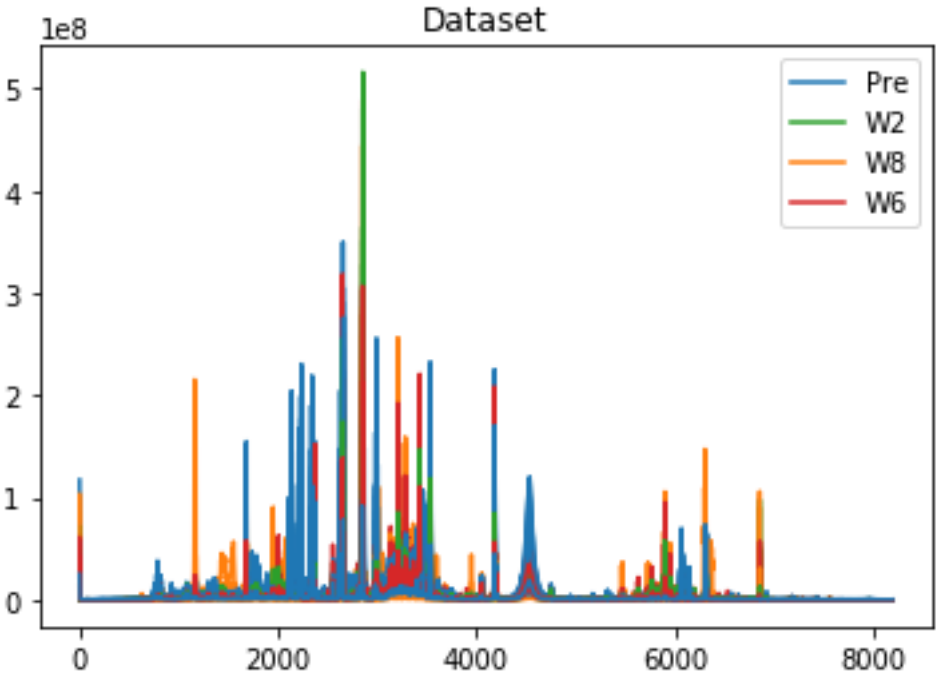
Spectroscopy signal classification

In this document presented an investigation of signal with chemical shift. The first aim is to see if unsupervised clusteristion can present convincing results. Second - if it is possible to set up a NN-based signal classifier.

1. Data overview

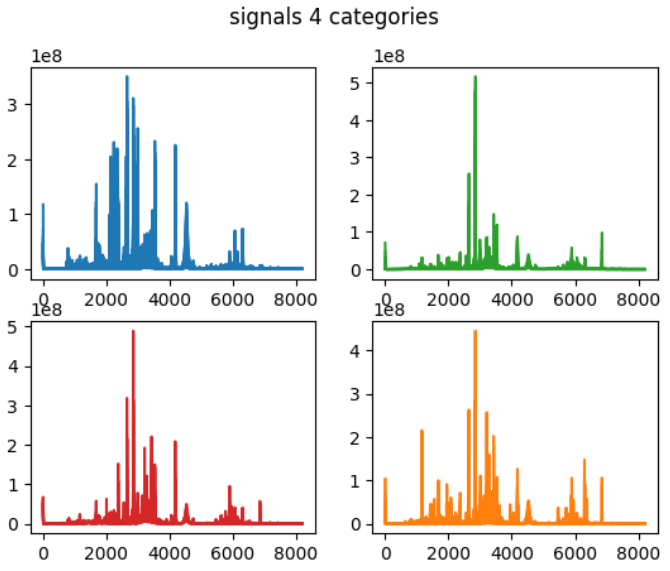
The disposable dataset has 59 samples belonging to 4 categories: 'Pre'(19 samples), 'W2'(14 samples), 'W8'(13 samples), 'W6'(13 samples).



