Boosted decision trees to identify hadronic decay products of tau leptons

Talia Rahall, Hrishi Shah
High Energy Physics Group, Faculty of Natural Sciences, Imperial College London

1. Lepton Flavour Universality Violation

The Standard Model dictates that the three flavours of leptons (e, μ, τ) couple to all types of electroweak bosons identically. The LHCb experiment measures rare decays involving B mesons to test for physics Beyond the Standard Model.

- In 2021 the LHCb published the ratio of branching fractions:

$$\frac{\mathcal{B}(B \to K^+\mu^+\mu^-)}{\mathcal{B}(B \to K^+e^+e^-)} = 0.846^{+0.044}_{-0.041}$$

  which gives a 3.1σ tension with the Standard Model's prediction of 1.00 ± 0.01 [1].
- Measurements of B meson decays to $\tau^+\tau^-$ are the next ingredients in the verification of Lepton Flavour Universality Violation at the LHCb.

2. Motivation

The τ lepton has a short lifetime and decays inside the beam pipe. It can decay into both leptons and hadrons, with respective branching fractions 35% and 65%. Identifying the hadronic decay products with good performance is essential.

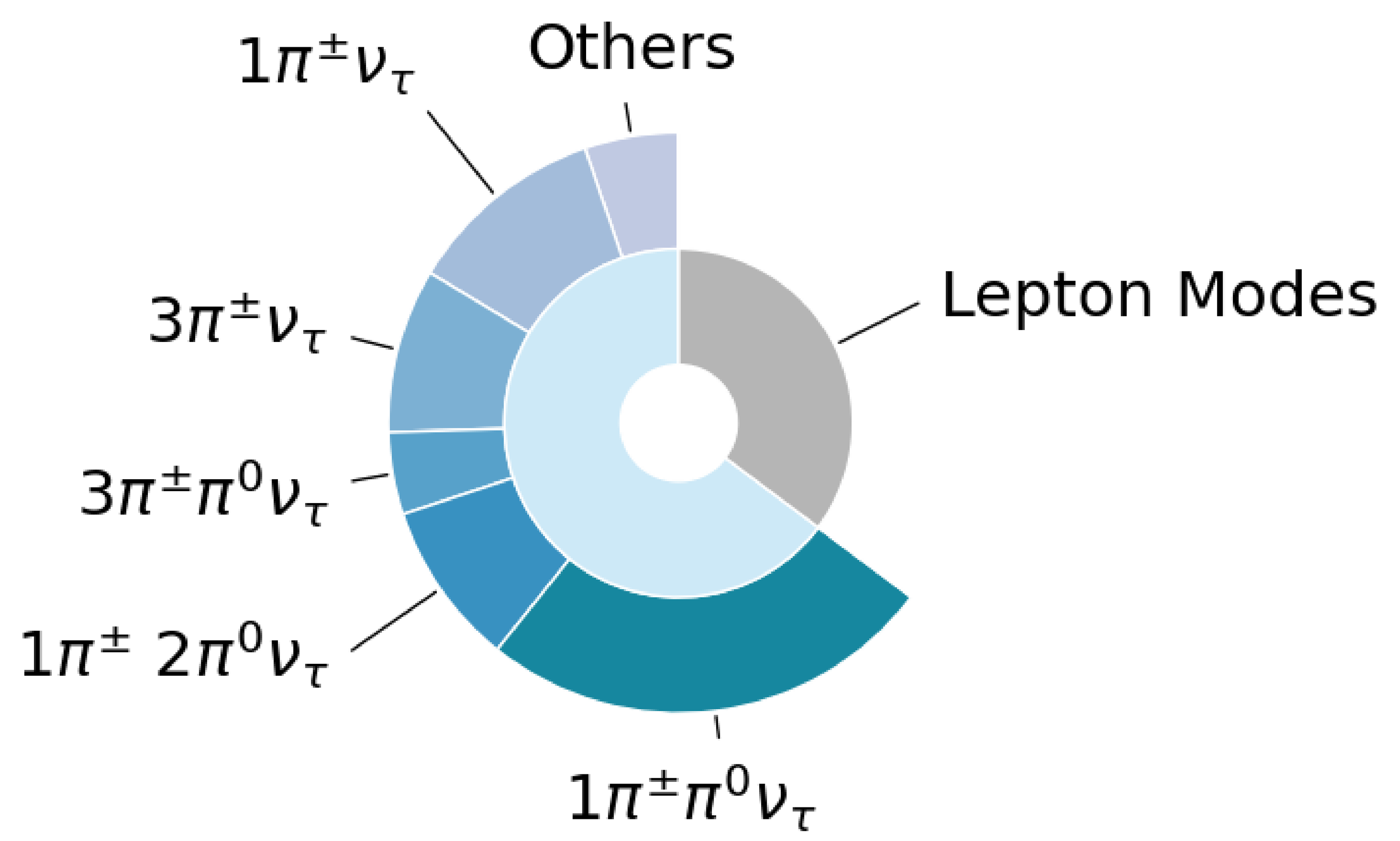

Figure 1: The branching fractions of the hadronic τ lepton decay modes.

