Journal of Nuclear Science and Technology

ISSN: 0022-3131 (Print) 1881-1248 (Online) Journal homepage: www.tandfonline.com/journals/tnst20

Multi-task deep neural network for predicting both nuclear fission yields and their experimental errors in peak-shaped data

M. Ueno, E. Zhang, K. Fuchimoto, S. Chiba, J. Chen & C. Ishizuka

To cite this article: M. Ueno, E. Zhang, K. Fuchimoto, S. Chiba, J. Chen & C. Ishizuka (05 Apr 2026): Multi-task deep neural network for predicting both nuclear fission yields and their experimental errors in peak-shaped data, Journal of Nuclear Science and Technology, DOI: 10.1080/00223131.2026.2652374

To link to this article: https://doi.org/10.1080/00223131.2026.2652374

JOURNAL OF NUCLEAR SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/00223131.2026.2652374

RESEARCH ARTICLE

Multi-task deep neural network for predicting both nuclear fission yields and their experimental errors in peak-shaped data

M. Ueno[a], E. Zhang[a], K. Fuchimoto[b], S. Chiba[c], J. Chen[d] and C. Ishizuka[d]

[a]Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan; [b]Research Division, The National Center for University Entrance Examinations, Tokyo, Japan; [c]NAT Research Center, Tokyo, Japan; [d]Institute of Innovative Research, Institute of Science Tokyo, Tokyo, Japan

ABSTRACT

The fission product yield (FPY) is crucially important information for numerous nuclear applications. However, the peak-shaped characteristics of FPY data present important challenges for predicting unobservable FPY data. To address these challenges, after applying multi-task learning models to fission product yield data, and their experimental error estimates, we introduce a novel loss function along with the incorporation of the odd–even effect. Our approach is intended to predict unknown fission yields and the associated experimental errors. To demonstrate the effectiveness of our proposed method, we compared our proposed method with conventional methods that learn each dataset independently. Our findings demonstrate that the proposed methods can predict peak-shaped data with experimental error estimates more effectively than earlier methods can.

ARTICLE HISTORY
Received 3 December 2025
Accepted 21 March 2026

KEYWORDS
Fission product yield; multi-task learning; machine learning in nuclear physics; JENDL-5; nuclear fission

1. Introduction

Nuclear fission data are vitally important in the field of nuclear energy [1]], including areas of applications such as design, simulation, evaluation, and safety. These data provide fundamentally important information for developing nuclear energy technologies. Fission product yield (FPY) data are particularly valuable. However, prominent nuclear data libraries (JENDL [2], CENDL [3], ENDF [4], JEFF [5], etc.) have FPY data only for thermal neutron energies between 0.5 and 14 MeV. This situation engenders the need to predict incomplete FPY data at different energies, especially for fast reactors.

Phenomenological and semimicroscopic models are known to fit existing experimentally obtained data well. Nevertheless, they have poor predictive capability for unobservable FPY data at different energy levels [6]. Although fully microscopic nuclear fission models have recently been proposed in studies, such as those of Time-Dependent Hartree–Fock–Bogoliubov (TDHFB) [7] and Time-Dependent Generator Coordinate Method (TDGCM) [8], their computational costs are daunting. Another theoretical approach to predict FPY is the macroscopic–microscopic (mac-mic) description of nuclear fission. The mac-mic model, which has been developed for more than 60 years, has come to predict FPY using recent models [9–12] to achieve good agreement with experimentally obtained data. Even for those frontier models in nuclear physics, predicting FPY with sufficient accuracy necessary for nuclear data remains very challenging.

Recently, machine learning has come to provide a powerful approach to leverage complex big data for knowledge acquisition and for predictive insights. In the field of nuclear physics, machine learning has become an important tool, with applications in elucidating nuclear structures [13–18] and nuclear reactions [19–22], providing insightful perspectives and approaches to nuclear physics research and demonstrating amazing promise [23]. Among the array of machine learning methods, Bayesian Neural Networks (BNNs), distinguished by their excellent predictive capabilities and their ability to provide uncertainty quantification, are applied frequently for diverse predictive tasks, at which they frequently demonstrate their effectiveness. These applications include predicting the charge yields of fission fragments [24] and fission yields [25,26] and providing the uncertainty quantification simultaneously.

