Flavour Anomalies with ML @LHC Run3

LIP Summer Internship Final Workshop

Gabriela Sousa, Teresa Escária

Supervisors: Alessio Boletti, Nuno Leonardo

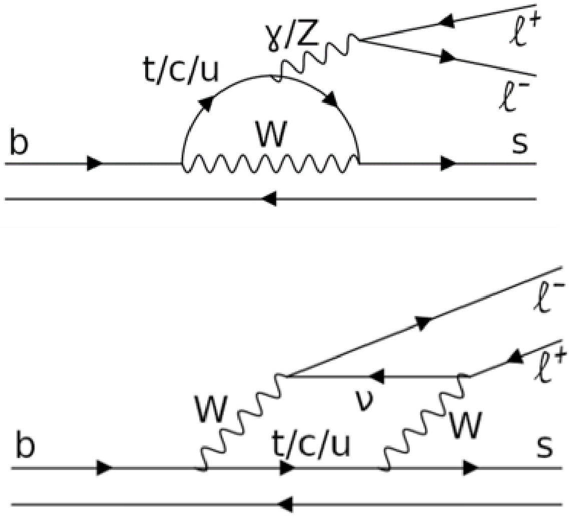
5/09/2025

Flavour Anomalies

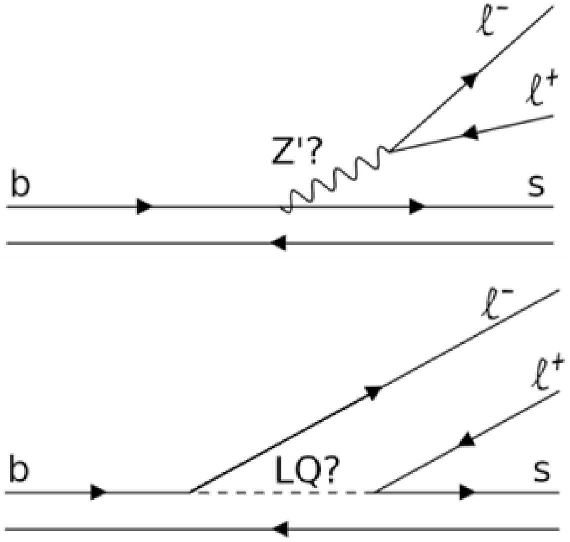
$b \rightarrow sl^+l^-$ transitions

- Flavour-changing neutral current (FCNC) decays
- Forbidden at tree level in the SM

Standard Model (SM)



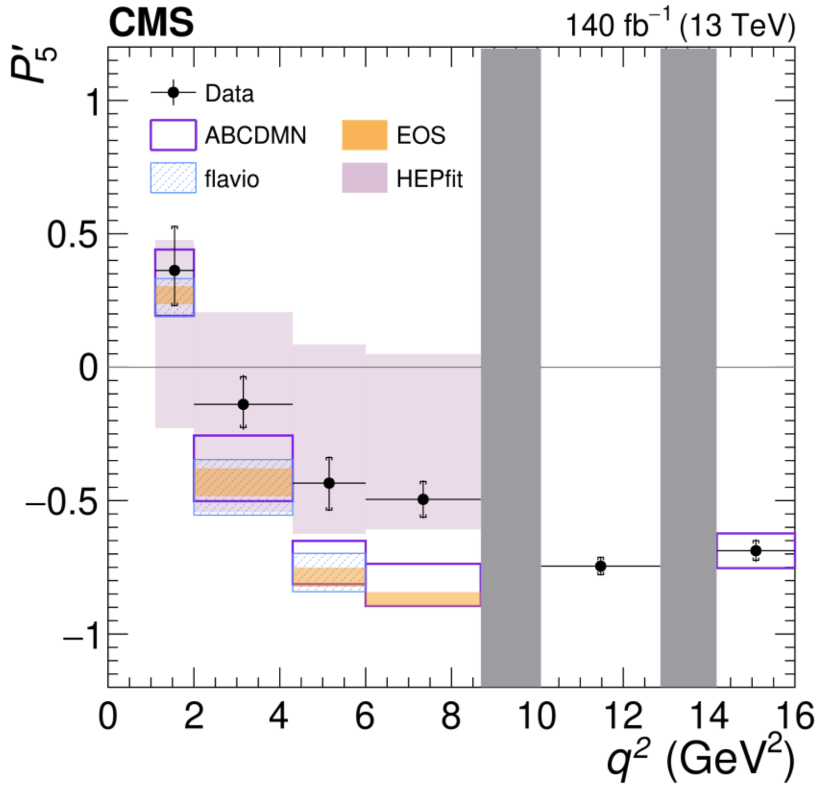
New Physics? (NP)



New Physics particles may:

- Enhance the decay rate;
- Alter the angular distribution of final-state particles;
- Couple differently to different leptons species (LFUV)

Angular Analysis



Recent (2025) result by CMS (*Phys. Lett. B 864 (2025) 139406*) based on Run2 data indicate deviation from a set of SM predictions
Corresponding deviations were previously reported also by LHCb

- Sensitive to New Physics through angular observables
- Famous anomaly: P5′
- Deviations in data vs SM predictions

Next steps:

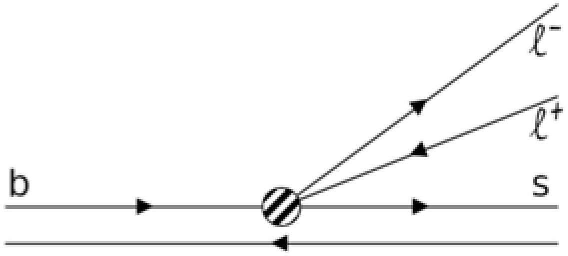
- Improve experimental precision usind latest LHC Run3 data;
- Improve theory calculations.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay

