



# Machine learning and event classification

SOTARRIVA ALVAREZ ISAI ROBERTO

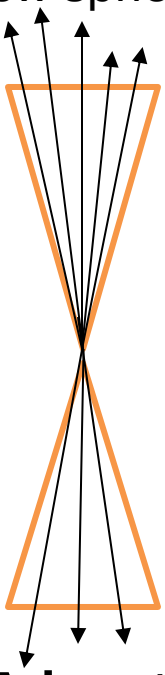
*Advisor Dr. Antonio Ortiz Velasquez*

# Motivation

**Machine learning actual applications**:

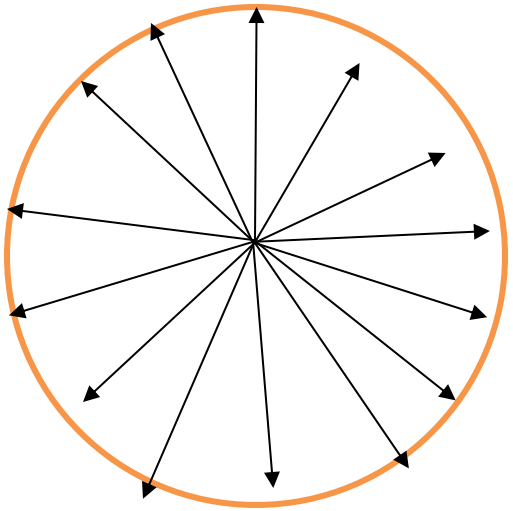
- Image classification
- Medical advisors
- Security
- Financial markets and stocks trading.
- Translation
- Etc.

Jet like
(low spherocity)



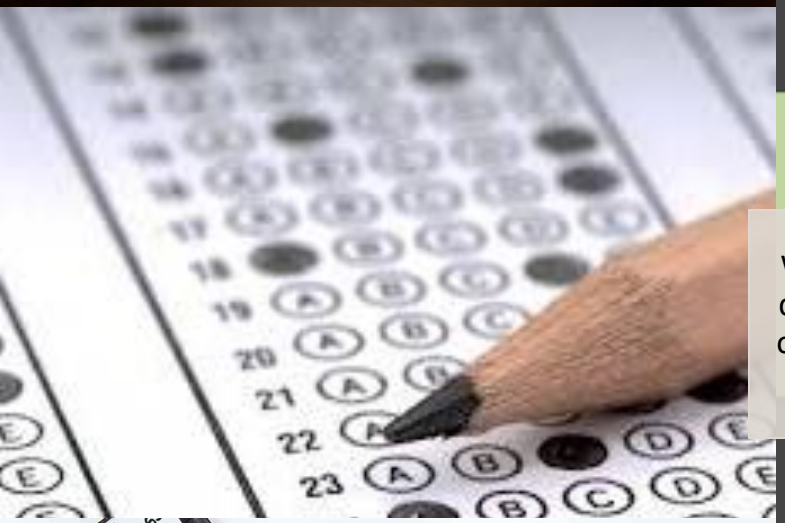
OR

Sphere like
(High spherocity)



**Advantage of machine learning**

More information taken into account → Better predictions



| Machine **learning** | **Human learning** |
| --- | --- |
| **Training:** | |
| We show examples of classified objects to the algorithm. The algorithm learns from them. | We go to school, read books, do some exercises. |
| **Testing:** | |
| We ask the algorithm to classify a new set of data we already know the answers for. Based on the answers of the algorithm we can tell if the algorithm was a good student or not. | We do exams to measure how good we have become after studying. |
| **Evaluation:** | |
| We ask the algorithm to work on unclassified sets of data. | We apply what we learned on the daily life or at work. |

# Isolation of real spherical events

- The algorithms trained were: MLPBNN, FDA_GA, BDT (using Adaptative boost) and LD.
- The methods are trained and tested using MC information MC production anchored to LHC15f pass 2 (pp collisions @ 13 TeV) 50% for training and 50% for testing.
- Standard event and track selection.

