Validating Explainable AI Techniques through High Energy Physics Data

Mariagrazia Monteleone[1], Francesca Camagni[1], Simone Gennai [2], Pietro Govoni [2,3], Chiara Paganelli[1]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

[2] INFN - Sezione di Milano Bicocca, Italy

[3] Università degli Studi Milano Bicocca, Milano, Italy

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Background

Within the Compact Muon Solenoid (CMS) Collaboration, various Deep Neural Networks (DNNs) and Machine Learning (MLs) approaches have been employed to investigate the production of a new massive particle that undergoes decay into Higgs Boson pairs (HH) which further decay into a pair of b-quarks and a pair of tau leptons and discriminate the HH signal from the backgrounds.

However, the mentioned models which are employed are often **complex** and considered black boxes, making it challenging **to interpret** how the task was performed and the data analysis review process.

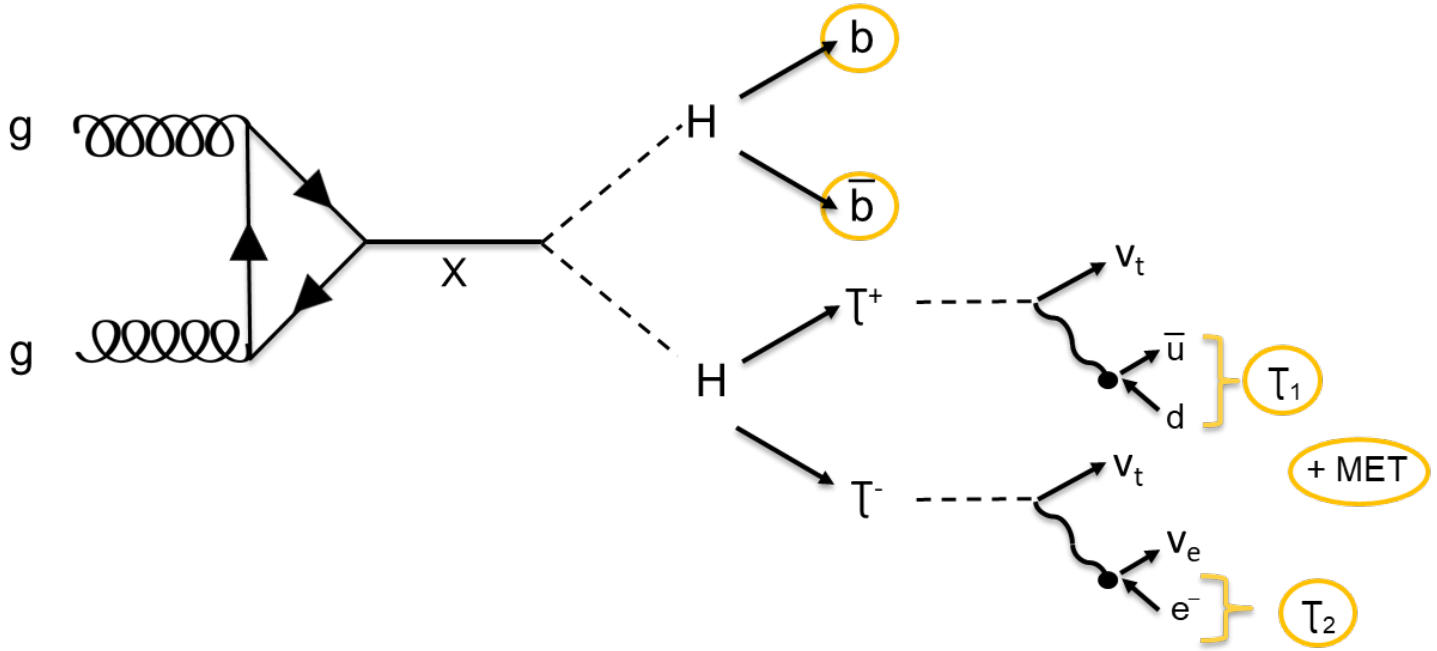

Fig. 1: Decay process outline

Aim of the work

This work aimed therefore to provide **a better understanding** of how the models work by validating an established **Explainable Artificial Intelligence (XAI)** technique such as SHapley Additive exPlanations (SHAP) [1], aiming for **more interpretable, trustworthy models** and predictions.

Workflow

A data pre-processing pipeline was established to select important features Recursive Feature Elimination based on SHAP values. This led to fine-tuning XGBoost for a classification task, whose features were compared with PCA results for validation and interpretation.

