Tau-shower classification in ARCA

Group Meeting 01-04-2021

G.J.P.W. Vermariën, Universiteit Leiden & Nikhef.

Universiteit Leiden
The Netherlands

Nikhef

Task

- Task: classify tau shower (0) vs electron CC(1)
- ∼100 000 samples (approximately balanced, 1:2 ν_τ−**shower** : $\nu_{e,CC}$ ratio)
- Every event file is subsampled if over 5000 hits such that it becomes 5000 hits long
- Full ARCA detector
- `Input: (x,y,z,ct*, dir_x, dir_z, dir_y)` *experimented with leaving it out
- `Output: [p_is_tau,p_is_not_tau]`

Feature Visualisation

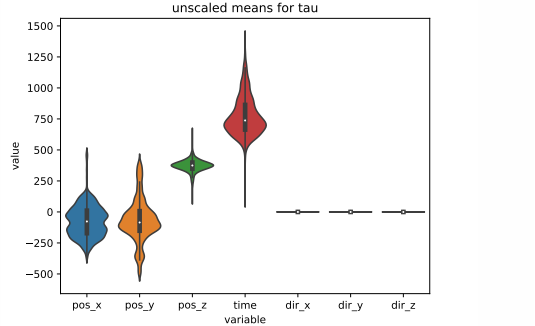


Figure: ν_τ *feature means for 5000 samples*

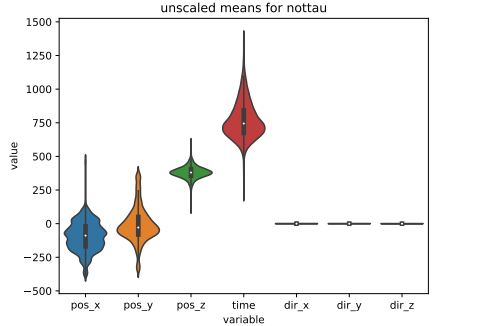


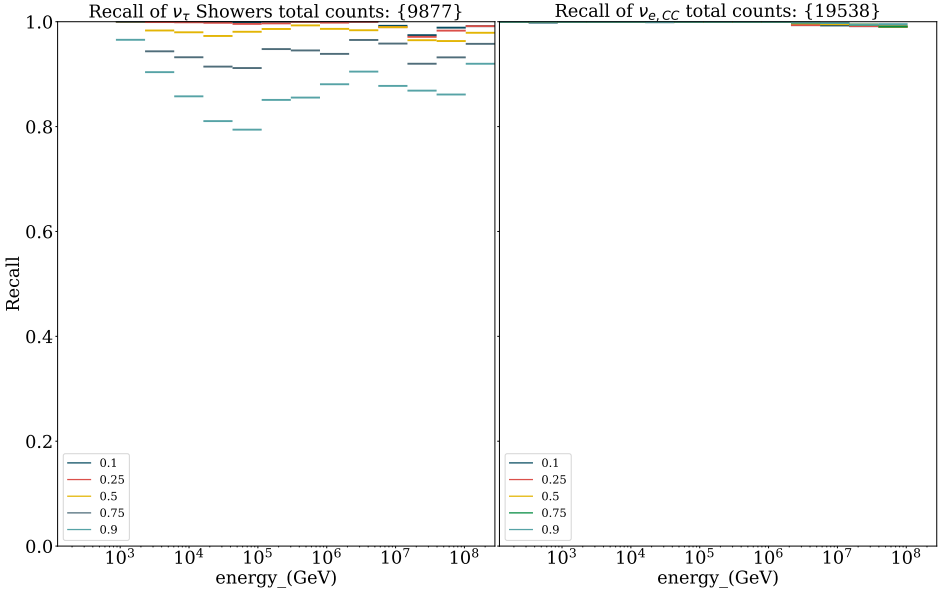
Figure: $\nu_{e,CC}$ *feature means for 5000 samples*

