as output a score (anomaly score) expressing the degree to which an instance is a member of a class [18] (rather than a yes or no decision as in the discrete case). The anomaly scores are sorted and used as score threshold. An instance is positive (abnormal) if the score is higher than the threshold. Else, if the score is lower than the threshold, the instance is considered negative (normal). We then compute the TPR and FPR for every threshold. Therefore, a non-discrete classifier is presented by an ROC curve. In **Fig.11.b**, we are giving an example for such case.

In the case of discrete classifier, to compare a classifier C1 to a classifier C2, we compare their position in the ROC space as shown in **Fig.12.a**. C1 is to northwest of the curve from C2, therefore C1, having higher TPR and lower FPR, is better than C2. In the case of non-discrete classifier, we need to reduce the ROC curve down to a single value, the AUC [18]. The AUC takes then a value between 0.0 and 1.0 and thus, can be expressed in percentage. AUC is considered as the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [18]. Therefore, to compare a classifier C1 to a classifier C2, we compare their AUC values as shown in **Fig.12.b**. The AUC of C1 is higher than the AUC of C2, therefore C1 is better than C2.

In the current case, and as used in task 2 of the DECASE 2020 challenge [21], the anomaly scores of normal test samples are used as the threshold. According to [19], the AUC is given by (6):

$$AUC = \frac{1}{N_- N_+} \sum_{1 \le i \le N_-} \sum_{1 \le j \le N_+} H(A(x_j^+) - A(x_i^-)) \tag{6}$$

Where:

- $\{x_j^+\}_{j=1}^{N_+}$ and $\{x_i^-\}_{i=1}^{N_-}$: the abnormal and normal test data vectors, respectively, and had been sorted that their anomaly scores are in descending order.
- $A(x_j^+)$ is the anomaly score of an abnormal test data x_j^+.
- $A(x_i^-)$ is the anomaly score of a normal test data x_i^-.
- $H(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{else.} \end{cases}$
- N_+ and N_-: the number of normal and abnormal test data, respectively

The pAUC was additionally required for practical requirements [19]. The pAUC by definition, is calculated as the AUC over a low false-positive-rate (FPR) range $[0, p]$. For a predictive solution, it is important to increase the TPR under low FPR conditions to avoid false alerts resulting in an unnecessary intervention. The pAUC is given by (7), according to [19]:

$$pAUC = \frac{1}{\lfloor pN_- \rfloor N_+} \sum_{1 \le i \le \lfloor pN_- \rfloor} \sum_{1 \le j \le N_+} H(A(x_j^+) - A(x_i^-)) \tag{7}$$

Where $\lfloor . \rfloor$ is the flooring function and p is set to 0.1.

Our results metrics

To effectively compare the AUC and pAUC of our system to the related works submitted to the task 2 of DECASE 2020 challenge, we calculated our AUC and pAUC from the obtained anomaly scores, using the same implementation by sklearn [22] as used in the baseline system [23] proposed by the organizers of the challenge. We evaluated the AUC and pAUC for each machine ID and each machine type and the values are given in **Table 3**.

Using the values obtained for each machine ID, we were able to compare our results to the baseline system presented in [19] over the 16 machines individually. **Fig.13** is a comparative figure between our results and the baseline system results in terms of AUC and pAUC. In total, our approaches outperformed the baseline system in 13 out of 16 machines in terms of AUC and 15 out of 16 machines in terms of pAUC. The *FFR_approach* outperformed the baseline system in 11 out of 16 machines in terms of AUC and 14 out of 16 machines in terms of pAUC. On the other hand, the *HPF_approach* outperformed the baseline system in 11 out of 16 machines in terms of AUC and 11 out of 16 machines in terms of pAUC. By averaging the obtained AUC and pAUC values for all machine IDs, and then for all machine types, our results are considered better than the baseline system results for the development set.

Using the average AUC and pAUC values for each machine type, we were able to position our results among the results of the participating teams, including the baseline system. **Fig.14** shows the comparison between our results and the results of the participating teams [19, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62] on each machine type, in terms of AUC and pAUC. **Fig.15** shows the overall position of our results compared to the results of the teams, by averaging the values of AUC and pAUC from all machine types given that achieving high AUC and pAUC for all machines is considered important [19].

The process evaluation

Compared to the baseline system and to the submissions in the challenge, our method is the only method that did 0-training. The rest of the submissions did train the anomaly score so that the AUC and pAUC are maximized. This training phase is nevertheless part of the challenge and done by only normal data. Additionally, several methods applied de-noising techniques and complex filers. In our work, we did not use any complex de-noising technique. Moreover, several methods had different system configurations for each machine type to optimize the anomaly scores and maximize the metrics. In our work, we used the exact same

scheme among all the 16 target machines with the same parameters.

Therefore, the achieved results in AUC and pAUC values, are satisfactory given that we had no training, no complex preprocessing phase and no dependency on the target machine. Instead, we kept the process as simple as possible to be fully understood and better interpreted by operators even without a data science background to have a solution that can be reproducible and thus used in industry. It is true that better results can be obtained by applying more sophisticated de-noising techniques or further tuning. However, to be applied and integrated in industry, a compromise should be made between the complexity of the proposed system, its understandability and the accuracy of detection. By the current work, we believe that our proposal satisfies these criterias.

Consequently, our solution, based on our scheme, is provided as a white box and has following advantages: (1) it can be in operation right after a normal data set is obtained usually in a few days and (2) it can be built into conventional operation and maintenance systems without advanced background in artificial intelligence or data science.

Discussion

In this paper, we did not propose a novel predictive scheme, but instead we applied our predictive scheme on composite and more complex systems compared to our previous work [13]. Our goal is to validate the scalability and generalization of our solution and its applicability to industry. In our previous work [13], we were able to achieve an accurate anomaly detection, tracking and prediction using vibration data from bearings as singular systems. In the current work, using the same predictive scheme as in [13], we were able to achieve an accurate anomaly detection and prediction using the audio data from valves, pumps, fans and slide-rails as composite systems.

Table 4 summarizes the differences between the dataset analyzed in [13] and the dataset analyzed in the current work. On the other hand, **Table 5** summarizes the analysis parameters and process for both works.

Despite having different data source, different data characteristics, different complexity level of the target monitored machine, we were able to maintain the same core technology, and not re-design or re-develop our method to adapt to the challenges raised by these differences. Furthermore, we used the original measurements length as segment length.

In the [13], we applied our scheme on the IMS bearing dataset [14, 15]. The dataset is a run-to-failure test data. Therefore, we had two research directions: to differentiate between anomalous data and normal data and we also implemented our proposal as a real time data tracker (RTDT) to detect early signs of degradation and predict failure. In the MIMII dataset, the anomalies were deliberately caused in the machines and therefore we applied our scheme only to detect anomalous sound data from normal data rather than a real-time tracker. Theoretically, if the machines in the MIMII dataset were degrading naturally rather than deliberately damaged, the data would move gradually out of the normal (safe) zone and the warning factor would increase as we showed in [13]. But such feature cannot be verified given the current data. However, in both works, we effectively detected and recognized the anomalies, given normal data only as reference and without any training.

To be also noted, the dimension reduction and visualization technique used in both works, is toorPIA. However, as we discussed in [13], t-SNE can be also used instead, with the same data vector structure, to generate a t-SNE-based RM.

Consequently, we confirm that our solution is scalable and generic.

Conclusion and future works

The results presented in this paper prove that our scheme: (1) is scalable by accurately detecting and recognizing anomalous data, given normal data only, on a composite system and (2) can be generalized. We did not just have relatively high AUC and pAUC in all the machines, but we used the same scheme and same anomaly formula, without training, for all the machine IDs and types.

Therefore, we confirmed that our solution is generic and applicable on the preventive maintenance field from different perspectives and on different scales:

- Element in a machine [10, 13].
- A facility: valve, pump, fan and slide-rail as in the current work.
- Infrastructure diagnosis: Road Damage Diagnosis using vibration data measured from an in-vehicle monitoring system.

Accordingly, the field of application can further be extended to:

- Intelligent robotic systems: monitoring the motion of manipulators to prevent possible unexperienced anomalous behavior that can lead to catastrophic accidents in the workplace.
- Building facilities: monitoring the vibration in lifts and escalators to prevent (deadly) faults.
- Autonomous systems in Simultaneous Localization and Mapping (SLAM) applications: detection of unexperienced abnormalities and obstacles in an unknown environment.

- Wearable devices for health monitoring: Heart Rate Variability (HRV) vs. stress level and sleep quality.

Our solution, based on our scheme, is provided as low-intrusive white box. Users will have the following merits: (1) our solution can be in operation right after a normal data set is obtained usually in a few days and (2) our solution can be built into conventional operation and maintenance systems without advanced background in artificial intelligence or data science.

Acknowledgements
The authors would like to thank Dr. Yoshio Takaeda the CTO of Toor Inc. for providing us with the license to use toorPIA as mapping engine.

Funding
This work was not funded by any organization.

Availability of data and materials
The MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection data that support the findings of this study are available in Zenodo with the identifier(s) http://doi.org/10.5281/zenodo.3384388.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
ST and TK proposed the analysis scheme and the predictive scheme. ST implemented and tested the analysis scheme and the predictive scheme on the MIMII dataset. TK and YH assisted with the approach of the research, led and assisted the research progress and revised and refined the manuscript. All authors read and approved the final manuscript.

Author details
[1]Department of Robotics, Graduate Faculty of Engineering, Tohoku University, Aoba, Sendai, Japan. [2]D'isum Co., Ltd., Tokyo, Japan.

References

1. Windau, J., Itti, L.: Inertial machine monitoring system for automated failure detection. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 93–98 (2018). doi:10.1109/ICRA.2018.8461266
2. Kankar, P.K., Sharma, S.C., Harsha, S.P.: Fault diagnosis of ball bearings using machine learning methods. Expert Systems with applications **38**(3), 1876–1886 (2011)
3. Zhang, Z., Wang, Y., Wang, K.: Fault diagnosis and prognosis using wavelet packet decomposition, fourier transform and artificial neural network. Journal of Intelligent Manufacturing **24**(6), 1213–1227 (2013). doi:10.1007/s10845-012-0657-2. 10.1007/s10845-012-0657-2
4. Galloway, G.S., Catterson, V.M., Fay, T., Robb, A., Love, C.: Diagnosis of tidal turbine vibration data through deep neural networks (2016)
5. Koizumi, Y., Saito, S., Uematsu, H., Harada, N.: Optimizing acoustic feature extractor for anomalous sound detection based on neyman-pearson lemma. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 698–702 (2017). doi:10.23919/EUSIPCO.2017.8081297
6. Kawaguchi, Y., Endo, T.: How can we detect anomalies from subsampled audio signals? In: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6 (2017). doi:10.1109/MLSP.2017.8168164
7. Oh, D.Y., Yun, I.D.: Residual error based anomaly detection using auto-encoder in smd machine sound. Sensors **18**(5), 1308 (2018)
8. Koizumi, Y., Saito, S., Uematsu, H., Kawachi, Y., Harada, N.: Unsupervised detection of anomalous sound based on deep learning and the neyman–pearson lemma. IEEE/ACM Transactions on Audio, Speech, and Language Processing **27**(1), 212–224 (2019). doi:10.1109/TASLP.2018.2877258
9. Jardine, A.K., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing **20**(7), 1483–1510 (2006)
10. Talmoudi, S., Kanada, T., Hirata, Y.: An iot-based failure prediction solution using machine sound data. In: 2019 IEEE/SICE International Symposium on System Integration (SII), pp. 227–232 (2019). doi:10.1109/SII.2019.8700357
11. toor Inc. https://www.toor.jpn.com/
12. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(86), 2579–2605 (2008)
13. Talmoudi, S., Kanada, T., Hirata, Y.: Tracking and visualizing signs of degradation for an early failure prediction of a rolling bearing. arXiv preprint arXiv:2011.09086 (2020)
14. Lee, J., Qiu, H., Yu, G., Lin, J., (2007), R.T.S.: Bearing Data Set. NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository). IMS, University of Cincinnati, NASA Ames Research Center, Moffett Field, CA
15. Qiu, H., Lee, J., Lin, J., Yu, G.: Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of Sound and Vibration **289**(4), 1066–1090 (2006). doi:10.1016/j.jsv.2005.03.007
16. Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., Kawaguchi, Y.: MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. doi:10.5281/zenodo.3384388. https://doi.org/10.5281/zenodo.3384388
17. Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., Kawaguchi, Y.: Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. arXiv preprint arXiv:1909.09347 (2019)
18. Fawcett, T.: An introduction to roc analysis. Pattern Recognition Letters **27**(8), 861–874 (2006). doi:10.1016/j.patrec.2005.10.010. ROC Analysis in Pattern Recognition
19. Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., Purohit, H., Suefusa, K., Endo, T., Yasuda, M., et al.: Description and discussion on dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring. arXiv preprint arXiv:2006.05822 (2020)
20. Koizumi, Y., Saito, S., Uematsu, H., Harada, N., Imoto, K.: Toyadmos: A dataset of miniature-machine operating sounds for anomalous sound detection. In: 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 313–317 (2019). doi:10.1109/WASPAA.2019.8937164
21. DECASE 2020, task 2. http://dcase.community/challenge2020/task-unsupervised-detection-of-anomalous-sounds
22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
23. Baseline implementation. https://github.com/y-kawagu/dcase2020_task2_baseline
24. Agrawal, V.K., Maurya, S.S.: Unsupervised detection of anomalous sounds for machine condition monitoring. Technical report, DCASE2020 Challenge (July 2020)
25. Ahmed, F., Nguyen, P., Courville, A.: An ensemble approach for detecting machine failure from sound. Technical report, DCASE2020 Challenge (July 2020)
26. Alam, J., Boulianne, G., Gupta, V., Fathan, A.: An ensemble approach to unsupervised anomalous sound detection. Technical report, DCASE2020 Challenge (July 2020)
27. Bai, J., Chen, C., Chen, J.: Bai_lfxs_nwpu_dcase2020_submission. Technical report, DCASE2020 Challenge (July 2020)
28. Chaudhary, N.K., Mathew, J., Sivadas, S.: Convolutional auto encoder for machine condition monitoring from acoustic signatures. Technical report, DCASE2020 Challenge (July 2020)
29. Chen, Y., Song, Y., Cheng, T.: Anomalous sounds detection using a new type of autoencoder based on residual connection. Technical report, DCASE2020 Challenge (July 2020)

