# Particle identification in MPD at NICA using machine learning


Pavel Gordeev, National Research Centre "Kurchatov Institute"
National Research Nuclear University "MEPhI"



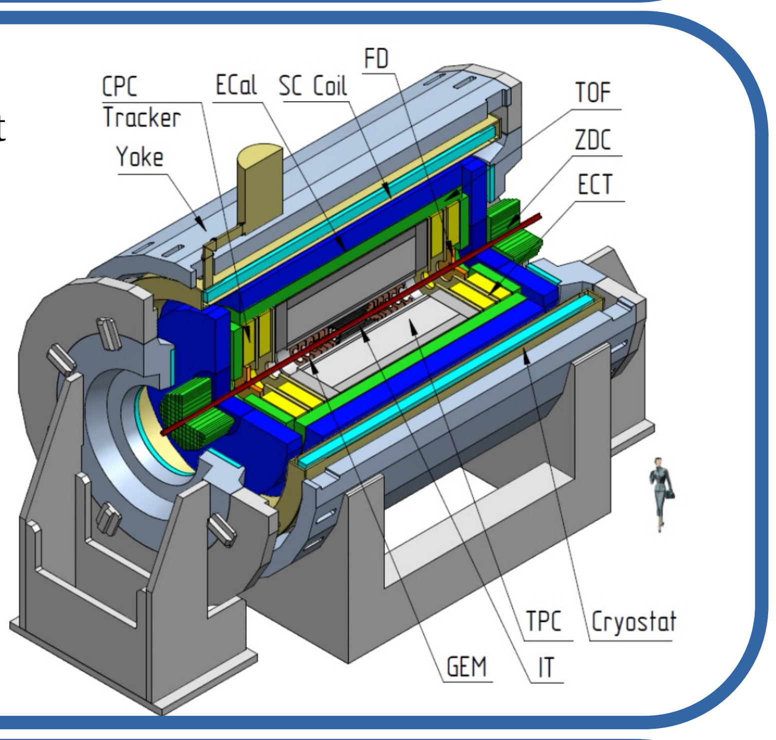
The 7th International Conference on Particle Physics and Astrophysics (ICPPA-2024)


## 1. Motivation

- Measurements of neutral mesons, direct photon spectra and correlations require precise photon identification. This analysis is devoted to improvement of the photons identification in MPD experiment with Machine Learning (ML) approach
- Photons can be reconstructed in MPD experiment in two ways: either in electromagnetic calorimeter ECAL or via reconstruction of e+e- pair created in the photon conversion in the material of inner detectors
- The e+e- pair provides the high-dimensional set of variables (individual track parameters, secondary vertex geometry, topological cuts etc.) with complex correlations between them. The ML techniques is often provides better performance than traditional cut-based algorithms
- The Boosted Decision Trees (BDTs) approach is widely used in HEP, providing accurate classification methods for distinguishing between signal and background. In this work, BDT was implemented for identification of the photon conversion pairs

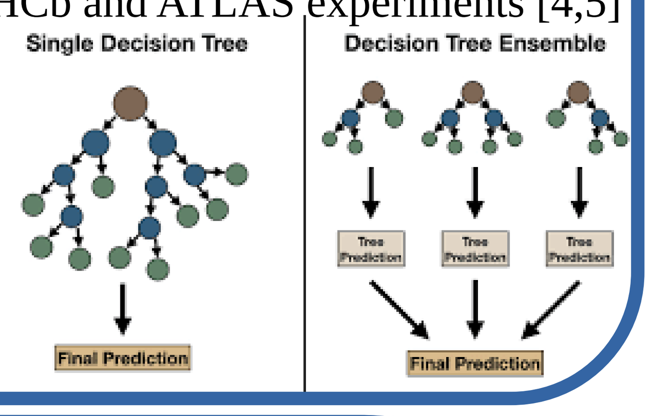
## 2. The MPD detector

- The Multi-Purpose Detector (MPD) experiment at NICA aims to study the properties of dense nuclear matter especially in the region of high baryon density [1]
- At the first stage the setup of the MPD experiment includes Time Projection Chamber, providing tracking of charged particles, Time Of Flight detector to improve their identification and electromagnetic calorimeter ECAL providing photon and electron reconstruction and identification in full coverage in azimuth and wide range in rapidity



