Application of CMS DeepJet to ILD

Mareike Meyer

Meeting on advanced jet flavour tagging, 15/03/2023

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

Introduction

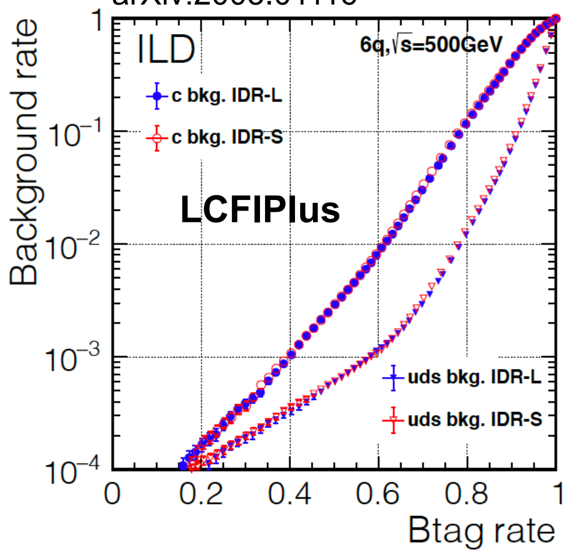
- current standard: **LCFIPlus**
  - arXiv:1506.08371, https://github.com/lcfiplus/LCFIPlus
- based on TMVA (BDTs)

➡ Can the **heavy flavour tagging** be **improved** by replacing the BDTs used in LCFIPlus **with (deep) NNs**?

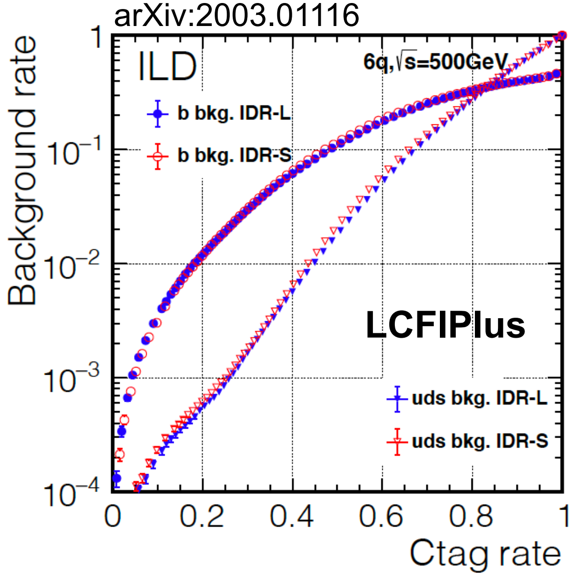
this work:

- application of **„CMS DeepJet“** to ILD
  - „Jet Flavour Classification Using DeepJet“, arXiv:2008.10519
  - „Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV“, arXiv:1712.07158

arXiv:2003.01116

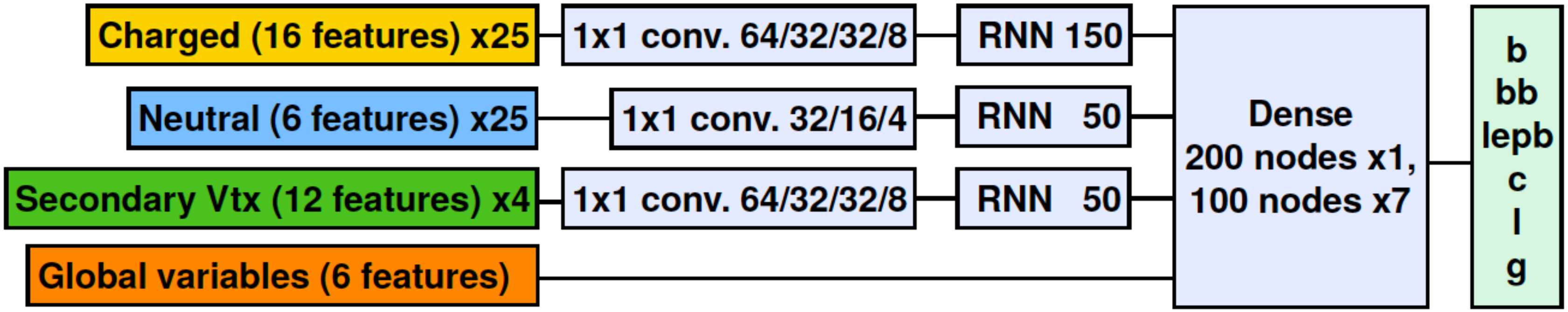


arXiv:2003.01116

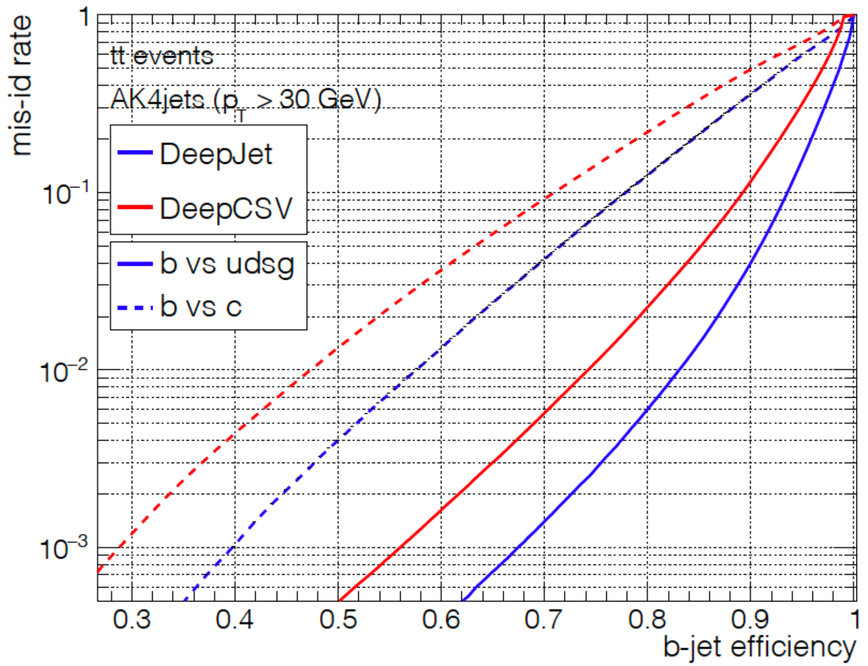


CMS DeepJet

arXiv:2008.10519, arXiv:1712.07158



- successfully applied in many recent CMS analyses
- allows for usage of low-level features from many jet constituents
- able to deal with variable length sequence of inputs
- allows for ordering of particles according to their assumed importance
- large gain in performance compared e.g. to FCNN (DeepCSV)



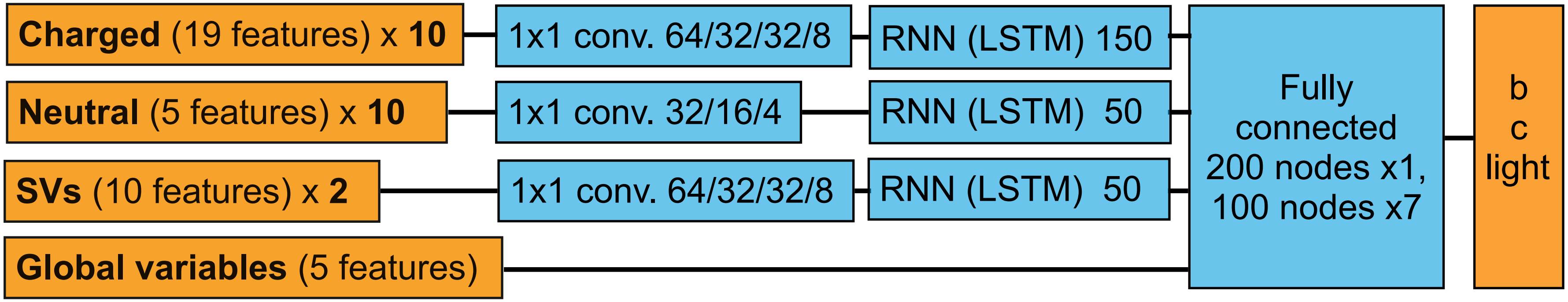
Training data

- train on **events with 6 jets** (b,c,u,d,s)
  - /pnfs/desy.de/ilc/prod/ilc/mc-opt-3/ild/dst-merged/500-TDR_ws/flavortag/ILD_l5_o1_v02/v02-00-01/
- run PV & SV finder, jet clustering and vertex refinement of LCFIPlus
- number of jets:
  - b jets: 578844
  - c jets: 645114
  - light jets: 1932975
- under-sampling of light and c jets performed to get same number of b,c and light jets for training and testing

➡ **total number of jets**: 3 * 578844 = 1736532

- split sample into training, validation and test (75% / 12.5% / 12.5%)