The decay modes involving π^0 mesons account for 60% of hadronic modes. These modes are often overlooked in reconstruction methods since there is a high probability of misidentifying π^0 signatures in the detectors.

We are motivated to increase the performance of π^0 classification at the LHCb through multivariate analysis and machine learning.

3. Experimental Setup

The LHCb is a forward arm spectrometer best suited for the study of B mesons due to their high rapidity.

- The ECAL, HCAL and RICH detectors allow identification of decay products.

We use the LHCb software DaVinci to track simulated τ decays within these detectors.

[1] LHCb Collaboration et al.
Test of lepton universality in beauty-quark decays, Mar 2021.

4. Multivariate Analysis

The output from DaVinci is used to create a dataset containing feature variables of π^0 candidates, which belong to a background or signal class.

- We conduct feature engineering to find differentiating properties between the classes, for example the system invariant mass is seen to obey τ physics.

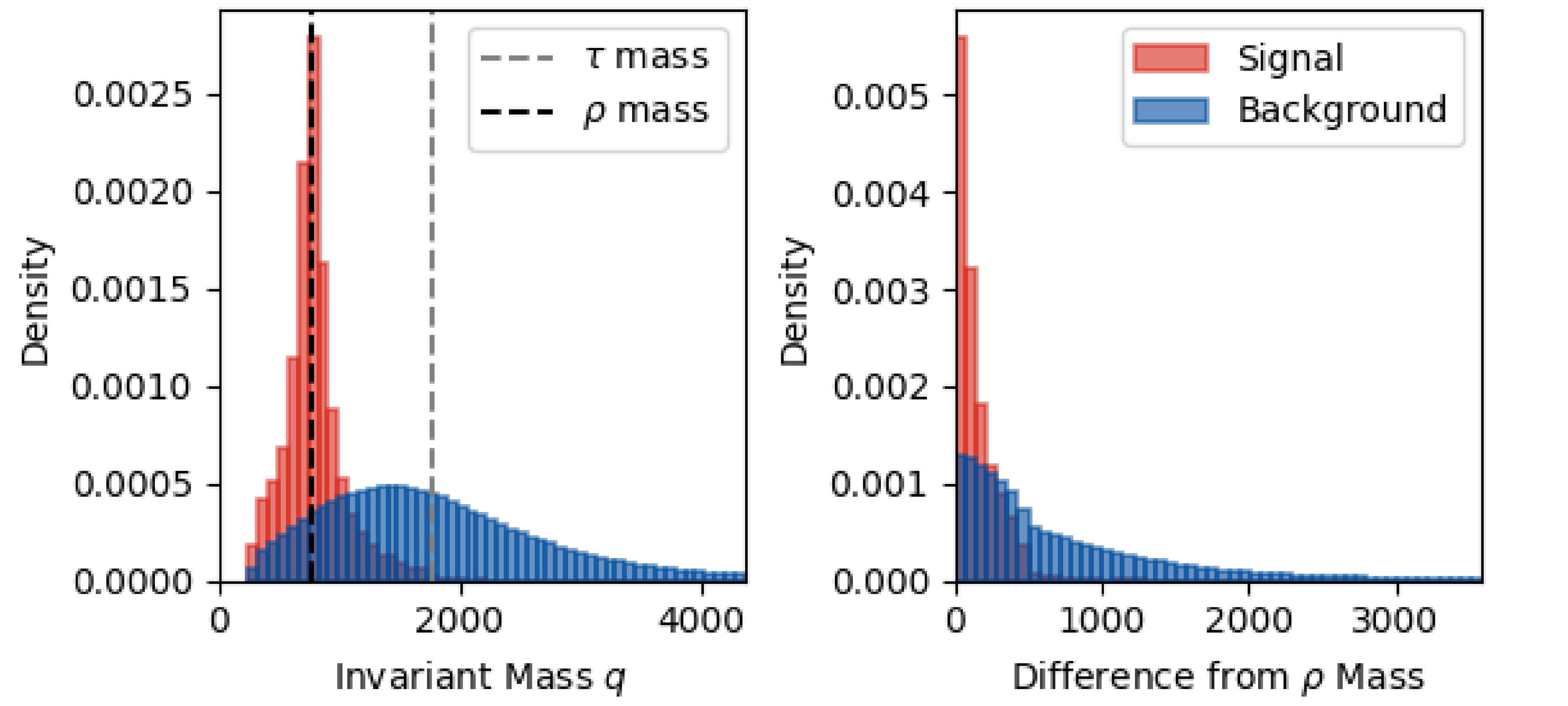

Figure 2: a) System invariant mass: 40% of Background π^0s are above the τ mass, b) Difference in invariant mass from ρ mass resonance for $1\pi^{\pm}\pi^0\nu_\tau$ mode.

