

Meson width predictions and symmetry emergence within the deep neural network

Xin Tong,¹ Wei Feng,² Weiwei Xu^{3,4} Chao-Hsi Chang,^{5,6} Guo-Li Wang,⁷ Qiang Li^{1*}

¹*School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China*

²*School of Information Mechanics and Sensing Engineering, Xidian University, Xi'an, 710071, China*

³*Shandong Institute of Advanced Technology (SDIAT), 250100, Jinan, China*

⁴*Shandong University, 250100, Jinan, China*

⁵*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

⁶*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

⁷*Department of Physics, Hebei University, Baoding 071002, China*

E-mail: tongxin@mail.nwpu.edu.cn, wfeng@xidian.edu.cn,
Weiwei.Xu@cern.ch, zhangzx@itp.ac.cn, wgl@hbu.edu.cn,
liruo@nwpu.edu.cn

ABSTRACT: We build a deep neural network model to predict meson widths from quantum numbers and masses based on the Transformer architecture. A Gaussian Monte-Carlo data enhancement method is adopted to enhance the meson data by considering the experimental errors, which significantly increase the data samples and improve the robustness and generalization performance of the model. With the meson widths ranging from $\sim 10^{-14}$ to 625 MeV, the relative errors of the predictions behave 0.07%, 1.0%, and 0.14% in the training set, the test set, and all the data, respectively. The width predictions are presented for the currently discovered mesons and some theoretically predicted states. We also use the model as a probe to study the quantum numbers and inner structures for some undetermined states. Furthermore, this data-driven model is investigated to show well charge conjugation symmetry and approximate isospin symmetry, which is consistent with the physical phenomena. The results indicate that the deep neural network has powerful learning and inference abilities to describe and explore the hadron structures and the complicated interactions in particle physics.

KEYWORDS: Meson width predictions; Hadron identification; Deep neural network; Symmetry emergence

Contents

1	Introduction	1
2	Meson encoding methods and data enhancement	3
2.1	Features used to denote the mesons	4
2.1.1	Quark contents encoding	4
2.1.2	Complete input features	5
2.2	Gaussian Monte-Carlo data enhancement	6
3	Deep neural network model	8
3.1	Backbone of the model	8
3.2	Loss function to describe the mesons widths	10
3.3	Training and validation	10
4	Numerical results of the meson widths predictions	11
4.1	Overall performance of the model	11
4.2	Meson width predictions and quantum numbers identification	13
4.3	Symmetries reproduced by the neural network	14
4.3.1	Charge conjugation symmetry	15
4.3.2	Isospin symmetry	17
5	Summary	19
A	Spectra of the predicted meson widths	20

1 Introduction

Particle physics aims to understand the essence of fundamental particles and their interactions in nature. Quarks and gluons are bounded by strong interactions to form hadrons, namely, mesons and baryons. For the quark confinement [1, 2], hadron properties play key roles to understand the strong interactions, which are directly determined by the dynamic of quarks and gluons inside. The strong interactions are described by the quantum chromodynamics (QCD). In the high-energy region (short distance), QCD's perturbation calculations have achieved great success, and its theoretical predictions are highly consistent with experimental results, which becomes one of the most important experimental verifications of the Standard Model for particle physics. However, in the low-energy region (long distance), the non-perturbative effects of QCD make analytical calculations quite difficult, which leads to predictions of hadron lifetime quite difficult in theory.

Mesons as the quark-antiquark bound states play important roles for studying the non-perturbative QCD interactions. The meson masses and decay widths are hence the direct

reflection of the inside interaction dynamics. Generally speaking, the total width of a meson is a summation of all the possible decay channels, while the decay behaviors are directly dependent on the meson quantum numbers and the involved interactions. Researches of mass spectra and total widths provide rich information on the internal structure and interactions of mesons.

There exist two main goals in hadron physics. One is to obtain the hadron mass and decay behaviors based on the quantum numbers, while the other is to obtain the quantum numbers from the observed mass and lifetime (or total width equivalently). With the QCD-inspired potentials and quark model, by solving the corresponding bound states equation, mass spectra for traditional hadrons can be well established in theory [3–5] though some exotic hadrons discovered in recent years are not included [6–10]. However, it is quite difficult to obtain the total widths based on the traditional calculations of Standard Model dynamics for the extreme theoretical complications. Hence it is quite necessary and helpful to predict lifetime spectra from a new reliable method.

In recent years, with the continuous operation of high-precision experiments such as LHCb, CMS, ATLAS, Belle II, and BESIII, significant progress has been achieved in hadron spectroscopy researches, and large number of new hadron states have been discovered including quite a lot of exotic state candidates [11]. These new discoveries greatly enrich our understanding of the hadron family, but at the same time pose a serious challenge to the current theoretical framework. The internal structure, quantum number, mass, and decay width of many particles have not yet been fully determined [6–12]. Therefore, systematic research on the meson spectra, especially reliable predictions for the total widths has become quite an important problem in hadron physics. All these challenges mentioned above highlight the necessity of exploring alternative data-driven methods.

On the other hand, deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction [13]. In the last ten years the deep neural networks have become a powerful paradigm for many complex problems since several key breakthroughs happened in this field. The application of back-propagation algorithm laid the theoretical foundation for effectively training neural networks [14]. Then a series of powerful architectures then have made great achievements in computer science areas, such as the recurrent neural network (RNN) [15, 16], the long short-term memory network (LSTM) [17] and neural probabilistic representation [18] for natural language and other sequential data processing, and the convolutional neural network (CNN) [19] for image recognition. The residual networks (ResNet) overcame the bottleneck of gradient vanishing in deep networks, making it more effective to construct and train extremely deep networks [20]. This series of advancements ultimately gave birth to the Transformer architecture, which is considered as the most advanced and powerful architecture characterized by its self-attention mechanism [21].

Currently the deep neural networks have also become a powerful paradigm for solving complex problem in high energy physics [22–26]. Its core capability lies in extracting physical information from high-dimensional complex experimental data, which applies in particle track reconstruction [27], jets classification [28–30], the search for new physics [31, 32], and mass spectra of atomic nuclei [33, 34], etc. Encouraged by these advances, several recent

researches have also begun to directly focus on meson width topic. A deep neural network (DNN) model has been developed to predict mass and width of meson states [35]; another work compared the performance of various machine learning algorithms on this problem [36]. These studies have validated the feasibility of data-driven approaches on this issue, but they also face several core challenges that constrain their development, for example, the performance of deep learning models is highly dependent on accurate hadron encoding and large-scale datasets, while the experimental hadron data is relatively sparse and also exists a large number of degeneracy.

This work aims to systematically explore a data enhancement strategy based on physical principles, and examine whether a deep neural network model trained in this way can successfully predict the meson widths and further reveal the underlying physical symmetries behind the data. To this end, we have developed a deep learning framework that starts with a systematic feature engineering approach to address meson encoding. The framework adopts the Feature Tokenizer Transformer (FT-Transformer) architecture [37] as the core model and enhances the data by the Gaussian Monte-Carlo method to effectively train the model and improve its robustness.

The trained model demonstrates strong predictive ability within a huge range spanning nearly 17 orders of magnitude, from about 10^{-14} MeV to about 625 MeV. The relative error on an unprecedented conventional meson test set reaches the level of one percent, demonstrating excellent generalization performance. And we further verified the physical consistency of the model. Our research shows that this data-driven model can successfully reproduce the fundamental symmetry in particle physics without any prior physical constraints. The predicted results not only strictly follow the charge conjugation symmetry, but also quantitatively reflect the approximate isospin symmetry for meson widths.

In addition, the model exhibits different response to particles with different internal structures, which also provides us a new insight to explore hadrons. We find that although the model is able to learn the patterns of traditional mesons and some exotic states contained in the training data, its predictions exhibit significant biases to the well-known exotic state candidates such as $D_{s0}^*(2317)$ and $\chi_{c1}(3872)$. It then suggests that these states may have different internal structures than the particles in the training set, which also suggests that the model's predictions may serve as a data-driven probe to assist in identifying and classifying potential non-traditional hadron states.

This paper is organized as follow. In Section 2, we first state the used data, data enhancement method, and the data encoding scheme. Then in Section 3 we introduce the neural network model we used to deal with the meson width problems, the loss function, the train methods and also the hyper parameters. The obtained results, predictions, and discussions are presented in Section 4. Finally we give a brief summary and outlook.

2 Meson encoding methods and data enhancement

In a predicting task based on deep neural network, the performance of the model is highly dependent on quality of the input data and representation to the data features. In this section, we introduce the complete procedure of the data processing for the meson width

predictions. First, we construct a complete feature vector which can definitely and precisely representing quantum states of mesons; and then, we explore data enhancement strategy to solve the important issue of limited data for deep learning.

2.1 Features used to denote the mesons

A particle is represented by a set of quantum numbers, such as spin J , parity P , charge conjugation C , G , isospin I and its third component I_3 , the quark contents, mass and total width (or lifetime equivalently). These quantum numbers will also be called features within the machine learning terminology. The non-relativistic quantum numbers, inner spin S and orbital angular momentum L , are often used to denote the mesons in literature, however, the two ones are neither the good quantum numbers in describing the particles nor the physical observable in experiments. We will not use the non-relativistic S and L as the meson features to train the model.

According to the conventions in deep learning, we categorize the aforementioned features into two groups. One group describing the inner symmetry of the mesons consists of the categorical features, including the P , C and G ; and the other group consists of the numerical features, including the I , I_3 , J , and the quark flavor coefficients. The features in the former one are assumed to be numerical-independent and different values just indicate different categories, while the latter ones are assumed to be numerical-dependent. It should be pointed out that the above category scheme is not rigorous.

Also for some mesons, the information of the experimental data are not complete while some quantum numbers may be missing. We apply the following scheme to deal with the case of feature missing. For the categorical features, the inapplicable features and the undetermined features are considered as two different independent discrete features respectively, for example, the values of C -parity for π^+ (inapplicable) and $X(4160)$ (undetermined currently) are assigned as two different categorical labels. For the numerical features, we developed a dual-path embedding framework that decouples a feature's value from its certainty. We first transform the original numerical value x_{num} into a number-mask pair as

$$(x'_{\text{num}}, x_{\text{mask}}) = \begin{cases} (0, 0) & \text{undetermined } x_{\text{num}}, \\ (x_{\text{num}}, 1) & \text{otherwise.} \end{cases} \quad (2.1)$$

Then every numerical feature is coupled with an independent feature which gives the undetermined data a specific mask value 0 to distinguish it from the usual sample. This pair is then processed by an augmented embedding module. Namely, the embedding of this number-mask pair will then be fed into a linear layer to give a d -dimensional fused feature vector which encoded both the value and its certainty.