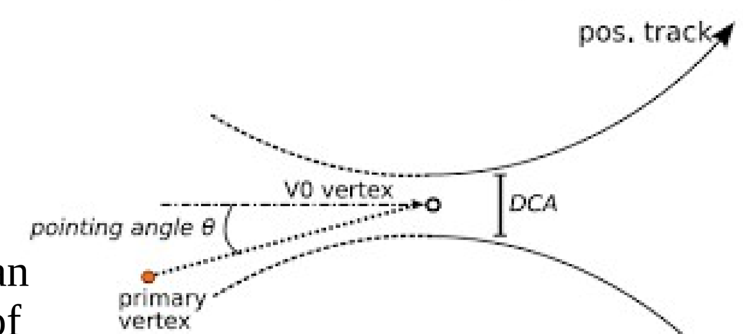

## 3. Boosted Decision Trees

- Boosted Decision Trees are an ensemble learning method that combines multiple weak decision trees into a strong classifier [2,3]
- The "boosting" process improves performance by sequentially training trees, where each tree corrects the mistakes of the previous one
- The BDT approach is widely used in ALICE, CMS, LHCb and ATLAS experiments [4,5]
- Main configuration parameters of BDT:
- Number of trees in the forest: 500
- Boosting type for the trees in the forest: AdaBoost
- Max depth of the decision tree allowed: 3

