Identifying Charm Quark Jets with the LHCb Experiment

Samuel Alibocus

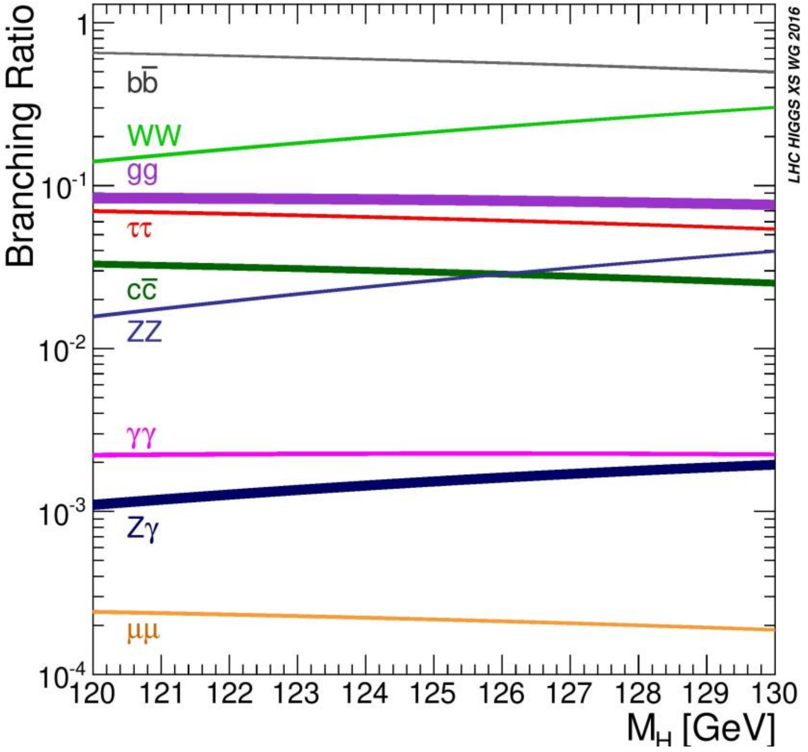
UNIVERSITY OF LIVERPOOL

Motivation – $H \rightarrow c\bar{c}$

- Higgs field couples to massive particles.

- The larger the mass the stronger the coupling.

$$V(\phi) = m^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$$



- $H \rightarrow b\bar{b}$ makes up 58% of the total branching fraction and has been measured to 5.6σ.

- Branching fraction for decays to γγ, ZZ, WW and ττ have also been measured.

- This leaves $H \rightarrow c\bar{c}$.

- It is important to check theoretical predictions by experimentally measuring branching ratios.

- A disagreement could indicate new physics.

UNIVERSITY OF LIVERPOOL

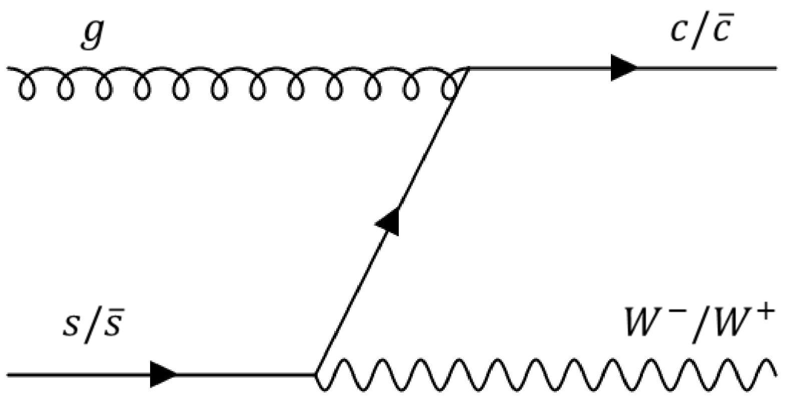
Motivation – Strange Quark Contribution to the Proton

- The proton consists of three valence quarks that carry most of the momentum and a sea of lower momentum quark-antiquark pairs that carry the rest of the momentum.

- The flavour symmetry group SU(3) suggests that the three light sea quarks (u, d and s) are equally distributed.