2.1.1 Quark contents encoding

The first challenge to use neural network to deal with the hadron spectroscopy is the encoding of quark contents inside hadrons. Within the traditional quark model, mesons consist of quark-antiquark pair. A direct encoding scheme is to use a 10-dimensional binary vector to represent the five quarks and the corresponding antiquarks, namely, quark

contents can be expressed as

$$\mathbf{q} = (u, \bar{u}, d, \bar{d}, s, \bar{s}, c, \bar{c}, b, \bar{b}), \quad (2.2)$$

where the i th element $q_i = 1$ denotes that the corresponding quark exists in the meson while $q_i = 0$ not; we did not take into account the (anti)top quark. However, this naive binary encoding cannot describe the mesons in flavor superposition states. The quark contents of many neutral mesons are the superposition of several quark-antiquark pair, for example, the flavor wave function of neutral pion is usually expressed as

$$\pi^0 = \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}). \quad (2.3)$$

In order to represent the information of these flavor superposition states, we directly pass the corresponding flavor coefficients into the quark content vector \mathbf{q} , for example, the π^0 and η are encoded as

$$\mathbf{q}_{\pi^0} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0, 0, 0), \quad (2.4)$$

$$\mathbf{q}_{\eta} = (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0, 0, 0, 0). \quad (2.5)$$

Notice that here we only care about the absolute values of the coefficients which are aimed to reflect the weights of the corresponding quarks inside the mesons, while the phase information is omitted. This encoding scheme for the quark contents can simplify the structure of the input features, and then the task for grasping deep symmetry information of the mesons can then be finished by the deep neural network with powerful self-attention mechanism.

2.1.2 Complete input features

Another challenge is that the aforementioned quantum numbers cannot definitely and uniquely represent a particle state. Several hadrons may occupy the same quantum numbers (J, P, C, G, I, I_3) and quark contents but correspond to different masses and widths. There exists degeneracy if we just use the features above. For example, $\rho(770)$ and $\rho(1450)$ share exactly the same quantum numbers $I^G(J^{PC}) = 1^+(1^{--})$ and quark contents $[u\bar{d}]$ but have quite different masses and widths. If these degenerate features were used as input, it would lead to the one-to-many ambiguous mappings and lower the performance of the model.

Within the quark model, these ambiguities may originate from several aspects. From one hand, the different principle quantum numbers cause different particle masses, such as J/ψ , $\psi(2S)$, $\psi(3770)$, $\psi(4040)$, $\psi(4160)$, $\psi(4260)$, $\psi(4360)$, and $\psi(4415)$, $\psi(4660)$. The ψ family mesons all share the same $J^{PC} = 1^{--}$ and quark content $[c\bar{c}]$. Although the usual principle quantum number n may be used to label different states, there are several reasons prevent us using this feature. First, n is a non-relativistic model-dependent quantity; second, n is not a directly experimental observable; and finally, it is difficult to precisely

establish all the principle quantum numbers of these mesons, since there may also exist the so-called $2S$ - $1D$, $3S$ - $2D$, ..., mixing effects, namely, even the non-good principle quantum number n is missing for most mesons. On the other hand, the unnatural parity ($J^P = 1^+, 2^-, 3^+, \dots$) mesons also have the extra inner freedom which makes the corresponding masses undetermined from the current features, for example, the $J^P = 1^+$ charmed mesons $D_1(2420)$ and $D_1(2430)$, and the charmed-strange meson $D_{s1}(2536)$ and $D_{s1}(2460)$. These mesons are usually considered as the two physical states from the non-relativistic 1L_L - 3L_L mixing, and are distinguished by an extra introduced mixing angle parameter. However, this parameter is also a model-dependent quantity and cannot be directly measured from experiments. Hence a definite quantum number is also missing to represent the inner freedom of unnatural parity mesons, which makes it impossible to determine the masses and widths together for such mesons from a set of good quantum numbers.

All these ambiguities make it difficult to directly use the aforementioned features to train a deep neural network model to predict the meson masses and widths. On the other hand, the meson mass spectra can be well obtained by the quark model calculations [3], and a more urgent and difficult problem is to calculate the meson widths. Then the mass itself can behave as an powerful input feature and be used to solve the degenerate issues above. Compared to the total width, the meson masses are much easier to obtain based on the quark model calculations or the lattice QCD simulation, which has also become part of the motivation to predict the meson widths while not the masses. By the way, we also introduce an auxiliary feature N with continuous values to represent the mass ordering, which is defined by normalizing the central value of the meson mass with m_{π^0} . Then the feature N behaves as an extensible relative one and any new discovered meson can be easily embedded into this system. Notice that N is not an independent feature but a auxiliary feature derived from the meson mass, designed to help the model better interpret the mass spectrum. Finally, we can encode a meson by a feature vector as

$$\mathbf{v} = (J, P, C, G, I, I_3, u, \bar{u}, d, \bar{d}, s, \bar{s}, c, \bar{c}, b, \bar{b}, m, N). \quad (2.6)$$

And our aim is to use the above feature vector \mathbf{v} to predict the meson total width Γ based on the deep neural network.

2.2 Gaussian Monte-Carlo data enhancement

The data used for training and test in this work are from the experimental meson data in PDG 2024 and 2025 update [11], which include 400 mesons. The most important challenge for this task is the lack of data. Though more than 400 mesons are discovered in experiments, it is still quite small for training a deep neural network with good generalization ability. To overcome the bottleneck of data sparsity, data enhancement is considered an effective strategy. There are two main implementation paths in machine learning: one is to use deep generative models to learn and generate new synthetic samples [38]; the second is to generate physically reasonable pseudo-data by Monte-Carlo sampling of experimental data. The latter one can expand the limited datasets and also improve model robustness.

Therefore, it is widely used in particle physics, nuclear physics, and astrophysics [39–43]. These two paths provide complementary perspectives for addressing data sparsity issues.

To overcome this problem, we implement the Gaussian Monte-Carlo data enhancement based on the uncertainties of experimental data in this work. The experimental measurement of a meson mass m is always accompanied by the experimental uncertainties Δm from both the statistical error and the systematic error, which in turn define the confidence interval of the corresponding experimental data. The core idea of the Gaussian Monte-Carlo data enhancement is to produce the Monte-Carlo (MC) mass data for every hadron based on the Gaussian distribution with the standard deviation σ defined by the experimental error Δm .

The experimental measurement of a hadron mass in PDG is usually denoted as $m = (M_c^{+\Delta m_+}_{-\Delta m_-})$ with M_c representing the central value while Δm_{\pm} the asymmetry errors. The concrete procedures are as follows. First, we naively average the asymmetry errors Δm_+ and Δm_- to obtain the standard deviation $\sigma = \frac{1}{2}(\Delta m_+ + \Delta m_-)$. Then we assume the mass data subject to the Gaussian distribution $\mathcal{N}(M_c, \sigma^2)$. Finally, for each meson, we implement the MC sampling to obtain N_{aug} mass data within the range of $M_c \pm k\sigma$, where k is the sampling constants used to control the distribution range of meson mass, and $k = 1$ and $N_{\text{aug}} = 500$ are set in the current research.

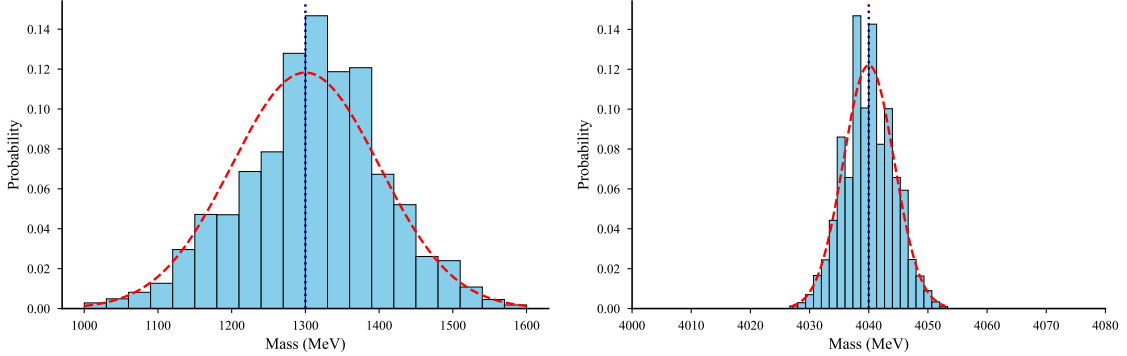


Fig. 1: The mass distribution of $\pi(1300)$ and $\psi(4040)$ after the Gaussian enhancement.

All the data points produced by MC sampling from an original meson share the same quantum numbers, with only the mass feature being varied. After this data enhancement, each meson state is effectively expanded, and the total data sample is increased from ~ 400 to ~ 200000 . This process simulates the statistical nature of experimental detections and efficiently enhances the robustness of the neural network model. In Fig. 1 we show the resulting mass distributions for $\pi(1300)$ and $\psi(4040)$. While the augmented mass values introduce a necessary variance, this can also obscure the precise hierarchy between different meson states. The auxiliary feature N is therefore particularly important in this context, as it provides a stable, normalized index of each meson’s position in the mass spectrum, complementing the fluctuating mass value.

3 Deep neural network model

After finishing the feature construction and data enhancement, we now choose a suitable neural network architecture and build the deep learning model to deal with the task for meson width predictions. We will introduce the model architecture, the loss function, and the training and test strategy.

3.1 Backbone of the model

Among all these deep neural network architectures, Transformer [21] model has an outstanding performance in many aspects for its self-attention mechanism. Compared with the traditional fully connected network or the convolutional neural network, the Transformer model is able to process the input data in parallel and can also capture the long distance dependencies of data. Though designed for the natural language process, its core idea has also been successfully applied to many science fields to deal with the deeply complicated inner mapping and dynamic. Considering the fact that we are dealing with the tabular data with complicated heterogeneous features, we choose the Feature Tokenizer Transformer (FT-Transformer) as backbone of the model. The FT-Transformer is designed specifically for tabular data and can effectively capture complex nonlinear dependence between different features by the self-attention mechanism. The main parts of the FT-Transformer is shown in Fig. 2.

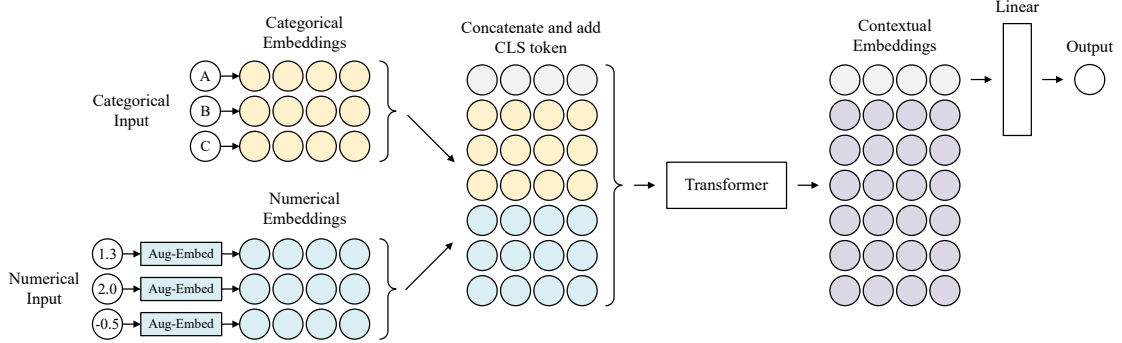


Fig. 2: The main parts of the FT-Transformer architecture. Both the numerical and categorical features are first embedded to d -dimensional vectors, which are then together with the [CLS] token fed into the typical Transformer encoder layer.