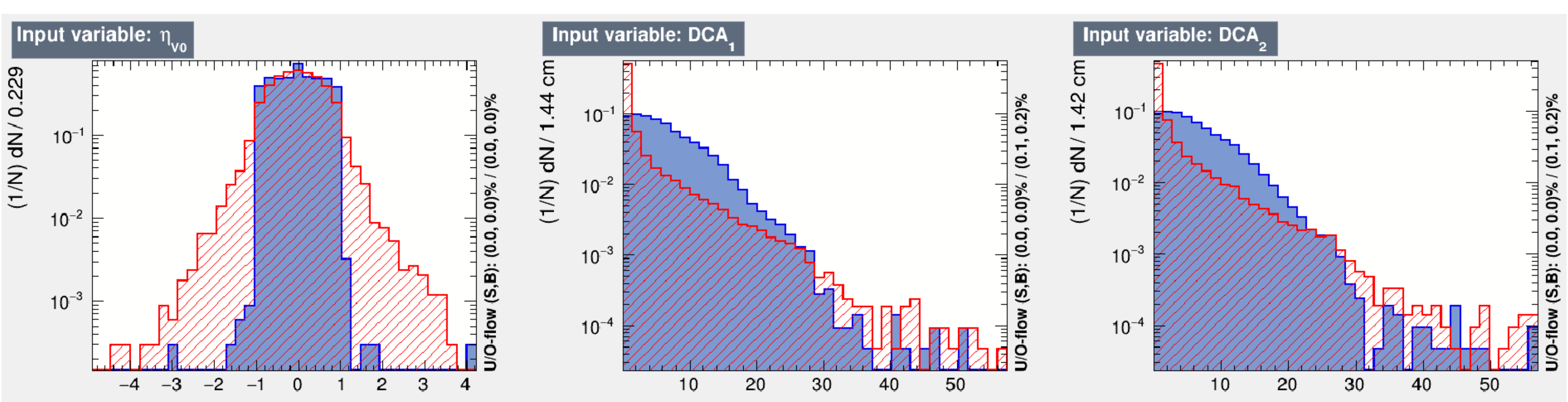



## 4. Variables selection of photon conversion

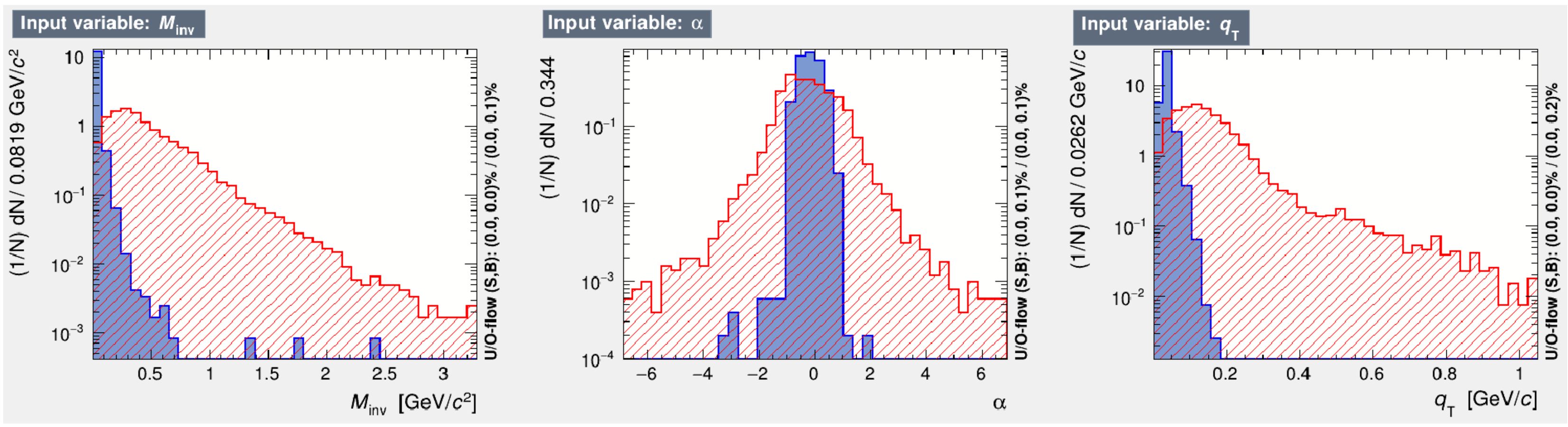
- Variables for training are combined in 4 groups: Track quality, V0 topology, Electron/Positron PID and Photon PID
- Number of hits in TPC serve as Track quality criteria
- Positively and negatively charged tracks can form V0 vertex (secondary vertex) and one can access topological variables such as: Cosine of Pointing Angle (CPA), Distance of Closest Approach (DCA) for tracks and V0 itself, and others

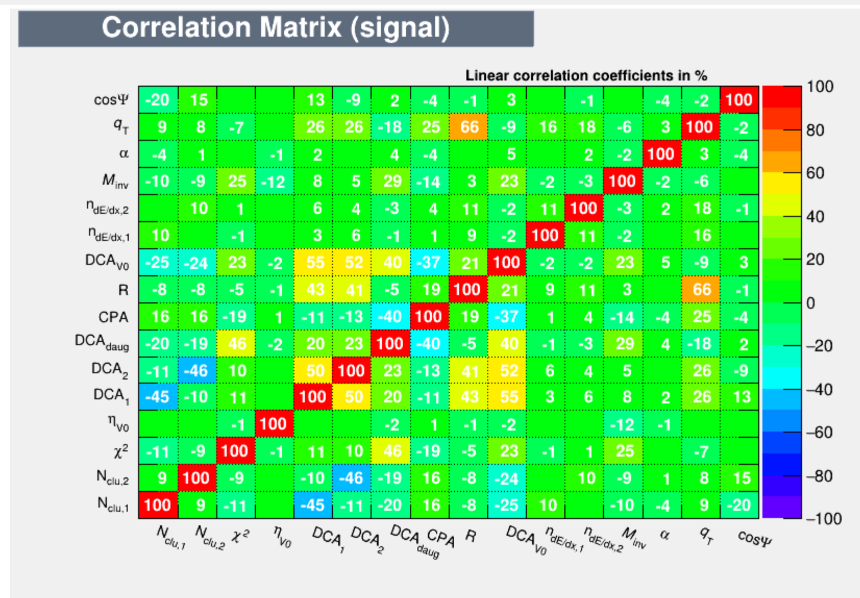



- For electron/positron identification specific energy loss dE/dx in TPC is used (time of flight in case of reconstruction in TOF detector, energy to momentum ratio in case of reconstruction in ECAL)
- For photon selection invariant mass, Armenteros-Podolanski kinematic variables and cos(ψ) (to suppress Dalitz decays) are used




