AI/ML Tools for analysis of hadron spectroscopy data

13th International Workshop
on the Physics of Excited Nucleons

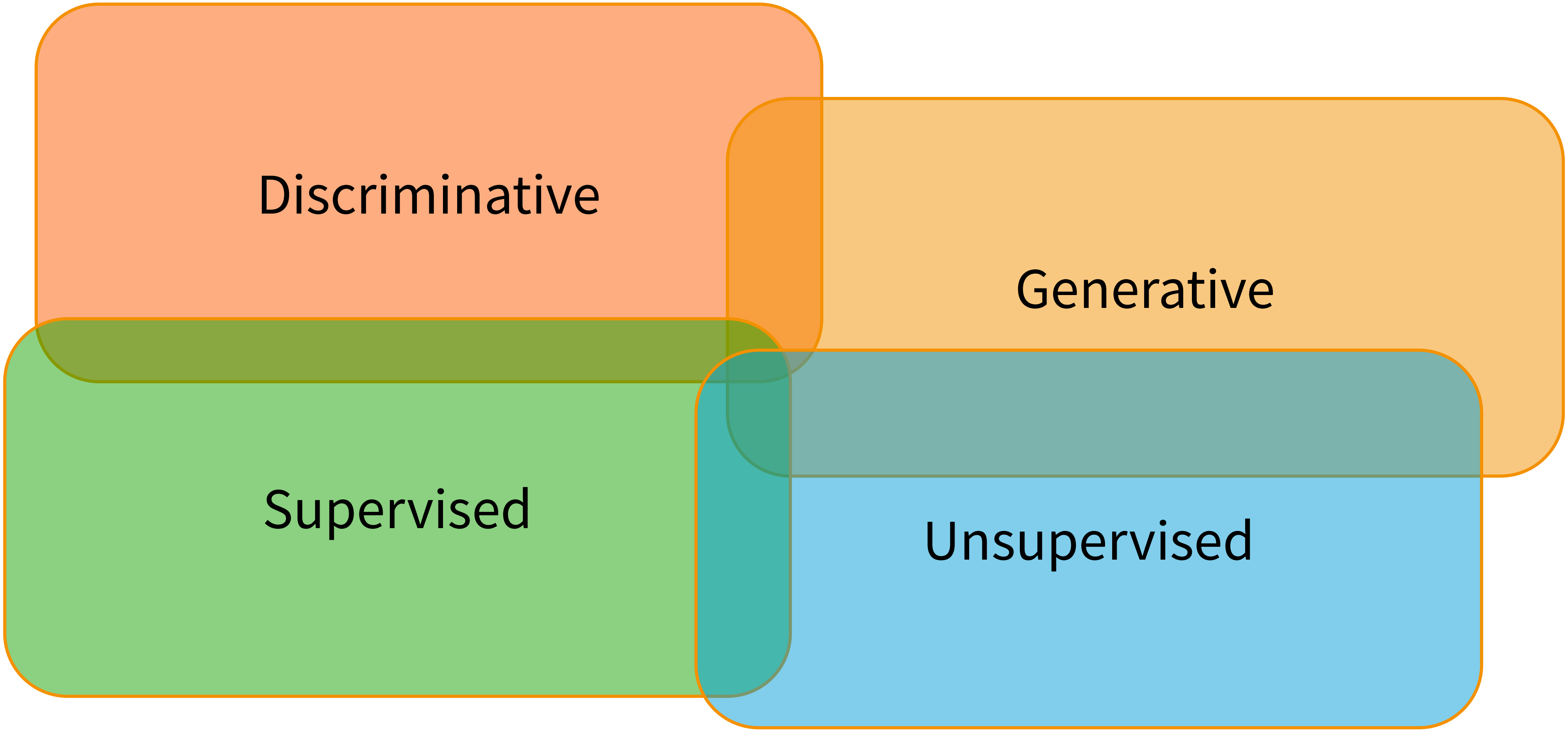
Santa Margherita Ligure, Oct 17-21, 2022

Łukasz Bibrzycki
Pedagogical University of Krakow
on behalf of the JPAC collaboration

Outline

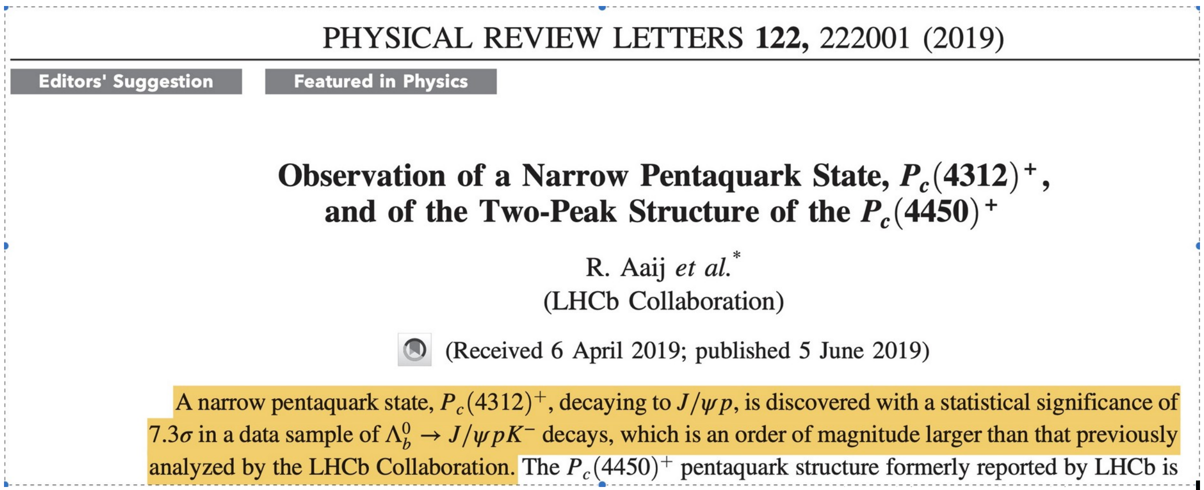
- Motivation and Physical model
- ML model
- Feature refinement
- Model predictions and explanation
- Going beyond discriminative model
- Summary

Types of ML models



Motivation

Plethora of potentially multiquark states observed in last decade



PHYSICAL REVIEW LETTERS **122,** 222001 (2019)

Editors' Suggestion | Featured in Physics

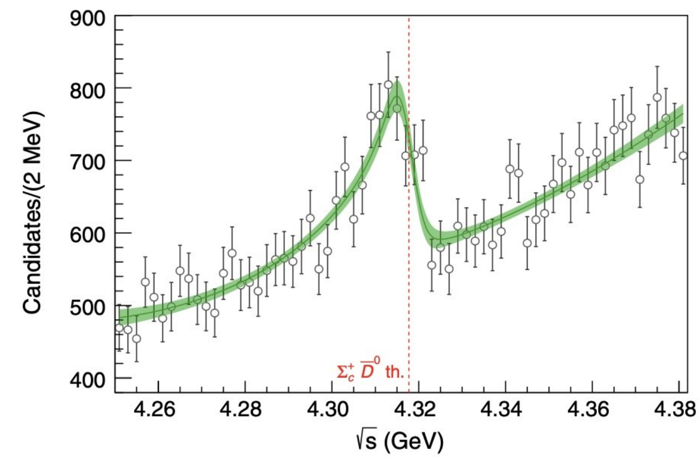
Observation of a Narrow Pentaquark State, $P_c(4312)^+$, and of the Two-Peak Structure of the $P_c(4450)^+$

R. Aaij *et al.**
(LHCb Collaboration)

(Received 6 April 2019; published 5 June 2019)

A narrow pentaquark state, $P_c(4312)^+$, decaying to $J/\psi p$, is discovered with a statistical significance of 7.3σ in a data sample of $\Lambda_b^0 \to J/\psi p K^-$ decays, which is an order of magnitude larger than that previously analyzed by the LHCb Collaboration. The $P_c(4450)^+$ pentaquark structure formerly reported by LHCb is

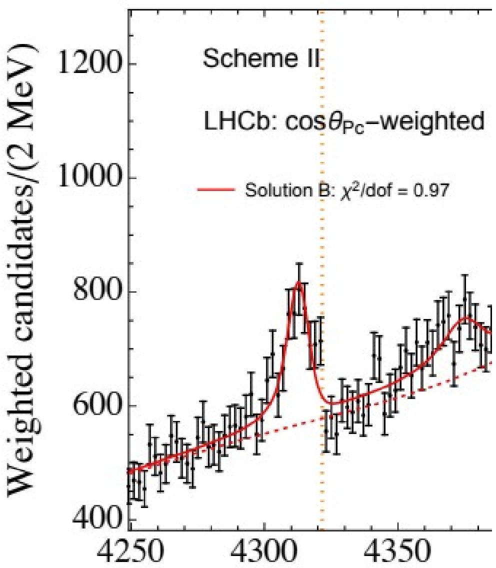
Possible interpretation as *duuc$\bar{c}$* pentaquark