Non-resonant Channel

$B^0 \to K^{*0}\mu^+\mu^- \to K^+\pi^-\mu^+\mu^-$

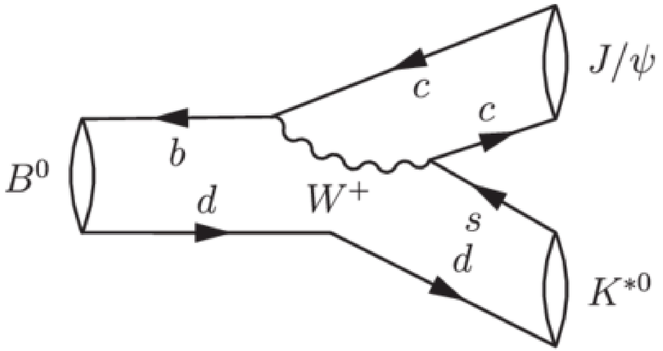
- Possibility of New Physics



Resonant Channel

$B^0 \to K^{*0}J/\Psi \to K^+\pi^-\mu^+\mu^-$

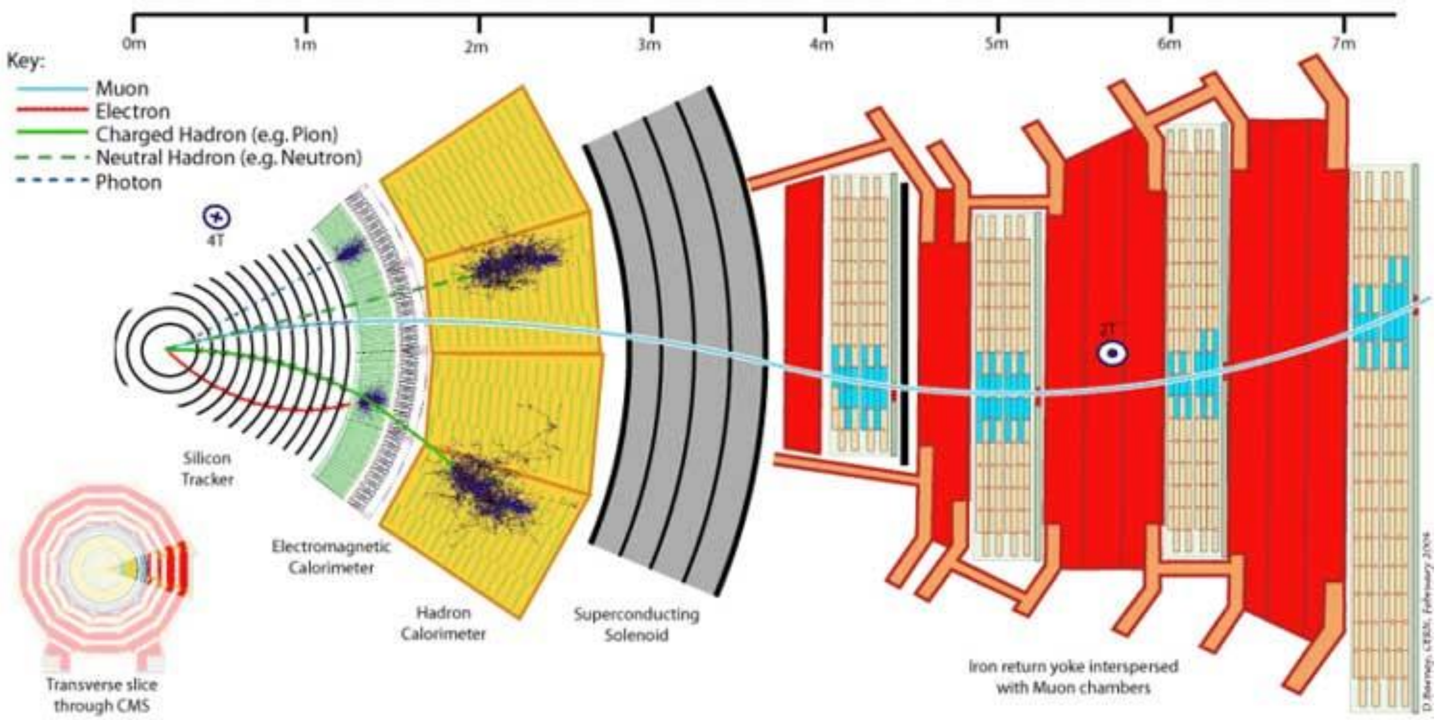
- Well understood within the SM
- Used as a control channel
- Branching fraction is precisely measured
- Channel that we focused on



Compact Muon Solenoid CMS

CMS

General purpose detector at LHC



Two Key Subsystems:

Silicon Tracker

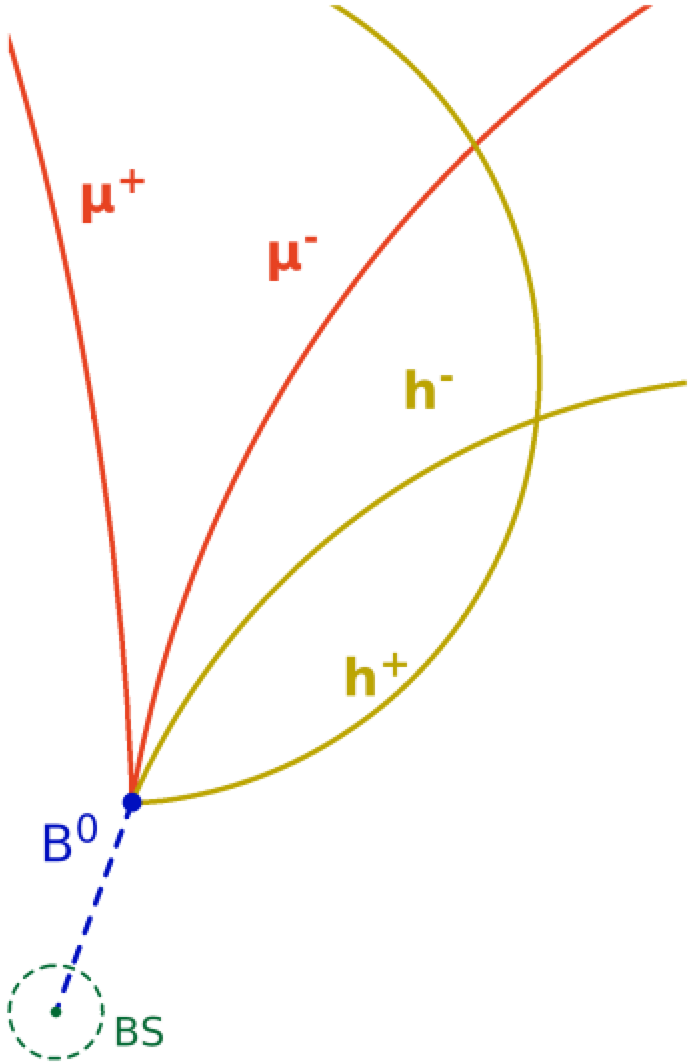
- Reconstructs charged particle trajectories, by collecting discrete hits
- Can fit precise tracks, determine momentum and reconstruct vertices for our desired decay