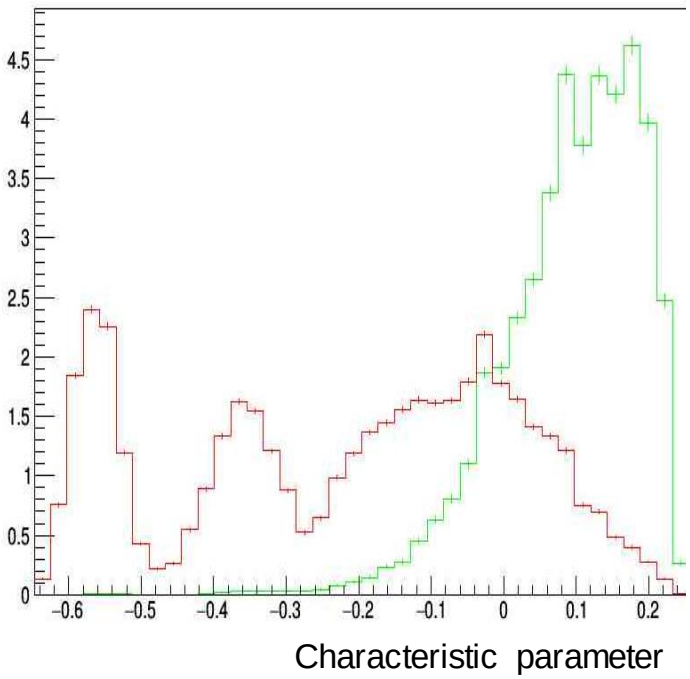
# Method response
## High multiplicity

- Isolation of events with a large number of charged particles isotropically distributed

