

Model picking beyond the χ^2 fitting

Klarence Tomas R. Cervantes* and Denny Lane B. Sombillo

National Institute of Physics, University of the Philippines Diliman, 1101 Quezon City, Philippines

ABSTRACT

The study addresses the challenge of interpreting near-threshold enhancements seen in scattering data, specifically focusing on the $P_{c\bar{c}}(4312)^+$ pentaquark signal. While the K -matrix formalism provides a robust framework for maintaining S -matrix unitarity by construction, multiple parametrizations can often produce statistically equivalent χ^2 fits despite having fundamentally different pole structures. In an attempt to resolve this ambiguity, a supervised deep neural network (DNN) was trained on 40,000 synthetic spectra generated from four candidate K -matrix families. Using a fixed held-out test set and five independent training seeds, the classifier achieved a mean accuracy of 82.61% and mean macro-F1 of 82.37%. The repeated DNNs predominantly assigned perturbed realizations of the experimental spectrum to the two-resonance model with nonresonant K -matrix terms (Class 3). However, this assignment was not stable across classifier families, and performance was sensitive to noise and the analyzed energy interval. The machine-learning result is therefore interpreted as conditional evidence within the four simulated model families and nominal analysis assumptions rather than as a posterior probability or definitive identification of the physical pole structure.

INTRODUCTION

Interpreting peaks in scattering data as genuine resonances is a central problem in hadron spectroscopy. The usual approach involves associating the peak with a pole of the scattering amplitude. However, this interpretation becomes more subtle when the enhancement appears close to a channel threshold. Near a threshold, the opening of a channel can distort the lineshape and produce structures that resemble genuine resonance behavior. Therefore, the presence of a peak alone is not sufficient to establish the nature of the state. This issue is especially important for exotic hadrons like the pentaquark candidate $P_{c\bar{c}}(4312)^+$, which was observed by the LHCb Collaboration in the $J/\psi p$ invariant mass spectrum (Aaij et al. 2019). Since the $P_{c\bar{c}}(4312)^+$ lies close to the $\Sigma_c^+ \bar{D}^0$ threshold, its observed enhancement may be strongly influenced by near-threshold phenomena. Several analyses on the signal gave different interpretations such as a compact state (Chen

et al. 2022), a hadronic molecule (Xiao et al. 2019), and a virtual state (Ruangyoo et al. 2022, Silva-Castro 2020). A model that reproduces the experimental peak may therefore still admit different physical interpretations depending on whether the enhancement is a compact resonance or a hadronic molecule state. For this reason, amplitude-based approaches are needed not only to fit the data but also to examine the pole topology associated with the fitted model.

A useful framework for this problem is the K -matrix formalism, which allows both resonance and nonresonant contributions to be incorporated while maintaining a unitary S -matrix, which is essential for adhering to the conservation of probability (Chung et al. 1995). In this study, the $P_{c\bar{c}}(4312)^+$ signal was described using a two-channel K -matrix model involving the $J/\psi p$ and $\Sigma_c^+ \bar{D}^0$ channels, with thresholds around 4035.17 MeV and 4317.49 MeV, respectively. A previous analysis of the same pentaquark signal used a model-independent approach and determined a $[1, 1, 1]$ pole configuration, corresponding to one pole in each relevant unphysical sheet, which is seen via analytic continuation (Santos et al. 2025). While this provides a physically motivated description of the $P_{c\bar{c}}(4312)^+$ signal, the interpretation remains incomplete as the model-independent approach does not give the actual value of the pole and requires an additional model-dependent step to determine pole value across the different Riemann sheets. Ultimately, the interpretation remains tied to the choice of model, and in this case, the choice of the K -matrix parametrization. This becomes problematic when alternative parametrizations are also able to reproduce the same experimental data with comparable chi-squared. A related analysis by Pagayon et al. (2026) obtained substantially different pole-width estimates from two unitary parametrizations despite comparably good fits to the same near-threshold signal, providing additional motivation for the present investigation of model-selection ambiguity beyond chi-squared fit quality.

Although chi-squared measures the numerical agreement between a model and the data, it does not necessarily identify whether the fitted model represents the most physically meaningful mechanism. Different parametrizations may compensate one another during fitting, especially when additional free parameters are introduced. For example, a model with two resonance terms may achieve a fit of quality comparable to that of a one-resonance

*Corresponding author
Email Address: krcervantes1@up.edu.ph
Date received: 31 May 2026
Date revised: 25 July 2026
Date accepted: 07 August 2026
DOI: <https://doi.org/10.54645/202619SupXGG-59>

KEYWORDS

fitting, machine learning, pentaquark, pole, resonance, scattering

model with nonresonant terms, even though the two descriptions imply different physical pictures. Thus, several models may appear equally acceptable from the standpoint of goodness-of-fit while leading to different pole interpretations.

