Study of the $B^+ \to \pi^+\mu^+\mu^-$ decay with the LHCb experiment

Author: Marc Salvador Carreté
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisors: Carla Marín Benito and Ernest Olivart Pino

Abstract: In this work, a study of the $B^+ \to \pi^+\mu^+\mu^-$ decay is performed in the new data collected during Run 3 of the LHCb experiment. A Boosted Decision Tree (BDT) algorithm is used to select the events corresponding to this decay and mass fits are performed to extract its yield. The $B^+ \to \pi^+ J/\psi(\to \mu^+\mu^-)$ decay is used as a control channel. Using the yield of the control channel, the expected signal yield is found to be 1.64 ± 0.22. No signal events have been found by performing a fit directly, which is compatible with the expectations.
Keywords: Particle Physics, High Energy Physics, Data Analysis, Machine Learning, Boosted Decision Tree
SDGs: Industry, innovation and infrastructure

I. Introduction

A. The Standard Model

The Standard Model of particle physics (SM) is our current best theory in high energy physics (HEP) that describes the fundamental components of matter along with three of their interactions: the electromagnetic, strong, and weak forces.

In this model, particles are divided into two large groups: fermions and bosons. Fermions are particles with half-integer spin and define the fundamental particles of matter. They are divided into two groups: quarks and leptons, each divided into three pairs, called generations. The first generation is the lightest and most stable one of all three. Quarks and leptons differ in the forces they experience. Quarks interact via electromagnetic, strong, and weak interactions, whereas leptons only interact via the electromagnetic and weak forces. On the other hand, bosons are particles with integer spin and act as force carriers. There are six bosons: the photon (carrier of the electromagnetic force), the gluon (carrier of the strong force), two W bosons (W^+ and W^-) and the Z boson (all carriers of the weak force), and the Higgs boson, which gives mass to the other particles.

Even though the SM is the most complete theory we have in particle physics, and has successfully predicted a great variety of experimental results, it is still incomplete. Examples of problems the SM does not solve are: the gravitational force, the nature of dark matter and dark energy, the large matter-antimatter asymmetry, and the nonzero neutrino mass. Models beyond the standard model (BSM) have been theorized, but more experimental measures are needed to see if physics BSM exists.

B. The LHCb experiment at CERN

The Large Hadron Collider (LHC) is the world's largest particle accelerator, with a 27 km ring in which protons are accelerated to a nominal energy of 6.8 TeV thus achieving a total collision energy of 13.6 TeV. The ring is located in a tunnel 100 meters underground at CERN (*Conseil Européen pour la Recherche Nucléaire*), on the Franco-Swiss border. The accelerator holds four main experiments: ALICE, ATLAS, CMS and LHCb. Since the end of its construction in 2008, there have been three runs in which experimental data has been recorded: Run 1 (2009-2013), Run 2 (2015-2018), and Run 3 (2022-2026).

The Large Hadron Collider beauty (LHCb) is an experiment that uses the LHC to explore the interactions of particles containing the *beauty* (b) quark and the *charm* (c) quark [1]. The detector is a single-arm forward spectrometer 21 meters long, and it is oriented towards the direction of the beam. It is oriented this way because hadrons containing b and $\bar{b}$ quarks are produced mostly in the same forward or backward cone [2]. The detector is made of three main parts: the tracking system, the Particle Identification system (PID), and the trigger system.

The tracking system consists of the Vertex Locator system (VELO), the Upstream Tracker (UT) and the Scintillating Fibre Tracker (SciFi). The tracking system is used to reconstruct the particle trajectories and measure their momenta. The PID system consists of two Ring-Imaging Cherenkov detectors (RICH), four muon stations (M2-5) and a calorimeter system, and is in charge of identifying the detected particles. Lastly, the trigger system consists of a High Level Trigger (HLT), which is a software trigger designed to reject uninteresting events, and is used to reduce the amount of data stored [3].

C. The $B^+ \to \pi^+\mu^+\mu^-$ decay

One of the decays studied in LHCb is a decay of the B^+ meson, made of an *up* quark and a *beauty* antiquark ($u\bar{b}$) to a pion (π^+) made of an *up* quark and a *down* antiquark ($u\bar{d}$) and two muons, which are leptons. The decay is expressed as follows: $B^+ \to \pi^+\mu^+\mu^-$.