The FT-Transformer first makes feature tokenization to map the heterogeneous raw input features (both the categorical and continuous features) to a unified high-dimensional vector space. For each categorical feature (such as particle quantum numbers P , C and G), the model converts the discrete integer index into a dense feature vector $\mathbf{e}_{\text{cat}} \in \mathbb{R}^d$ through an embedding layer with d representing the embedding dimension. The embedding is implemented by a lookup table for the categorical features.

For each continuous feature (such as meson mass), the model employs an augmented embedding module detailed in Sec. 2.1 and illustrated in Fig. 3(a), which performs a

linear transformation on each continuous feature value and then maps it into the same d -dimensional space, namely,

$$\mathbf{e}_{\text{num}} = x_{\text{num}}\mathbf{w} + \mathbf{b}, \quad (3.1)$$

where $\mathbf{w}, \mathbf{b} \in \mathbb{R}^d$ denote the weight and bias vector exclusive to the corresponding feature x_{num} . In addition, in order to enable the model to effectively utilize all the information of the raw data, we designed a mask value for each numerical feature to denote whether this value exists. Then we concatenate the embedding vectors of each feature-mask pair, and then fuse them through a linear layer to generate a single vector, namely,

$$\mathbf{e}_{\text{fused}} = \text{Linear}[(\mathbf{e}_{\text{raw}}; \mathbf{e}_{\text{mask}})], \quad (3.2)$$

where the symbol ‘;’ in above bracket denotes the concatenation operation. The above process ensures that the model can deal with a feature with determined or undetermined value.

All tokenized feature vectors will be concatenated with a special classification token [CLS] in the sequence dimension to form the final input sequence. The sequence is then fed into a Transformer encoder consisting of L identical layers stacked together. Each layer of the encoder consists of two core submodules: a Multi-Head Self-Attention module and a feedforward network, which is graphically shown in Fig. 3(b). Residual connection

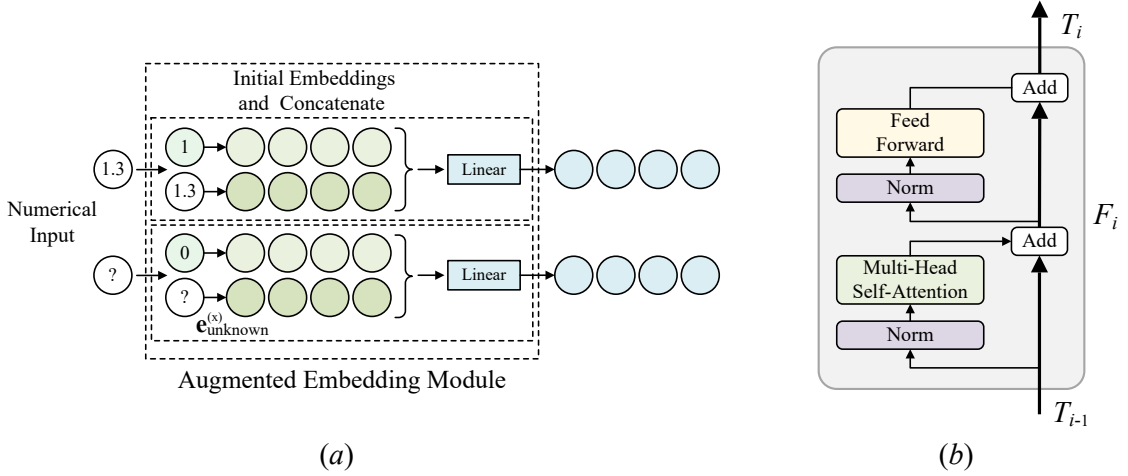


Fig. 3: (a) The Augmented Embedding module to deal with the numerical features; (b) one Transformer layer.

and layer normalization are also applied after each submodule to ensure training stability. The multi-head self-attention mechanism is the core of the model, which evaluates the mutual importance between all tokens in the sequence by calculating the scaled dot product between the query (Q), key (K), and value (V), namely,

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V, \quad (3.3)$$

where d_K is the dimension of K vector. This allows every feature token to interact with and incorporate context from all other features in the feature vector.

After being processed by the L layer encoders, the final output vector corresponding to the [CLS] token is considered as a highly aggregated representation of the entire input feature sequence. The vector is ultimately fed into a simple multi-layer neural network to predict the target physical quantity (particle width here).

3.2 Loss function to describe the mesons widths

Loss function behaves as the ultimate principle for the neural network model. The core task of the neural network designed before is to precisely predict the total widths of mesons based on the input quantum numbers. The prediction precision for the i th particle can be represented by the relative error ϵ_i as

$$\epsilon_i = \frac{\hat{y}_i - y_i}{y_i}, \quad (3.4)$$

where \hat{y}_i and y_i denote the predicted value and true value for the i th sample, respectively. To obtain the globally optimal description, we use the following mean square relative error (MSRE), namely,

$$\text{Loss} = \epsilon^2 = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right)^2, \quad (3.5)$$

where the summation is over all the training data samples; also we define ϵ to represent the overall relative error. The loss function insures that each particle has the same weight in the training.

3.3 Training and validation

The division of our dataset into a training and a test set was a deliberate, physically-motivated choice, not a random split. The training set is composed of the vast majority of conventional $q\bar{q}$ mesons and a subset of relatively well-established exotic candidates. The test set comprises the remaining particles, resulting in an approximate 90%:10%. The training set is used for updating model parameters. The test set serves a dual purpose: it is used to monitor for potential overfitting during training, such as learning rate scheduling and early stopping, and to evaluate the final model's generalization performance.

In the training implementation, we use the AdamW optimizer [44] to update the model parameters with an initial learning rate $\eta = 10^{-4}$ and weight decay 10^{-4} . To achieve stable learning rate adjustment, we combined a hybrid strategy of linear warm-up and the ReduceLROnPlateau [45]: in the first 10 epochs, η smoothly increased from 10^{-5} to the initial value; then the learning rate will decay by a factor of 0.75. The entire training process is conducted in a batch size of 64 with a maximum of ~ 500 epochs. The loss value is continuously monitored during the entire training process.

4 Numerical results of the meson widths predictions

4.1 Overall performance of the model

The research focuses on task of describing and predicting the meson widths ranging from 10^{-14} MeV to 625 MeV, which poses a serious challenge to any predictive model. In order to systematically evaluate the performance of the model on meson width predictions, we investigate the prediction accuracy from both macro and micro perspectives.

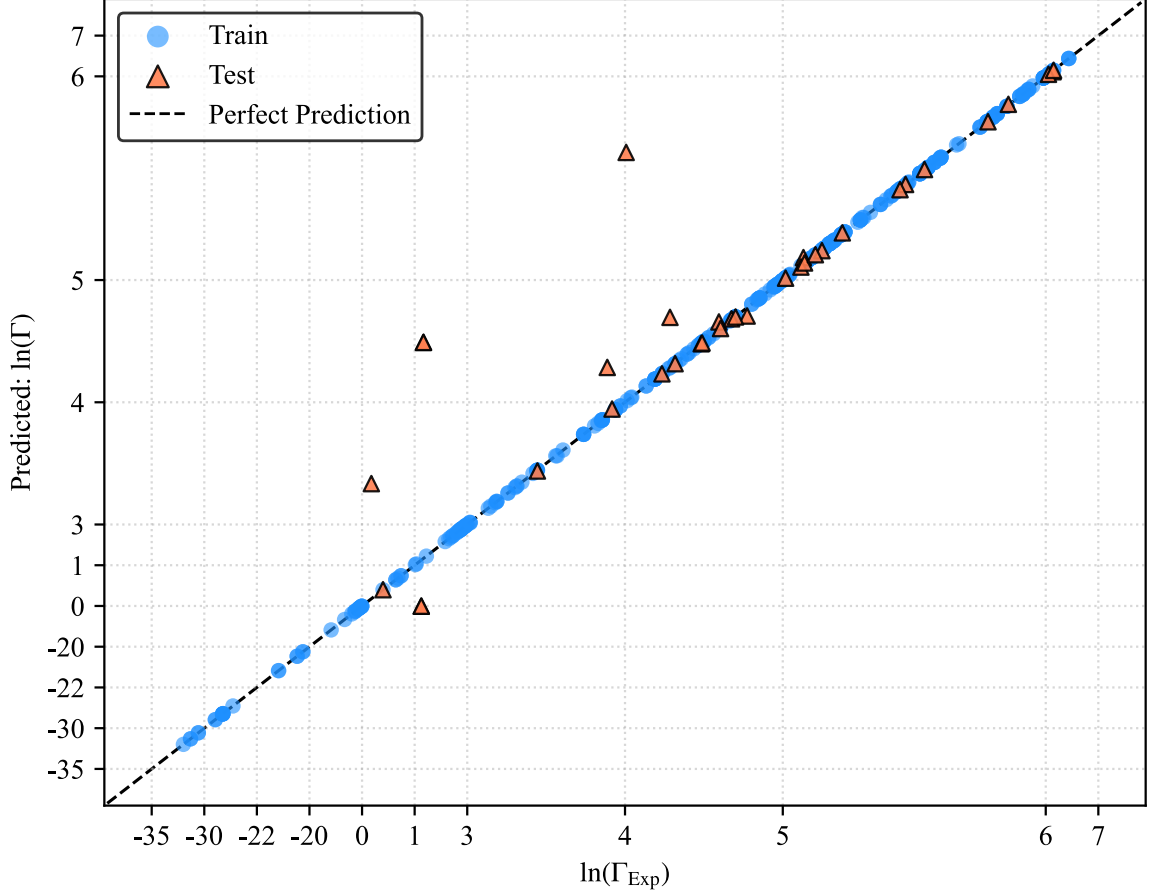


Fig. 4: The predicted meson widths versus the experimental values in units of MeV under the natural logarithm. The horizontal axis represents the experimental widths reported by PDG, while the vertical axis represents the predicted values. The blue dots denote the results of the training data, while the orange ones denote results of the test sample.

In Fig. 4 we show the overall performance of the meson width predictions for all the data including both the training and the test data. To present the results more clearly across multiple orders of magnitude, we have taken the natural logarithm of the meson widths in units of MeV. The x -axis represents the experimental widths reported by PDG [11], while the vertical y -axis represents the predicted values. The blue dots denote the results of the training data while the orange ones denote that of the test sample. It can be seen that the model's predicted values are all closely distributed around the ideal line $y = x$, indicating that the model has successfully captured the overall trend and correlation. To quantita-

tively evaluate the linear correlation, we calculated the coefficient of determination for the log-transformed values, which is $R^2 = 0.9987$. This macroscopic performance validates the effectiveness of the data processing scheme and the model architecture.