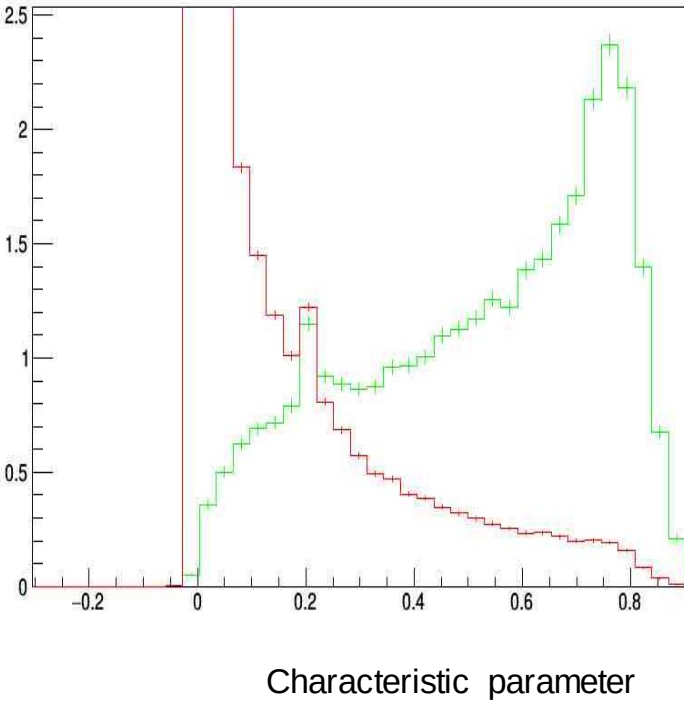
signal spherocity true>0.8

background spherocity true<=0.8

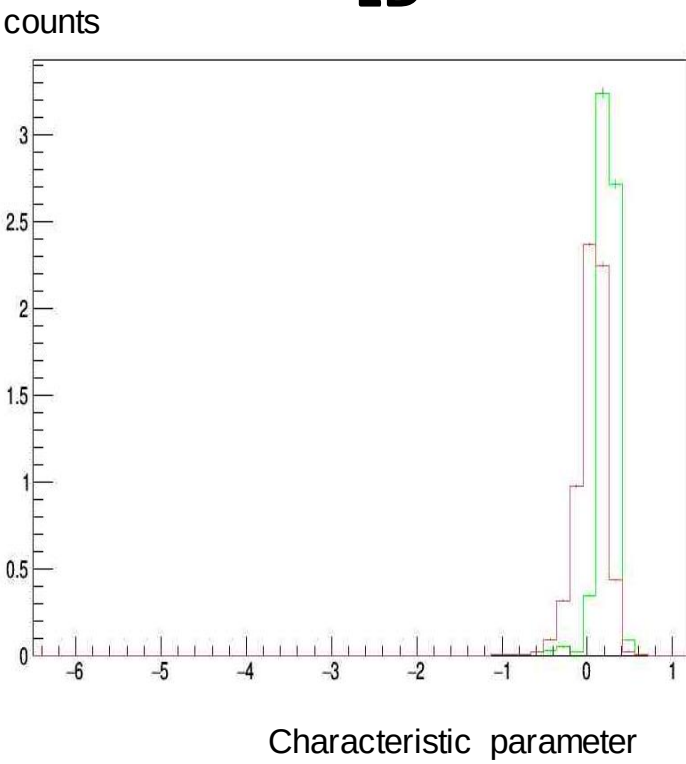
**BDT**



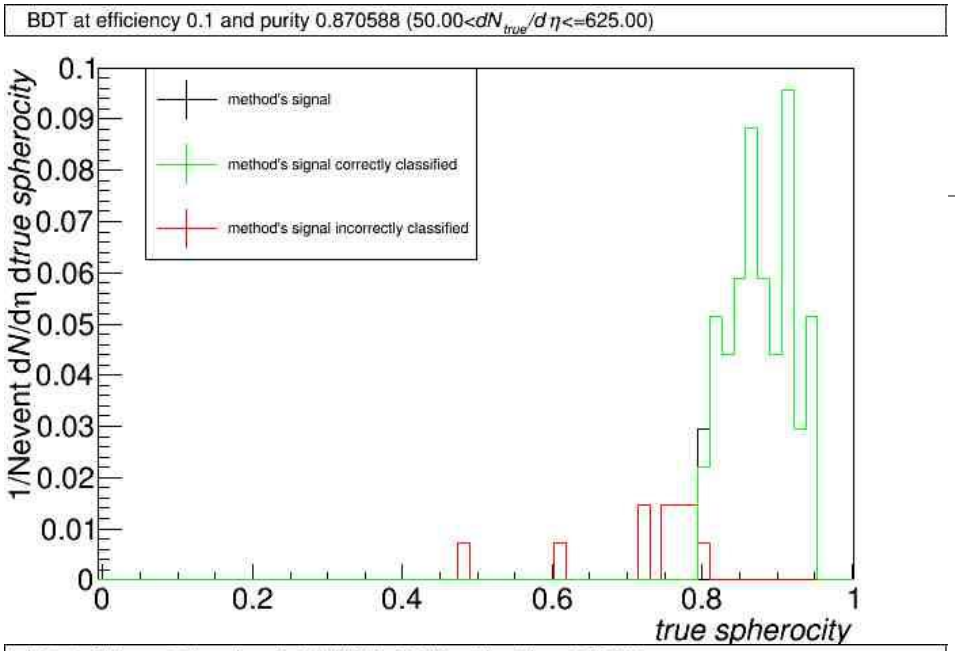