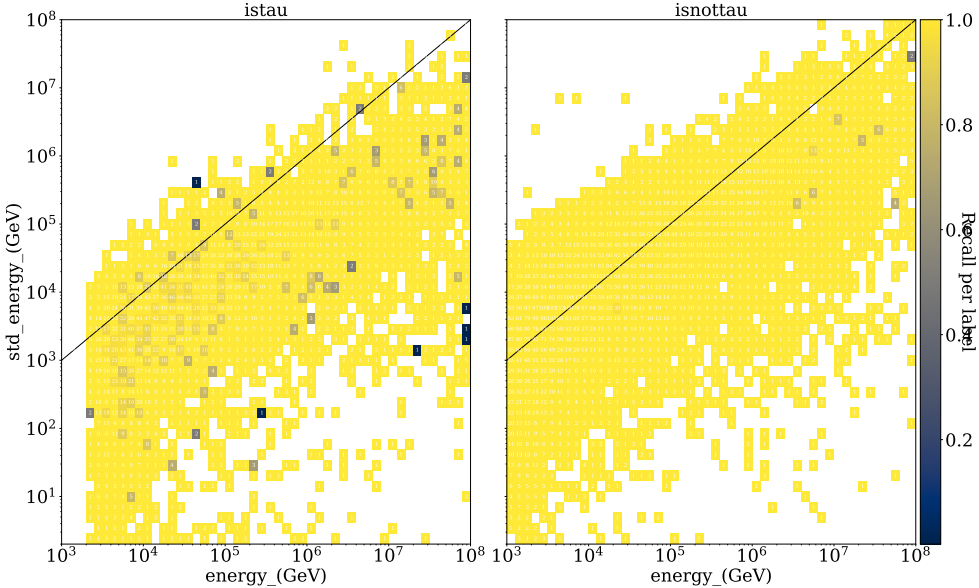
First with all the features

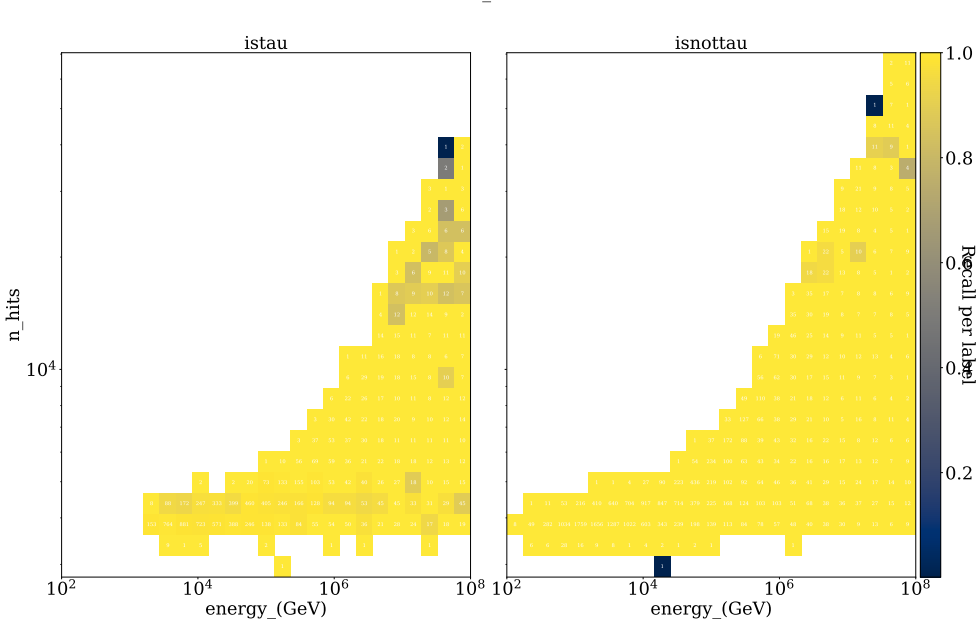
- Reached high accuracy (99%), but convergence properties of the algorithm are still suboptimal

