DeepMODE – A Neural Network for Tau Decay Mode Classification

Adam Brzosko and Fin Osman Sellwood
Supervisor: Professor David Colling
Imperial College London
High Energy Physics (HEP)

1. Motivation

The CP-state of the Higgs boson is an important indicator of new physics, as many theories predict the Higgs not to be CP-even [1].

However, the Higgs is extremely short-lived, and can only be observed by its decay products. The two most frequent fermionic Higgs decays are $H \rightarrow b\bar{b}$ and $H \rightarrow \tau\bar{\tau}$ [2], and the tau decay was the focus of this project. Tau particles also decay too quickly to be measured by a detector, and the five leading tau decay modes are observed instead (Table 1).

Mode	$l^{\pm}$	$h^{\pm}$	$h^{\pm}\pi^{0}$	$h^{\pm}\pi^{0}\pi^{0}$	$h^{\pm}h^{\mp}h^{\pm}$
B(%)	35.2	11.5	25.9	9.5	9.8
Symbol	e, μ	π	ρ	a_1^{1pr}	a_1^{3pr}

Table 1: The five most frequent tau decay modes, and their branching ratios as percentages. Here $l^{\pm}$ indicates a lepton ($e^{\pm}$ or $\mu^{\pm}$) and $h^{\pm}$ represents a meson, usually a pion ($\pi^{\pm}$). Table inspired by [3].

The current best value for the CP state of the Higgs is from the CMS group at the LHC, which found the **Higgs' CP mixing angle to be $\boldsymbol{\phi_{\tau\tau} = (4 \pm 17)}$°** [3], where an angle of 0° would indicate a fully CP-even Higgs, and 90° fully CP-odd. This method used the CP-sensitive angles between tau decay planes and required the decay modes of the two tau particles to be known.

In this project, we have collaborated with the CMS group to reduce the uncertainty on this value, focusing on improving the classification of tau decay modes using new image-based machine learning techniques.

2. Machine Learning Techniques

Boosted Decision Trees

In the previous analysis, tau decay modes were identified using a boosted decision tree (BDT), a type of machine learning model that processes a set of input values, or 'features', to return probabilities that an event is one of the given decay modes.

Neural Networks

Neural networks (NNs), like BTDs, take a series of input features and return a set of outputs. However, they are also more versatile, capable of accepting multiple inputs with different datatypes (images, lists etc.) and returning multiple outputs.

Our Work

We have developed and fine-tuned a deep NN with added image processing (convolutional) layers.

- This allowed us to not only incorporate 'high-level' parameters, like particle momentum, but also images created from the angular positions of particles registered by the detector.