Muon Chambers

- Detect muons exiting the calorimeters

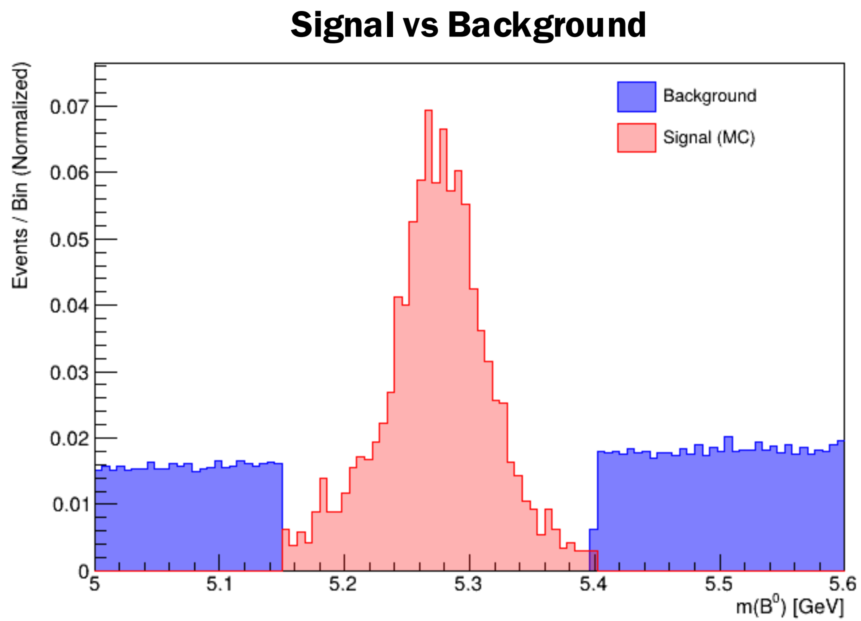
Candidate Reconstruction

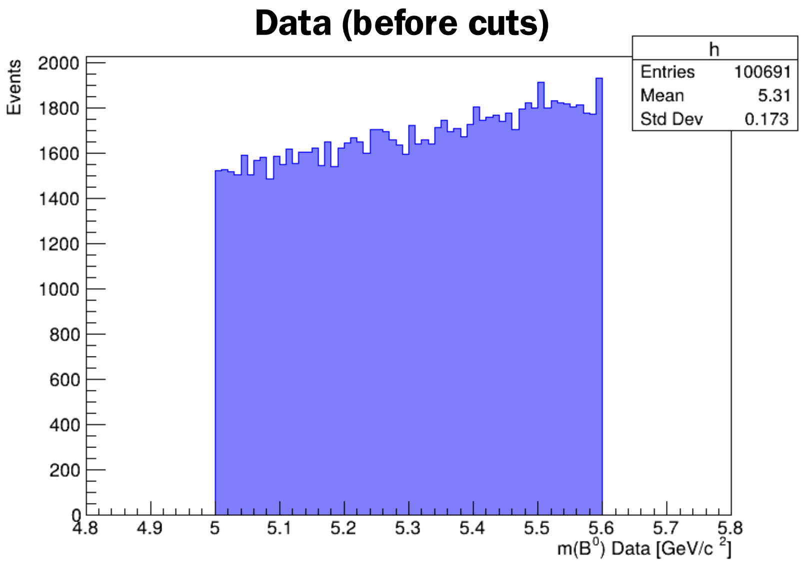
- **Candidate building:** combineboth muons with two opposite-sign tracks; fit a common displaced vertex; form K ^{0}from the hadron pair and $B^0$from K^0+$\mu^+\mu^-$; apply basic track/ID, vertex-probability, and displacement cuts.
- Both $B^0 \rightarrow \boldsymbol{K}^+\boldsymbol{\pi}^-\mu^+\mu^-$ and $\bar{B}^0 \rightarrow \boldsymbol{K}^-\boldsymbol{\pi}^+\mu^+\mu^-$ decays occur
- CMS has no hadron PID:
  - the K/π combination closest to $m_{PDG}(K^{*0})$ is associated to the event (a fraction of events can be mis-tagged)
- B^0 has a long lifetime, leading to a secondary vertex displaced from the primary
- **Combinatorial Background:**
  - Tracker and muon system may reconstruct uncorrelated tracks that fake B^0 candidates
  - Contributes to a smooth, non-peaking background in the B^0 mass distribution
- **Physics Background:**
  - Other real B decays with similar final states can mimic the signal

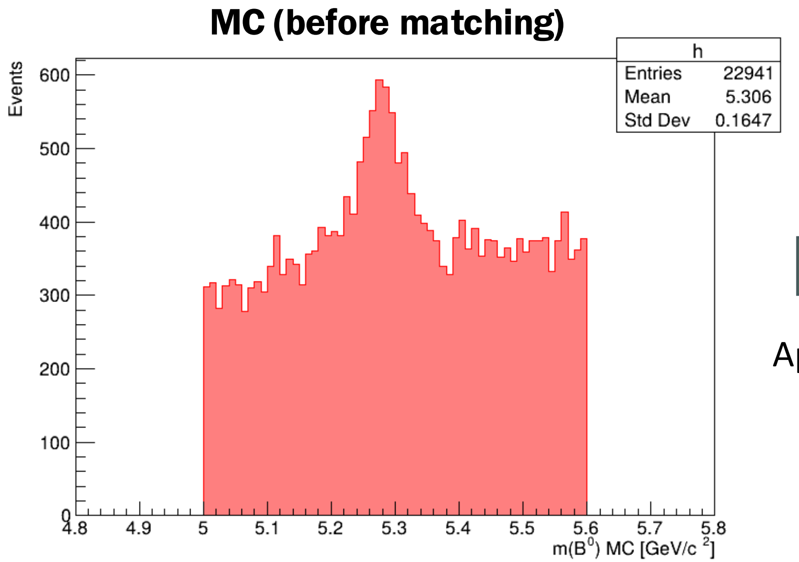


Data Preparation

- **Signal sample:** from Monte Carlo truth-matched
- **Background sample:** from Data sidebands
  - ~3.5 σ away from mean of the signal peak