Intensity in the $P_c(4312)$ neighbourhood and the JPAC fit *C. Fernandez-Ramirez Phys.Rev.Lett. 123 (2019) 9, 092001*

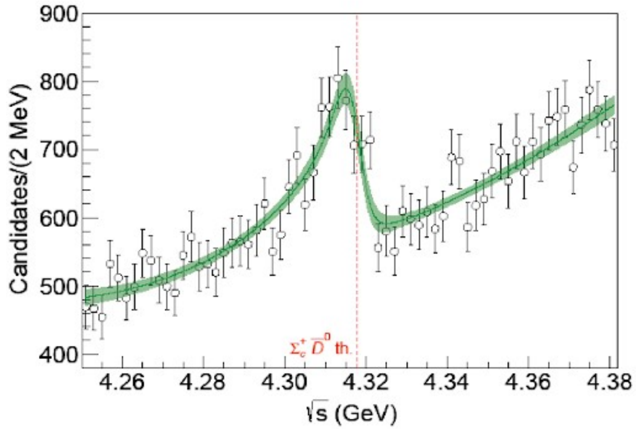
- There is a close relation between QCD spectrum and the analytic structure of amplitudes (production thresholds → branch points, resonances/bound states → poles)
- Currently this relationship is impossible to derive from first principles of QCD (top down approach)
- One can utilize the general properties of amplitudes, like unitarity, analyticity or crossing symmetry, but then model parameters must be derived from data – bottom up approach

Discrepant interpretations of the $P_c(4312)$ nature



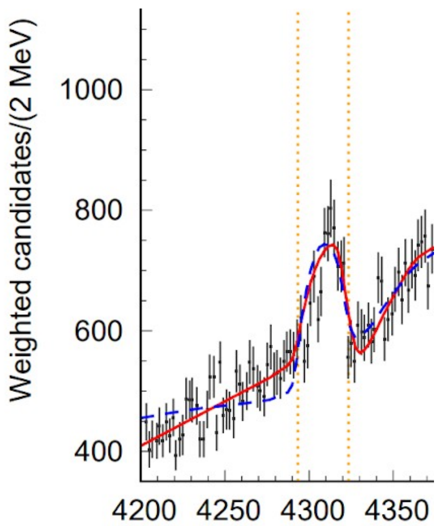
Molecule

Du et al., 2102.07159



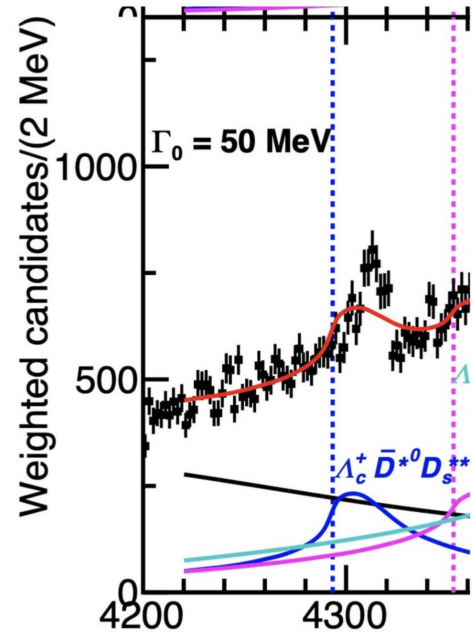
Virtual

C. F-R et al. (JPAC), Phys. Rev. Lett. 123, 092001 (2019)



Double-triangle (w. complex coupl. in the Lagrangian)

Nakamura, Phys. Rev. D 103, 111503 (2021)



Single triangle (ruled out)

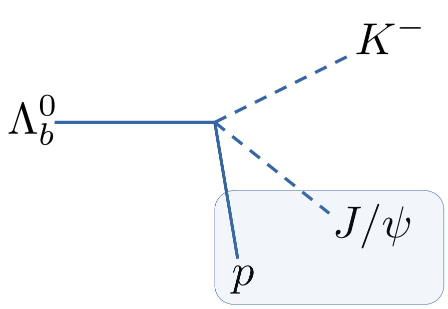
LHCb, Phys. Rev. Lett. 122, 222001 (2019)

We want to use ANN to:

- Go beyond the standard χ^2 fitting
- Specific questions:
  - Can we train a neural network to analyze a line shape and get as a result the probability of each possible dynamical explanation ?
  - If possible, what other information can we gain by using machine learning techniques?
- First attempts to use Deep neural networks as model classifiers for hadron spectroscopy:

Sombillo et al., 2003.10770, 2104.141782, 2105.04898

Physics model



- $P_c(4312)$ seen as a maximum in the pJ/ψ energy spectrum
- $P_c(4312)$ has a well defined spin and appears in single partial wave
- Background contributes to all other waves
- $\Sigma^+_c \bar{D}^0$ channel opens at 4.318 GeV -coupled channel problem

- Intensity

$$\frac{dN}{d\sqrt{s}} = \rho(s)\left[|P_1(s)T_{11}(s)|^2 + B(s)\right]$$

where

$$\rho(s) = pqm_{\Lambda_b}$$ phase space

$$p = \lambda^{\frac{1}{2}}(s, m^2_{\Lambda_b}, m^2_K)/2m_{\Lambda_b}, \;\; q = \lambda^{\frac{1}{2}}(s, m^2_p, m^2_\psi)/2\sqrt{s}$$

$$P_1(s) = p_0 + p_1 s$$ production term

$$B(s) = b_0 + b_1 s$$ background term

Physics model

- Coupled channel amplitudes

$$T_{ij}^{-1} = M_{ij} - ik_i\delta_{ij} \quad \text{where} \quad k_i = \sqrt{s - s_i}$$

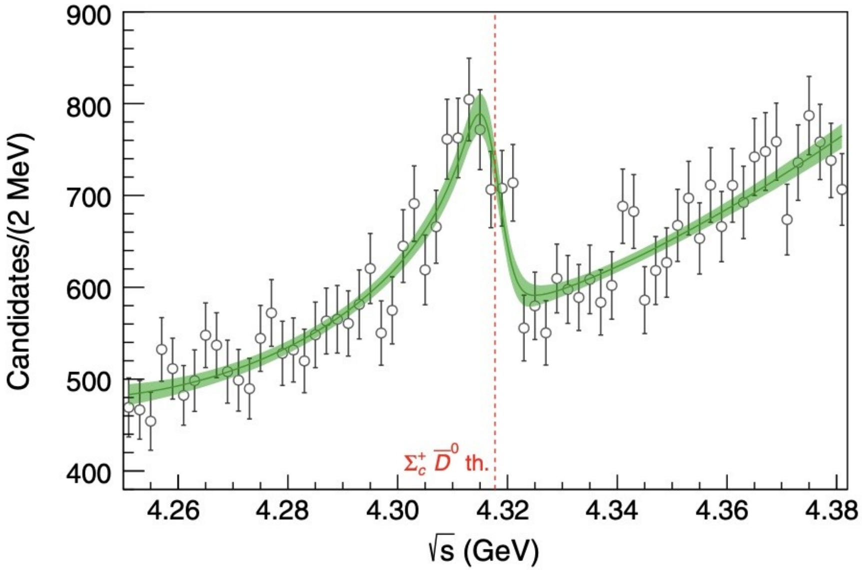
$$s_1 = (m_p + m_{J/\psi})^2 \quad \text{and} \quad s_2 = (m_{\Sigma_c^+} + m_{\bar{D}^0})^2$$

- Unitarity implies that M_{ij} is free from singularities near thresholds s_1 and s_2 so that it can be Taylor expanded *Frazer, Hendry Phys. Rev. 134 (1964)*

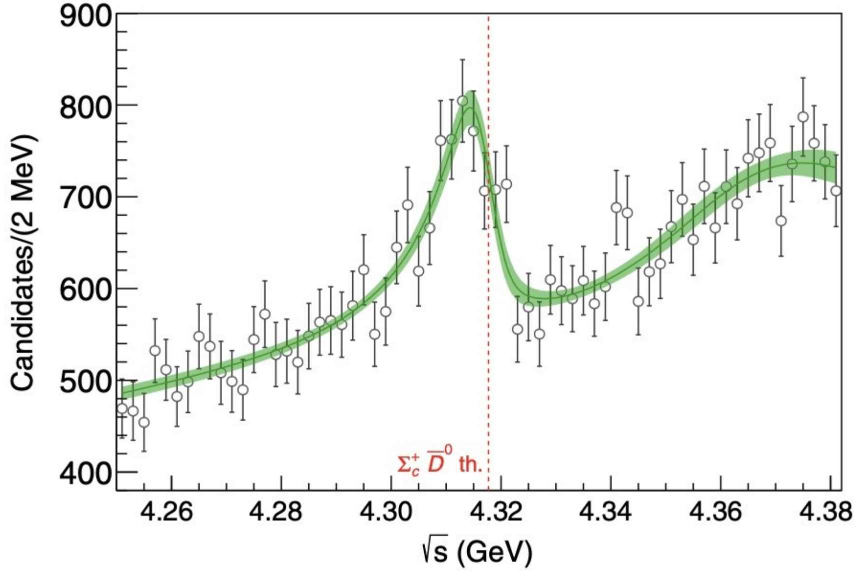
$$M_{ij}(s) = m_{ij} - c_{ij}s$$

- In principle the off-diagonal term of the amplitude $P_2(s)T_{21}$ could be included but we are interested in the analytical structure ("denominator") – so it's effect can be absorbed to the background and production terms.