- Examples of normalized signal and background distributions for variables, having the best resolving power are shown

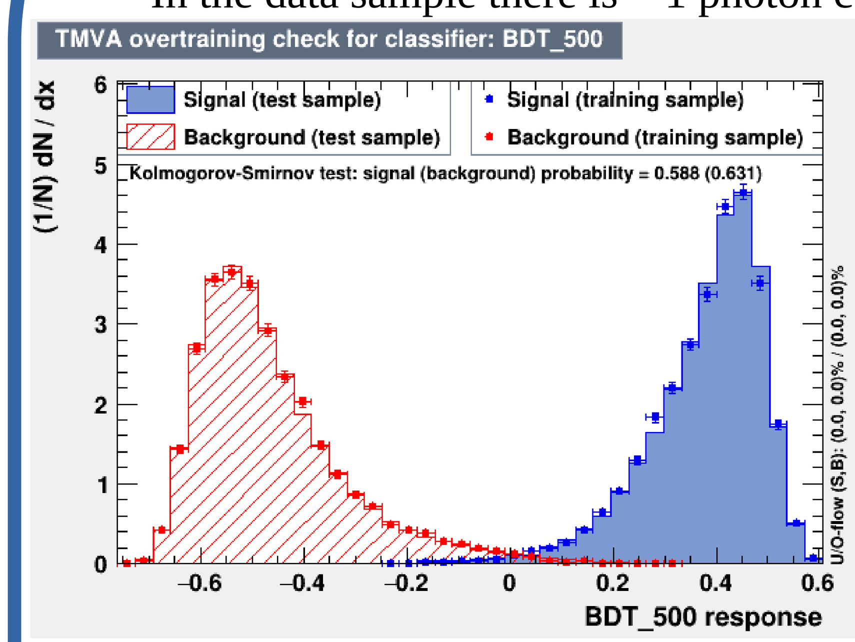


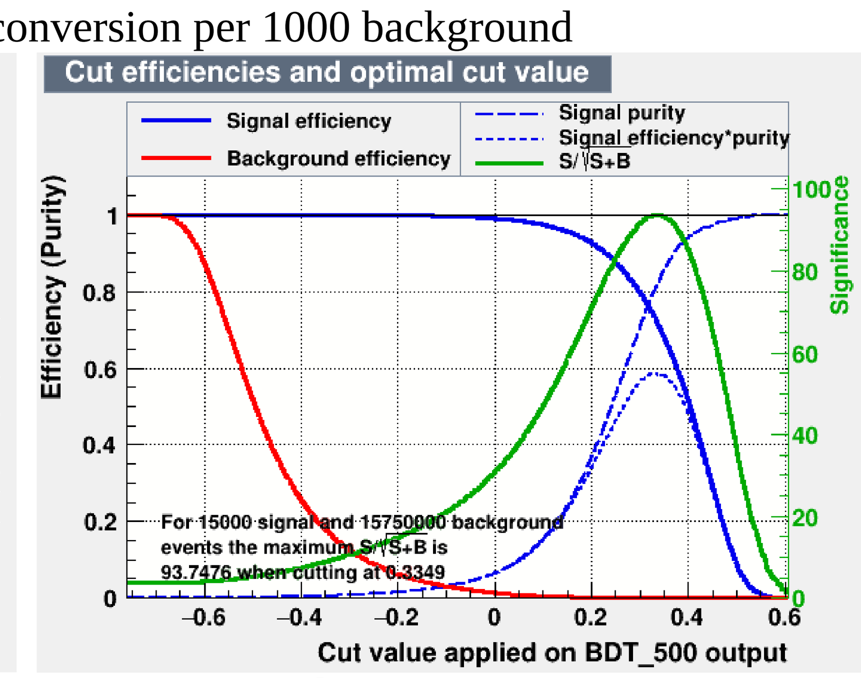




- The best performance of ML approach is reached when all input variables are uncorrelated
- V0 topology variables show some (anti-)correlations. It means that two correlated features may provide nearly the same information and, as a result, tree may make redundant splits without gaining significant new information