**MLPBNN**

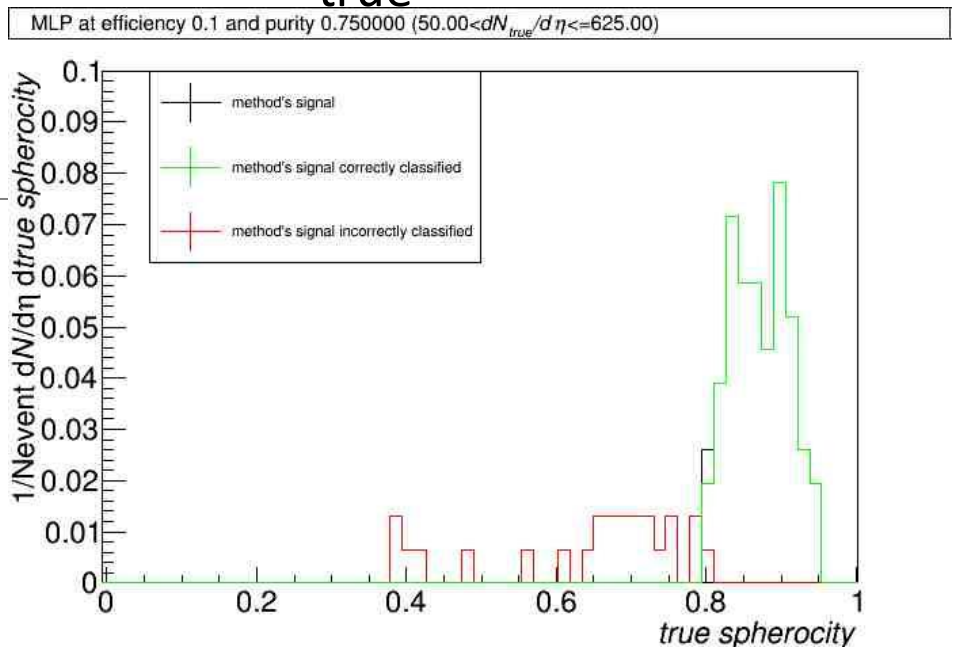


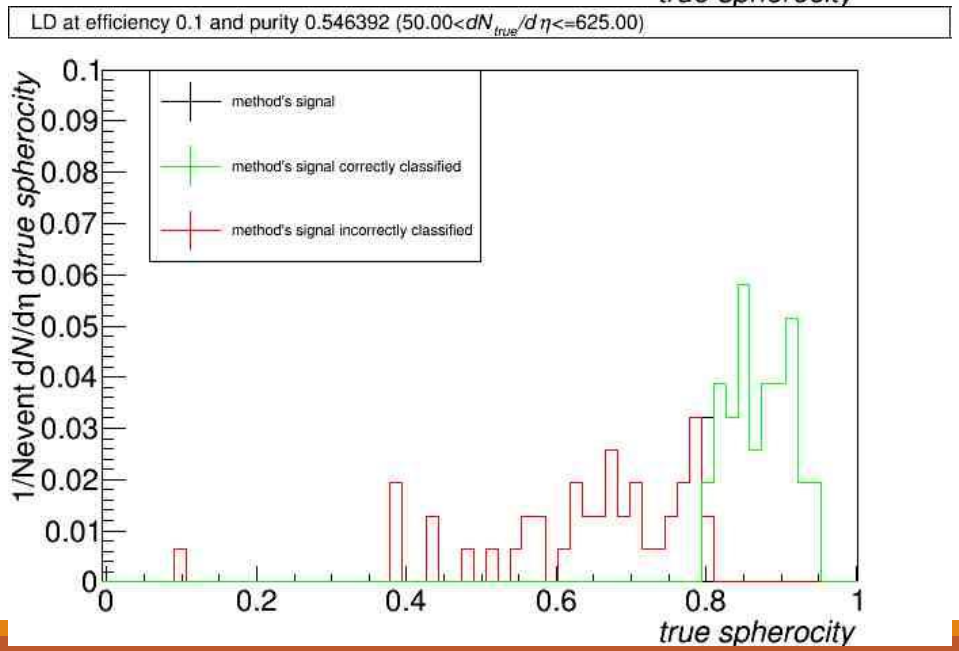

**LD**

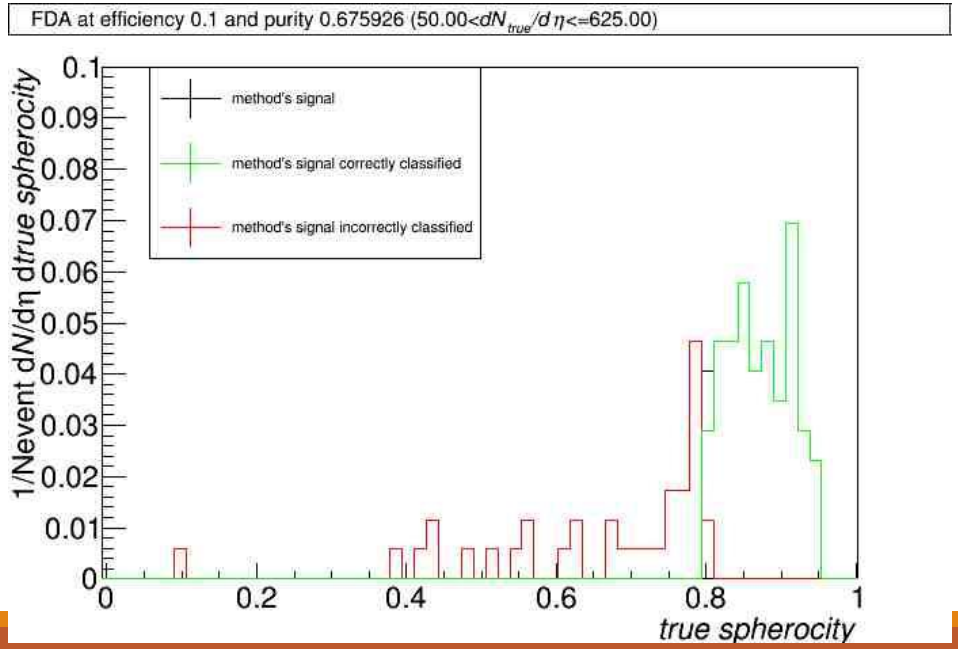


# True spherocity at 10% efficiency