This decay is a $b \to d$ flavor-changing neutral-current process (FCNC), which is suppressed in the SM because the $b \to dl^+l^-$ does not include tree-level contributions, proceeding only through higher-order loop diagrams [4].

This decay could be affected by the presence of particles predicted by extensions of the SM, especially in models in which the flavor structure is different from that of the SM [4]. Therefore, this type of decay can be a good test for physics BSM.

This decay has already been detected and studied in previous runs of the LHCb experiment [4]. Therefore, the main objective of this work is to develop a dedicated selection for the particular conditions of the data collected during Run 3 (2024) using Machine Learning (ML) algorithms.

D. Data classification: Boosted Decision Trees

In HEP, to separate the events of interest (called Signal) from the rest of measured events (called Background), supervised learning methods are typically used. In these algorithms, a model is trained to classify signal and background events. In this work, the separation algorithms used are Boosted Decision Trees (BDT). A decision tree is structured as a collection of nodes that can be connected to other nodes in a tree-like pattern. Each node is divided into other nodes based on an input parameter, thus forming node branches. The last nodes of each branch (called leaves) give the final assigned probabilities [5]. BDTs are an ensamble of multiple decision trees therefore being able to perform more complex categorizations.

E. Datasets

In this work, two datasets (DS), or tuples, are used: the experimental DS, which includes all the registered data from the LHCb experiment; and the simulation DS, which includes data from simulated events using Montecarlo (MC) simulations.

These tuples are stored in ROOT files [6], which are structured in a table-like manner, in which each row indicates a given event and each column a variable for that event (for example, the mass of the B^+ meson or its momentum).

In this work, the datasets are divided into two modes, depending on the value of the squared invariant mass of both muons $q^2 = (p_{\mu^+} + p_{\mu^-})^2$:

- **Rare mode.** Corresponding to the region in which $q^2 < 8\,\mathrm{GeV}^2$. The desired decay is found in this region, therefore, the BDT is trained in this mode.
- **J/ψ mode.** Corresponding to the region in which $8\,\mathrm{GeV}^2 < q^2 < 11\,\mathrm{GeV}^2$. This channel corresponds to the $B^+ \to \pi^+ J/\psi(\to \mu^+\mu^-)$ decay and will be used as a control channel.

The branching ratios of both decays are [7]:

$$\mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-) = (1.78 \pm 0.23) \times 10^{-8}$$
$$\mathcal{BR}(B^+ \to \pi^+ J/\psi) = (3.92 \pm 0.09) \times 10^{-5}$$

Therefore, the ratio between them is:

$$\frac{\mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-)}{\mathcal{BR}(B^+ \to \pi^+ J/\psi)} = (4.54 \pm 0.60) \times 10^{-4} \qquad (1)$$

II. Methodology

A. Preselection

To improve the BDT performance, several cuts can be applied to the data before training the algorithm, so that some of the background can be eliminated while losing the least amount of signal possible. This process is known as a preselection. In this study, three types of preselection cuts have been applied: cuts to "PROBNN" variables, mass swap cuts and *Triggered on Signal* (TOS) cuts.

To aid with the selection of the variables used for the preselection, a large list of variables from the tuples has been plotted, superposing both the experimental data and the MC simulation data. These are called discrimination plots and they have been done in the rare mode. Examples of these plots can be found in appendix A. Using them, a cut can be chosen in which a large part of background is removed while keeping the signal almost intact.

"PROBNN" variables are the output of a Neural Network that classifies the detected particles in different species (π, μ, etc.). They represent the probability that a particle that is reconstructed is indeed that particle, or is a different one. For example, the variable "Mup_PROBNN_E" gives the probability that a particle reconstructed as a μ^+ is actually a e^+. To decide where these cuts should be applied, these variables have been plotted and a cut has been chosen such that the events in which the particle is the least likely to be the one detected are left out. All "PROBNN" cuts can be found in table III, in appendix B.