Might be due to the usage of two different simulations of data

- 'ML_mcv5.1_res.genhen_nueCC.km3_AAv1.jte.jchain.aashower.*.h5'
- 'ML_mcv5.1_res.genhen_nutau-CCHEDIS-shower_1e3-1e8GeV.sirene.jte.jchain.aashower.*.h5'

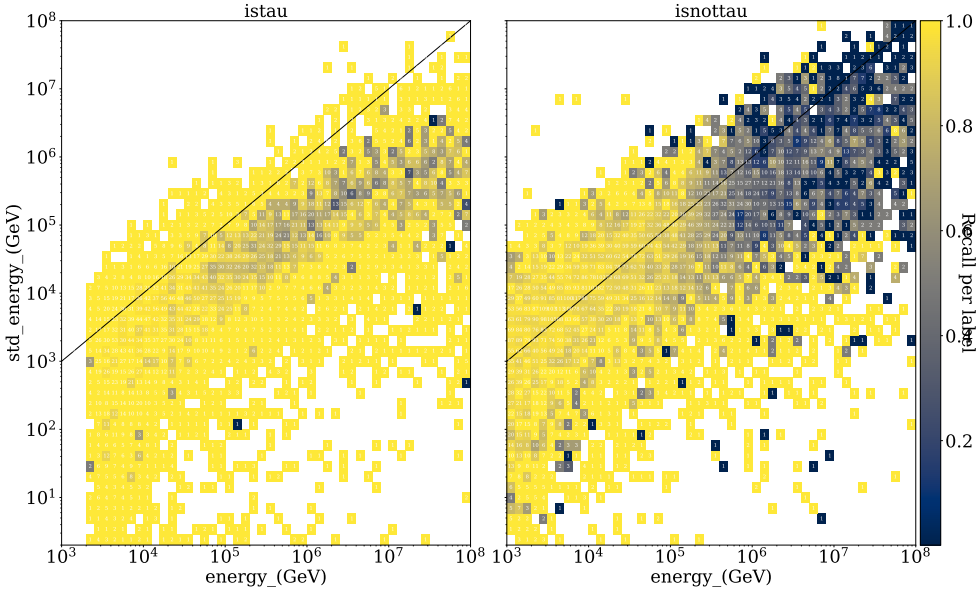

Recall of ν_τ Showers total counts: {9877}
Recall of $\nu_{e,CC}$ total counts: {19538}
Recall
1.0
0.8
0.6
0.4
0.2
0.0
10^3
10^4
10^5
10^6
10^7
10^8
energy_(GeV)
0.1
0.25
0.5
0.75
0.9