30. Daniluk, P., Gozdziewski, M., Kapka, S., Kosmider, M.: Ensemble of auto-encoder based systems for anomaly detection. Technical report, DCASE2020 Challenge (July 2020)
31. Durkota, K., Michael, L., Martin, L., Jan, T.: Neuron-net: Siamese network for anomaly detection. Technical report, DCASE2020 Challenge (July 2020)
32. Giri, R., Tenneti, S.V., Helwani, K., Cheng, F., Isik, U., Krishnaswamy, A.: Unsupervised anomalous sound detection using self-supervised classification and group masked autoencoder for density estimation. Technical report, DCASE2020 Challenge (July 2020)
33. Grollmisch, S., Johnson, D., AbeBer, J., Lukashevich, H.: Iaeo3 - combining openl3 embeddings and interpolation autoencoder for anomalous sound detection. Technical report, DCASE2020 Challenge (July 2020)
34. Haunschmid, V., Praher, P.: Anomalous sound detection with masked autoregressive flows and machine type dependent postprocessing. Technical report, DCASE2020 Challenge (July 2020)
35. Hayashi, T., Yoshimura, T., Adachi, Y.: Conformer-based id-aware autoencoder for unsupervised anomalous sound detection. Technical report, DCASE2020 Challenge (July 2020)
36. Hoang, T., Nguyen, H., Pham, G.: Unsupervised detection of anomalous sound for machine condition monitoring using different auto-encoder methods. Technical report, DCASE2020 Challenge (July 2020)
37. Jalali, A., Schindler, A., Bernhard, H.: Dcase challenge 2020: Unsupervised anomalous sound detection of machinery with deep autoencoders. Technical report, DCASE2020 Challenge (July 2020)
38. Jiang, H., Lan, B., Li, H.: Abnormal sound detection system based on autoencoder. Technical report, DCASE2020 Challenge (July 2020)
39. Kaltampanidis, Y., Thoidis, I., Vrysis, L., Dimoulas, C.: Unsupervised detection of anomalous sounds via protopnet. Technical report, DCASE2020 Challenge (July 2020)
40. Lapin, D., Khubbatulin, M., Klychnikov, V.: The study of anomalous machine sound detection based on cyclostationarity model. Technical report, DCASE2020 Challenge (July 2020)
41. Lopez, J., Hong, L., Lopez-Meyer, P., Nachman, L., Stemmer, G., Huang, J.: A speaker recognition approach to anomaly detection. Technical report, DCASE2020 Challenge (July 2020)
42. Morita, K., Yano, T., Tran, K.Q.: Anomalous sound detection by using local outlier factor and gaussian mixture model. Technical report, DCASE2020 Challenge (July 2020)
43. Naranjo-Alcazar, J., Perez-Castanos, S., Zuccarello, P., Cobos, M.: Task 2 dcase 2020: Anomalous sound detection using unsupervised and semi-supervised autoencoders and gammtone audio representation. Technical report, DCASE2020 Challenge (July 2020)
44. Park, J., Sooyeon, Y., Yoo, S.: Unsupervised detection of anomalous machine sound using various spectral features and focused hypothesis test in the reverberant and noisy environment. Technical report, DCASE2020 Challenge (July 2020)
45. Phan, D., Jones, D.: Dcase 2020 task 2: Unsupervised detection of anomalous sounds for machine condition monitoring. Technical report, DCASE2020 Challenge (July 2020)
46. Primus, P.: Reframing unsupervised machine condition monitoring as a supervised classification task with outlier-exposed classifiers. Technical report, DCASE2020 Challenge (July 2020)
47. Ribeiro, A., Matos, L., Pereira, P., Nunes, E., Ferreira, A., Cortez, P., Pilastri, A.: Deep dense and convolutional autoencoders for unsupervised anomaly detection in machine condition sounds. Technical report, DCASE2020 Challenge (July 2020)
48. Sakamoto, Y., Miyamoto, N.: Anomaly calculation for each components of sound data and its integration for dcase 2020 challenge task2. Technical report, DCASE2020 Challenge (July 2020)
49. Shao, R., Han, Y., Lin, D.: A method of anomalous sound detection with multi-dimensional audio feature inputs. Technical report, DCASE2020 Challenge (July 2020)
50. Shinmura, Z.: Dcase2020 task2 self-supervised learning solution. Technical report, DCASE2020 Challenge (July 2020)
51. Tian, K., Fu, G., Li, S., Tang, G., Shao, X.: Anomaly machine detection algorithm based on semi variational auto-encoder of mel spectrogram. Technical report, DCASE2020 Challenge (July 2020)
52. Tiwari, P., Jain, Y., Avila, A., Monteiro, J., Kshirsagar, S., Amr, G., Falk, T.H.: Modulation spectral signal representationand i-vectors for anomalous sound detection. Technical report, DCASE2020 Challenge (July 2020)
53. Uchikoshi, M.: Anomaly detection using the middle layer of the cnn-classification model. Technical report, DCASE2020 Challenge (July 2020)
54. Vinayavekhin, P., Inoue, T., Morikuni, S., Wang, S., Hoang Trong, T., Wood, D., Tatsubori, M., Tachibana, R.: Detection of anomalous sounds for machine condition monitoring using classification confidence. Technical report, DCASE2020 Challenge (July 2020)
55. Wang, Y., Zhang, X., He, L.: Anomalous sound detection based on a novel autoencoder. Technical report, DCASE2020 Challenge (July 2020)
56. Wei, Q., Liu, Y.: Auto-encoder and metric-learning for anomalous sound detection task. Technical report, DCASE2020 Challenge (July 2020)
57. Wen, H., Shi, S., Shi, C.: Unsupervised detection of anomalous sounds using abnormal sound simulation algorithm and auto-encoder classifier. Technical report, DCASE2020 Challenge (July 2020)
58. Wilkinghoff, K.: Anomalous sound detection with look, listen, and learn embeddings. Technical report, DCASE2020 Challenge (July 2020)
59. Xiao, Y.: Unsupervised detection of anomalous sounds technical report. Technical report, DCASE2020 Challenge (July 2020)
60. Zhang, C., Yao, Y., Zhou, Y., Fu, G., Li, S., Tang, G., Shao, X.: Unsupervised detection of anomalous sounds based on dictionary learning and autoencoder. Technical report, DCASE2020 Challenge (July 2020)
61. Zhao, S.: Acoustic anomaly detection based on similarity analysis. Technical report, DCASE2020 Challenge (July 2020)
62. Zhou, Q.: Arcface based sound mobilenets for dcase 2020 task 2. Technical report, DCASE2020 Challenge (July 2020)

Figures

Figure 1 The process flow of our predictive scheme

Figure 2 Placing the newly acquired data on the RM

Figure 3 The evolution of an audio-clip through the preprocessing phase **(a)** time domain data extracted from the .wav file **(b)** original spectrum **(c)** smoothed spectrum

Figure 4 The obtained RMs for all machine types and IDs with $FFR\_approach$

Figure 5 Spectral distribution of the RMs of valve 00, pump 00, fan 00 and slider 00

Figure 6 The obtained RMs for all machine types and IDs with $HPF\_approach$

Tables

Figure 7 The obtained anomaly scores for valve

Figure 8 The obtained anomaly scores for pump

Figure 9 The obtained anomaly scores for fan

Figure 10 The obtained anomaly scores for slide-rail

Figure 11 Presentation in the ROC space of **(a)** a discrete classifier and **(b)** a non-discrete classifier

Figure 12 Comparison of **(a)** discrete classifiers and **(b)** non-discrete classifiers

Figure 13 Comparison of our obtained AUC and pAUC values for each machine ID to the baseline system results

Figure 14 Positioning of our results compared to the participating teams for each machine type **(a)** valve, **(b)** pump, **(c)** fan and **(d)** slide-rail

Figure 15 Overall positioning of our work compared to the participating teams

Table 1 Digest of the development dataset

Target machine	Development set	
	Training set	Test set
Valve00	891	219
Valve02	608	220
Valve04	900	220
Valve06	892	220
Pump00	906	243
Pump02	905	211
Pump04	602	200
Pump06	936	202
Fan00	911	507
Fan02	916	459
Fan04	933	448
Fan06	915	461
Slide-rail00	968	456
Slide-rail02	968	367
Slide-rail04	434	278
Slide-rail06	434	189

Table 2 Summary of the multidimensional space for the similarity quantification to generate the RM

Case no.	Machine type and ID	Multidimensional space structure	
		$FFR\_approach$	$HPF\_approach$
1	Valve 00	(891×372)	(891×302)
2	Valve 02	(608×372)	(608×302)
3	Valve 04	(900×372)	(900×302)
4	Valve 06	(892×372)	(892×302)
5	Pump 00	(906×372)	(906×302)
6	Pump 02	(905×372)	(905×302)
7	Pump 04	(602×372)	(602×302)
8	Pump 06	(936×372)	(936×302)
9	Fan 00	(911×372)	(911×302)
10	Fan 02	(916×372)	(916×302)
11	Fan 04	(933×372)	(933×302)
12	Fan 06	(915×372)	(915×302)
13	Slide-rail 00	(968×372)	(968×302)
14	Slide-rail 02	(968×372)	(968×302)
15	Slide-rail 04	(434×372)	(434×302)
16	Slide-rail 06	(434×372)	(434×302)

Table 3 Our AUC and pAUC values compared to the baseline system by the organizers

Machine type and ID	$FFR\_approach$		$HPF\_approach$		Baseline	
	AUC	pAUC	AUC	pAUC	AUC	pAUC
Valve00	87.22	76.78	98.51	92.43	68.76	51.70
Valve02	89.78	73.72	97.06	91.97	68.18	51.83
Valve04	77.95	63.77	90.83	80.78	74.30	51.97
Valve06	65.98	55.13	72.20	56.84	53.90	48.43
Avg.	**80.23**	**67.35**	**89.65**	**80.50**	**66.28**	**50.98**
Pump00	86.37	76.73	70.25	55.53	67.15	56.74
Pump02	66.90	63.11	67.21	61.87	61.53	58.10
Pump04	70.61	58.52	90.49	83.68	88.33	67.10
Pump06	73.58	70.53	62.03	49.32	74.55	58.02
Avg.	**74.36**	**67.22**	**72.49**	**62.6**	**72.89**	**59.99**
Fan00	57.41	50.32	56.07	50.48	54.41	49.37
Fan02	77.51	60.63	79.32	56.47	73.40	54.81
Fan04	73.74	59.58	51.82	52.13	61.61	53.26
Fan06	81.35	60.91	68.44	54.52	73.92	52.35
Avg.	**72.50**	**57.86**	**63.91**	**53.4**	**65.83**	**52.48**
Slide-rail00	94.10	86.04	76.54	52.83	96.19	81.44
Slide-rail02	71.57	52.73	69.14	52.53	78.97	63.68
Slide-rail04	93.40	82.52	98.25	95.59	94.30	71.98
Slide-rail06	72.71	58.42	84.48	69.84	69.59	49.02
Avg.	**82.94**	**69.92**	**82.10**	**67.69**	**84.76**	**66.53**

Table 4 The differences between the datasets in the application of [13] and the current application

	IMS dataset [14, 15]	MIMII dataset [16, 17]
Data type	Vibration	Audio
Target	Bearings	Valve, pump, fan and slide-rail
Target complexity	Singular system	Composite system
Sensor attachement	Directly	Contactless
Sampling rate	20 kHz	16 kHz
Operation	Monotone at constant rotation speed	Valve: repeated open/close with different timing Pump: suction from and discharging to a water pool Fan: normal work Slide-rail: repeated slide on different speeds
Test scenario	Test-to-failure	Seeded anomalies
Target condition	Normal/Degraded/Failure	Normal/Anomalous
Labels	Not provided	Provided

Table 5 The analysis process and parameters for both applications

		Application of [13]	Current application
Process	RM generation	(1)Preprocessing (Segmentation FFT, smoothing) (2)Similarity quantification (3)Visualization	(1)Preprocessing (Segmentation, FFT, smoothing) (2)Similarity quantification (3)Visualization
	Adding test data	Geometry-based algorithm with zero-data as origin	Geometry-based algorithm with the gravity center of the RM as origin
Parameteres	Segment length	1 sec	10 sec
	Spectrum size	128	512
	Frequency resolution	78.13 Hz	19.58 Hz