Physics model – final version



Scattering length approximation



Effective range approximation

See *C. Fernandez-Ramirez Phys.Rev.Lett. 123 (2019) 9, 092001*

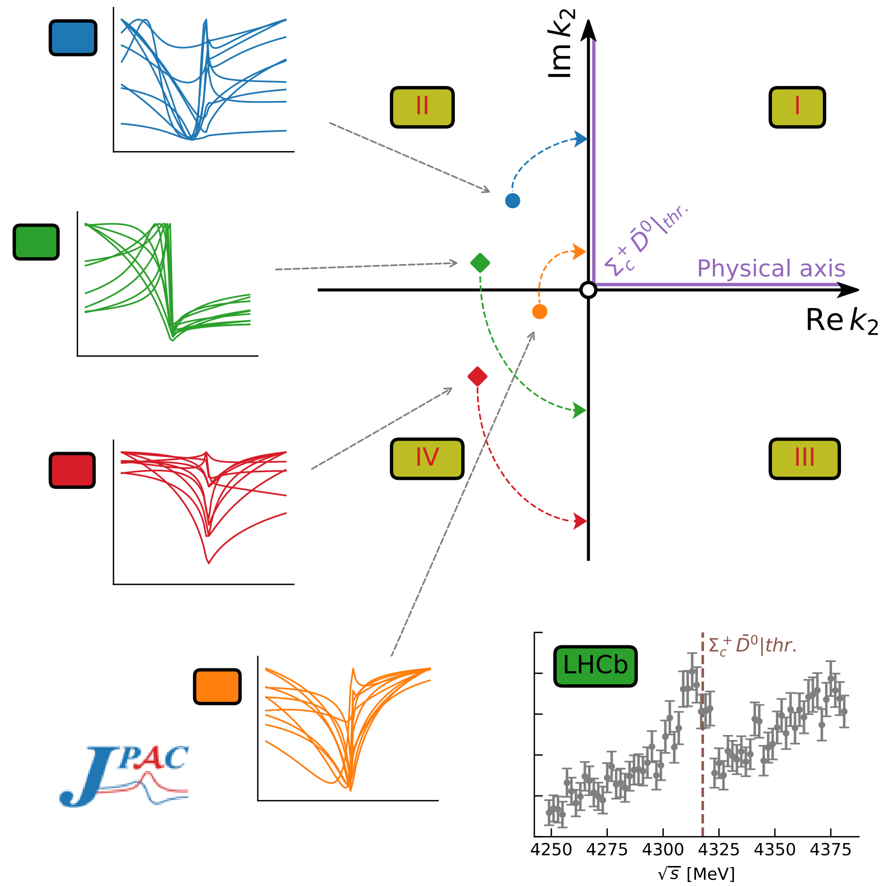
Finally we use the scattering length approximated amplitude as the basis for ML model

$$T_{11} = \frac{m_{22} - ik_2}{(m_{11} - ik_1)(m_{22} - ik_2) - m_{12}^2}$$

7 model parameters in total: m_{11}, m_{22}, m_{12}, p_0, p_1, b_0, b_1.

ML model – general idea

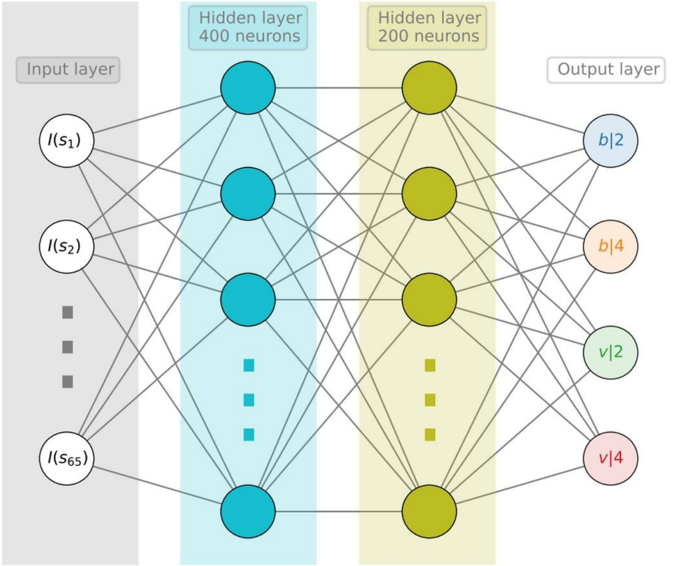
- From the physical model we produce:
  - Sample intensities (computed in 65 energy bins) – produced with randomly chosen parameter samples – **examples**
  - For each parameter sample we are able to compute the **target class** – one of the four: b|2, b|4, v|2, v|4
  - Symbolically:



$$K : \{[I_1, ..., I_{65}](m_{11}, m_{22}, m_{12}, p_0, p_1, b_0, b_1)\} \to \{b|2, b|4, v|2, v|4\}$$

ML model – MLP

Layer	Shape in	Shape out
Input		(B, 65)
Dense	(B, 65)	(B, 400)
Dropout(p=0.2)	(B, 400)	(B, 400)
ReLU	(B, 400)	(B, 400)
Dense	(B, 400)	(B, 200)
Dropout(p=0.5)	(B, 200)	(B, 200)
ReLU	(B, 200)	(B, 200)
Dense	(B, 200)	(B, 4)
Softmax	(B, 4)	(B, 4)



Training dataset preparation:

1. Parameters were uniformly sampled from the following ranges: b_0 = [0 ; 700], b_1 = [-40 ; 40], p_0 = [0 ; 600], p_1 = [-35 ; 35], M_{22} = [-0.4 ; 0.4], M_{11} = [-4 ; 4], M_{12}^2 = [0 ; 1.4]
2. The signal was smeared by convolving with experimental LHCb resolution:

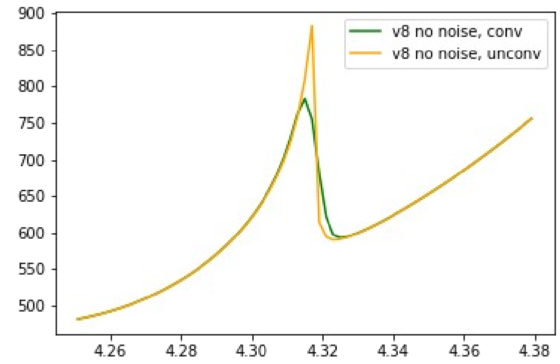
$$I(s) = \int_{m_\psi + m_p}^{m_{\Lambda_b} - m_K} I(s')_{\text{theo}} \exp\left[-\frac{(\sqrt{s}-\sqrt{s'})^2}{2R^2(s)}\right] d\sqrt{s'} \Big/ \int_{m_\psi + m_p}^{m_{\Lambda_b} - m_K} \exp\left[-\frac{(\sqrt{s}-\sqrt{s'})^2}{2R^2(s)}\right] d\sqrt{s'}\,,$$

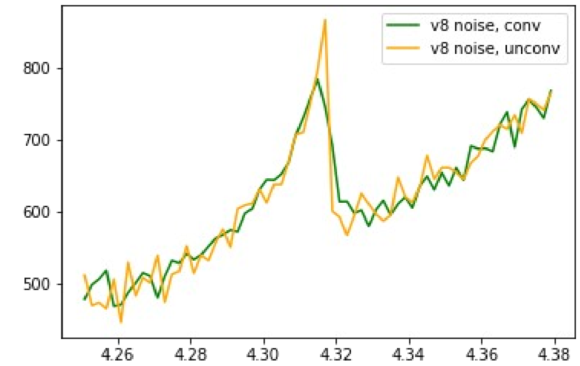
$$R(s) = 2.71 - 6.56 \times 10^{-6-1} \times \left(\sqrt{s} - 4567\right)^2$$

3. To account for experimental encertainty the 5% gaussian noise was added

ML model - training

- Input examples (efect of energy smearing and noise):