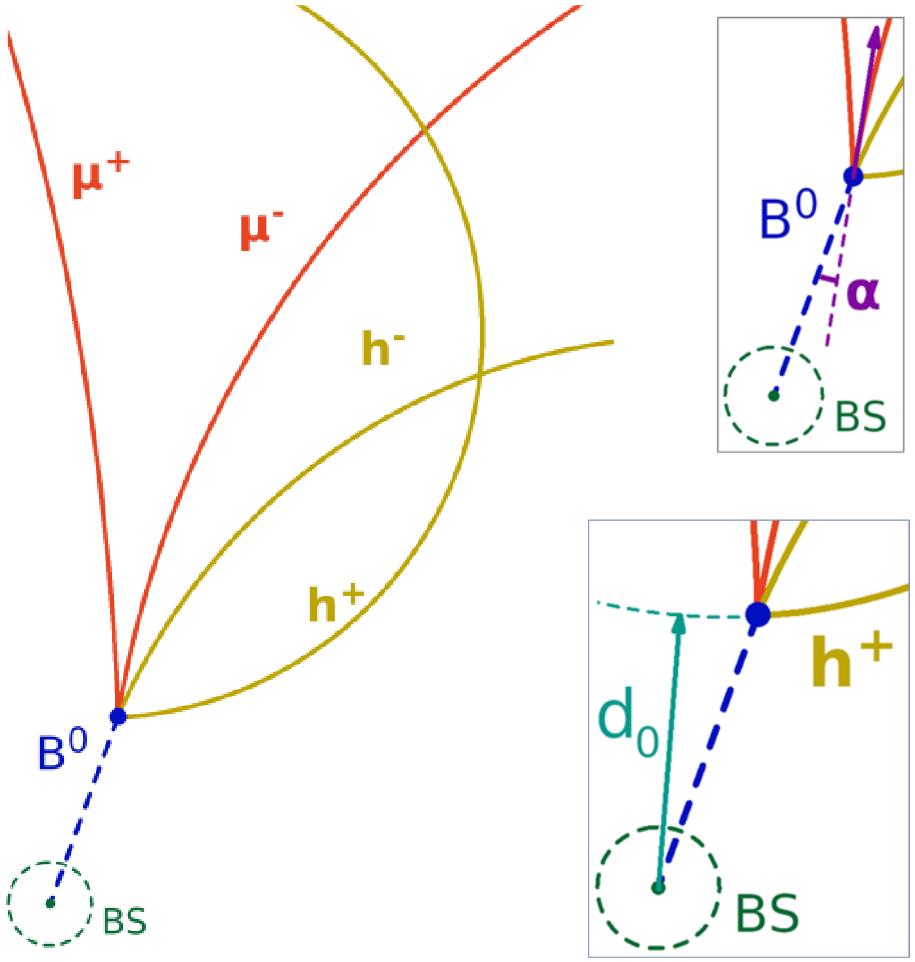






Apply Matching and Cuts

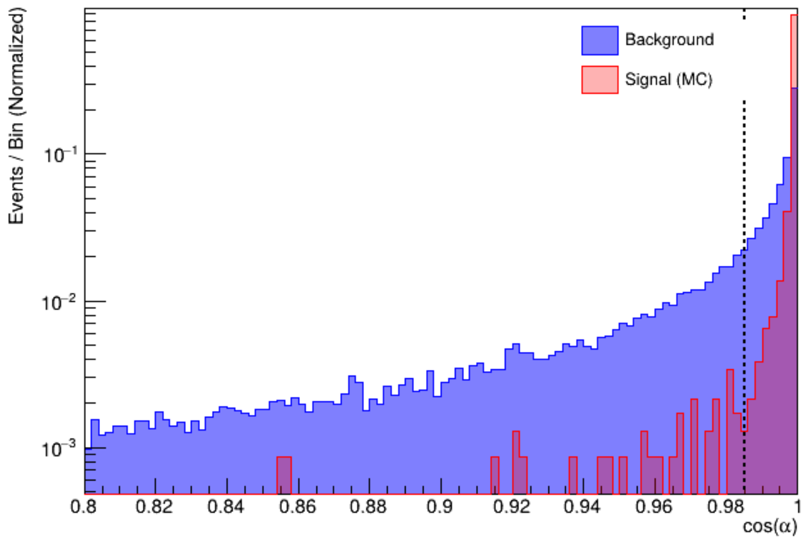
Discriminating Variables

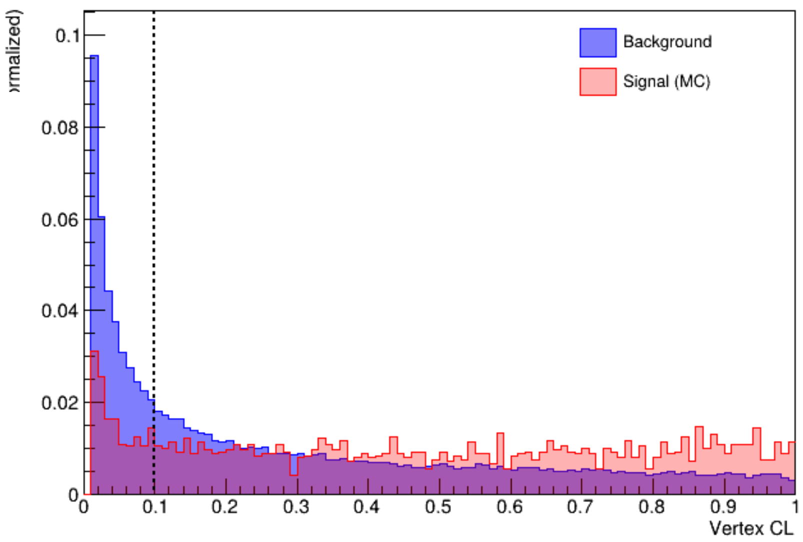


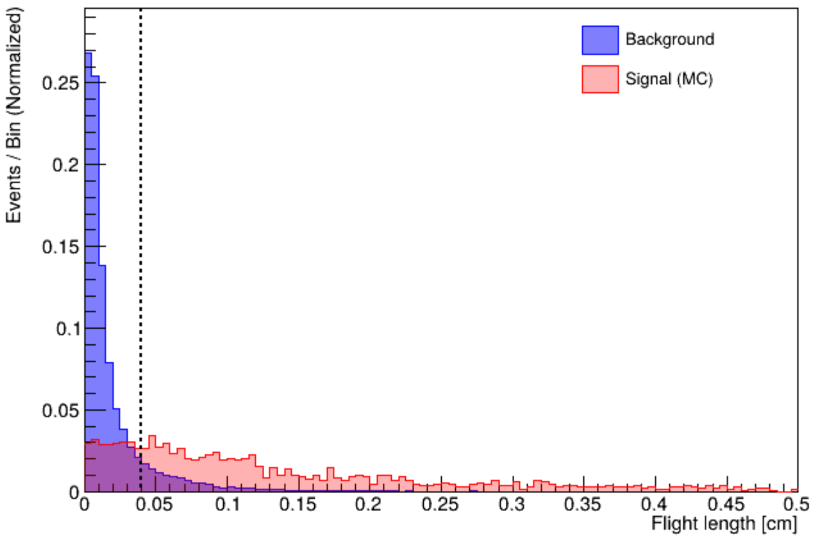
There are several kinematic and geometric variables that can help:

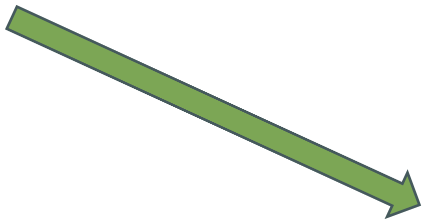
- **Flight Length:** distance between the primary vertex and the secondary vertex (B^0 decay vertex)
- **Cos(α):** cosine of the angle between the reconstructed B^0 meson momentum and its flight direction
- **Vertex CL:** measure of how likely the four tracks come from the same decay point
- **DCA:** how close a track comes to the beamline
- **p_T, η, ϕ:** transverse momenta, pseudorapidities and azimuthal angle of various variables