To address this ambiguity, this work treated the selection of competing two-channel K -matrix parametrizations as a supervised machine learning classification problem. Instead of using the DNN to directly extract pole positions, the network is trained to recognize lineshape patterns generated by different K -matrix model families. Four classes are considered: one resonance with nonresonant terms, one resonance only, two resonances with nonresonant terms, and two resonances only. Synthetic lineshapes generated from these classes are used to train and validate the classifier, after which the trained model is applied to the $P_{c\bar{c}}(4312)^+$ experimental data. Through this approach, the study aimed to provide a model-selection tool that complements conventional χ^2 -based fitting and helps distinguish physically distinct mechanisms that may otherwise produce similarly good fits.

MATERIALS AND METHODS

Two-channel K -matrix formalism

The scattering amplitude was modeled using a two-channel K -matrix formalism. The K -matrix parametrization used in this study contains resonance terms and, in some classes, nonresonant term, with a general form seen in Equation 1.

$$K_{ij} = \sum_{\alpha} \frac{\gamma_{\alpha i} \gamma_{\alpha j} m_{\alpha} \Gamma_{\alpha}^0}{m_{\alpha}^2 - s} + c_{ij} \quad (1)$$

The i and j label the scattering channels, α labels the resonance term, m_{α} is the bare resonance mass, Γ_{α}^0 is the resonance width parameter, $\gamma_{\alpha i}$ is the coupling of resonance α to channel i , c_{ij} represents the nonresonant contribution. The two channels considered are the $J/\psi p$ and $\Sigma_c^+ \bar{D}^0$ channels. The use of these specific two channels is motivated by the location of the pentaquark signal which is near the $\Sigma_c^+ \bar{D}^0$ threshold. This K -matrix equation is directly related to the scattering matrix T which is used to generate the invariant mass distribution plots using Equation 2 (Chung et al. 1995).

$$T = K(I - iK)^{-1} \quad (2)$$

where I is the 2×2 identity matrix for this case. Using these, four K -matrix model families were constructed. These models differ in the number of resonance terms and in whether nonresonant terms are included in their K -matrix form.

Table 1: Definition of each K -matrix class used in the classification framework.

Class	K -matrix form	Description
1	$K_{ij} = \frac{\gamma_{\alpha i} \gamma_{\alpha j} m_{\alpha} \Gamma_{\alpha}^0}{m_{\alpha}^2 - s} + c_{ij}$	One resonance with nonresonant term
2	$K_{ij} = \frac{\gamma_{\alpha i} \gamma_{\alpha j} m_{\alpha} \Gamma_{\alpha}^0}{m_{\alpha}^2 - s}$	One resonance term only
3	$K_{ij} = \frac{\gamma_{\alpha i} \gamma_{\alpha j} m_{\alpha} \Gamma_{\alpha}^0}{m_{\alpha}^2 - s} + \frac{\gamma_{\beta i} \gamma_{\beta j} m_{\beta} \Gamma_{\beta}^0}{m_{\beta}^2 - s} + c_{ij}$	Two resonances with nonresonant term
4	$K_{ij} = \frac{\gamma_{\alpha i} \gamma_{\alpha j} m_{\alpha} \Gamma_{\alpha}^0}{m_{\alpha}^2 - s} + \frac{\gamma_{\beta i} \gamma_{\beta j} m_{\beta} \Gamma_{\beta}^0}{m_{\beta}^2 - s}$	Two resonance terms only

This class definition is important because it allows the DNN to test two competing sources of model flexibility, namely, additional resonances and nonresonant interactions. Classes 1 and 2 isolate the role of the nonresonant term in a one-resonance model. Classes 3 and 4 test whether adding another resonance improves the description. Classes 1 and 3 compare whether the nonresonant model requires one or two resonances. Classes 2 and 4 compare purely resonant descriptions of different complexity.