Reference Map generation from early acquired data

Adding the newly collected data in real-time

In $\mathbb{R}^n$

Preprocessing phase

Similarity estimation

Dimension reduction by toorPIA/t-SNE

Visualization

In $\mathbb{R}^2$

Reference map

Decision

No alert

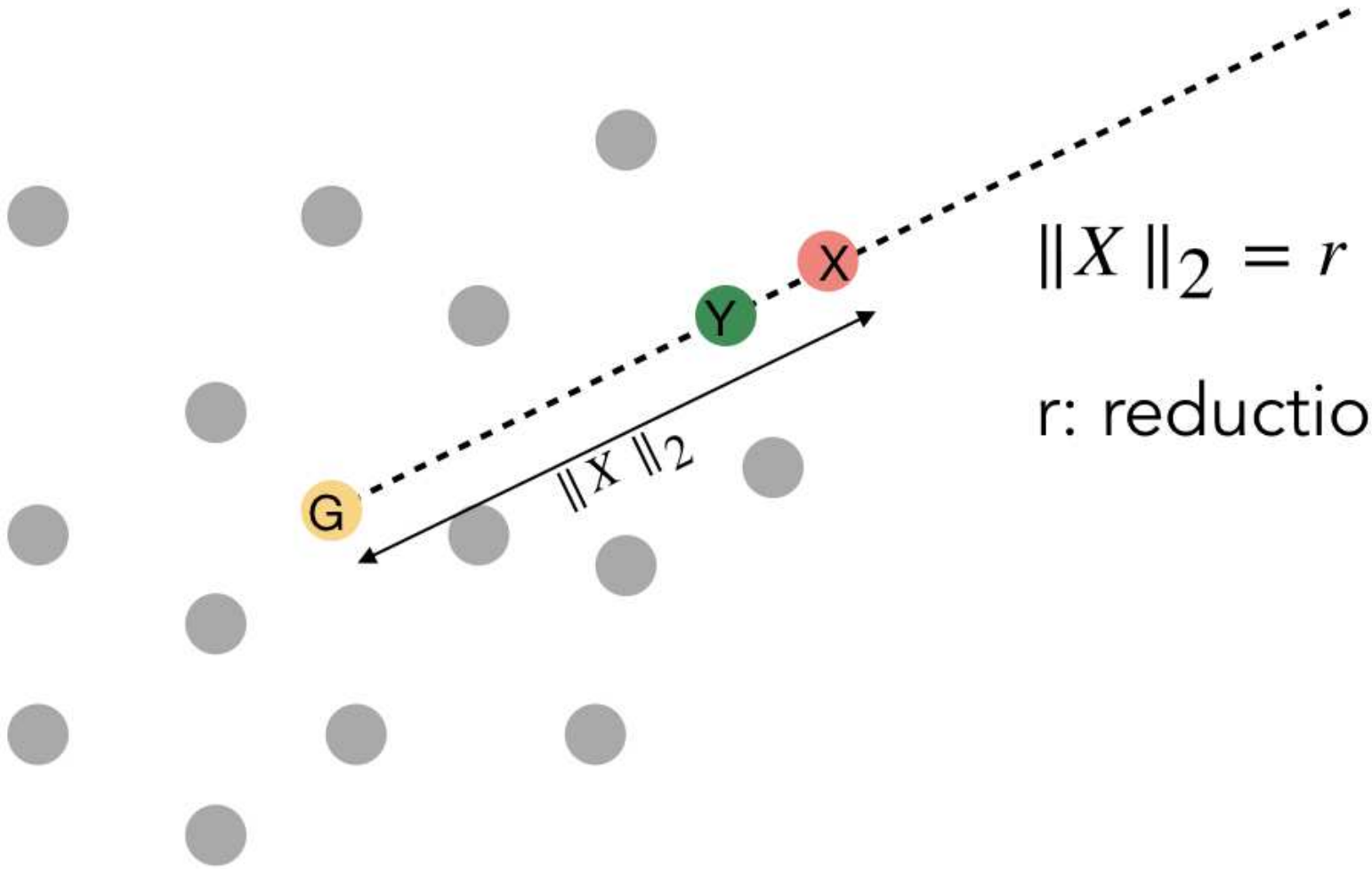
Alert

Interfere

Preprocessing phase

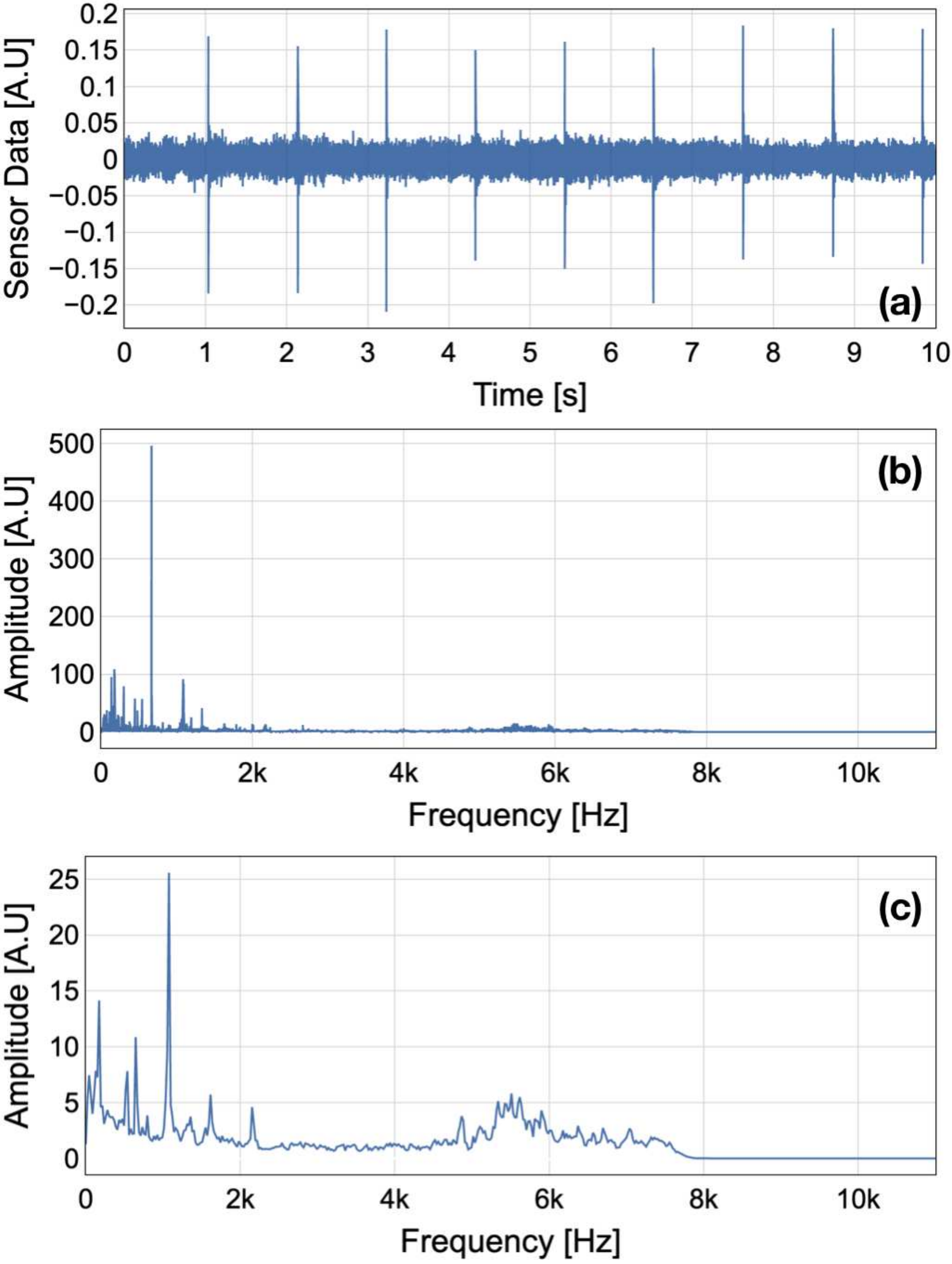
Data plotting

Continue monitoring

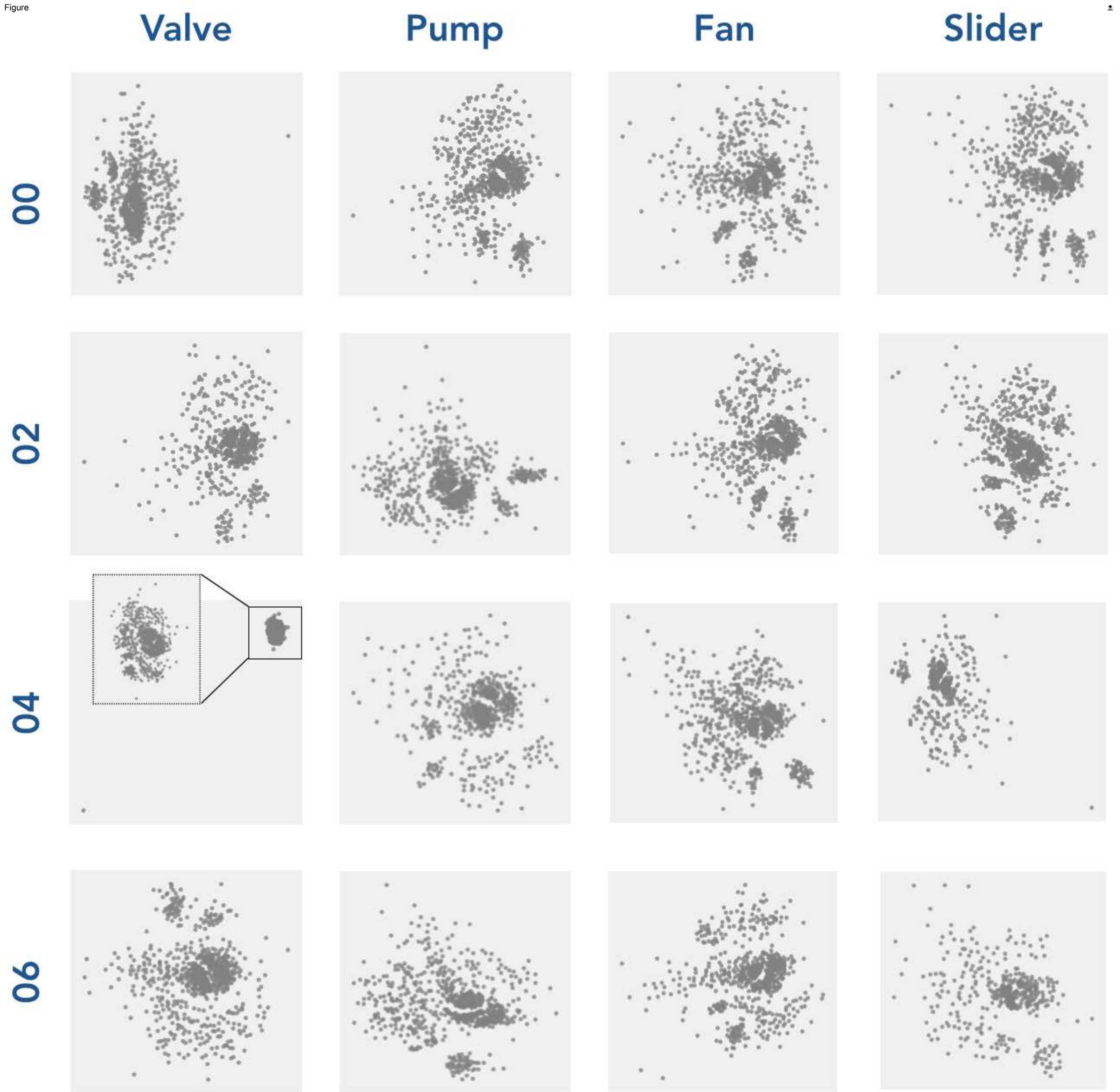


$$\|X\|_2 = r \times \|X\|_n$$

r: reduction ration

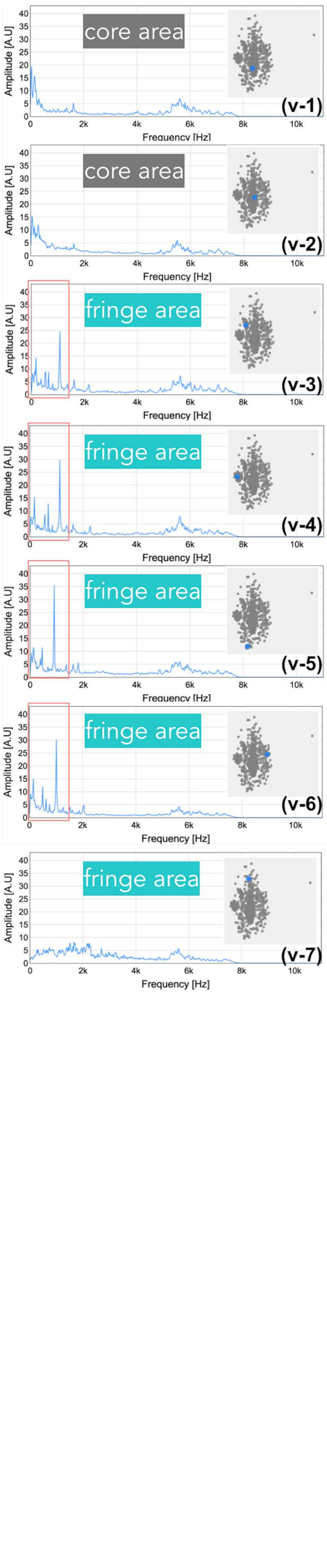

Sensor Data [A.U]
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
(a)
0
1
2
3
4
5
6
7
8
9
10
Time [s]
Amplitude [A.U]
500
400
300
200
100
0
(b)
0
2k
4k
6k
8k
10k
Frequency [Hz]
Amplitude [A.U]
25
20
15
10
5
0
(c)
0
2k
4k
6k
8k
10k
Frequency [Hz]

Figure

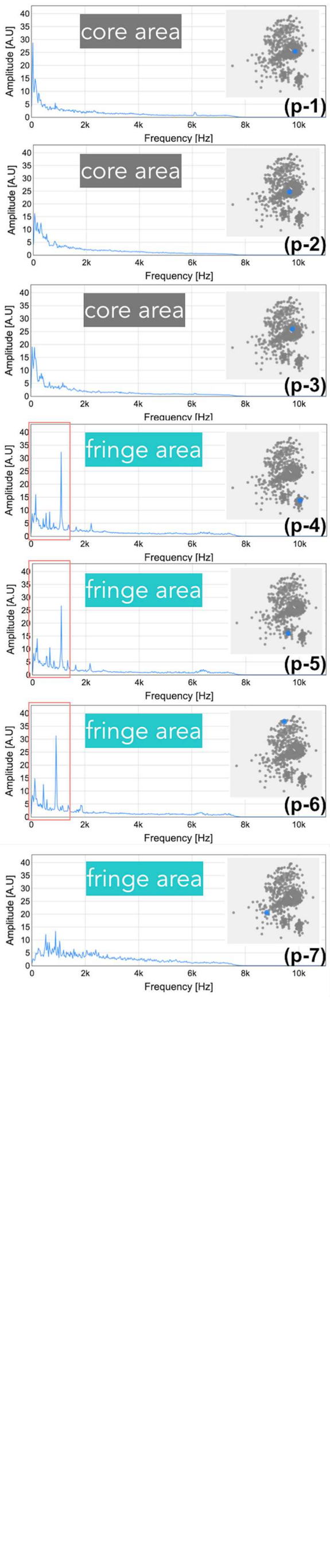


Figure

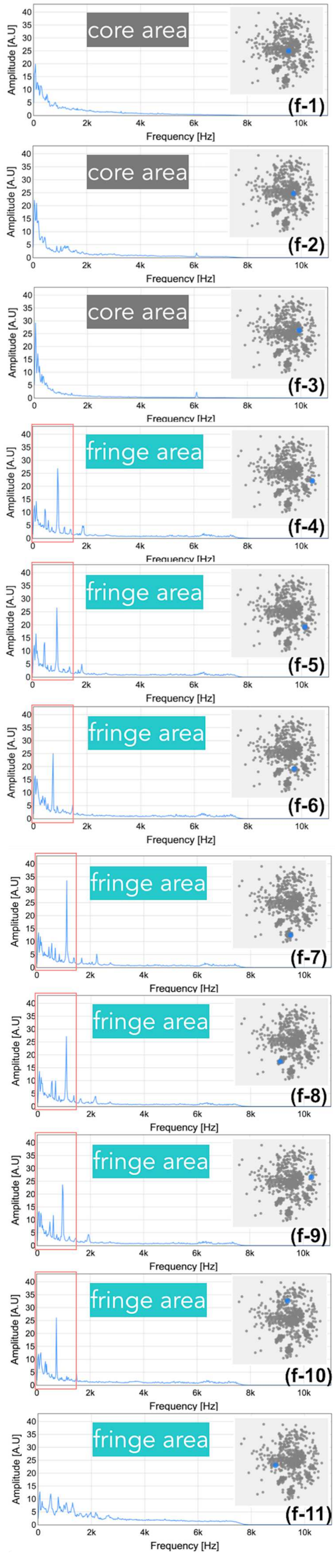
Valve00



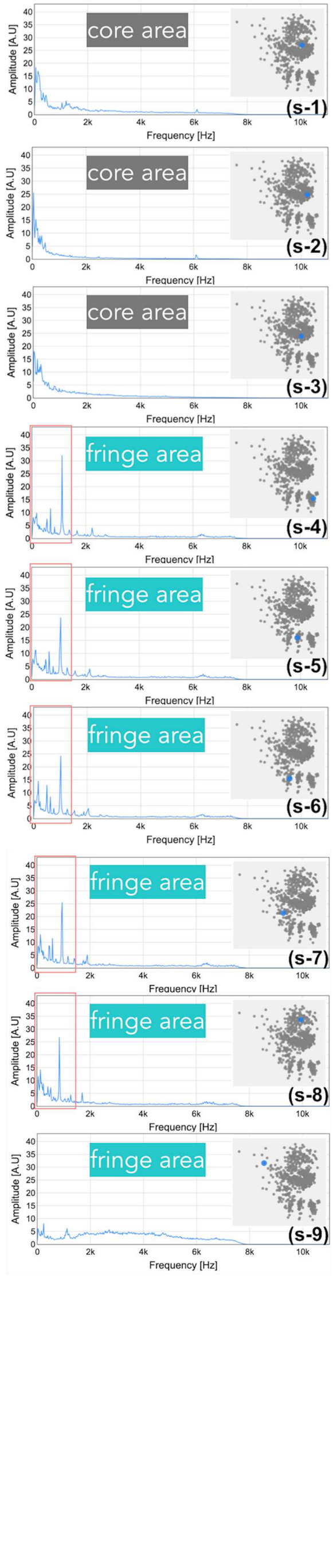
Pump00



Fan00

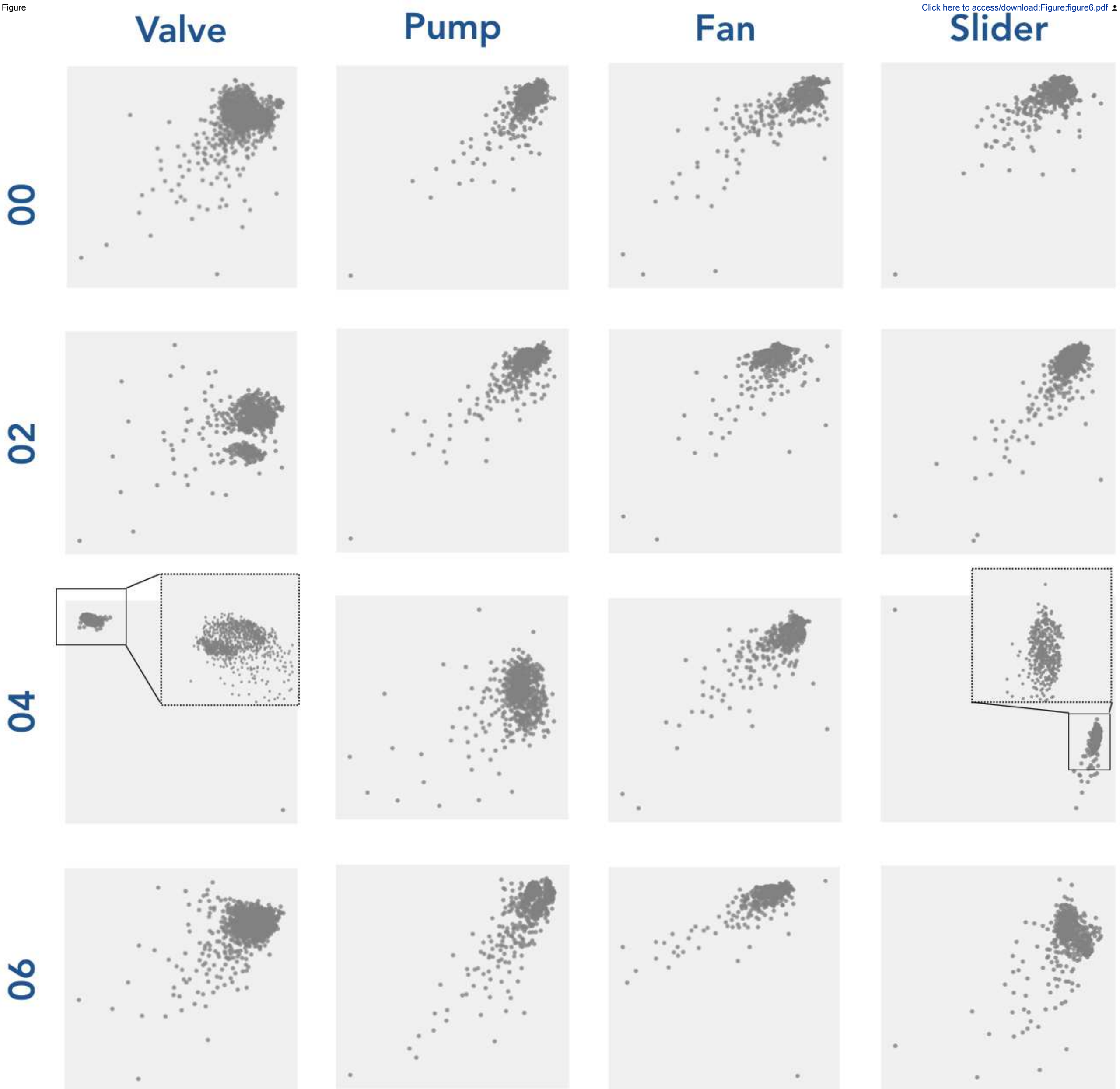


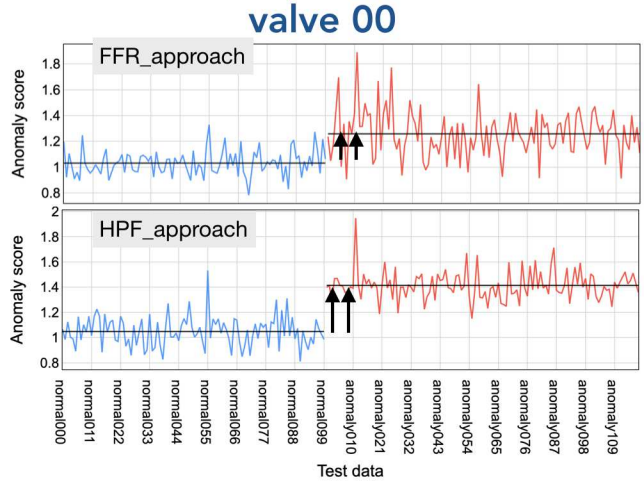
Slider00

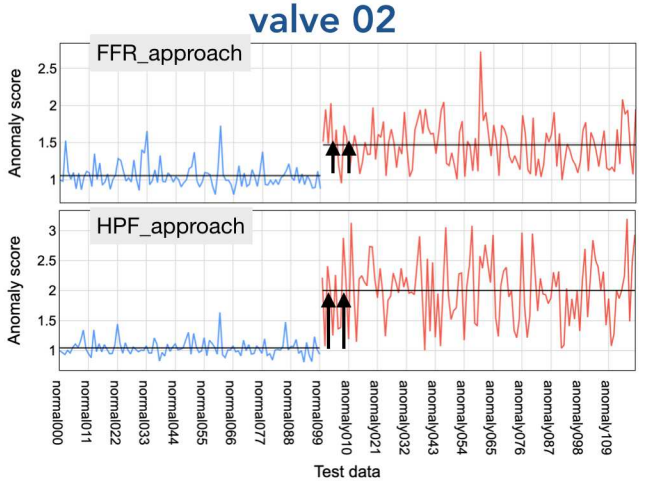


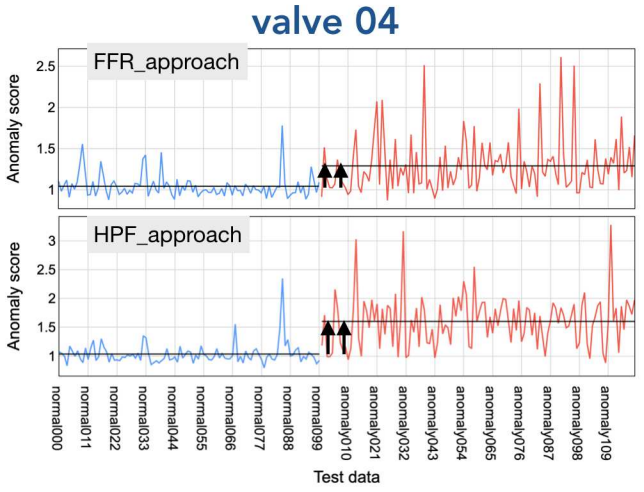
Figure

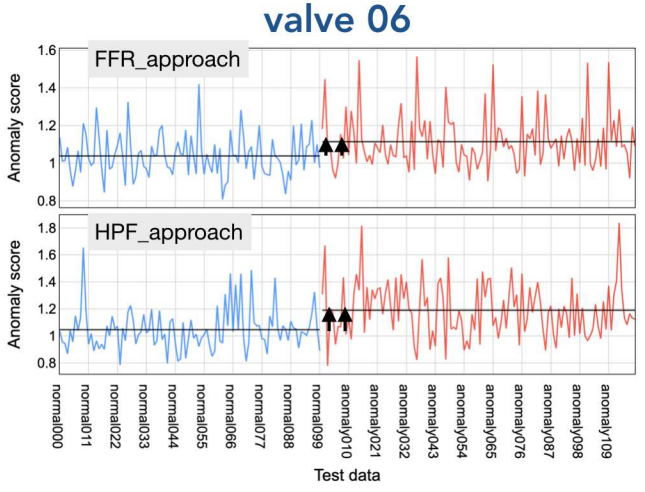
Click here to access/download;Figure;figure6.pdf

