Heavy Flavor Decay Classification with Deep Neural Networks

Tanner Mengel

September 25, 2024

Abstract

Deep learning methods are applied to discriminate decay channels of heavy quarks produced in pp collisions at $\sqrt{s_{NN}}$ =200 GeV. Previous applications of deep learning techniques to hadron decays have successfully reconstructed decay properties through use of Deep Neural Networks (DNN) [Liu et al., 2022]. This project shows binary classification of prompt and non-prompt charm quark products is possible using a DNN.

1 Introduction

2 Data Set

3 Model

4 Benchmarking/Testing/Optimizing

Heavy Flavor Decays

Heavy-quarks are produced in hadron-hadron collisions by parton hard-scattering. Open Heavy-Flavor particles containing $c-$quarks, such as the D^0 ($\bar{D}^0$) mesons, provide information on charm quark production and partonic energy loss within the medium produced by the collisions. $c-$quarks can be done by measuring the D^0 meson's produced in initial collisions. The D^0 meson is short lived and must be indirectly measured through reconstructing stable decay daughters. One of the the primary decay channels used to reconstruct D^0 mesons is $D^0 \to K^-\pi^+$ and π^+ which has a branching ratio of $\approx 3.9\%$. This channel contains sources of background from D^0 meson's produced form beauty-quarks rather than charm quarks. Contributions from non-prompt D^0's must subtracted which is typically done through cuts to topological decay variables. An example of the contribution from non-prompt background sources is shown in Figure 1.

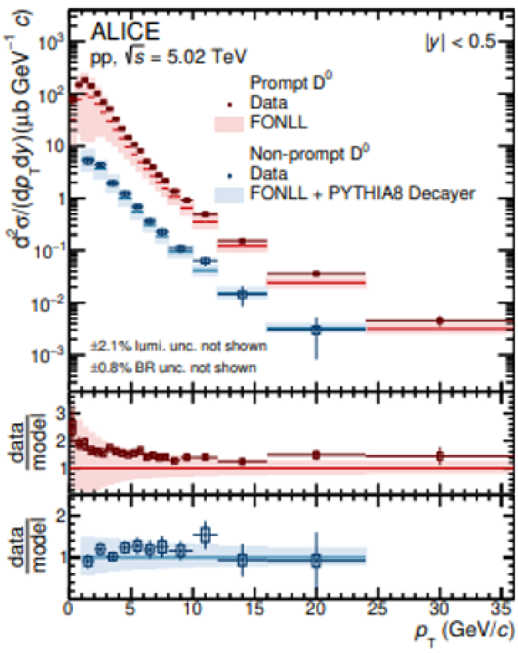

Figure 1: Contribution from prompt and non-prompt D^0 to invariant yield measured in pp collisions at $\sqrt{S_{NN}} = 5.02$ TeVS. Acharya et al., 2021.

Data Generation

The topological variables that are typically cut on are shown in Figure 2. These variables are generated using Pythia8 in proton-proton collisions where $\sqrt{S_{NN}} = 200$ GeV. The branching ratios are set to 1 for the following decays:

$$D^0 \to K^-\pi^+ \qquad B^- \to D^0\pi^- \qquad \bar{B}^0 \to D^0\pi^+\pi^-$$

Following simulation, the 'reconstructed' variables are fast-simmed to reproduce detector like vertex resolution and momentum smearing. This fast-sim technique is based of the method outlined in Xiaolong Chen, n.d. Detector efficiency values were inspired by the sPHENIX detector as well as the psuedorapidity acceptance of $|\eta| < 1.1$.

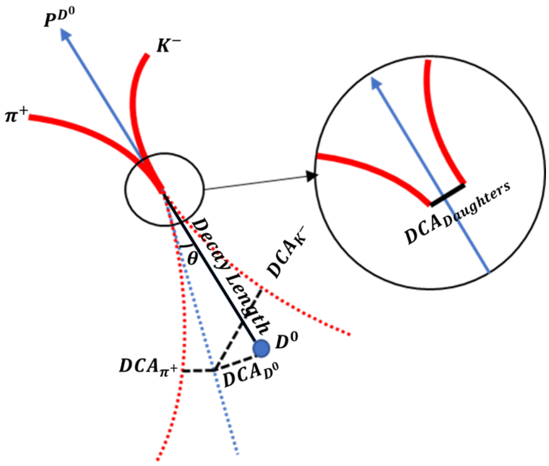

Figure 2: Topological decay variables fast-simmed in Pythia8.

