New approach for centrality determination with FHCal in BM@N experiment

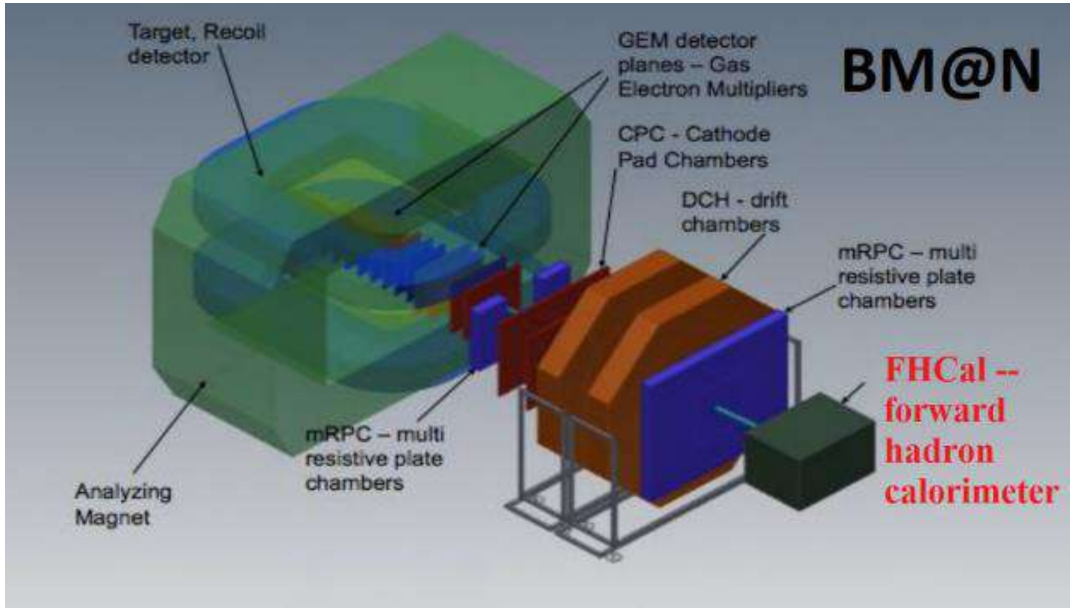
Nikolay Karpushkin
INR RAS

This work is supported by RFBR grant 18-02-40081

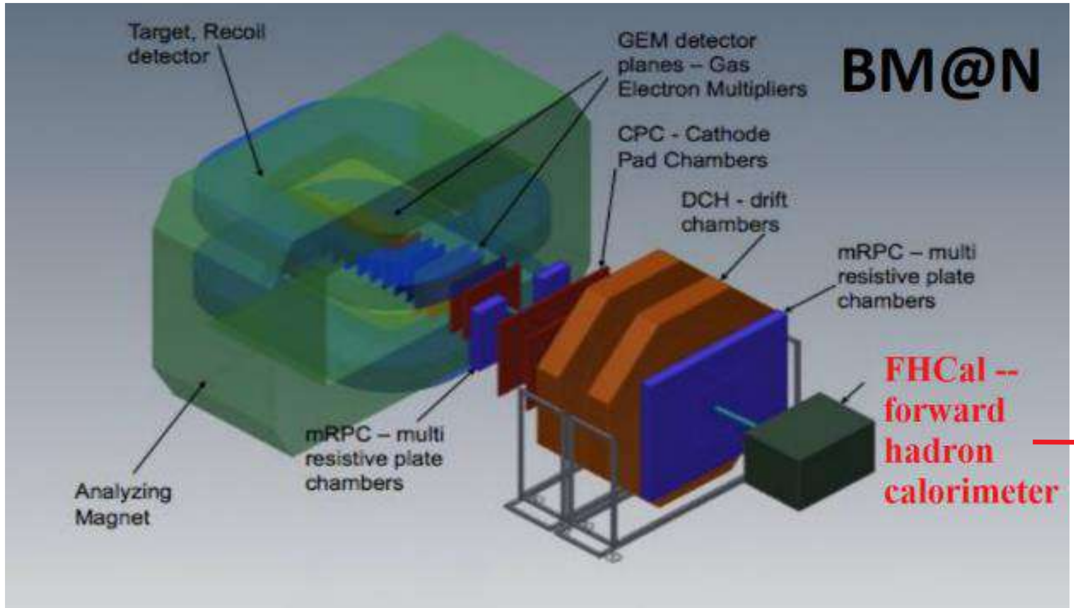
RFBR grants for NICA, 21 October 2020

Outline

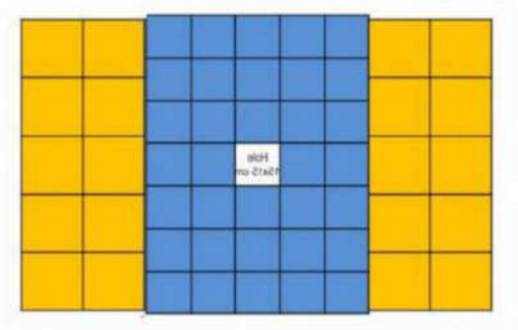
- FHCal of BM@N centrality problem statement
- Proposed solution
- Supervised&Unsupervised ML approaches
- Application to the simulation files



Determination of centrality using hadron calorimeters by ML methods

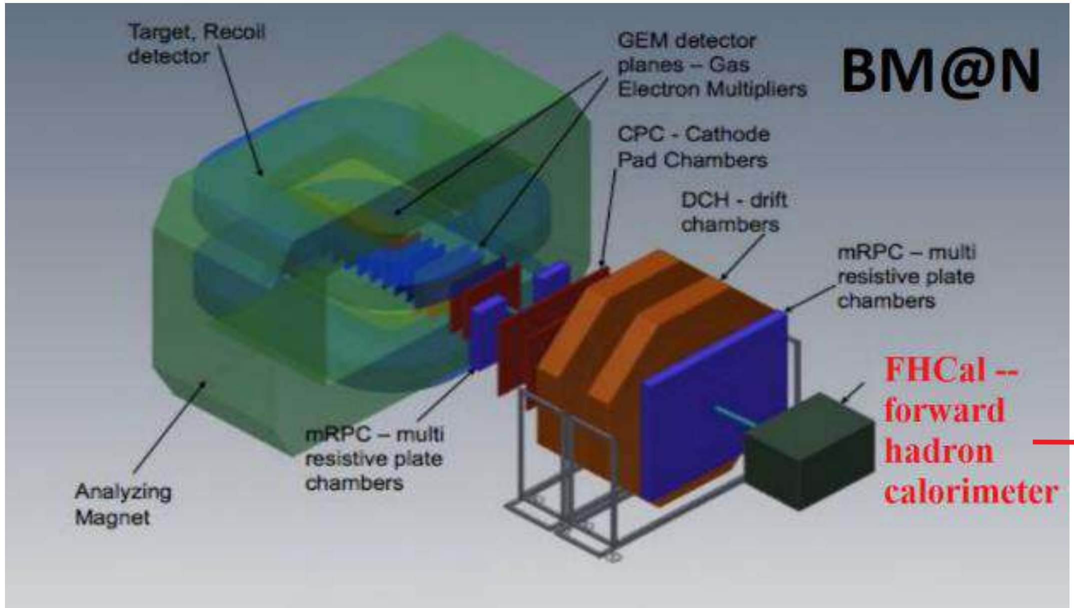


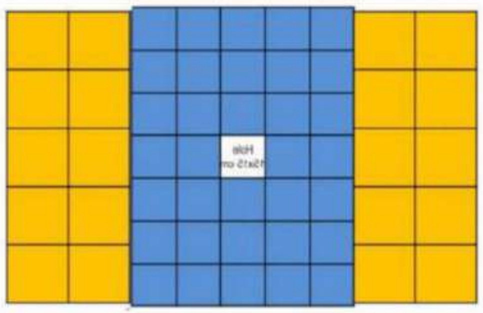
Determination of centrality using hadron calorimeters by ML methods



34 inner modules with sizes 15*15
+ 20 outer modules with sizes 20*20
Beam hole 15*15
Total weight – 17t