- $s/\bar{s}$ sea quark content can be studied by observing decays where they fuse with gluons.

- This happens through two processes; $sg \longrightarrow W^-c$ and $\bar{s}g \longrightarrow W^+\bar{c}$.



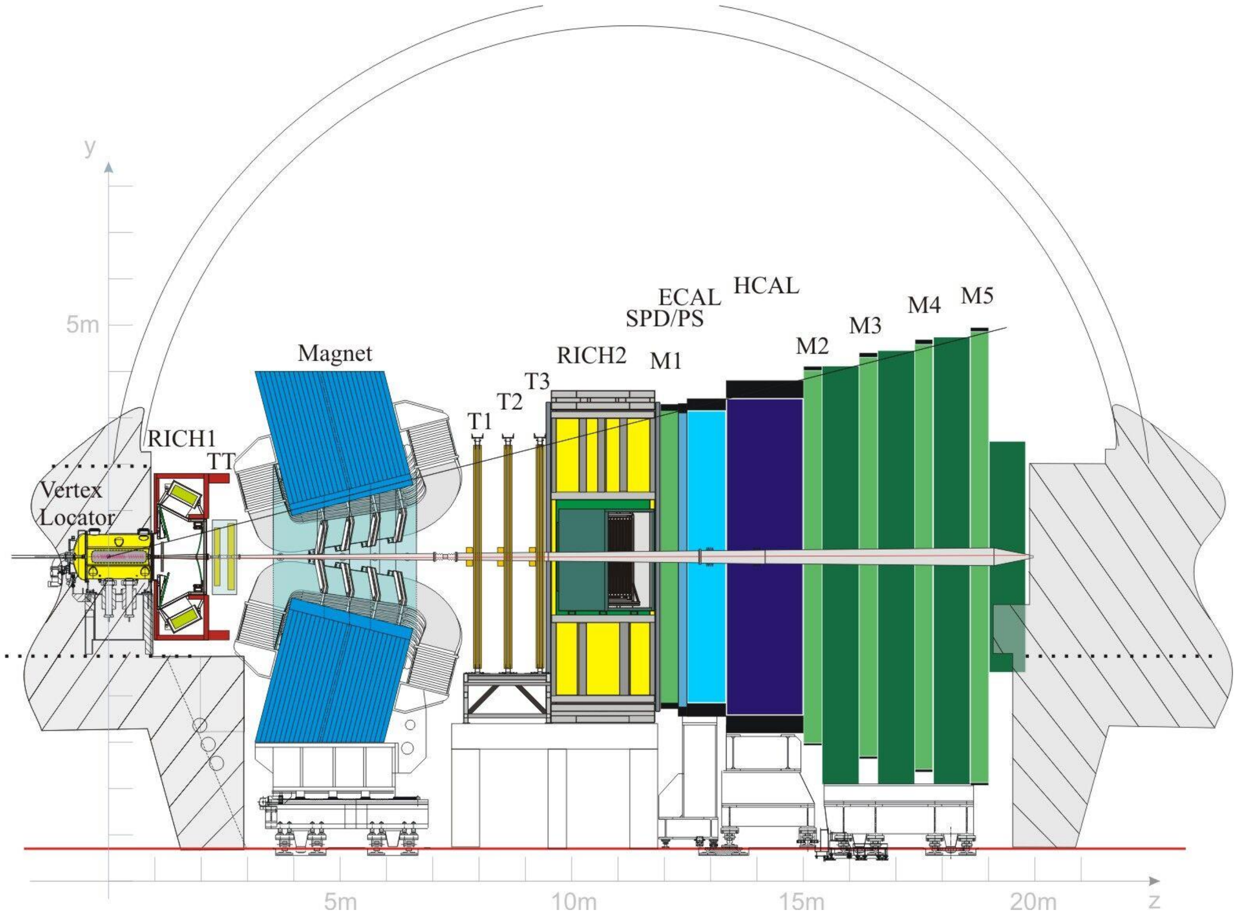
UNIVERSITY OF LIVERPOOL

LHCb Detector

- Identifying the secondary vertex of a jet is crucial to understanding heavy flavour jets (b and c).

