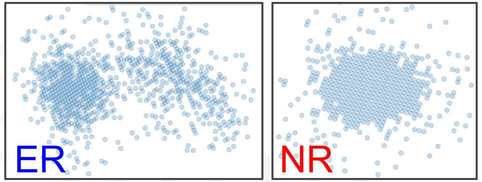
Electron and Nuclear Recoil Discrimination Studies

Atul Prajapati
Supervisor: Elisabetta Baracchini

Discriminating Variables

Discriminating Variables

- Standard Deviation of Charge Distribution 2D(SDCD_2D):

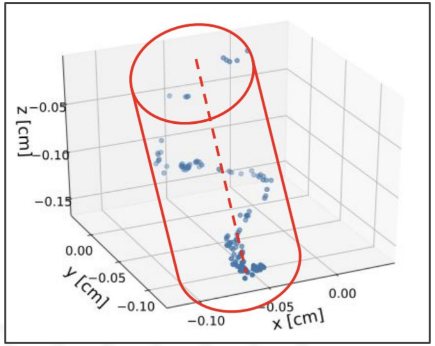


$$SDCD = \sqrt{\frac{\sum_{i=1}^{N}(\mathbf{r_i} - \bar{\mathbf{r}})^2}{N}}.$$

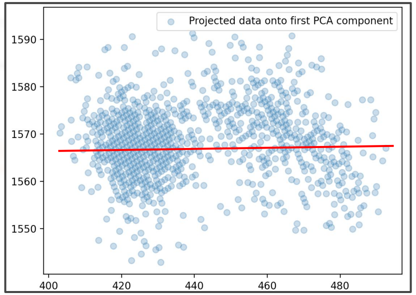
  - Electron recoils (ER) are longer, so the spread of charge is higher for ER when compared to Nuclear recoils (NR).
- Charge Uniformity 2D (ChargeUnif_2D):
  - For each point within the charge distribution, find the average distance to all other points.
  - ChargeUnif_2D is standard deviation of values computed in step 1.
  - Electron recoils tend to have charge distribution which is dense in some areas and sparse in other areas, while nuclear recoils are generally uniform.
- Maximum Density 2D (MaxDen_2D):
  - MaxDen is the value of most intense pixel from the image after rebinning it by a factor 2.
  - Electrons lose their energy at a slower rate than nuclei, this suggests that electron recoils are travel greater distance between interactions resulting in more sparse energy distribution.

[1] Observables for recoil identification in gas TPCs, arXiv:2012.13649v1

Discriminating Variables



- ❖ Cylindrical Thickness 2D (CylThick_2D):
  - ➢ For each charge, calculate the squared distance from the principal axis.
  - ➢ CylThick is the sum of all squared distances.
  - ➢ It is a measure of how much a recoil track deviates from the trajectory approximated by the principal axis.
  - ➢ Electrons experience far more scattering compared to nuclei, so principal axis approximates NR's trajectory much more accurately than it does for ER.

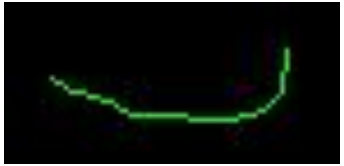


- ❖ Length Along Principal Axis 2D (LAPA_2D):
  - ➢ Project all the points in the charge distribution on to the principal axis.
  - ➢ LAPA is the difference between maximum and minimum projected value.
  - ➢ ER are longer compared to NRs, therefore projection is also longer.