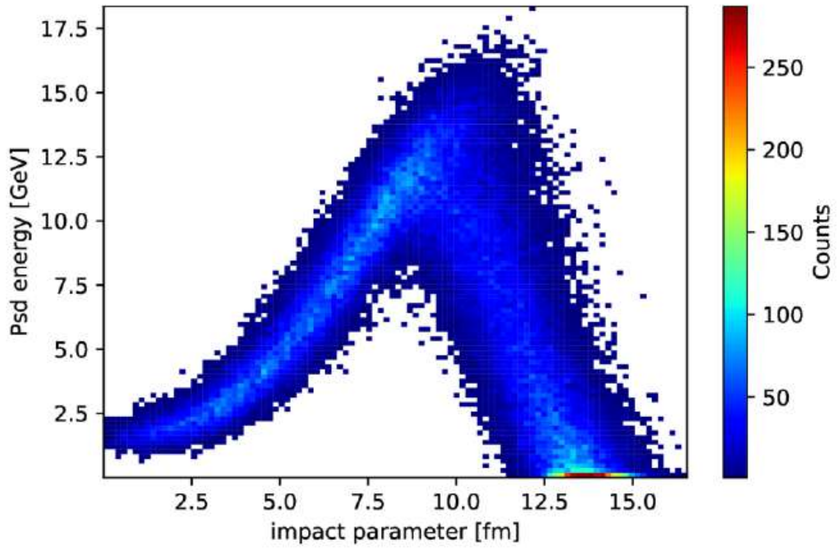
Determination of centrality using hadron calorimeters by ML methods



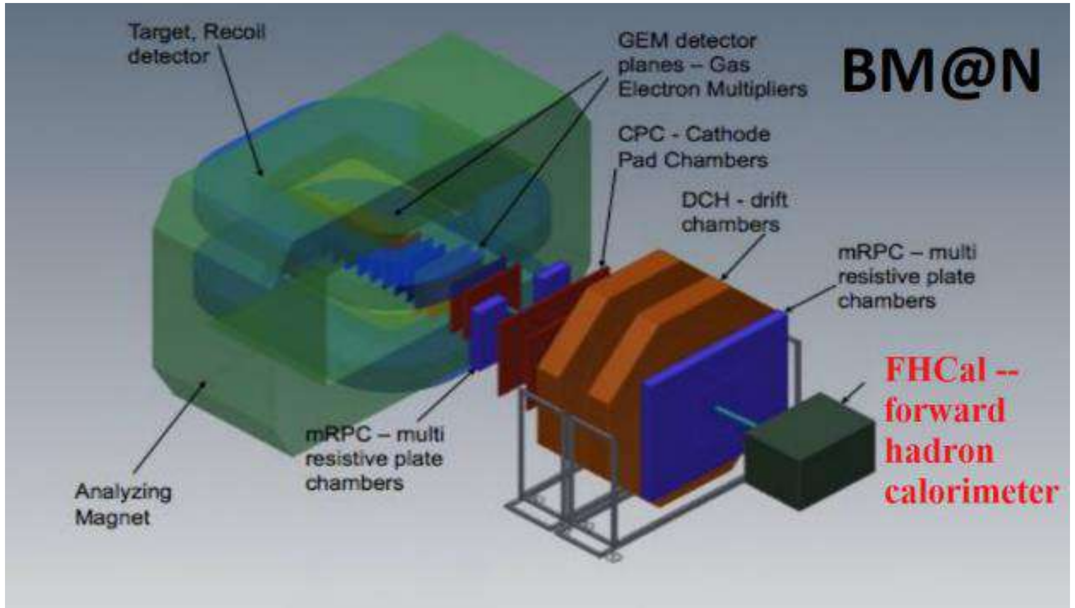


34 inner modules with sizes 15*15
+ 20 outer modules with sizes 20*20
Beam hole 15*15
Total weight – 17t

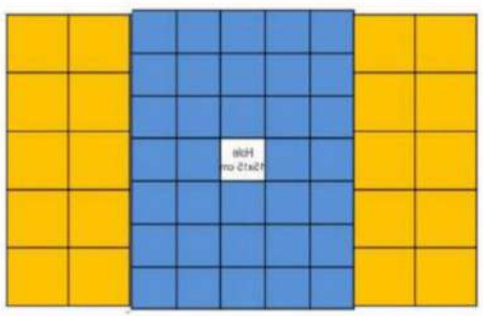
ambiguity

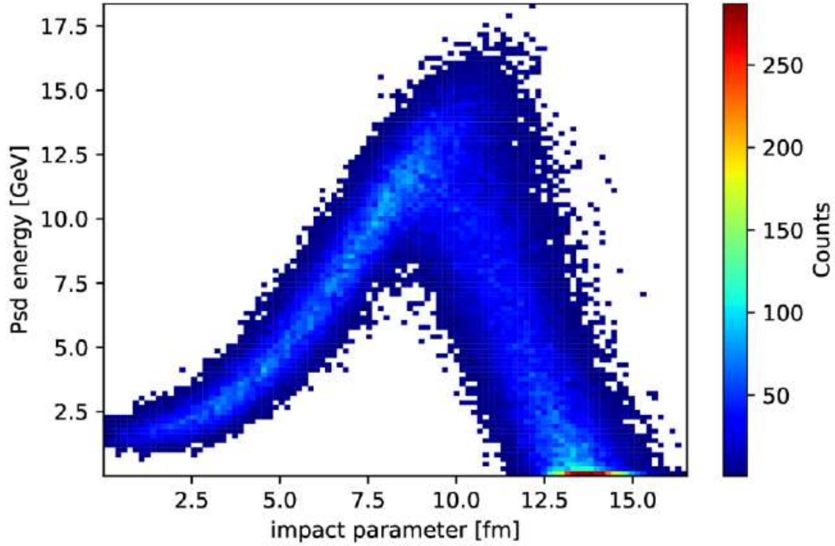


BM@N FHCal hole15cm DCMQGSM
AuAu 4.5AGeV 85k events



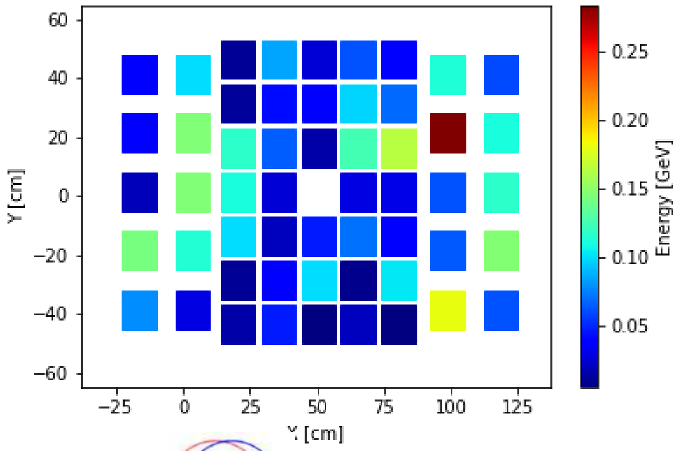
Determination of centrality using hadron calorimeters by ML methods

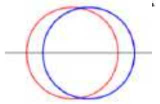




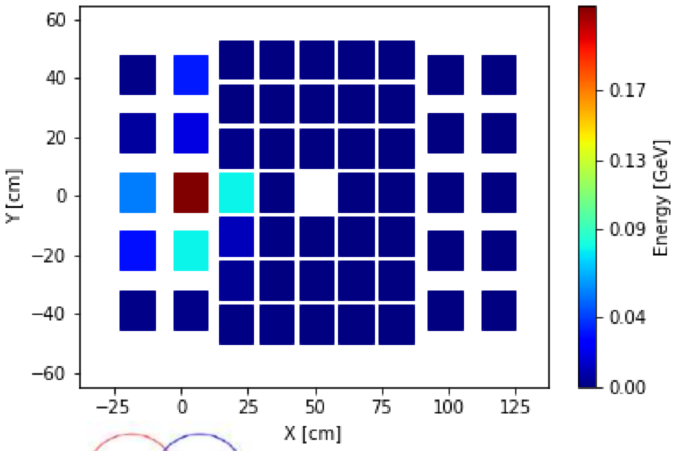
BM@N FHCal hole15cm DCMQGSM
AuAu 4.5AGeV 85k events

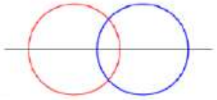
Calorimeter energy surface (single event)