Multivariate analysis is used to classify the π^0 via a boosted decision tree algorithm:

- XGBoost is an optimised algorithm where models are trained sequentially by minimising the residuals of prior models using gradient descent; high performing models have boosted influence.

5. Results

There is a 760:1 ratio in the number of the background versus signal π^0 candidates. The ROC curve provides information about the true and false positive rates across a range of decision thresholds.

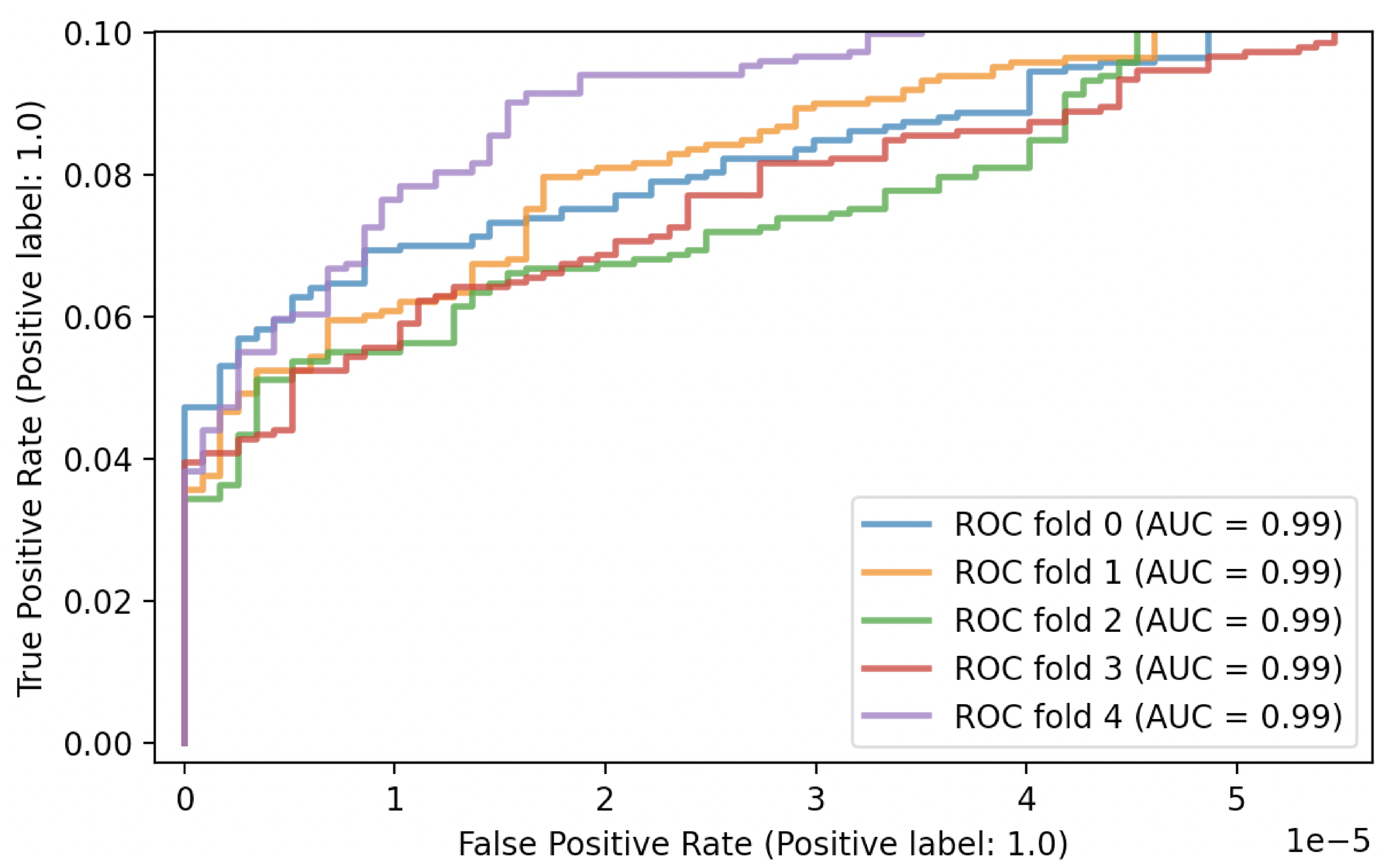

Figure 3: ROC Curve for five cross validations

The AUC is a good single number metric for imbalanced datasets. We select the best classifier with AUC = 0.987 and a 0.047 true positive rate at false positive rate equal to zero.

6. Future Work

- The next step in τ reconstruction is the identification of the charged hadronic decay products eg. $\pi^{\pm}$
- The DaVinci program currently identifies these for us, which is not reflective of LHCb capabilities.

tr4018@ic.ac.uk 10th March, 2022 hgs18@ic.ac.uk