Discriminating Variables

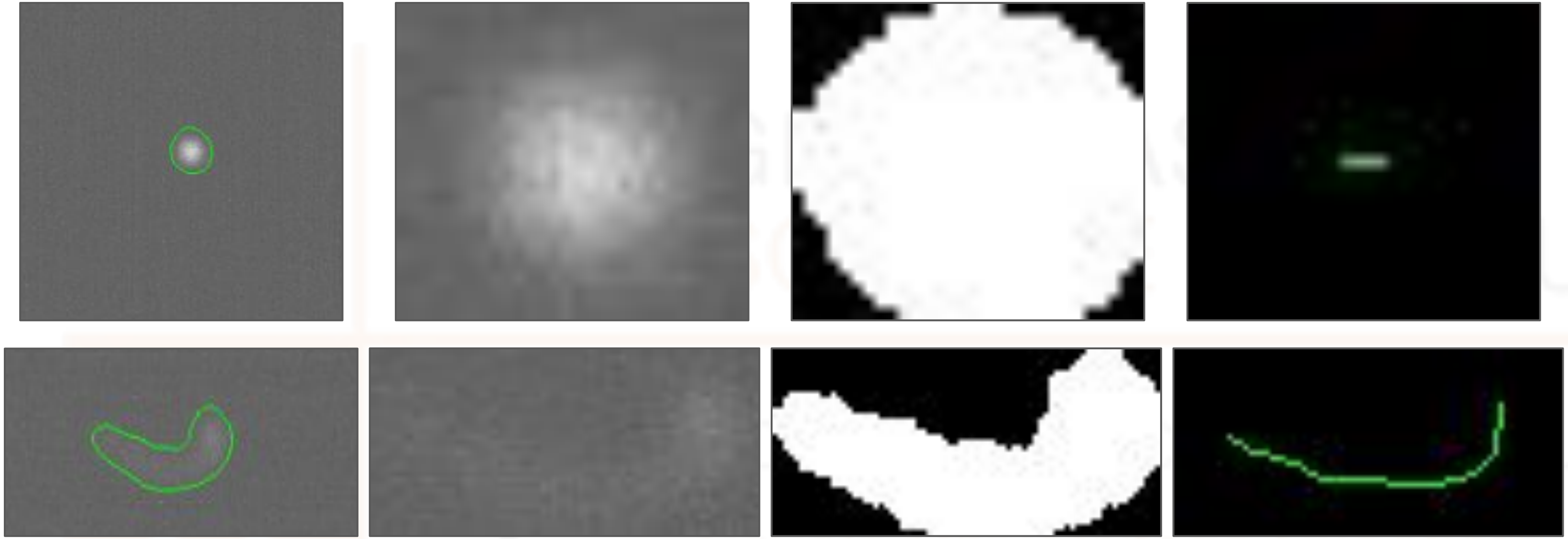
- eta:
  - MaxDen_2D divided by length (found by skeletonization)
- Light Density (delta):
  - Integral of the track divided by number of pixels in the track.
  - NR deposit higher energy over a short distance, therefore Light Density is higher for NR.
- Slimness:
  - Ratio of minor over major axis of the ellipse which bounds the track.
  - Electrons recoils suffer more scattering, so minor axis of the bounding ellipse is bigger when compared to NR which are generally straight.

- Skeleton length (thin_track):
  - Length in pixels found by thinning procedure.



[1] GEM-based TPC with CCD Imaging for Directional Dark Matter Detection, arXiv:1510.02170v3
[2] E Baracchini et. al., “Identification of low energy nuclear recoils in a gas TPC with optical readout”, arXiv:2007.12508v1