Architecture & data pre-processing

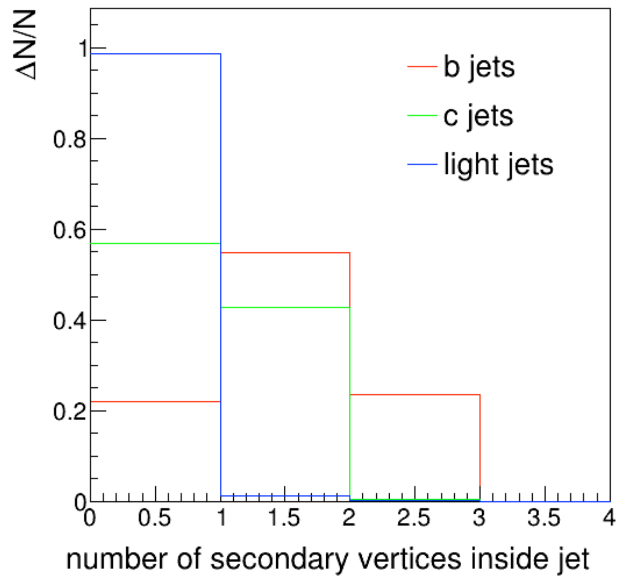


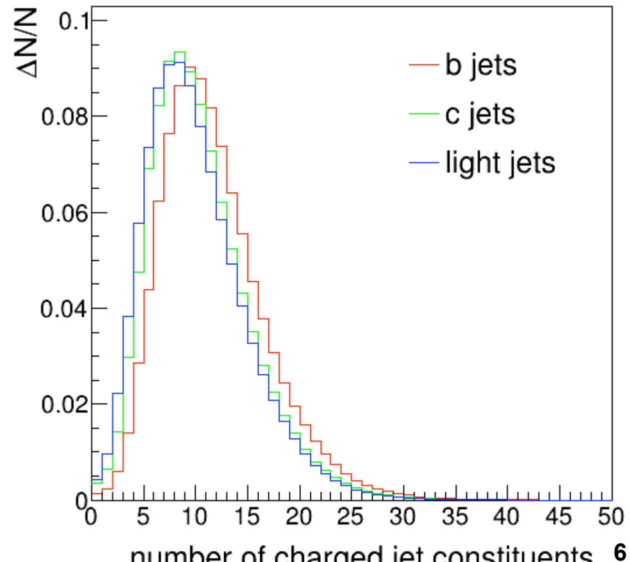
- classify jets into **three classes**: b jets, c jets & light jets
- **ordering of input particles** by (as applied in CMS)
  - impact parameter significance for charged jet constituents
  - shortest angular distance to a secondary vertex (by momentum if there is no secondary vertex) for neutral jet constituents
  - flight distance significance for secondary vertices
- if a value of a features is not available, the value is set to -10
- **normalize input features** to mean 0, std 1

Input features - global variables

- jet momentum
- jet transverse momentum
- number of charged jet constituents
- number of neutral jet constituents
- number of secondary vertices

5 input features

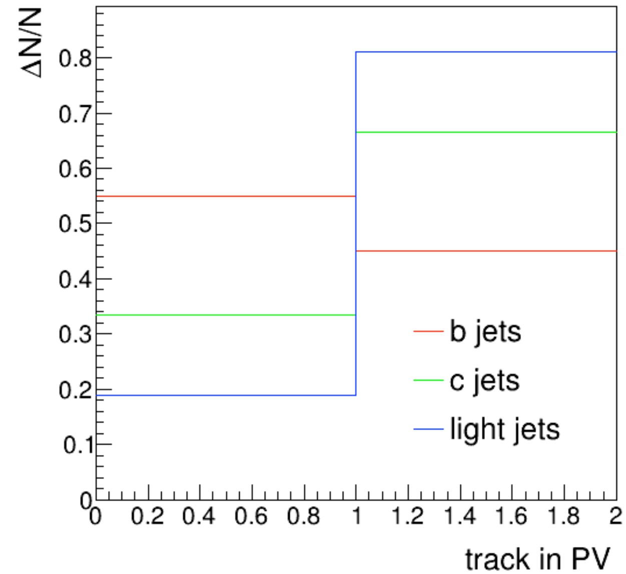


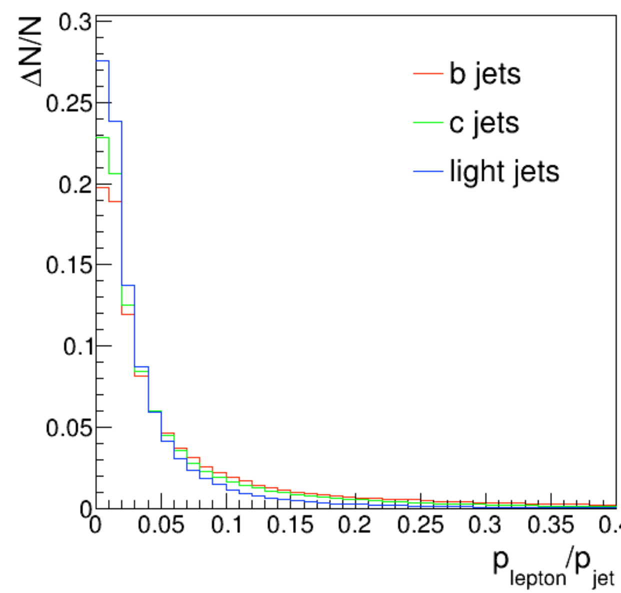


Input features - charged jet constituents

- track momentum / jet momentum
- transverse track momentum relative to jet
- dot product of jet and track momentum w.r.t. jet momentum
- ΔR(track, jet),
- d0, d0 significance
- Z0, Z0 significance
- 3D impact parameter, 3D impact parameter significance
- track reconstructed in PV?
- is electron?, is muon?, lepton momentum relative to jet, lepton transverse momentum relative to the jet, lepton momentum / jet momentum
- kaon-ness of charged particles, track momentum fraction weighted with kaon-ness
- $\chi 2$/ndf

19 input features

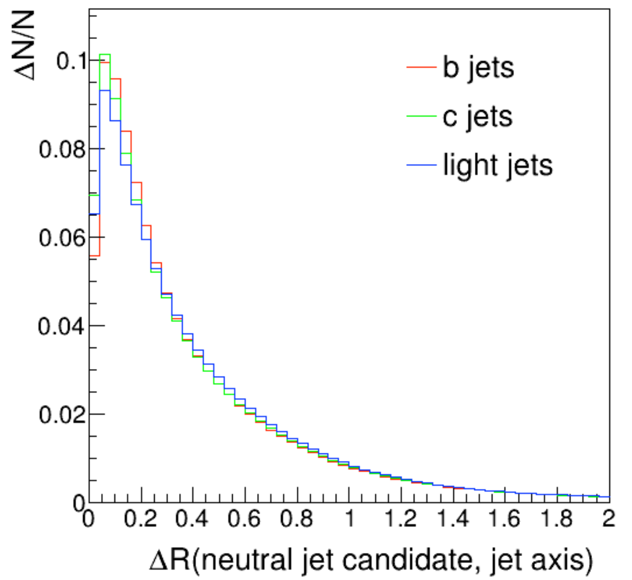


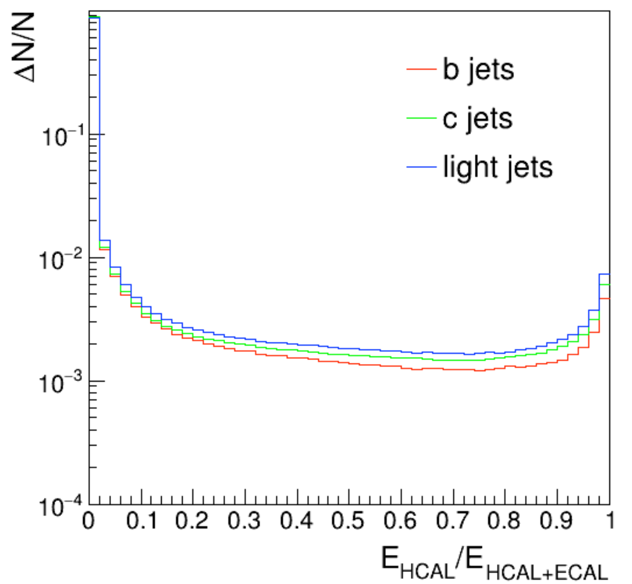


Input features - neutral jet constituents

- momentum of neutral jet constituent
- fraction of the jet momentum carried by neutral jet constituent
- ΔR(jet axis, neutral candidate),
- is photon?
- fraction of neutral candidate energy deposited in the hadronic calorimeter

