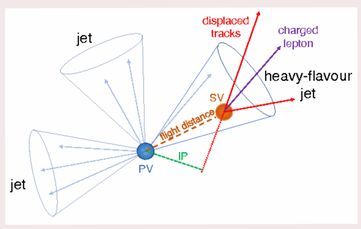
MACHINE LEARNING TECHNIQUES FOR HEAVY FLAVOUR IDENTIFICATION

Chazin B. On behalf of the CMS Collaboration

Heavy-flavour jets identification algorithms in CMS

- **The most performant algorithms for heavy-flavour jet identification are based on multivariate combination of the B or C hadrons properties in the jet:**



Long lifetime:
Displaced decays result in tracks with large impact parameter (IP) and secondary vertex (SV)

Large mass and hard fragmentation:
Decay products have larger transverse momentum (p_T) relative to the jet axis than the other jet constituents.

Semi-leptonic decays:
Presence of soft muons or electrons in the jet

Algorithms developments in Run2

- **The Run2 algorithms use techniques as Boosted Decision Tree (BDT) or Artificial Neural Network (ANN) to combine larger number of variables than Run1 algorithms:**

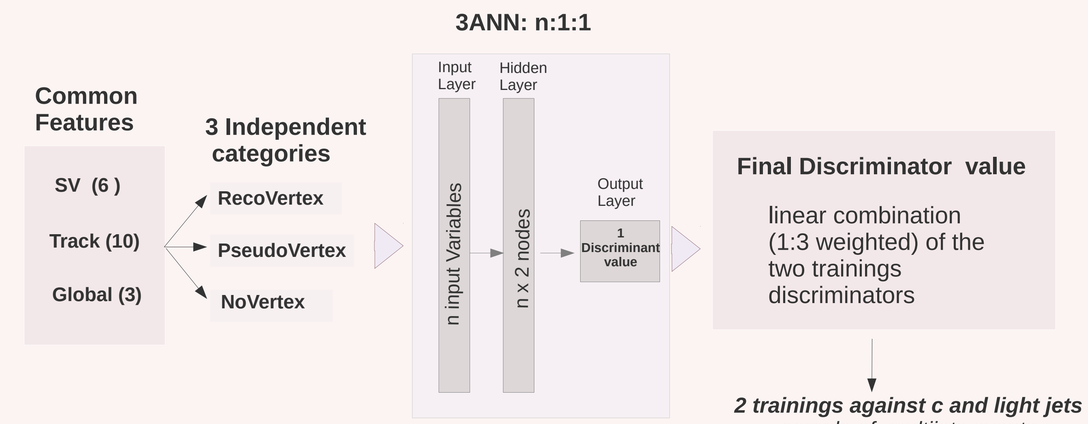
CSVv2: based on CSV algorithm (Run1), combines larger number of variables using the secondary vertex and track-based lifetime information in a neural network

cMVAv2: based on cMVA algorithm (Run1), uses as input 6 b-jet discriminators outputs of CSVv2 and other simpler taggers in a gradient boosting classifier (GBC) as BDT

DeepCSV, DeepFlavour: use of deep neural networks which allow for multi-classification providing an output probability for each jet flavour hypothesis