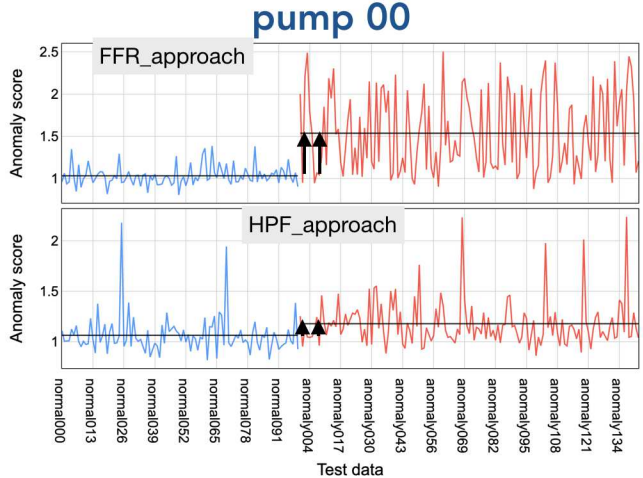


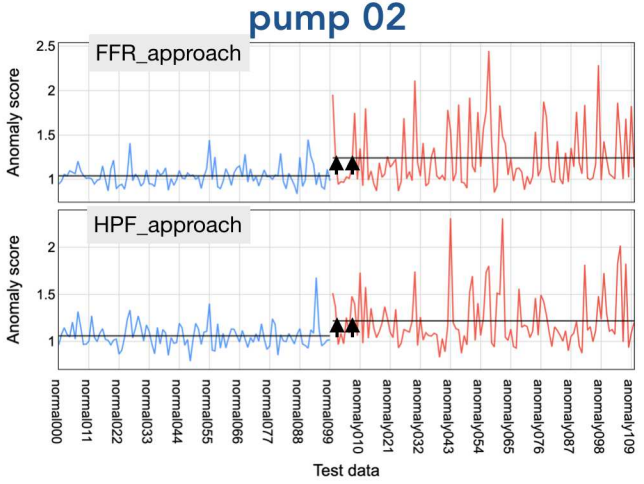

valve 00
FFR_approach
HPF_approach
Anomaly score
Test data
normal000
normal011
normal022
normal033
normal044
normal055
normal066
normal077
normal088
normal099
anomaly010
anomaly021
anomaly032
anomaly043
anomaly054
anomaly065
anomaly076
anomaly087
anomaly098
anomaly109

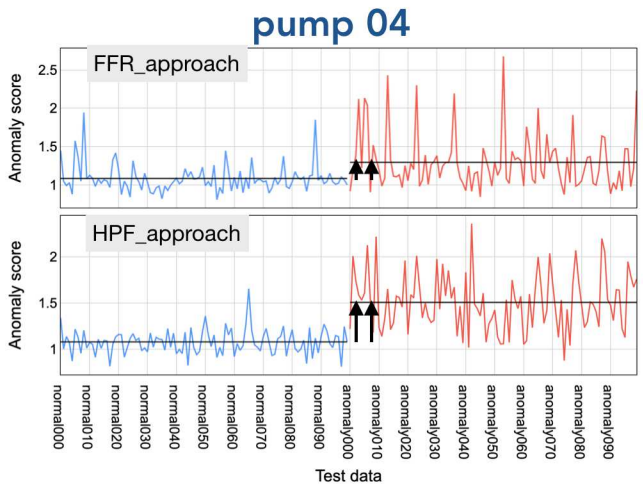

valve 02
FFR_approach
HPF_approach
Anomaly score
Test data
normal000
normal011
normal022
normal033
normal044
normal055
normal066
normal077
normal088
normal099
anomaly010
anomaly021
anomaly032
anomaly043
anomaly054
anomaly065
anomaly076
anomaly087
anomaly098
anomaly109

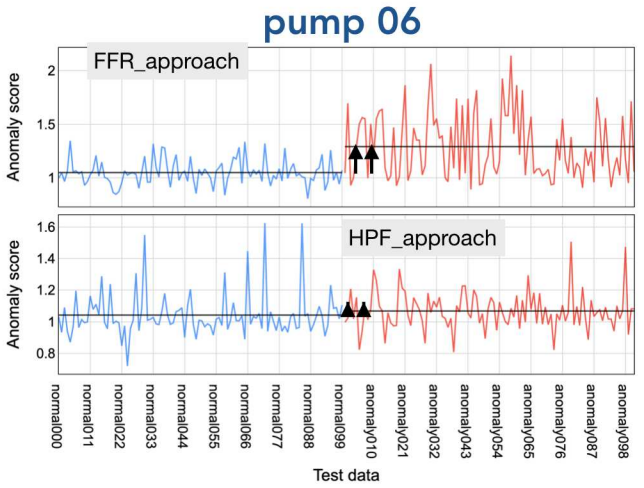

valve 04
FFR_approach
HPF_approach
Anomaly score
Test data
normal000
normal011
normal022
normal033
normal044
normal055
normal066
normal077
normal088
normal099
anomaly010
anomaly021
anomaly032
anomaly043
anomaly054
anomaly065
anomaly076
anomaly087
anomaly098
anomaly109

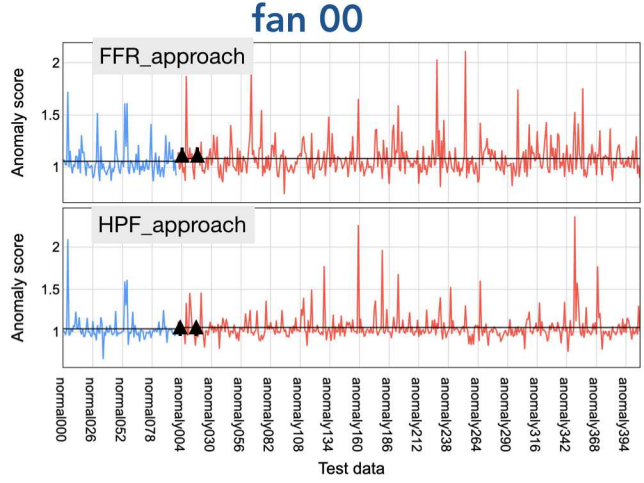

valve 06
FFR_approach
HPF_approach
Anomaly score
Test data
normal000
normal011
normal022
normal033
normal044
normal055
normal066
normal077
normal088
normal099
anomaly010
anomaly021
anomaly032
anomaly043
anomaly054
anomaly065
anomaly076
anomaly087
anomaly098
anomaly109