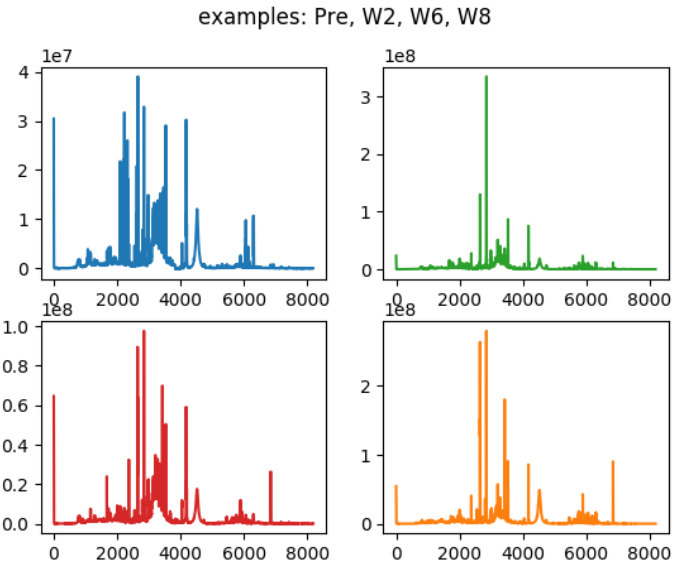
Pic.1

As seen on the Pic.1, there is no clear pattern for any of signal types. Pic.2 and Pic.3 show all the signal groups plotted separated and and a random example of each type of signal respectively.

All signals	Signals plotted by class



Pic.2



Pic.3

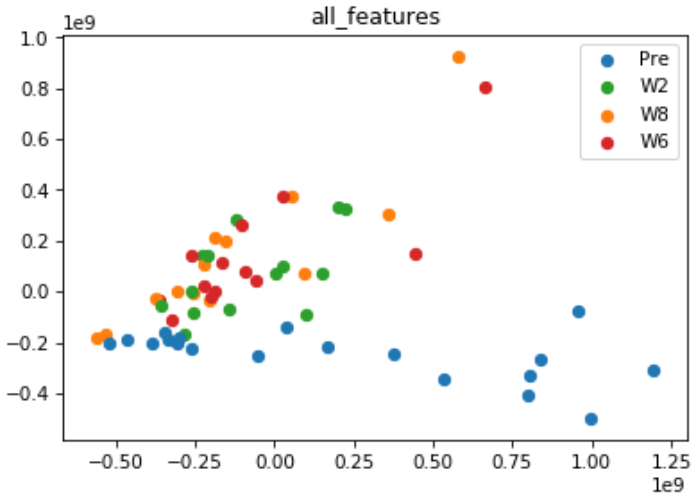
As seen, there is no obvious clear pattern which allows visually discern the signals and define which class they belonging to.

2. Clusterisation

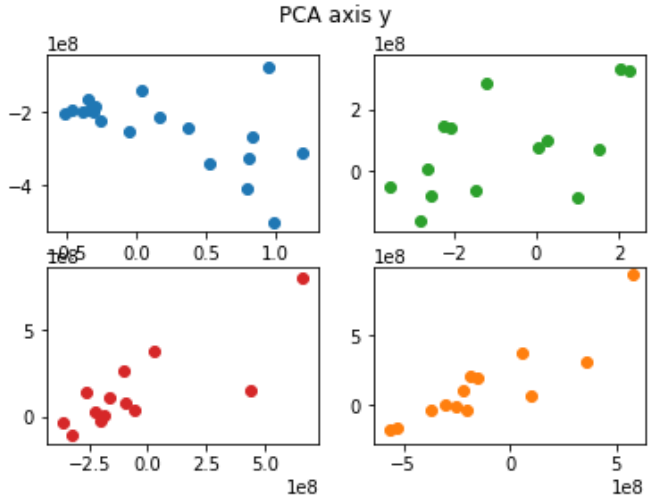
2.1 PCA

PCA was the first unsuperwised clusterisation method, apllied to the dataset:

All signals	Signals plotted by class



Pic.4



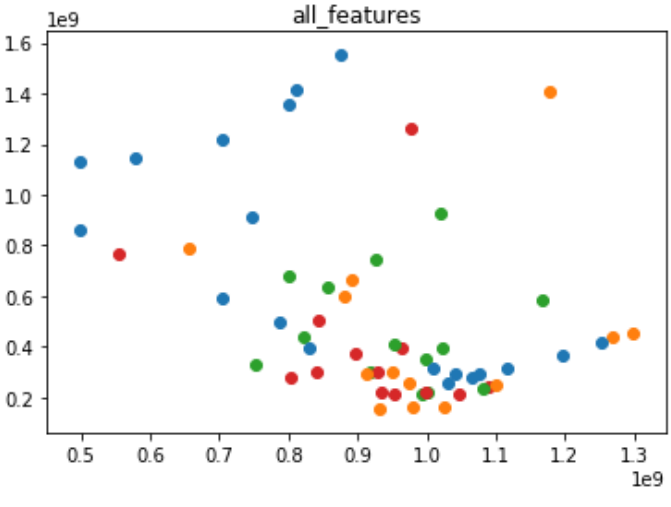
Pic.5

Only the class "Pre" shows a clear pattern in PCA. For the other three is impossible to define affiliation to any of other classes with accetable error rate.