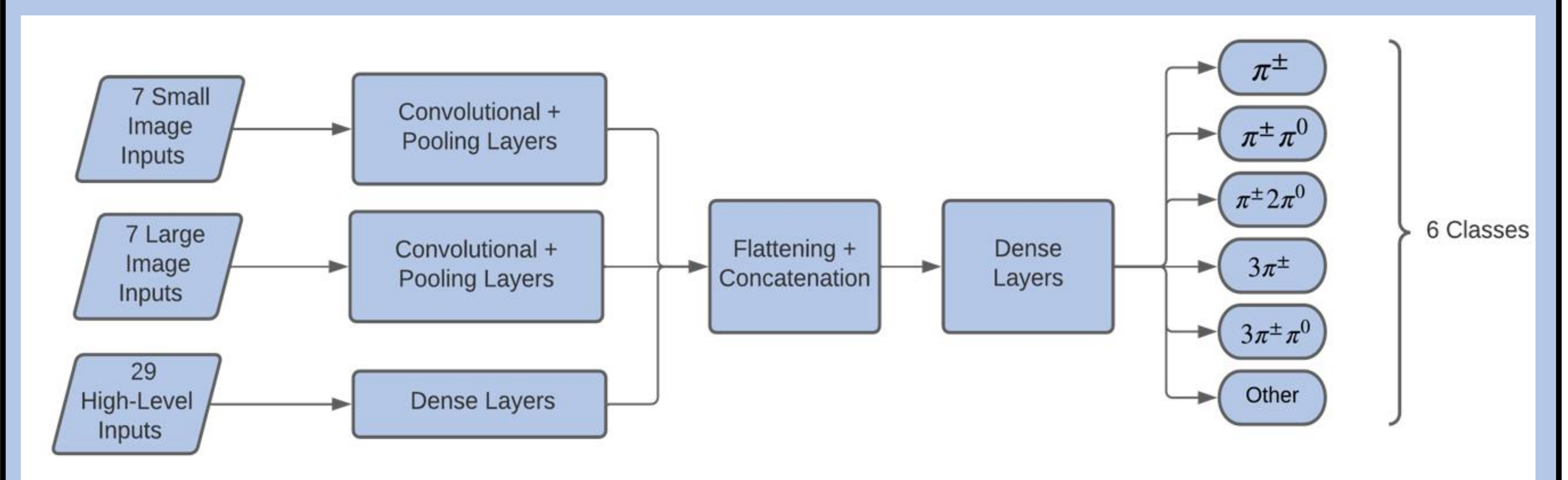

Fig. 2: A flowchart representing the general structure of our NN

- Our NN accepted 29 high level parameters and seven pairs of large and small images, representing the distribution of energy, momentum, and different particle types within the CMS detector.
- The output was a list of probabilities that a given event belongs to any of the six classes shown on the flowchart above.
- We found better results when training two different models, one to deal with singly-charged modes ($\pi, \rho, a_1{}^{1pr}$) and one for modes with three charged particles ($a_1{}^{3pr}, a_1{}^{3pr} + \pi^0$).
- These models would usually take ~1hr per epoch and 24hrs for a fully-trained model, although more complex structures (especially more convolutional layers) increased training time to around a week.

3. Results

We evaluated the performance of our NN models using Receiver Operating Characteristic (ROC) curves. These show how number of correctly classified events (true positive) and misclassified events (false positive) are correlated. The ideal classifier would be a right-angle with the corner at (0,1).