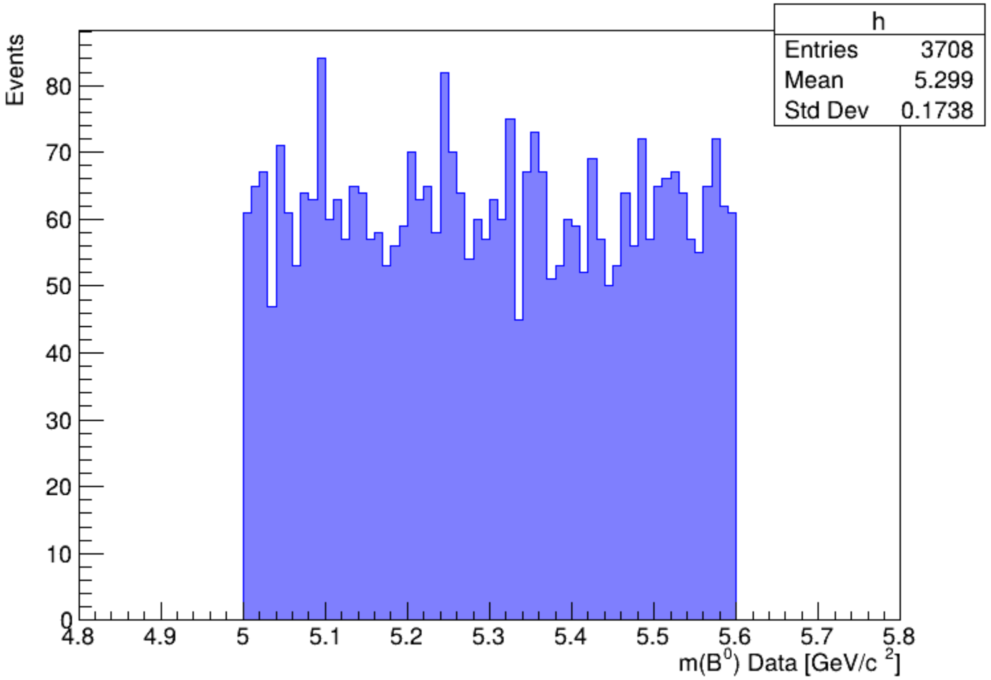
Applying Cuts





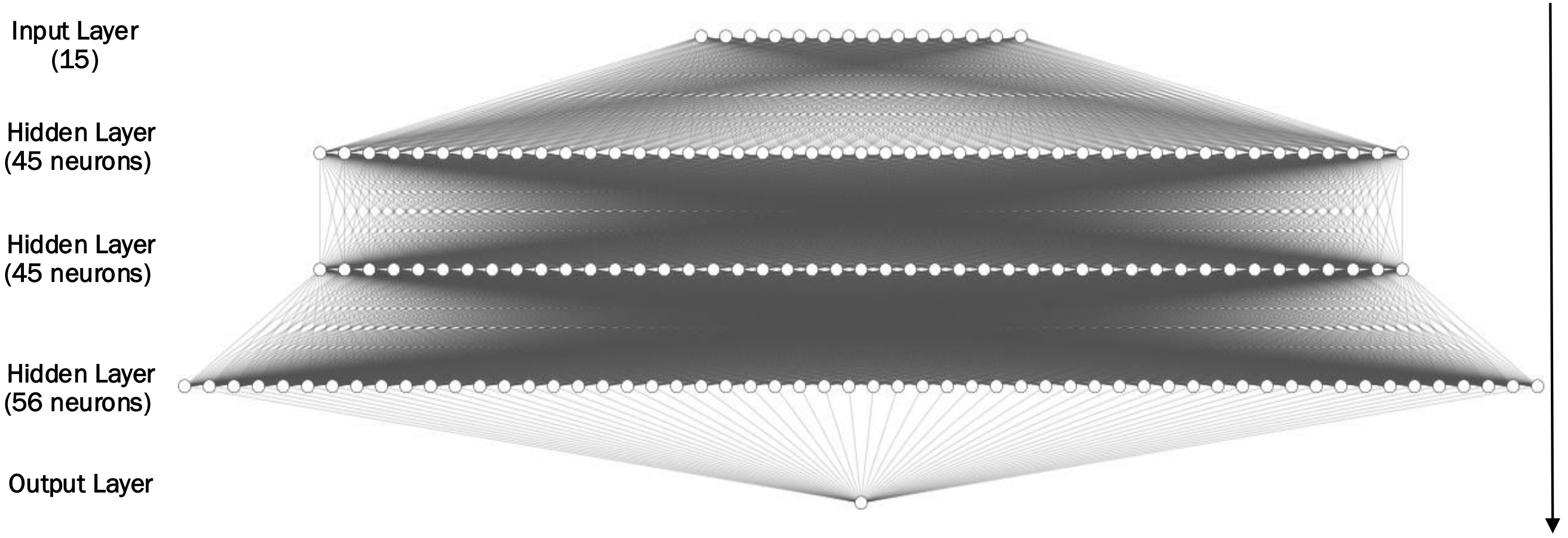




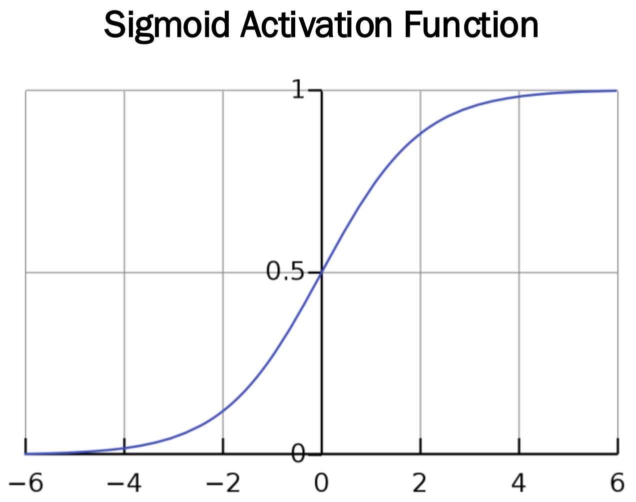


Neural Network Classifier

- Architecture:
  - Three fully connected hidden layers ;45, 45 and 56 neurons, respectively (optimized with Optuna); with ReLU activation
  - Dropout layers (10%) to prevent overfitting
  - Sigmoid output for binary classification (signal vs background)



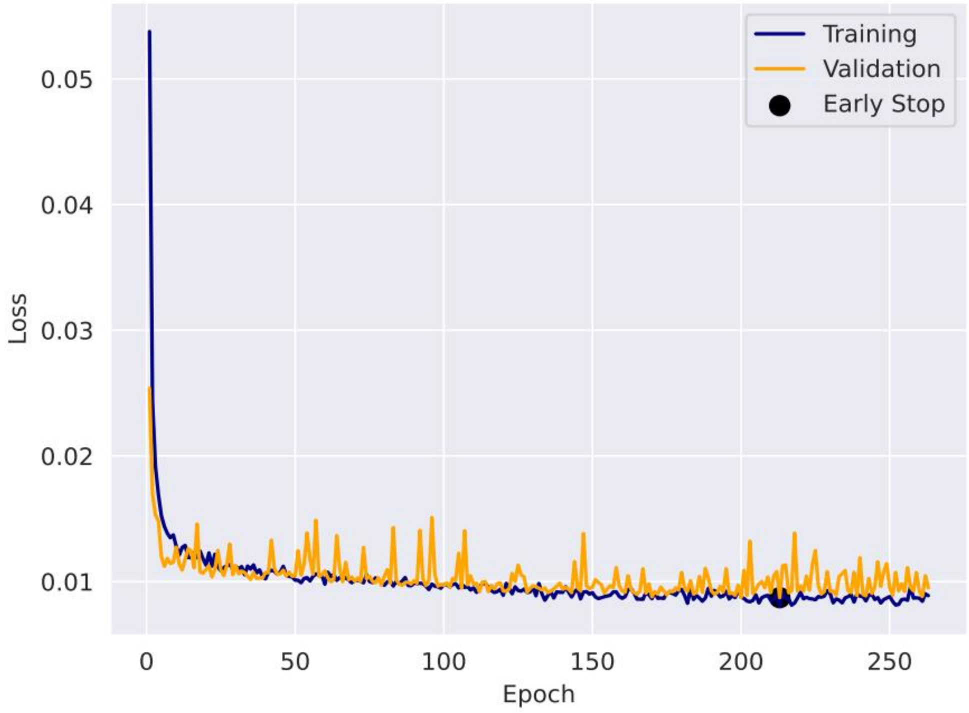
$$\frac{1}{1+e^{-x}}$$



Loss Function and Class Imbalance

- **BalancedLoss (Weighted BCE):**
  - Extends BCE by introducing class weights via alpha
  - Helps address class imbalance by up-weighting underrepresented class
  - Ensures the model doesn't become biased toward predicting background only