- Computing target classes:
  - $m_{22}>0$ – bound state, $m_{22}<0$ – virtual state
  - To localize the poles on Riemann sheets we need to find zeros of the amplitude denominator in the momentum space:

$$p_0 + p_1\, q + p_2\, q^2 + p_3\, q^3 + q^4 = 0$$

with

$$p_0 = (s_1 - s_2)\, m_{22}^2 - \left(m_{12}^2 - m_{11} m_{22}\right)^2$$

$$p_1 = 2\,(s_1 - s_2)\, m_{22} + 2 m_{11}\left(m_{12}^2 - m_{11} m_{22}\right)$$

$$p_2 = -m_{11}^2 + m_{22}^2 + s_1 - s_2$$

$$p_3 = 2 m_{22}$$

Then poles appear on sheets defined with (η_1, η_2) pairs:

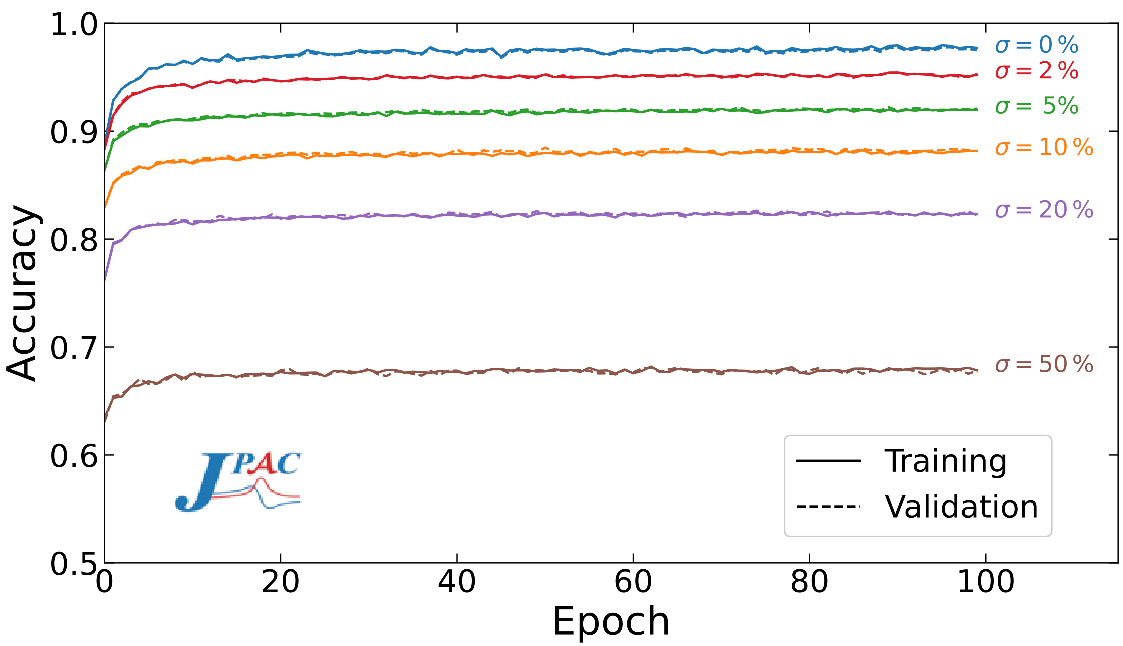
(-,+) - II sheet

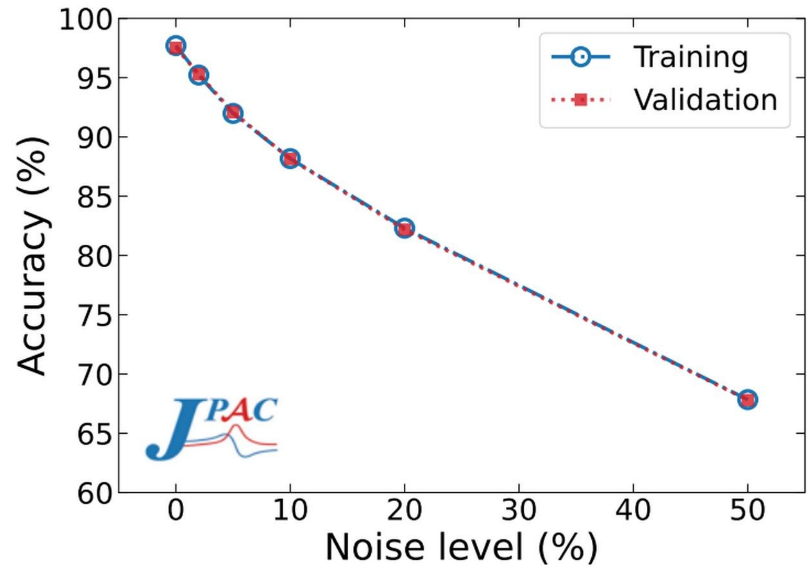
(+,-) - IV sheet

$$\eta_1 = \text{Sign Re}\left(\frac{m_{12}^2}{m_{22}+q} - m_{11}\right) \quad \eta_2 = \text{Sign Re} q$$