The data is sectioned into two categories shown in Figure 3. The prompt and non-prompt reconstructed D^0 candidates are labeled with truth labels and then mixed prior to training. The data set consists of 7 million events with an identified D^0 from either prompt or non-prompt sources. The data is binned by p_t hard-min accordingly: [5,15,25, 35,45,55,10000] GeV/c. The data is truth labelled and then randomly split into test/train data sets (50/50). The training and testing datasets are then pre-processed and feed into TMVA.

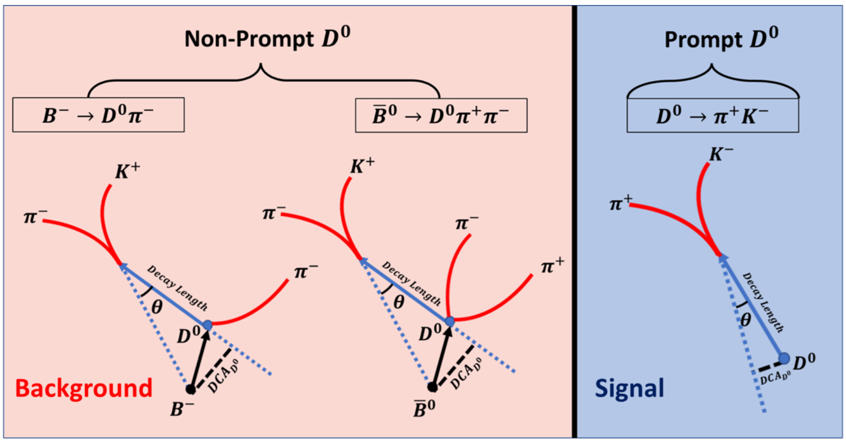

Figure 3: Categories of reconstructed D^0 mesons where signal is taken to be prompt charm from $D^0 \rightarrow K^-\pi^+$ and background is non-prompt.

Deep Neural Networks

A neural network is simply a non-linear map from an input space to an output space. The neural networks that are used in the binary classification application are feed-forward neural networks. This means information move in the positive direction and weights and basis must be updated retroactively using back-propagation. A diagram of a feed-forward neural network is shown in Figure 4 which is taken from TMVA's user manualHoecker et al., 2007. This is a shallow neural network, because it only has 1 hidden layer. The networks implemented in this project are considered Deep neural networks because they are dense in neurons and have multiple hidden layers.

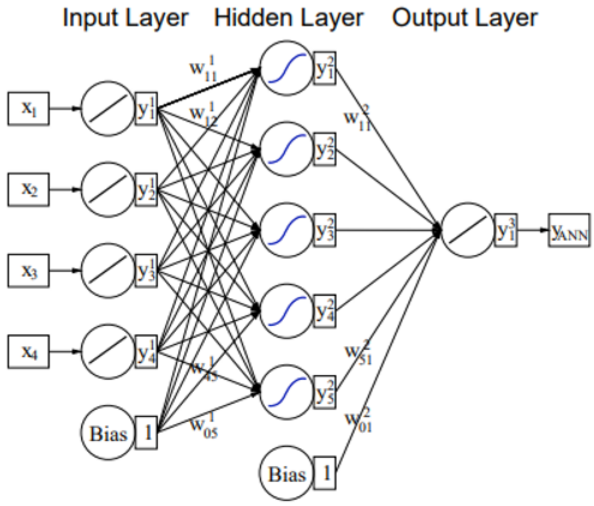

Figure 4: Feed Forward neural network with 1 hidden layer Hoecker et al., 2007

The activation functions for the neural networks implemented in this project are both hyperbolic tangent and sigmoid functions. The shape of the activation curve greatly impacts the neurons willing-ness to activate and therefore must but tuned for optimal performance.

TMVA

The tool used to implement DNN's and preprocess generated data is ROOT's TMVA (ToolKit for Multivariate Analysis). This interface allows from easy model generation and training. TMVA also has built in interface with several of the leading machine learning libraries such as Tensorflow and PyTorch. Full documentation of TMVA is provided Hoecker et al., 2007.