Length with Skeletonization



60 keV He recoil

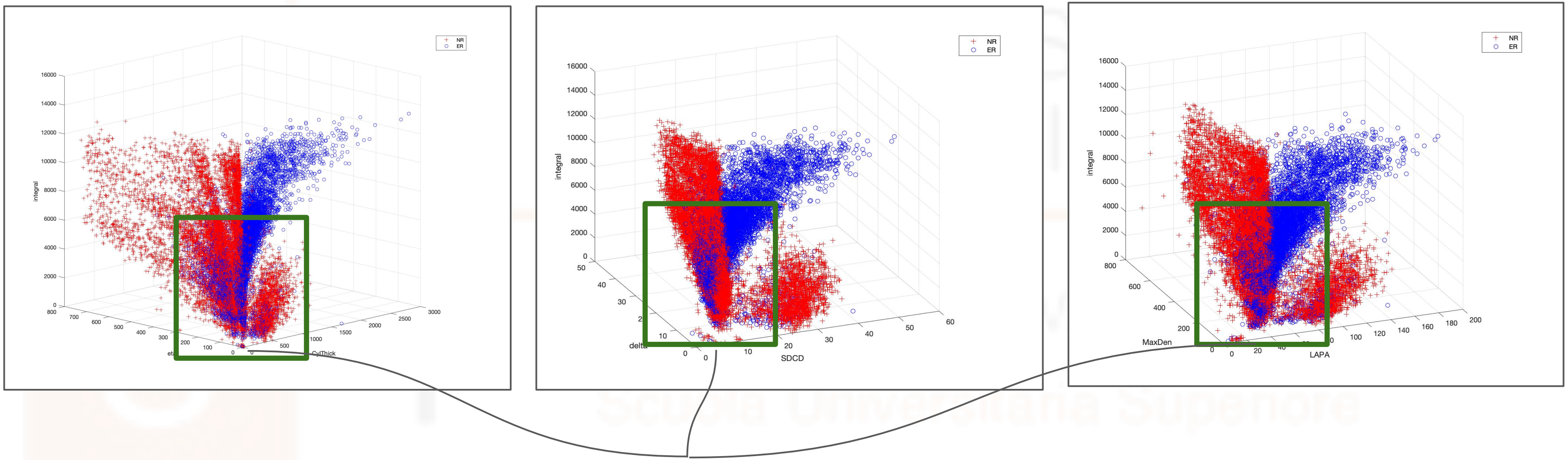
30 keV electron recoil

(a) Supercluster (b)Track (c)Thresholded track (d) Skeleton

Typically, directional detector use the diffused track length as length estimate, but that the REAL track length is more discriminating because is not affected by the diffusion.

Discrimination at low energies using traditional approach:

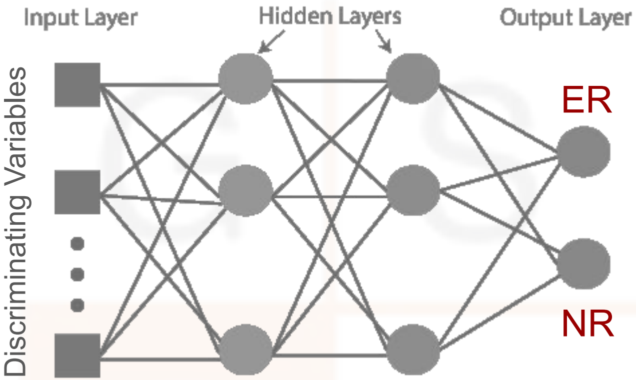
eta vs CylThick | delta vs SDCD | MaxDen vs LAPA

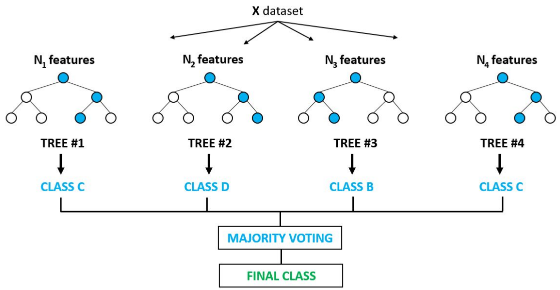


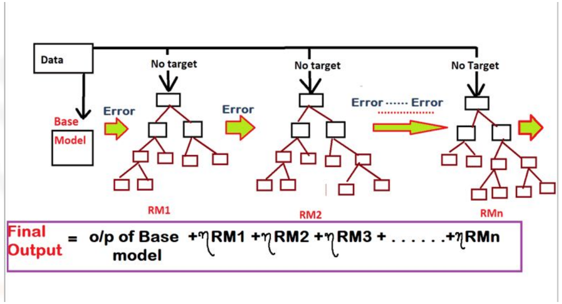
At low energies, it is very difficult to discriminate using traditional approach (applying a cut on the variables).

Deep Learning Models

Deep Learning Models







1) Deep Neural Network

- ❖ Weights of the network is optimised iteratively
- ❖ Result is the output of the last layer.

2) Random Forest Classifier

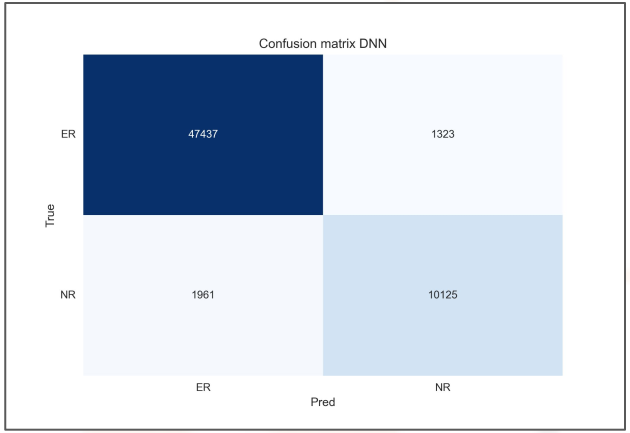
- ❖ It can build each tree independently.
- ❖ Results are combined at the end of the process.