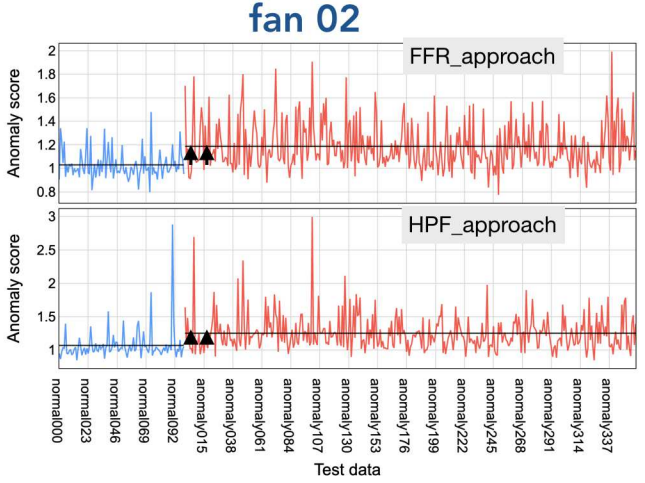

pump 00
FFR_approach
HPF_approach
Anomaly score
Test data

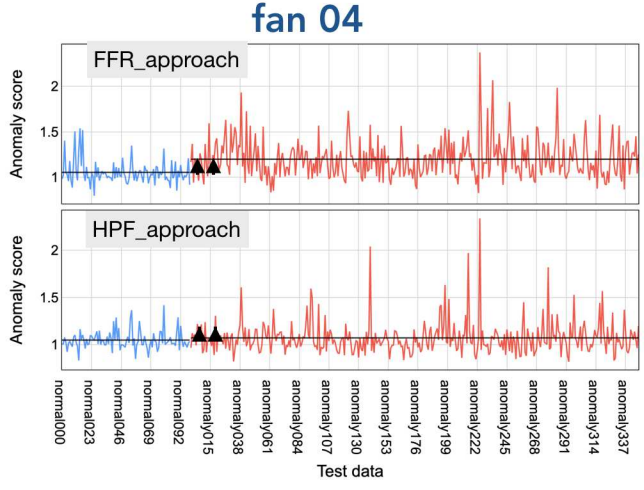

pump 02
FFR_approach
HPF_approach
Anomaly score
Test data

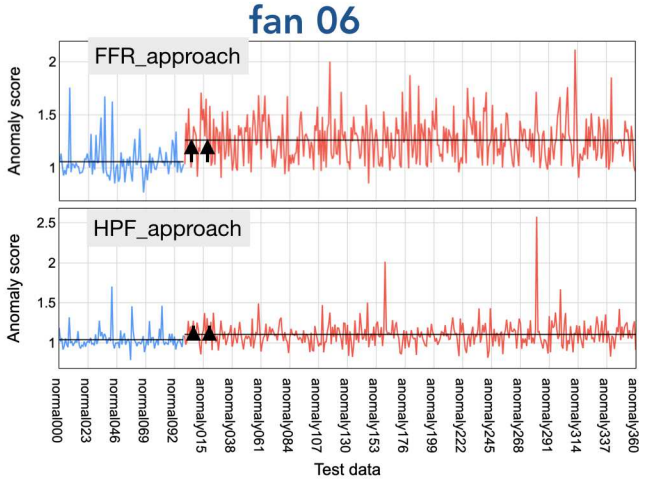

pump 04
FFR_approach
HPF_approach
Anomaly score
Test data

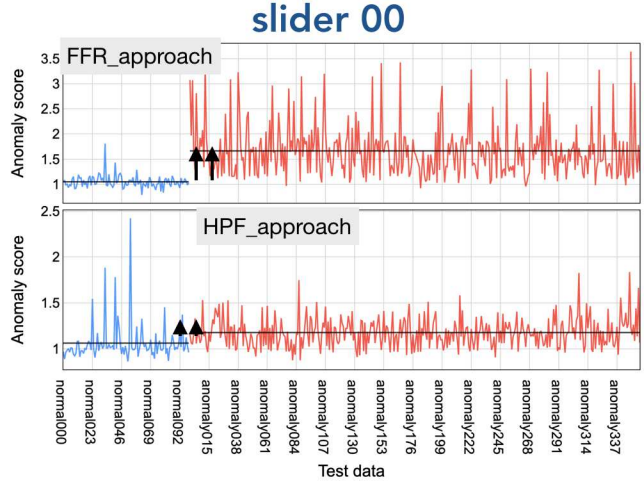

pump 06
FFR_approach
HPF_approach
Anomaly score
Test data

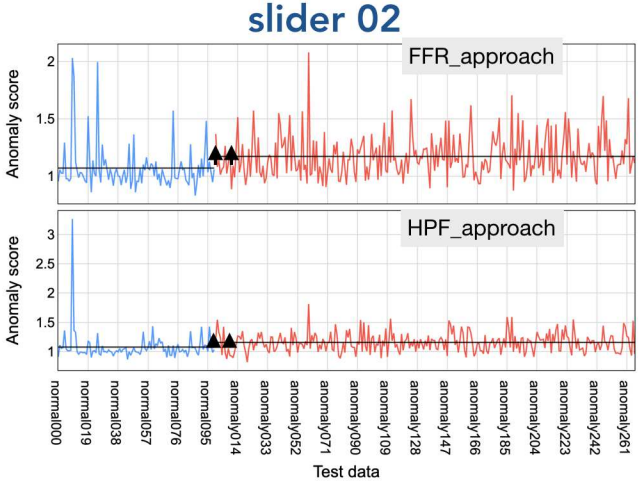

fan 00
FFR_approach
HPF_approach
Anomaly score
Test data

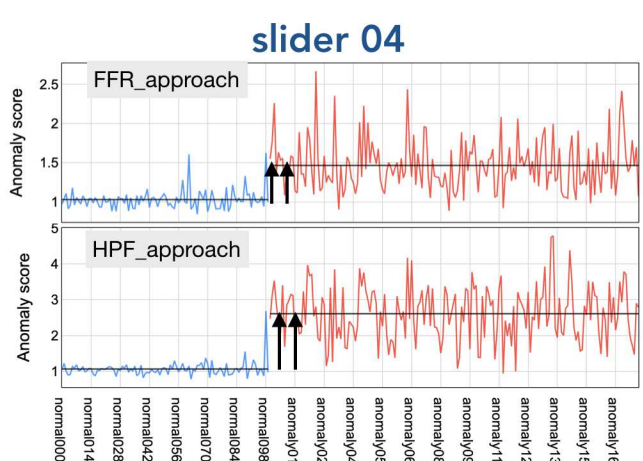

fan 02
FFR_approach
HPF_approach
Anomaly score
Test data

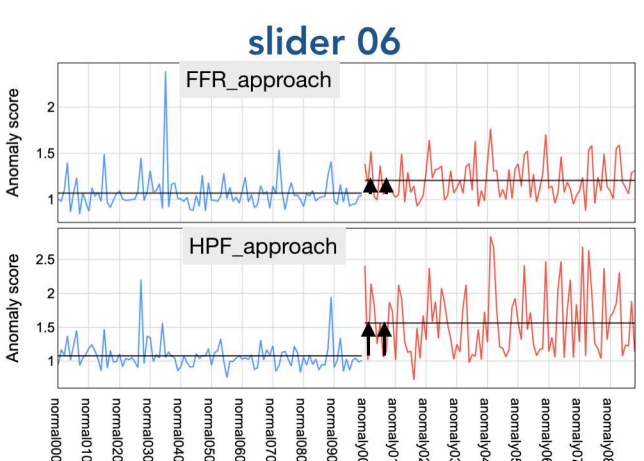

fan 04
FFR_approach
HPF_approach
Anomaly score
Test data

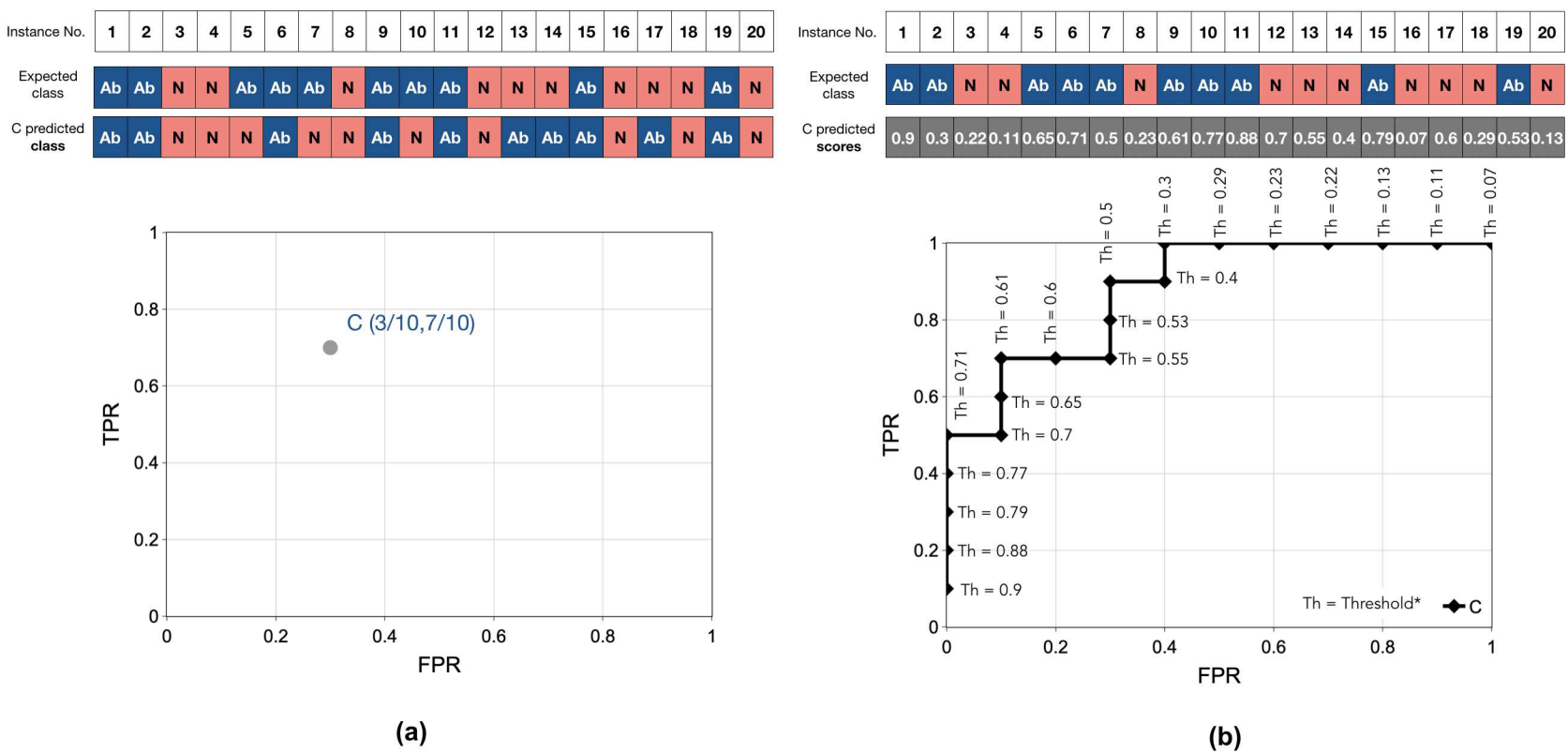

fan 06
FFR_approach
HPF_approach
Anomaly score
Test data


slider 00
FFR_approach
HPF_approach
Anomaly score
Test data

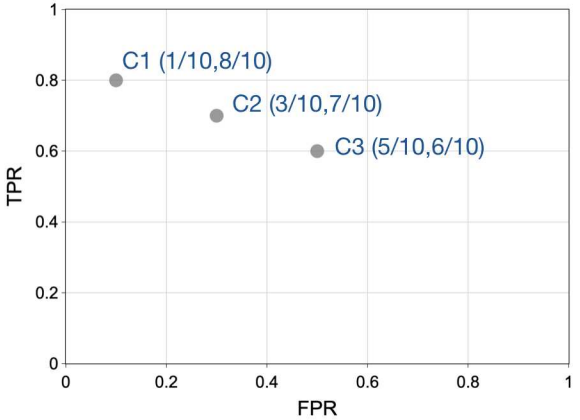

slider 02
FFR_approach
HPF_approach
Anomaly score
Test data


slider 04
FFR_approach
HPF_approach
Anomaly score
Test data


slider 06
FFR_approach
HPF_approach
Anomaly score
Test data

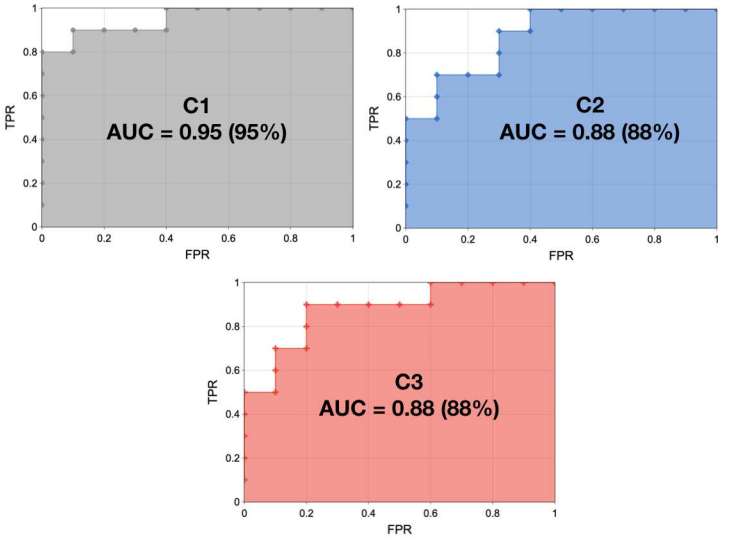

Instance No. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Expected class Ab Ab N N Ab Ab Ab N Ab Ab Ab N N N Ab N N N Ab N
C predicted class Ab Ab N N N Ab N N Ab N Ab N Ab Ab Ab N Ab N Ab N
C (3/10,7/10)
TPR
FPR
0 0.2 0.4 0.6 0.8 1
(a)
Instance No. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Expected class Ab Ab N N Ab Ab Ab N Ab Ab Ab N N N Ab N N N Ab N
C predicted scores 0.9 0.3 0.22 0.11 0.65 0.71 0.5 0.23 0.61 0.77 0.88 0.7 0.55 0.4 0.79 0.07 0.6 0.29 0.53 0.13
Th = 0.3
Th = 0.29
Th = 0.23
Th = 0.22
Th = 0.13
Th = 0.11
Th = 0.07
Th = 0.5
Th = 0.4
Th = 0.61
Th = 0.6
Th = 0.53
Th = 0.55
Th = 0.71
Th = 0.65
Th = 0.7
Th = 0.77
Th = 0.79
Th = 0.88
Th = 0.9
Th = Threshold*
C
TPR
FPR
0 0.2 0.4 0.6 0.8 1
(b)

Instance No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Expected class	Ab	Ab	N	N	Ab	Ab	Ab	N	Ab	Ab	Ab	N	N	N	Ab	N	N	N	Ab	N
C1 predicted **class**	Ab	N	N	N	Ab	Ab	N	N	Ab	Ab	Ab	N	Ab	N	Ab	N	N	N	Ab	N
C2 predicted **class**	Ab	Ab	N	N	N	Ab	N	N	Ab	N	Ab	N	Ab	Ab	Ab	N	Ab	N	Ab	N
C3 predicted **class**	N	Ab	Ab	Ab	N	Ab	N	N	Ab	N	Ab	Ab	N	Ab	Ab	N	Ab	N	Ab	N

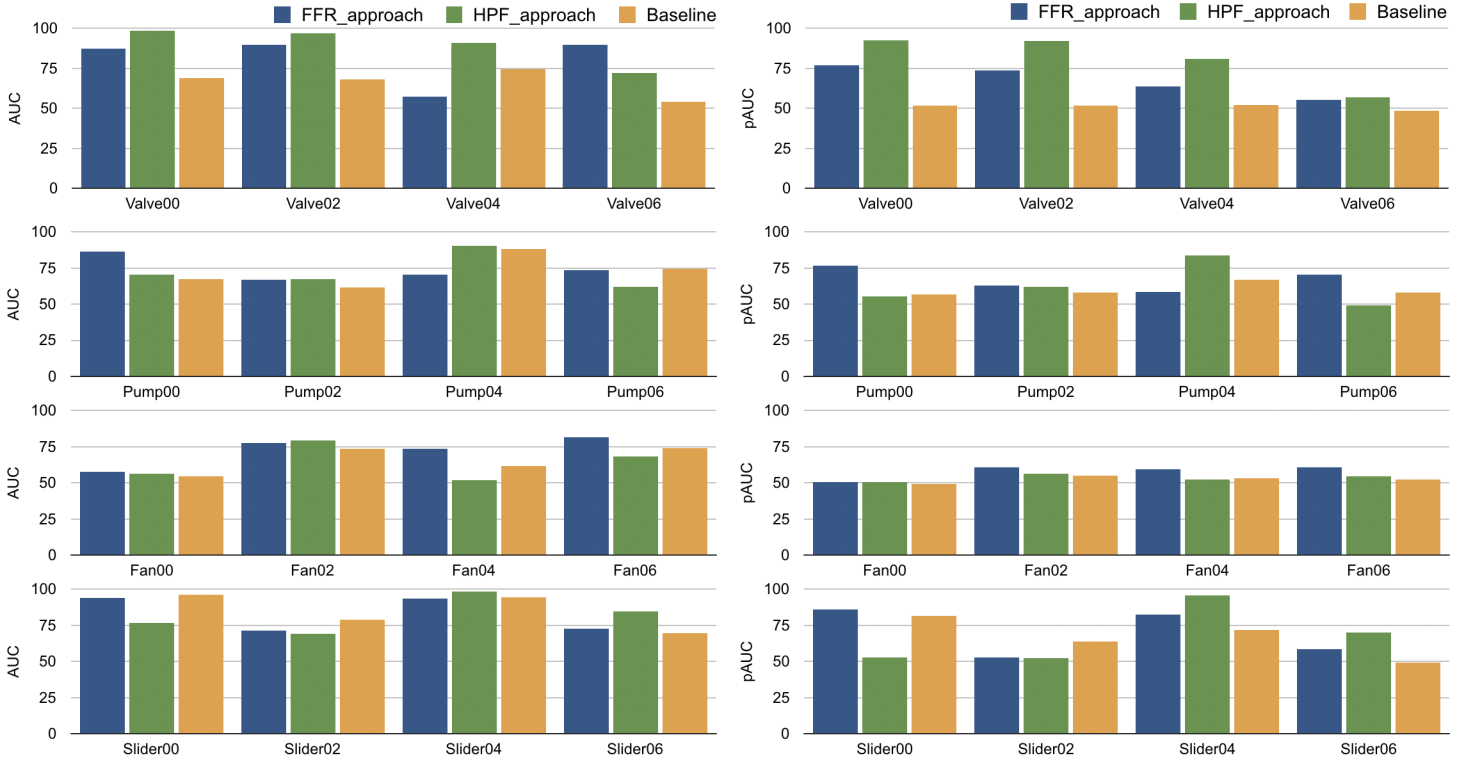


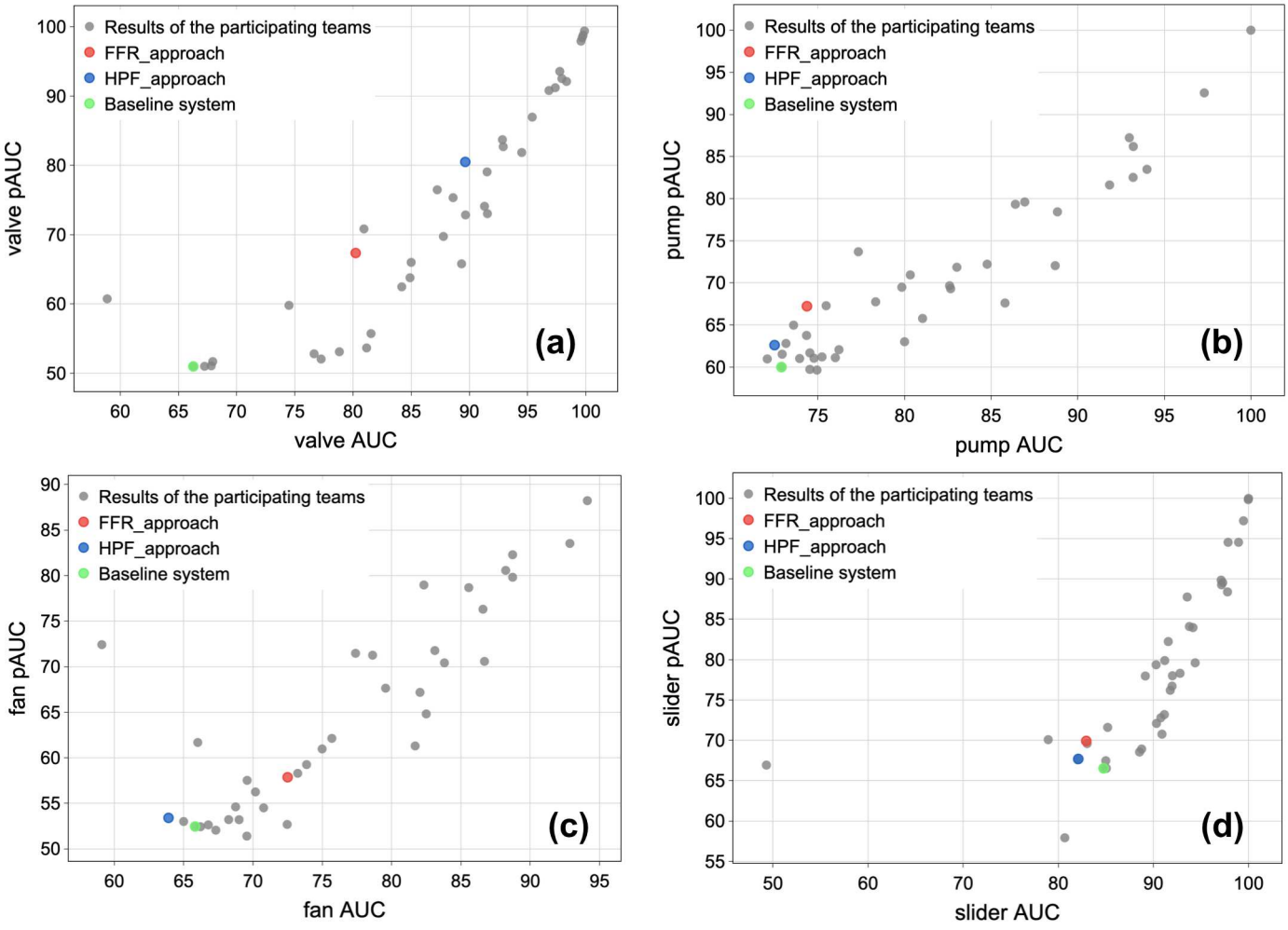
(a)

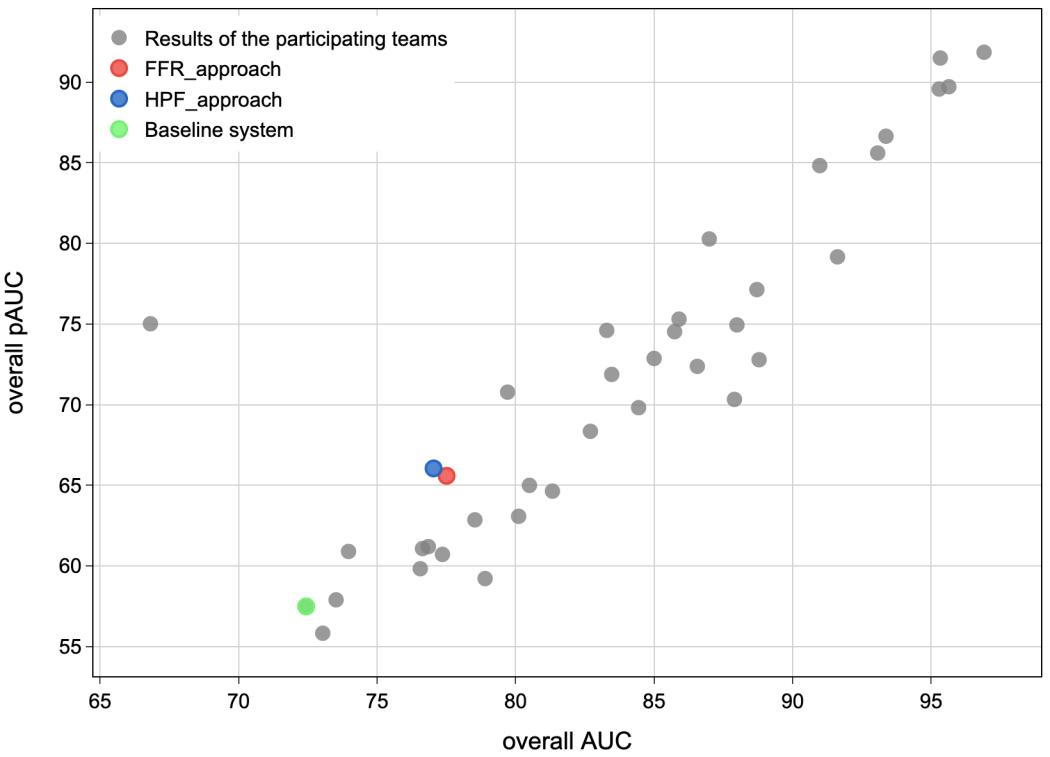
Instance No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Expected class	Ab	Ab	N	N	Ab	Ab	Ab	N	Ab	Ab	Ab	N	N	N	Ab	N	N	N	Ab	N
C1 predicted **scores**	0.9	0.85	0.1	0.2	0.73	0.63	0.56	0.4	0.61	0.3	0.88	0.6	0.5	0.21	0.79	0.13	0.43	0.26	0.71	0.09
C2 predicted **scores**	0.9	0.3	0.22	0.11	0.65	0.71	0.5	0.23	0.61	0.77	0.88	0.7	0.55	0.4	0.79	0.07	0.6	0.29	0.53	0.13
C3 predicted **scores**	0.8	0.7	0.74	0.32	0.5	0.9	0.88	0.31	0.66	0.3	0.89	0.4	0.34	0.11	0.77	0.29	0.05	0.27	0.53	0.6



(b)


FFR_approach
HPF_approach
Baseline
AUC
pAUC
0
25
50
75
100
Valve00
Valve02
Valve04
Valve06
Pump00
Pump02
Pump04
Pump06
Fan00
Fan02
Fan04
Fan06
Slider00
Slider02
Slider04
Slider06


Results of the participating teams
FFR_approach
HPF_approach
Baseline system
valve pAUC
valve AUC
(a)
pump pAUC
pump AUC
(b)
fan pAUC
fan AUC
(c)
slider pAUC
slider AUC
(d)


Results of the participating teams
FFR_approach
HPF_approach
Baseline system
overall pAUC
overall AUC
55
60
65
70
75
80
85
90
65
70
75
80
85
90
95