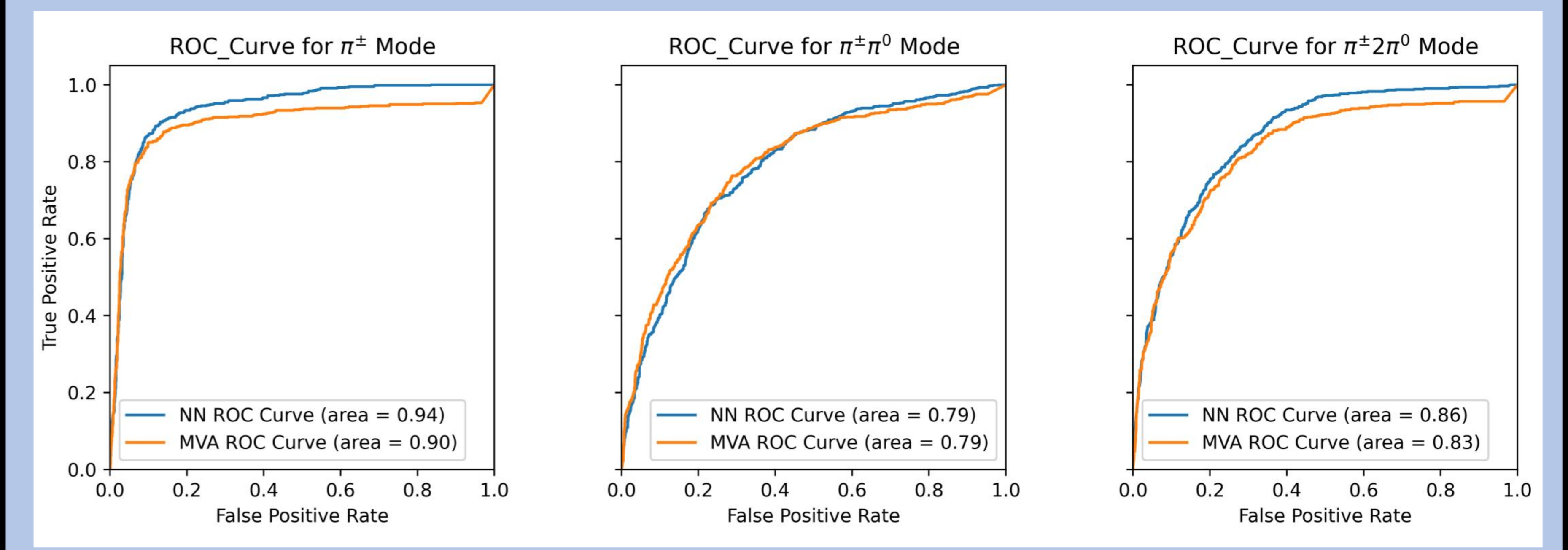

Fig. 3: ROC curves for the singly-charged tau decay modes for both our best NN (blue) and the BDT (orange). Similar curves exist for modes with three charges.

We also measured the purity and efficiency of each mode $\boldsymbol{i}$, defined as;

$$Efficiency = \frac{Correctly\ classified\ \boldsymbol{i}\ events}{Total\ genuine\ \boldsymbol{i}\ events} \qquad Purity = \frac{Correctly\ classified\ \boldsymbol{i}\ events}{Total\ classified\ \boldsymbol{i}\ events}$$

Purity and efficiency have a direct impact on the uncertainty of the final mixing angle. These are compared to the BDT, and to CMS's HPS algorithm in Fig. 4. This shows the NN had **higher efficiency but lower purity** than the BDT model.