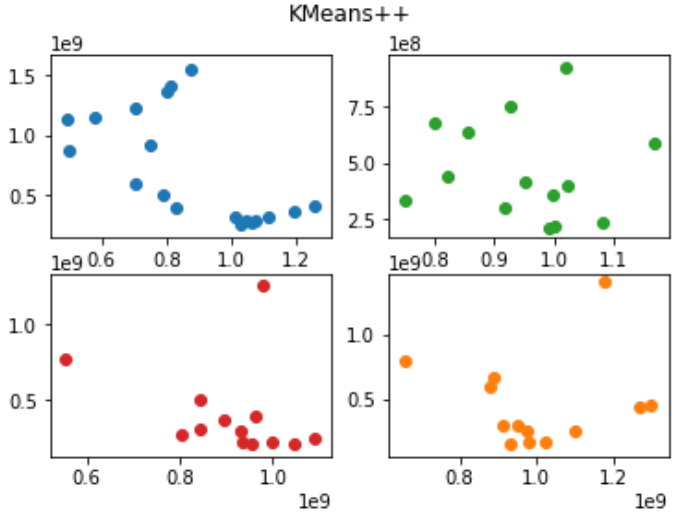
2.2 K-Mean Neighbours

K-Mean neighbours has neither shown satisfied results (Pic.6 and Pic.7)

All signals	Signals plotted by class



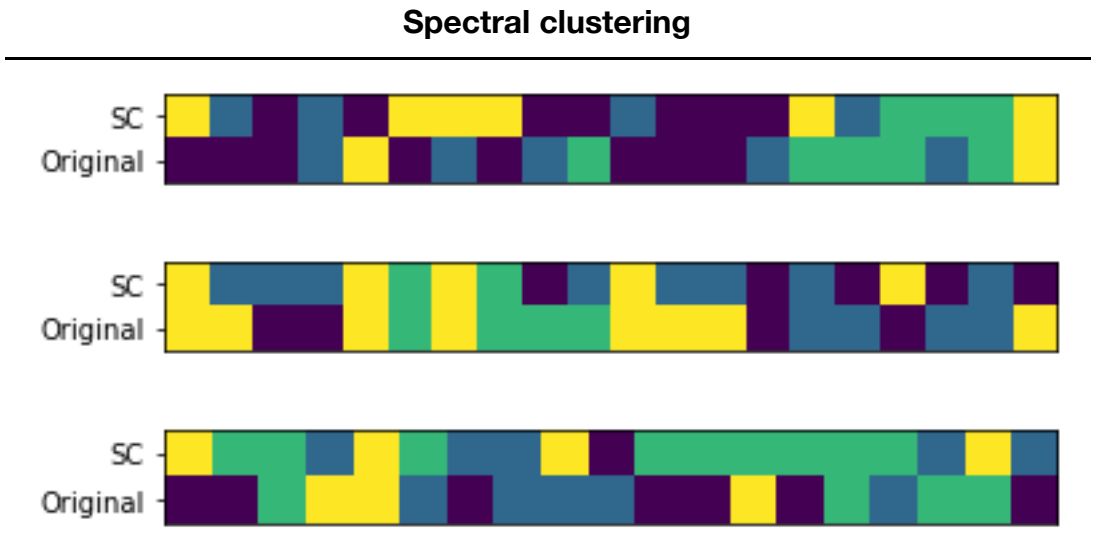
Pic.6



Pic.7

2.3 Spectral clustering

Try to project K-Means on Laplacian: spectral clustering (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.165.9323):



Pic.8

Resulting accuracy is about 33%

2.4 Clusterisation Results and Discussion

Presented clusterisation methods do not provide satisfactory results. Means either they do not have needed feature sensivity or there are simply no sistematic elements in presented classes. The next logical step would be to apply geometrical feature analysis to the dataset.