Mass swap cuts refer to the removal of the events in which one or more of the children particles have been misidentified. This can happen because in the reconstruction of the events, the masses of the different children particles are assumed, not measured. This can cause that particles that don't belong in the studied decay are considered as particles that do. In this case, three of these particle swaps have been studied, all coming from the decay of a D^0 meson: $D^0 \to h^+h^-$ where h is either K or π and has been misidentified as a μ. For these three swaps, the mass of the $\mu^+\mu^-$ pair has been calculated changing the mass hypothesis (m_{swap}) and a veto $|m_{\mathrm{swap}} - M_{D^0}| < 12\,\mathrm{MeV}$ has been applied.

Lastly, TOS variables are boolean (True or False) variables in the tuples that express whether an event sets off specific software triggers. Triggered on signal means that the selection criteria have to be fulfilled by candidate particles of the decay of interest. Two of these triggers have been selected for the preselection: "Bu_Hlt1TwoTrackMVADecision_TOS" and "Bu_Hlt1TrackMuonMVADecision_TOS". Both triggers require that the detected particles have enough energy and transverse momentum. The first variable is a trigger for pairs of tracks, while the second one is for particles detected in the muon chambers.

B. Classification using a BDT

In order to classify the experimental data with a BDT, the algorithm has to be trained and then, applied to the experimental DS.

The first step in training the BDT is selecting which of the many variables present in the DSs will be used. These variables are called features. In order to choose them, a large set of variables used in previous analyses of the studied decay have been plotted in discrimination plots (see appendix A for an example) using data outside the signal window (which is the region in the reconstructed B^+ mass in which signal is expected). Then, the variables in which the experimental data and the simulation show more discrimination power have been selected. In total, 22 features have been selected. The selected features are related to the dynamics of the particles and to the quality of the reconstruction of their trajectories. A more detailed description of the types of features used can be found in appendix C.

Once all the variables have been selected, a BDT is trained using the python library `xgboost` [8]. To train the BDT, data from the experiment outside the signal window is used as background, and data from simulations as signal. The full data sample is split in two uneven parts: 80% of the data is used for training and the other 20% to test the trained model. Because the studied decay is in the rare mode, the BDT is trained and tested in this mode.

After training, a collection of outputs is plotted. These outputs are:

- **Feature importance.** This plot shows how much a feature is used in the decision trees (F score).
- **Probability plots.** These plots are histograms that show the probability, assigned by the BDT, that each event is signal or background. These plots are shown for both the signal and background samples and for the training and testing samples. Ideally, the probabilities of the background samples should be close to zero, while those of the signal samples should be close to one.
- **ROC curve and AUC score.** A ROC (Receiver Operating Characteristic) curve is a True Positive Rate (TPR) vs False Positive Rate (FPR) plot that shows the relationship between these two values for different BDT cuts. The area under each curve (or AUC) is also determined. Ideally, the AUC should be close to 1. This plot is done for both the training and testing samples.

C. Optimization of the BDT cut using a Figure of Merit

With the BDT trained and all tests done, an optimal cut to the probability needs to be found. To find this cut, a Figure of Merit (FoM) is used. A FoM is a function that depends on the signal and background yields measured for different BDT cuts. Finding the best BDT cut is equivalent to optimizing the FoM. The FoM used in this work is the significance function, and is expressed in equation 2. In this equation, S and B are the expected signal and background yields, respectively, for each BDT cut.

$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}} \tag{2}$$

In order to find the signal yield for each cut, the cut efficiency is calculated by dividing the number of events in the MC simulation after the given cut by a reference value, which is the number of MC events for a cut at 0.2. Then, the signal yield is found by multiplying this efficiency by the reference signal yield, which is obtained from data using the cut at 0.2.

To find the background yield, a fit is done outside the signal window in the experimental data for each cut and the yield in the signal window is extrapolated.

For simplicity, the FoM is optimized in the J/ψ mode.

D. Mass Fit

Even after optimizing the BDT cut and applying it, there is some background in the data. In order to find the actual signal yield, fits to the different signal and background contributions are done both in the rare mode and in the J/ψ mode.

