PROCEEDINGS OF SCIENCE

Machine Learning for New Physics Searches in $B^0 \to K^{*0}\mu^+\mu^-$ Decays

S. Dubey,[a,c,*] **T.E. Browder,**[a] **S. Kohani,**[a] **R. Mandal,**[b] **A. Sibidanov**[a] **and R. Sinha**[a]

[a] *Department of Physics and Astronomy, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA*

[b] *Department of Physics, Indian Institute of Technology Gandhinagar, Gujarat 382055, India,*

[c] *Center for the Fundamental Physics of the Universe, Department of Physics, Brown University, Providence, RI 02912, USA*

E-mail: shawn_dubey@brown.edu

We report on a novel application of computer vision techniques to extract beyond the Standard Model (BSM) parameters directly from high energy physics (HEP) flavor data. We develop a method of transforming angular and kinematic distributions into "quasi-images" that can be used to train a convolutional neural network to perform regression tasks, similar to fitting. This contrasts with the usual classification functions performed using ML/AI in HEP. As a proof-of-concept, we train a 34-layer Residual Neural Network (ResNet) to regress on these images and determine the Wilson Coefficient C_9 in MC (Monte Carlo) simulations of $B^0 \to K^*\mu^+\mu^-$ decays. The technique described here can be generalized and may find applicability across various HEP experiments and elsewhere.

42nd International Conference on High Energy Physics
18-24 July 2024
Prague Congress Centre, Prague, Czech Republic

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(ICHEP2024)1034

1. Introduction

Recently, there have been possible hints of new physics beyond the Standard Model (SM) in the observed angular distributions of $B^0 \to K^{*0}\mu^+\mu^-$. These can be more clearly identified in angular asymmetries, such as the forward-backward asymmetry (A_{FB}), S_5, and others, described in Ref. [1]. Determining the scenario from which these apparent anomalies originate — via SM interactions with unaccounted-for hadronic effects, or beyond the SM (BSM) physics — is a key experimental problem.

Performing high-dimensional fits in all the angular and kinematic variables/observables to determine this can become complicated in the presence of backgrounds and detector resolution. To solve this, we propose a machine learning solution. We recast the problem as a computer vision problem and use a neural network (NN) to regress on images created using the angular and kinematic information to predict BSM paramters.

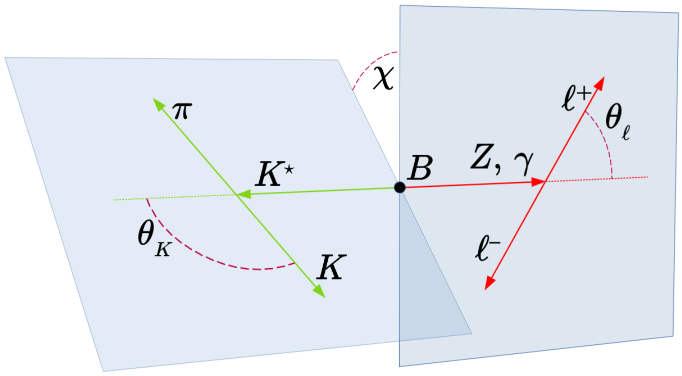


Figure 1: The $B \to K^*\ell^+\ell^-$ general decay topology showing the observables [2]. For this study we only consider the di-muon channel where $\ell = \mu$.

To train the NN we use generator-level Monte Carlo (MC) generated from an effective field theory model we have implemented [2] in the EvtGen framework [3]. The model uses the operator product expansion formalism in terms of Wilson Coefficients C_i (WCs).

Our model is parameterizable in terms of the WCs' deviation from their SM values, i.e. $\delta C_i \equiv C_i^{\mathrm{BSM}} - C_i^{\mathrm{SM}}$. Choosing a non-zero δC_i has the effect of altering the correlations between four variables: $q^2 \equiv M(\ell^+\ell^-)$, the cosine of the lepton helicity angle $\cos\theta_\ell$, the cosine of the helicity angle of the K^*, $\cos\theta_K$, and the angle χ between the decay planes of the di-lepton and K^* decay planes. Figure 1 shows the decay topology and the full set of angular observables. Our goal is to train a NN regression model to predict δC_9 from unseen images, generated by the decay observables given above.

2. Creating the Images

We generate high-statistics MC samples of $B \to K^*\mu^+\mu^-$ events, with $\delta C_9 \in [-2.0, 1.1]$ for several different values of δC_9; 500000 events per δC_9 value. Images are produced by binning the average, normalized, q^2 value of each event in 50 equal-width bins of $\cos\theta_\mu$, $\cos\theta_K$, and χ, so that approximately 250 ab^{-1}-equivalent events are in each (quasi-)image. The images are input as tensors to the NN and each have shape (50, 50, 50, 1), where the value 1 denotes the number of channels. Figure 2 shows two examples of these images for different values of δC_9.