- Heavy flavour quarks decay initially to a heavy flavour meson (b or d meson) which then decay into a jet.

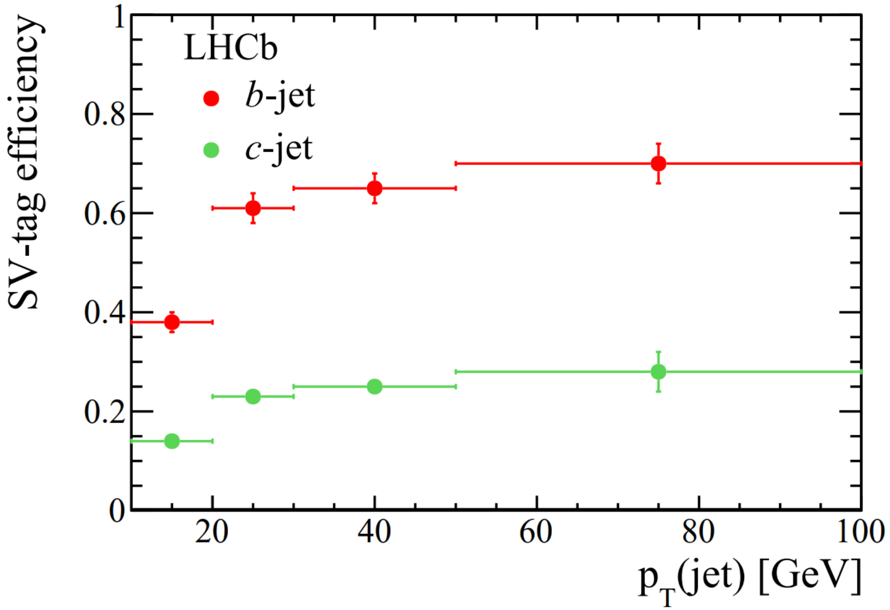
- Due to the nature of the LHCb dectector and the VELO, these secondary vertices can be accurately measured.

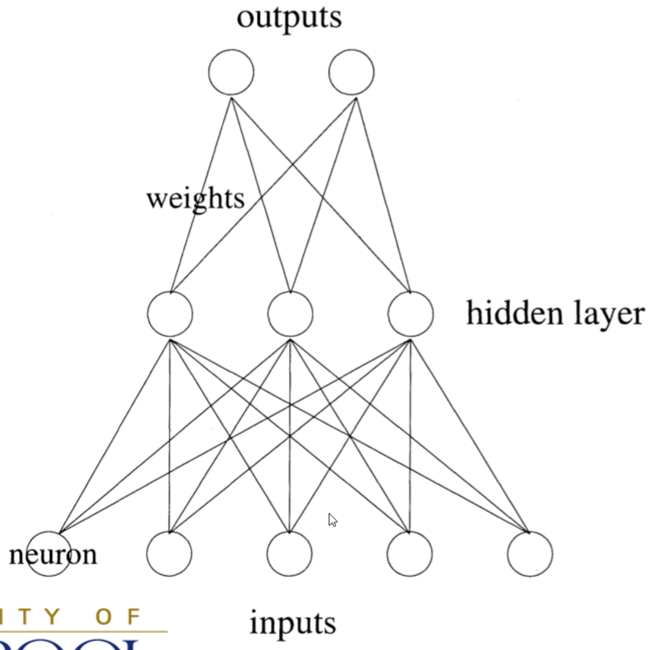


UNIVERSITY OF LIVERPOOL

Machine Learning Approach

- Jet tagging is the identification of jets originating from quarks and gluons.
- Using a secondary vertex tagger the mean efficiency for a jet with $p_T > 20$ GeV and $2.2 < \eta < 4.2$ is roughly 65% for b-jets and 25% for c-jets has previously been obtained.