Initial χ^2 fitting via iminuit

Before applying the DNN classifier, each of the four K -matrix classes was fitted directly to the experimental invariant-mass distribution. This part is essential because it demonstrates the original problem, which is that the four models can produce visually similar lineshapes and comparable chi-squared values. Each K -matrix model was fitted to the experimental invariant-mass spectrum by minimizing a chi-square objective function. Equation 1 is converted to the scattering T -matrix to produce the count or intensity values of the invariant mass distribution (Equations 3 and 4; Santos & Sombillo 2023)

$$F(E) = |s_1 T_{11} + s_2 T_{12}|^2 \quad (3)$$

$$\text{count or intensity} = A[P * F(E) + B(E)] \quad (4)$$

where s are the amplitude coefficients, A is a scaling parameter, P is the production constant, and $B(E)$ is some background polynomial. The resonance mass m_{α} and width parameter Γ_{α}^0 were treated as free parameters within the restricted intervals based on the experimentally reported mass and width of the pentaquark signal in Aaij et al. (2019). These restrictions were imposed to

confine the minimization to the experimentally relevant resonance region. For models having a second resonance, the parameters m_{β} and Γ_{β}^0 were also treated as free parameters within specific energy and value intervals. Other parameters such as the channel couplings γ were treated as free parameters but with broader bounds. For models that contained nonresonant terms (c_{ij}), these terms were also varied with bounds, but it was ensured that $c_{12} = c_{21}$ to enforce the symmetry of the K -matrix. The minimization was performed using iminuit (James and Roos 1975), which provides numerical optimization tools commonly used for nonlinear least-squares fitting. The resulting best-fit curves were then compared through their chi-squared values and visual agreement with the data. The critical observation is not merely that the fits are good, but that physically distinct models achieve comparable goodness-of-fit values. A one-resonance model with a nonresonant term and a two-resonance model may both reproduce the peak region, but they imply different mechanisms. This motivates the use of a classifier trained to recognize model-family-level patterns beyond a single best-fit chi-squared.

Deep Neural Network (DNN) architecture and application

The dataset used for DNN training and validation consisted entirely of synthetic lineshapes generated from the four K -matrix classes defined in Table 1. The LHCb experimental data were excluded in the training, validation, hyperparameter optimization, and were used only during the final inference stage. For each synthetic sample, the free parameters of the corresponding K -matrix class were randomly drawn from the distributions in Table 2. We

generated 10,000 spectra per class, giving a balanced nominal dataset of 40,000 spectra and a class prior of 0.25. Here, nominal refers to the reference dataset generated using the baseline parameter ranges, common energy interval, equal class proportions, and no additional measurement noise. Equal representation of the four classes prevents the classifier from obtaining an artificially high accuracy score by favoring a more frequent class. Every spectrum was evaluated at the same 75 experimental data central energy coordinates satisfying $4200 \leq E < 4350$ MeV. Since the energy coordinates were identical for all classes, only the 75 intensity values were used as classifier features. The common energy coordinates were retained as metadata. No interpolation or additional additive noise was applied to the nominal dataset. Using the same energy grid for all spectra ensures that classification is based on differences in the intensity lineshapes rather than on differences in the input coordinates. The coordinates were retained only to associate each intensity with its physical mass and were not supplied as classifier features. All continuously varied parameters were sampled independently from uniform distributions within the ranges listed in Table 2. For the classes containing nonresonant terms, the c_{11} , c_{12} , and c_{22} parameters were varied while imposing the symmetry condition $c_{12} = c_{21}$. These terms were set to zero for the classes without nonresonant contributions. For each sampled parameter set, the counts were calculated from Equation 4.

Table 2: Distributions of the relevant parameters used in generating the synthetic lineshapes.

Parameter	Sampling Distribution	Classes
m_α	Uniform (4311.2, 4311.9)	1-4
Γ_α^0	Uniform (2.6, 16.2)	1-4
m_β	Uniform (4321.7, 4325.7)	3 and 4
Γ_β^0	Uniform (2.6, 16.2)	3 and 4
$\gamma_{\alpha 1}, \gamma_{\alpha 2}$	Uniform (-7,7)	1-4
$\gamma_{\beta 1}, \gamma_{\beta 2}$	Uniform (-7,7)	3 and 4
c_{11}, c_{12}, c_{22}	Uniform (-5,5)	1 and 3

The dataset and neural-network training pipeline were implemented using PyTorch (Paszke et al. 2019). The balanced nominal dataset was divided once using a fixed stratified 70/15/15 split into 28,000 training, 6,000 validation, and 6,000 held-out test spectra. Each subset contained equal class proportions, and the subsets were verified to be disjoint. The training subset was used for backpropagation and weight updates, the validation subset was used for hyperparameter and checkpoint selection, and the held-out