3) Gradient Boosted Classifier

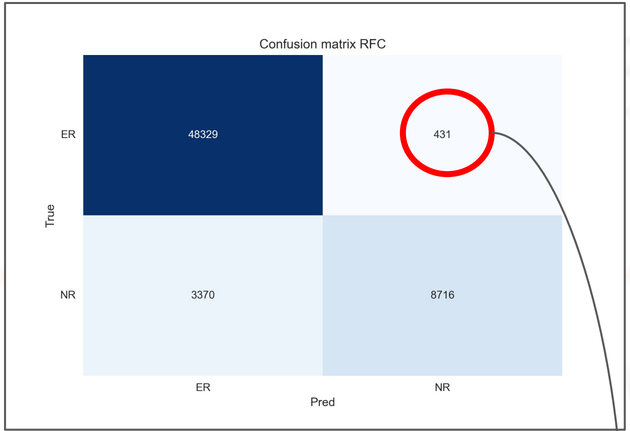
- ❖ It builds one tree at a time.
- ❖ It combines results along the way.

Results

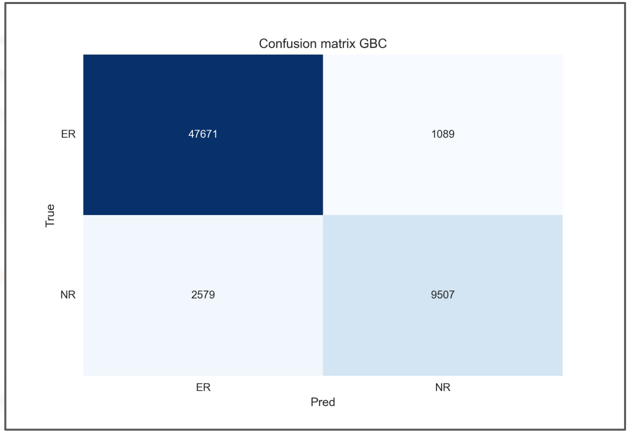
Results: Confusion Matrix



DNN



RFC

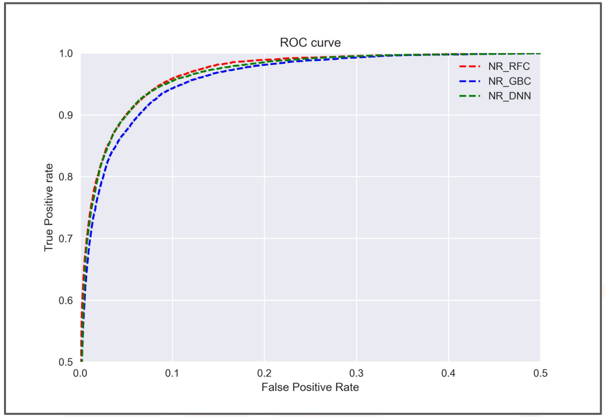


GBC

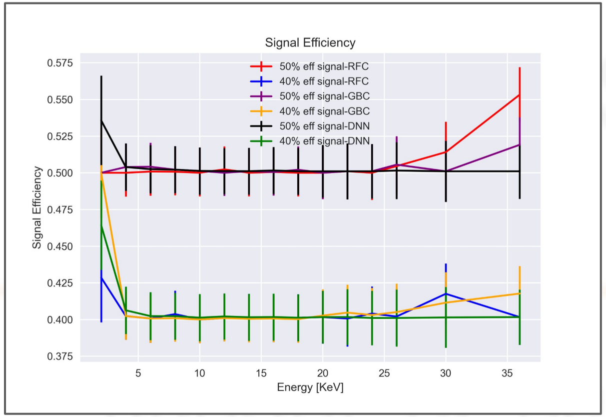
Less number of ER are classified as nuclear recoil in Random Forest Classifier compared to other two models.

I already have a larger sample and I’m working on it.