PoS(ICHEP2024)1034

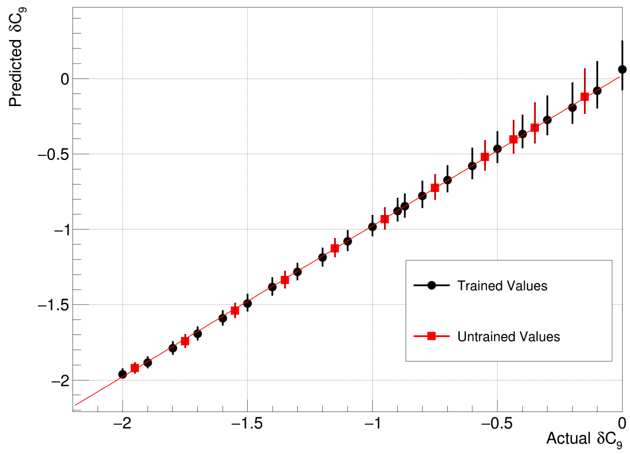
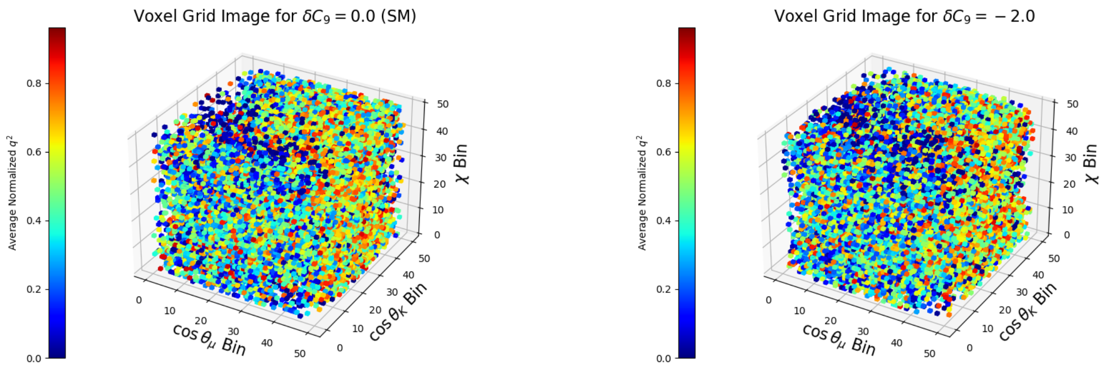


Figure 3: Linearity test from the test set. The black points are from distributions where the images are generated according to δC_9 values the ResNet has been trained with; the red points are from those where the images are generated according to δC_9 values with which the ResNet has *not* been trained.



(a) Voxel grid image for $\delta C_9 = 0.0$ (SM) **(b)** Voxel grid image for $\delta C_9 = -2.0$

Figure 2: Voxel grid images used for training and evaluation of the ResNet. Examples for the cases of $\delta C_9 = 0.0$ (SM) and $\delta C_9 = -2.0$ are shown.

3. The Neural Network

As we are now studying computer vision with 3D images, we employ a 3D version of the ResNet34 model [4], known for performing well on computer vision tasks. Stochastic gradient descent is used for optimization. The loss function is the mean absolute error (MAE) [5]. At the end of network there is one fully-connected layer employing 1000 neurons, followed by a drop-out layer with a 50% drop-out probability. The final layer is a dense layer with one neuron and a linear activation function that performs the regression task to predict δC_9 values directly from the images

4. Results and Discussion

To test the trained model, we examine ensembles of MC simulation experiments. For each of the 22 WCs that were used to generate training images, statistically independent samples of 900

PoS(ICHEP2024)1034

images are generated for testing, which are evaluated by the trained NN model. The mean and 16% and 84% error bars are determined from resulting distributions of predictions are. We also perform this test using δC_9 values that are between the ones used to generate the training images. The fit results are plotted against their generated δC_9 values to obtain a linearity plot, shown in Fig. 3. This demonstrates that the NN is able to learn how to predict δC_9 from images created using the information from $B^0 \to K^*\mu^+\mu^-$ events. For further details, link to the code, and acknowledgements, we refer the reader to Ref. [6]. This work supersedes our previous work.

References

[1] R. Aaij et al., Journal of High Energy Physics **2017**, 55 (2017)

[2] A. Sibidanov et al., Journal of High Energy Physics **08**, 151 (2024)

[3] D.J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001)

[4] K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition* (2015), arXiv:1512.03385

[5] A. Geron, *Hands-on machine learning with scikit-learn, keras, and TensorFlow*, 2nd edn. (O'Reilly Media, Sebastopol, CA, 2019)

[6] S. Dubey et al. (2023), arXiv:2311.13060

PoS(ICHEP2024)1034