5 input features

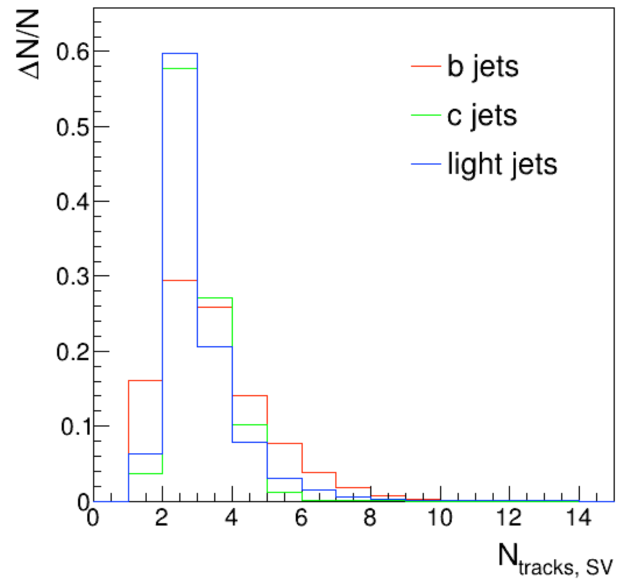


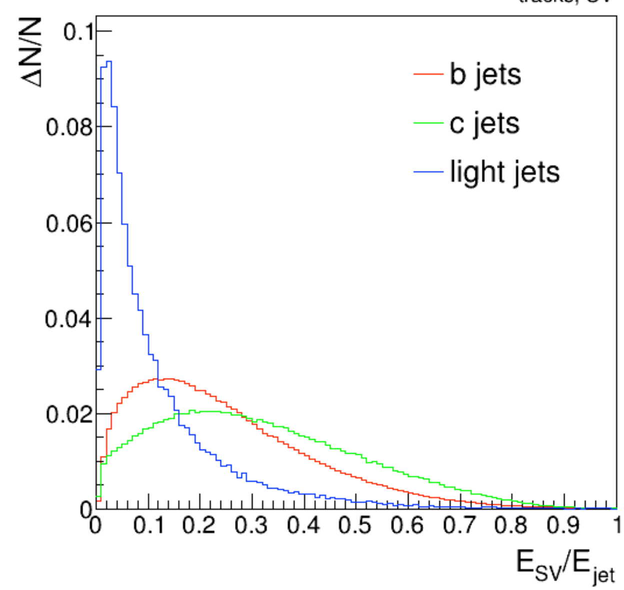


Input features - secondary vertices

- SV mass
- number of tracks in SV
- ΔR(SV, jet)
- SV energy / jet energy
- SV energy
- cosine of the angle between the secondary vertex flight direction and the direction of the secodary vertex momentum
- 3D impact parameter, 3D impact parameter significance
- χ2, ndf

10 input features

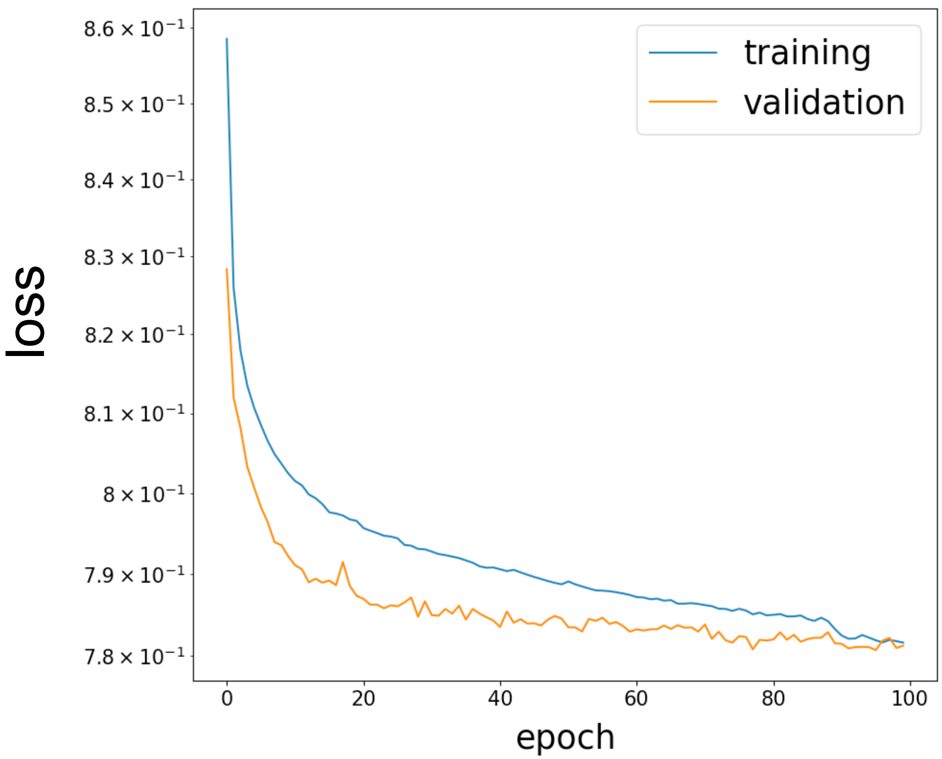


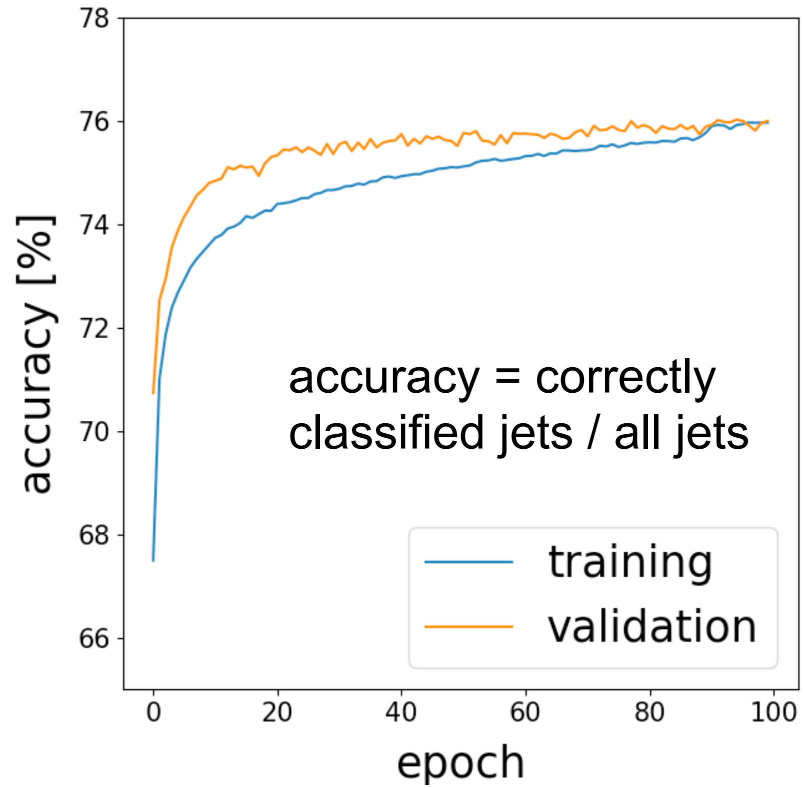


Training

- activation functions: relu / softmax (last layer)
- cross entropy loss
- optimizer: Adam
- regularization: batch normalization, dropout (0.1)
- batch size: 200
- learning rate: 0.0003
- learning rate is halved if validation loss stagnates for 10 epochs
- number of epochs: 100
- Xavier weight initialization