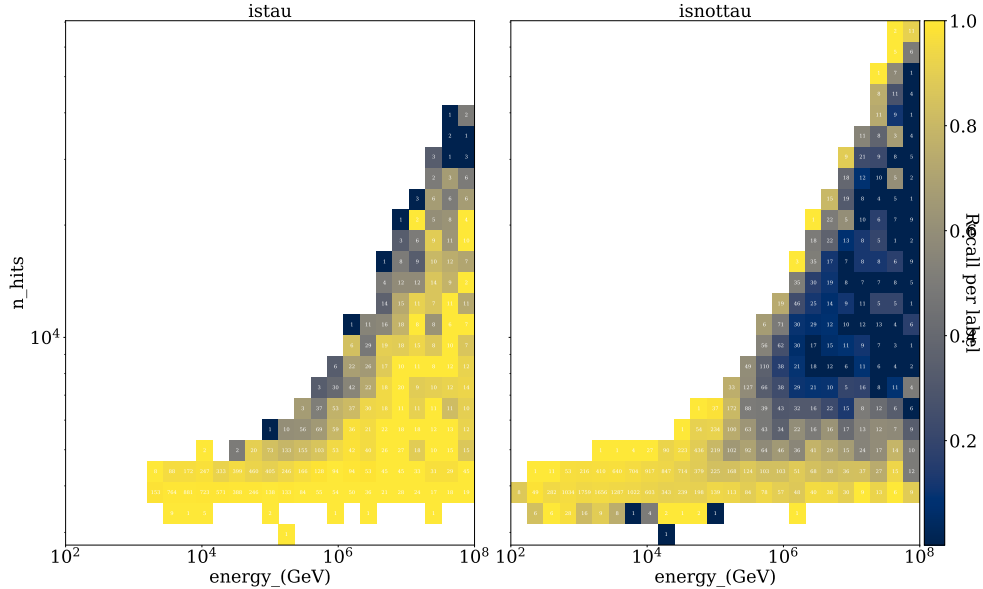

istau
isnottau
std_energy_(GeV)
energy_(GeV)
energy_(GeV)
Recall per label
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
0.2
0.4
0.6
0.8
1.0

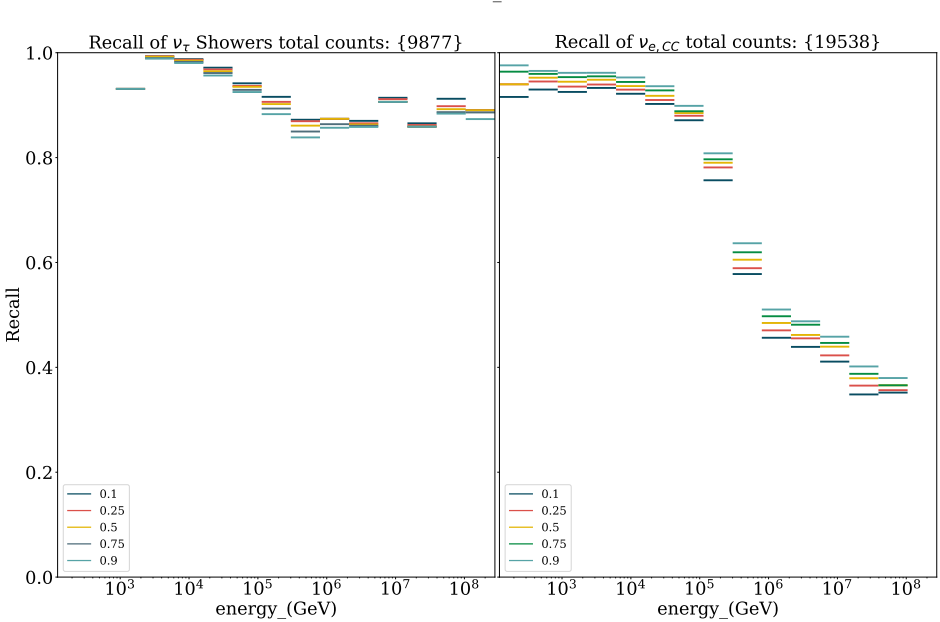

with_time
istau
isnottau
n_hits
10^4
10^2
10^4
10^6
10^8
energy_(GeV)
Recall per label
1.0
0.8
0.6
0.4
0.2

See how the relatively large values of time influence the classification

- What if we remove the time feature?


without_time
istau
isnottau
std_energy_(GeV)
energy_(GeV)
energy_(GeV)
Recall per label
1.0
0.8
0.6
0.4
0.2


without_time
istau
isnottau
n_hits
10^4
10^2
10^4
10^6
10^8
energy_(GeV)
Recall per label
1.0
0.8
0.6
0.4
0.2


without_time
Recall of ν_τ Showers total counts: {9877}
Recall of $\nu_{e,CC}$ total counts: {19538}
Recall
1.0
0.8
0.6
0.4
0.2
0.0
10^3
10^4
10^5
10^6
10^7
10^8
energy_(GeV)
0.1
0.25
0.5
0.75
0.9

Outlook

- See if standardizing features yields better results ($\tilde{\boldsymbol{x}} = (\boldsymbol{x} - \boldsymbol{\mu})/\boldsymbol{\sigma}$)
- Use files from the MC 6 production
- Expand analysis of results based on features such as contained in detector and tau decay length