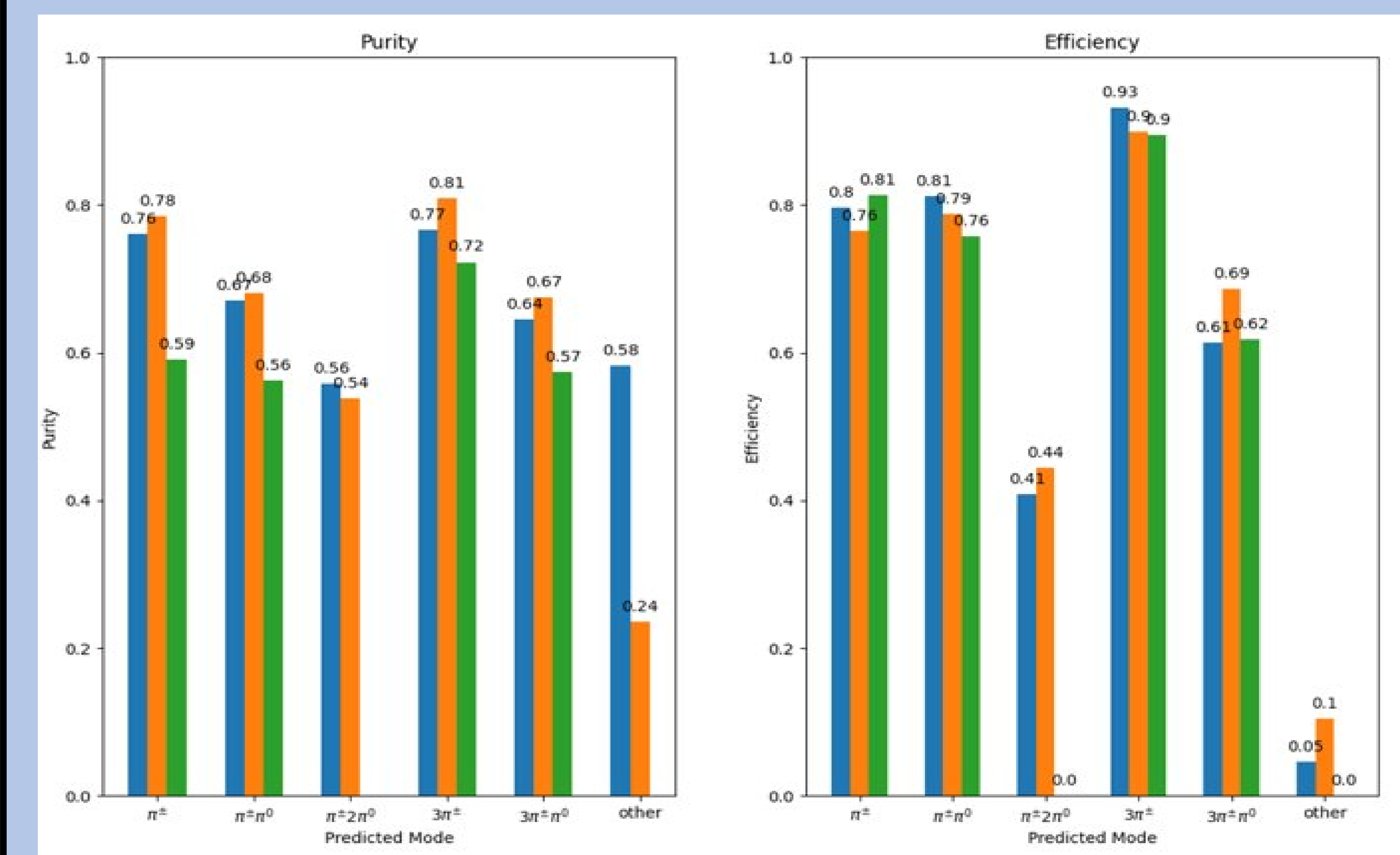

Fig. 4: Comparison of purity and efficiency of the classification between our NN (blue), the BDT (orange) and the default HPS algorithm (green).

We also found that by varying the cutoff probability for different modes we could **improve efficiency at the expense of losing data**. The effects of this on the full CP-state calculation are not yet known.

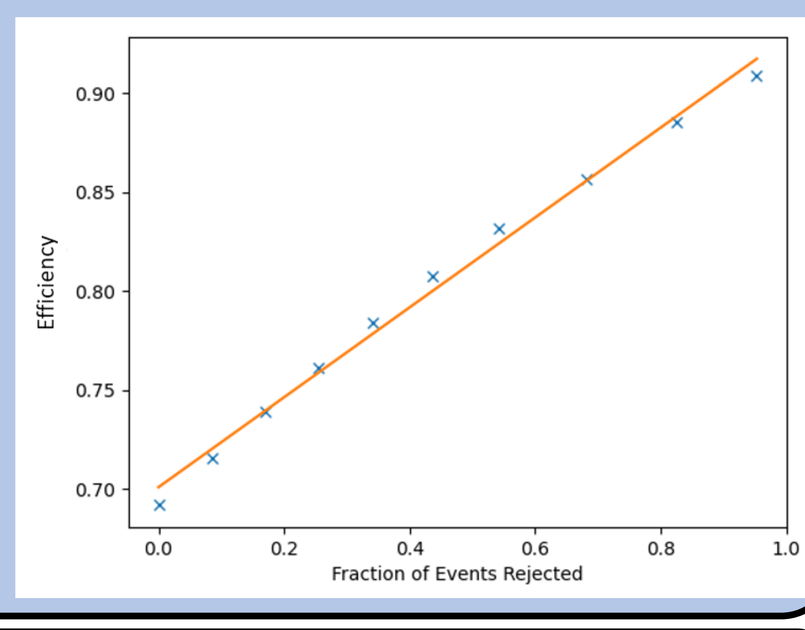

Fig. 5: Efficiency of classification against the fraction of events rejected

4. Conclusion

- Our neural network can be used for classification of past and future data from the CMS experiment and is at least as powerful as existing BDT models.
- A more complex network and longer training will yield better results, although how much better is currently unknown.

Next Steps

- More research should be done into how varying classification cutoffs affects overall results, as this can improve results in a more significant way (~10%) than optimising a network can (~1%).
- By focusing on distinguishing the ρ and $a_1{}^{1pr}$ modes we can develop models that are overall more useful, as this is the least well-defined sector.
- Training could also focus on better resolving π^0 particles, which are currently badly defined.
- Novel network architectures, like graph networks, may also be researched.

References

[1] D. Fontes et al. "The C2HDM revisited". In: Journal of High Energy Physics 2018.2 (2018), pp. 1–40.
[2] SM Higgs Branching Ratios and Total Decay Widths, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR, Accessed 01/03/22.
[3] CMS collaboration et al. Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at s= 13 TeV. Tech. rep. CMS-PAS-HIG-20-006, 2020.