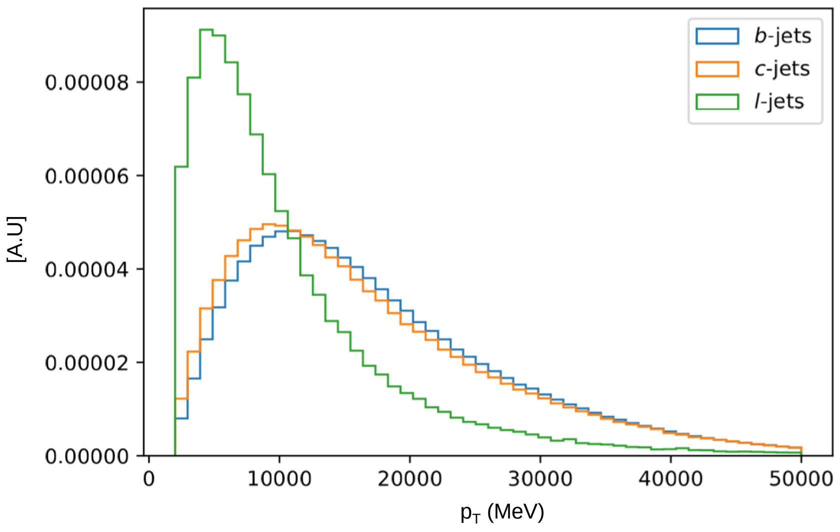


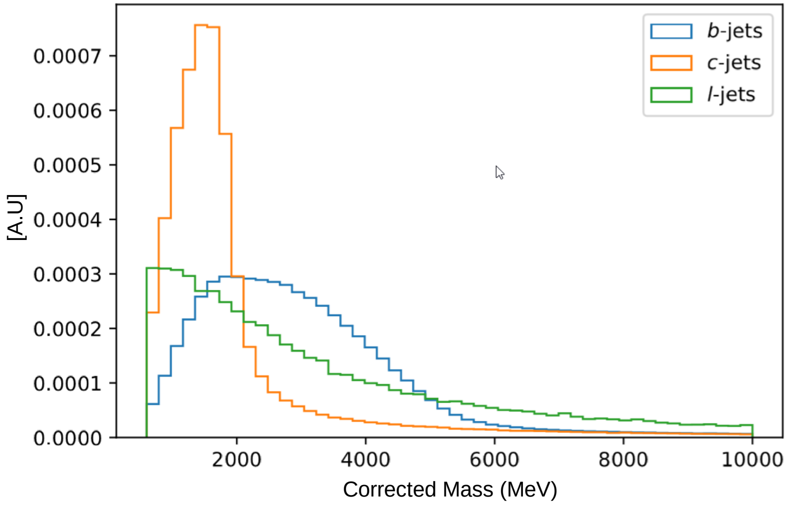
- This project aimed at using an artificial neural network so multiple input variables could be used in a bid to increase the efficiency of the c-tagger.
- For this the PyTorch machine learning library was used.
- The neural network was run using simulated data from the PYTHIA simulation.

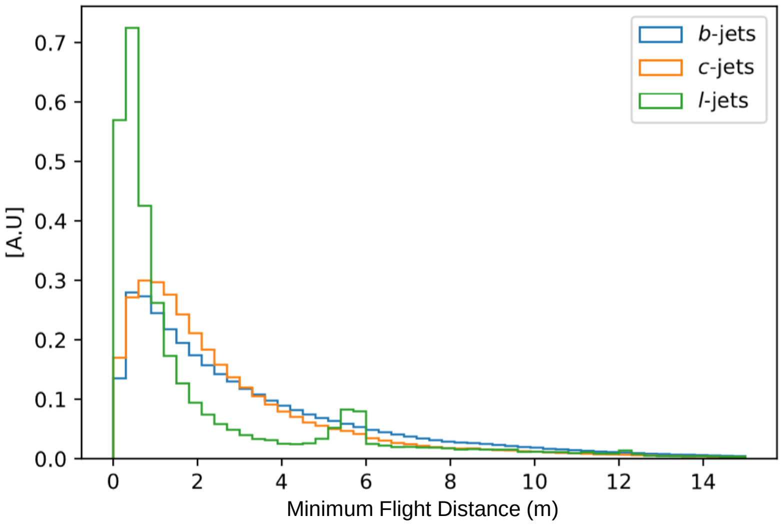
UNIVERSITY OF LIVERPOOL

Tagger Performance – Input Variables

- Input variables were selected to feed into the neural network.
- These were selected based on how well they separated the different jets.
- Only variables with a secondary vertex cut were used.