test subset was reserved for final evaluation. This separation prevents the reported test performance from being optimized indirectly through repeated model selection. Feature-wise standardization was fitted using the training subset only and applied unchanged to the validation, test, experimental, baseline, robustness, and out-of-distribution samples. Fitting the standardization transformation using the training subset alone prevents information from the validation or test subsets from influencing model development. Standardization also places the 75 input features on comparable numerical scales, which facilitates stable neural-network optimization. Hyperparameters were optimized with Optuna (Akiba 2019) using only the training and validation subsets. A Tree-structured Parzen Estimator (TPE) sampler initialized with seed 314159 was used together with a median pruner having five startup trials. The optimization objective was the mean best validation macro-F1 across training seeds 101 and 202. Macro-F1 assigns equal importance to the F1 score of each class and was therefore used to avoid selecting a configuration that performed well only for the easiest class. Averaging the objective across two training seeds reduced the probability of selecting a configuration because of one unusually favorable random initialization. Sixty usable trials were recorded, comprising 35 completed and 25 pruned trials. Each trial used at most 75 epochs with an early-stopping patience of 10 epochs. The search space included one to four hidden layers, 32-512 units per layer in increments of 32, learning rates from 10^{-5} to 3×10^{-3} , weight decay from 10^{-7} to 10^{-3} , dropout from 0.0 to 0.4 in increments of 0.1, and batch sizes of 32, 64, 128, or 256. The selected DNN contained four hidden layers with 416, 416, 416, and 480 units, corresponding to 580,644 trainable parameters. The network used rectified linear-unit activations, a learning rate of 0.000282869, weight decay of 0.000377569, zero dropout, a batch size of 256, cross-entropy loss, and the AdamW optimizer. After the configuration was frozen, final models were independently trained using seeds 11, 22, 33, 44, and 55 for at most 250 epochs with an early-stopping patience of 25 epochs. These five runs used the same frozen architecture, hyperparameters, and data partitions but different random initializations. Repeated training allowed the variability of the DNN performance and experimental assignments to be quantified, avoiding reliance on a single potentially favorable initialization. Validation loss determined checkpoint selection. Each selected checkpoint was then evaluated once on the untouched test subset. No backpropagation, checkpoint selection, or hyperparameter adjustment was performed using the test subset. Its metrics therefore estimate generalization to unseen spectra generated under the nominal simulation assumptions. The analysis used Python 3.12.4, PyTorch 2.5.0 with CUDA 12.1, NumPy 2.0.1, scikit-learn 1.6.0, and Optuna 4.5.0. The resulting DNN architecture is illustrated schematically in Figure 1.

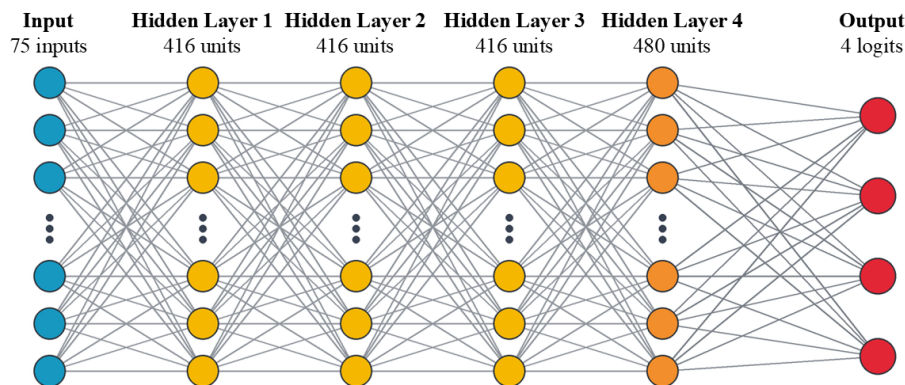


Figure 1: Schematic diagram of the implemented DNN architecture. The network receives 75 standardized intensity values and contains four fully connected hidden layers with 416, 416, 416, and 480 units, respectively, followed by four output logits corresponding to the four model classes.

To propagate the reported measurement uncertainties through the frozen classifiers, repeated plausible realizations of the measured spectrum were generated. Each realization was formed by sampling one possible intensity in every experimental bin and did not involve resampling the K -matrix model parameters. For experimental inference, 3,000 normal and 3,000 uniform perturbations of the measured spectrum were generated. Within each of the 75 selected bins, an energy coordinate was sampled uniformly within the reported lower and upper bin boundaries and retained as metadata. Under Gaussian perturbation, the intensity was drawn from a normal distribution centered on the reported experimental intensity, with a standard deviation equal to one-third of the reported upper statistical uncertainty. Under uniform perturbation, the intensity was drawn uniformly from the interval between the reported intensity minus its upper statistical uncertainty and the reported intensity plus its upper statistical uncertainty. The upper statistical uncertainty applies to both tails since the experimental uncertainties are symmetric. Fixed random seeds 7001 and 7002 were used for the normal and uniform prescriptions, respectively. Since the classifiers used the common energy grid, only the 75 sampled intensity values entered the classifiers after application of the standardization fitted using the training subset. The same experimental realizations were reused for every independently trained deep neural network and baseline classifier. The purpose of these realizations was to determine whether uncertainty-level changes in the measured intensities altered the assigned class.