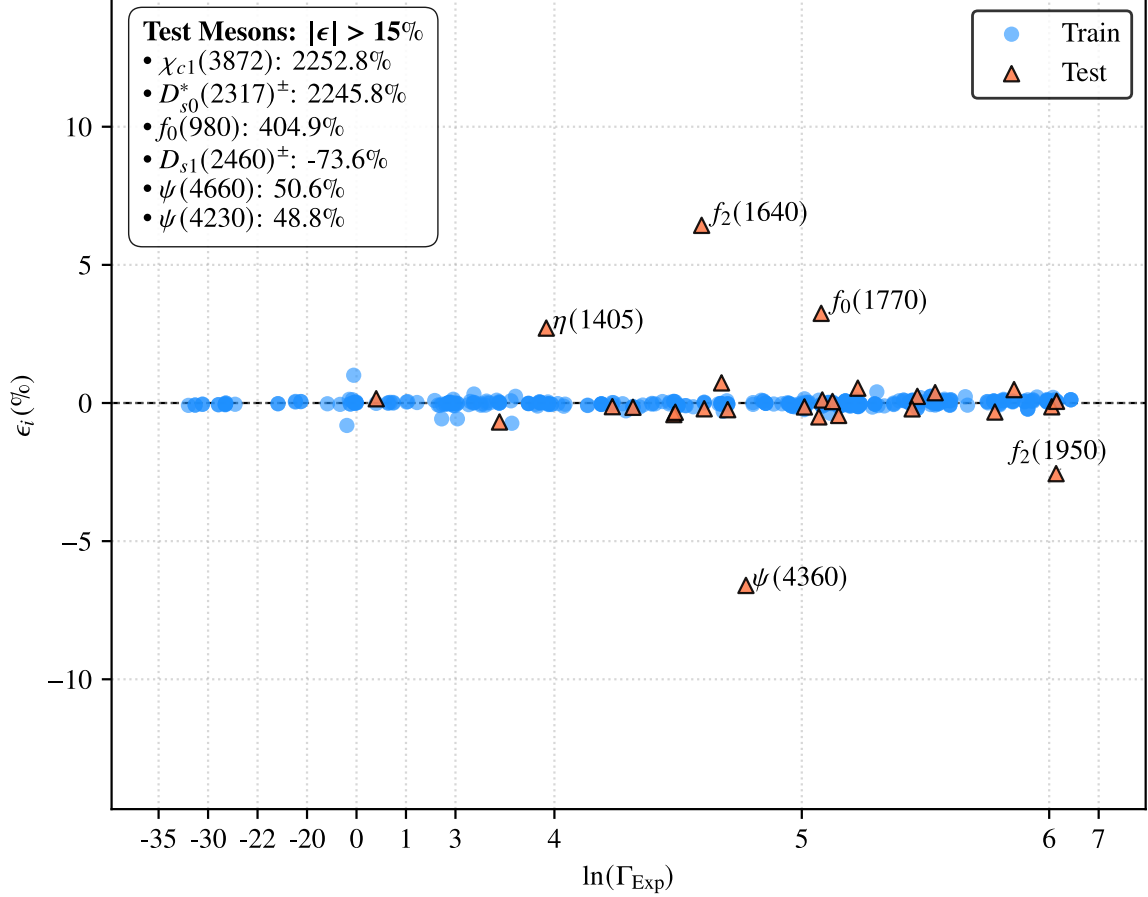


Fig. 5: The relative error ϵ_i versus the log widths in units of MeV. The blue (orange) denotes the results of the training (test) data.

To investigate the prediction accuracy on the original physical scale, we also plot the distribution of the relative error ϵ_i (defined in Eq. 3.4) versus the log width in units of MeV in Fig. 5. The distribution for the training data (blue) are almost clustered around the zero error line. After excluding several states with $\epsilon_i > 48\%$, the obtained relative errors behave $\epsilon = 0.07\%, 1.0\%$, and 0.14% for the training set, the test set, and all the data, respectively. More specifically, among the 370 training mesons, 99% of the samples had a relative error $< 1\%$, while 80% of the samples even had a relative error of less than 0.1% , which indicate that model can well describe the training data. The generalization ability of the model is more reflected by the predicting performance for the test data. As shown in Fig. 5, the error distribution of the test set (orange) exhibits high differentiation. On the one hand, most meson in the test set still show quite low relative errors. After excluding a few outliers with significant errors ($> 48\%$), the mean relative error of the remaining 27 mesons in test data is about 1% . If we further exclude the f and η families, whose flavor-

mixed states require an approximation in the quark content encoding, the remaining 22 traditional mesons have the mean relative error of less than 0.3%, which achieves the high-precision level comparable to the training set. The above results clearly reveal that the model has good generalization ability in predicting the widths of the traditional mesons.

On the other hand, Fig. 5 also show several mesons with significant relative errors, for example, $\chi_{c1}(3872)$ with $\epsilon \sim 2250\%$, $D_{s0}^*(2317)$ with $\epsilon \sim 2240\%$, and $D_{s1}^*(2460)$ with $\epsilon \sim 70\%$, which are more likely to be the exotic states instead of the traditional $q\bar{q}$ mesons and to have more complicated inner structure. Since the model are mainly trained on the traditional meson data, its predictions show significant deviations on these exotics states, which is reasonable and also reveals the boundaries of the applicability of the current model.

In summary, the analysis in this subsection indicate that the model show high accuracy in predicting the widths of the traditional mesons and well generalization performance. The same time it show systematically significant errors in dealing with the potential exotic states. However, this result in turn provides us a tentative probe to identify and reexamine the potential exotic states from their total widths.

4.2 Meson width predictions and quantum numbers identification

First we list the meson width spectra predicted by the deep neural network in Figs. 9 to 15 of the appendix A.

In Fig. 6, we provided quantitative width predictions for 15 unmeasured mesons (green) in PDG, as well as 32 new mesons (orange) that were theoretically predicted by the GI model [3] but not yet experimentally discovered. These results provide helpful references and tests for future experimental searches and theoretical studies.

Besides predicting the unknown meson widths, the model also provides us a tentative probe to explore the possible quantum numbers for undetermined mesons based on the comparison between experimental width and predicted one. We conducted a series of hypothesis tests on various particles with unknown quantum numbers using the trained model. Namely, the possible quantum numbers are fed into the model to predict the corresponding widths which are then compared with the experimental values. The comparison results provide the model's suggested quantum numbers for those states, which are listed in Tab. I.

For the $X(4160)$ state, the PDG favors the $J^{PC} = 2^{-+}$ assignment with a significance of over 4σ . Under the assumption of $2^{-+} [c\bar{c}u\bar{u}]$ tetraquark state, the predicted width 138.5 MeV is well consistent with the experimental value 136 MeV. For $X(4630)$, the PDG favors the 1^- assignment over the 2^- , the model shows that under the $[c\bar{c}s\bar{s}]$ hypothesis, both the 1^{-+} and 2^{-+} yield predictions consistent with the experimental width. For $T_{c\bar{c}}(4100)$, the PDG suggests the assignment 0^+ or 1^- . The model's prediction 164.3 MeV under the 1^{-+} hypothesis is quite close to the experimental value 150^{+80}_{-70} MeV and then favors the latter assignment. The model shows well consistency between its predictions and the preliminary conclusions drawn from experimental analyses by the PDG, which in turn provides independent, data-driven support for these findings.

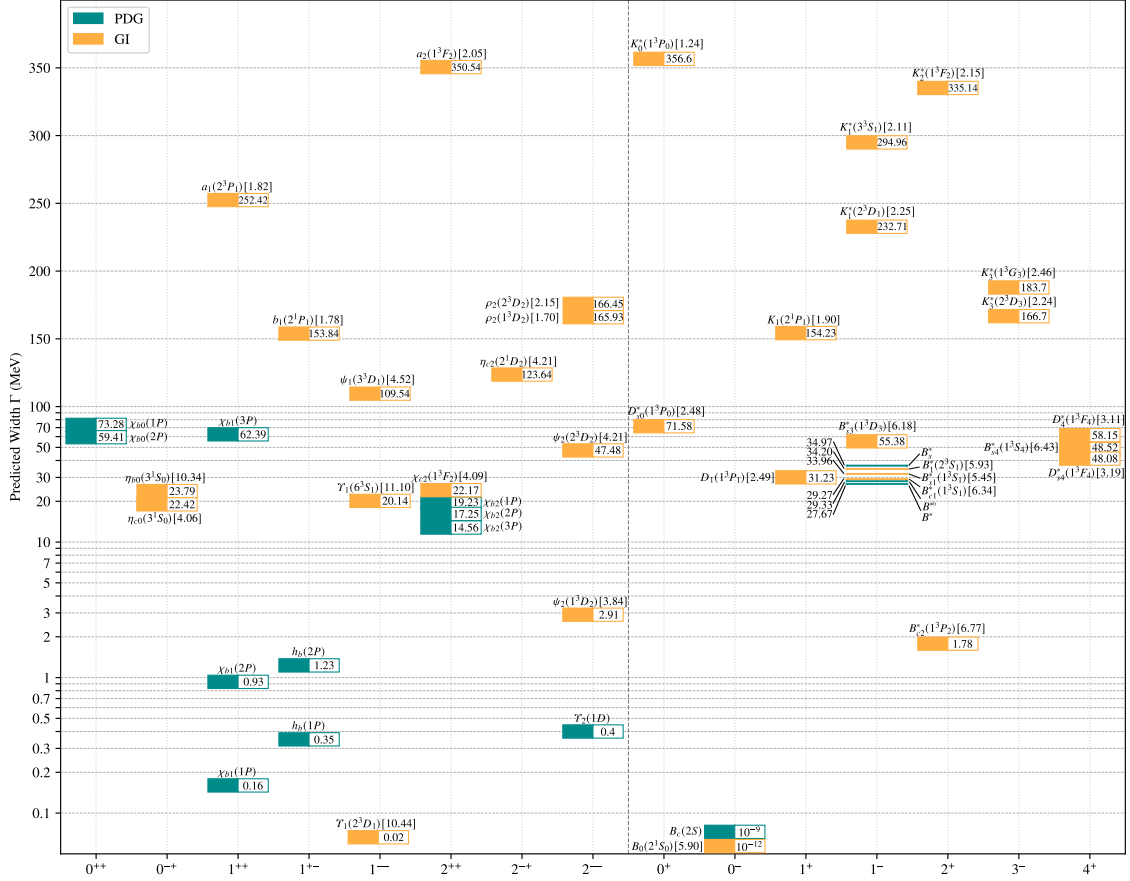


Fig. 6: The predicted widths for 47 meson states, where the green labels denote the states observed in experiments and the orange ones denote the states predicted by the GI model [3].

For particles where the PDG has not provided a clear preference, the model offers independent assignments for their quantum numbers. For example, $J^P = 0^+$ hypothesis for $D(3000)$ yields the predicted width with about a 2% relative error from the experimental value. For the tetraquark $T_{c\bar{c}}(4050)$, under the 3^{-+} assignment the model gives a predicted width that is quite close to the experimental value. Plausible quantum number assignments were also identified for several other particles, such as $X(1750)$ and $B_{sJ}^*(5850)$.

For the K , D and B states, if only the parity P is specified, the predictions show well consistency with the experimental data, which is a manifestation of the strategy for the missing features. In these cases, our analysis provides support for the particles to have specific P values, while the determination of their precise spin J values is beyond the scope of the current predictive model.