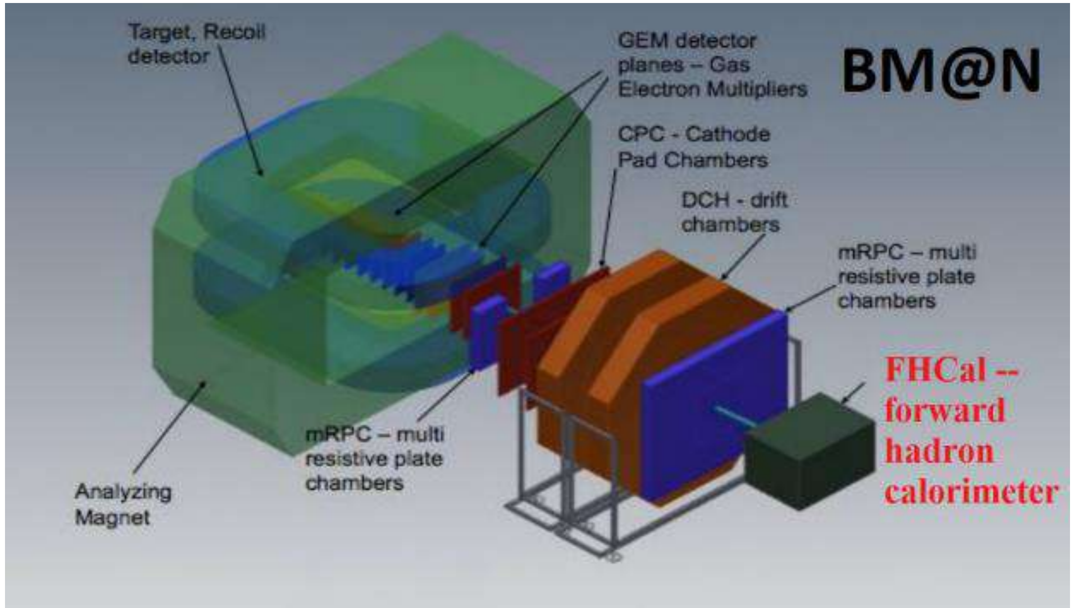


Central

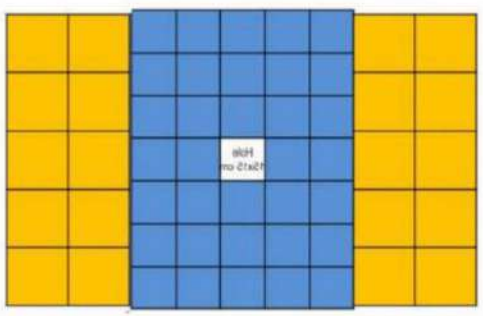




Peripheral



Determination of centrality using hadron calorimeters by ML methods



54 “pixels” to train ML algorithm

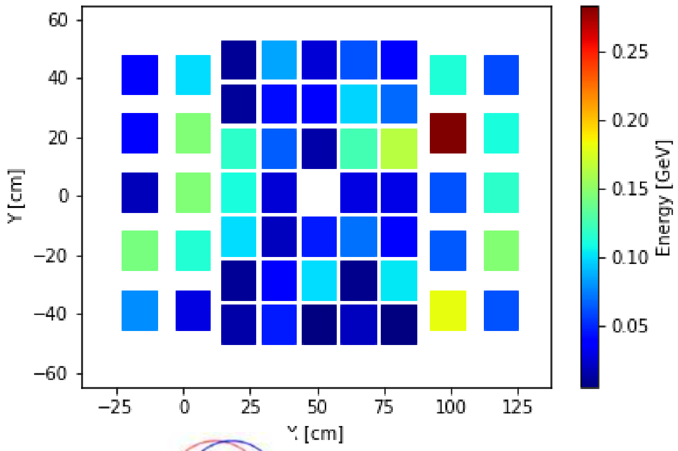
Use of simulation files:

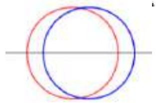
Input parameters – modules positions and energy depositions

Target variable – impact parameter

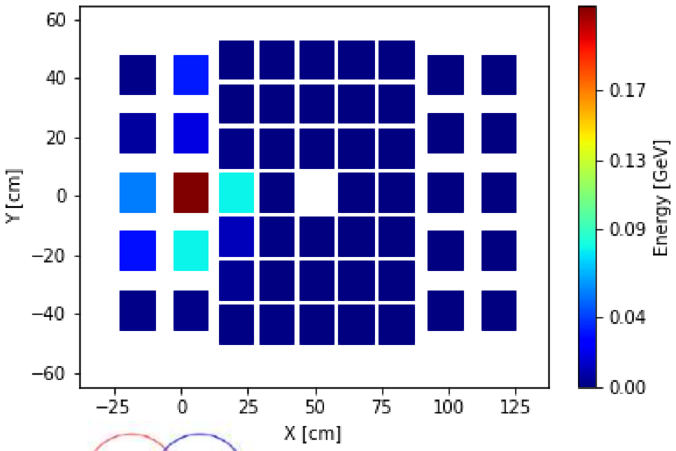
Expected result: online trigger for centrality estimation

Calorimeter energy surface (single event)





Central



Peripheral

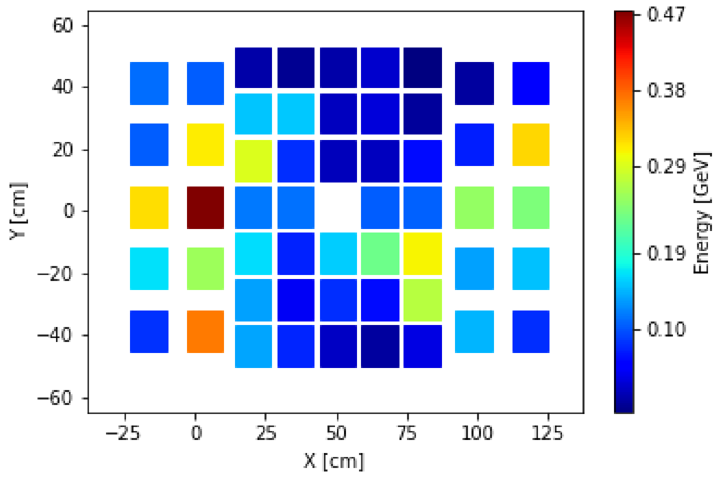
Supervised approach

Main goal:
Confirm approach capabilities.
Not to be used on real data.

1. Train-test split
2. Train the model:

   Inputs:

   - 1D arrays of energy depositions in calorimeter modules (Energy surface)
   - Centrality class index (impact parameter label)