Figures

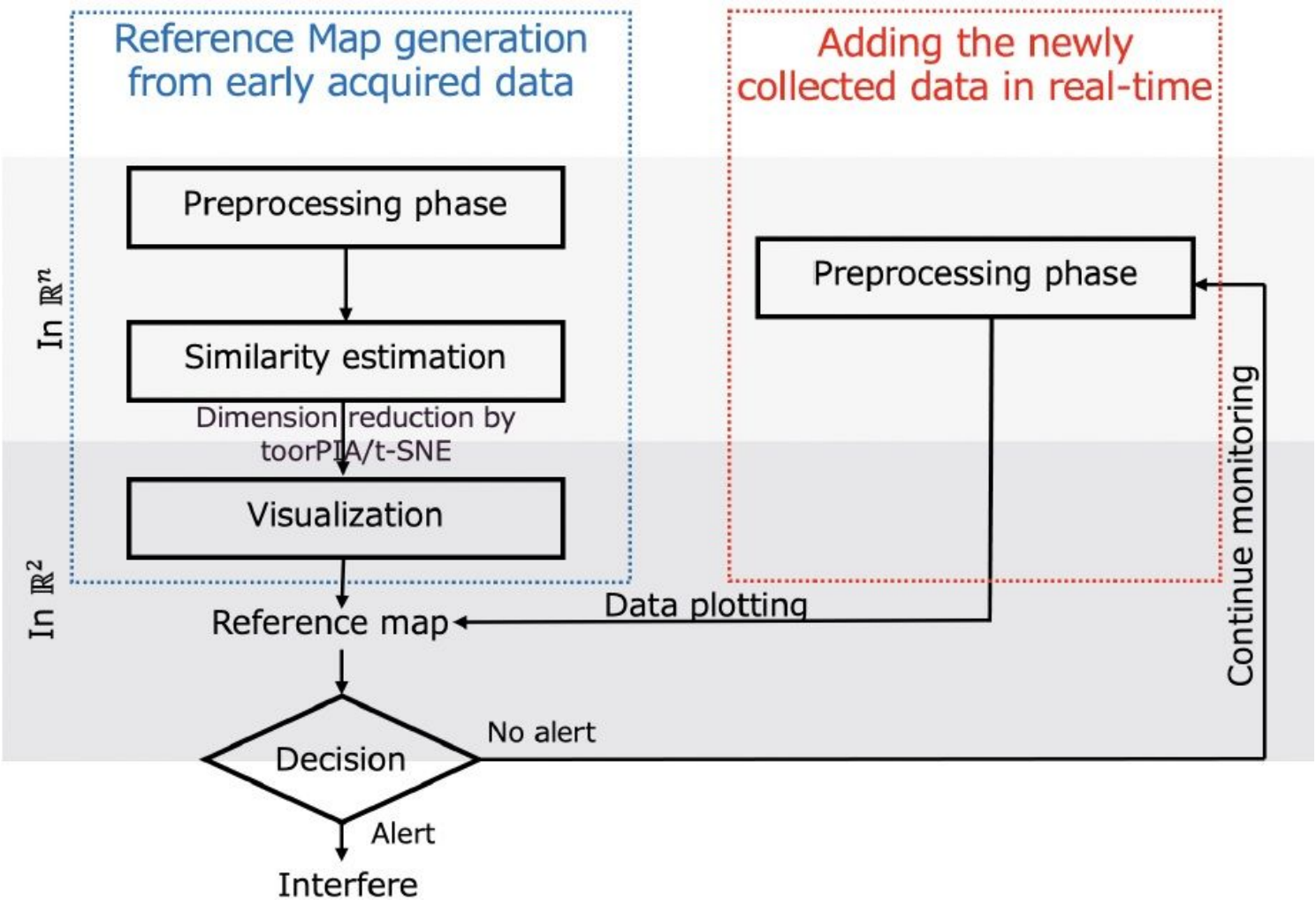


Figure 1

The process flow of our predictive scheme

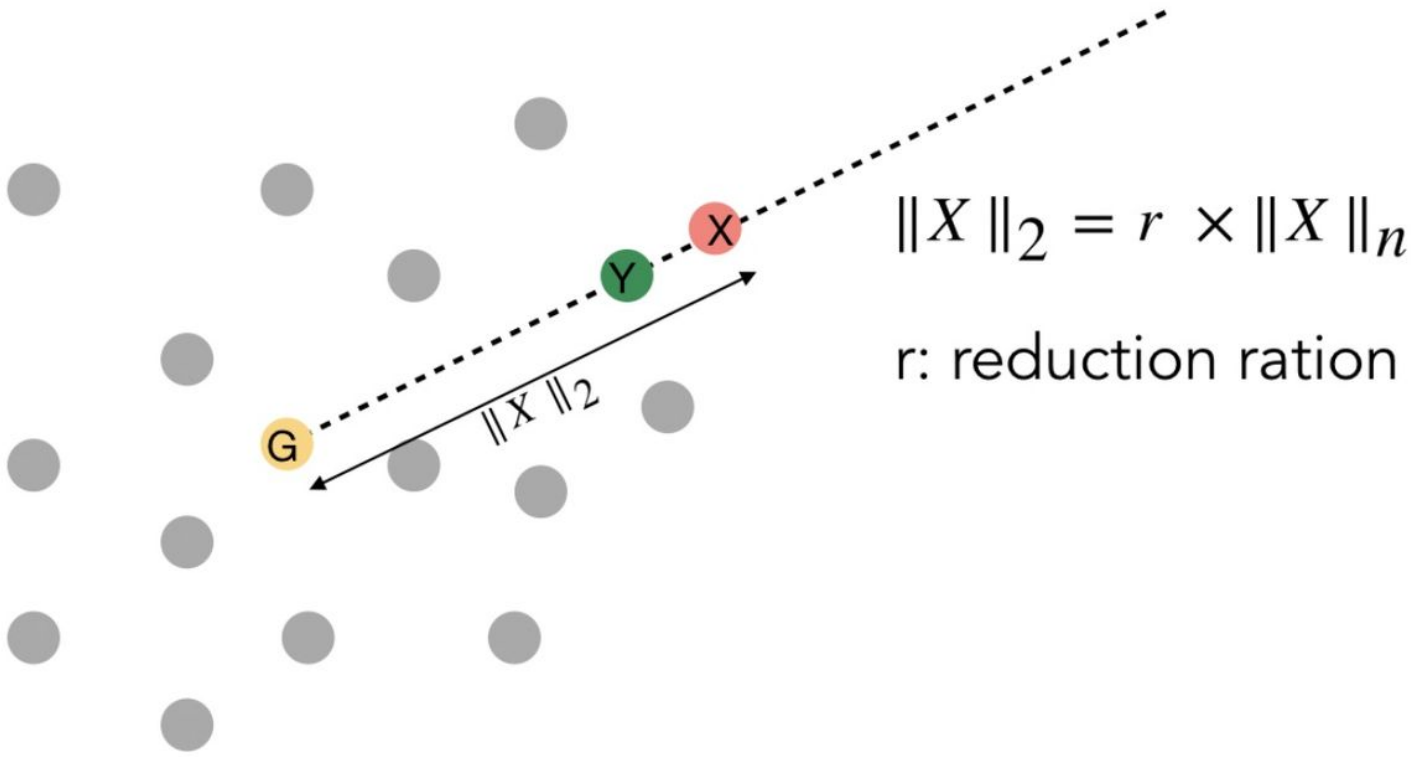


Figure 2

Placing the newly acquired data on the RM

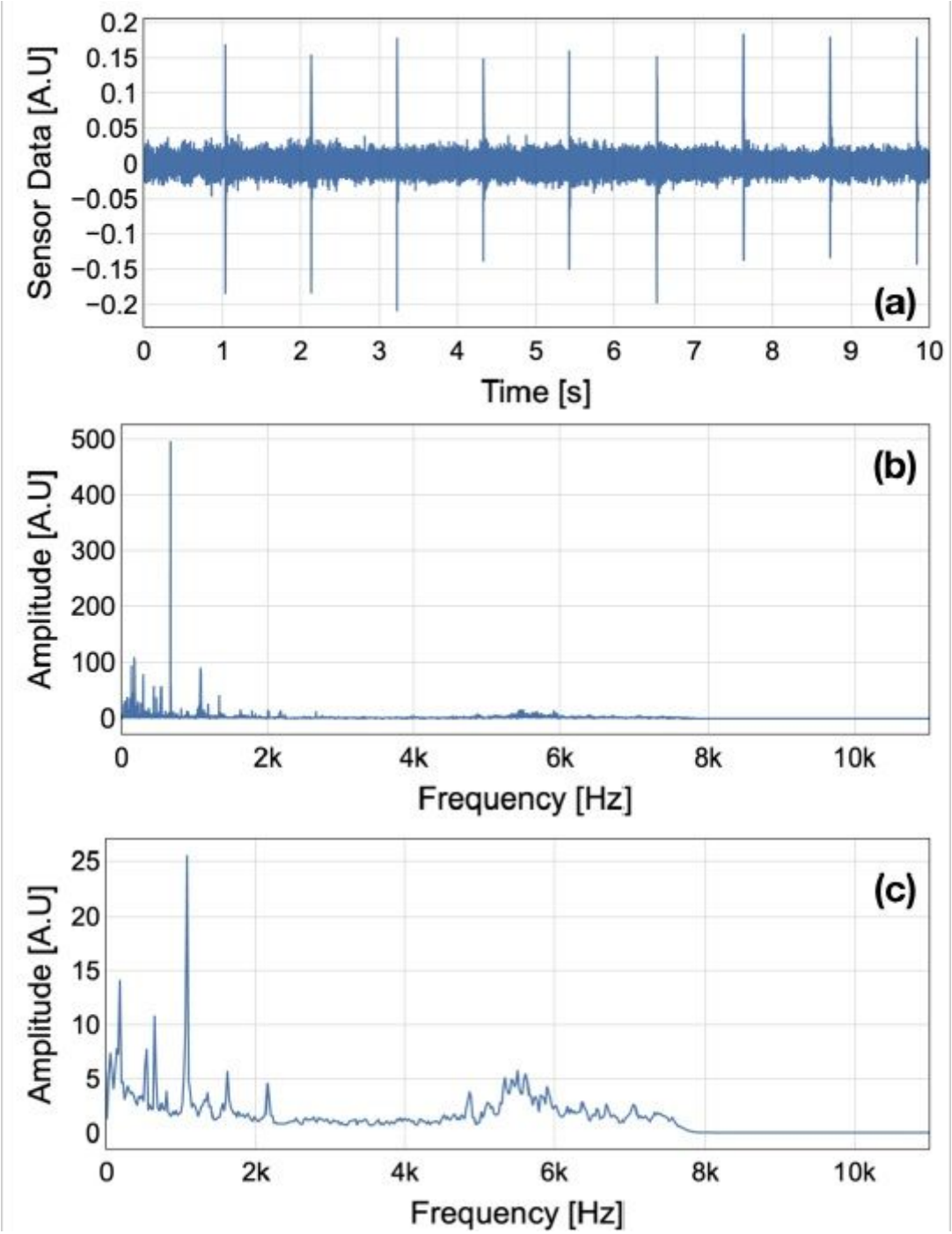


Figure 3

The evolution of an audio-clip through the preprocessing phase (a) time domain data extracted from the .wav file (b) original spectrum (c) smoothed spectrum