4.3 Symmetries reproduced by the neural network

A successful model should not only have high accuracy on predicting the numerical task, but it should also comply with the deep physics principles and symmetries. In this subsection

Tab. I: The model's predicted widths and favored quantum numbers for the undetermined mesons.

States	$I^{(G)}(J^{P(C)})$	Γ_{Exp}	$I^{(G)}(J^{P(C)})$	Content	Γ_{Pre1}	$I^{(G)}(J^{P(C)})$	Content	Γ_{Pre2}
$X(1750)$	$?^-(1^{--})$	120 ± 10	$0^-(1^{--})$	$s\bar{s}$	149.1	$0^-(1^{--})$	$u\bar{u}s\bar{s}$	149.1
$K(1630)$	$\frac{1}{2}(?^?)$	16_{-16}^{+19}	$\frac{1}{2}(?^-)$	$d\bar{s}$	16.4	-	-	-
$K(3100)$	$?^?(?^?)$	42 ± 16	$\frac{1}{2}(?^+)$	$u\bar{s}$	41.5	$\frac{3}{2}(2^+)$	$s d \bar{u} \bar{u}$	38.5
$D^*(2640)$	$\frac{1}{2}(?^?)$	< 15	$\frac{1}{2}(?^-)$	$c\bar{d}$	16.9	-	-	-
$D(3000)$	$\frac{1}{2}(?^?)$	190 ± 80	$\frac{1}{2}(0^+)$	$c\bar{u}$	186.1	-	-	-
$D_{sJ}(3040)$	$0(?^?)$	240 ± 60	$0(?^-)$	$c\bar{s}$	177.3	-	-	-
$X(3940)$	$?^?(?^{??})$	43_{-18}^{+28}	$0^-(3^{+-})$	$c\bar{c}$	36.1	$0^+(2^{++})$	$c\bar{c}$	35.4
$X(4160)$	$?^?(?^{??})$	136_{-35}^{+60}	$0^+(2^{-+})$	$u\bar{u}c\bar{c}$	138.5	$0^+(2^{-+})$	$d\bar{d}c\bar{c}$	160.9
$X(4630)$	$0^+(?^{?+})$	170_{-80}^{+140}	$0^+(2^{-+})$	$s\bar{s}c\bar{c}$	167.3	$0^+(1^{-+})$	$s\bar{s}c\bar{c}$	152.9
$B_J^*(5732)$	$?(?^?)$	128 ± 18	$\frac{1}{2}(?^-)$	$d\bar{b}$	126.5	-	-	-
$B_J(5840)^+$	$\frac{1}{2}(?^?)$	220 ± 80	$\frac{1}{2}(?^-)$	$u\bar{b}$	190.7	-	-	-
$B_J(5840)^0$	$\frac{1}{2}(?^?)$	130 ± 40	$\frac{1}{2}(?^-)$	$d\bar{b}$	123.9	-	-	-
$B_J(5840)^+$	$\frac{1}{2}(?^?)$	62 ± 20	$\frac{1}{2}(?^-)$	$u\bar{b}$	74.5	-	-	-
$B_J(5970)^0$	$\frac{1}{2}(?^?)$	81 ± 12	$\frac{1}{2}(?^-)$	$d\bar{b}$	90.5	-	-	-
$B_{sJ}^*(5850)$	$?(?^?)$	47 ± 22	$0(1^-)$	$s\bar{b}$	36.9	-	-	-
$B_{sJ}(6063)$	$0(?^?)$	26 ± 6	$0(2^-)$	$s\bar{b}$	22.9	-	-	-
$B_{sJ}(6114)$	$0(?^?)$	66 ± 28	$0(4^-)$	$s\bar{b}$	82.7	-	-	-
$T_{c\bar{c}}(4050)$	$1^-(?^{?+})$	82_{-28}^{+50}	$1^-(3^{-+})$	$u\bar{d}c\bar{c}$	82.8	$1^-(2^{-+})$	$u\bar{d}c\bar{c}$	105.0
$T_{c\bar{c}}(4055)$	$1^+(?^{?-})$	45 ± 13	$1^+(3^{+-})$	$u\bar{d}c\bar{c}$	49.1	$1^+(2^{+-})$	$u\bar{d}c\bar{c}$	35.9
$T_{c\bar{c}}(4100)$	$1^-(?^{?+})$	150_{-70}^{+80}	$1^-(1^{-+})$	$u\bar{d}c\bar{c}$	164.3	$1^-(0^{++})$	$u\bar{d}c\bar{c}$	230.4
$T_{c\bar{c}}(4250)$	$1^-(?^{?+})$	180_{-70}^{+320}	$1^-(1^{-+})$	$u\bar{d}c\bar{c}$	175.2	$1^-(0^{-+})$	$u\bar{d}c\bar{c}$	172.0
$T_{cc\bar{c}\bar{c}}(6900)$	$0^+(?^{?+})$	137 ± 21	$0^+(4^{-+})$	$cc\bar{c}\bar{c}$	158.9	$0^+(1^{++})$	$cc\bar{c}\bar{c}$	125.0
$T_{b\bar{s}}(5568)$	$1(?^?)$	19_{-7}^{+9}	$1(3^+)$	$u\bar{d}s\bar{b}$	23.2	$1(1^+/2^+)$	$u\bar{d}s\bar{b}$	25.7

we aim to verify whether the deep learning model of neural network can autonomously understand and reproduce the fundamental symmetries in particle physics, for example, the charge conjugation symmetry and the approximate isospin symmetry of meson widths.

4.3.1 Charge conjugation symmetry

The CPT theorem behaves as one of the deepest results of quantum field theory, which states that the combined operation of time reversal, charge conjugation, and parity in any order is an exact symmetry of any interaction. An important experimental implication of the CPT theorem claims that every particle must have precisely the same mass and

lifetime (width) as its antiparticle, which has been tested by the K^0 - \bar{K}^0 mass difference fraction to be less than 6×10^{-19} [11]. In above research, the particle and antiparticle are trained by the same neural network model and the only difference is the feature vector of the data set. In meson feature vector \mathbf{v} , a particle and its antiparticle are directly different by a charge conjugation transformation to the quark contents, for example,

$$\mathbf{q}_{B_c^+} = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1) \Rightarrow (0, 0, 0, 0, 0, 0, 0, 1, 1, 0) = \mathbf{q}_{B_c^-}. \quad (4.1)$$

If the trained model could also predict the same width for a particle and its antiparticle, we can deduce that the neural network model is invariant under the charge conjugation transformation.

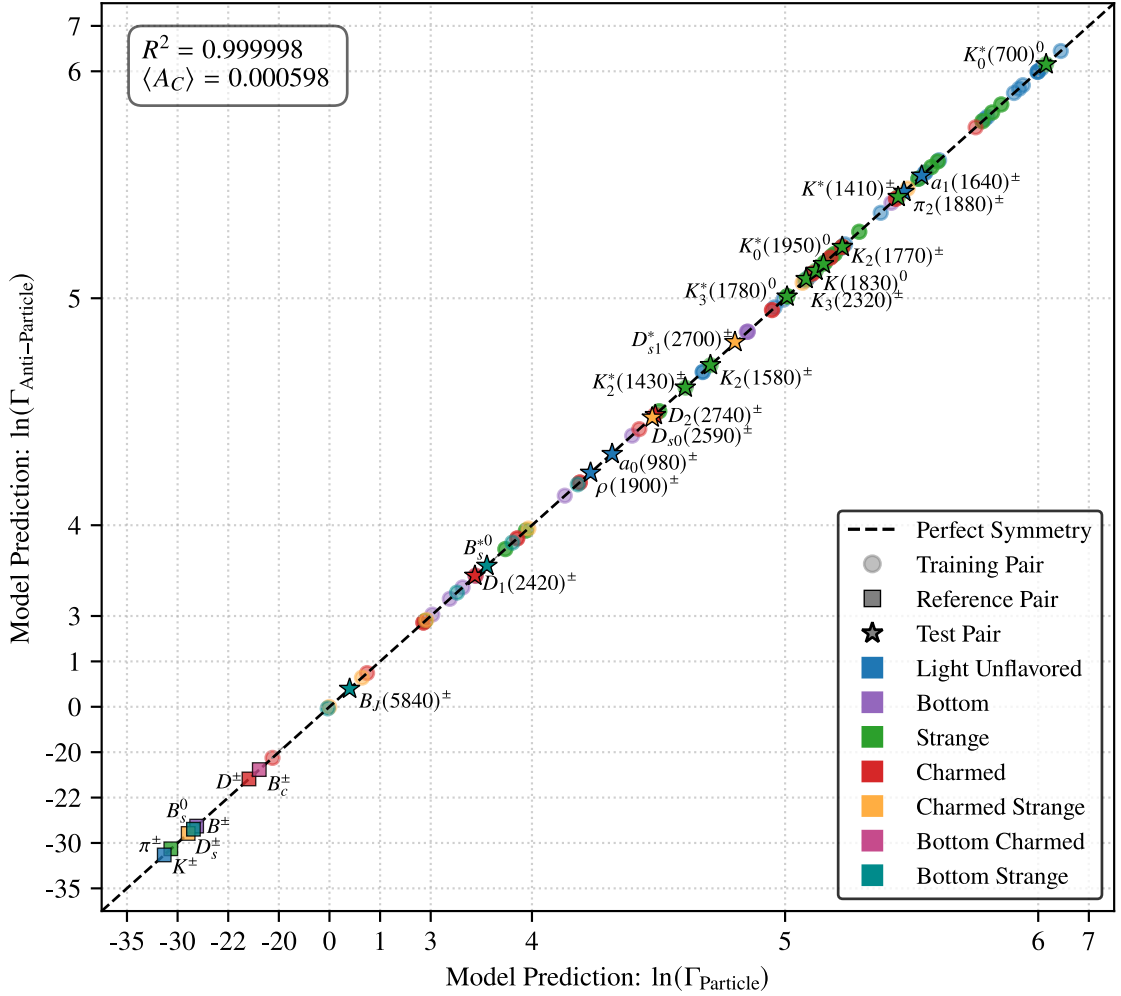


Fig. 7: The predicted widths for particle-antiparticle pairs, plotted on a logarithmic scale. The points in the test set are represented by pentagrams, while those in the training set are represented by circles. Key reference particles within the training set are highlighted with squares. Different colors are used to denote mesons with different quark contents.

In Fig. 7, we show the invariance of charge conjugation transformation clearly. The

x -axis denotes the predicted log widths for particles, while the y -axis represents the log widths for the corresponding antiparticles. Then each point in this plot represents a two-dimensional coordinate $(\ln \Gamma_{\text{Particle}}, \ln \Gamma_{\text{Antiparticle}})$. If the model was really invariant under the transformation for particle and antiparticle, all the points should locate closely along the line $y = x$. From Fig. 7, we can see that all the data points, regardless of their quark composition (color distinction) or role in the dataset (shape distinction), are strictly listed along the $y = x$ line. We also calculate the determination coefficient to be $R^2 = 0.9999$. In order to reflect the degree of this symmetry violation for the neural network model, we introduce the violation coefficient as

$$A_{Ci} = \frac{\Gamma_i - \Gamma_{\bar{i}}}{\Gamma_i + \Gamma_{\bar{i}}}, \quad (4.2)$$

where Γ_i and $\Gamma_{\bar{i}}$ denote the widths of the particle i and its antiparticle, respectively. Then we can define the root mean square violation coefficient A_C to denote the overall violation of charge conjugation symmetry of the model for all mesons,

$$A_C = \sqrt{\langle A_{Ci}^2 \rangle} = \sqrt{\frac{1}{n} \sum A_{Ci}^2} = 0.06\%. \quad (4.3)$$

The obtained result indicates that the model makes almost the identical width predictions for particles and their antiparticles. From above analysis, it can be deduced that the trained neural network model is invariant under the charge conjugation transformation.