5 Implementation

The neural networks implemented within this project are all multilevel perceptron layed (MPL) neural networks. Details regarding each of the 4 networks are as follows:

$$MLP1 : [N + 1, N + 10, 1], \tanh, Regulated, BP$$

$$MLP2 : [N + 1, N - 1, N + 5, 1], \tanh, Regulated, BP$$

$$MLP3 : [N + 1, N - 1, N + 10, N], Sigmoid, Regulated, BP$$

$$TMlp : [N + 1, N, 1], Sigmoid, CrossValidated, BE$$

This can be interpreted as, architecture, activation function, over training precautions and weight updating scheme. All are updated via back propagation (BP) with the exception of the TMlp which is updated biased of Bayesian probability estimation (BE). The over training precautions are Regulated to sample the loss/epoch curve every 5 epochs to cut training early if the network exceeds optimal discrimination. The architecture of MLP1 and MLP3 is shown in Figure 5 and Figure 6, respectively.

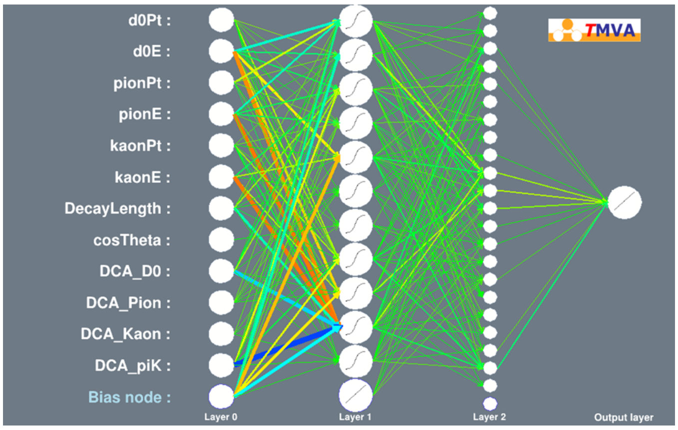

Figure 5: Network architecture of MLP1. [N+1,N+10,1]

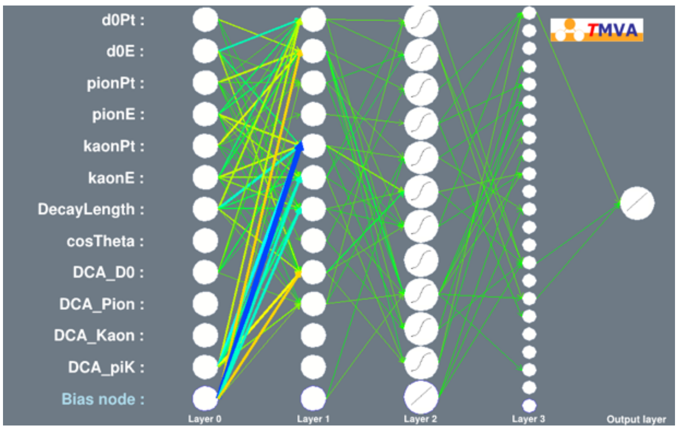

Figure 6: Network architecture of MLP3. [N+1,N-1, N+10,N]

Input variables are plotted against background in Figure 7. This gives a good idea over the overlapping of the input space and therefore the difficulty to discriminate within that space.

Luckily TMVA offers a variety of tools to help differentiate this input space. Decorrelation and principle component analysis help limit out input dimensional to orthogonal representations. Furthermore, over correlated coordinates can be taken as access information and reduced in the training of the DNN's. Examples of TMVA's correlation and decorrelation data preprocessing is shown in Figure 8