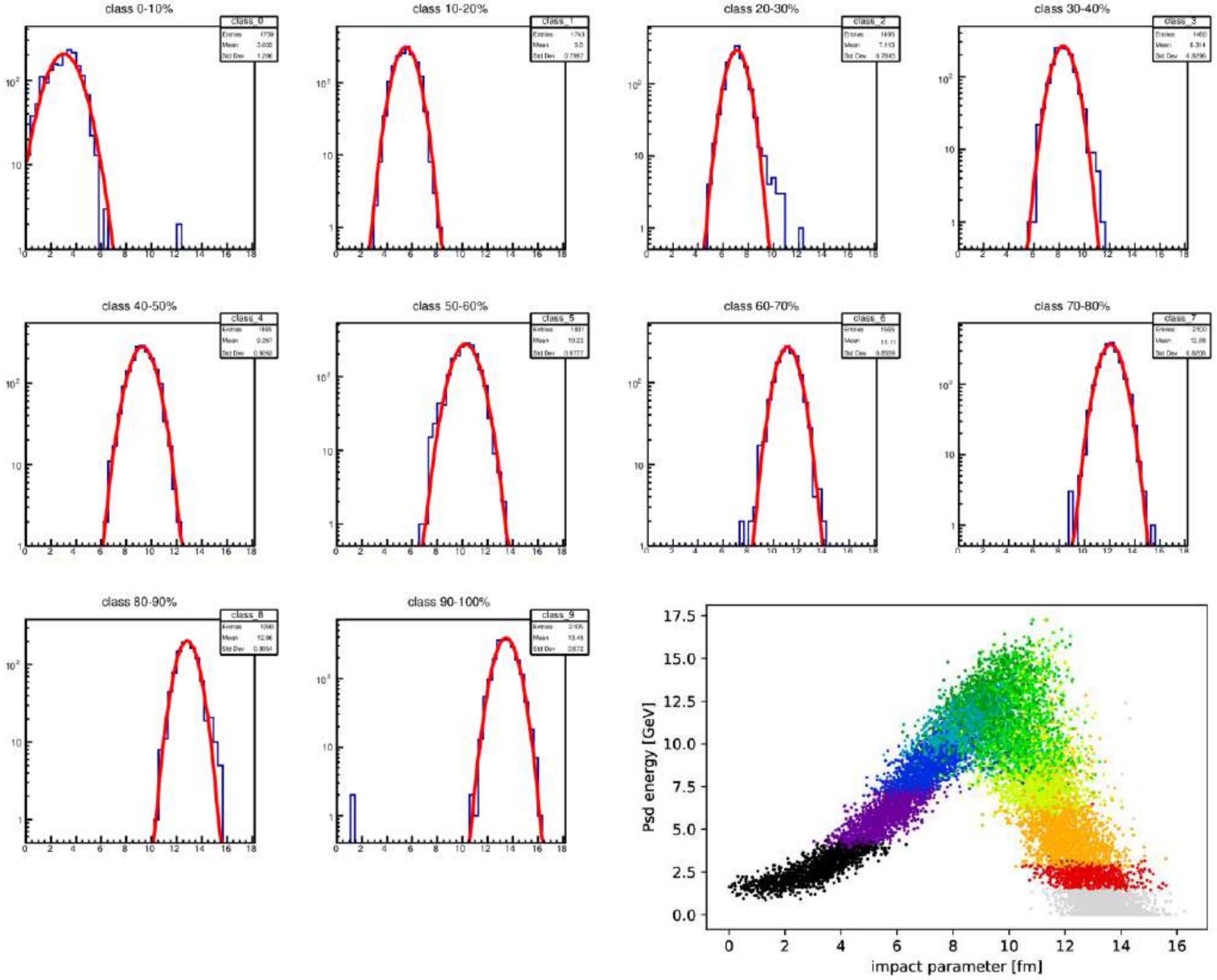


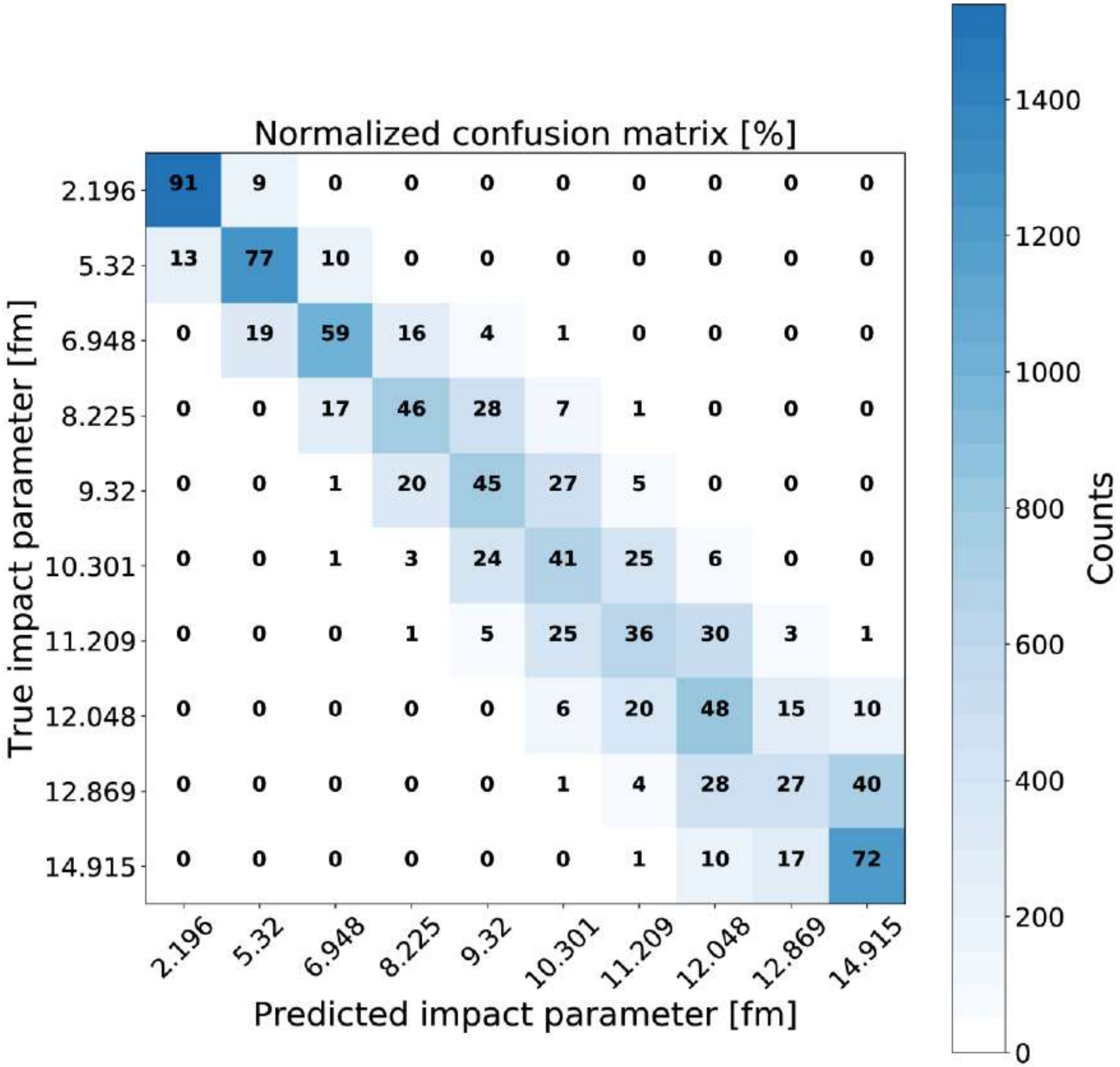
   Model architecture:

   dense_input: InputLayer → dense: Dense → dense_1: Dense → dense_2: Dense → dense_3: Dense → dense_4: Dense