4.3.2 Isospin symmetry

Within the dynamics of the Standard Model, the strong interaction behaves much larger than the electromagnetic interaction and the weak interaction. Meson widths are then usually determined by the corresponding strong decay behaviors except for those that can only electroweakly decay. On the other hand, the strong interaction is approximately invariant within the isospin transformation, and then the widths for mesons that can strongly decay also have approximate isospin symmetry. In this subsection we investigate whether the deep neural network model can capture this deep abstract symmetry.

In order to check this point, we plot the predicted widths versus the averaged widths for those isospin multiplets in Fig. 8. The result shows the approximate isospin symmetry of the model by comparing the predicted meson width (y -axis) with the average one of its isospin multiplets (x -axis). As shown in Fig. 8, the data points predicted by the model are also clustered along the $y = x$ line which stands for the perfect isospin symmetry of meson widths. The corresponding determination coefficient of this isospin symmetry are determined to be $R_I^2 = 0.9950$. One can see that the model correctly identifies the isospin multiplets and maintain well consistence within their internal members. The behaviors of the test samples marked with black boxes in the figure validate the well generalization performance of the model.

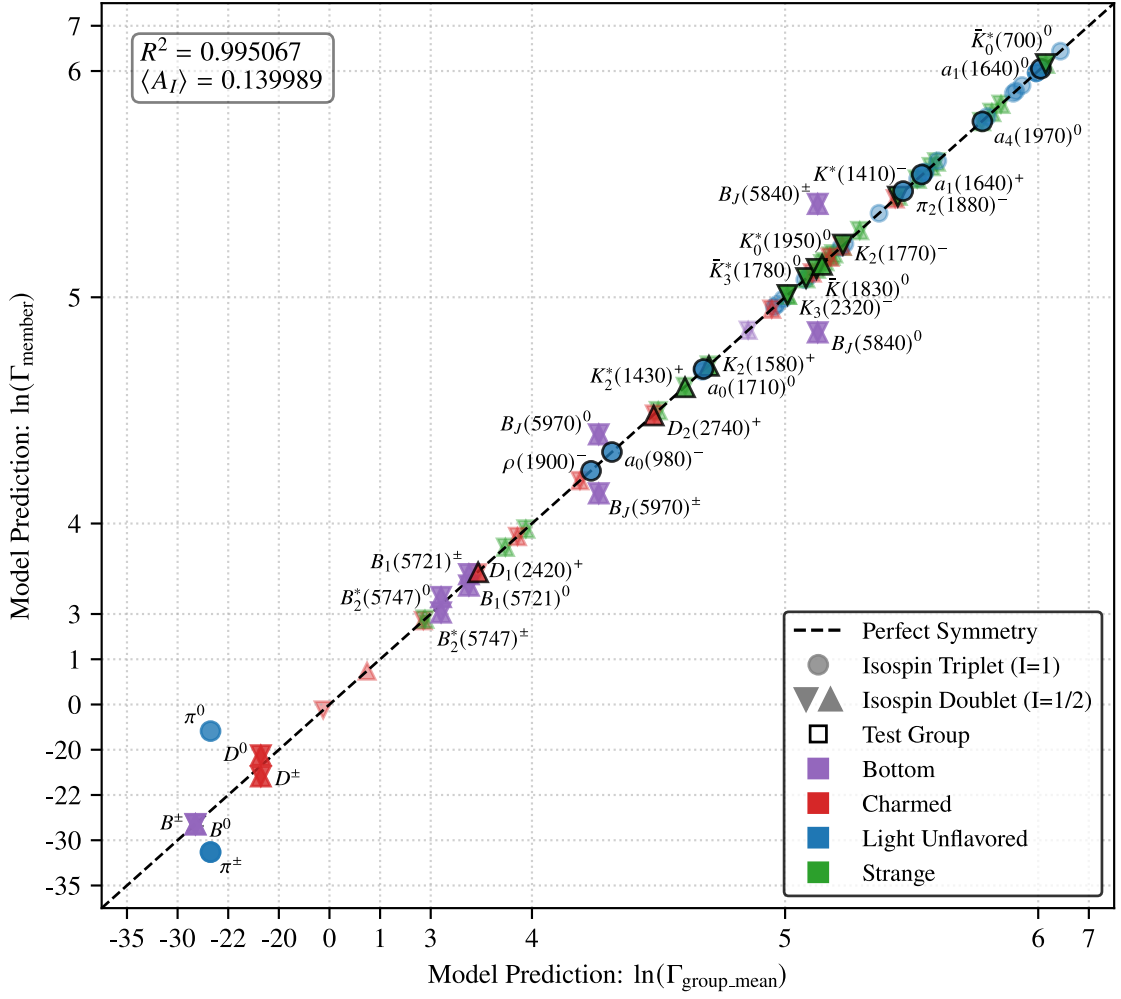


Fig. 8: The predicted width versus the averaged width of the isospin multiplets for mesons with isospin $I > 0$. The x -axis denotes the log average widths of the isospin multiplets, while y -axis denotes the predicted width of a single state. The dots represent results for the isospin triplets $I = 1$, while the triangles represent those for the isospin doublets $I = 1/2$.

The violation coefficient A_{Ii} for isospin symmetry can be introduced similarly as

$$A_{Ii} = \frac{\Gamma_i - \langle \Gamma_i \rangle_I}{\Gamma_i + \langle \Gamma_i \rangle_I}, \quad (4.4)$$

where Γ_i denotes the width for meson i , and $\langle \Gamma_i \rangle_I$ denotes the average one of the isospin multiplets. After calculating the root mean square for A_{Ii} , we obtain the violation coefficient of isospin symmetry for the model as

$$A_I = \sqrt{\langle A_{Ii}^2 \rangle} = \sqrt{\frac{1}{n} \sum A_{Ii}^2} = 14.0\%, \quad (4.5)$$

where the summation is over all the mesons of isospin $I > 0$. This value reveals that the

isospin symmetry is significantly violated. A closer inspection reveals that the deviation is not uniformly distributed across all multiplets. The pion multiplet (π^\pm, π^0) is the predominant source of this large value. If excluding the pions, the violation coefficient for the remaining multiplets drops to 5.0%. A less pronounced effect is observed for the D meson doublets ($D^\pm, D^0/\bar{D}^0$); removing them as well reduces the coefficient to a remarkable 3.0%. This final value is mainly a reflection of the more subtle deviations within the B mesons sector. Notice the isoscalar mesons are not included in the A_I calculation. It can be seen clearly from above results that the model still shows reasonable isospin symmetry but this symmetry behaves obviously and much more violated compared with the charge conjugation symmetry.

In fact, the model had also captured the violation effects for different symmetries during the training process, and hence reproduced the effects in the inference process. For the mesons in the training set, the outputs of the model reflect the patterns it has learned from the input data. For example, the model provides almost the same widths for π^+ and π^- but quite different width values for π^0 and π^+ ; for K , D , and B family mesons, the predicted widths behave slightly different within the isospin multiplets but behave quite close for the particles and their corresponding antiparticles.

The difference in the determination coefficients R^2 (0.9999 vs. 0.9950) can be taken as a measure of the strictness of model symmetries. The result is consistent with the physical reality since the charge conjugation symmetry is much more precise than the isospin symmetry. Combining the results in Fig. 7 and Fig. 8, we can conclude that the output of the model not only qualitatively follows the physical symmetry, but also quite precisely reflects the accuracy of different symmetries at a quantitative level. The results above indicate that the constructed feature system and neural network have well learning and inference abilities to distinguish the detailed physics determined by different interactions or input parameters.

5 Summary

In this work, we build a deep neural network model to predict the total widths of mesons based on the Transformer architecture, which can also be used a tentative probe to determine the meson quantum numbers. The input feature vector is constructed from the meson quantum numbers and mass. To overcome the lack of data samples in training the deep learning model, we adopt the Gaussian Monte-Carlo data enhancement method to enhance the meson data by considering the experimental errors, which also significantly improve the robustness and generalization performance of the model. Within the widths range from $\sim 10^{-14}$ to ~ 600 MeV, the relative errors of the model's predictions behave $\epsilon = 0.07\%$, 1.0% , and 0.14% in the training set, the test set, and all the meson data, respectively. We present the width predictions for the current discovered mesons and some theoretical predicted ones. The obtained width spectra can provide useful references to both the experimental and theoretical researches and can also be tested by the future experiments. Its predictive deviations further serve as a valuable a probe to study the quantum numbers and inner structures for some undetermined mesons. Furthermore, the

pure data-driven model is also investigated to have the well charge conjugation symmetry and the approximate isospin symmetry. The results indicate that the deep neural network has powerful learning and inference abilities to describe and explore the complicated interactions and physical laws in particle physics.

A Spectra of the predicted meson widths

We list the spectra of the predicted meson widths in Fig. 9 ~ Fig. 15. Fig. 9 shows the predicted widths for the isovector mesons with quark contents $[u\bar{d}, \sqrt{1/2}(u\bar{u} - d\bar{d}), d\bar{u}]$; Fig. 10, the strange mesons $(u\bar{s}, d\bar{s})$; Fig. 11, the isoscalar mesons, $(u\bar{u}, d\bar{d}, s\bar{s})$; Fig. 12, the charmonia mesons $(c\bar{c})$; Fig. 13, the charmed mesons $(c\bar{d}, c\bar{u}, c\bar{s})$; Fig. 14, the bottomed mesons $(b\bar{u}, b\bar{d}, b\bar{s}, b\bar{c})$; and finally, Fig. 15, the Υ family $(b\bar{b})$.