Results: loss & accuracy

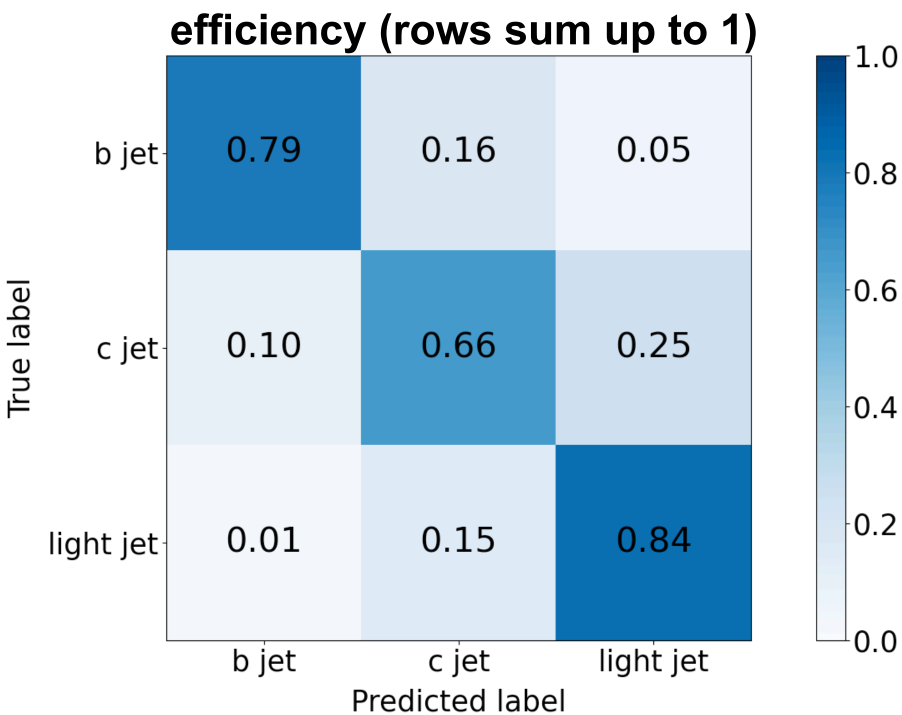


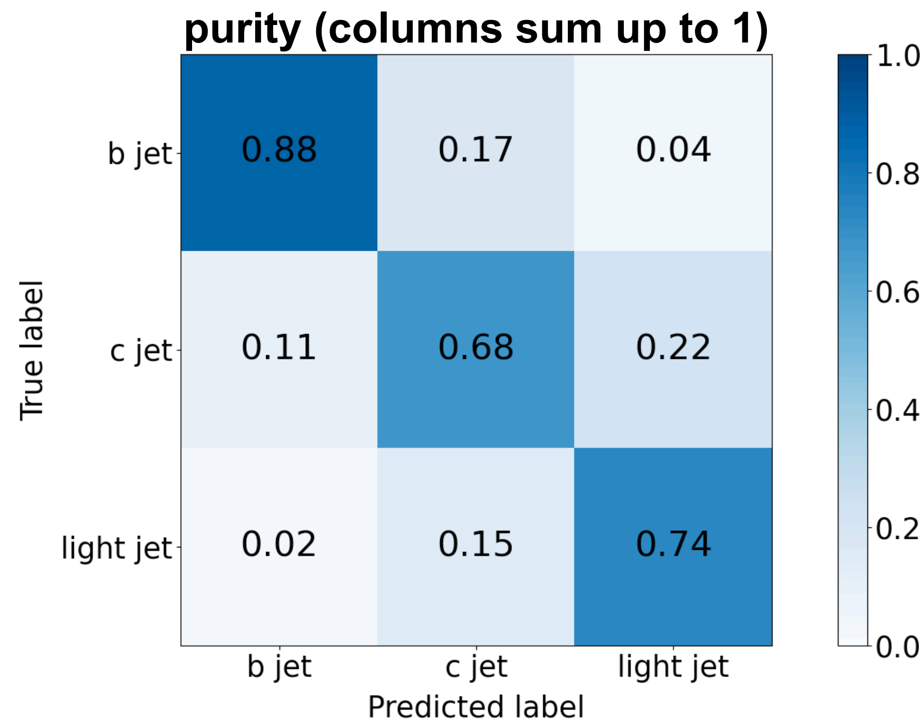


- accuracy ~76% in training & validation data
- epoch 89: learning rate halved
- train longer? decrease dropout rate?

Results: confusion matrices

validation data

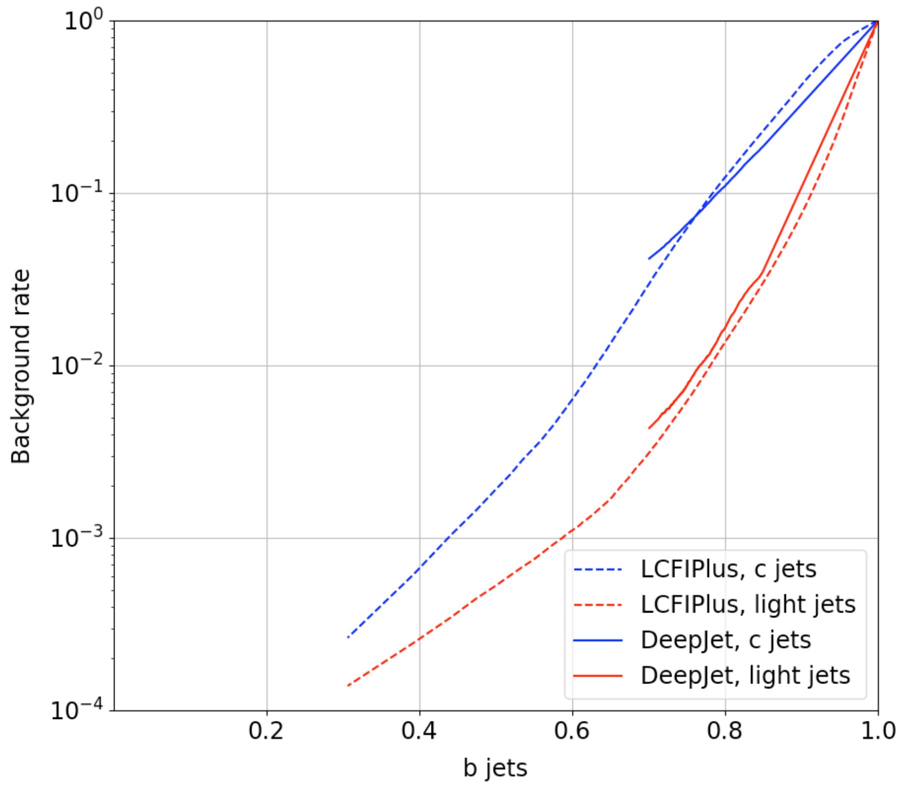


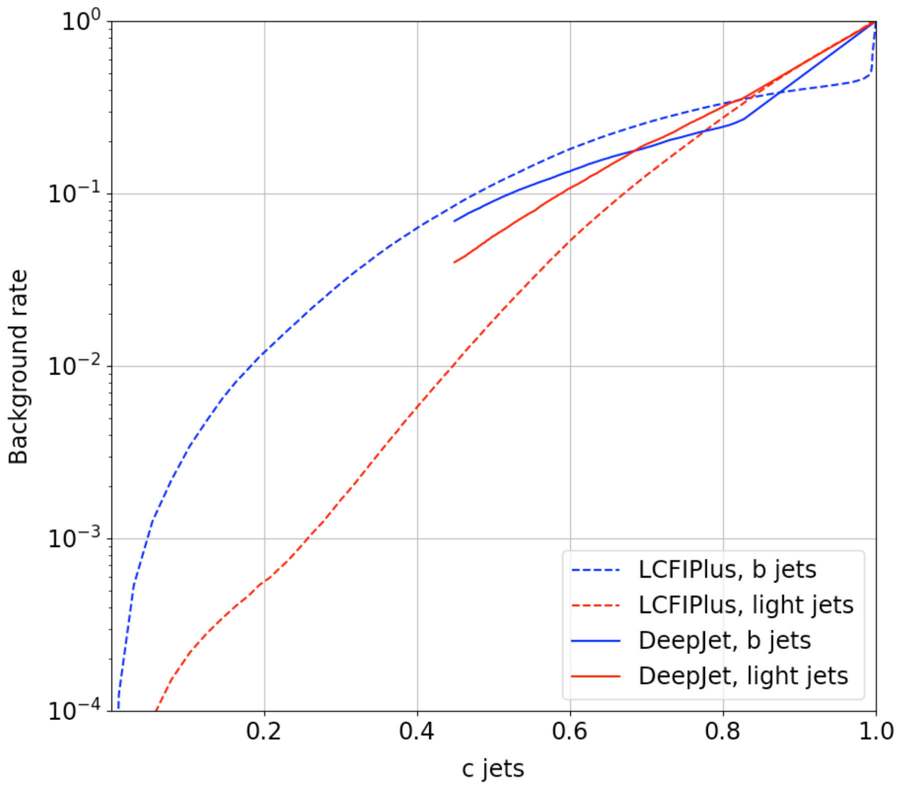


- identification efficiencies ~80% for b jets & light jets
- c jet identification quite low (66%)
- especially separation between c jets and light jets should be improved

ROC curves - comparison to LCFIPlus

validation data





- slightly better performance for b jet identification vs. c jet background
- better performance for c jet efficiencies vs. b jet background below ~90% c jet identification efficiency
- worse performance in b jet / c jet identification vs. light jets (especially for c jets)

Summary, next steps & outlook

- application of CMS DeepJet tagger to ILD
- slightly improved separation between c jets and b jets, worse performance for b jet / c jet identification vs. light jets

Next steps:

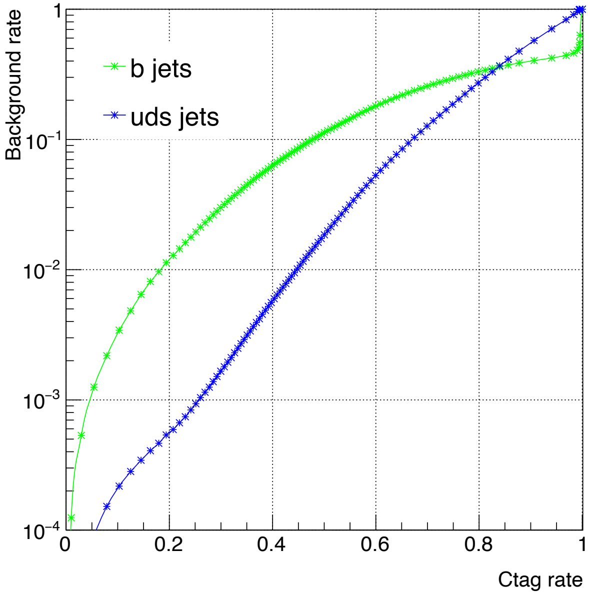
- comparison to LCFIPlus variables
- over-sampling, sample weights in loss function?
- optimization of hyperparameters