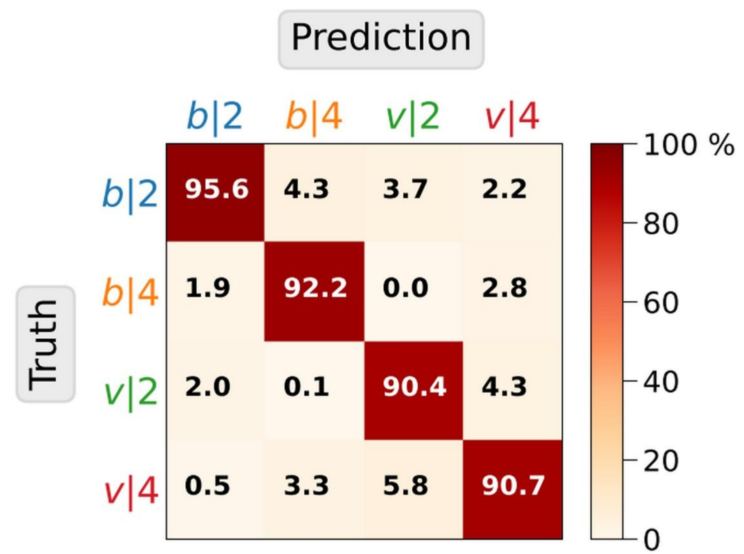
ML model – training results

Accuracy for various noise levels



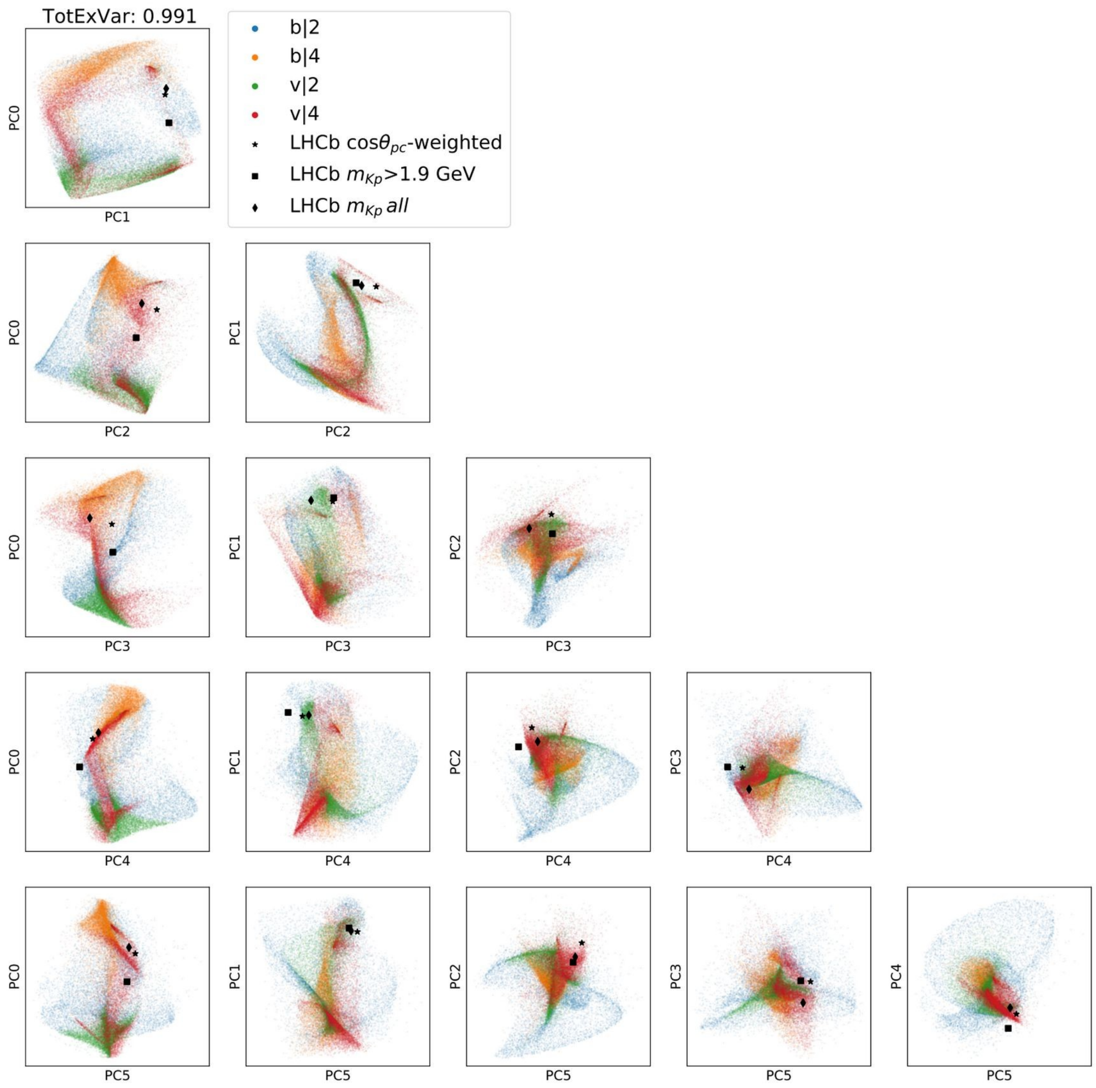


Confusion matrix for the 5% noise



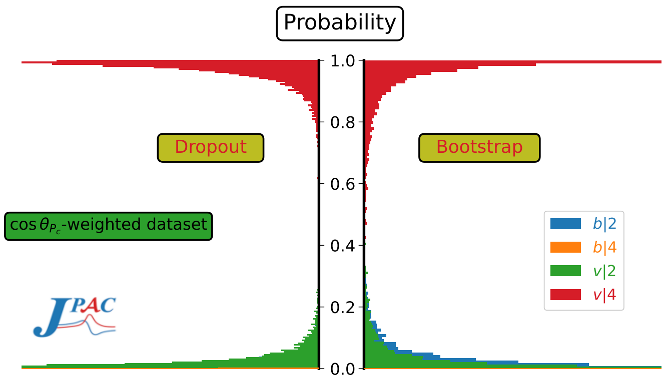
Feature refinement

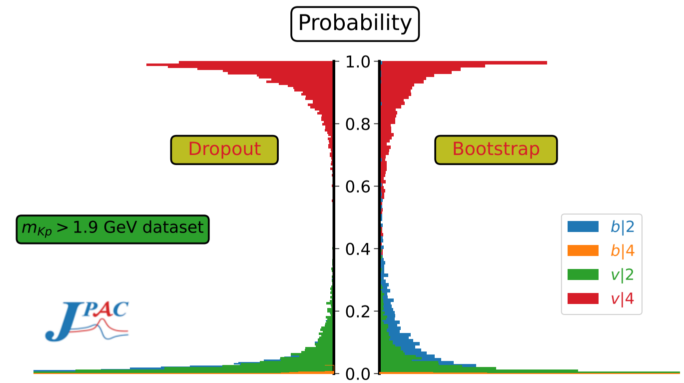
- Dimensionality reduction - Principal Component analysis
- Over 99% of the variance can be explained with just 6 features
- Experimental data projected onto principal components are well encompassed within the training dataset

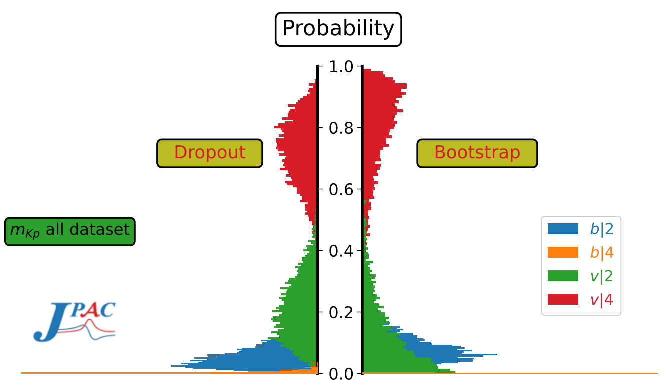


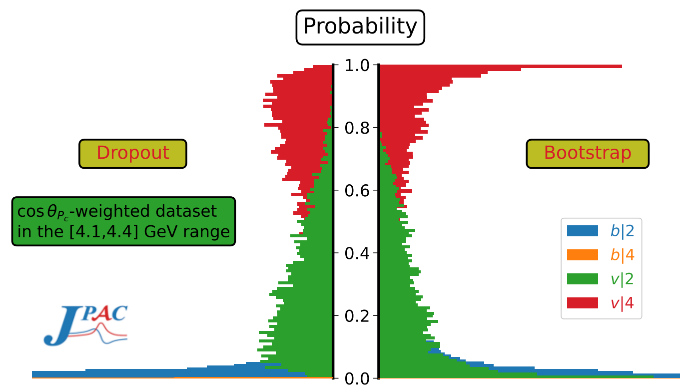
Model predictions – statistical analysis

- The distribution of the target classes was evaluated with
  - the bootstrap (10 000 pseododata based on experimental mean values and uncertainties) and
  - dropout (10 000 predictions from the ML model with a fraction of weights randomly dropped out)





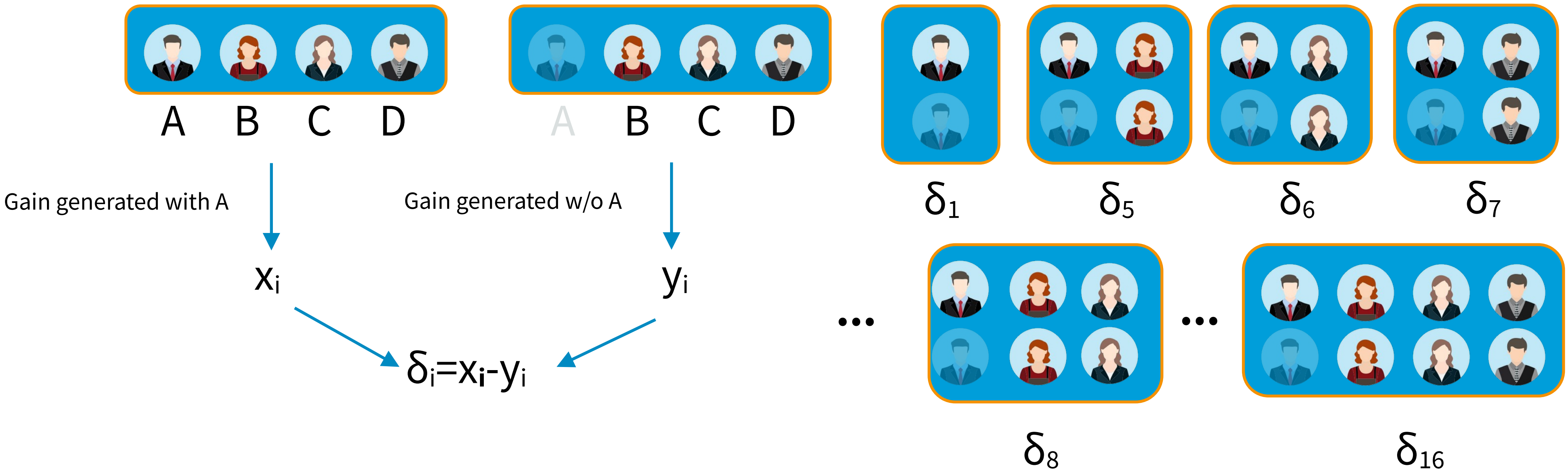




Model explanation with SHAP

- Shapley values and Shapley Additive Explanations

Shapley, Lloyd S. "Notes on the n-Person Game -- II: The Value of an n-Person Game" (1951)