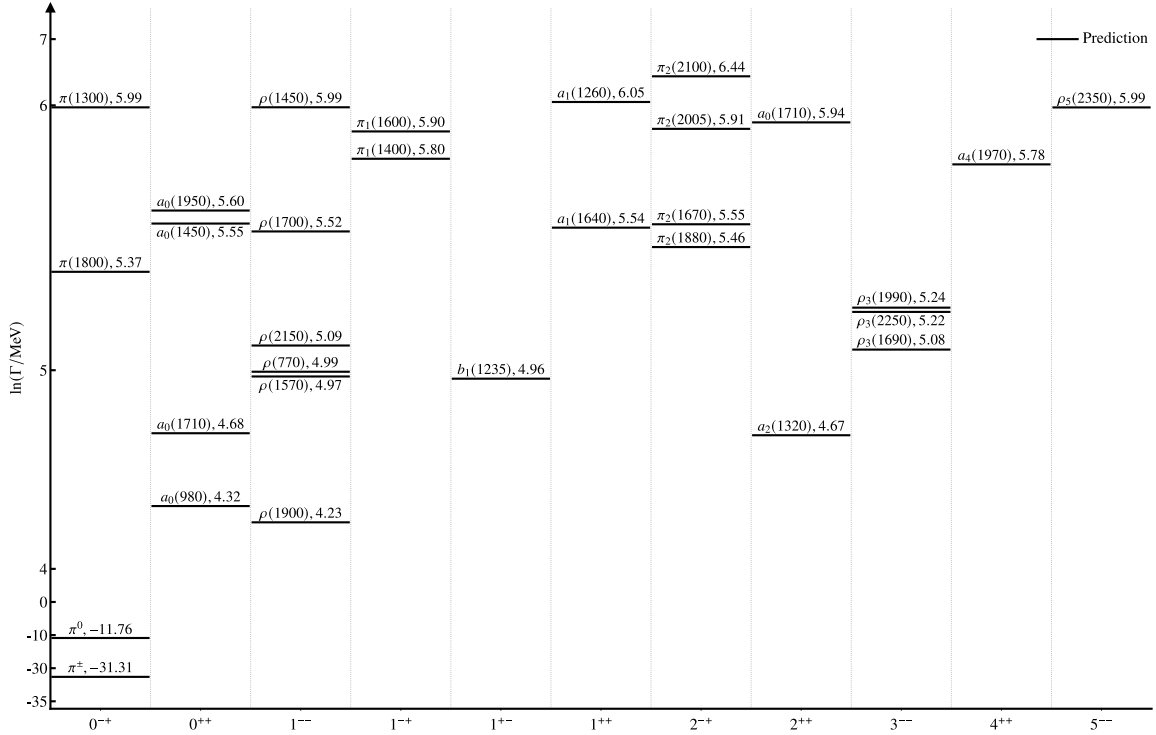


Fig. 9: The predicted log widths for mesons with quark contents $[u\bar{d}, \sqrt{1/2}(u\bar{u} - d\bar{d}), d\bar{u}]$. The y -axis denotes the log widths $\ln \Gamma$ with Γ in units of MeV.

Acknowledgments

This work is supported by the National Key R&D Program of China (2022YFA1604803), and the Natural Science Basic Research Program of Shaanxi (No. 2025JC-YBMS-020). It is also supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 12047503, 12575097, 12005169, 12075301, 11821505, and 12047503.

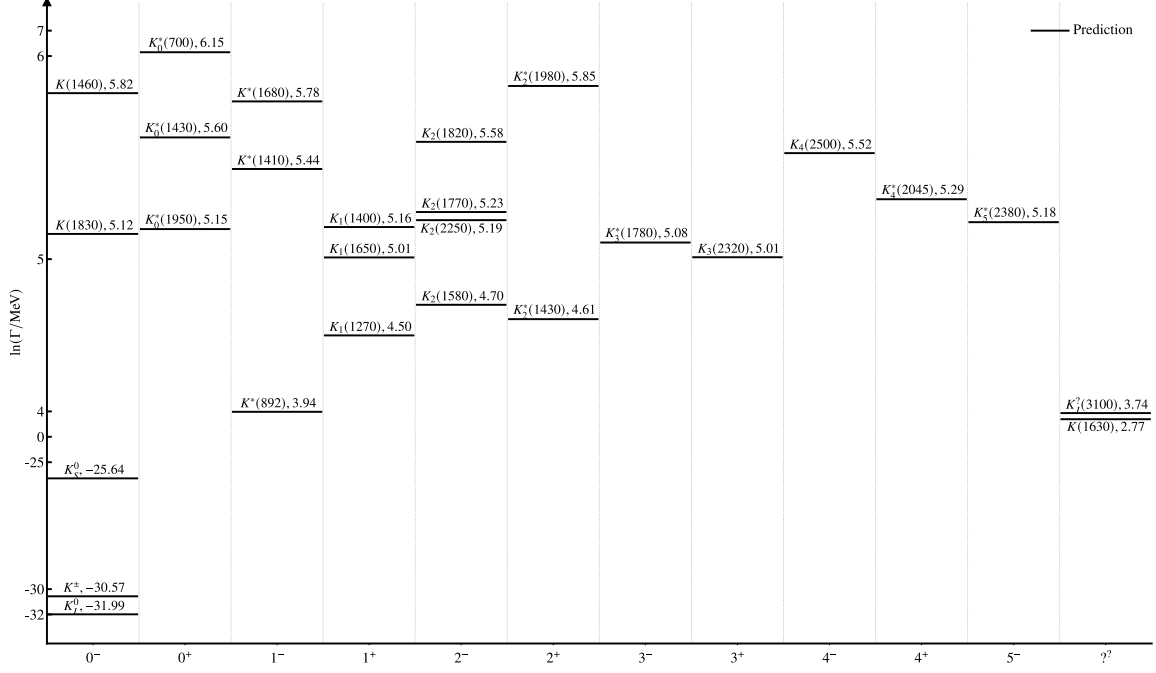


Fig. 10: The predicted log widths for strange mesons ($u\bar{s}, d\bar{s}$). The legend is as for Fig. 9.

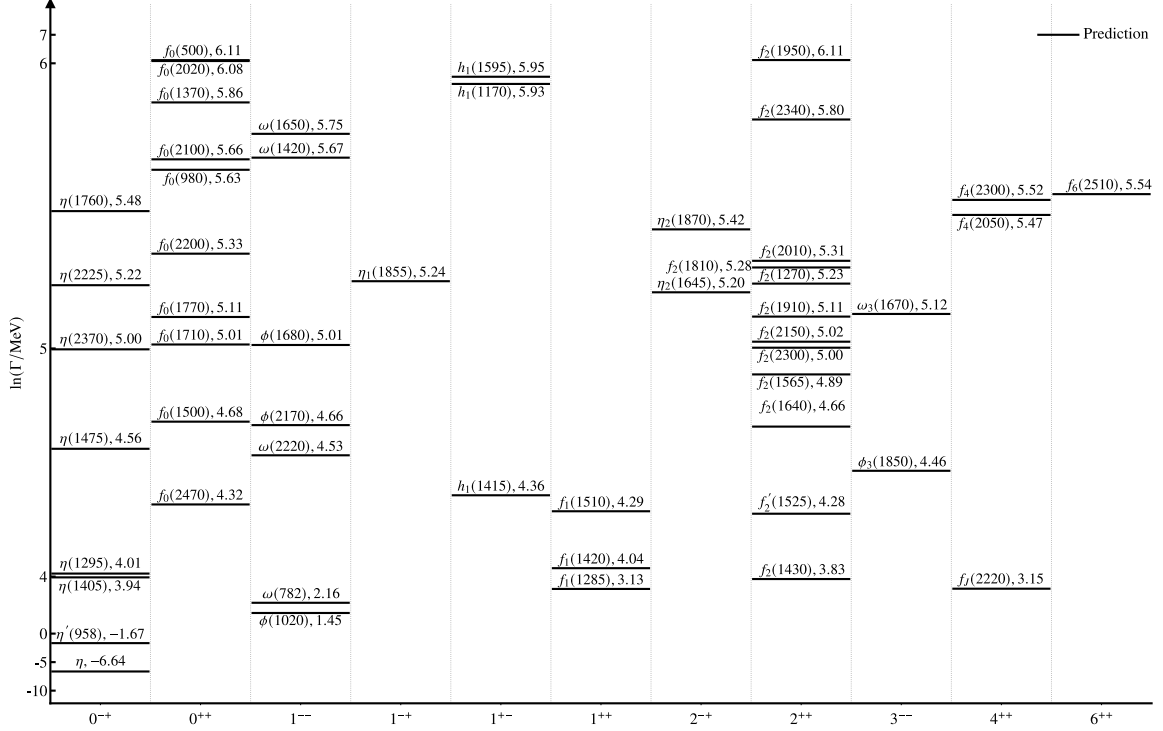


Fig. 11: The predicted log widths for isoscalar mesons ($u\bar{u}, d\bar{d}, s\bar{s}$). The legend is as for Fig. 9.

References

- [1] M.Y. Han and Y. Nambu, *Three Triplet Model with Double $SU(3)$ Symmetry*, *Phys. Rev.* **139** (1965) B1006.

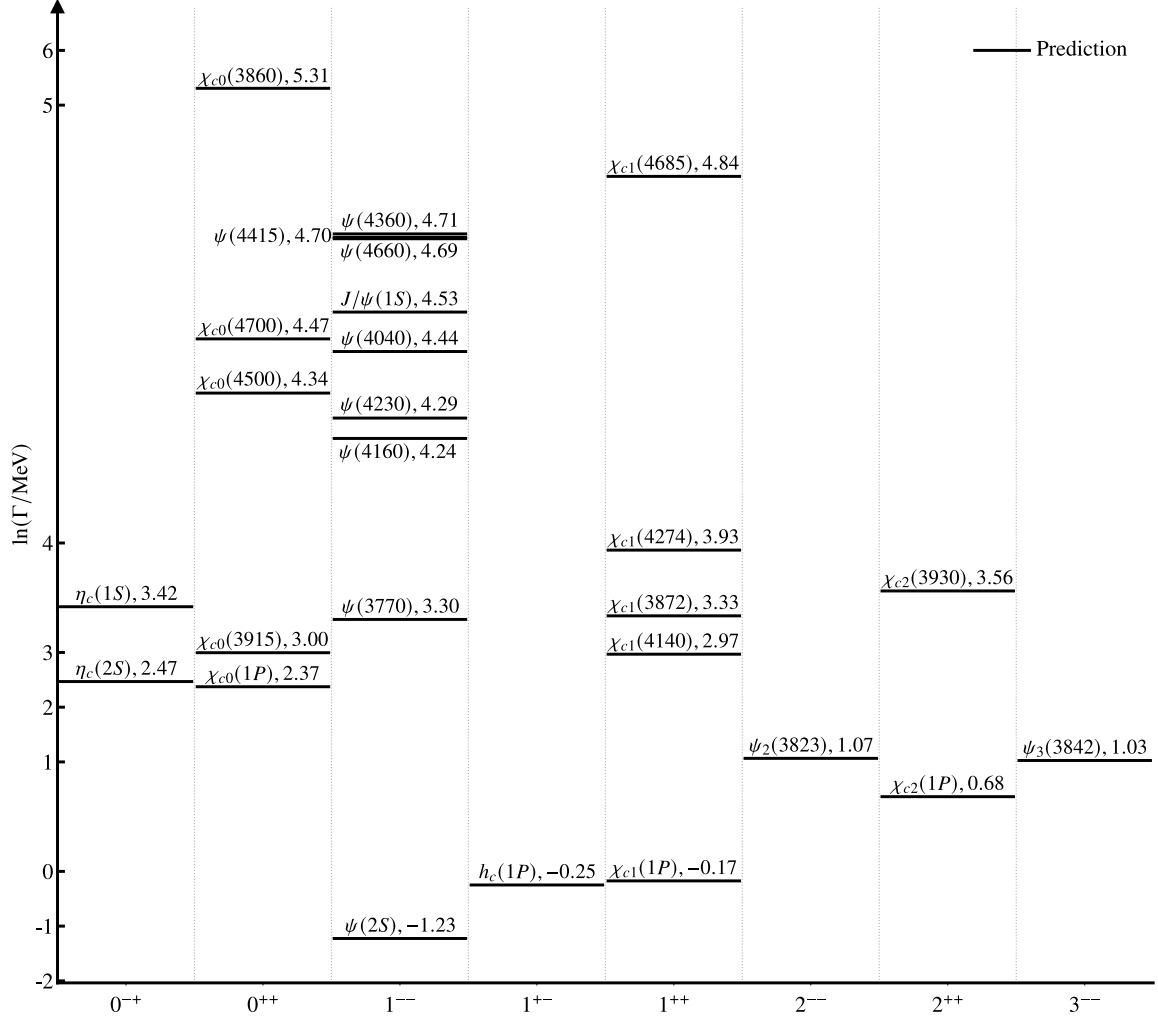


Fig. 12: The predicted log widths for charmonia mesons ($c\bar{c}$). The legend is as for Fig. 9.