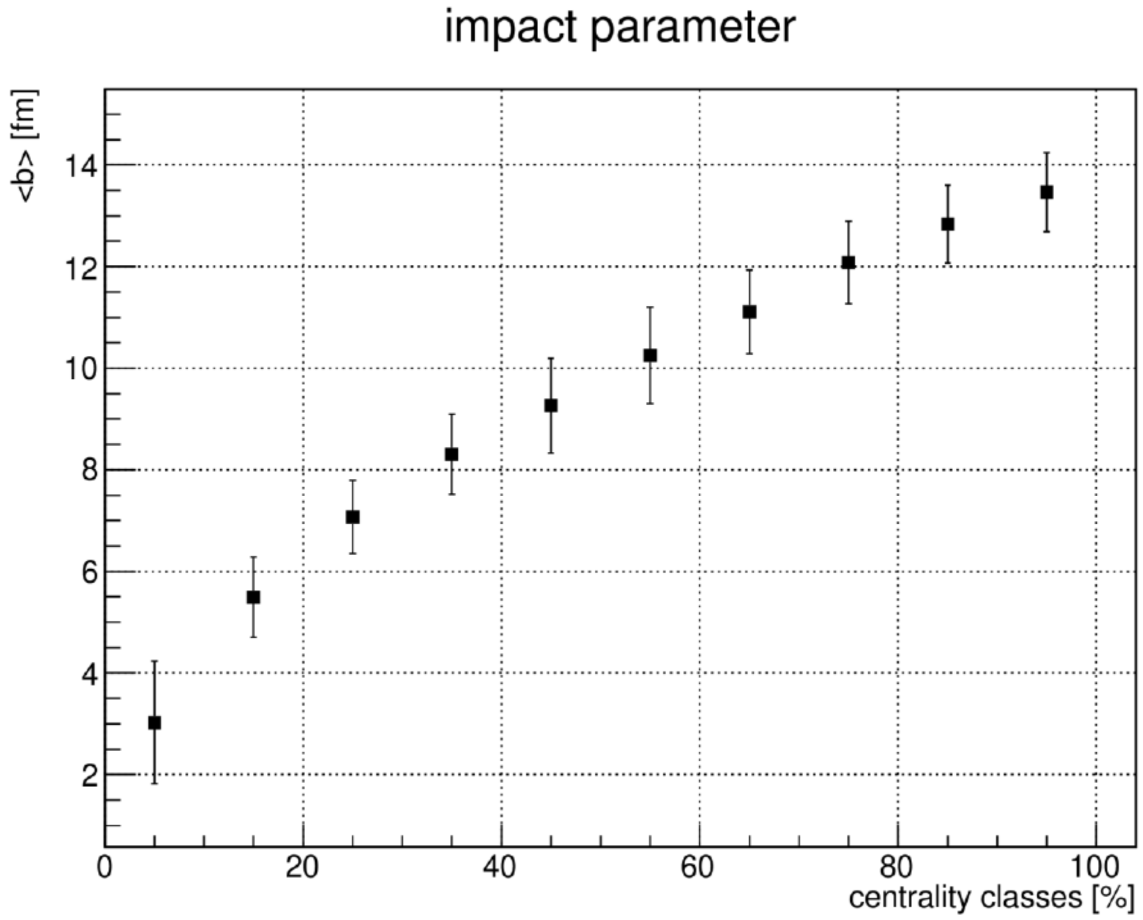
3. Test model accuracy

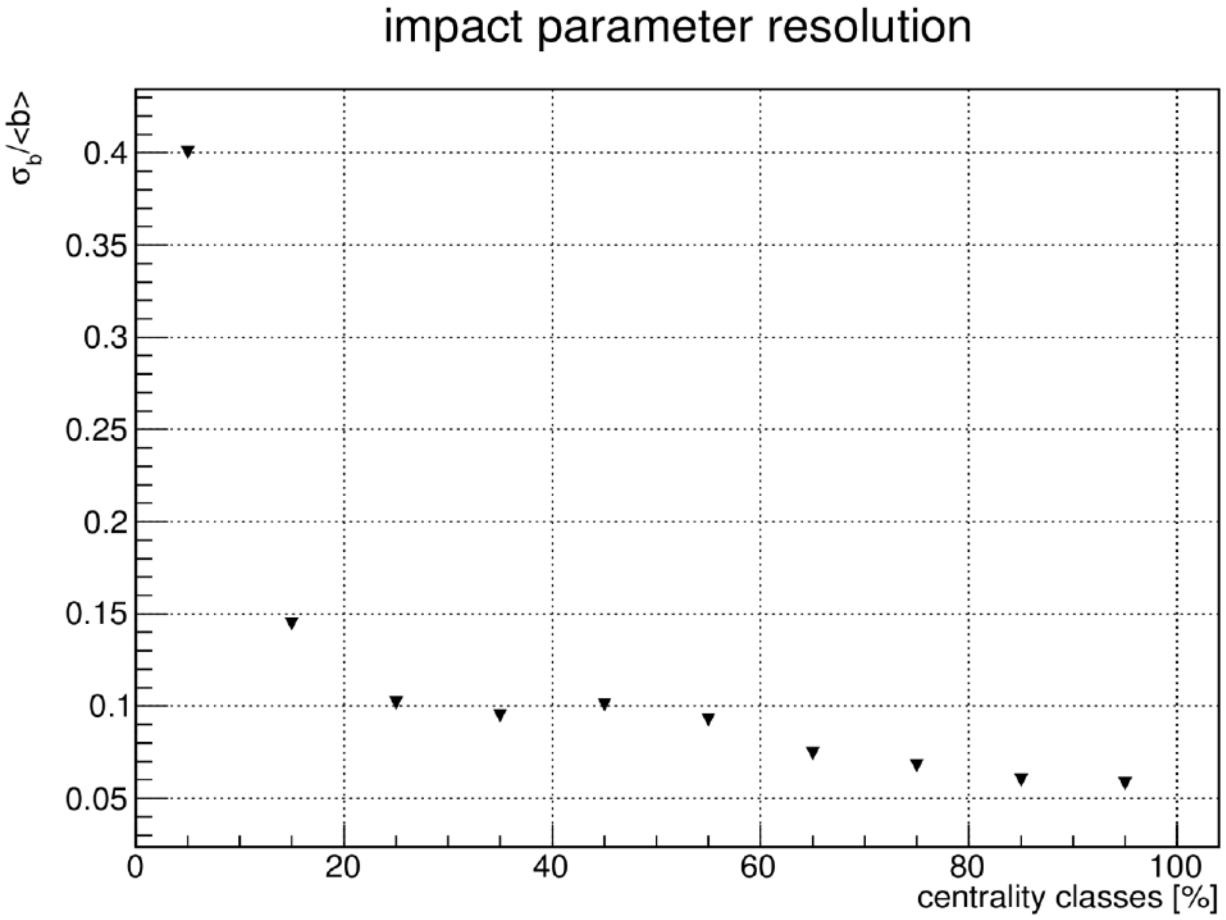
AuAu 4.5AGeV DCMQGSM Supervised





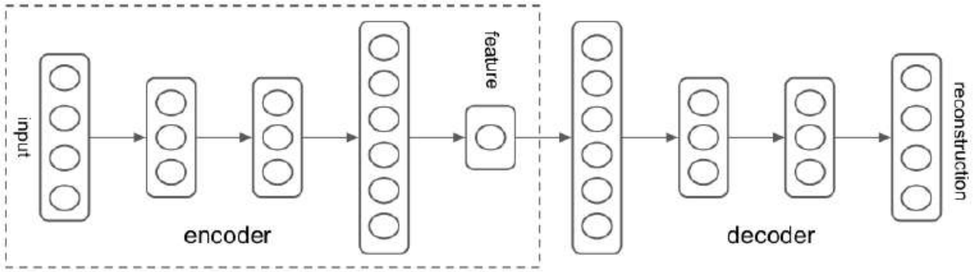
AuAu 4.5AGeV DCMQGSM Supervised





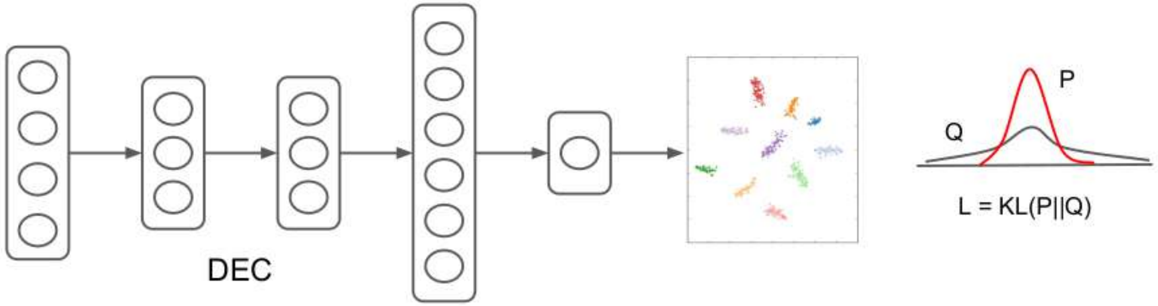
Unsupervised approach – Deep Embedded Clustering

1. Train autoencoder



2. Estimate cluster centroids: Encode data + TSNE + KMeans

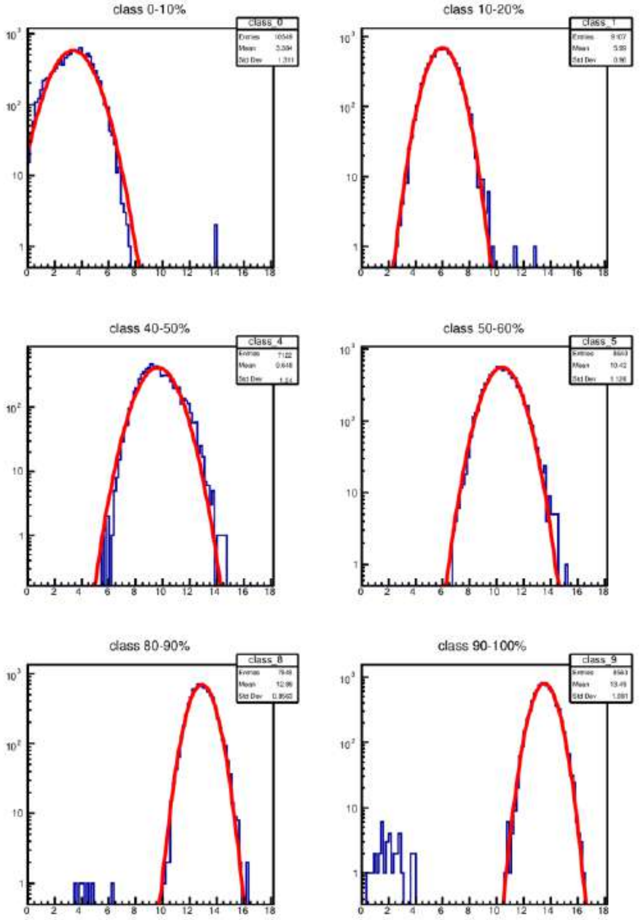
3. Deep Embedded Clustering (link):