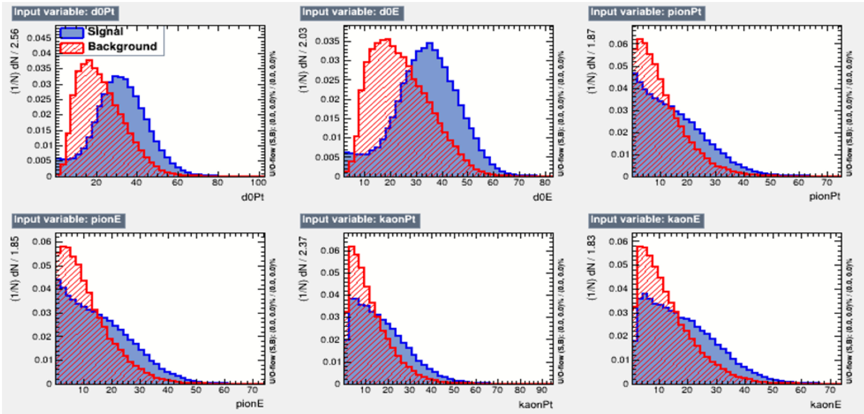

Figure 7: Set of input variables separated into signal and background

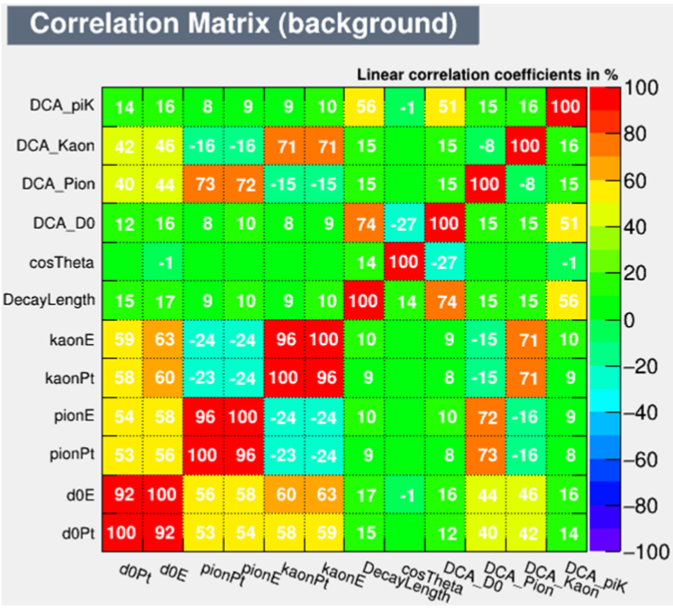

Figure 8: Correlation matrix from TMVA with decay topological variables.

6 Results

Figure 9 shows a normalized response to the binary classification problem. There is clear separation in method response and cuts can be made to optimize the signal purity.

The following is the classification results of each neural network. The first 4 classification responses are not normalized making interpretation less than clear. This is solved by including the Receiver Operating Characteristic (ROC) curve which details all information regarding binary classifier.