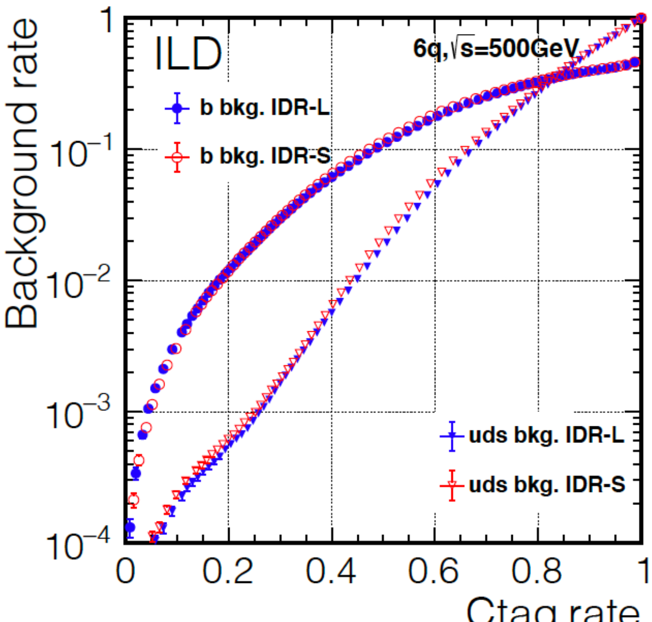
Outlook:

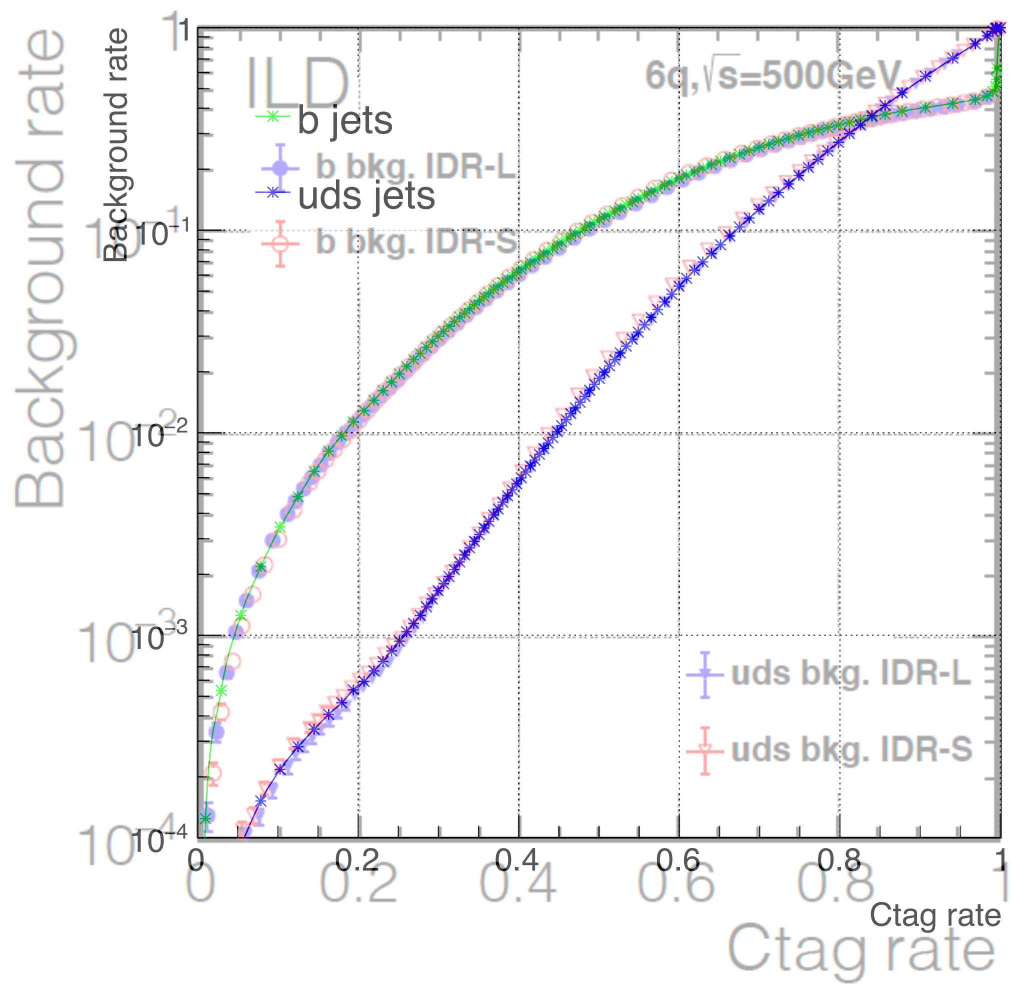
- study other architectures (e.g. particle net)
- integrate into iLCSoft