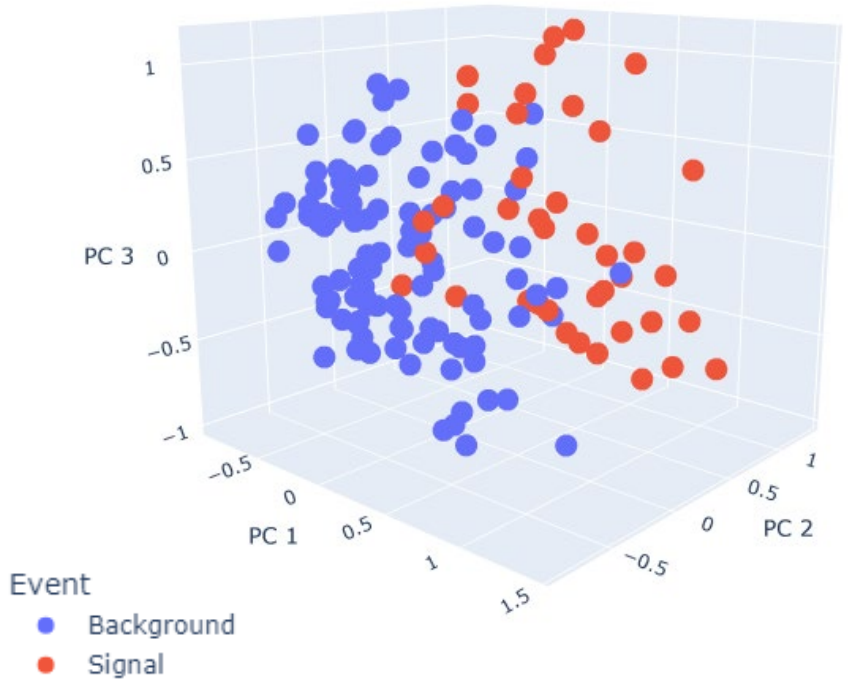

Fig. 2: PCA data visualization

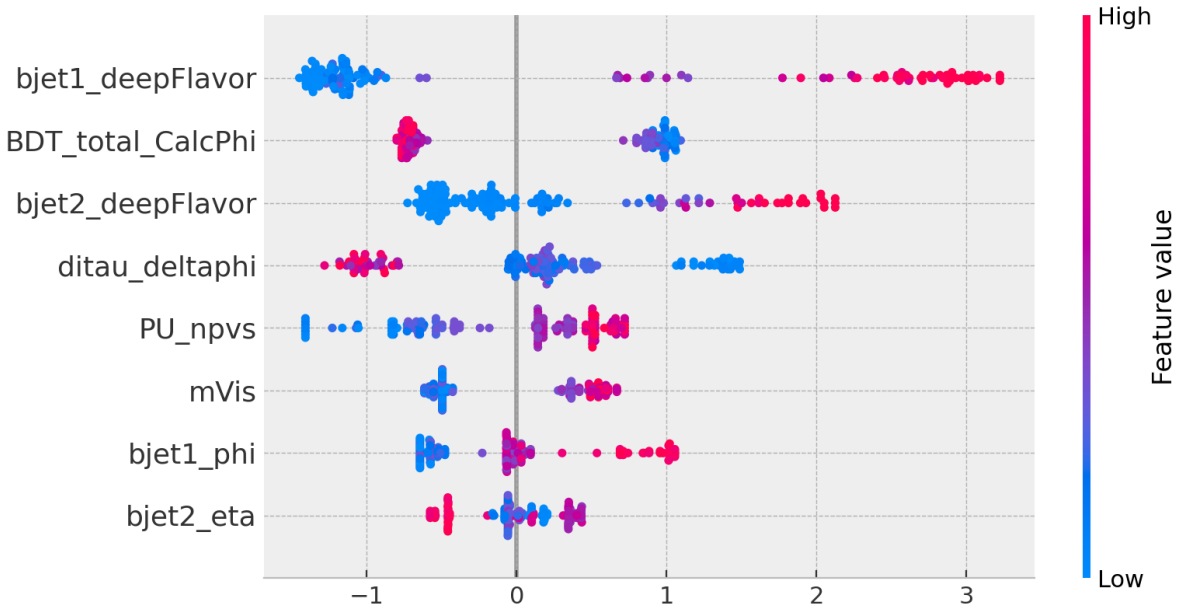

Fig. 3: SHAP value, impact on model output

Conclusions

The results obtained with SHAP and PCA agreed on the importance of some of the features , the combination of the two techniques confirmed the reliability of SHAP as an established tool, but also the potential of High Energy Physics (HEP) domain as a **new technical validation tool.**

Mariagrazia Monteleone, MSc