# True multiplicity 50.0<$dN_{true}/d\eta$

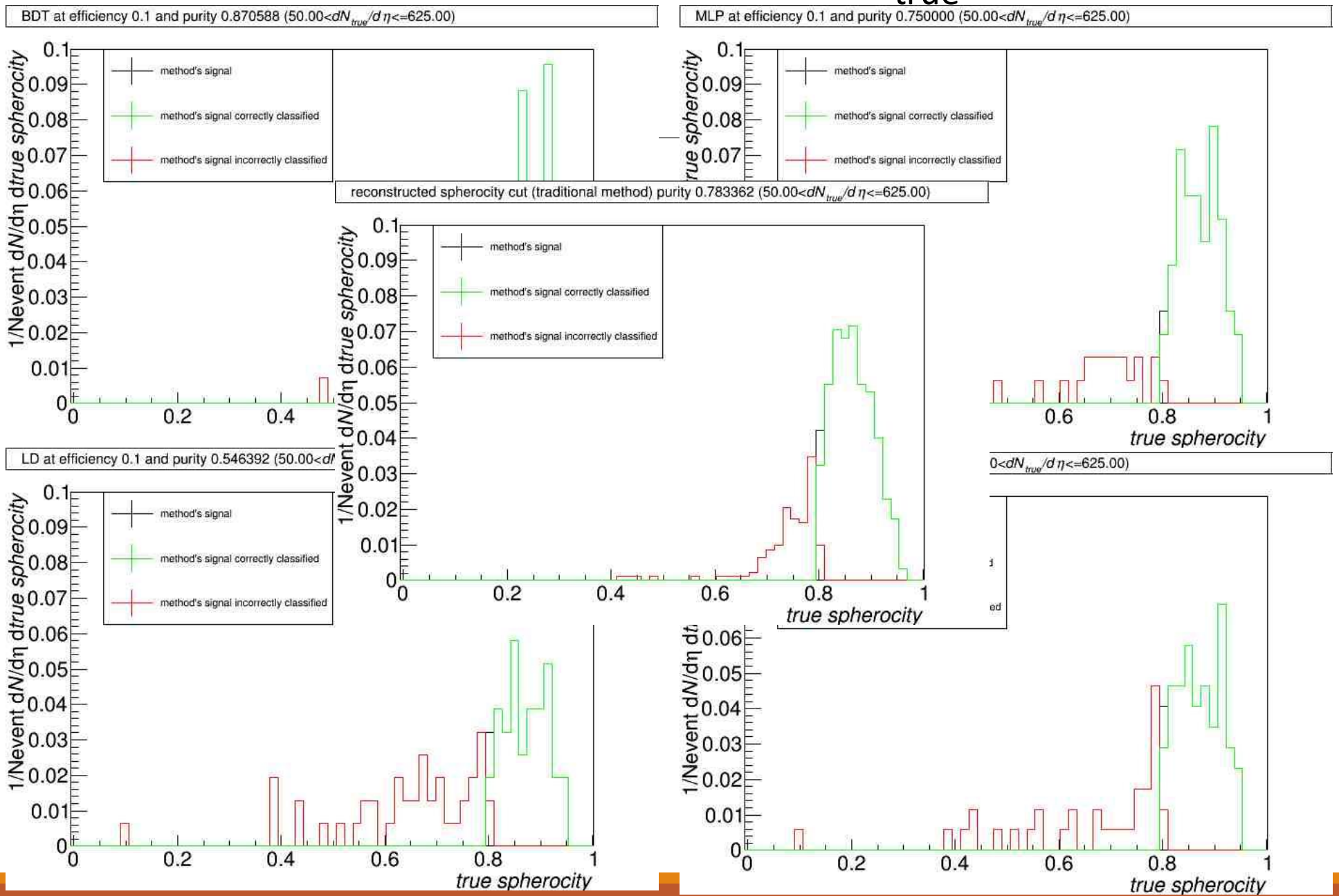


# True spherocity at 10% efficiency
# True multiplicity 50.0<$dN_{true}/d\eta$

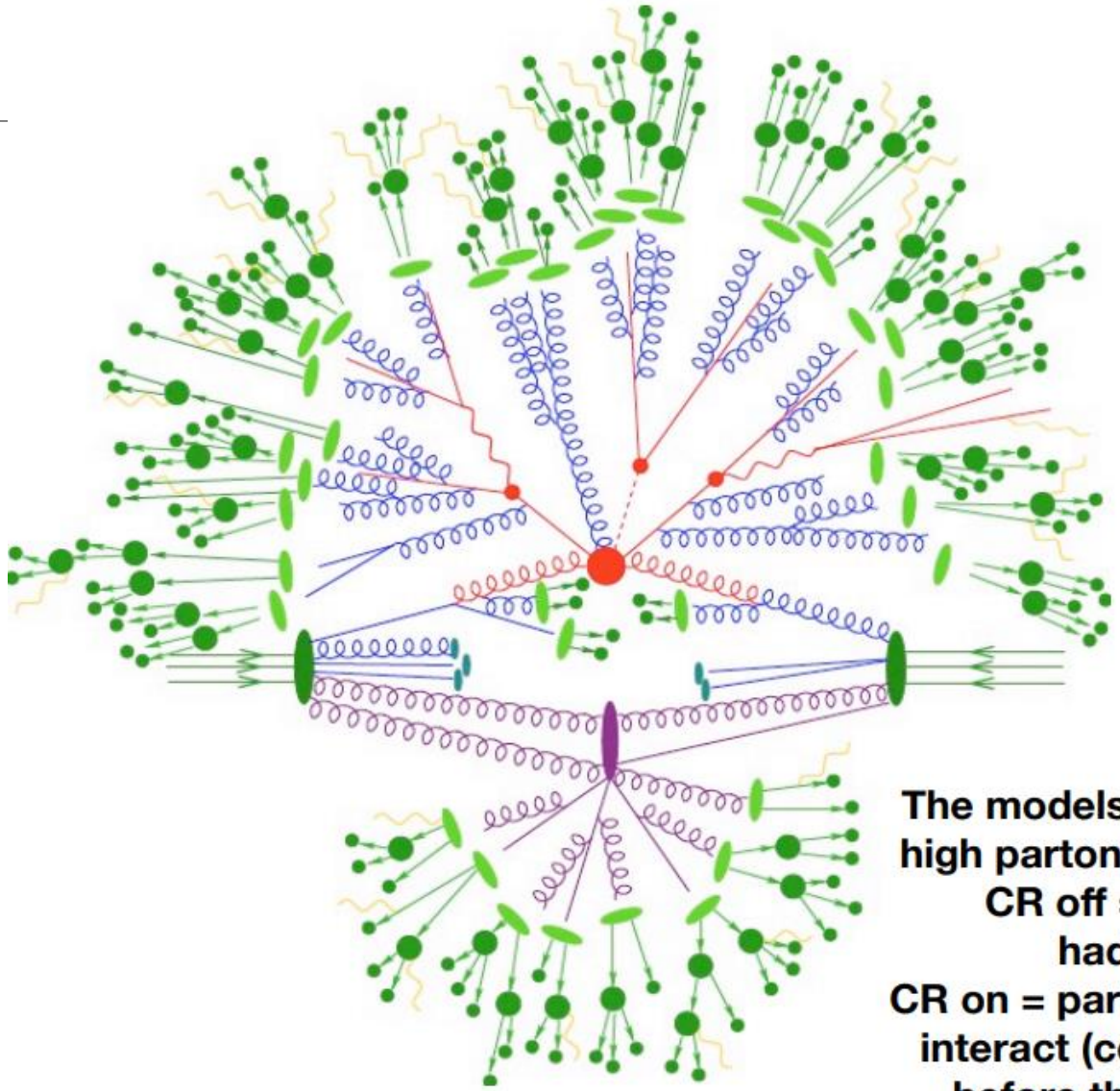


# NMPI classification

TESIS PROJECT

- Objetive: We want to improve the isolation of events with high number multiparton interactions using only reconstructed quantities.

