NA61/SHINE online noise filtering using machine learning methods

Anna Kawęcka

Wojciech Bryliński, Manjunath Omana Kuttan, Olena Linnyk, Janik Pawlowski, Katarzyna Schmidt, Marcin Słodkowski, Jakub Zieliński

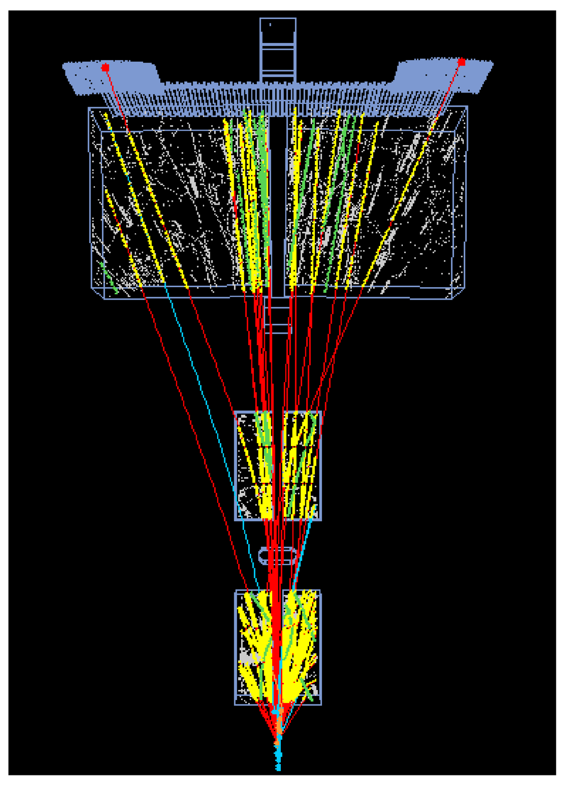
Motivation and methodology

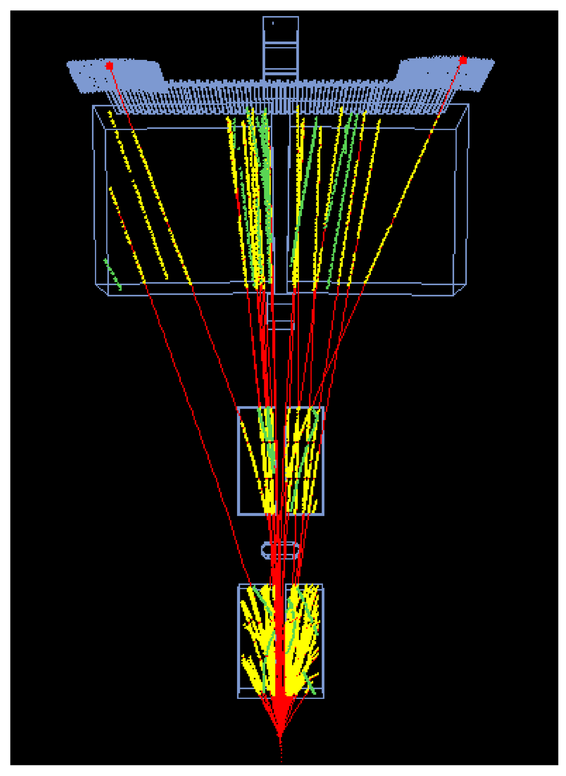
Motivation:

- Ongoing NA61/SHINE upgrade will allow for tenfold increase of data taking rate – up to 1 kHz,
- The fast and efficient tool for online noise filtering is needed to be able to handle the amount of data,
- One can use information from reconstructed TPC clusters and machine learning techniques as the classification tool for clusters reconstructed from noise data.

Methodology:

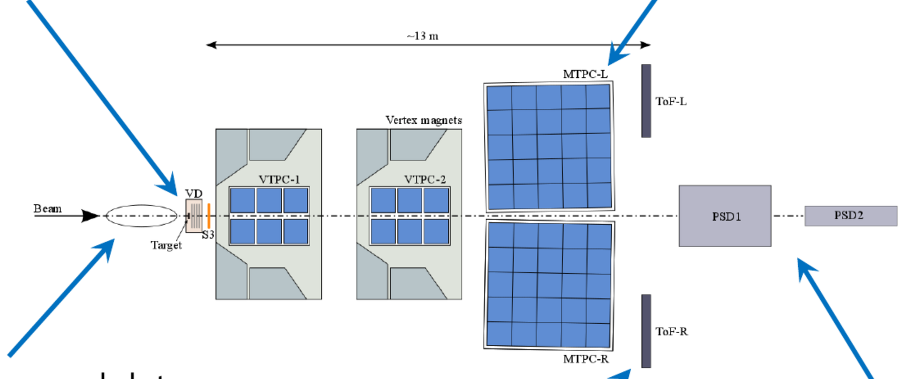
- Decision tree and dense neural networks (NN) based on reconstructed clusters features like total charge, positions, maxADC etc.
- Convolutional NN based on properties of clusters raw data.