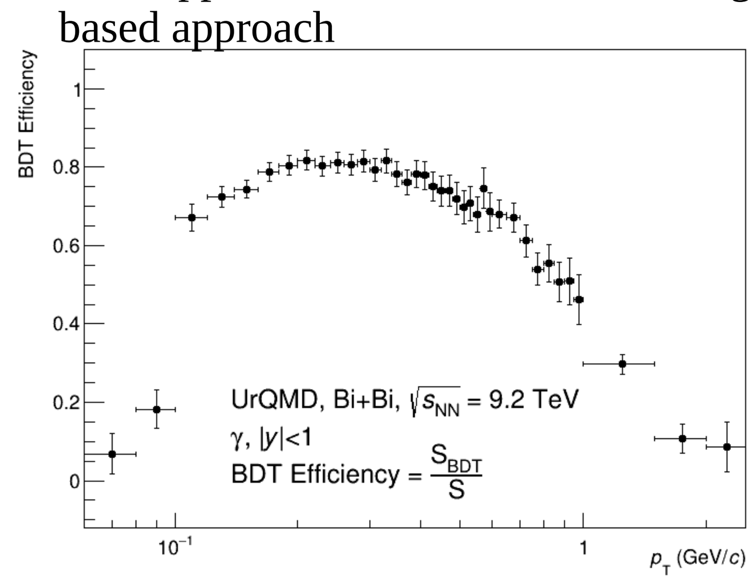
## 5. BDT performance. Efficiency and purity

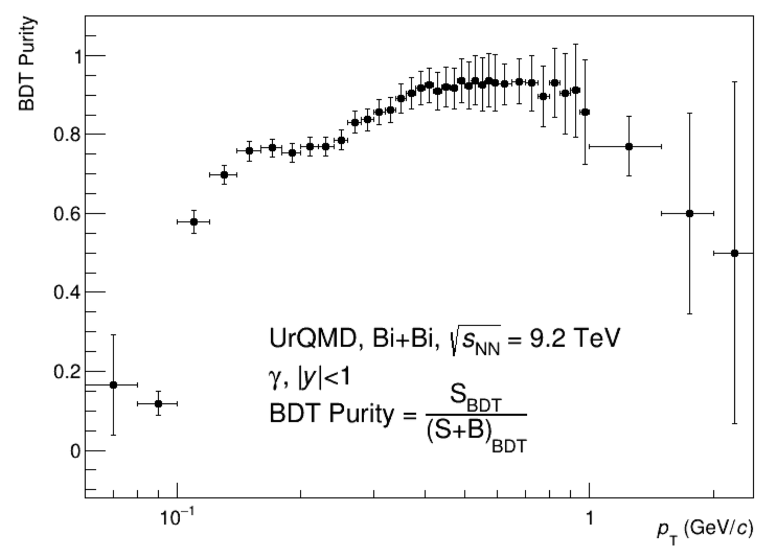
- BDT response and different cut efficiencies are shown below
- In the data sample there is ~ 1 photon conversion per 1000 background