UNIVERSITY OF LIVERPOOL

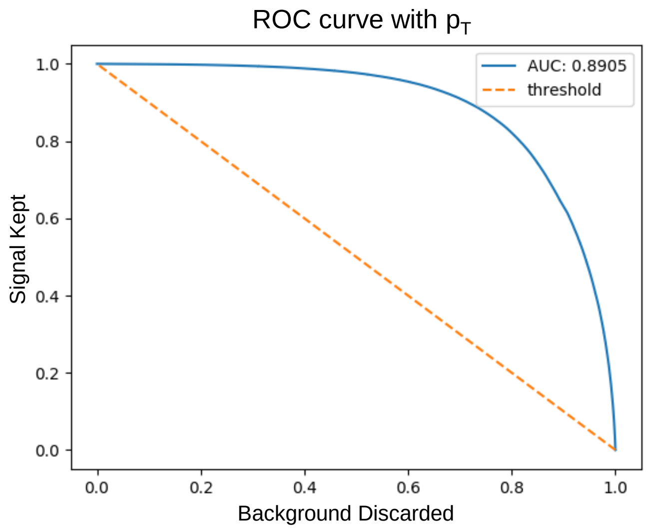
Tagger Performance – Assessing impact of p_T

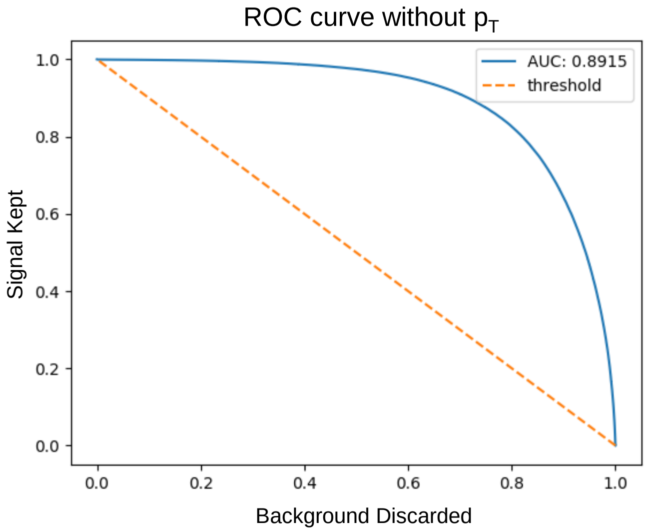
- p_T is not a variable that the neural network should rely on as there could be l-jets with particularly high momenta and this is not a property characteristic of heavy flavour jets.

- The neural network was run both with and without the p_T variable included and the ROC curves were compared.

- The ROC curve shows the signal kept against background discarded.

- It can be seen that removing the p_T variable had little effect on the plot.