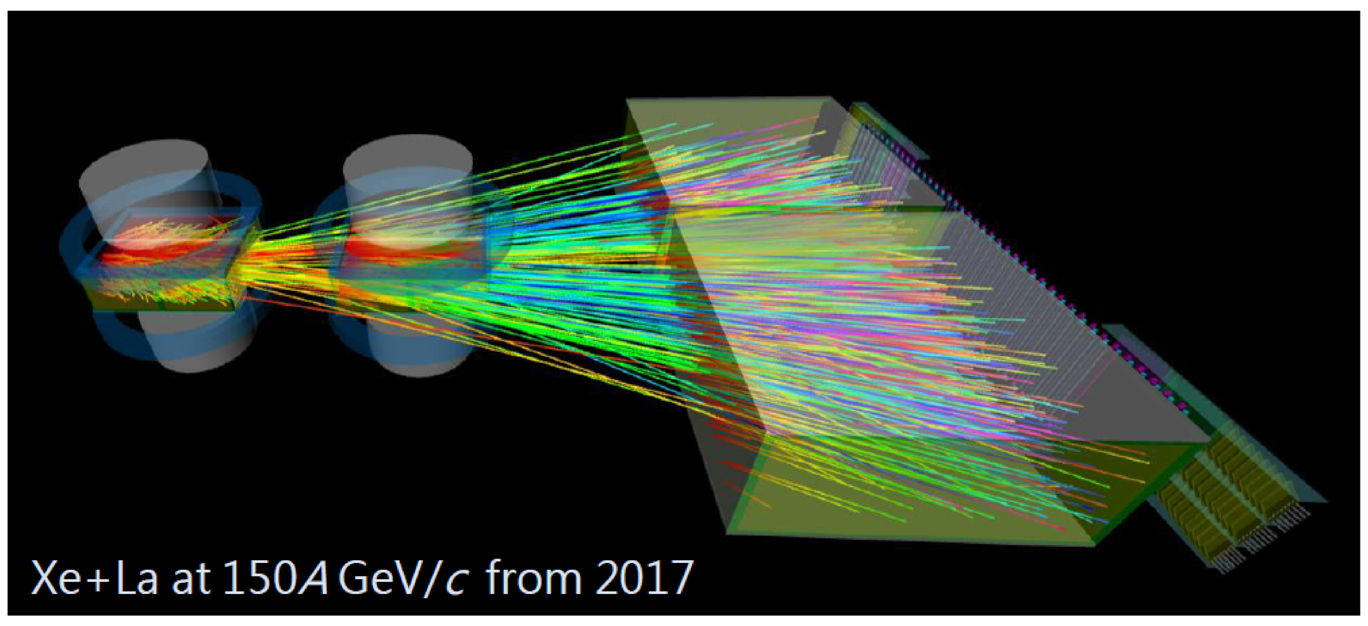




NA61/SHINE

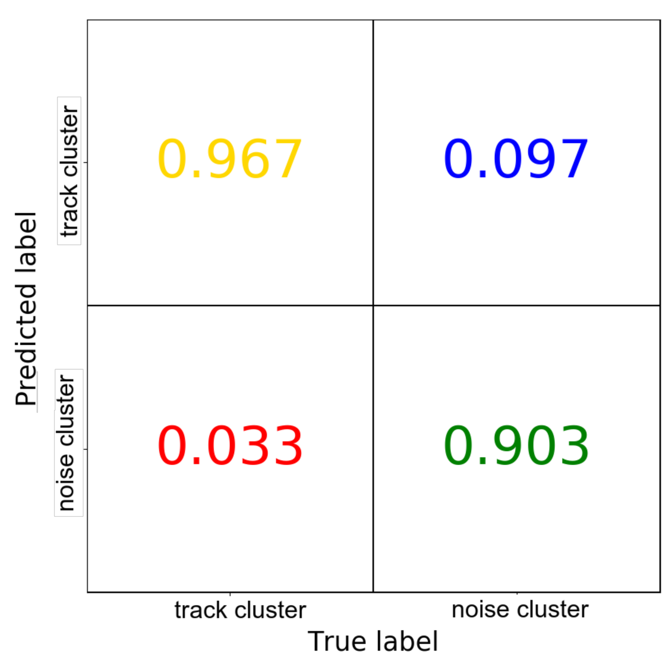
NA61/SHINE (SPS Heavy Ion and Neutrino Experiment) is a multi-purpose spectrometer optimised to study the hadron production in different types of collisions: p+p, p+A, A+A. The experiment is currently undergoing a significant upgrade which will improve the acceptance and efficiency of the detector and allow for tenfold increase of data taking rates:





Confusion matrix

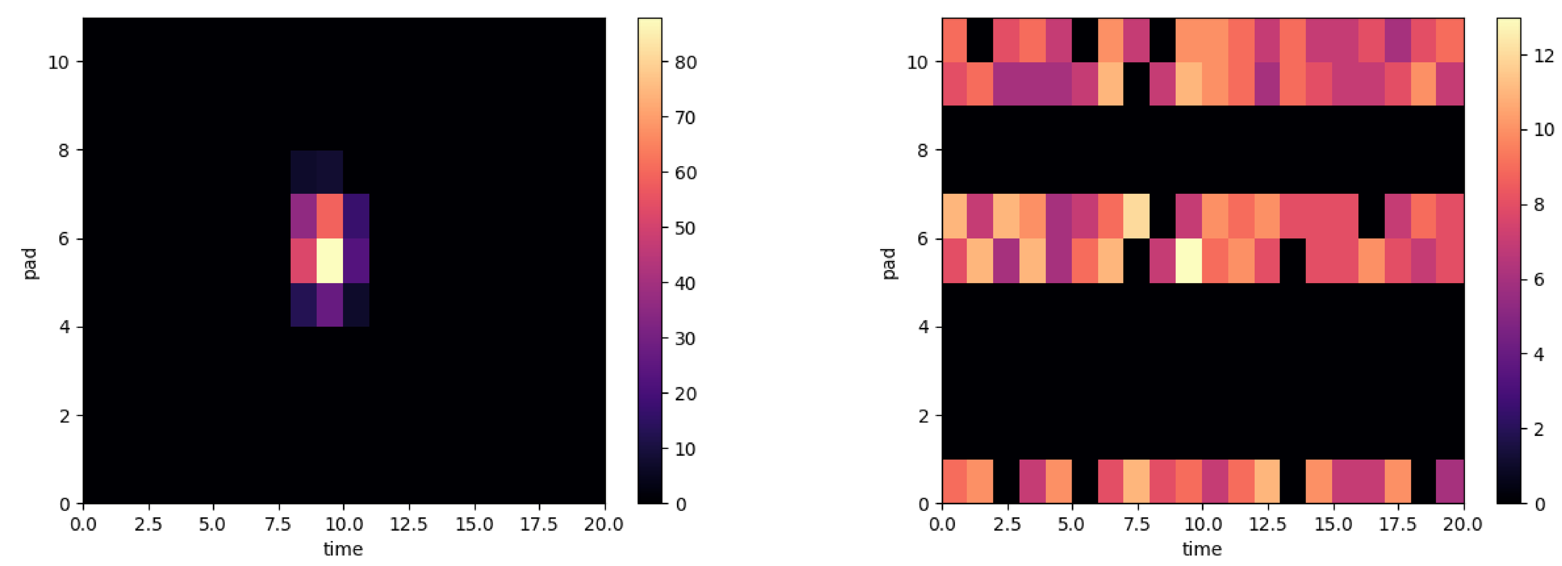
The main goal is to keep the high accuracy of good clusters recognition and keep the fraction of good clusters classified as noise (false negatives) as low as possible. The example confusion matrix for convolutional NN is shown below.