RESULTS AND DISCUSSION

Comparable χ^2 values from different K -matrix models

The initial iminuit fits show that all four tested K -matrix parametrizations can reproduce the $P_{c\bar{c}}(4312)^+$ region with comparable chi-squared values as seen in Figure 2.

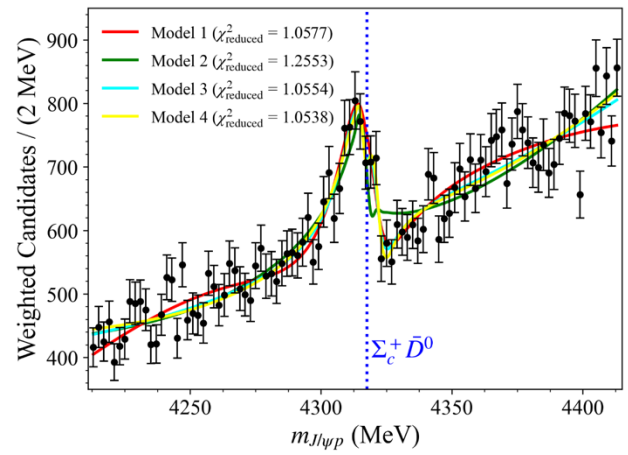


Figure 2: Best-fit lineshapes from four competing two-channel parametrizations using iminuit chi-square minimization. The models produce comparable agreements with the data despite representing different mechanisms which illustrate the limitations of traditional fitting.

The numerical fit results are summarized in Table 3 where the analysis used $N = 101$ experimental data points. The number of varied parameters differs among the four models due to the additional resonant and nonresonant terms.

Table 3: Summary of best-fit results for the four K -matrix models.

Model	N	k	Degrees of Freedom	χ^2_{reduced}
1	101	12	89	1.0577
2	101	9	92	1.2553
3	101	16	85	1.0554
4	101	13	88	1.0538

Table 4: Best-fit K -matrix parameters and fitting uncertainties for the four model classes.

Model	m_α	Γ_α^0	m_β	Γ_β^0	$\gamma_{\alpha 1}$	$\gamma_{\alpha 2}$	$\gamma_{\beta 1}$	$\gamma_{\beta 2}$	c_{11}	c_{12}	c_{22}	s_1	s_2
1	4314.9 ± 1.8	9.1 ± 4.8	-	-	-2.56 ± 1.12	-1.70 ± 1.60	-	-	1.77 ± 0.77	3.78 ± 0.64	-4.3 ± 0.8	0.7 ± 0.05	0.3 ± 0.07
2	4310.90 ± 8.1	16.2 ± 8.9	-	-	-9.3 ± 1.8	-29.2 ± 6.0	-	-	-	-	-	0.0004 ± 8.43	1.2 ± 0.11
3	4310.9 ± 6.7	16.0 ± 5.2	4323.9 ± 3.2	16.0 ± 5.1	-2.10 ± 0.71	-5.91 ± 2.20	0.35 ± 0.27	-4.16 ± 1.3	-0.37 ± 0.75	-0.03 ± 0.16	1.00 ± 1.8	0 ± 0.19	1.0 ± 0.22
4	4311.4 ± 4.4	7.6 ± 5.9	4323.90 ± 4.0	9.5 ± 5.8	3.4 ± 0.8	10.4 ± 3.0	0.18 ± 0.3	-6.0 ± 1.8	-	-	-	0.35 ± 0.15	-1.8 ± 0.21

The corresponding best-fit parameters and their fitting uncertainties are reported in Table 4. Uncertainties correspond to the parameter errors obtained from the iminuit fitting procedure. This result confirms the model-selection ambiguity motivating the present work. Since the four models differ in their physical content, similar goodness-of-fit values do not imply equivalent physical interpretations. In particular, additional resonance terms can increase fitting flexibility, while nonresonant terms can absorb threshold-related distortions in the lineshape. Therefore, chi-squared alone is insufficient to distinguish whether the observed enhancement is better represented by additional resonant structure or by a simpler resonance model supplemented by nonresonant dynamics.