Nevertheless, despite the effective predictive capabilities of BNNs, the peak-shaped distribution characteristics of FPY data present important challenges for predicting unobservable FPY data because the FPY data within the peak structure predominantly exhibit a jagged, non-smooth distribution. BNNs provide uncertainty quantification by Gaussian-model-based confidence intervals (CIs), which mainly reflect the network estimation uncertainties. Because a BNN estimates a Gaussian distribution to approximate the data

CONTACT M. Ueno ueno@ai.lab.uec.ac.jp The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan

distribution, predicting complex structures such as peak structures is difficult.

As this study focuses primarily on improving the predictive performance for the peak-shaped FPY distribution, we do not pursue the estimation of CIs in this work. Instead, this study addresses the FPY error, which represents another type of error distinct from the network estimation uncertainty. The Japanese Evaluated Nuclear Data Library ver. 5 (JENDL-5) [2] provides not only the FPY value but also the associated FPY error [27], which is estimated via error propagation to obtain the experimental error. In the JENDL evaluation, the uncertainties of fission product yields are estimated using a generalized least-squares method. Specifically, the covariance matrix of the evaluated yields is estimated so as to reproduce the standard deviations of the mass-yield distribution given in the well-established England–Rider systematics. The FPY error values are obtained as the square roots of the diagonal elements of this covariance matrix.

These facts indicate that FPY and its FPY errors are strongly correlated, with information from one exerting a significant influence on the other. Consequently, in general, nuclides with higher FPY values tend to have lower relative FPY errors due to this strong correlation. This correlation prominently reflects the characteristics of the relevant fission product nuclides, much in the same way that the correlation between nuclear mass and deformation represents underlying nuclear properties. Multi-task models are known to enable supplementary prediction of multiple different mutually related types of data while assisting their learning from each other. Recently, the multi-task model has been used to learn multiple physics observations simultaneously, such as nuclear masses and deformations [28,29], as well as nuclear number densities and corresponding binding energies for various nuclear shapes [30]. Similarly, predicting both the value and the associated FPY errors simultaneously using a multi-task model might be more effective than using conventional methods to predict each dataset independently.

To improve the multi-task model prediction accuracy, a multi-gate mixture-of-experts (MMoE) [31] has been proposed. After MMoE structurally learns shared representations from input data through multiple feedforward neural networks, it allocates task-specific sub-networks to learn task-dependent features. In fact, because of its efficient information-sharing capabilities, MMoE has demonstrated superior results for multi-task prediction across various fields such as aerospace [32], industrial fault detection [33], and traffic data analysis [34].

Inspired by this important benefit, this study proposes using the MMoE to learn and predict FPY and FPY errors simultaneously. The proposed method is expected to increase the prediction accuracies of both FPY and FPY error data because MMoE is known to be effective when the multiple tasks affect one another. A key advantage of MMoE in contrast to BNNs, is its superior capability in predicting complex structures like peak data.

Furthermore, to predict peak data accurately, we introduce a novel loss function that assigns weights to the training data's loss function according to the FPY values. Additionally, we incorporate the odd–even effect [6] as auxiliary input information.

The MMoE framework is introduced because FPY and FPY errors are strongly correlated targets, and joint learning can exploit this correlation to improve both predictions. In contrast, the weighted loss function is introduced to address the underfitting of the peak region caused by standard MSE-based training, thereby improving the reproduction of the peak-shaped FPY distribution.

The results demonstrate that the proposed methods can predict the peak-shaped FPY distribution with experimental error estimates more effectively.

2. Theoretical framework

2.1. FPY and FPY error data

This study utilizes the fission product yields (FPYs) and their associated errors from JENDL-5. Here, the term 'FPY error' refers to the evaluated uncertainty assigned to each independent fission product yield in the JENDL evaluation [27], which is derived through a generalized least-squares (GLS) procedure under five physical constraints, which are conservation of mass number, conservation of charge number, normalization of the independent yields, normalization of the heavier-mass yields, and consistency with the calculated mass-chain yields. Specifically, the updated yields θ_{upd} and their covariance matrix $\mathbf{V}_{\mathrm{upd}}$ are obtained by the GLS method as

$$\theta_{\mathrm{upd}} = \theta_{\mathrm{a}} + \mathbf{V}_{\mathrm{a}}\mathbf{S}^{\mathrm{T}}\left(\mathbf{S}\mathbf{V}_{\mathrm{a}}\mathbf{S}^{\mathrm{T}} + \mathbf{V}\right)^{-1}(\boldsymbol{\eta} - \mathbf{S}\theta_{\mathrm{a}}), \qquad (1)$$