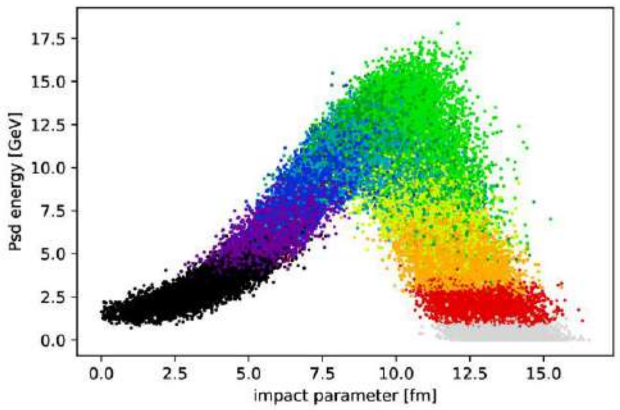


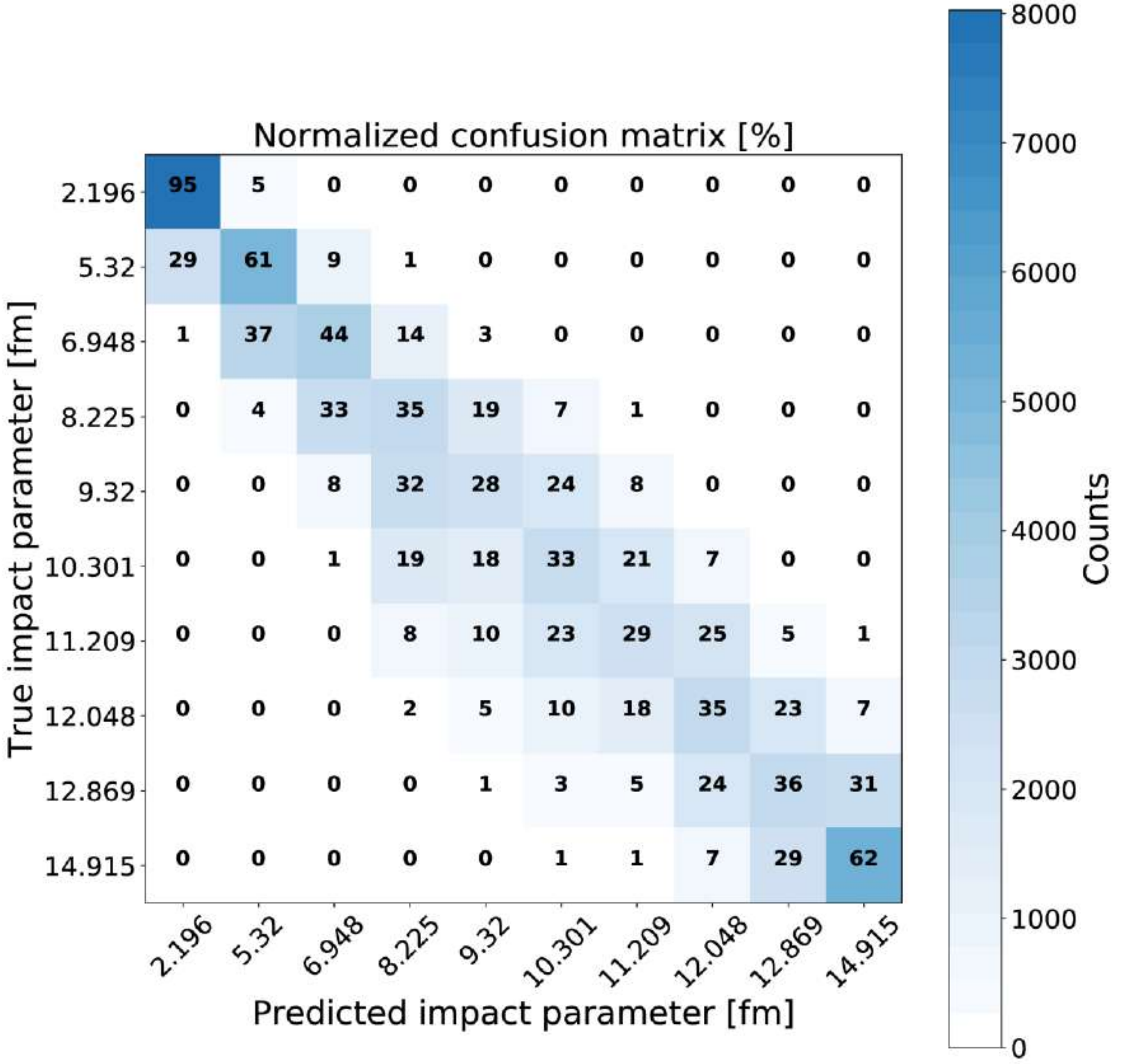
a) Soft clustering of encodded data by Student's t-distribution

b) Iteratively strengthen predictions by approximating the obtained distribution **Q** to the auxiliary target distribution **P**