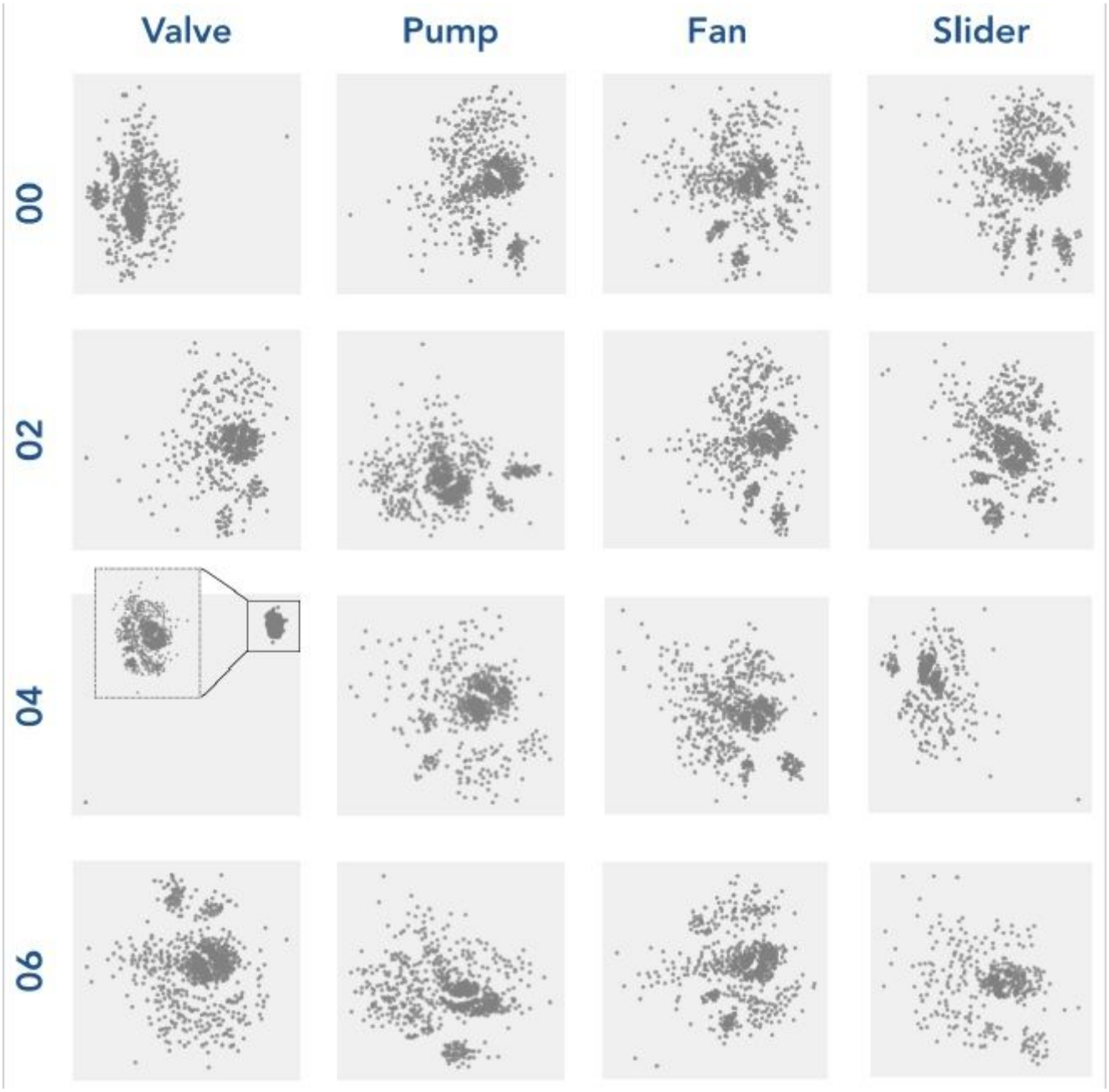


Figure 4

The obtained RMs for all machine types and IDs with FFR approach

Figure 5

Spectral distribution of the RMs of valve 00, pump 00, fan 00 and slider 00

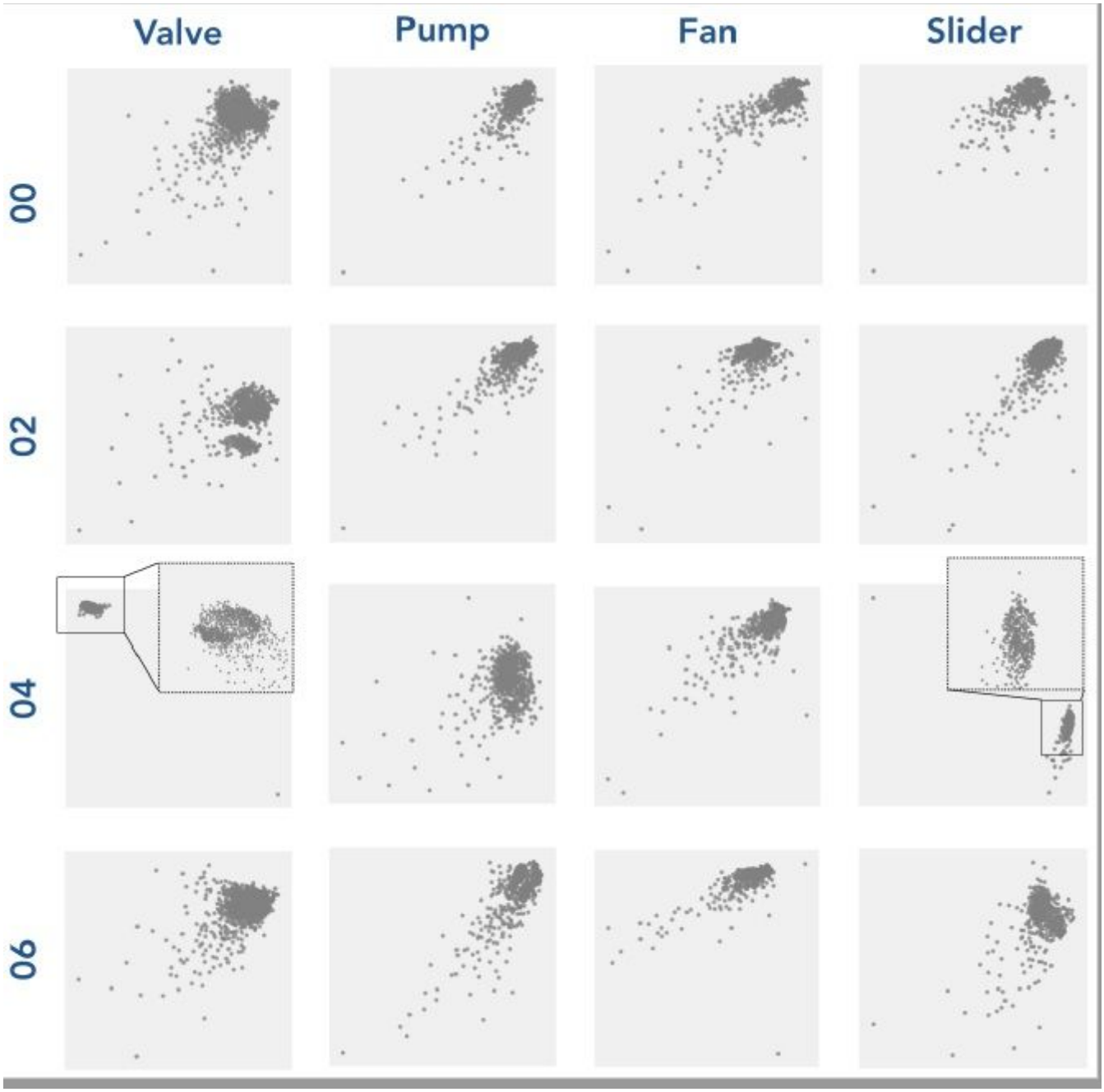


Figure 6

The obtained RMs for all machine types and IDs with HPF approach

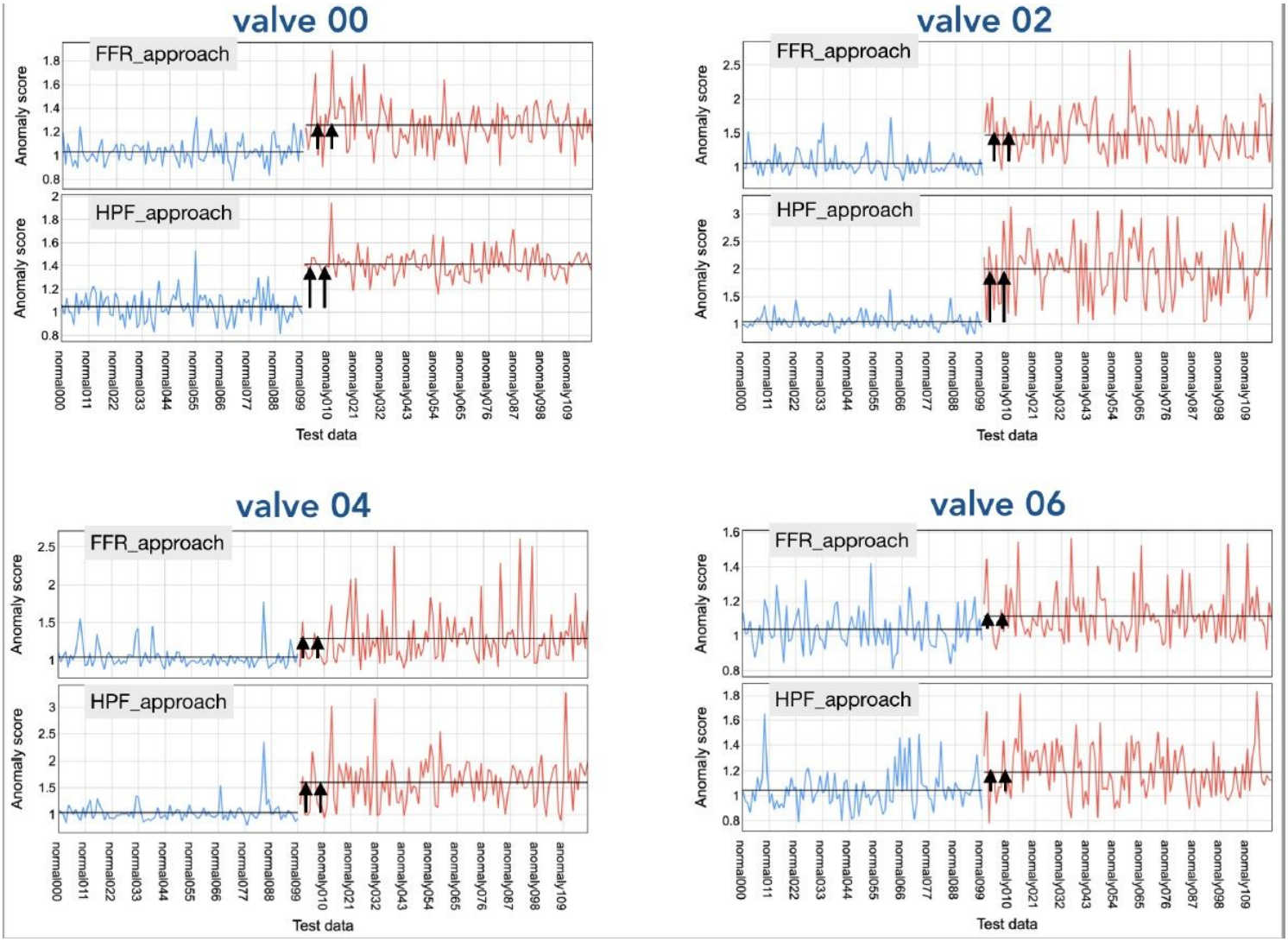


Figure 7

The obtained anomaly scores for valve

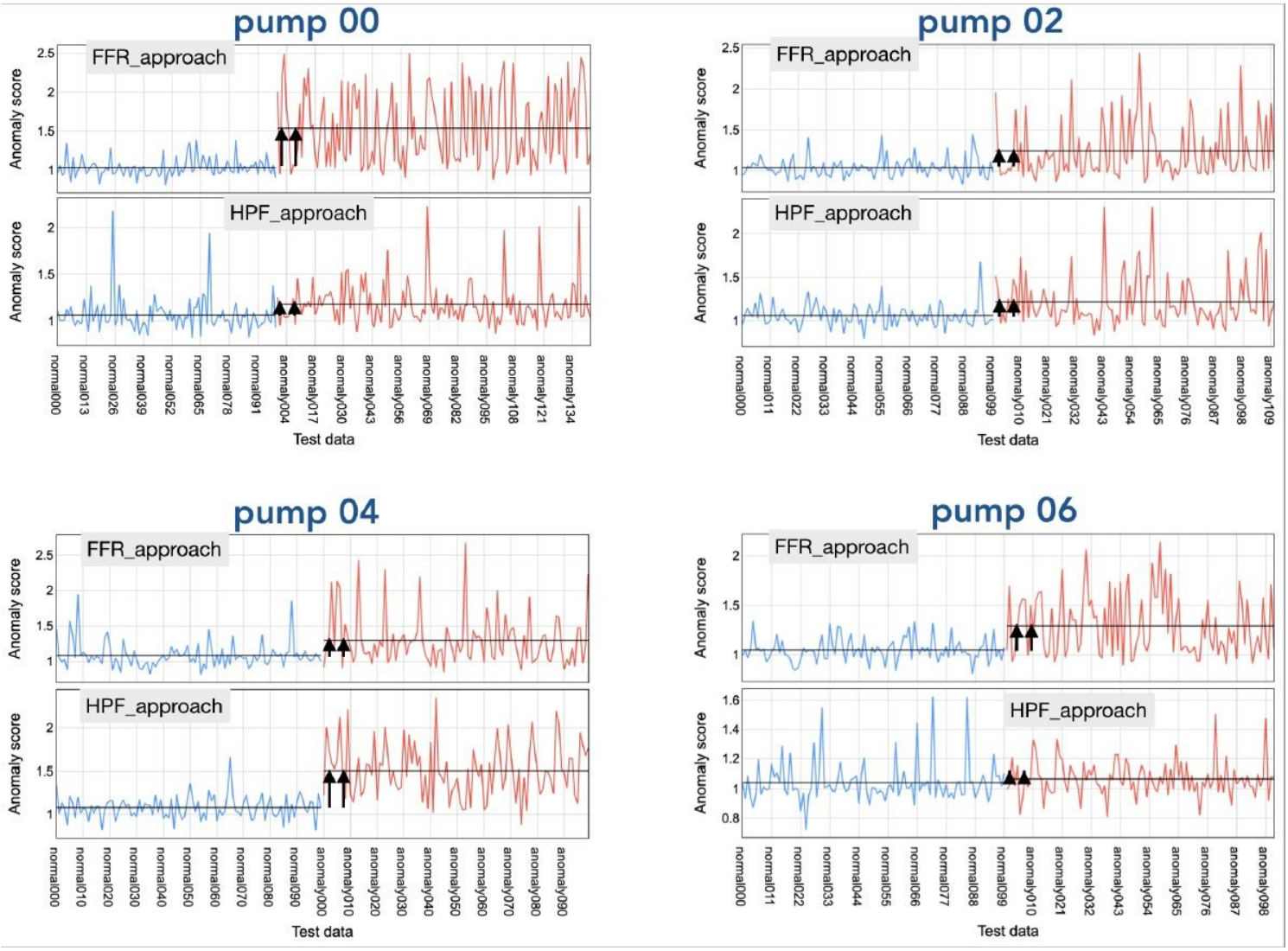


Figure 8

The obtained anomaly scores for pump

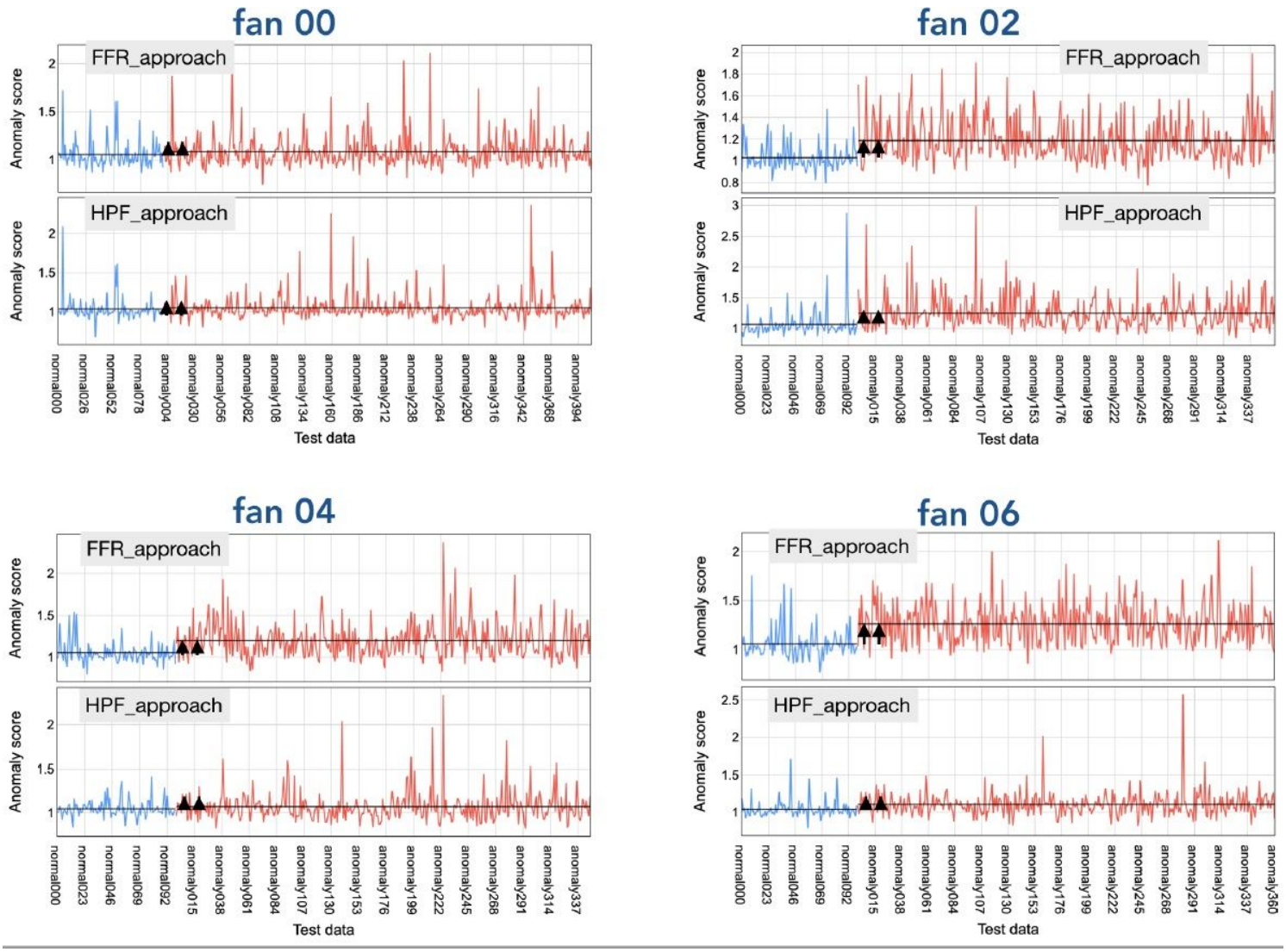


Figure 9

The obtained anomaly scores for fan

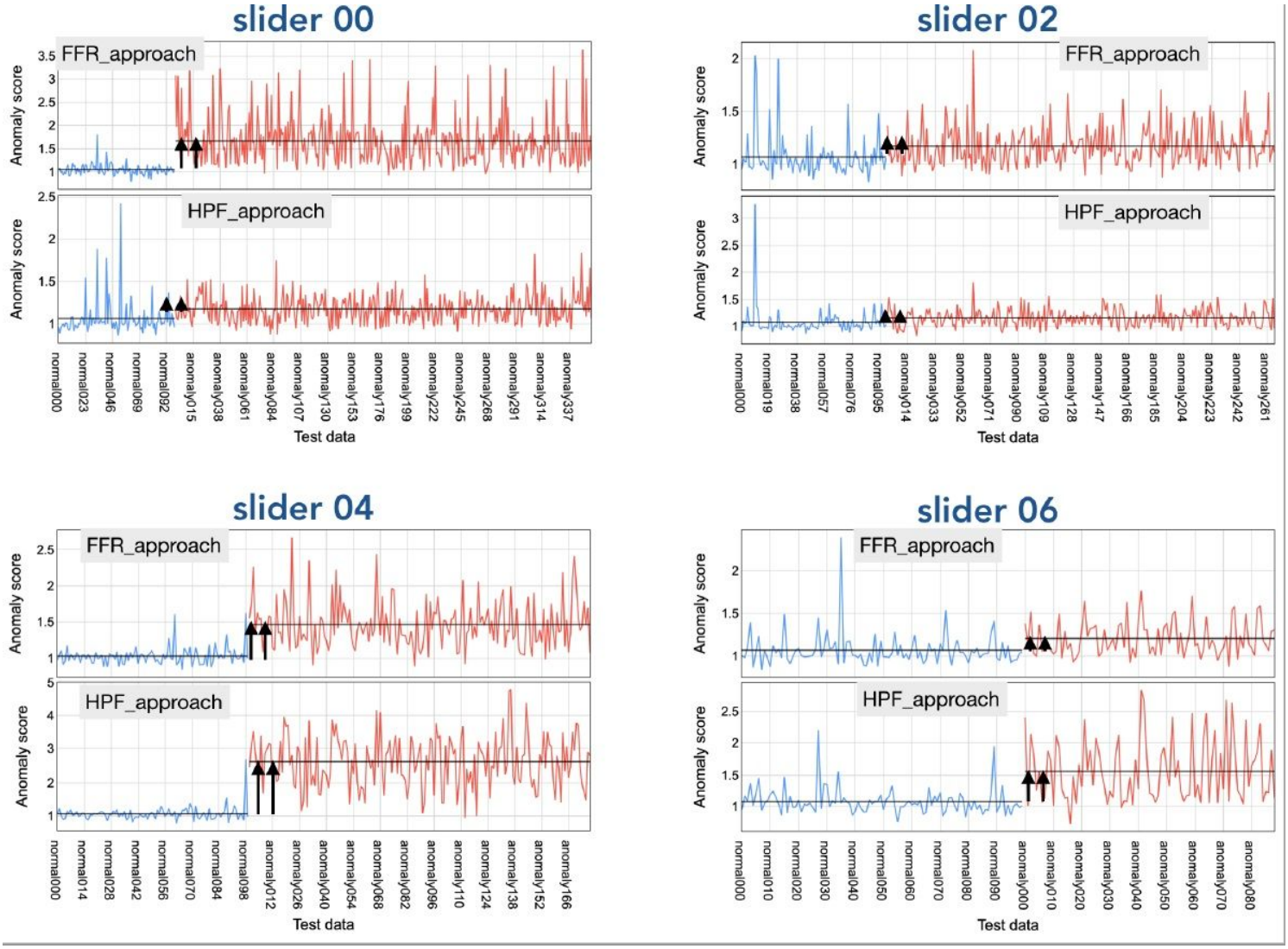


Figure 10

The obtained anomaly scores for slide-rail

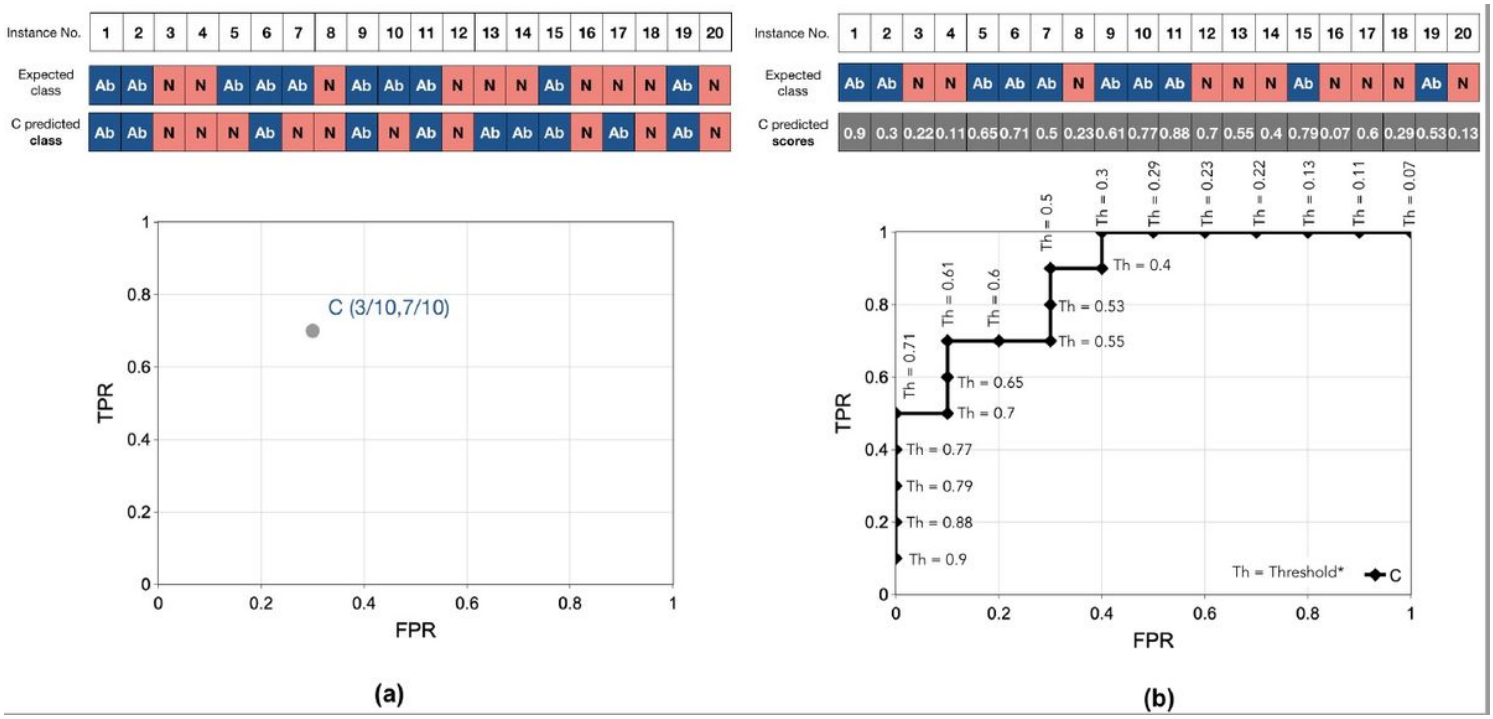


Figure 11

Presentation in the ROC space of (a) a discrete classifier and (b) a non-discrete classifier