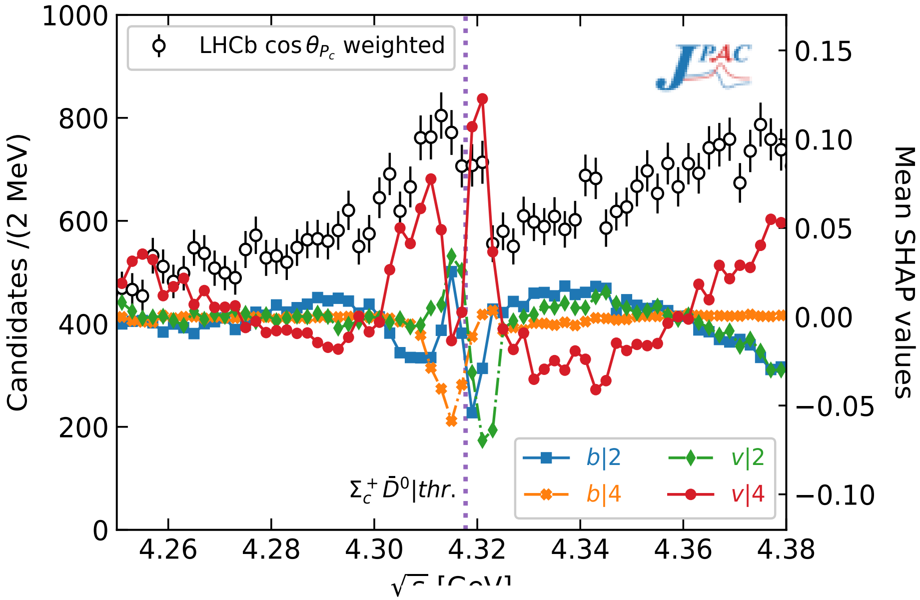
Shapley value for member A: $$\phi_A = \frac{\delta_1 + \delta_2 + \cdots + \delta_{16}}{16}$$

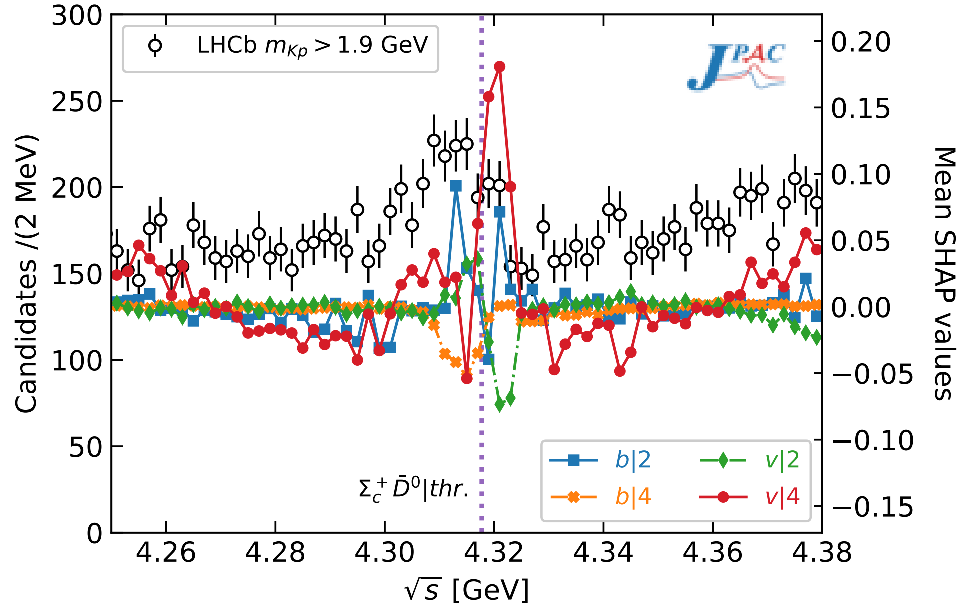
Model explanation with SHAP

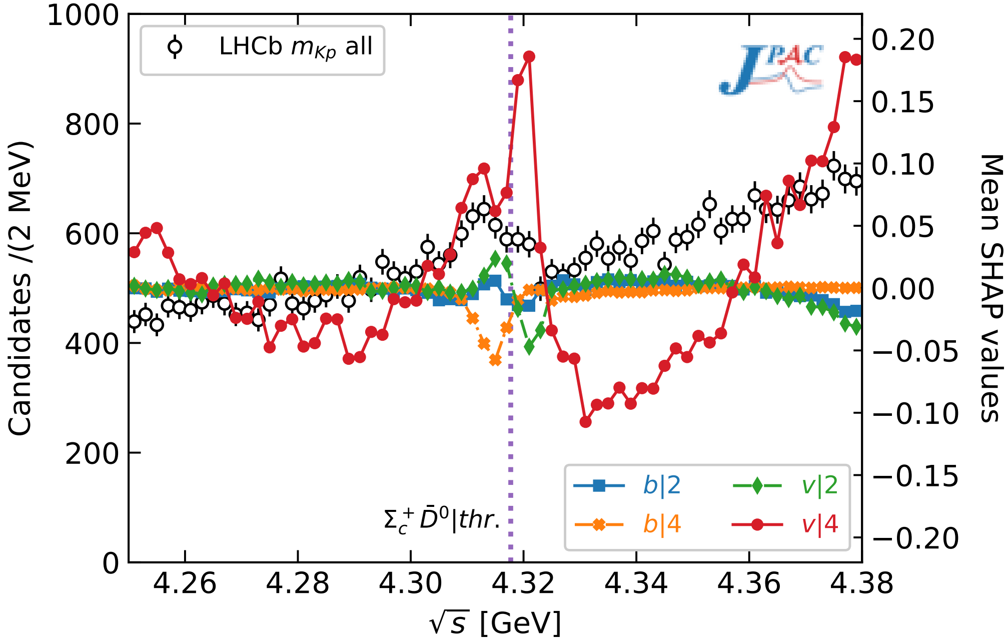
- By making an assotiation:
  - Member of a coalition → Feature (intensity in the energy bin)
  - Game → Function that generates classification/regression result
  - Gain → Prediction
  - We define the Shapley values for features
- Caveats:
  - A number of possible coalitions grows like 2^N
  - Prohibitively expensive computationally (NP-hard)

Solution: Shapley additive explanations (Lundberg, Lee, arXiv:1705.07874v2, 2017)

Model explanation with SHAP







Going beyond discriminative model

(work in progress under A(I)DAPT collaboration)

Marco Battaglieri, INFN
Alessandro Pilloni, University of Messina/JPAC
Tommaso Vittorini, University of Tor Vergata
Gloria Montana, University of Barcelona
Łukasz Bibrzycki, Pedagogical University of Krakow/JPAC
Nobuo Sato, JLab
Yaohang Li, Old Dominion University
Tereq Alghamdi, Old Dominion University
Hua Liu, Old Dominion University

Going beyond discriminative model

- Generative models learn to generate data instances drawn from pdf $\rho(x_1, \dots, x_n)$, which in turn is learnt from data
- Physically pdf-s are related to amplitudes of hadronic processes

$$\rho(x_1, \dots, x_n) \sim |A(x_1, \dots, x_n)|^2$$

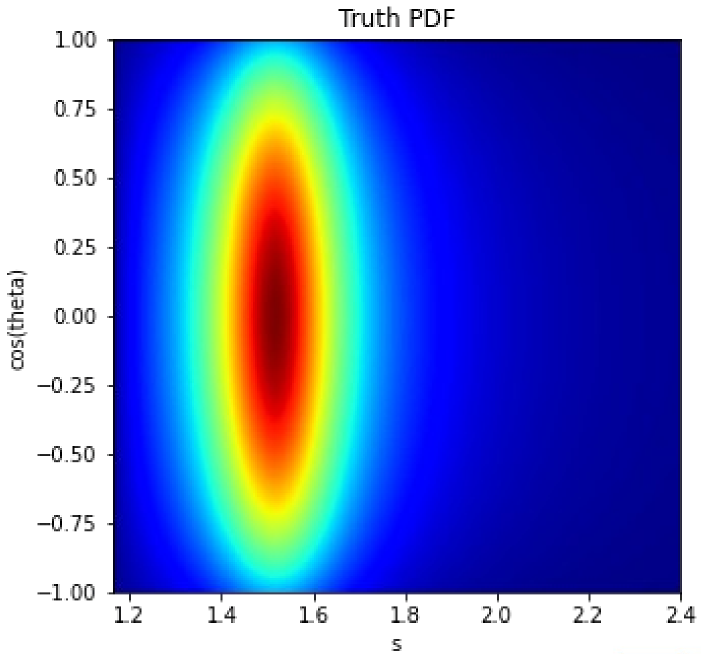
- One can think about obtaining the amplitude from the pdf obtained from the generative model
- This problem is, unfortunately, ill posed problem
- Can additional conditions (unitarity, dispersion relations) cure the situation ?