CSVv2

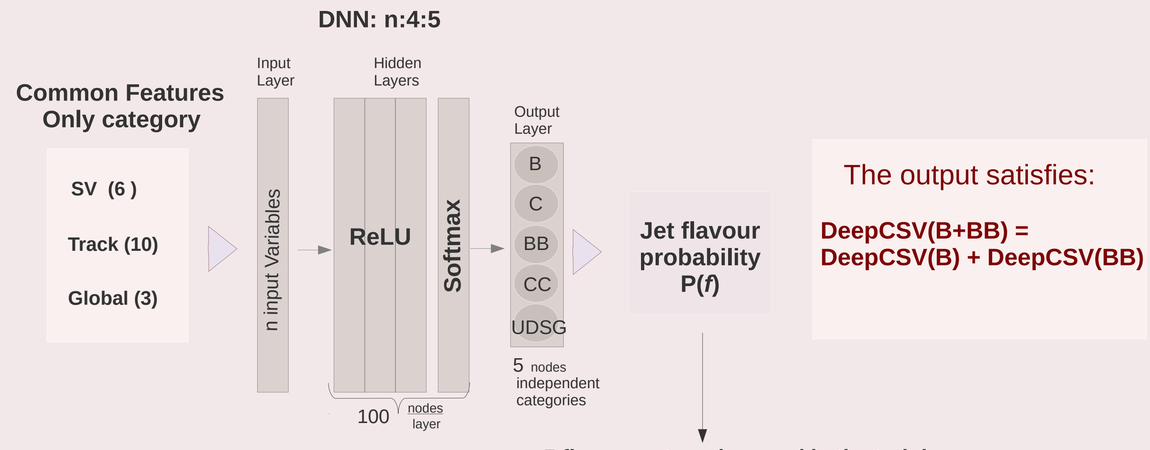
- 3 Feed-forward multilayer perceptron (n:1:1) depending on the vertex information and combined with a likelihood method



Activation functions: Sigmond
Learning method: Backpropagation
Loss function: Cross entropy

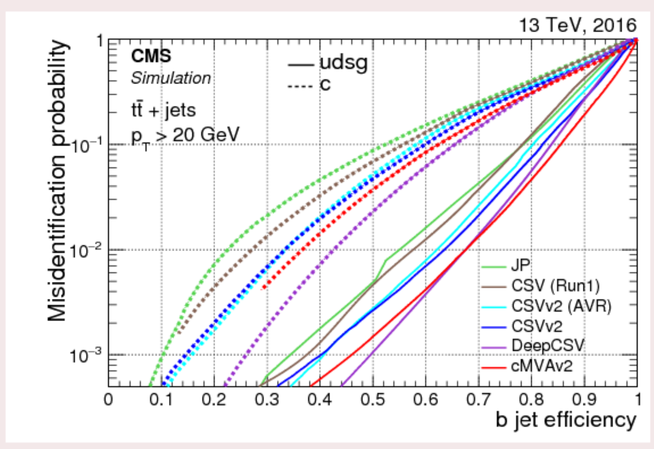
DeepCSV

- An extension of CSVv2 which inherits the common features but uses more charged particle tracks (up to 6) combined in a deep feed-forward neural network (n:4:5)



Activation functions: ReLU, Softmax
Learning method: Backpropagation
Loss Function: Categorical cross entropy

Performance of b-jet tagging algorithms in 2016

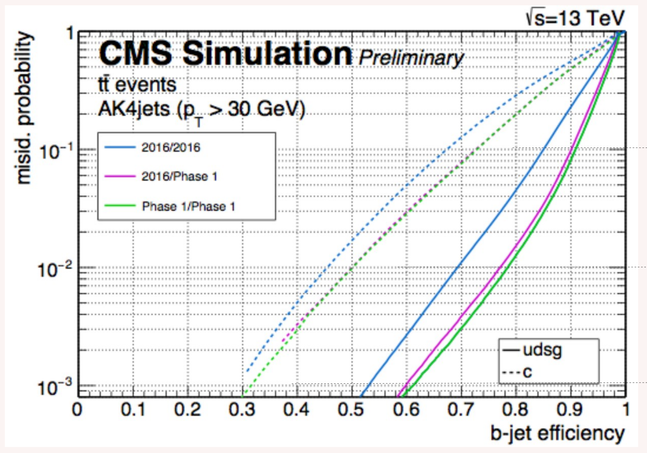


JINST 13 (2018) P05011

- The DeepCSV (P(*B*) +P(*BB*)) discrimination against c and light jets outperforms all other algorithms for b-tagging efficiencies below 70%
- cMVAv2 tagger performs better against light jets for for b-tagging efficiencies below 70%
- Both taggers improves the CSVv2 performance by ~ 4% in the medium (M) working point