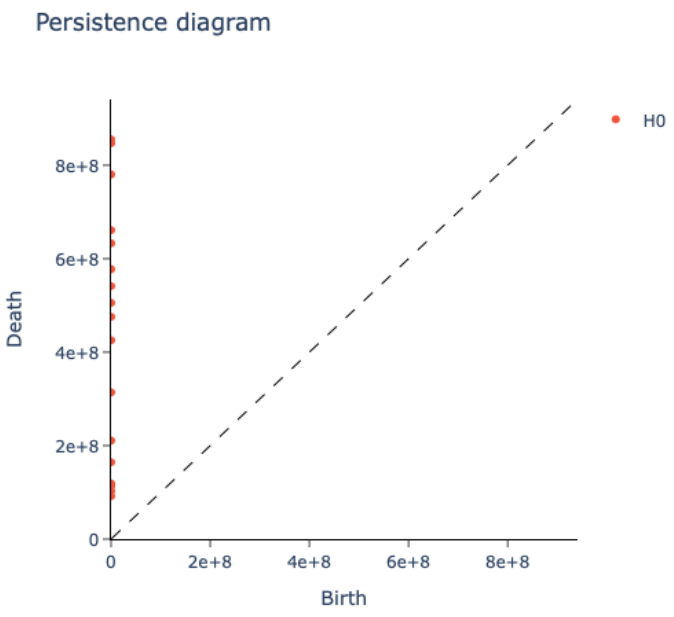
3. Geometrical Feature Search

In this analysis the presented dataset will be analysed class-wise to find out if there are some significant feature which could be enough for more sensitive clusterisation or classification.

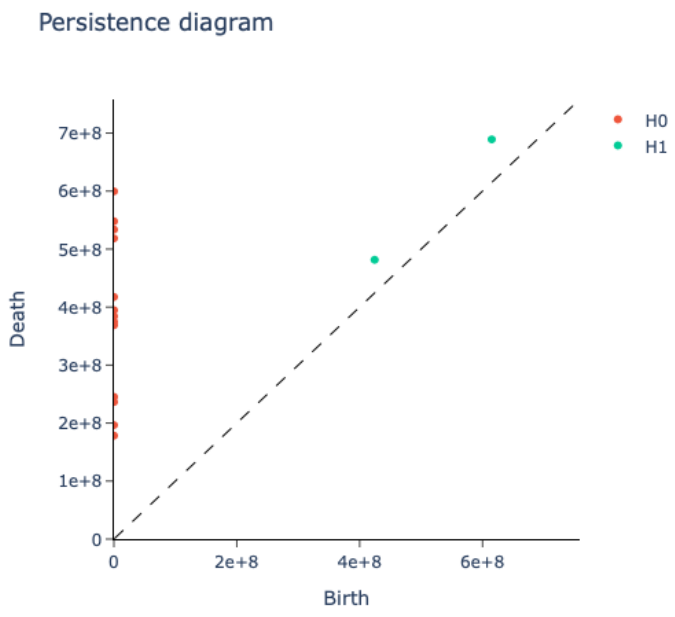
3.1 Persistence diagrams

Persistence homology (https://en.wikipedia.org/wiki/Persistent_homology) is well known as a tool which allows to extract true features rather than artifacts and/or noise. Results of homological analysis is presented on the Pic.9