π^0 photoproduction

- Consider the model for $\gamma p \to p\pi^0$

$$\frac{d\sigma}{d\Omega} \propto \frac{p_f}{p_i\, s} \frac{3\left|H_{3/2}\right|^2 + 5\left|H_{1/2}\right|^2 - 3\cos 2\theta\left(\left|H_{3/2}\right|^2 - \left|H_{1/2}\right|^2\right)}{(m_\Delta^2 - s)^2 + \Gamma_\Delta^2 m_\Delta^2}$$

- Putting physical values of model parameters we obtain the 2d pdf in (s, cosθ) variables



(Machine) Learning pdf-s

- Normalizing flows are the ML architectures designed to learn pdf-s from data
- Basic idea:
  - Let X and Z be *n*-dimensional random vectors and
  - $f : R^n \to R^n$ be the mapping between them so that

    $X = f(Z)$ and $Z = f^{-1}(X)$ (f must be invertible)

  - Then $\rho_X(x) = \rho_Z(f^{-1}(x)) \left| \frac{\partial f^{-1}(x)}{\partial x} \right|$

(Machine) Learning pdf-s

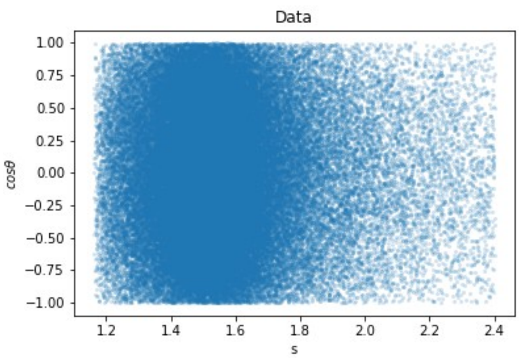
In practice

- f is implemented as a composition of several functions parametrized by neural network layers (flows) – activation functions must be invertible

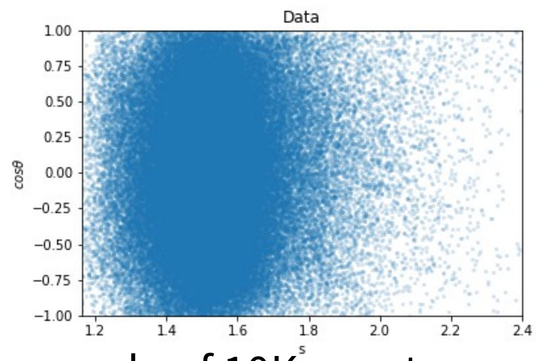
$$f = f_n \circ \cdots \circ f_2 \circ f_1$$

- Z is usually normal disributed, so we train the network to map unknown (either model or experimental) input distribution into a Gaussian noise
- Having trained the network (to parametrize f) we use it's invertibility to sample "unknown" distribution by feeding Gaussian noise.

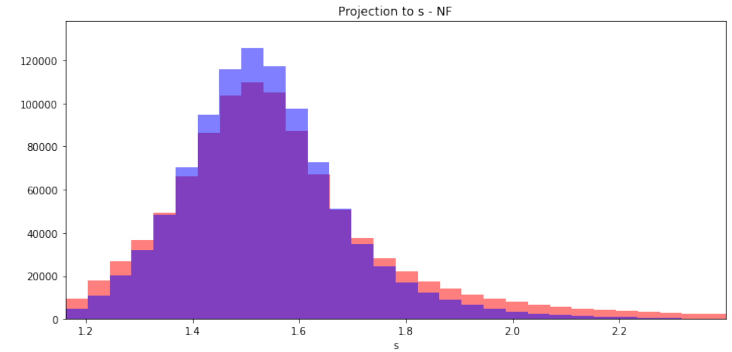
Results

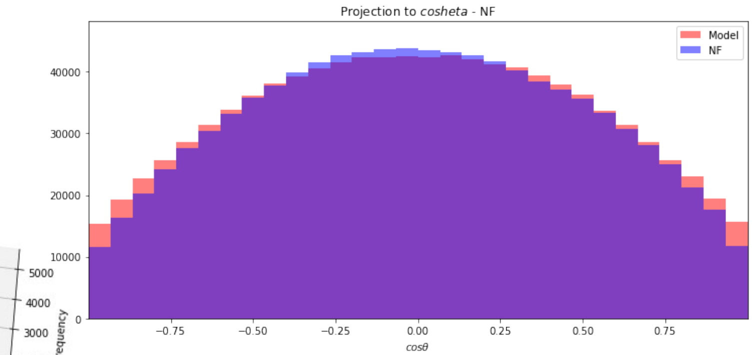


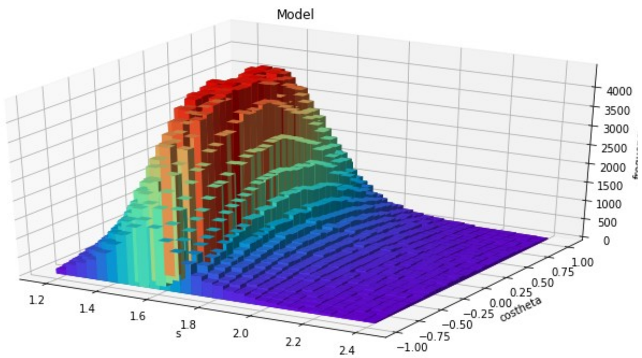
A sample of 10K events from the model distrubution



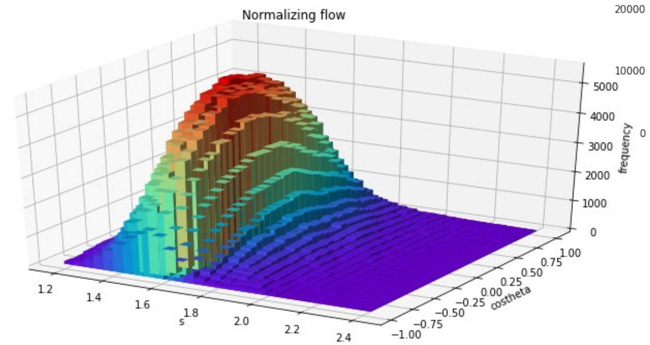
A sample of 10K events generated by the normalizing flow







Histogrammed model sample



Histogrammed normalizing flow sample

- Histograms projected in s and in $\cos\theta$
- Some discrepancies visible but they can be reduced by hyperparameter fine-tuning

- This was an easy part – obtaining the amplitudes requires imposing constraint of unitarity, eg. in the form of dispersion relations.
- It is still ahead of us.

Summary

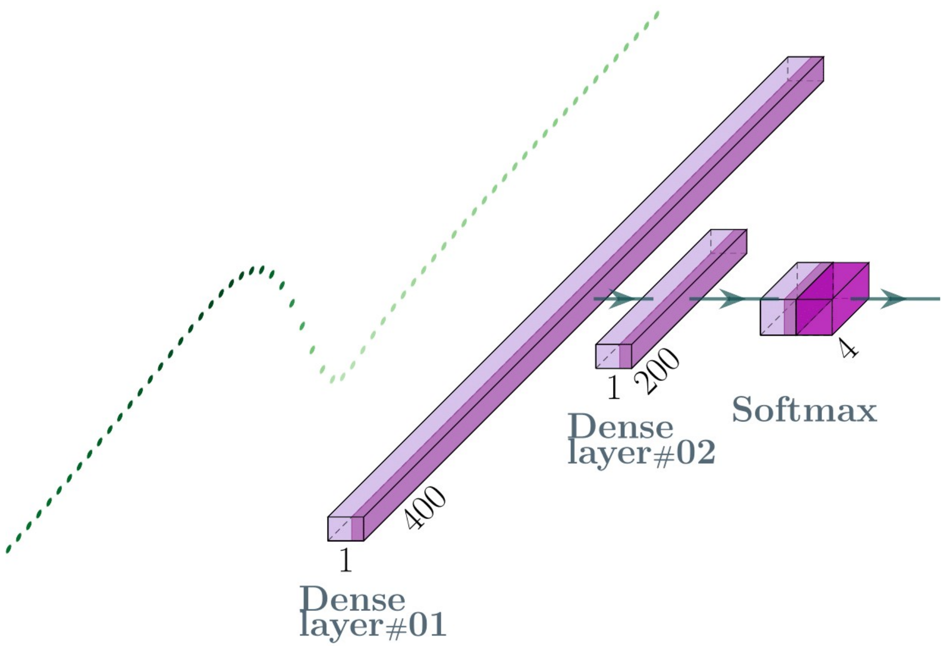
- Classification of $P_c(4312)$ poles:
  - Rather than testing the single model hypothesis with χ^2, we obtained the probabilities of four competitive pole assignments for the $P_c(4312)$ state
  - By the analysis of the SHAP values we obtained an *ex post* justification of our scattering length approximation
- Learning pdf-s (and possibly the amplitudes) with normalizing flows
  - With normalizing flows one can learn pdf-s from models (not a big deal) or experimental data(bigger deal)
  - Physical constraints of unitarity etc. have to be imposed on training in order to go from pdf-s to amplitudes.