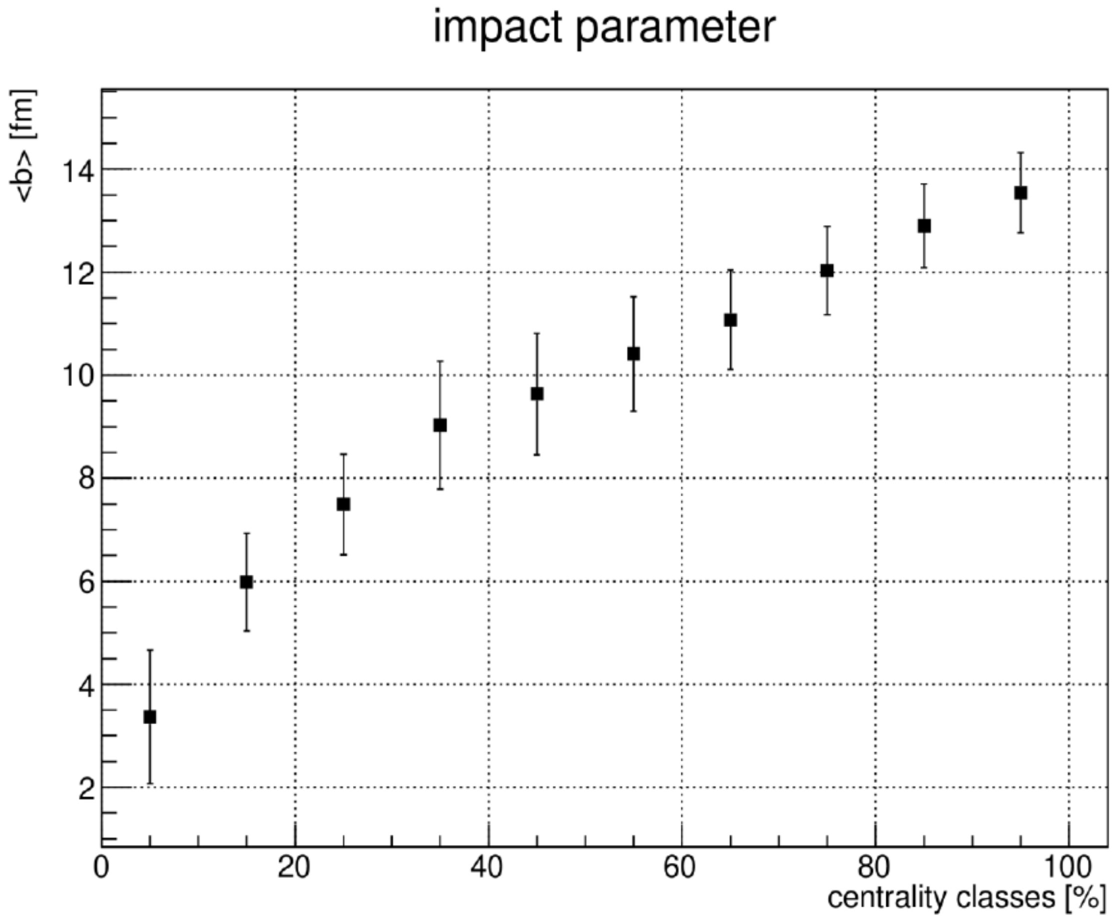
AuAu 4.5AGeV DCMQGSM Unsupervised

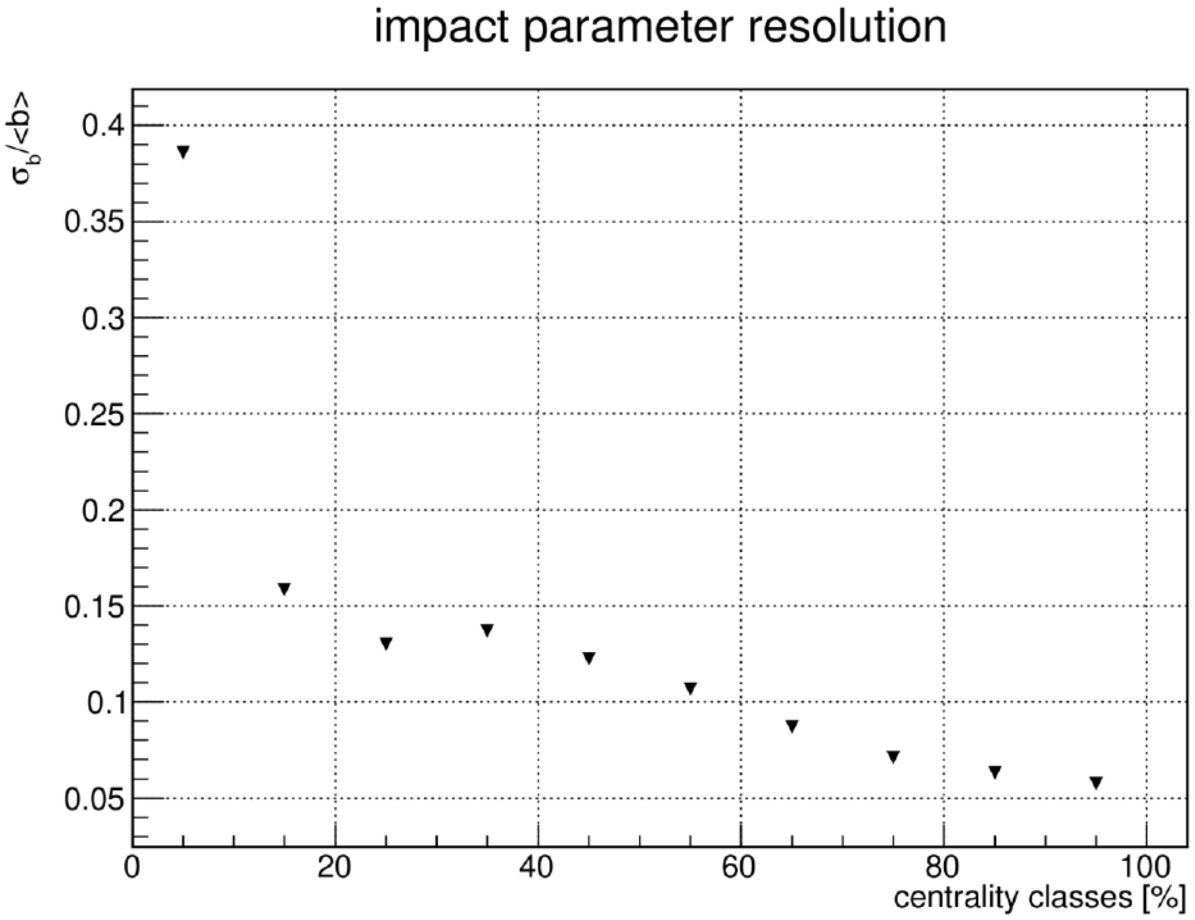




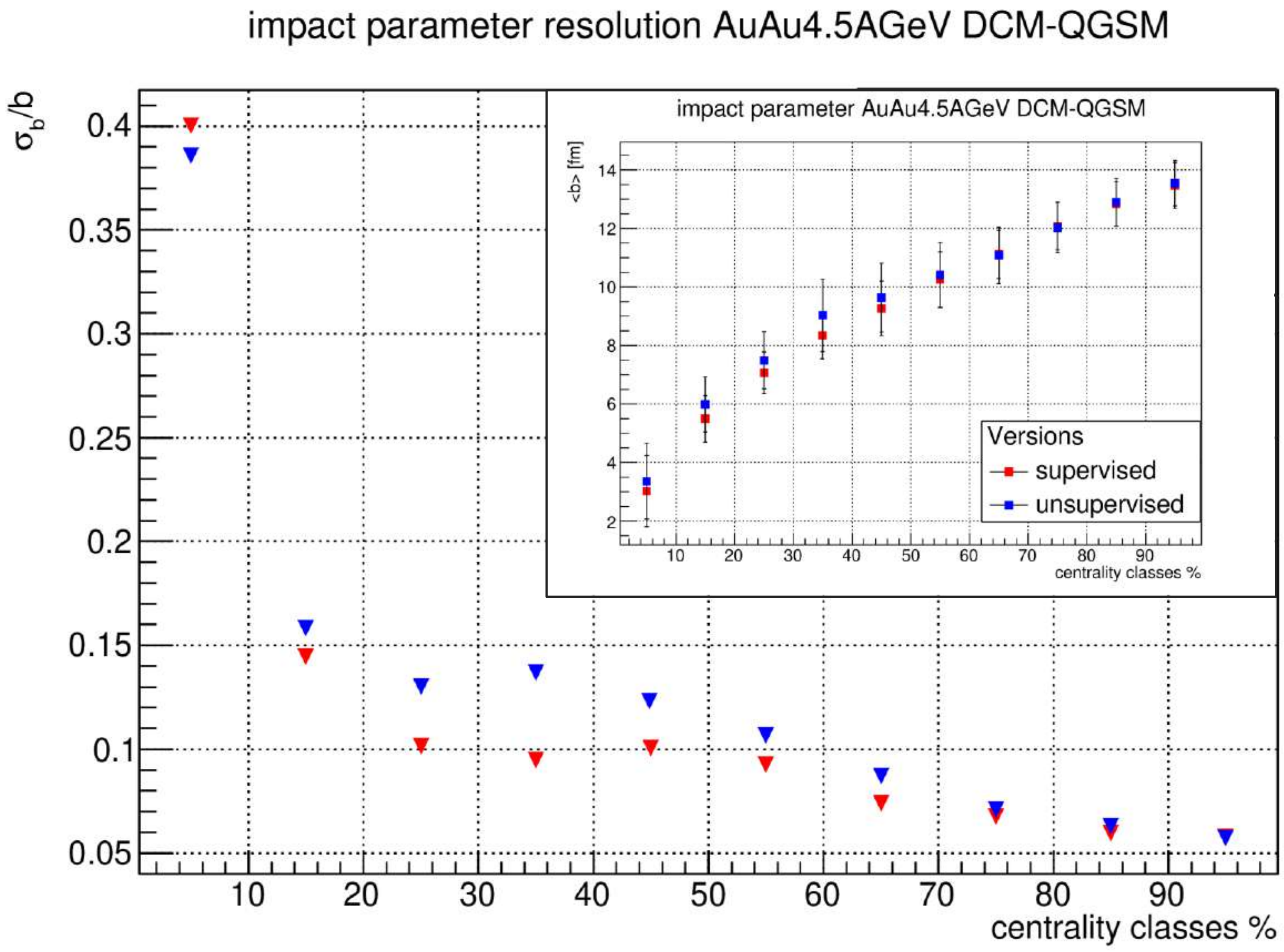


AuAu 4.5AGeV DCMQGSM Unsupervised

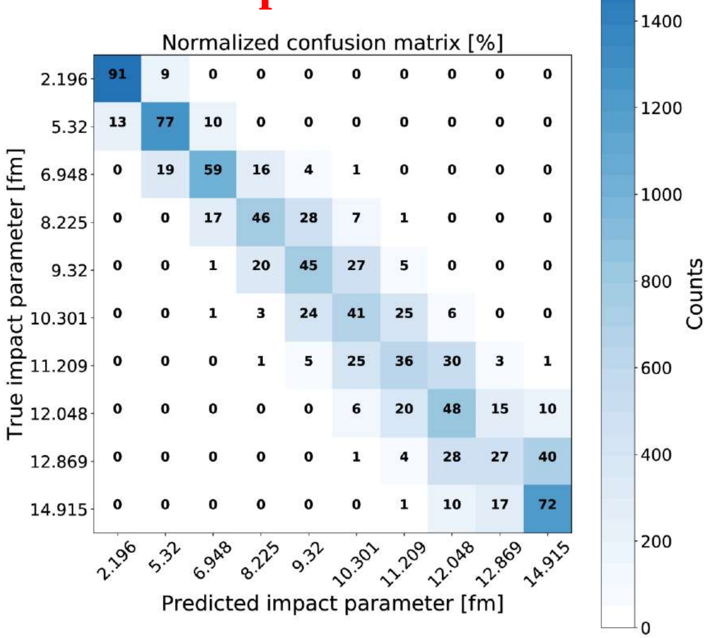




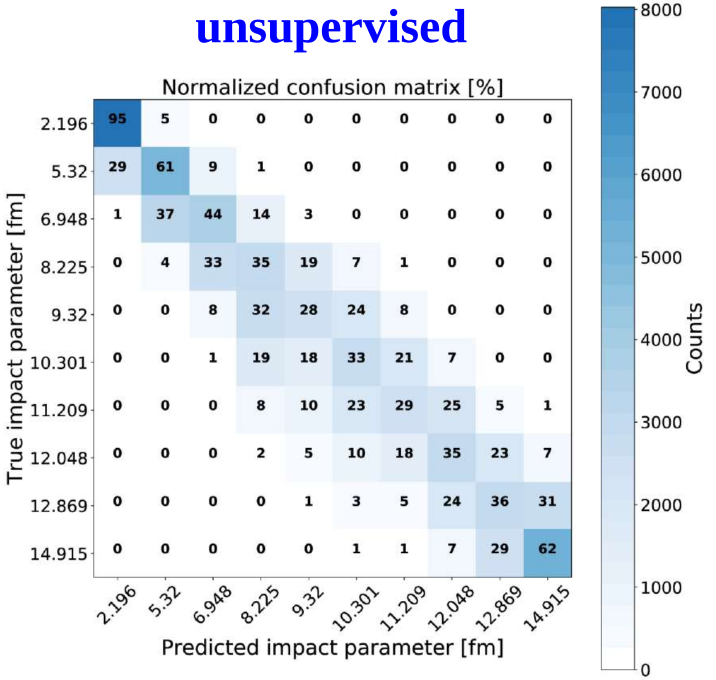
Resolution: supervised, unsupervised



supervised



unsupervised



Conclusions

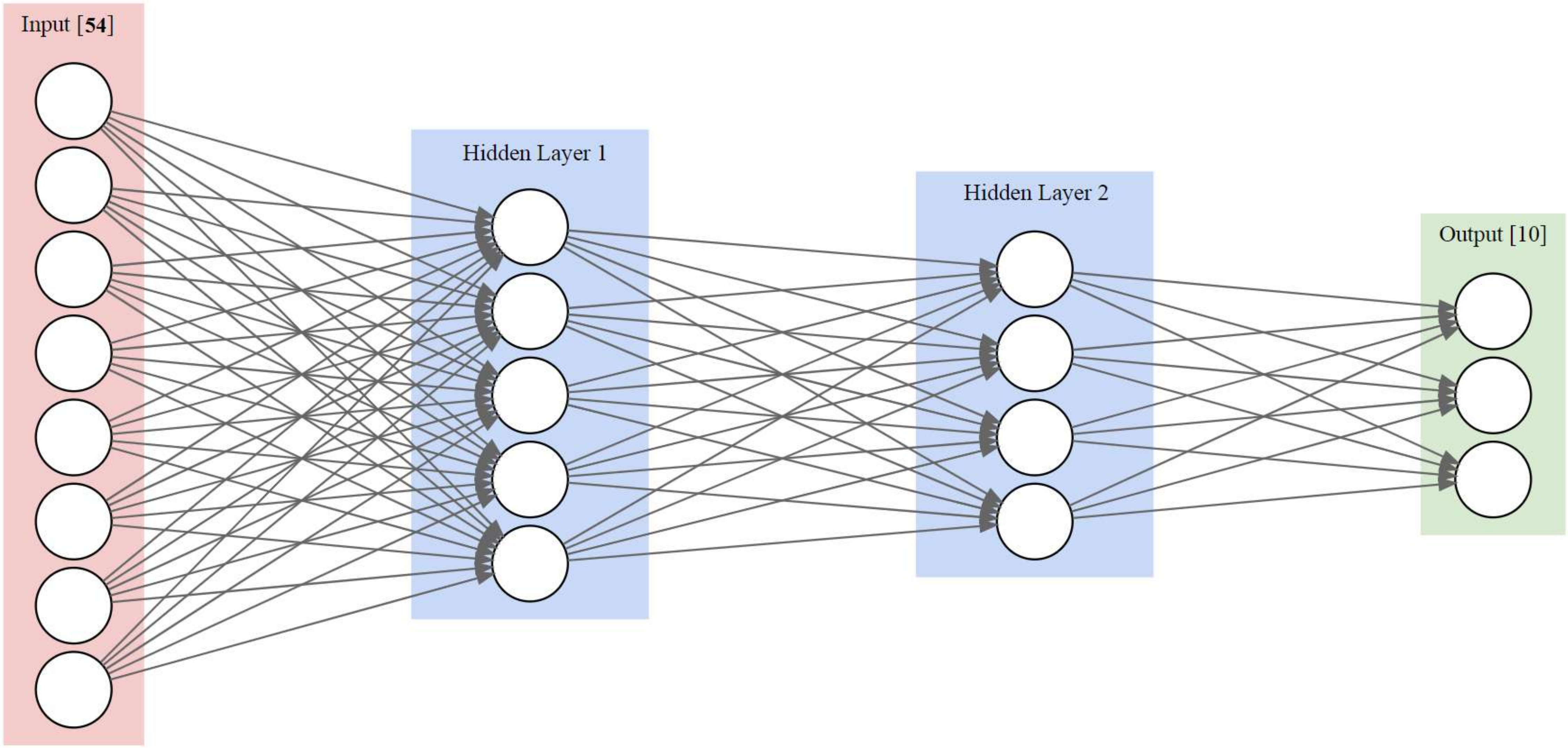
- Supervised&Unsupervised ML approaches are developped for centrality classes determination with forward hadron calorimeters with beam holes.
- The results of applying the approaches to BM@N simulation data with different collision energies were shown.
- The centrality resolution and impact parameters are shown for all centrality classes in each case.

Outlook

- Further improvement of methods will be carried out. Git repository: link
- The approaches will be tested to determine the centrality classes in the BM@N, NA61/SHINE@SPS and CBM@FAIR experiments.

Thank you for your attention!

BACKUP


Input [54]
Hidden Layer 1
Hidden Layer 2
Output [10]

Machine Learning

Machine Learning

Supervised

Unsupervised

Machine Learning

Supervised

- Train-test split of the same data.
- Need target variable to train (data labeling).
- Model dependent: if ML-model is trained with one physical model, the spatial distributions of another model will hardly be reproduced. As well as real physical data.
- May serve as a reference for unsupervised ML.

Unsupervised

Machine Learning

Supervised

- Train-test split of the same data.
- Need target variable to train (data labeling).
- Model dependent: if ML-model is trained with one physical model, the spatial distributions of another model will hardly be reproduced. As well as real physical data.
- May serve as a reference for unsupervised ML.

Unsupervised

- Uses all available data and clusters them.
- No need of target variable.
- Model independent: one can take real physical data and cluster them. No need to use MC data first.
- How to check? Use secondary particles multiplicity distributions in centrality classes selected by ML-model.