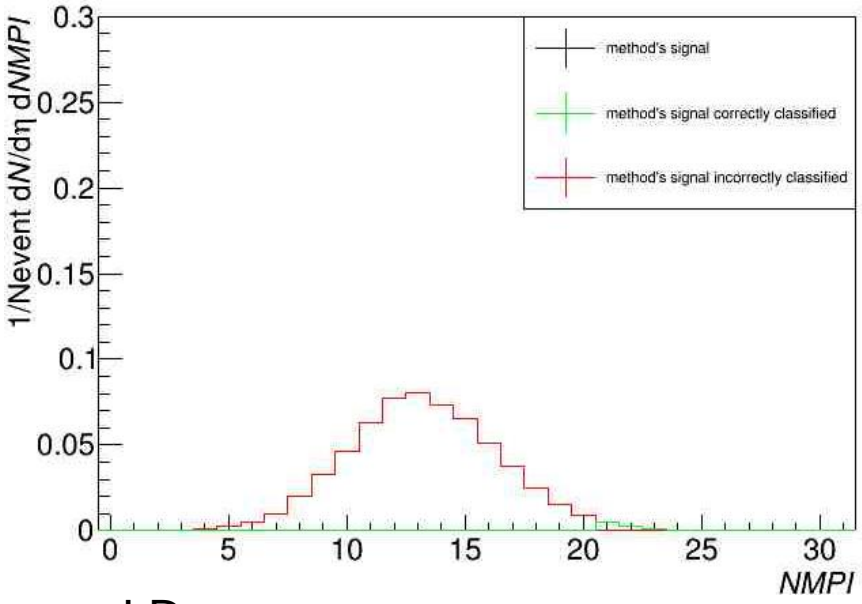
# Number of multiparton interactions



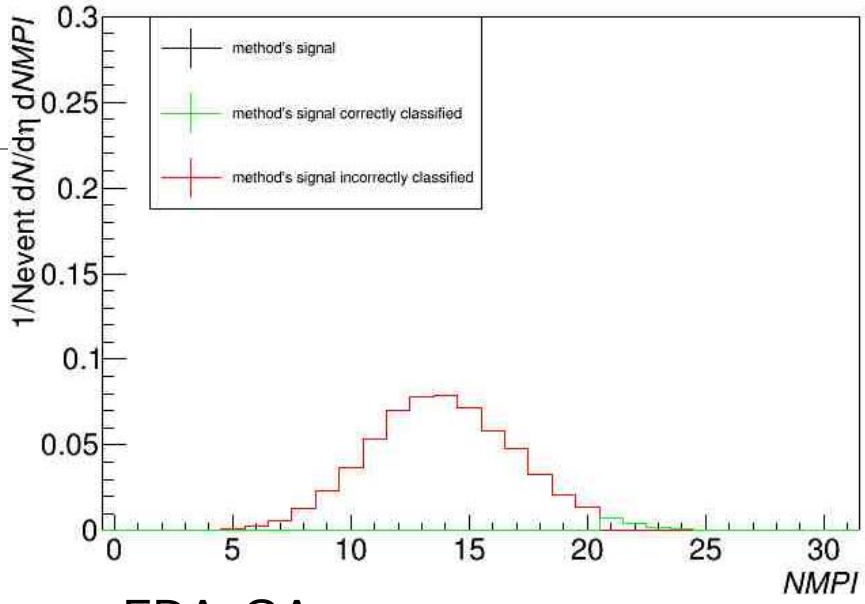
The models with MPI produce high partonic density systems
CR off = independent hadronization
CR on = partons are allowed to interact (connected system) before the hadronization