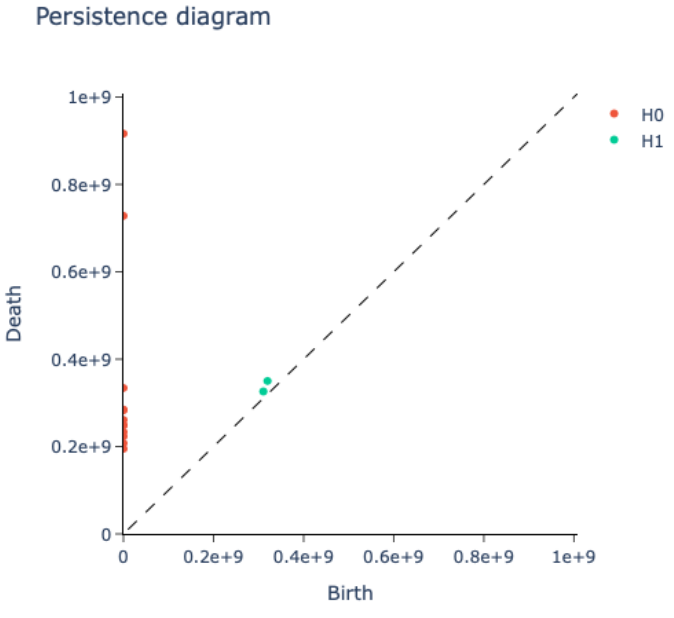
Pre



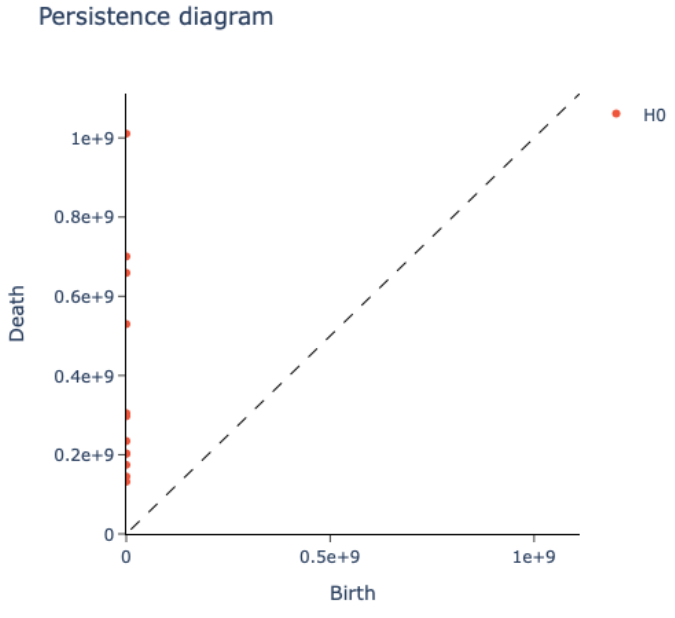
W2



W6



W8

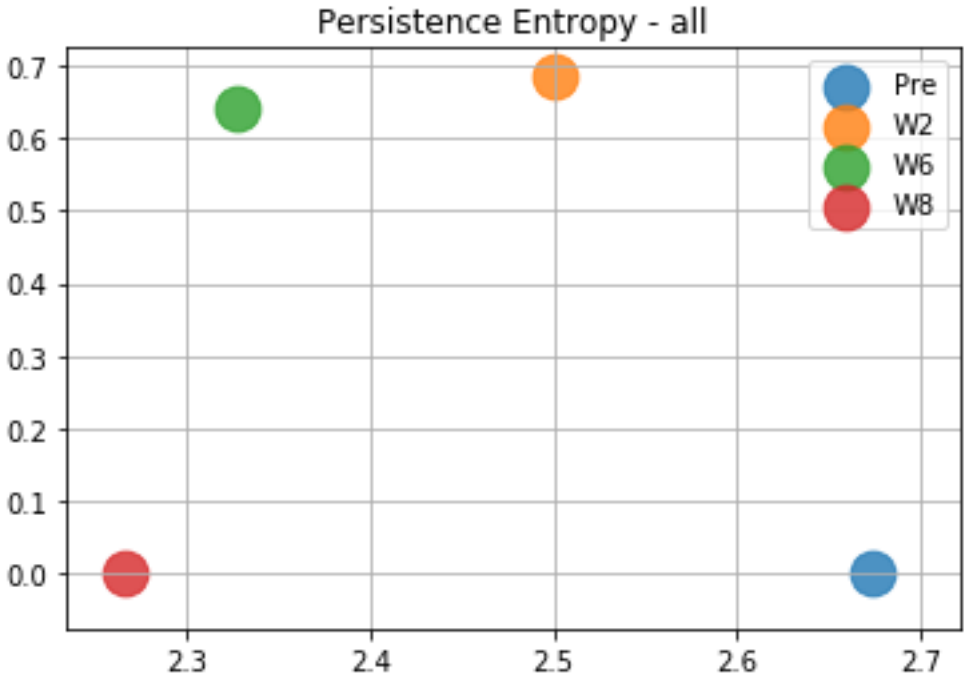


Pic.9

Persistent homology shows clear difference between presented samples.