DNN classification performance and robustness

Across five independent training seeds, the DNN achieved a mean held-out test accuracy of 82.61%, with a sample standard deviation of 0.72 percentage points and a 95% confidence interval of 81.72%-83.50%. Mean macro-F1 was 82.37%, with a sample standard deviation of 0.74 percentage points and a 95% confidence interval of 81.45%-83.29%. Mean balanced accuracy was 82.61%, mean weighted-F1 was 82.37%, and mean log loss was 0.4631. The validation-selected epochs were 68, 68, 64, 67, and 89 for seeds 11, 22, 33, 44, and 55, respectively. The training and validation trajectories for the representative seed 22 are shown in Figure 3. The loss, accuracy, and macro-F1 curves permit visual assessment of convergence and of the relationship between training and validation performance. Seed 22 was selected for visualization since its held-out macro-F1 score was closest to the median across the five independently trained DNNs. The model parameters used for this run were taken from the validation-selected epoch of 68.

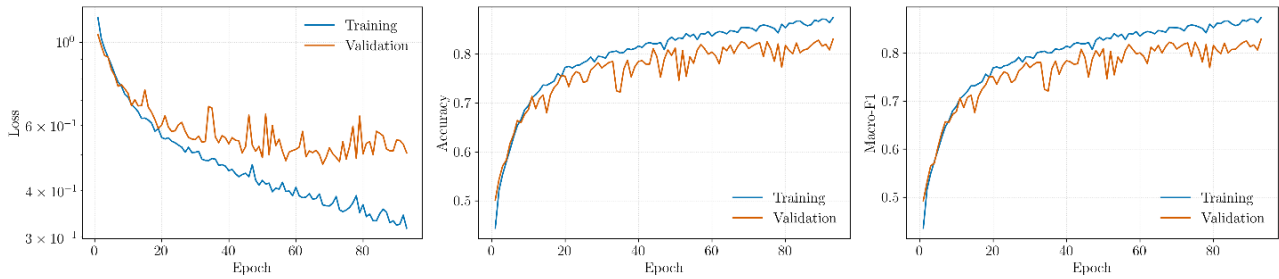


Figure 3: Training and validation trajectories for the representative DNN initialized with seed 22: (left) loss, (middle) accuracy, and (right) macro-F1 score. The final model parameters were selected at epoch 68 based on minimum validation loss. Seed 22 was selected for visualization because its held-out macro-F1 score was closest to the median across the five independently trained DNNs.

Class 2 had the highest mean F1 of 0.890, whereas Class 3 had the lowest mean recall of 0.706 and the lowest mean F1 of 0.759. Thus, the two-resonance model with nonresonant terms (Class 3) was the most difficult class to recover on the nominal held-out test set. A full summary of these details can be seen in Table 5.

Table 5: Mean precision, recall, and F1 scores on the 6,000-spectrum held-out synthetic test set, averaged across five independently trained DNNs. The test set contained 1,500 spectra from each class.

Class	Precision	Recall	F1
1	0.794	0.804	0.799
2	0.827	0.964	0.890
3	0.822	0.706	0.759
4	0.865	0.830	0.847

Figure 4 presents the row-normalized confusion matrix for seed 22 at its validation-selected epoch of 68. Seed 22 was selected for visualization because its macro-F1 score was closest to the median across the five prespecified runs. The confusion matrix confirms that the largest classification difficulty involved recovering Class 3, while Class 2 obtained the highest recall. The results in Figure 4 represent one illustrative seed, while all five runs contributed to the summary statistics reported above and in Table 5.

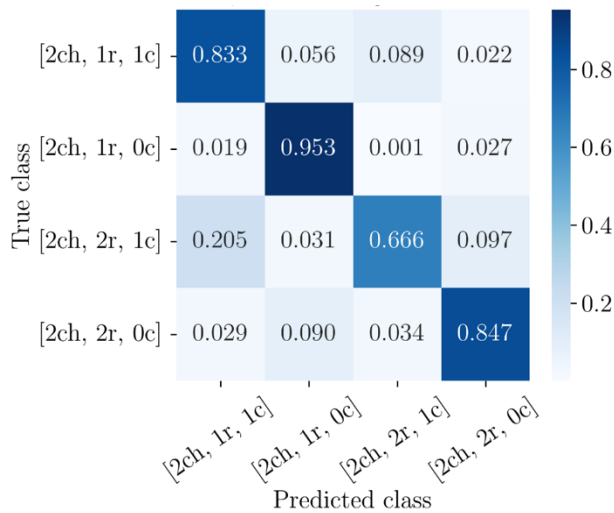


Figure 4: Row-normalized confusion matrix evaluated on the 6,000-spectrum held-out synthetic test set for seed 22 at validation-selected epoch 68.