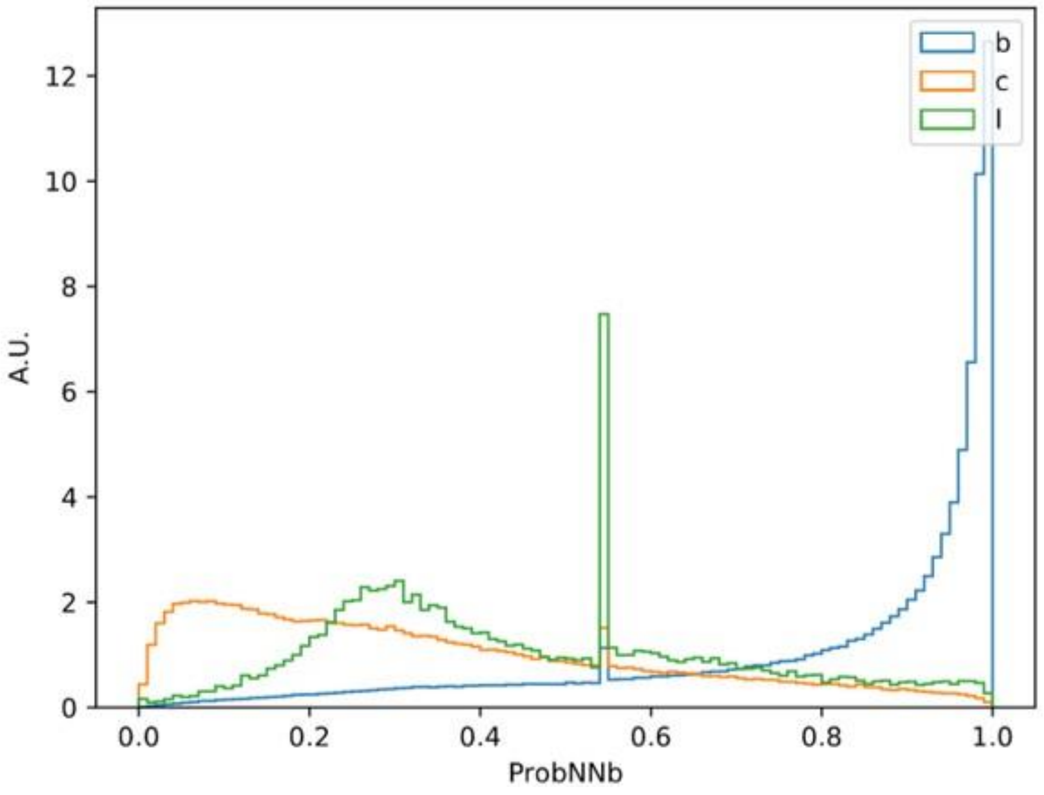




UNIVERSITY OF LIVERPOOL

Tagger Performance – Neural Network Distributions

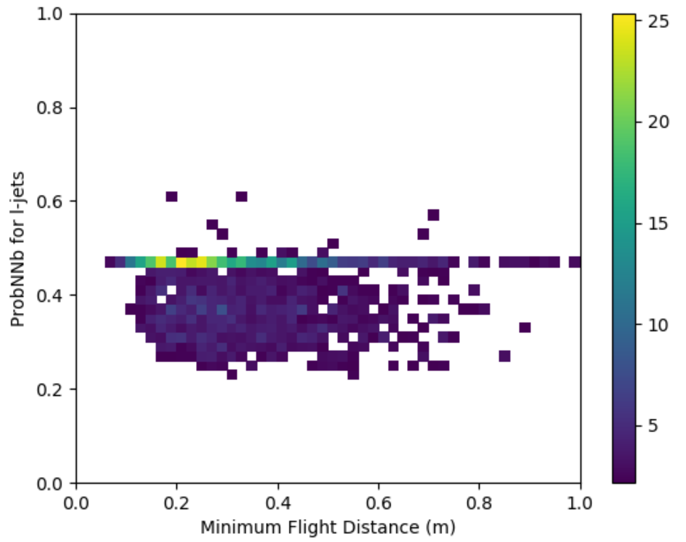
- Once the neural network was run, plots were made of the probability for each type of jet being tagged.
- ProbNNc is just a mirror image of ProbNNb.
- It can be seen that the neural net is doing well at tagging b-jets and also has a high ProbNNc for tagging c-jets.
- Spike in ProbNNb for l-jets means there is something causing l-jets to look like b-jets.