Data preparation

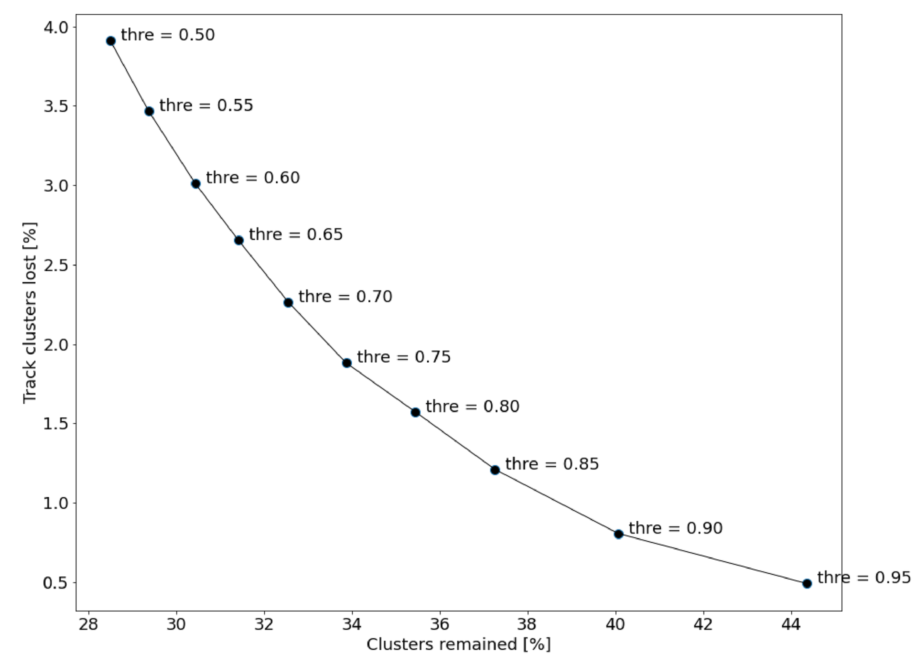
The first step of data preparation is the reconstruction of local tracks and division of the TPC clusters into 2 groups: good clusters (belonging to track) and noisy clusters. The divided clusters are used to train the classification algorithms.

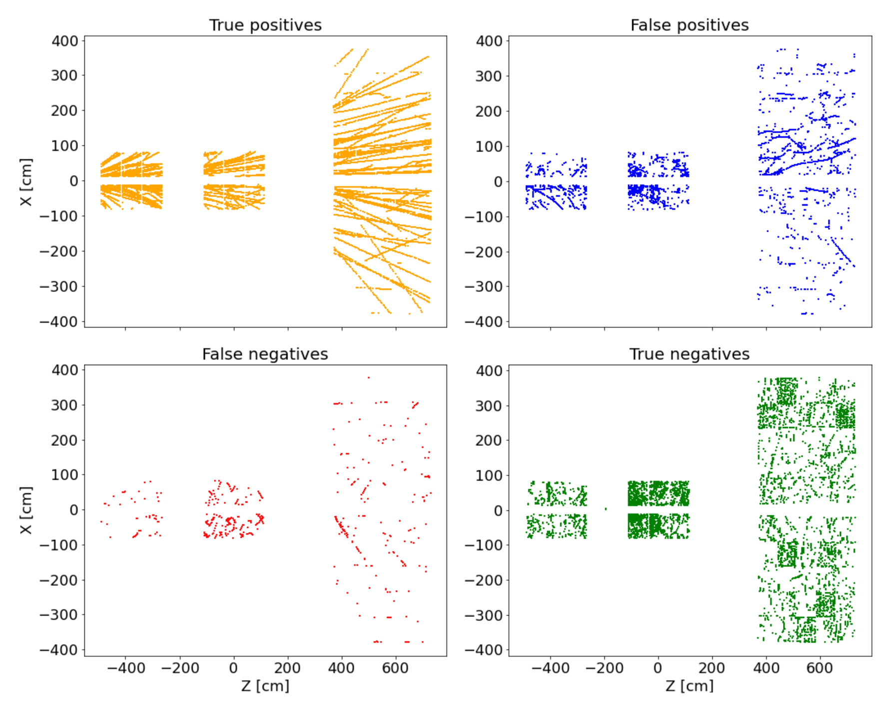
- The first approach uses the reconstructed properties of the clusters (total charge, positions, maxADC, etc.)
- The second approach is based on clusters raw data. For each cluster the pad ID and timestamp of max charge deposit is found and the signals from ±5 neighbouring pads and ± 9 timestamps are included. The example of good (left) and noisy (right) clusters are presented below.



Prediction results

The result of the classification for both methods is the probability that a given cluster was reconstructed from noise. If the probability is higher than a threshold, the cluster is removed from the data. The threshold can be selected in a way to minimise the loss of track clusters and maximise the reduction of noise. The example results for decision tree are presented in the pictures below. The left plot shows the percentage of lost track clusters plotted as a function of the percentage of remained clusters for different threshold values. With 80% probability threshold, the loss of track clusters is only around 1.5%, obtaining the data reduction by over 60%. The visualization of the confusion matrix for 80% probability threshold is shown in the right figure.





All details of the implemented algorithms together with the comparison of results can be found here:

Conclusions

- NA61/SHINE will need online noise filtering tool for future data taking and machine learning approach is a very efficient and robust solution.
- Two approaches were tested – using reconstructed and raw clusters information and both give good results with prediction accuracy over 90%.
- The algorithms are being implemented in the data acquisition (DAQ) software as online tools and the most efficient will be used during data taking.