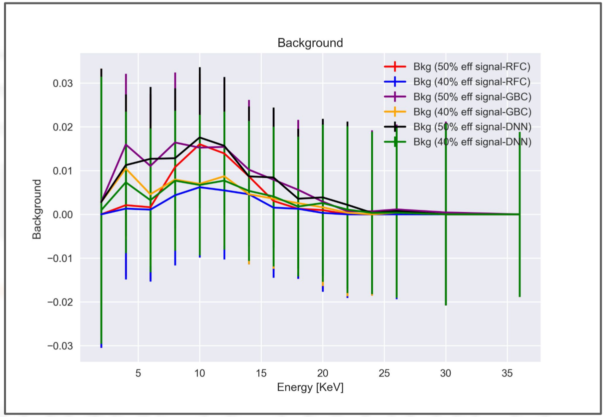
Results: ROC, Signal Efficiency and Background



ROC curve



Signal Efficiency

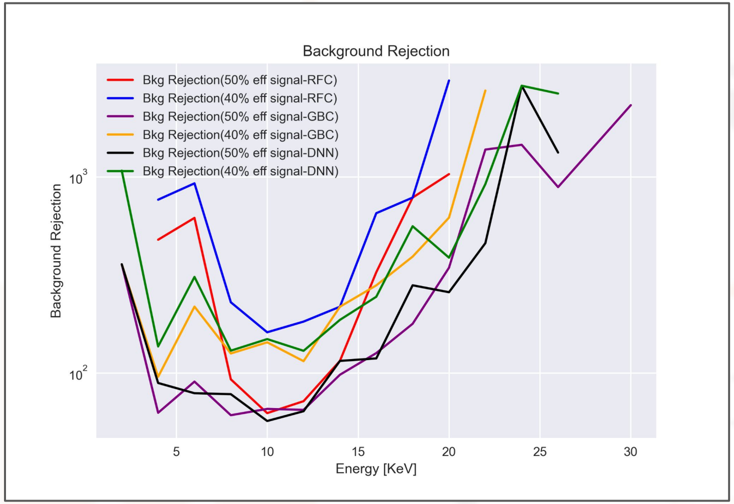


Background

Performance of all 3 network is very similar (Especially RFC and DNN).

A larger sample of data is needed for training.

Results: Background Rejection

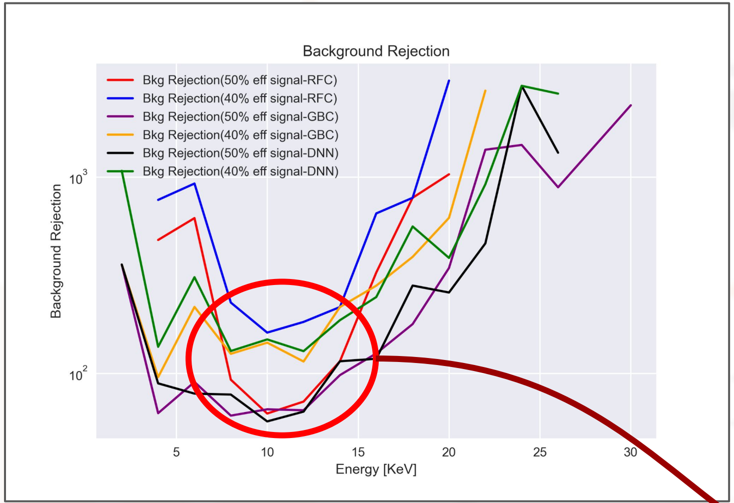


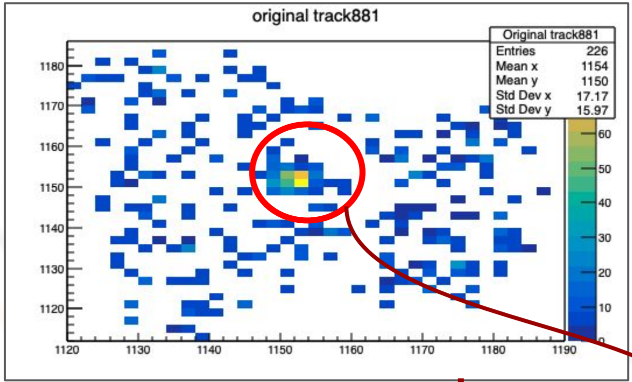
Models	Signal Eff.	Bkg. Eff.
RFC	0.40	0.0019
	0.50	0.0045
GBC	0.40	0.0041
	0.50	0.0082
DNN	0.40	0.0045
	0.50	0.0085
Selection on Delta	0.40	0.008
	0.50	0.035