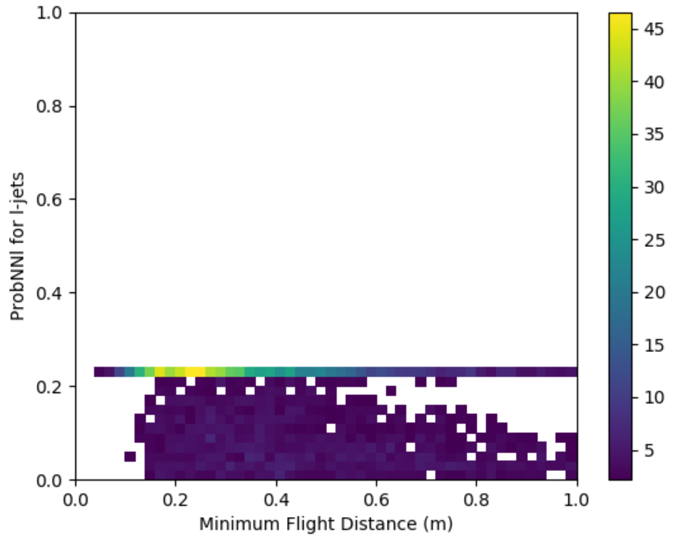


UNIVERSITY OF LIVERPOOL

Tagger Performance - Investigating ProbNNb for l-jets

- Spike in minimum flight distance for l-jets suggested that this could be the cause for the spike in ProbNNb for l-jets.
- 2D histograms for ProbNNb/ProbNNl for l-jets against the minimum flight distance.
- It can clearly be seen that most of the data lies at the 0.2m mark in both of these plots.
- This means that there is something at 0.2m in the detector causing l-jets to look b-jets.
- This could be an aspect of the VELO or the simulated data itself.





UNIVERSITY OF LIVERPOOL

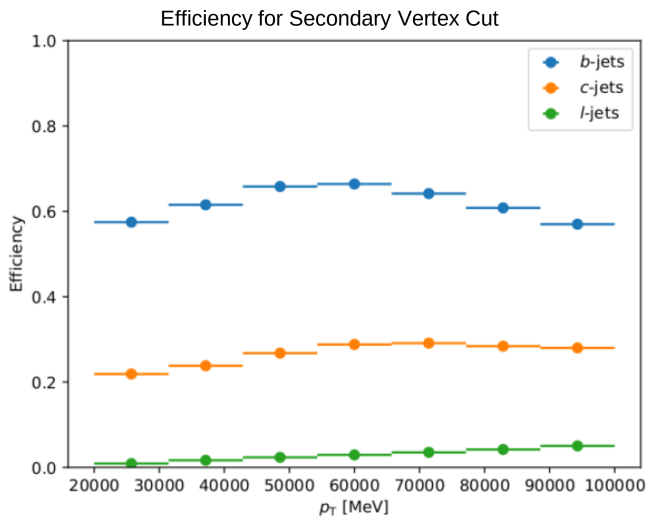
Tagger Performance – Tagger Efficiancy

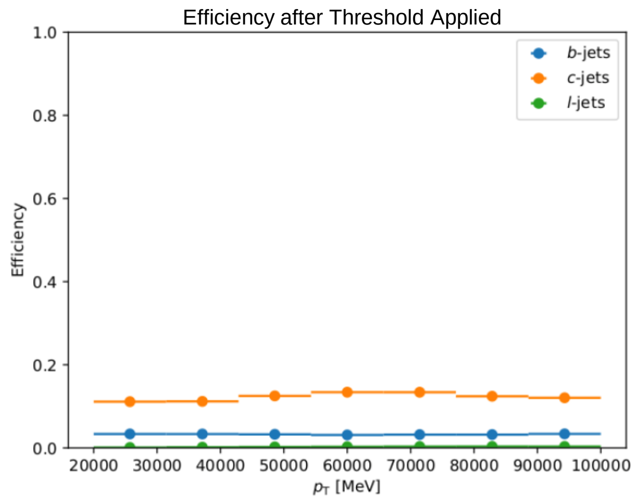
- Efficiency was calculated by diving the number of events successfully tagged as a particular flavour by the total number of events that were actually that flavour.

- This was first calculated just after the secondary vertex cut was applied.

- It was then calculated again after a threshold of ProbNNc > 0.6 and ProbNNl > 0.1 was applied.

- This produced a mean overall efficiency of 13% over a range of 25 GeV < p_T < 95 GeV.





UNIVERSITY OF LIVERPOOL

Tagger Performance – Improvements to Performance

- Compared to previous studies the mean overall efficiency achieved over a range of p_T for this project was just over 50% as efficient.

- This was achieved by only introducing the neural network to the input variables containing a secondary vertex tag.

- There are several ways this efficiency could be improved:
  1. Use input variables that do not contain a secondary vertex tag but still provide good separation.

  2. Adjust intrinsic neural network parameters to fine tune the neural network.

  3. Address the issue where l-jets are being mis-tagged.

Any Questions?