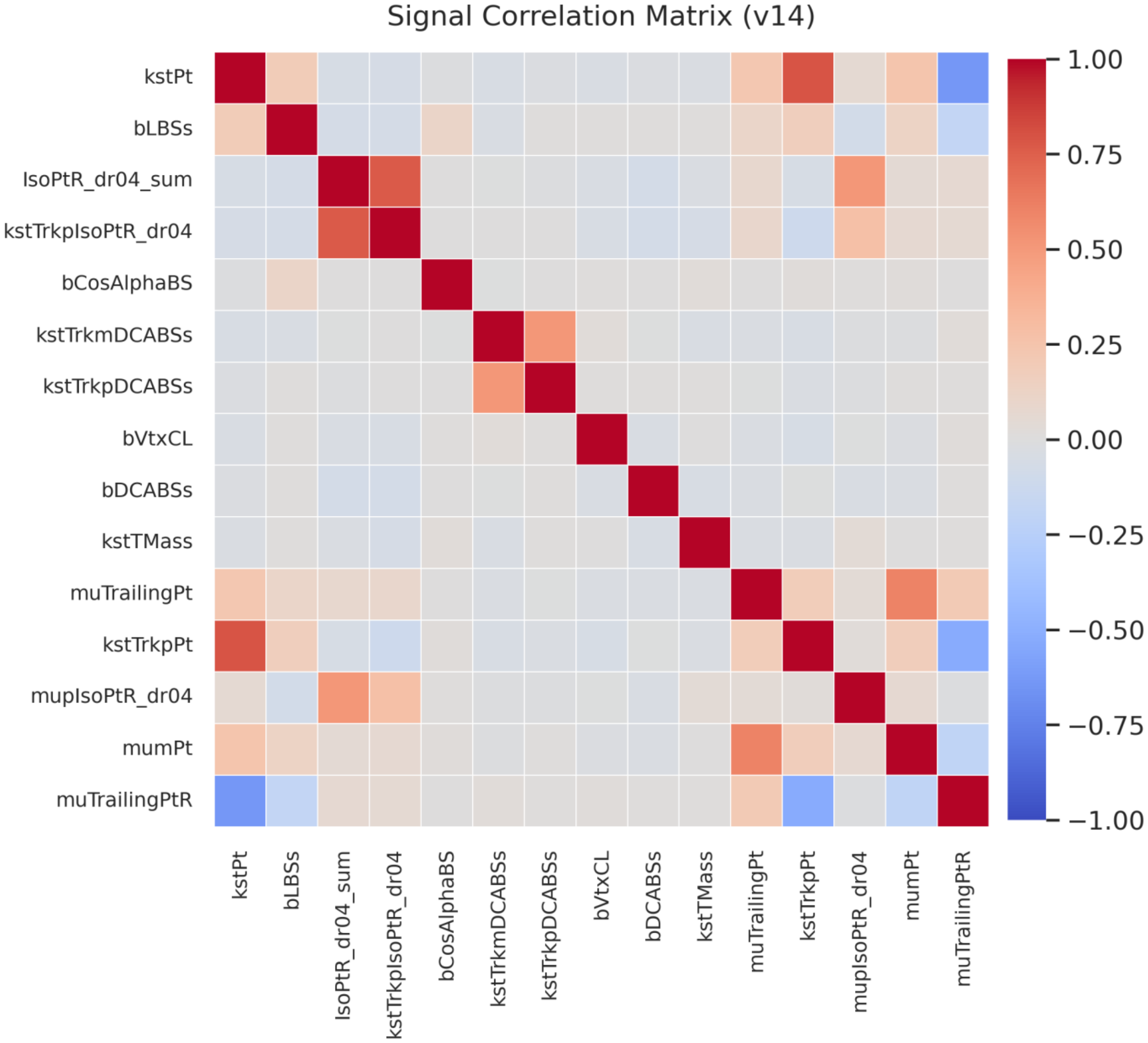
$$L_{Balanced} = -\left[\alpha_1 y \log(\hat{y}) + \alpha_0 (1-y) \log(1-\hat{y})\right]$$



Feature selection

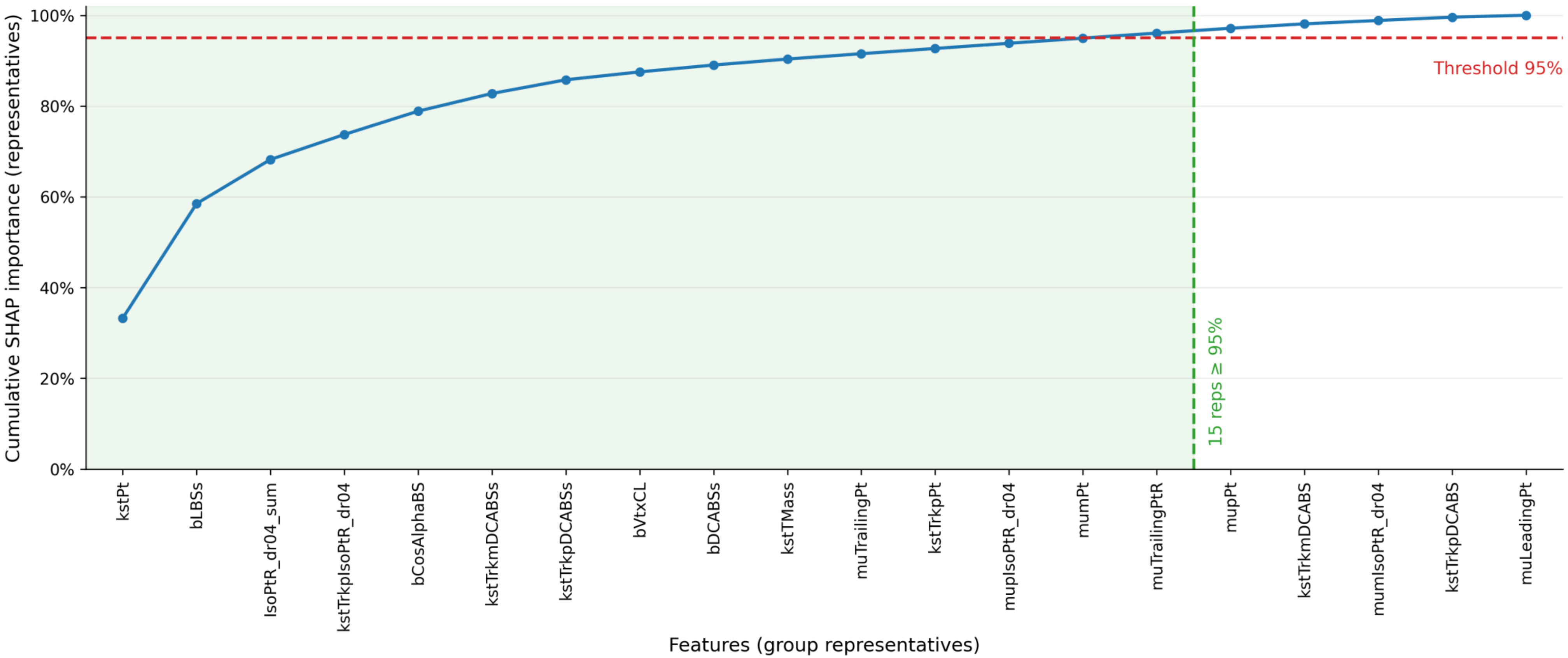
1. Correlation Matrix:

- Visualises how strongly variables are correlated (−1 to +1).
- Strong correlations mean variables carry overlapping information.
- We use this to identify and remove redundant features before training. (meter 14)