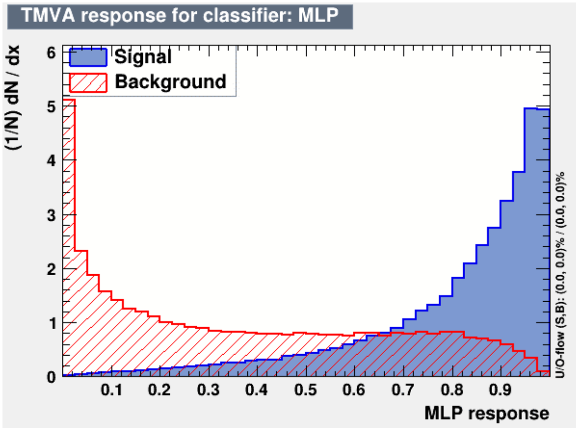

Figure 9: Normalized classifier Response

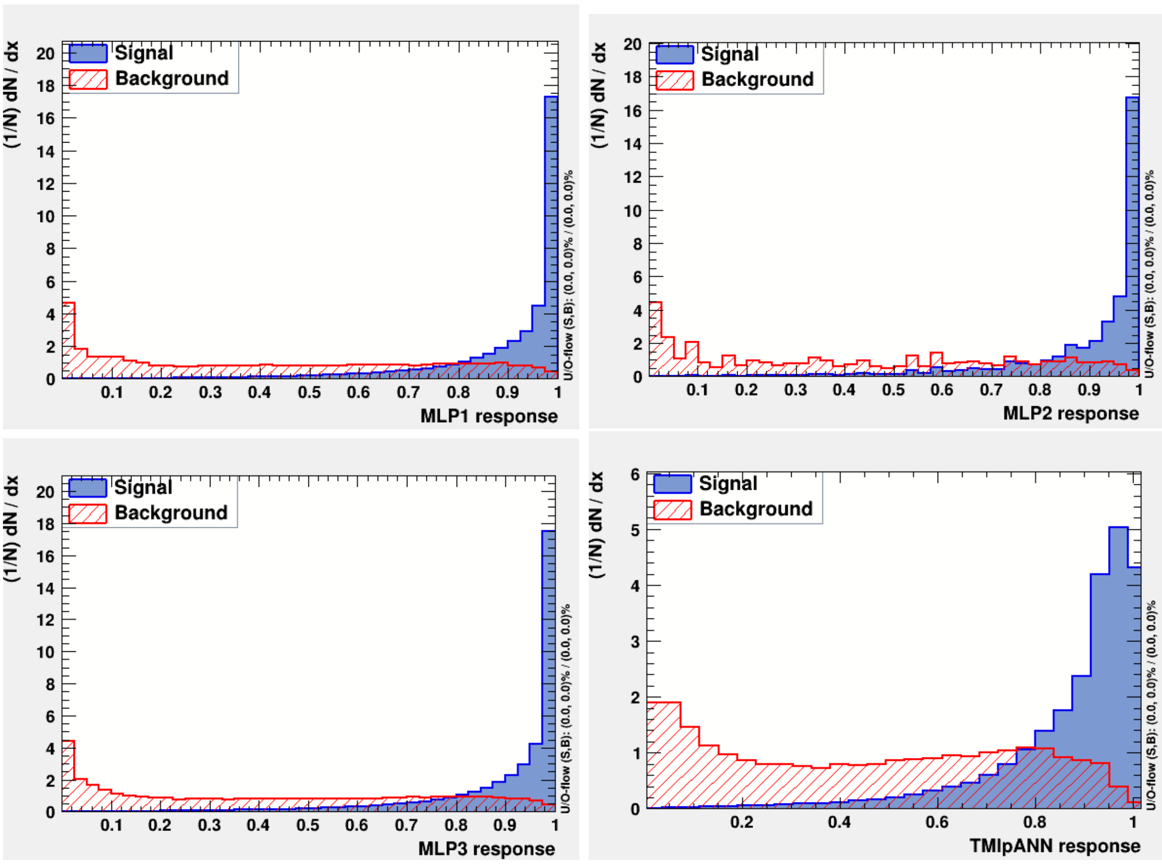

Figure 10: Classified responses for the four neural networks -raw output.

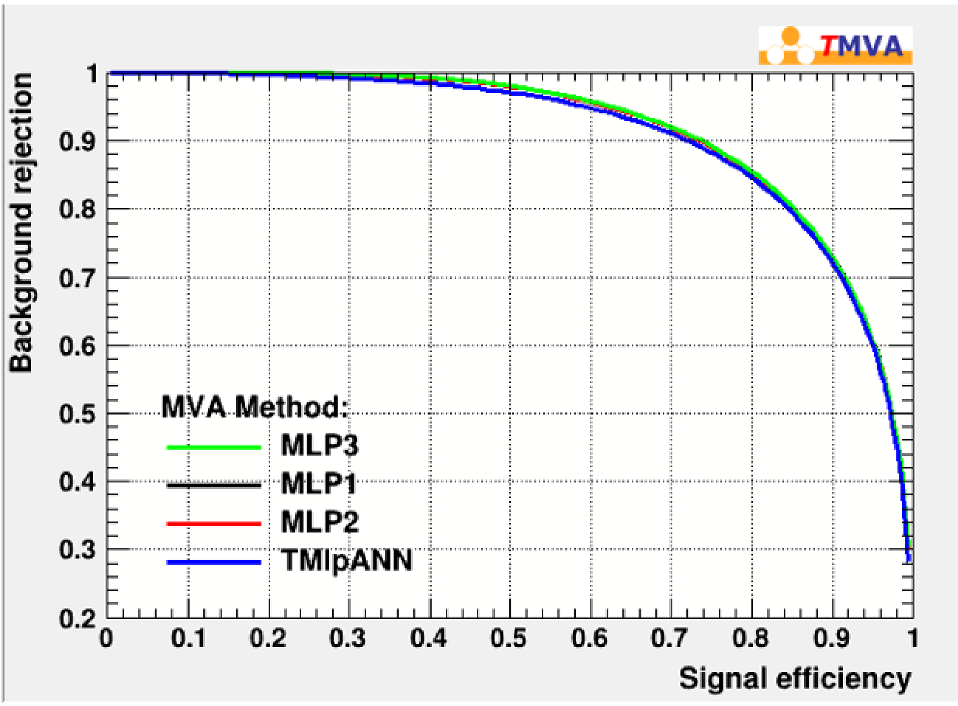

Figure 11: ROC curve showing clear classification skill in signal background discrimination.