3.2 Persistence entropy

To be sure about results obtained with persistence persistence diagrams, will be also applied [persistence entropy (https://link.springer.com/article/10.1007/s10844-017-0473-4)](https://link.springer.com/article/10.1007/s10844-017-0473-4) (Pic.10)

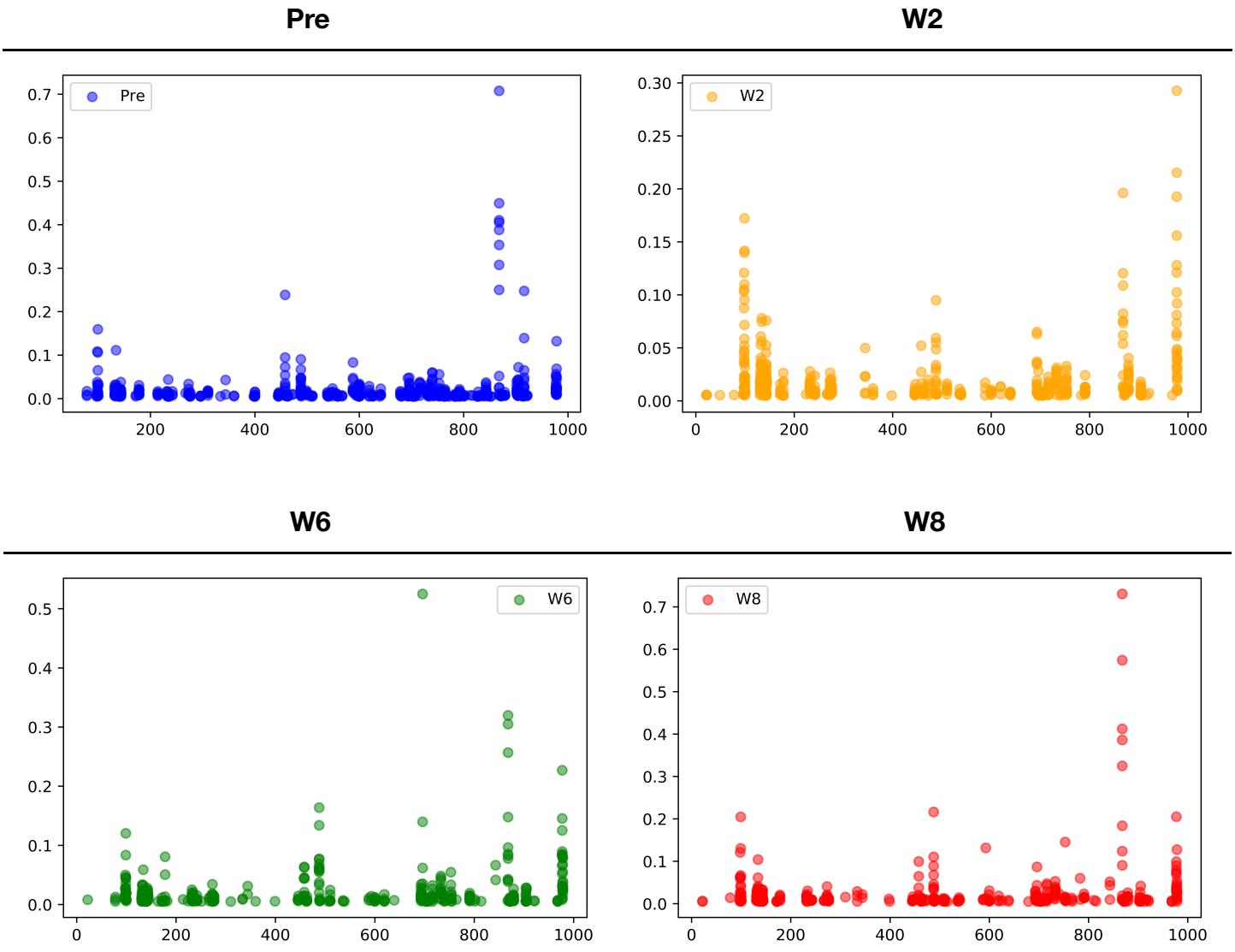


Pic.10

This method also shows clear feature difference for all the four classes.

4. Feature Extraction with Convolutional Neural Network

The next logical step would be prove if a deep learning method is able to catch the features of deformed signals in presented dataset (and experience says - yes, it can). But a check is needed. As feature extractor was used pretrained ResNet50, thresholded results are presented on the Pic.11:



Pic.11

NN seems to be sensitive to the features of signal (even with limited input, which in case of the most presented signal - "Pre" is only 19). Anyway, a classification worth to try.

5. Signal Classification with Neural Net

5.1 Dataset Augmentation

Unfortunately, the available dataset has limited number of samples (min 14, max 19) per class. Fortunately, data augmentation can be applied to the dataset.

In [ ]:

```
```