The trained DNNs were subsequently applied to perturbed realizations of the measured spectrum. As shown in Figure 5, the normally perturbed realizations were assigned to Class 1 with a mean frequency of 2.24% and to Class 3 with a mean frequency of 97.76%; the mean frequencies for Classes 2 and 4 were zero. The Class 3 frequency ranged from 88.8% to 100% across the five DNN seeds. Under uniform perturbations, the mean frequencies were

3.23% for Class 1, 0% for Class 2, 96.53% for Class 3, and 0.23% for Class 4. The Class 3 frequency ranged from 85.8% to 99.8%. These frequencies represent repeated classifier assignments for perturbations of one measured spectrum and should not be interpreted as posterior probabilities that a physical model is correct.

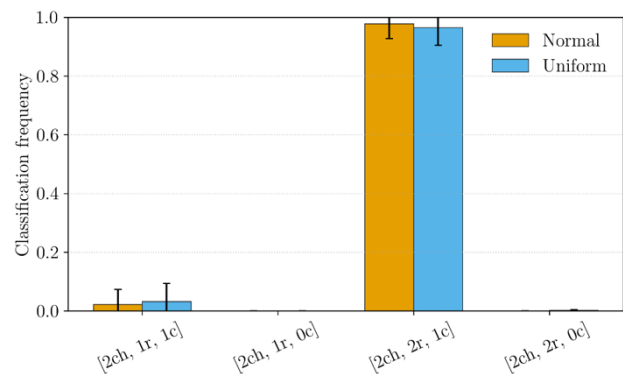


Figure 5: Mean DNN classification frequencies for 3,000 normally perturbed and 3,000 uniformly perturbed realizations of the measured spectrum. Error bars show the variation across the five independently trained DNNs.

To determine whether the complexity of the DNN provided a reproducible classification advantage, it was compared with simpler classifiers evaluated using exactly the same data partitions. The DNN result was compared with multinomial logistic regression and random forest using the same fixed training and held-out test partitions. As summarized in Figure 6, logistic regression achieved a test accuracy of 49.20% and macro-F1 score of 48.45%, while random forest achieved an accuracy of 62.67% and macro-F1 score of 62.30%. The DNN achieved the strongest performance, with a mean accuracy of 82.61% and mean macro-F1 score of 82.37%. However, the experimental assignment was not classifier-independent since the logistic regression assigned essentially all experimental realizations to Class 2, whereas random forest assigned most realizations to Class 1 with relatively diffuse class probabilities. The experimental DNN preference for Class 3 must therefore be interpreted as conditional on the selected classifier and nominal synthetic-data distributions.

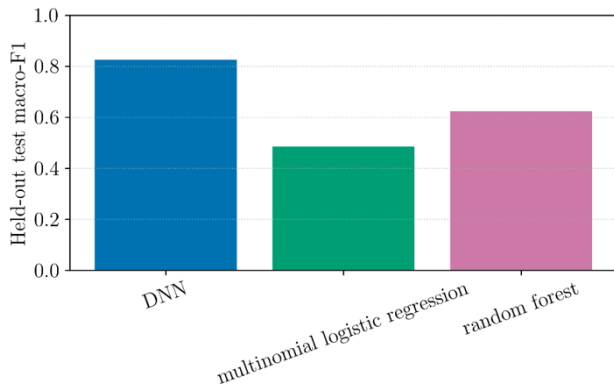


Figure 6: Held-out synthetic classification performance of the DNN, multinomial logistic regression, and random forest evaluated using the same fixed data partitions. DNN values show the mean across five independently trained networks.

The sensitivity of the DNN to departures from the nominal synthetic-data distribution is summarized in Table 6. Mean macro-F1 decreased from 0.8237 under nominal conditions to 0.797, 0.720, 0.551, and 0.407 after adding 1%, 2%, 5%, and 10% noise, respectively. Mean macro-F1 was 0.783 for wider resonance widths, 0.683 for wider resonance masses, 0.734 after retraining on an imbalanced subset, and 0.167 for the alternative energy interval. A macro-F1 value close to the nominal value of 0.824 indicates comparatively stable discrimination, whereas a substantial reduction indicates that the classifier does not generalize reliably to that modified condition. The value of 0.167 for the alternative energy interval therefore indicates severe degradation rather than ordinary run-to-run variation. The classifier was therefore comparatively stable under the tested background and coupling modifications but was not robust to all distribution shifts, particularly changes in the energy interval.