7 Conclusion

A DNN was applied to a binary classification problem which was non-prompt vs prompt D^0 meson decay channel discrimination. The neural network was able to successfully classify these decay channels which is clear in Figure 12 which compares skill of DNN to a traditional rectangular cut on the decay topology.

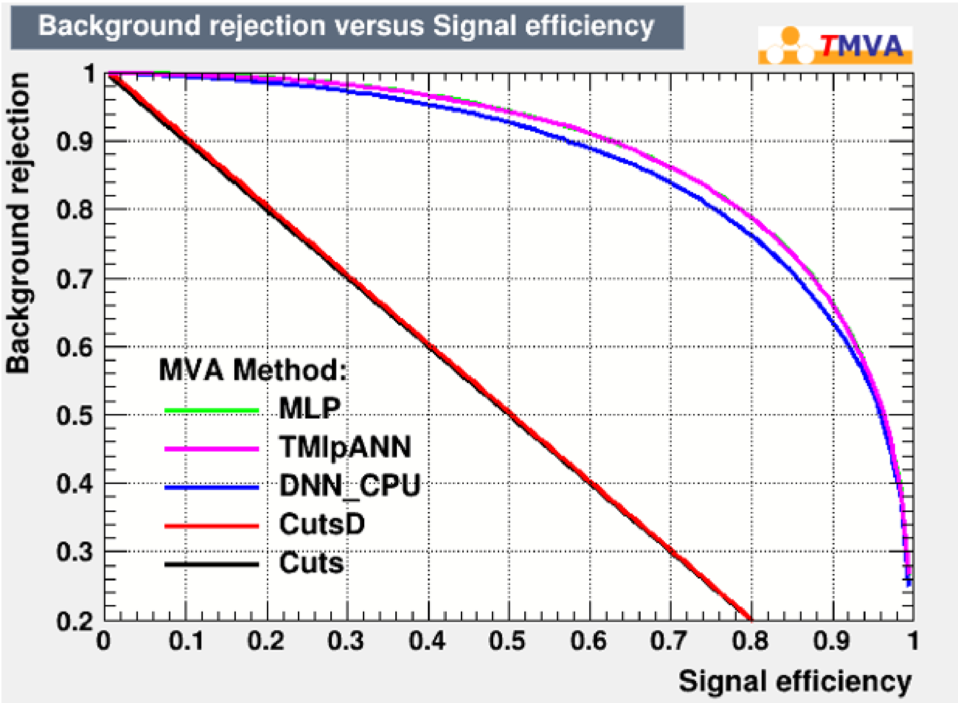

Figure 12: ROC curve showing clear classification skill in signal background discrimination vs rectangular cuts.

References

Hoecker, A. et al. (2007). *TMVA - Toolkit for Multivariate Data Analysis*. DOI: 10.48550/ARXIV.PHYSICS/0703039.

Liu, Jiahao et al. (Apr. 2022). "Study of exotic hadrons with machine learning". In: *Physical Review D* 105.7. DOI: 10.1103/physrevd.105.076013.

S. Acharya, and et al. (May 2021). "Measurement of beauty and charm production in pp collisions at \sqrt{s} = 5.02 TeV via non-prompt and prompt D mesons". In: *Journal of High Energy Physics* 2021.5. DOI: 10.1007/jhep05(2021)220.

Xiaolong Chen 5 Xin Dong, Guannan Xie (n.d.). *D0-meson and B+-meson production in Au+Au Collisions at s 3 NN = 200 4 GeV for sPHENIX*. URL: https://portal.nersc.gov/project/star/dongx/sPHENIX/sPH-HF-2017-002-v1.pdf.