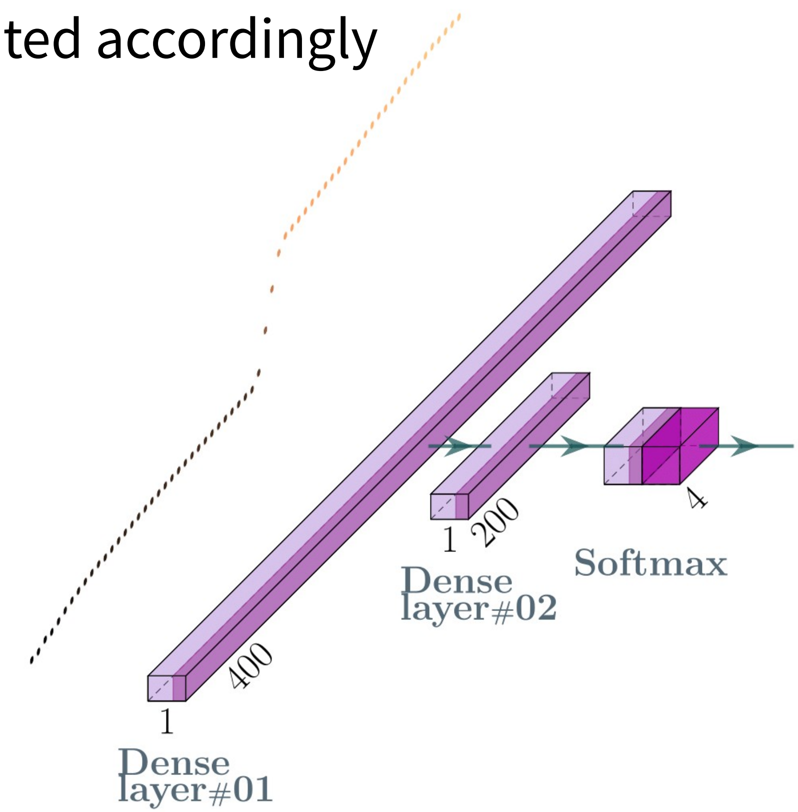
Thank you !

Back-up slides

Caveats (on using MLPs)

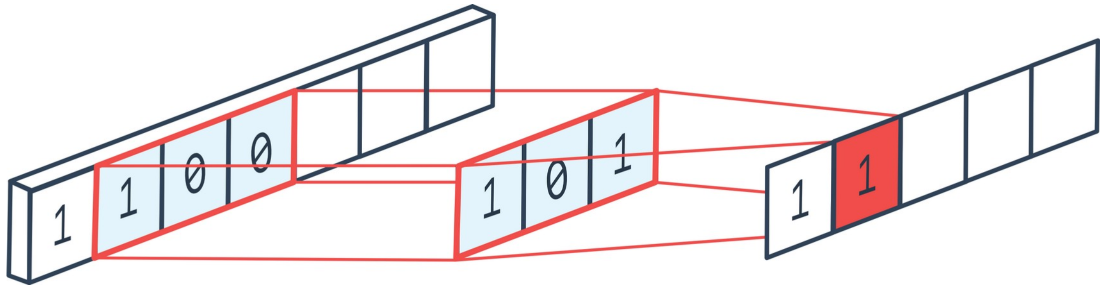
- Even though we want to recognize ordered sets (series) of data the MLP rather recognizes just sets
- One can permute the data arbitrarily and get basically the same classification quality
- Provided the prediction dataset is permuted accordingly

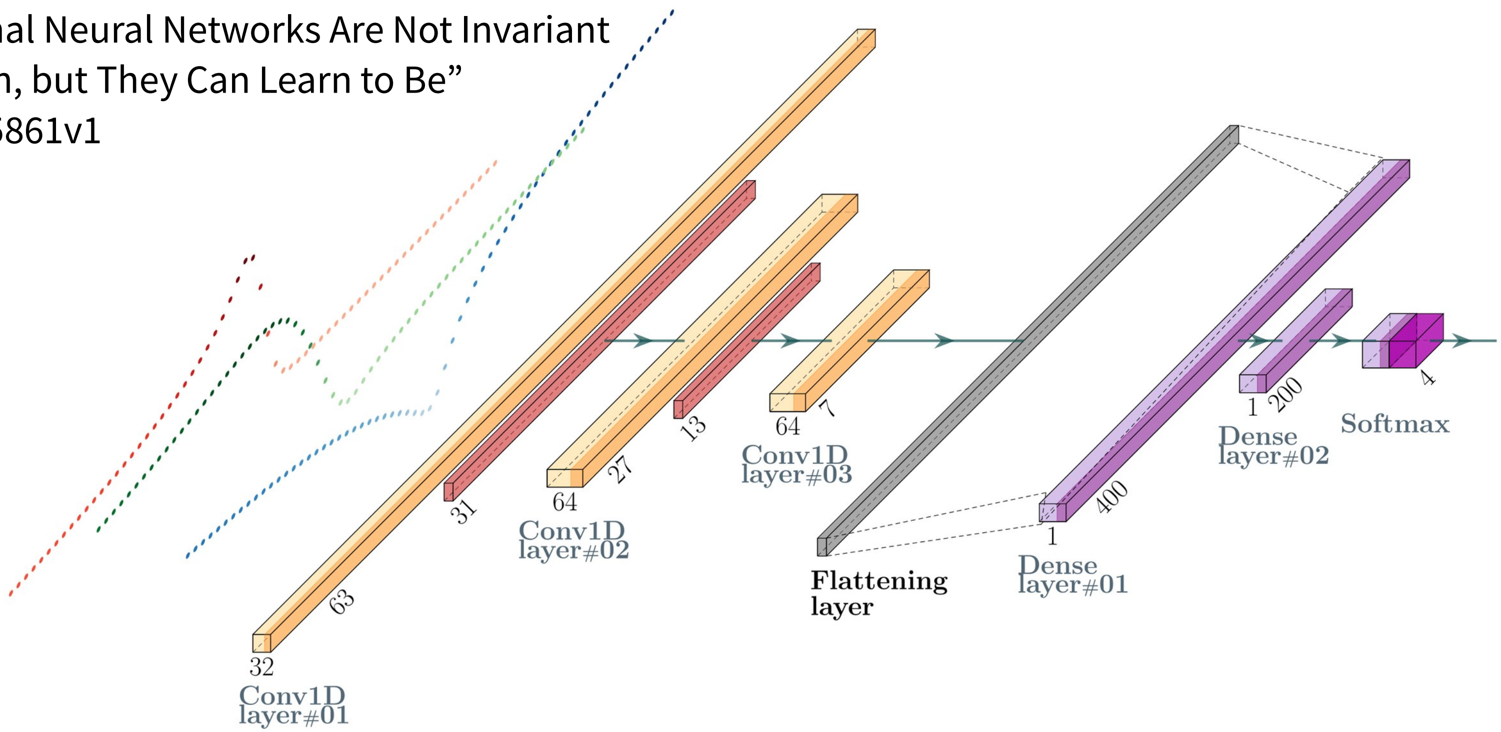




CNN as an alternative

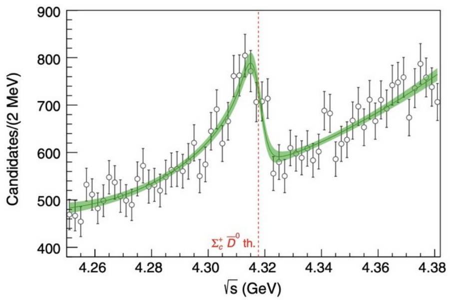
- Convolution neural network is able to detect local patterns
- Unfortunately it does it in fixed location (it's not translationally invariant)
- There are some (partial) remedies however: V. Biscione, J. S. Bowers, "Convolutional Neural Networks Are Not Invariant to Translation, but They Can Learn to Be" arXiv:2110.05861v1

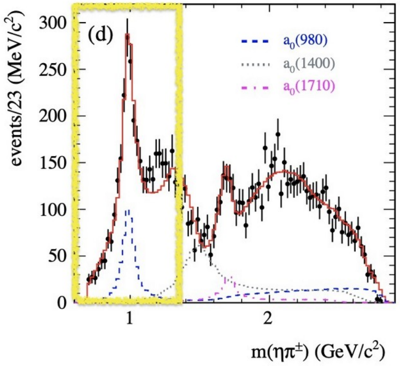




Questions to be addressed

- Going beyond the limited generalization power - applying the method for larger class of resonances, described by the same physics, eg. $a_0/f_0(980)$ or other resonances located near thresholds





- Eg. we believe that these two resonances can be described by the same physics
  - MLPs and CNNs require inputs of the same size – rebinning required (but also kinematics and resonance parameters change: masses, widths, thresholds, phase spaces,...)
  - Alternatively we can use the length of the signal as part of the input information for RNNs
  - Difference between the models is not always as clear as above (different Riemann sheets) – need for model selection criteria (discussed already on Wednesday)