# High NMPI classification efficiency =0.2

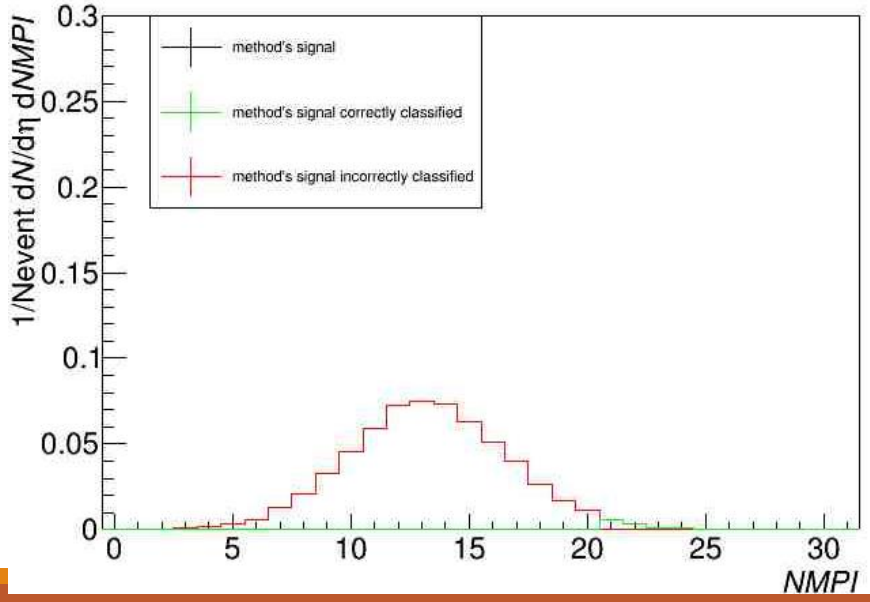
BDT



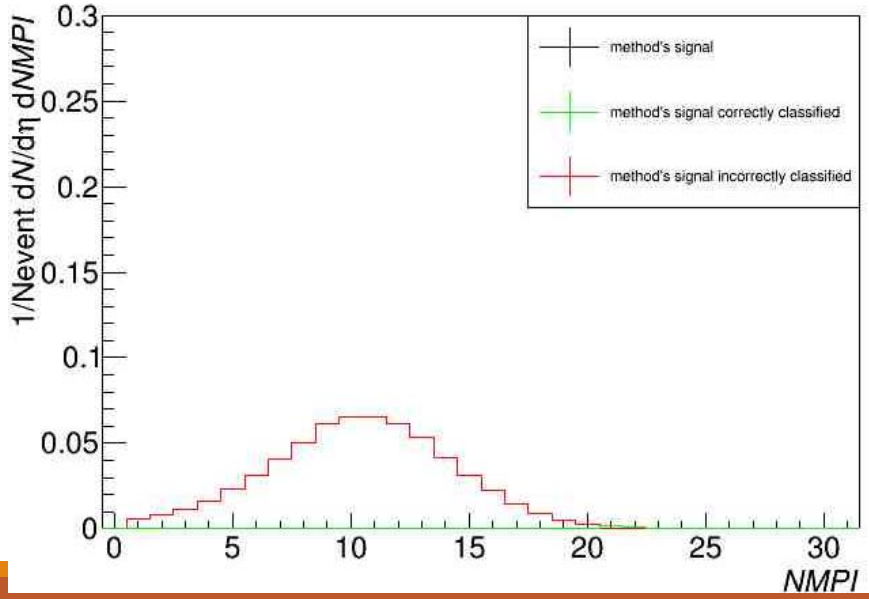

MLPBNN




LD

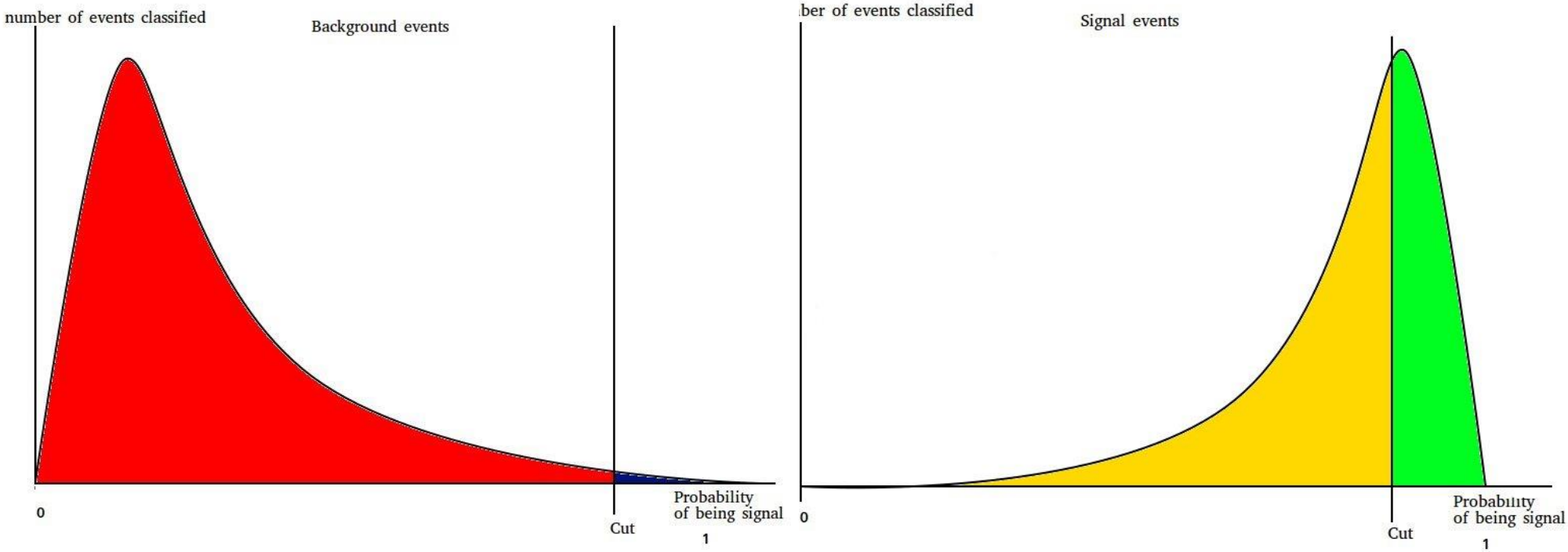



FDA_GA



# Back up slides

# ¿which method is better?




Signal efficiency= signal events classified as signal by the algorithm/ the total number of signal events=green/(green+yellow)

Signal purity=signal events correctly classified/Events classified as signal=(green/green+blue)

# Summary NMPI

- For number of multiparton interactions (NMPI) methods are trained using the MC production:LHC18f1(pp collisions @ 13 TeV) anchored to LHC16k for training.
- And MC production:LHC15g3c3(pp collisions @ 13 TeV) for testing.
- Standard event and track selection.

# Isolation of real spherical events

- *Objective: Classify on signal (true spherocity>0.8) and background (true spherocity<=0.8) using only reconstructed quantities.*

- Preclassified set according to true multiplicity in multiplicity classes.

- cuts $|\eta| < 0.8$, $0.15 < p_T$ and at least 3 MCparticles per event. Spherocity and sphericity require at least 3 particles to be calculated.

- Training variables (all of them after the simulated detector reconstruction):

- average $p_T$
- Sphericity
- Multiplicity
- Recoil (Momentum balance)
- $p_T$ leading (Sensitive to hard physics)