- [2] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, *Phys. Lett. B* **47** (1973) 365.
- [3] S. Godfrey and N. Isgur, *Mesons in a relativized quark model with chromodynamics*, *Phys. Rev. D* **32** (1985) 189.
- [4] S. Capstick and N. Isgur, *Baryons in a relativized quark model with chromodynamics*, *Phys. Rev. D* **34** (1986) 2809.
- [5] E.J. Eichten and C. Quigg, *Mesons with beauty and charm: Spectroscopy*, *Phys. Rev. D* **49** (1994) 5845 [[hep-ph/9402210](#)].
- [6] E. Klempt and A. Zaitsev, *Glueballs, Hybrids, Multiquarks. Experimental facts versus QCD inspired concepts*, *Phys. Rept.* **454** (2007) 1 [[0708.4016](#)].
- [7] H.-X. Chen, W. Chen, X. Liu, Y.-R. Liu and S.-L. Zhu, *A review of the open charm and open bottom systems*, *Rept. Prog. Phys.* **80** (2017) 076201 [[1609.08928](#)].
- [8] S.L. Olsen, T. Skwarnicki and D. Zieminska, *Nonstandard heavy mesons and baryons: Experimental evidence*, *Rev. Mod. Phys.* **90** (2018) 015003 [[1708.04012](#)].

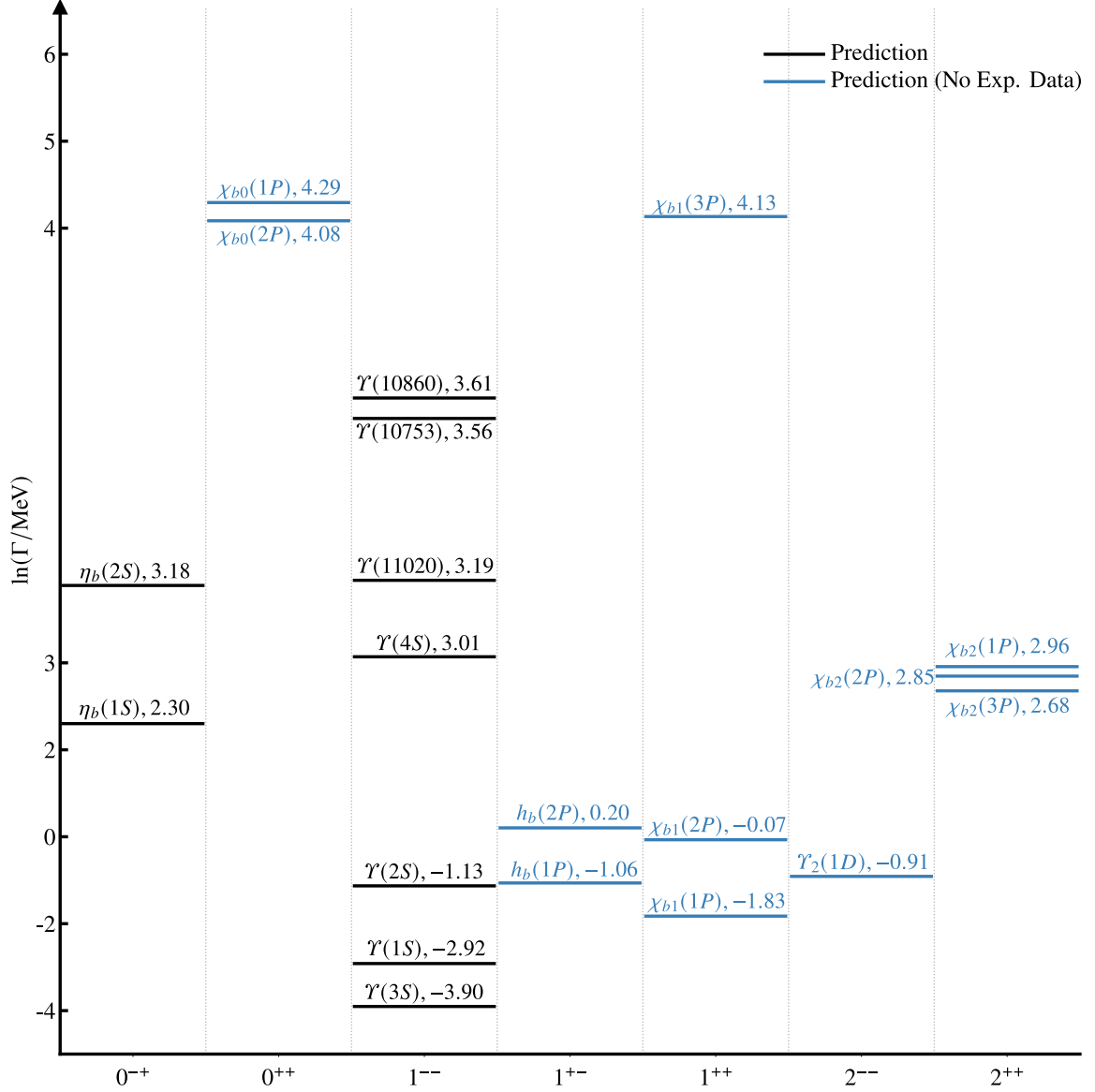


Fig. 15: The predicted log widths for $b\bar{b}$ mesons. The legend is as for Fig. 9.

states, *Rept. Prog. Phys.* **86** (2023) 026201 [2204.02649].

- [13] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436.
- [14] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning representations by back-propagating errors*, *Nature* **323** (1986) 533.
- [15] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, *Proc. Nat. Acad. Sci.* **79** (1982) 2554.
- [16] J.L. Elman, *Finding structure in time*, *Cognitive Science* **14** (1990) 179.
- [17] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural Computation* **9** (1997) 1735
[\[https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf\]](https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf).

- [18] Y. Bengio, R. Ducharme, P. Vincent and C. Jauvin, *A neural probabilistic language model*, *Journal of machine learning research* **3** (2003) 1137.
- [19] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, *Gradient-based learning applied to document recognition*, *Proceedings of the IEEE* **86** (1998) 2278.
- [20] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016, DOI.
- [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention Is All You Need*, in *31st International Conference on Neural Information Processing Systems*, 6, 2017 [[1706.03762](#)].
- [22] M. Krenn et al., *On scientific understanding with artificial intelligence*, *Nature Rev. Phys.* **4** (2022) 761 [[2204.01467](#)].
- [23] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel et al., *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (2018) 41.
- [24] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and its Application to LHC Physics*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161 [[1806.11484](#)].
- [25] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby et al., *Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91** (2019) 045002 [[1903.10563](#)].
- [26] K. Zhou, L. Wang, L.-G. Pang and S. Shi, *Exploring QCD matter in extreme conditions with Machine Learning*, *Prog. Part. Nucl. Phys.* **135** (2024) 104084 [[2303.15136](#)].
- [27] EXA.TRKX collaboration, *Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors*, in *33rd Annual Conference on Neural Information Processing Systems*, 3, 2020 [[2003.11603](#)].
- [28] L. Lonnblad, C. Peterson and T. Rognvaldsson, *Using neural networks to identify jets*, *Nucl. Phys. B* **349** (1991) 675.
- [29] E. Bols, J. Kieseler, M. Verzetti, M. Stoye and A. Stakia, *Jet Flavour Classification Using DeepJet*, *JINST* **15** (2020) P12012 [[2008.10519](#)].
- [30] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [[1810.05165](#)].
- [31] B.P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu and G. McGregor, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **543** (2005) 577.
- [32] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, *Eur. Phys. J. C* **76** (2016) 235 [[1601.07913](#)].
- [33] M.R. Mumpower, T.M. Sprouse, A.E. Lovell and A.T. Mohan, *Physically interpretable machine learning for nuclear masses*, *Phys. Rev. C* **106** (2022) L021301 [[2203.10594](#)].
- [34] W. He, Q. Li, Y. Ma, Z. Niu, J. Pei and Y. Zhang, *Machine learning in nuclear physics at low and intermediate energies*, *Sci. China Phys. Mech. Astron.* **66** (2023) 282001 [[2301.06396](#)].
- [35] M. Malekhosseini, S. Rostami, A.R. Olamaei, R. Ostovar and K. Azizi, *Meson mass and width: Deep learning approach*, *Phys. Rev. D* **110** (2024) 054011 [[2404.00448](#)].

- [36] T. Akan, *Exploring machine learning models for predicting meson mass and width*, *Phys. Scripta* **99** (2024) 125305.
- [37] Y.V. Gorishniy, I. Rubachev, V. Khrulkov and A. Babenko, *Revisiting deep learning models for tabular data*, in *Neural Information Processing Systems*, 2021, <https://api.semanticscholar.org/CorpusID:235593213> [2106.11959].
- [38] M. Malekhsosini, S. Rostami, A.R. Olamaei and K. Azizi, *Exploring fully-heavy tetraquarks through the CGAN framework: Mass and width*, *Nucl. Phys. B* **1018** (2025) 116977 [2503.00993].
- [39] ATLAS collaboration, *Measurement of the W-boson mass and width with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **84** (2024) 1309 [2403.15085].
- [40] LHCb collaboration, *Measurement of the CKM angle γ from a combination of $B^\pm \rightarrow Dh^\pm$ analyses*, *Phys. Lett. B* **726** (2013) 151 [1305.2050].
- [41] Y. Fujimoto, K. Fukushima and K. Murase, *Extensive Studies of the Neutron Star Equation of State from the Deep Learning Inference with the Observational Data Augmentation*, *JHEP* **03** (2021) 273 [2101.08156].
- [42] LIGO SCIENTIFIC, VIRGO collaboration, *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Phys. Rev. Lett.* **116** (2016) 061102 [1602.03837].
- [43] ATLAS collaboration, *Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $s=13$ TeV*, *Phys. Lett. B* **843** (2023) 137745 [2211.01216].
- [44] I. Loshchilov and F. Hutter, *Decoupled Weight Decay Regularization*, *arXiv e-prints* (2017) [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) [1711.05101].
- [45] A. Krizhevsky, I. Sutskever and G. Hinton, *Imagenet classification with deep convolutional neural networks*, in *NIPS*, 2012.