Backup

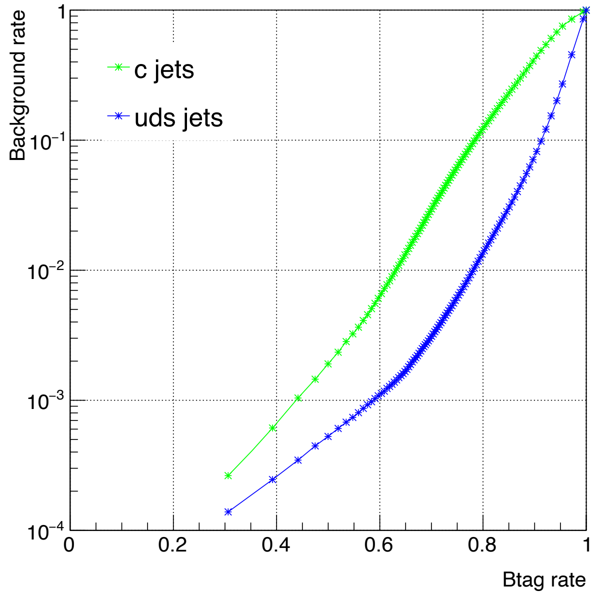
Performance LCFIPlus

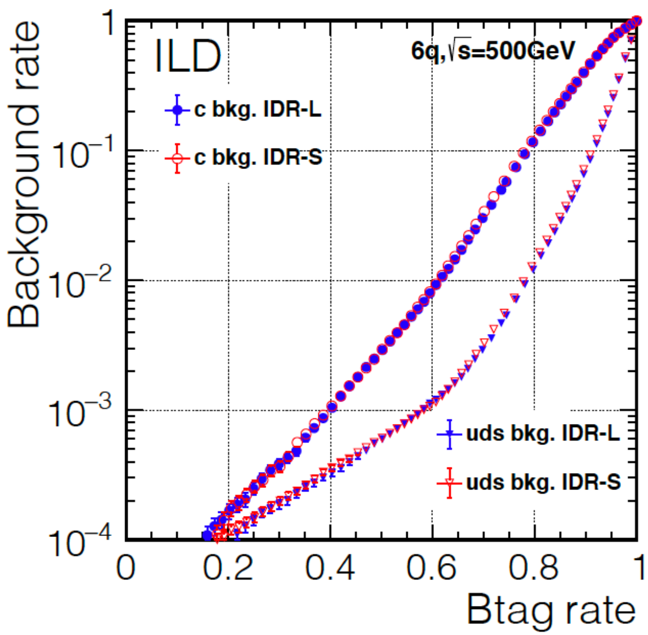


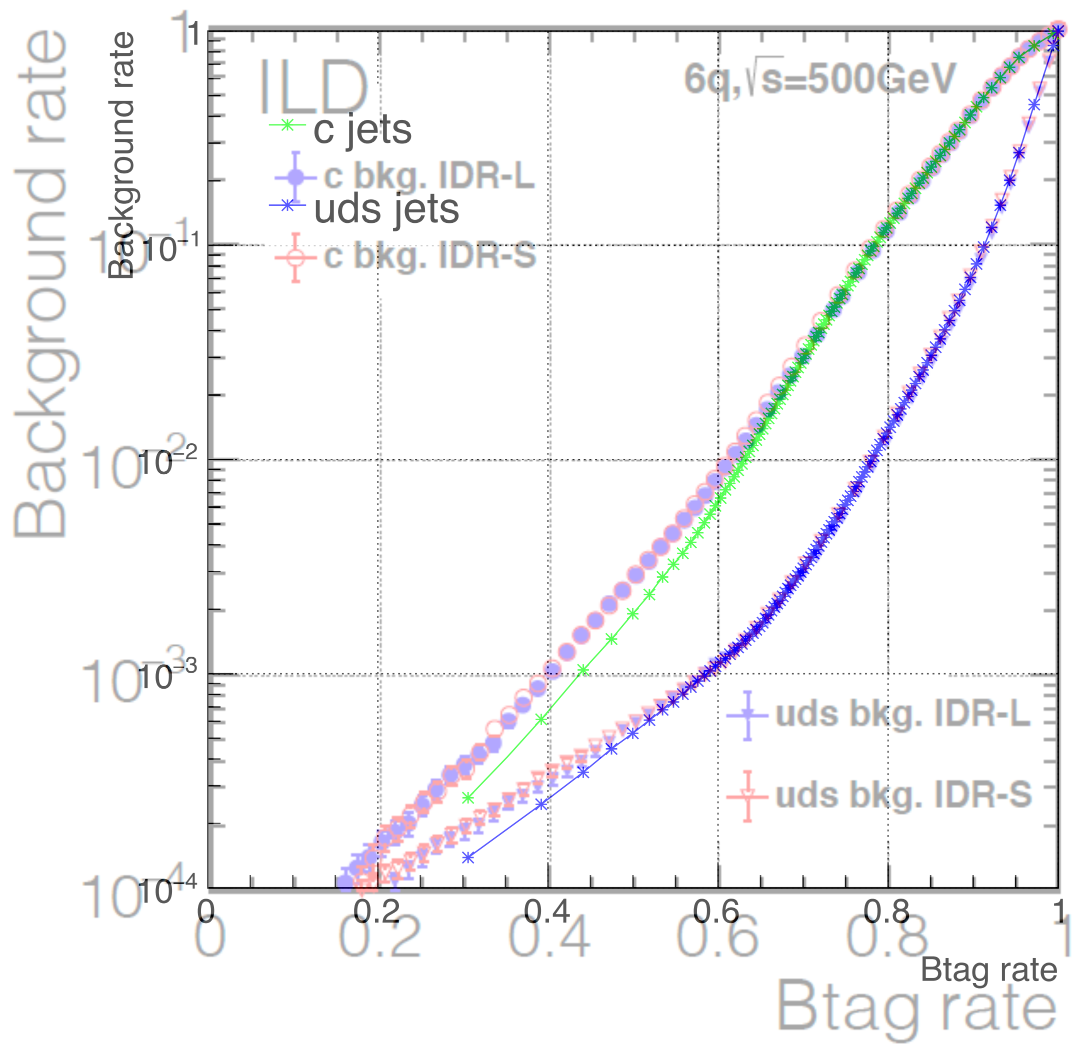




Performance LCFIPlus







CMS DeepJet: input variables

A.1 List of global variables

- Jet p_t
- Jet η
- The number of charged particle flow candidates in the jet
- The number of neutral particle flow candidates in the jet
- The number of secondary vertices in the jet
- The number of primary vertices in the event

A.3 List of neutral candidate variables

- Fraction of the jet momentum carried by the neutral candidate
- ΔR between the jet axis and the neutral candidate
- A integer flag indicating whether the neutral candidate is a photon.
- Fraction of the neutral candidate energy deposited in the hadronic calorimeter.
- ΔR between the neutral candidate and the closest secondary vertex
- The neutral candidates PUPPI weight

CMS DeepJet: input variables

A.4 List of secondary vertex variables

- Secondary vertex p_t
- ΔR between the jet axis and the secondary vertex
- Secondary vertex mass
- Number of tracks in the secondary vertex
- χ^2 of the secondary vertex fit
- Reduced χ^2 of the secondary vertex fit
- The secondary vertex 2D impact parameter value
- The secondary vertex 2D impact parameter significance
- The secondary vertex 3D impact parameter value
- The secondary vertex 3D impact parameter significance
- Cosine of the angle between the secondary vertex flight direction and the direction of the secondary vertex momentum.
- Ratio of the secondary vertex energy to the jet energy

CMS DeepJet: input variables

A.2 List of charged candidate variables

- Charged track η relative to the jet axis
- Charged track p_t relative to the jet axis
- Dot product of the jet and track momentum
- Dot product of the jet and track momentum divided by the magnitude of the jet momentum
- ΔR between the jet axis and the track
- The track 2D impact parameter value
- The track 2D impact parameter significance
- The track 3D impact parameter value
- The track 3D impact parameter significance
- The track distance to the jet axis
- Fraction of the jet momentum carried by the track.
- ΔR between the track and the closest secondary vertex
- An integer flag that indicate whether the track was used in the primary vertex fit.
- The charged candidates PUPPI weight
- χ^2 of the charged track fit.
- A integer flag which indicate the quality of the fitted track, based on number of detector hits used for the reconstruction as well as the overall χ^2 of the charged track fit.