In the J/ψ mode, four background contributions are considered:

- Combinatorial background: detections of a π^+, a μ^+ and a μ^- coming from other processes.
- $B^+ \to K^+ J/\psi$ where K^+ is misidentified as a π^+.
- $B^0 \to K^{*0} J/\psi$ where $K^{*0} \to K^+\pi^-$ and K^+ is not reconstructed.
- $B^0 \to \rho^0 J/\psi$ where $\rho^0 \to \pi^+\pi^-$ and π^- is not reconstructed.

The combinatorial background follows an exponential profile and therefore, it is modeled as such. On the other hand, the other contributions are fitted using a Double Sided Crystal Ball [9]. This probability density function (PDF) is a Gaussian function near its mean value, but decays as a different power-law in each of its tails.

In the rare mode, because the number of events remaining is low, the background is considered to be fully combinatorial. Again, this background is modeled as an exponential.

In both modes, the signal is fitted with a Gaussian curve.

In the J/ψ mode, the parameters of each PDF are fixed, leaving only the yield of each contribution variable. To find the parameters of the combinatorial background contribution, a fit is done outside the signal region. To extract the shapes of the distribution of the other contributions, fits are done in MC simulations of these background decays.

On the other hand, in the rare mode, because of the small number of events left after applying the BDT, all the parameters of both curves are left variable.

III. Results

A. BDT outputs

The BDT outputs explained in section II B are shown in this section for the trained model.

The feature importance plot is shown in figure 8, and can be found in appendix C along with an explanation of the different kinds of features it presents.

The probability plot (in logarithmic scale) is shown in figure 1. It can be seen that the Train and Test samples show very similar curves for both signal and background data. This is indicative that the BDT has not been overtrained, which would mean it has been overtuned to fit the training sample, therefore it would not be able to discriminate well enough when exposed to new data.

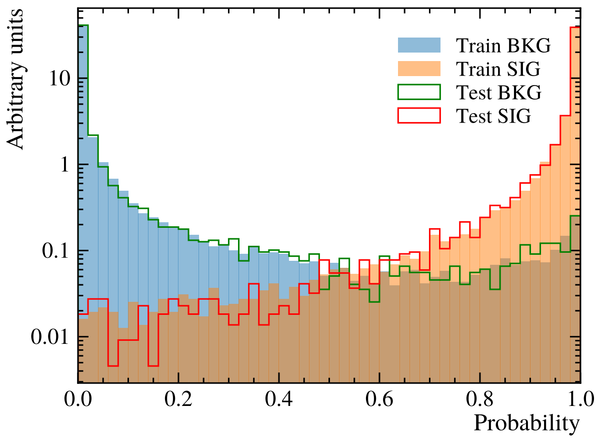

FIG. 1: Probability plot for the trained BDT in logarithmic scale.

Finally, the ROC curve and the AUC score can be seen in figure 2. The training and the testing samples show really similar curves and an AUC score close to 1, which is a good indicative that the BDT can perform a good classification.

B. FoM results

Figure 3 shows the figure of merit obtained for this BDT. It also shows a vertical line corresponding to the optimal cut, which has been found to be at 0.8774 ± 0.0002. The evolutions of S and B yields used to calculate the FoM can be found in appendix D.

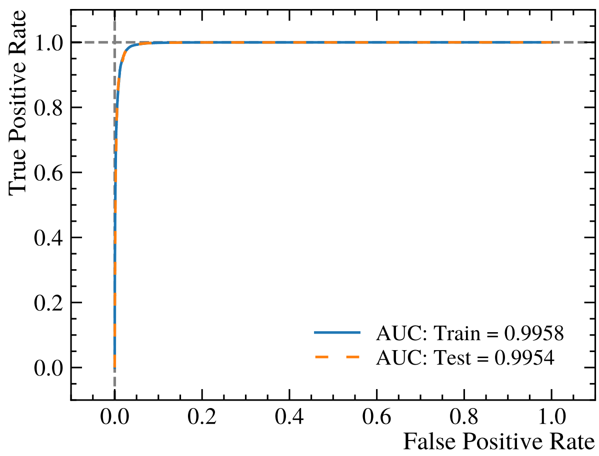

FIG. 2: ROC curve and AUC score for the trained BDT.