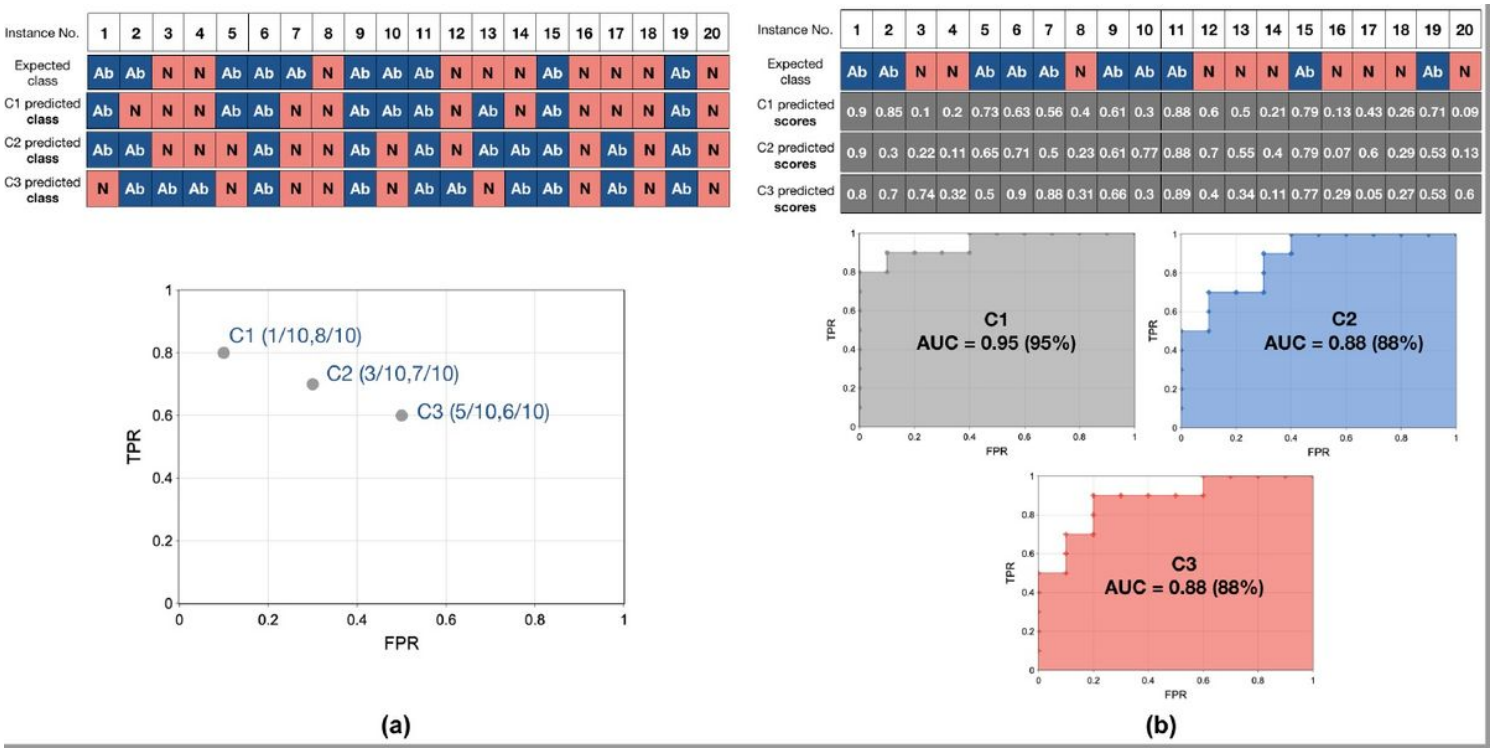


Figure 12

Comparison of (a) discrete classifier and (b) non-discrete classifier

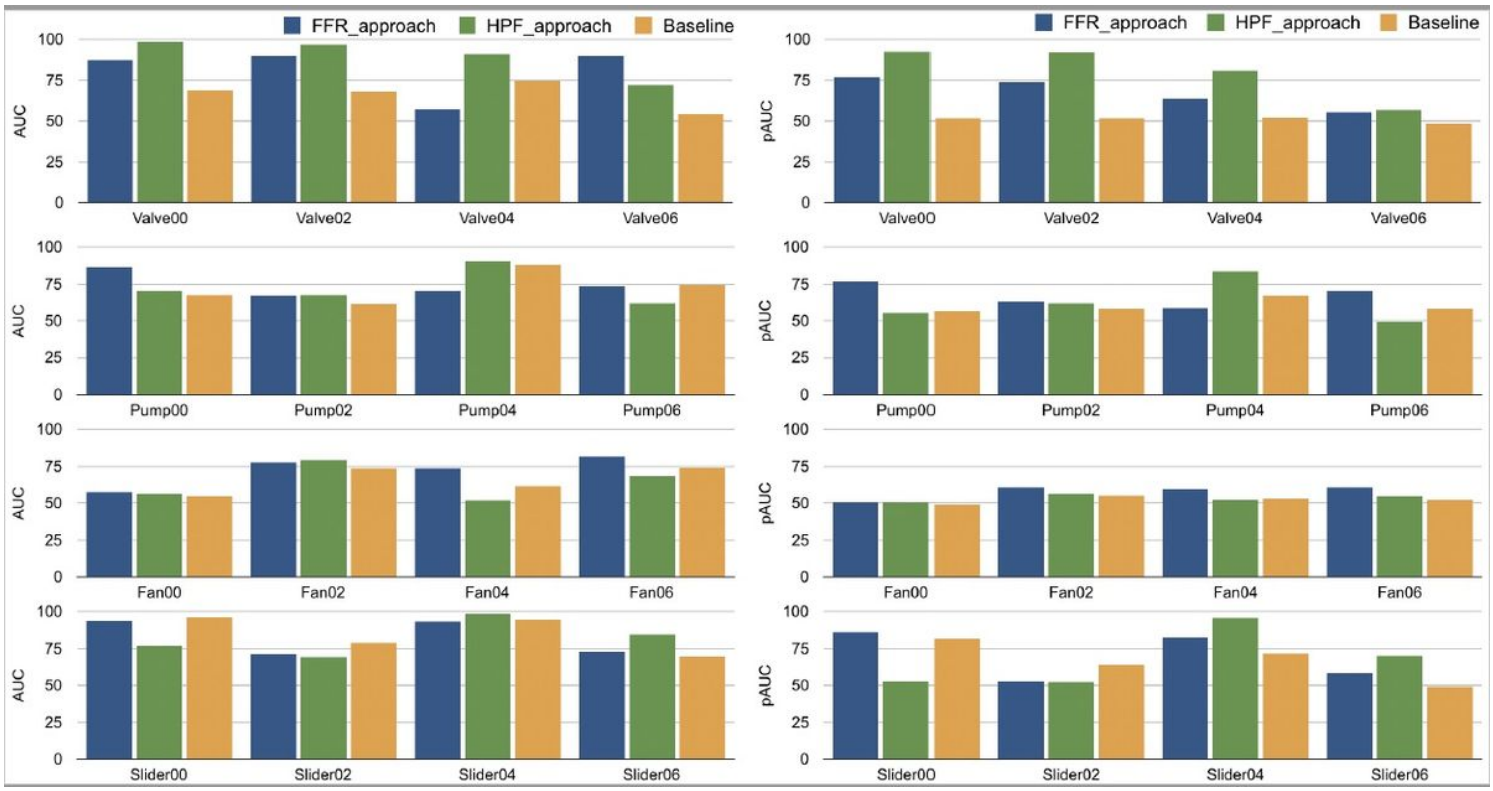


Figure 13

Comparison of our obtained AUC and pAUC values for each machine ID to the baseline system results

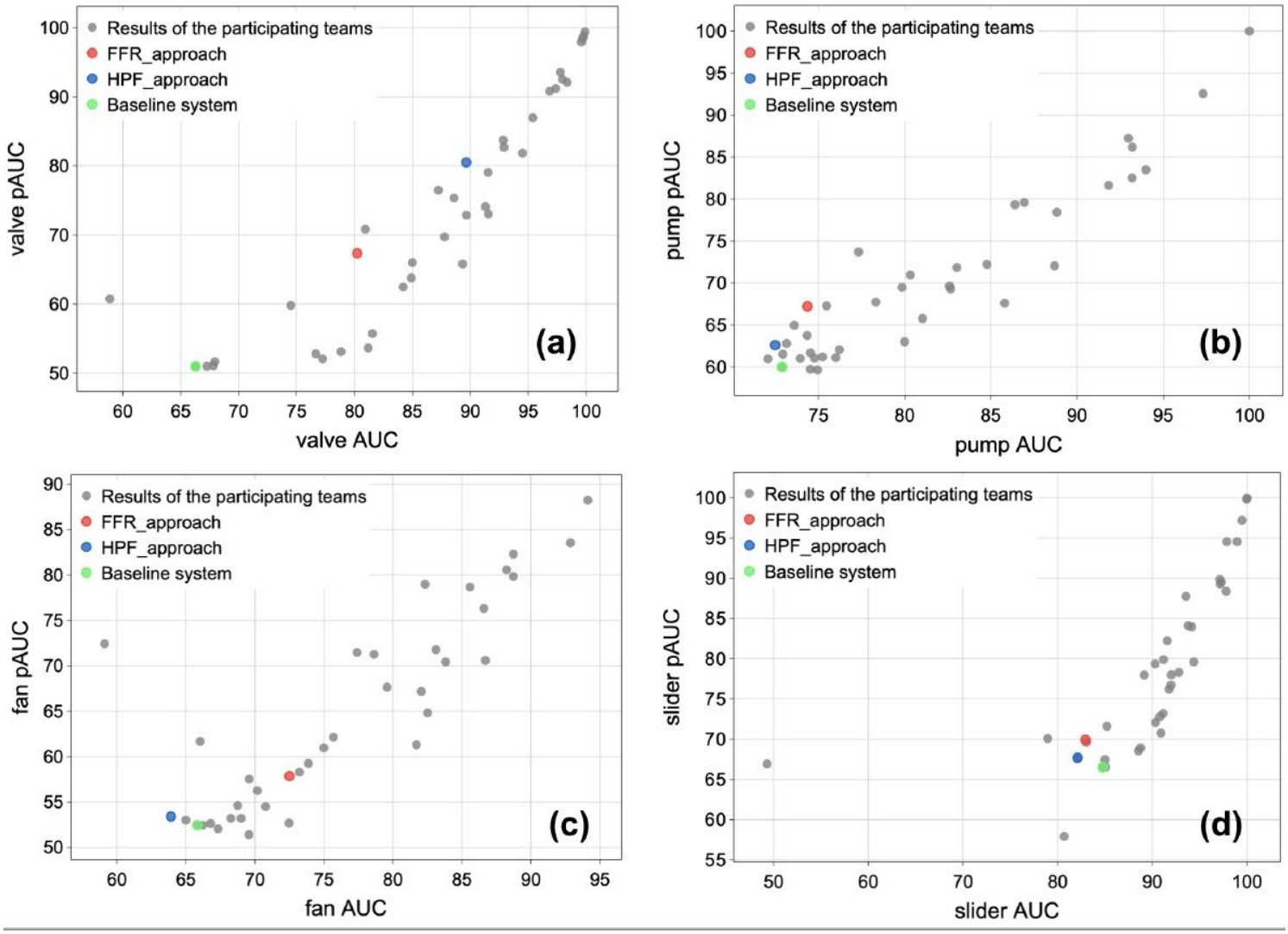


Figure 14

Positioning of our results compared to the participating teams for each machine type (a) valve, (b) pump, (c) fan and (d) slide-rail

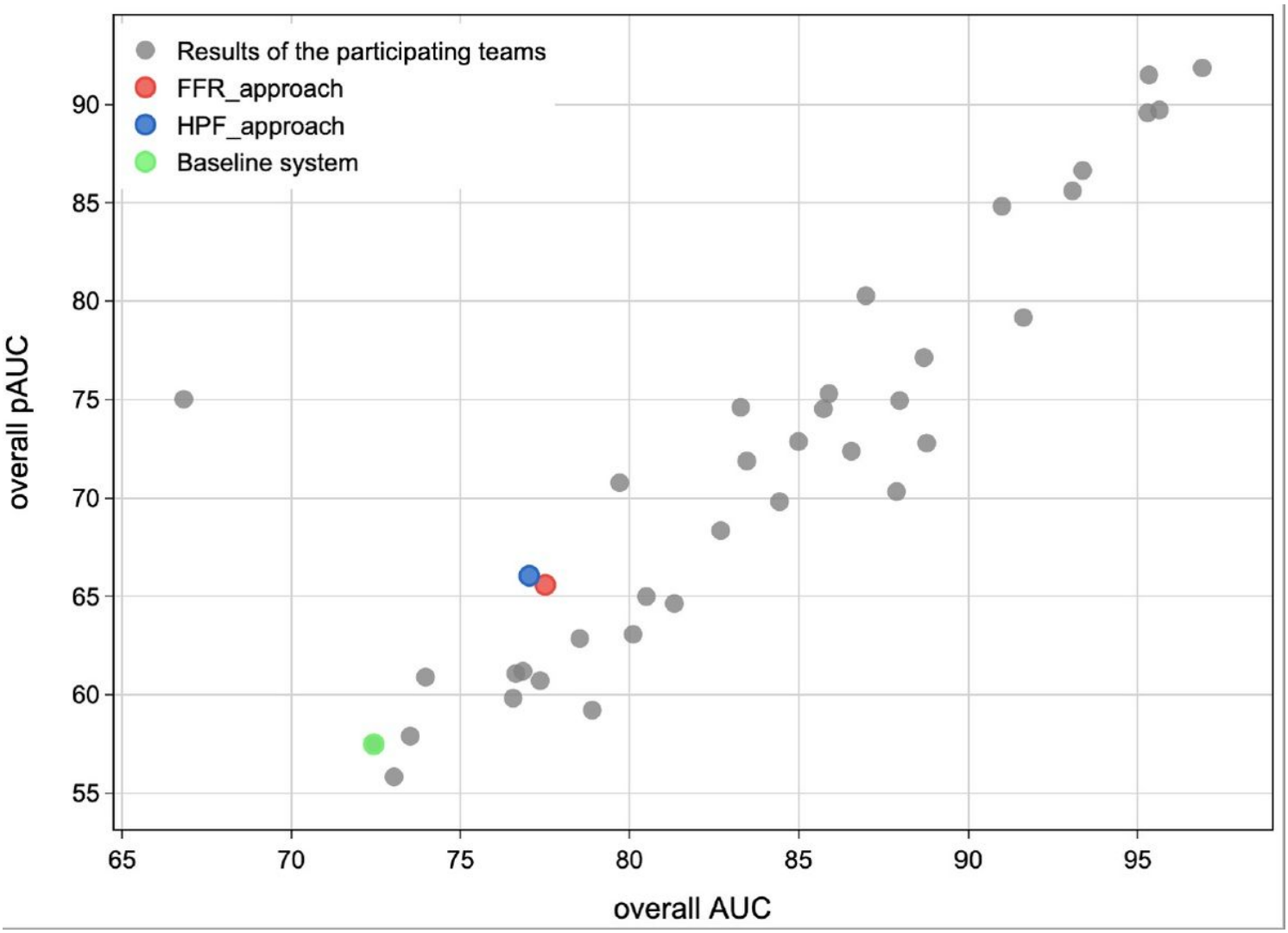


Figure 15

Overall positioning of our work compared to the participating teams