Table 6: Mean macro-F1 scores under nominal and modified synthetic-data conditions.

Condition	Mean macro-F1
Nominal	0.824
1% added noise	0.797
2% added noise	0.720
5% added noise	0.551
10% added noise	0.407
wider resonance widths	0.783
wider resonance mass	0.683
alternative energy interval	0.167

Since the DNN was trained only on four K -matrix families, it cannot identify mechanisms outside these classes such as a three-channel parametrization. The result may also depend on the parameter ranges and background prescriptions used in generating the synthetic lineshapes.

CONCLUSION

This work examined model-selection ambiguity among four two-channel K -matrix descriptions of the $P_{c\bar{c}}(4312)^+$ lineshape. Under a fixed stratified train-validation-test protocol, five independently trained DNNs achieved a mean held-out accuracy of 82.61% and mean macro-F1 of 82.37% on nominal synthetic spectra. The DNN outperformed multinomial logistic regression and random forest on the common synthetic test set. When applied to perturbed realizations of the experimental spectrum, the repeated DNNs predominantly assigned the samples to Class 3, the two-

resonance model with nonresonant terms. This assignment was reproducible across the five DNN initializations but was not classifier-independent, as logistic regression and random forest favored different classes. Performance also degraded under increasing noise, wider mass ranges, class imbalance, and especially an alternative energy interval. The DNN result should therefore be regarded as supportive, model-conditional evidence within the four simulated families, nominal parameter distributions, fixed energy interval, and classifier assumptions. It does not determine pole positions, establish a unique physical mechanism, or provide posterior model probabilities. Future work should extend the simulated model space and employ calibrated simulation-based inference alongside explicit amplitude and pole analyses.

ACKNOWLEDGMENTS

KTR Cervantes would like to acknowledge the support from the Nuclear and Hadron Physics (NHP) Team and the Pisayan organization.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

Klarence Tomas R. Cervantes: Writing – Original draft, Methodology, Investigation, Formal analysis. Denny Lane B. Sombillo: Methodology, Supervision, Writing – Review and Editing.

REFERENCES

- Aaij R, et al. Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$. *Physical Review Letters* 2019; 122(22):222001. <https://doi.org/10.1103/PhysRevLett.122.222001>.
- Chen C, Xie Y, Xu H, Zhang Z, Zhou D, She Z, Chen G. Exotic states $P_c(4312)$, $P_c(4440)$, and $P_c(4457)$ in pp collisions at $\sqrt{s}=7, 13$ TeV. *Physical Review D* 2022; 105(5):054013. <https://doi.org/10.1103/PhysRevD.105.054013>.
- Chung SU, Brose J, Hackmann R, Klempt E, Spanier S, Strassburger C. Partial wave analysis in K -matrix formalism. *Annalen der Physik* 1995; 4(5):404-430. <https://doi.org/10.1002/andp.1995570504>.
- James F, Roos M. Minuit: A system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications* 1975; 10(6):343-367. [https://doi.org/10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9).
- Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2019:2623–2631. <https://doi.org/10.48550/arXiv.1907.10902>.
- Pagayon JB, Cervantes KTR, Sombillo DLB. Deep learning topological inference-guided T_{cc}^+ pole parameter extraction. *Physical Review D* 2026; 114(1):014049. <https://doi.org/10.1103/nr94-jzjxj>.

- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* 2019;32:8024–8035. <https://doi.org/10.48550/arXiv.1912.01703>.
- Ruangyoo W, Phumphan C, Chen C, Limphirat A, Yan Y. P_c resonances in the compact pentaquark picture. *Journal of Physics G: Nuclear and Particle Physics* 2022; 49(7):075001. <https://doi.org/10.1088/1361-6471/ac58af>.
- Santos LM, Chavez VAA, Sombillo DLB. Pole structure of $P_\psi^N(4312)^+$ via machine learning and uniformized S -matrix. *Journal of Physics G: Nuclear and Particle Physics* 2025; 52(1):015104. <https://doi.org/10.1088/1361-6471/ad8ee3>.
- Silva-Castro JA. On the possible virtual state nature of the LHCb $P_c(4312)^+$ signal. *Journal of Physics Conference Series* 2020; 1610(1):012015. <https://doi.org/10.1088/1742-6596/1610/1/012015>.
- Xiao CJ, Huang Y, Dong YB, Geng LS, Chen DY. Exploring the molecular scenario of $P_c(4312)$, $P_c(4440)$, $P_c(4457)$. *Physical Review D* 2019; 100(1):014022. <https://doi.org/10.1103/PhysRevD.100.014022>.