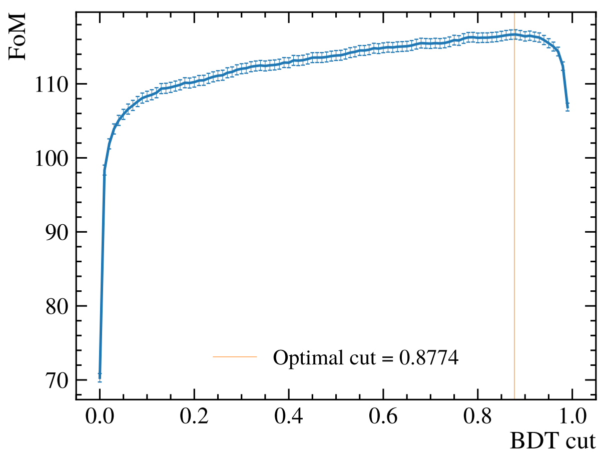

FIG. 3: Figure of Merit for the trained BDT.

C. Mass Fit results

1. J/ψ mode

After applying all the data selection, a total of 9467 entries remain in the J/ψ mode. The results of the mass fit in the J/ψ mode can be seen in figure 4 The yields of each contribution in this mode can be found in table I. Using equation 1, the expected yield in the rare mode can be computed to be 1.64 ± 0.22.

TABLE I: Yields of the different contributions in the J/ψ mode fit.

Channel	Yield
$B^+ \to \pi^+ J/\psi$	3620 ± 60
Combinatorial	2950 ± 90
$B^+ \to K^+ J/\psi$	720 ± 60
$B^{*0} \to K^{*0} J/\psi$	290 ± 170
$B^0 \to \rho^0 J/\psi$	1920 ± 150

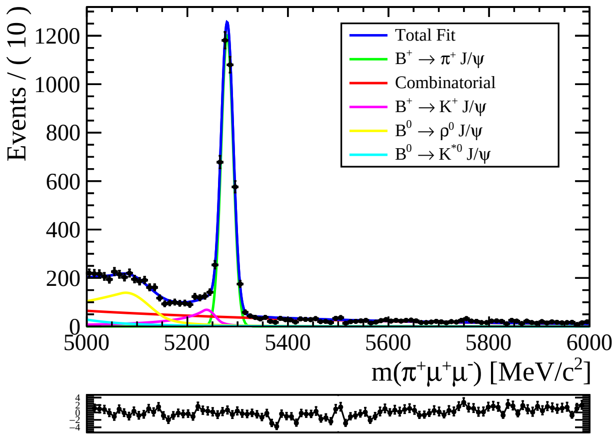

FIG. 4: Mass fit in the J/ψ mode after data selection.

2. *Rare mode*

After the data selection, a total of 135 events remain in the rare mode. The results of the mass fit in the rare mode can be seen in figure 5. The low number of remaining events causes the error bars to be much larger than in the J/ψ mode fit. The yields of each contribution in this mode can be found in table II. In this fit, no events have been found for the decay of interest. This result is compatible with the expected yield found with the J/ψ mode fit.

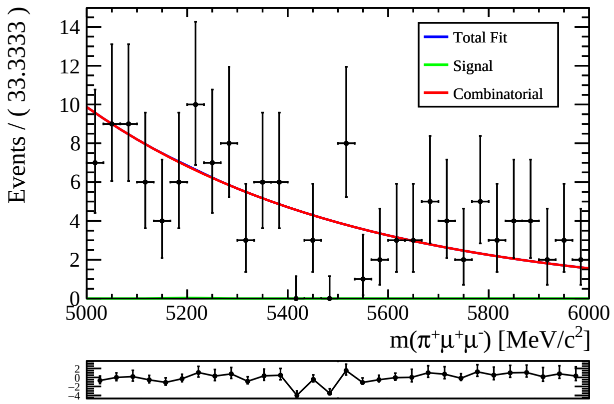

FIG. 5: Mass fit in the rare mode after data selection.