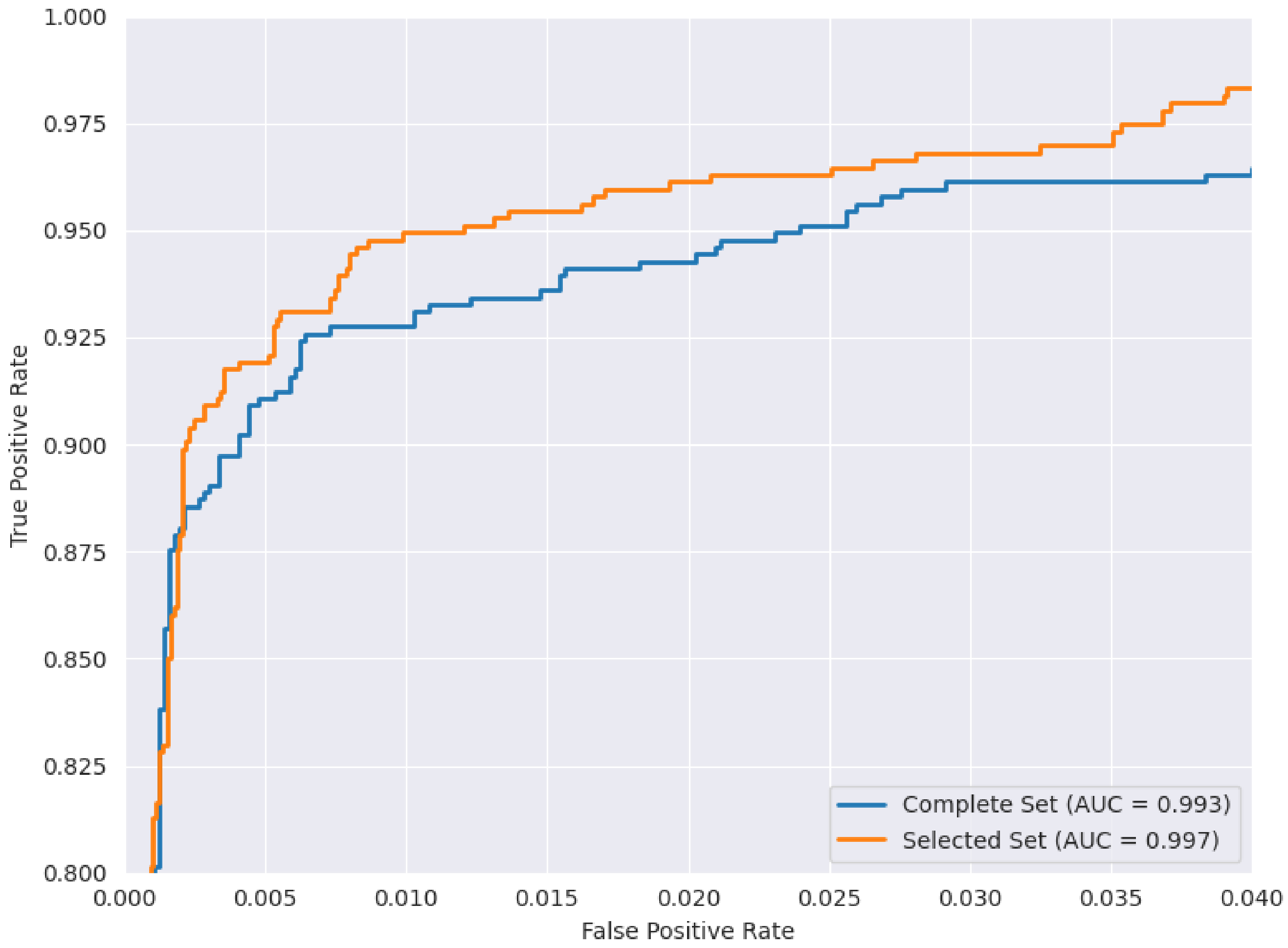


Feature selection

2. Cumulative SHAP Importance:



Model Evaluation



Used **Receiver Operating Characteristic (ROC)** curves to evaluate classifier performance

Plots True Positive Rate (TPR) vs. False Positive Rate (FPR) across all thresholds

- **TPR** = signal efficiency;
- **FPR** = background misclassification rate

Area Under the Curve (AUC) gives a single performance metric

- AUC = 1 perfect classification
- AUC = 0.5 random performance

Neural Network Output

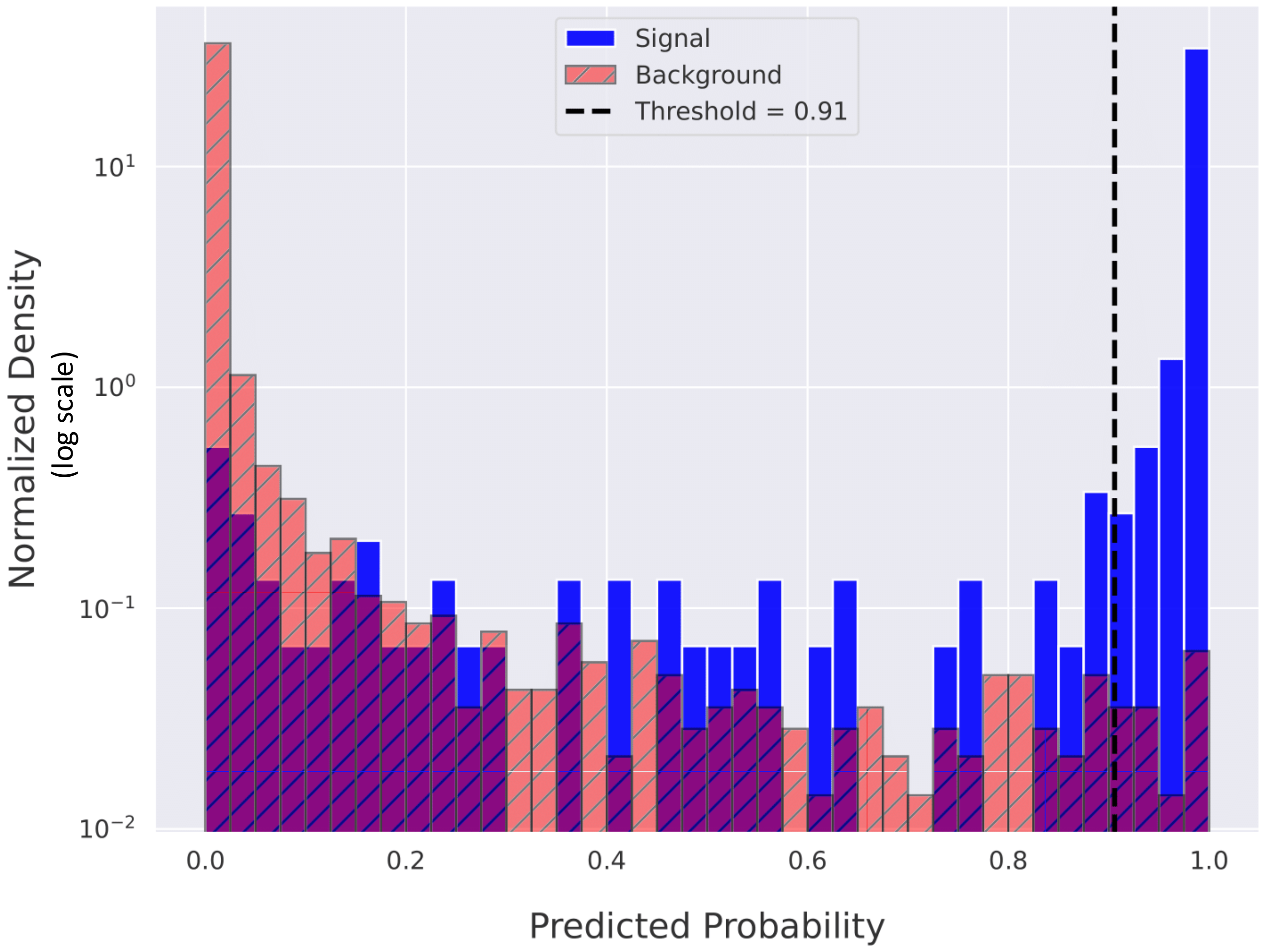
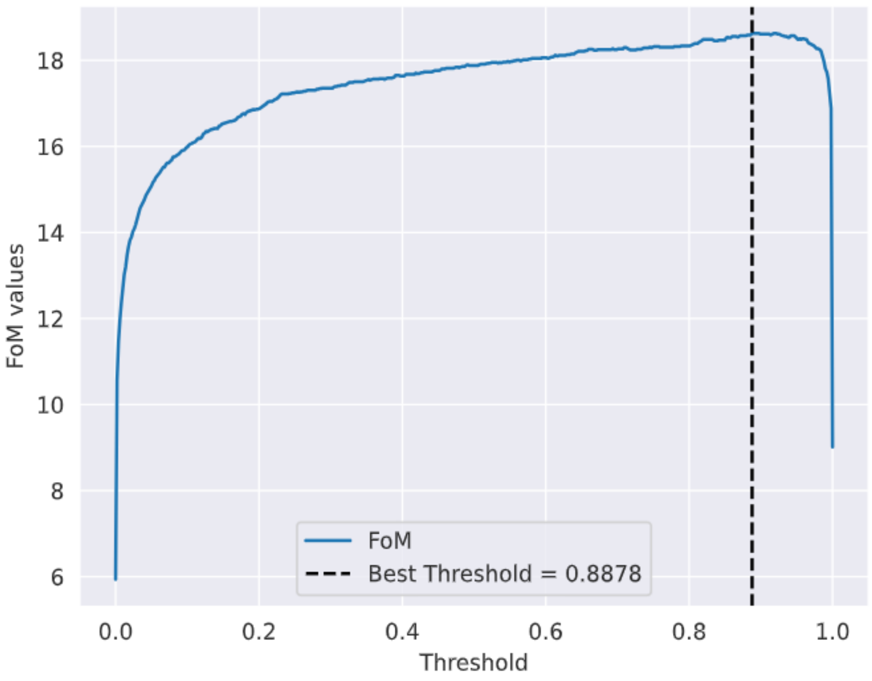