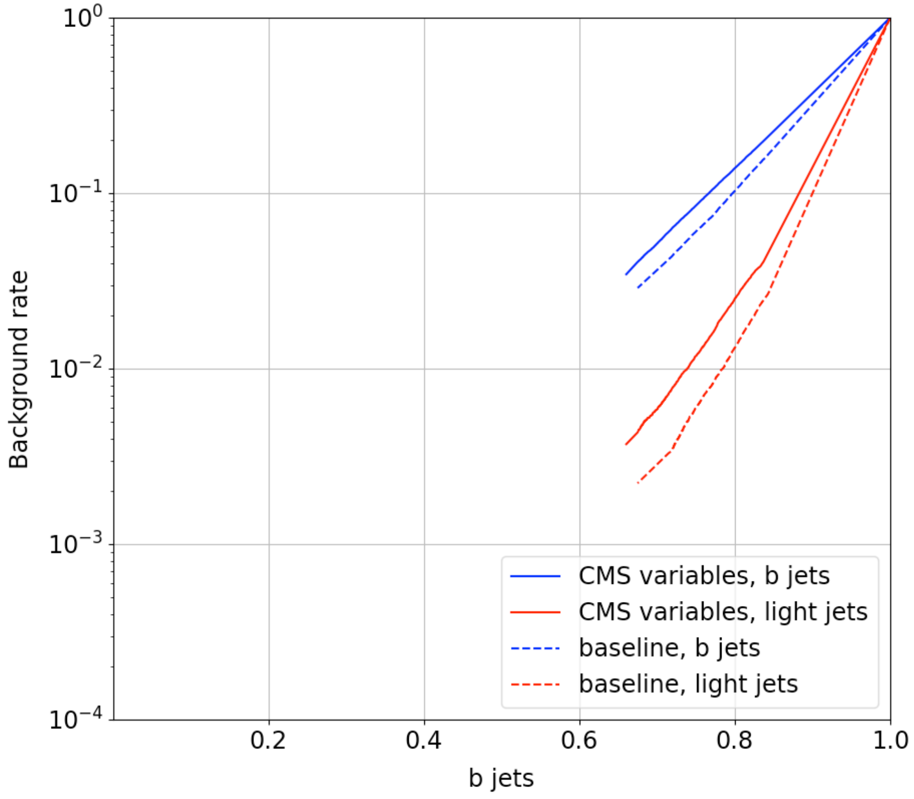
Variables used by LCFIPlus

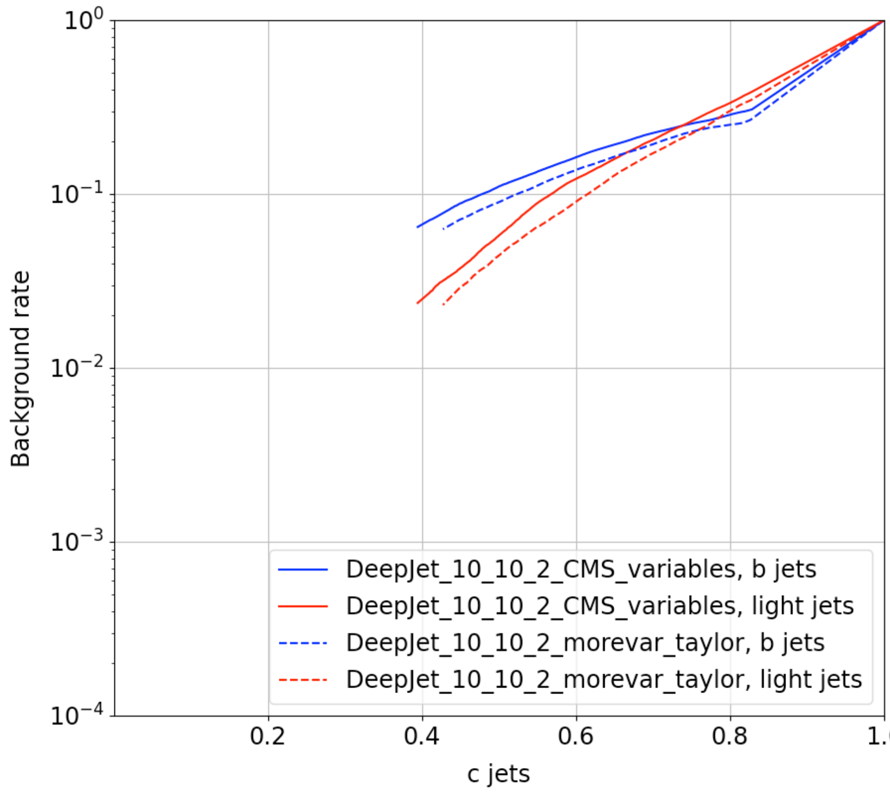
Name	Description	Normalization factor	Used by category
trk1d0sig	d0 significance of track with highest d0 significance	1	A, B, C, D
trk2d0sig	d0 significance of track with second highest d0 significance	1	A, B, C, D
trk1z0sig	z0 significance of track with highest d0 significance	1	A, B, C, D
trk2z0sig	z0 significance of track with second highest d0 significance	1	A, B, C, D
trk1pt	transverse momentum of track with highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
trk2pt	transverse momentum of track with second highest d0 significance	$1/E_{\mathrm{jet}}$	A, B, C, D
jprobr	joint probability in the r-phi plane using all tracks	1	A, B, C, D
jprobr5sigma	joint probability in the r-phi plane using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
jprobz	joint probability in the z projection using all tracks	1	A, B, C, D
jprobz5sigma	joint probability in the z projection using all tracks having impact parameter significance exceeding 5 sigma	1	A, B, C, D
d0bprob	product of b-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0cprob	product of c-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
d0qprob	product of q-quark probabilities of d0 values for all tracks, using b/c/q d0 distributions	1	A, B, C, D
z0bprob	product of b-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0cprob	product of c-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
z0qprob	product of q-quark probabilities of z0 values for all tracks, using b/c/q z0 distributions	1	A, B, C, D
nmuon	number of identified muons	1	A, B, C, D
nelectron	number of identified electrons	1	A, B, C, D
trkmass	mass of all tracks exceeding 5 sigma significance in d0/z0 values	1	A, B, C, D

Variables used by LCFIPlus

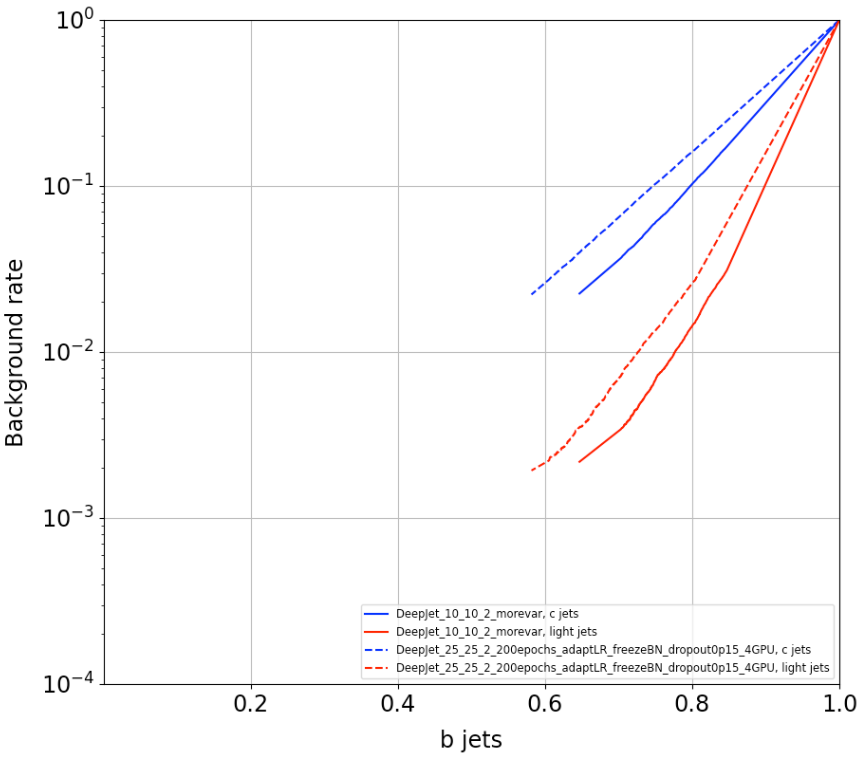
Name	Description	Normalization factor	Used by category
1vtxprob	vertex probability with all tracks associated in vertices combined	1	B, C, D
vtxlen1	decay length of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxlen2	decay length of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxlen12	distance between the first and second vertex (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig1	decay length significance of the first vertex in the jet (zero if no vertex is found)	$1/E_{\text{jet}}$	B, C, D
vtxsig2	decay length significance of the second vertex in the jet (zero if number of vertex is less than two)	$1/E_{\text{jet}}$	D
vtxsig12	vtxlen12 divided by its error as computed from the sum of the covariance matrix of the first and second vertices, projected along the line connecting the two vertices	$1/E_{\text{jet}}$	D
vtxdirang1	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the first vertex	E_{jet}	B, C, D
vtxdirang2	the angle between the momentum (computed as a vector sum of track momenta) and the displacement of the second vertex	E_{jet}	D
vtxmult1	number of tracks included in the first vertex (zero if no vertex is found)	1	B, C, D
vtxmult2	number of tracks included in the second vertex (zero if number of vertex is less than two)	1	D
vtxmult	number of tracks which are used to form secondary vertices (summed for all vertices)	1	D
vtxmom1	magnitude of the vector sum of the momenta of all tracks combined into the first vertex	$1/E_{\text{jet}}$	B, C, D
vtxmom2	magnitude of the vector sum of the momenta of all tracks combined into the second vertex	$1/E_{\text{jet}}$	D
vtxmass1	mass of the first vertex computed from the sum of track four-momenta	1	B, C, D
vtxmass2	mass of the second vertex computed from the sum of track four-momenta	1	D
vtxmass	vertex mass as computed from the sum of four momenta of all tracks forming secondary vertices	1	B, C, D
vtxmasspc	mass of the vertex with minimum pt correction allowed by the error matrices of the primary and secondary vertices	1	B, C, D
vtxprob	vertex probability; for multiple vertices, the probability P is computed as 1-P = (1-P1)(1-P2)...(1-PN)	1	B, C, D