TABLE II: Yields of the different contributions in the rare mode fit.

Channel	Yield
$B^+ \to \pi^+\mu^+\mu^-$	0 ± 10
Combinatorial	135 ± 12

IV. Conclusions

In this work, a BDT has been trained and applied to data collected during Run 3 in the LHCb experiment, in order to perform a selection of $B^+ \to \pi^+\mu^+\mu^-$ events. On top of that, in order to approximate the expected event yield, the $B^+ \to \pi^+ J/\psi$ channel has been used.

By applying the BDT and using mass fits of the different expected background contributions in the J/ψ mode, the expected signal yield has been found to be 1.64 ± 0.22. Doing the mass fits directly in the rare mode has led to no events being found, which is compatible with what was expected.

The results don't show a significant number of events for the studied decay, but a number of steps could be taken in order to improve the selection. Firstly, computing the FoM in the rare mode instead of the J/ψ mode could provide better results. On top of that, an optimization of the hyperparameters of the BDT could be performed.

Acknowledgments

I would like to thank both my advisors, Carla Marín and Ernest Olivart for their help and guidance during this project. I have learned a lot with you during these months.

I would also like to thank my family and friends for their unconditional support and patience. Without you this work would not have been possible.

[1] CERN, "LHCb", URL: `https://home.cern/science/experiments/lhcb`, 2025.
[2] LHCb Collaboration. "The LHCb Detector at the LHC", *JINST*, vol. 3, p. S08005, 2008.
[3] LHCb Collaboration. "The LHCb Upgrade I", *JINST*, vol. 19, p. P05065, 2024.
[4] LHCb Collaboration, R. Aaij *et al.*, "First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \to \pi^\pm\mu^+\mu^-$ decay", *JHEP*, vol. 10, p. 034, 2015.
[5] E. Olivart, "Optimisation and tuning of a BDT. IDAML - Research Work", 2024.
[6] "ROOT", URL: `https://root.cern/manual/python/`, 2025.
[7] "Particle Data Group (PDG)", URL: `https://pdg.lbl.gov/index.html`, 2025.
[8] XGBoost, "Introduction to Boosted Trees" , URL: `https://xgboost.readthedocs.io/en/stable/tutorials/model.html`, 2022.
[9] T. Skwarnicki, "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances. PhD thesis, Cracow, INP, 1986.

Estudi de la desintegració $B^+ \to \pi^+\mu^+\mu^-$ amb l'experiment LHCb

Author: Marc Salvador Carreté
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisors: Carla Marín Benito and Ernest Olivart Pino

Resum: En aquest treball, es realitza un estudi de la desintegració $B^+ \to \pi^+\mu^+\mu^-$ per les noves dades obtingudes durant la *Run 3* de l'experiment LHCb. S'utilitza un Arbre de Decisió Potenciat (en anglès: *Boosted Decision Tree*, BDT) per seleccionar els esdeveniments corresponents a aquesta desintegració i es realitzen ajustos a les masses per tal d'extreure'n el nombre d'esdeveniments de senyal. La desintegració $B^+ \to \pi^+ J/\psi$ s'utilitza com a canal de control. A partir del nombre d'esdeveniments trobats al canal de control, s'espera un total de 1.64 ± 0.22 esdeveniments d'interès. No s'ha trobat senyal realitzant l'ajust directament, fet que és compatible amb el resultat esperat.
Paraules clau: Física de Partícules, Física d'Altes Energies, Anàlisi de Dades, Machine Learning, Boosted Decision Tree
ODSs: Indústria, innovació i infraestructures

Objectius de Desenvolupament Sostenible (ODSs o SDGs)

1. Fi de la es desigualtats		10. Reducció de les desigualtats	
2. Fam zero		11. Ciutats i comunitats sostenibles	
3. Salut i benestar		12. Consum i producció responsables	
4. Educació de qualitat		13. Acció climàtica	
5. Igualtat de gènere		14. Vida submarina	
6. Aigua neta i sanejament		15. Vida terrestre	
7. Energia neta i sostenible		16. Pau, justícia i institucions sòlides	
8. Treball digne i creixement econòmic		17. Aliança pels objectius	
9. Indústria, innovació, infraestructures	X		

Aquest TFG es pot relacionar amb l'ODS 9, concretament amb l'objectiu 9.5, ja que és un treball enfocat en aportar a la recerca científica, específicament en el camp de la física d'altes energies.