Figure of Merit

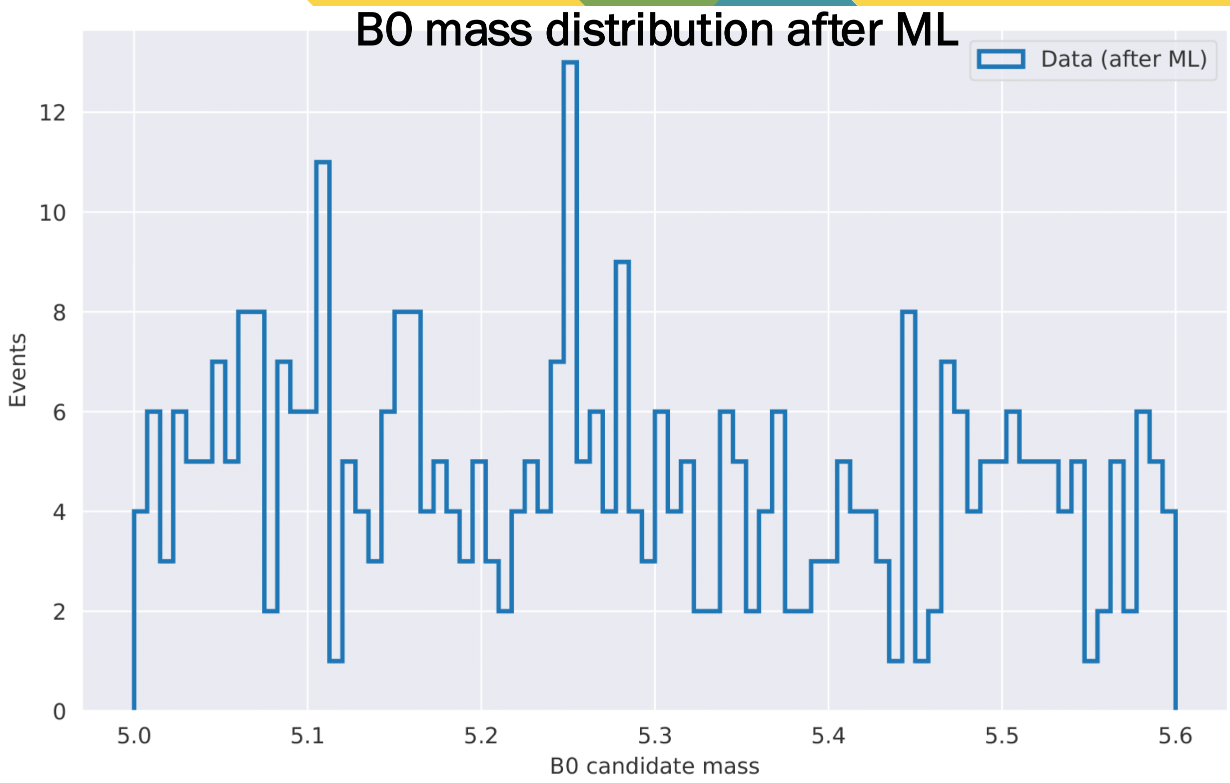
$$FOM = \frac{S}{\sqrt{S+B}}$$

- S/B: signal and background event counts above the threshold after applying estimated scaling



Summary

- Comparisons between **signal and background samples** from MC and data.
- Trained a **neural network classifier** with good separation power.
- Identified the **most relevant variables** (via correlation matrix and SHAP).
- Current dataset: **no visible B0 mass peak** → likely dominated by background noise.
- Next steps:
    - ○ process a larger data sample;
    - ○ apply trained model to the larger dataset;
    - ○ extract the B0 resonant signal (yield, angular variables);
    - ○ extend the to the non-resonant channel.



Backup Slides

Variables

- **muLeadingPt / muTrailingPt** – transverse momentum of the leading and trailing muons.
- **muIsoPR_dR04 / muIsoPR_dR04_sum** – isolation variables: how much extra activity (tracks or energy) is found within a cone of ΔR=0.4 around the muon.
- **kstTrkpPt / kstTrkmPt** – transverse momentum of the kaon and pion tracks.
- **kstTrkpDCABSs / kstTrkmDCABSs** – significance of the distance of closest approach of the kaon/pion to the beam spot (a measure of track pointing quality).
- **kstTMass** – reconstructed invariant mass of the K∗0
- **bVtxCL** – vertex fit confidence level of the B0 candidate. Higher means a better fit to a common vertex.
- **bLBSs** – significance of the decay length (how far the B0 secondary vertex is displaced from the primary vertex, normalized by its uncertainty).
- **bCosAlphaBS** – cosine of the pointing angle between the B0 momentum and the vector from primary to secondary vertex (close to 1 for well-reconstructed decays).
- **bDCABSs** – impact parameter significance of the B0 candidate relative to the beam spot.
- **kstKpIsoPR_dR04 / kstKpIsoPR_dR04_sum** – isolation variables for the kaon from the K∗0.

Threshold Optimisation

- Signal (MC) and background (data) do not correspond to same collision statistics:
  - Would expect MC to have the same amount of signal as dataset;
  - Need to scale S and B in FoM accordingly;
- Used **Youden's J statistic** to guide initial threshold choice:

$$J = TPR - FPR$$

- Maximising J gave initial optimal threshold
- Applied model and initial threshold to full dataset to derive background fraction under the peak
  - Np (backgorund events under peak region); Nl and Nh (backgrounf events on left and right sidebands respectively)
- Applied model and initial threshold to MC to get number of predicted signal events after selection
- Scaling applied in **Figure of Merit:**

$$w_B = \frac{N_p}{N_l + N_h}$$

$$w_S = \frac{S}{S_{MC}} = \frac{x * N_p}{S_{MC}}$$

$$FOM = \frac{S * w_s}{\sqrt{S * w_s + B * w_B}}$$