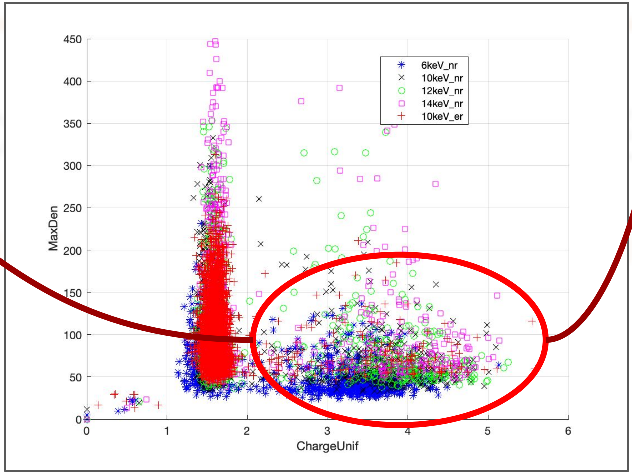
Rejection = Total No. of ER/ No. of ER left.

All the models perform better than the traditional technique used for classification (applying a cut on delta variable). Among these three models, RFC has better background rejection capability.

Performance at low energies:





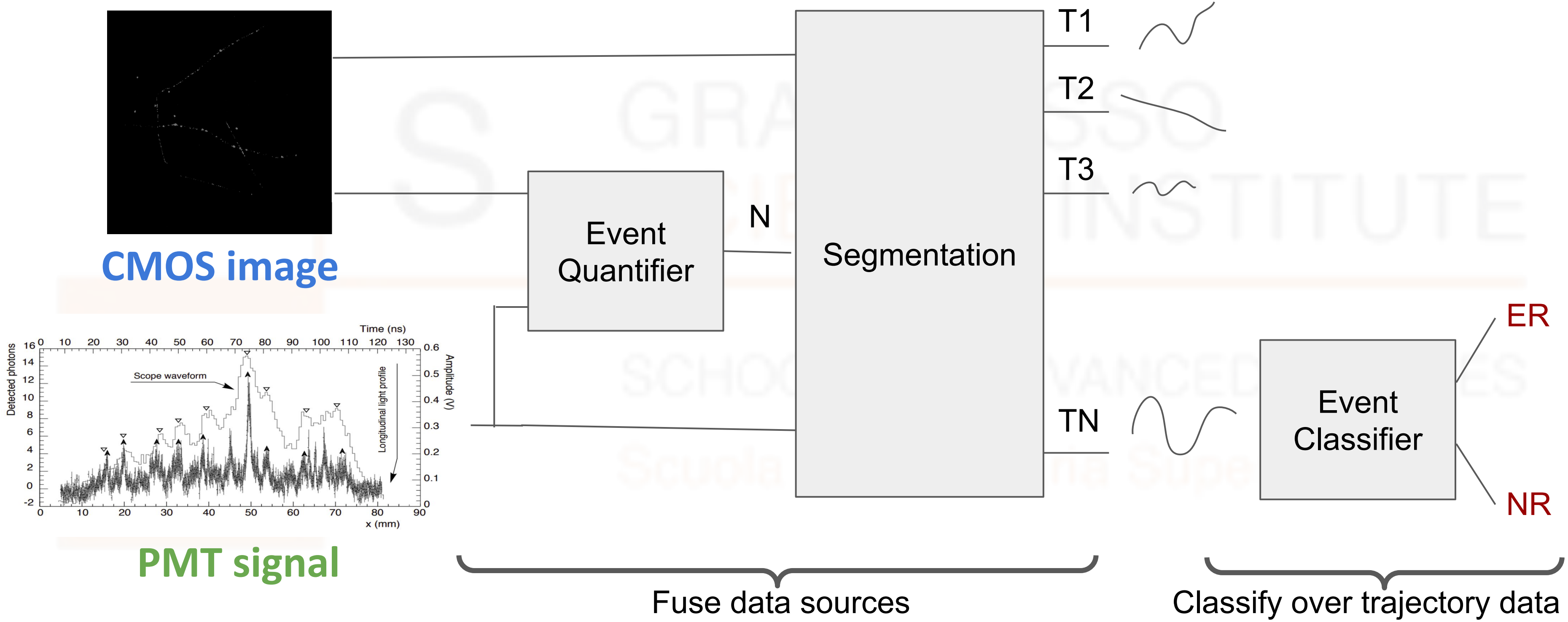


Actual track.

Performance of all the models deteriorates at low energy, because there is a lot of noise in reconstructed supercluster around the track.

Future work

Further work



This work is being done in collaboration with the CS department at GSSI.