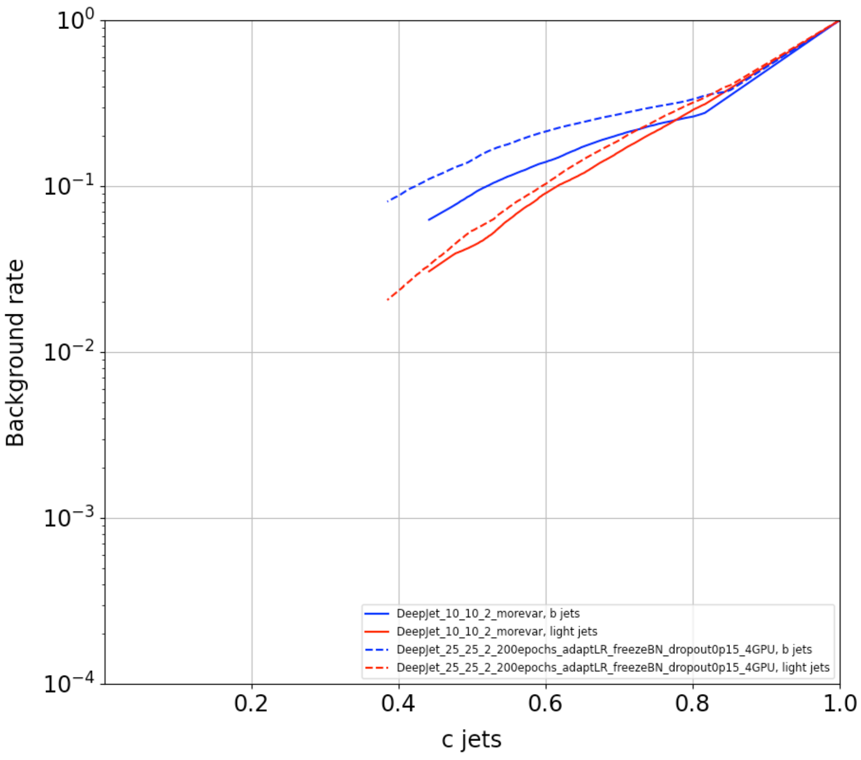
Training with CMS variables





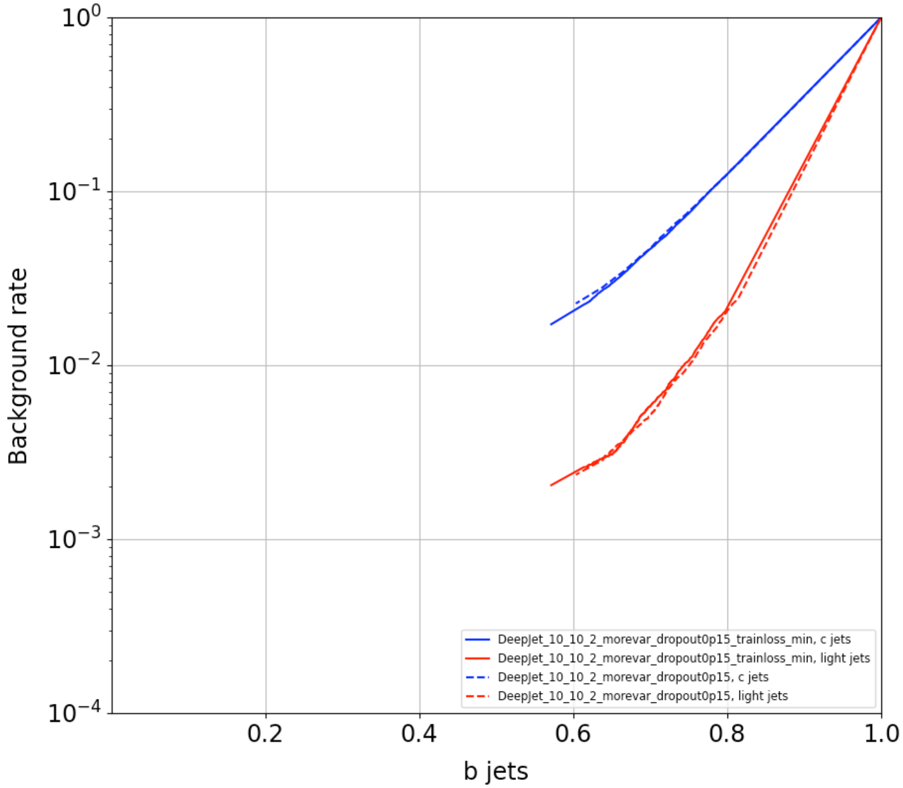
Results: training with more variables vs. more constituents

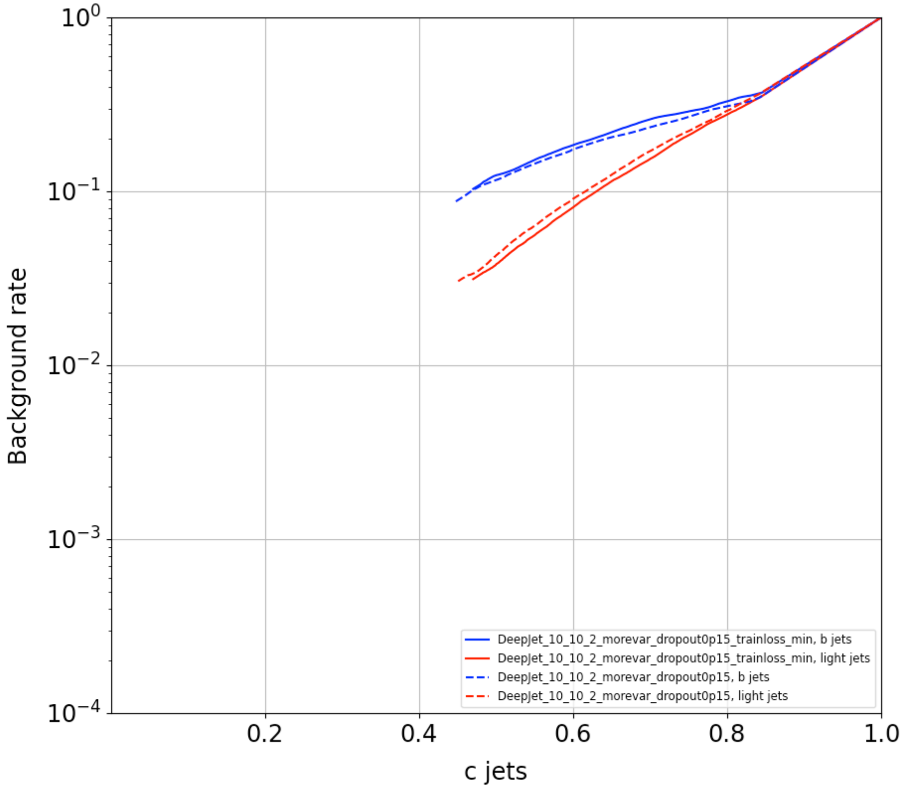


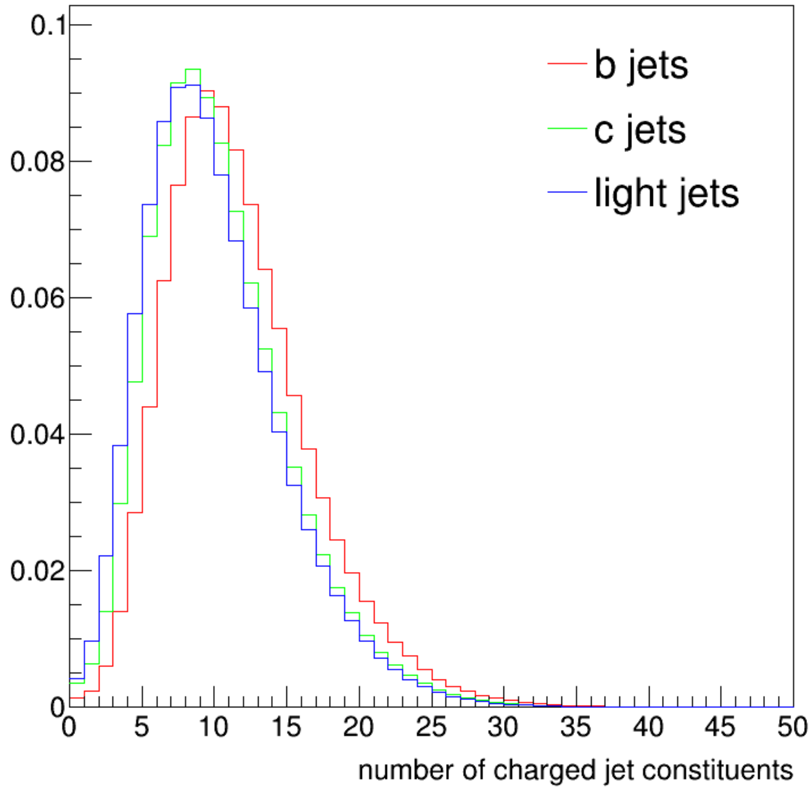


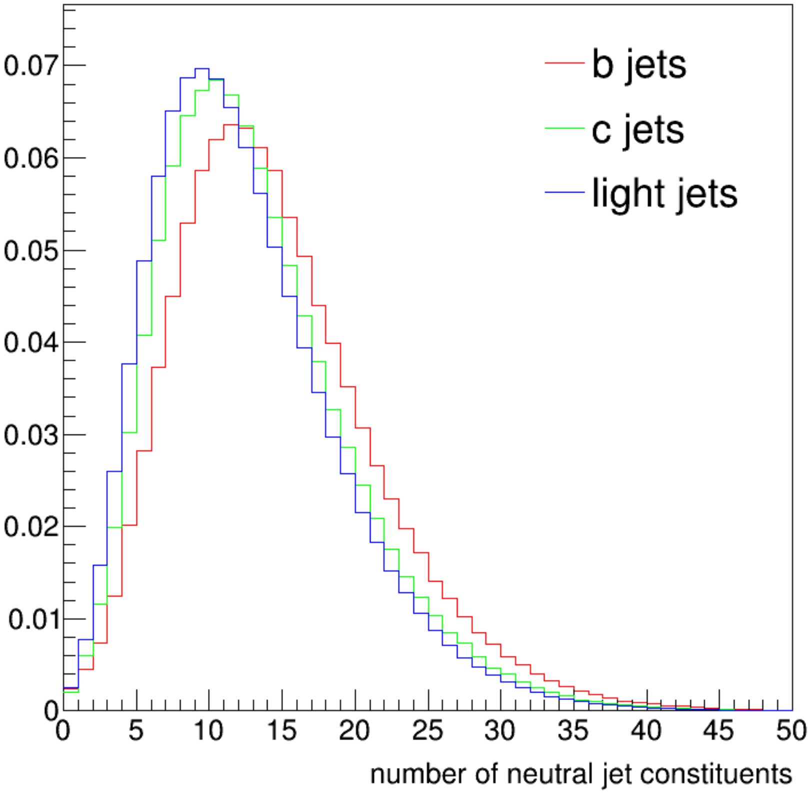
Comparison min training loss vs. max validation accuracy

- min training loss: epoch 297
- max validation accuracy: epoch 294






0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
45
50
number of charged jet constituents
b jets
c jets
light jets


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
35
40
45
50
number of neutral jet constituents
b jets
c jets
light jets