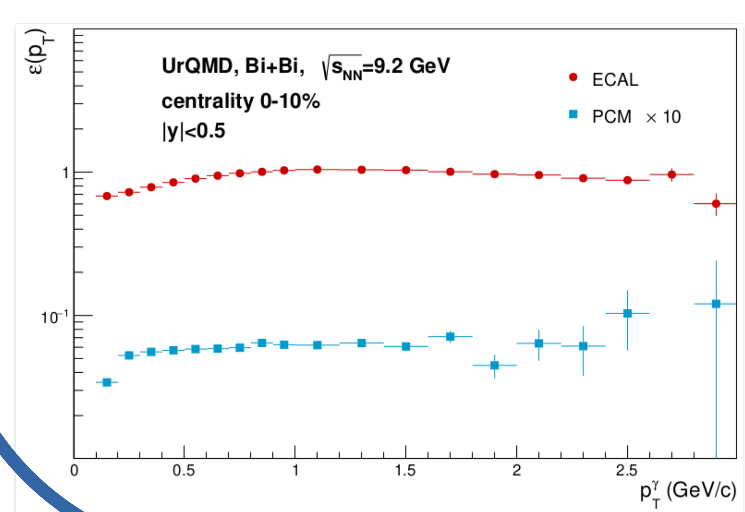


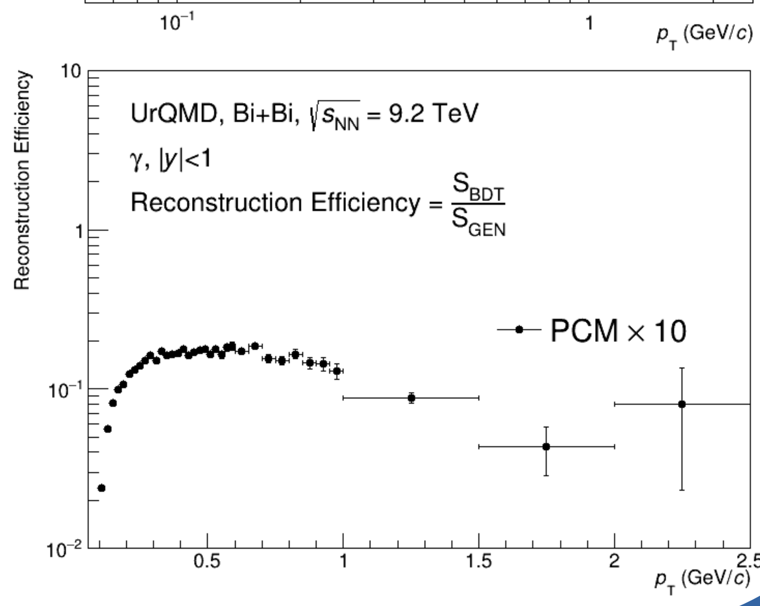




- BDT cut efficiency reaches ~ 80% at $p_T = 0.3$ GeV/c and purity rises up to ~ 90%
- Reconstruction efficiency is small due to low material budget
- BDT approach results in ~ 3 times larger reconstruction efficiency than default cut-based approach




## 6. Conclusion

- Using Machine Learning, and BDT in particular, the improvement in gamma reconstruction efficiency is increased by a factor of ~ 3 compared to default cut-based approach
- BDT efficiency and purity can be improved by performing a $p_T$ differential analysis
- Obtained BDT weights can be used further in neutral meson, direct photon and other measurements


- The work has been supported by the Ministry of Science and Higher Education of the Russian Federation, Project "Fundamental and applied research at the NICA megascience experimental complex" № FSWU-2024-0024



[1] - Golovatyuk, Vea, et al. "The Multi-Purpose Detector (MPD) of the collider experiment." The European Physical Journal A 52.8 (2016): 212.
[2] - Coadou, Yann. "Boosted decision trees and applications." EPJ Web of conferences. Vol. 55. EDP Sciences, 2013.
[3] - Hoecker, Andreas, et al. "TMVA-toolkit for multivariate data analysis." arXiv preprint physics/0703039 (2007).
[4] - CMS collaboration. "Measurement of properties of B $^ 0_\mathrm {s}\to\mu^+\mu^-$ decays and search for B $^ 0\to\mu^+\mu^-$ with the CMS experiment." arXiv preprint arXiv:1910.12127 (2019).
[5] - ATLAS Collaboration. "Search for diphoton resonances in the 66 to 110 GeV mass range using $ pp $ collisions at $\sqrt {s}= 13$ TeV with the ATLAS detector." arXiv preprint arXiv:2407.07546 (2024).