A. Preselection discrimination plots

In this appendix, two discrimination plots are shown.

The plot in figure 6 shows the discrimination plot for the variable "Mum_PROBNN_MU" which shows the probability that a particle reconstructed as a μ^+ is indeed a muon. It can be seen that applying a cut at "Mup_PROBNN_MU" > 0.8 removes a substantial amount of background events (in blue) without practically affecting the signal (in orange). In this case, this plot has been used to select a cut for the preselection.

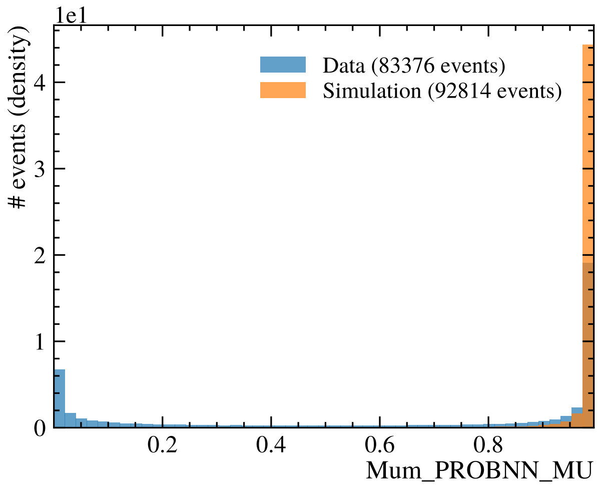

FIG. 6: Discrimination plot for the variable "Mum_PROBNN_MU".

The plot in figure 7 shows the discrimination plot for the variable "Bu_MAXDOCA". In this plot, the simulation (in orange) and data (in blue) distributions show a large discriminating power. Therefore, this variable has been used as a feature to train the BDT.

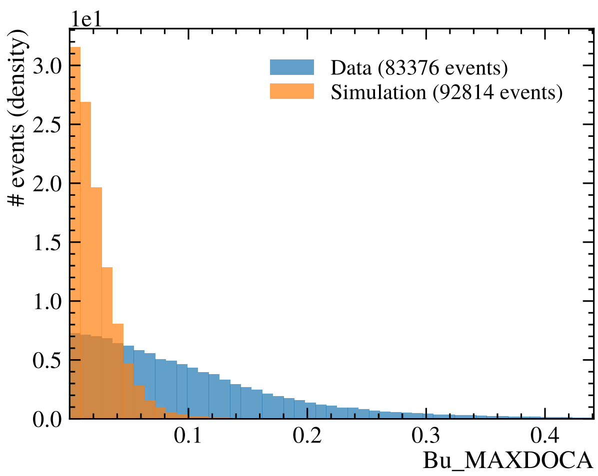

FIG. 7: Discrimination plot for the variable "Bu_MAXDOCA".

B. Preselection PROBNN cuts

A table with all the "PROBNN" cuts applied in the preselection can be found in table III. The cuts are divided by the reconstructed particle.

TABLE III: PROBNN cuts for each child particle.

Particle	Cut
π^+	Pip_PROBNN_GHOST < 0.2
	Pip_PROBNN_PI > 0.5
	Pip_PROBNN_K < 0.2
μ^+	Mup_PROBNN_E < 0.2
	Mup_PROBNN_GHOST < 0.2
	Mup_PROBNN_K < 0.2
	Mup_PROBNN_MU > 0.8
	Mup_PROBNN_PI < 0.4
μ^-	Mum_PROBNN_E < 0.2
	Mum_PROBNN_GHOST < 0.2
	Mum_PROBNN_K < 0.2
	Mum_PROBNN_MU > 0.8
	Mum_PROBNN_PI < 0.4

C. Feature importance plot

The following figure shows the F scores of all the selected features:

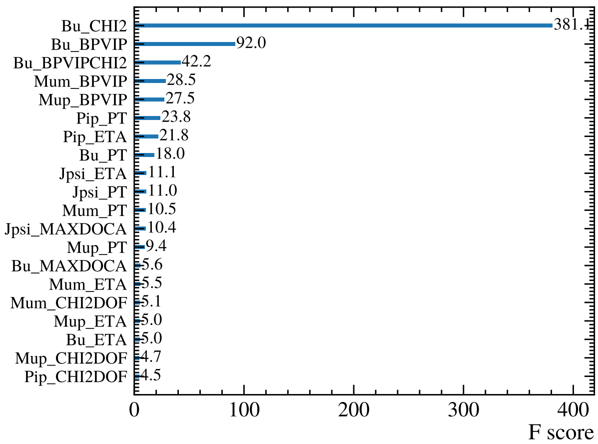

FIG. 8: Feature importance plot for the trained BDT.

In the figure, the following kinds of variables can be found:

- "_CHI2": A variable that assesses the quality of how the particle trajectories belong to a single event. It is the χ^2 value of the fit done to all the hits that form the trace.

- "_BPVIP": Best Primary Vertex Impact Parameter. It represents the closest approach of a particle trajectory to the primary vertex (the point where the initial collision of protons occurs).

- "_BPVIPCHI2": A variable similar to "_BPVIP" but accounting for experimental uncertainties. It is computed by fitting the primary vertex with and without the trace of interest and assessing how the χ^2 value changes.
- "_PT": The transverse momentum of the particle.
- "_ETA": The pseudorapidity (η) of the particle.
- "_MAXDOCA": The Maximum Distance of Closest Approach between pairs of children particles.
- "_CHI2DOF": A variable that assesses the quality of the reconstruction of particle traces. It is the value of the "_CHI2" variable divided by the number of degrees of freedom in the fit.

The particle that each feature refers to is specified at the beginning of the variable name.

D. S and B plots

The evolution of the expected signal (S) and background (B) yields for different BDT cuts can be found in figures 9 and 10, respectively.

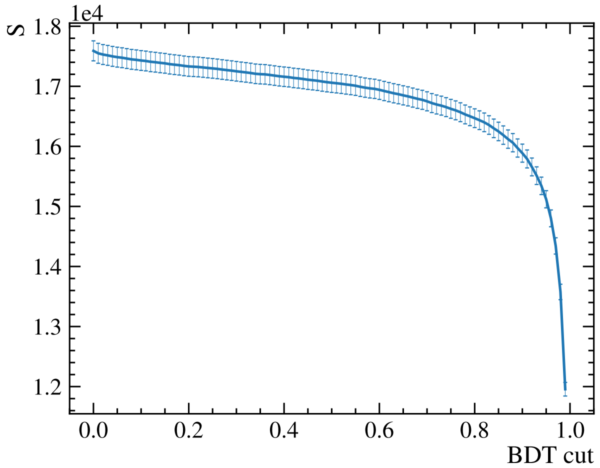

FIG. 9: Evolution of the signal yield for different BDT cuts.

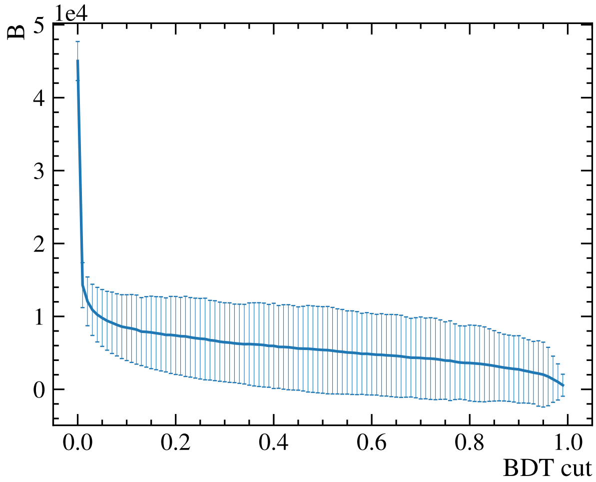

FIG. 10: Evolution of the background yield for different BDT cuts.