WP w.r.t. mis-Id of light jets

Algorithm performances in 2017



CMS DP-2017/013

- The CMS Phase 1 upgrade included a new pixel detector with an additional layer, closer to the beam spot

CMS-TDR-011

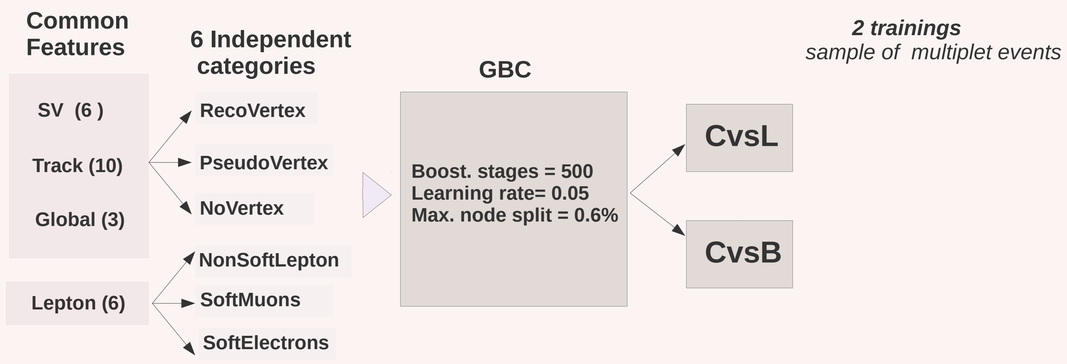
10% Loose (L)
1% Medium (M)
0.1% Tight (T)

- Comparison of DeepCSV performance with 2016 detector, Phase 1 detector and 2016 training, and with Phase 1 detector and new dedicated training

WP w.r.t. mis-Id of light jets

c-jet identification algorithm

- A **GBC** is used for two trainings to discriminate c jets against light (CvsL) and b (CvsB) jets. Similar inputs as defined in CSVv2 but adding soft lepton info (up to 2 tracks or leptons)



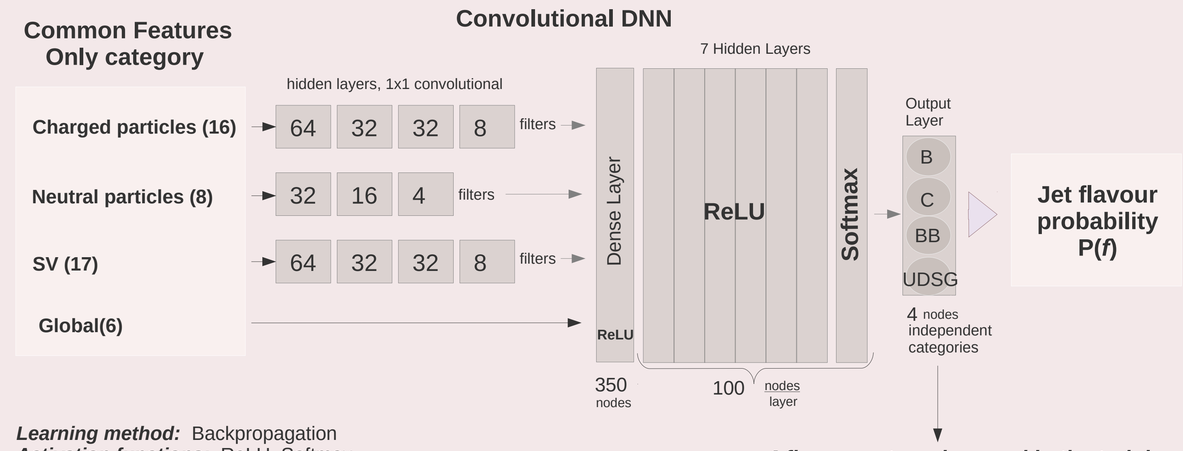
- c taggers can be built from **DeepCSV outputs**:

$$\text{DeepCSVCvsB} = \frac{P(c) + P(cc)}{1 - P(udsg)}$$

$$\text{DeepCSVCvsL} = \frac{P(c) + P(cc)}{1 - (P(b) + P(bb))}$$

DeepFlavour

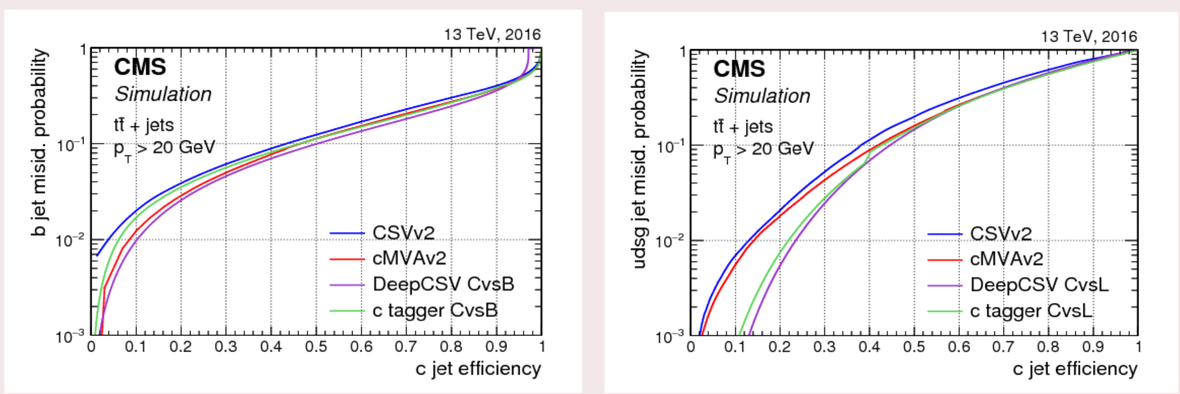
- A deep-neural-network algorithm is based on different properties of charged (≤ 25) and neutral (≤ 4) particle jet constituents, SV (≤ 4) and global jet variables



Learning method: Backpropagation
Activation functions: ReLU, Softmax
Loss Function: Categorical cross entropy

Performance of c-jet tagging algorithms in 2016

- DeepCSV is already outperforming the dedicated c-tagger



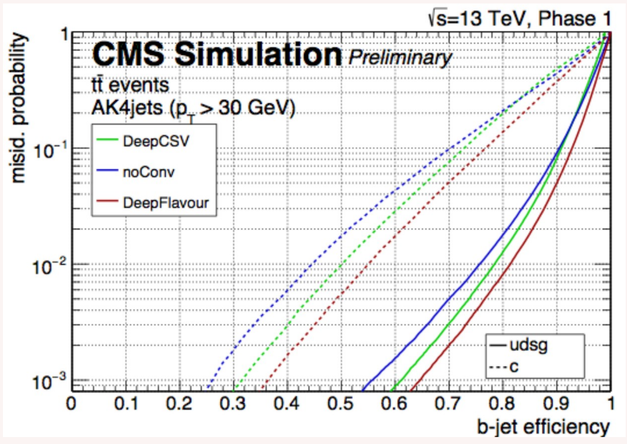
JINST 13 (2018) P05011

Performance of Deep-tagging algorithms

- DeepFlavour tagger gives a 4% absolute improvement in b-tag efficiency for a mistag rate of 0.1% against DeepCSV
  - Extended to gluon vs quark discrimination (DeepJet):

CMS DP-2017/027

- **NoConv**: DeepFlavour algorithm but trained without the convolutional layers - only for comparison.



CMS DP-2017/013