$$\mathbf{V}_{\mathrm{upd}} = \mathbf{V}_{\mathrm{a}} - \mathbf{V}_{\mathrm{a}}\mathbf{S}^{\mathrm{T}}\left(\mathbf{S}\mathbf{V}_{\mathrm{a}}\mathbf{S}^{\mathrm{T}} + \mathbf{V}\right)^{-1}\mathbf{S}\mathbf{V}_{\mathrm{a}}, \qquad (2)$$

where θ_{a} denotes the initial FPY vector, $\mathbf{S}$ denotes the sensitivity matrix associated with the imposed physical constraints, $\boldsymbol{\eta}$ denotes the vector of constraint values, and $\mathbf{V}$ denotes the uncertainty associated with the constraint quantities. The initial covariance matrix $\mathbf{V}_{\mathrm{a}}$ was taken from experimental uncertainties of the independent yields when available. When experimental uncertainties were not available, errors estimated from JENDL/FPY-2011 or the JEFF-3.3/FY library, incorporating improvements from the UKFY3.7 evaluation [35] were adopted as pseudo-experimental data.

Through this GLS adjustment, the evaluated FPY values and their covariance matrix become mutually consistent with conservation laws and experimental information. The diagonal elements of $\mathbf{V}_{\mathrm{upd}}$ correspond to the variances of FPY, and their square roots are defined as the 'FPY errors' in this study.

It should be emphasized that the FPY errors represent evaluated experimental uncertainties propagated and adjusted under physical constraints, rather than statistical fluctuations from a Poisson counting process. Although experimental fission yield measurements originate from counting statistics that are often approximated by Poisson distributions, the final FPY errors in JENDL are obtained after global adjustment and covariance propagation using the GLS method. Therefore, they reflect correlated uncertainties among nuclides, rather than independent counting errors.

This study proposes a multi-task neural network to simultaneously predict both the evaluated FPY values and these associated evaluated uncertainties.

2.2. MMoE for predicting both FPY and FPY error data

To learn and predict FPY and FPY error data, this study employs the MMoE model [31] as the core feature of the multi-task learning framework. The MMoE is built on the fundamental concept of assembling a set of feedforward neural networks for the collective learning of complex patterns and representations from data. At its core, MMoE has a gating mechanism that assigns weights to each feedforward neural network for a given input instance. These weights are learned to combine the multiple feedforward neural networks and thereby to increase prediction accuracy.

The MMoE is given as

$$\hat{y}^k = h^k(f^k(\boldsymbol{x})), \tag{3}$$

where $\hat{y}^k$ represents the output of the k-th task, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ represents the input vector, and n stands for the number of the input data, whereas h^k and $f^k(\boldsymbol{x})$ denote the feedforward neural networks for the k-th task.

$$f^k(\boldsymbol{x}) = \sum_{j=1}^{m} g_j^k(\boldsymbol{x}) f_j(\boldsymbol{x}), \quad \text{and} \tag{4}$$

$$g_j^k(\boldsymbol{x}) = \frac{\exp\left(\mathbf{W}_{g_j^k}\boldsymbol{x}\right)}{\sum_{j'=1}^{m} \exp\left(\mathbf{W}_{g_{j'}^k}\boldsymbol{x}\right)} \tag{5}$$

Each task corresponds to a task gate $g_j^k(\boldsymbol{x})$, where each task gate is a softmax layer. In addition, m, which denotes the number of feedforward neural networks, is set as 3. Here, $g_j^k(\boldsymbol{x})$ represents the corresponding weight of the j-th feedforward neural network given by the gate network. In addition, $f_j(\boldsymbol{x})$ represents j-th feedforward neural network. The purpose of the gate network for each task is to select and weight all the feedforward neural networks. Within this framework, $\boldsymbol{W}_{g_j^k}$ is the trainable matrix. Different tasks can match their respective loss functions and weights. Additional details about MMoE are presented in the relevant literature [31].

Figure 1 depicts the network architecture of the proposed MMoE. The proposed neural network has two outputs: FPY value as $\hat{y}^1$ and its FPY error as $\hat{y}^2$. This architecture enables the simultaneous prediction of the FPY values and their FPY errors using supervised learning from observed data from JENDL-5. For this study, we specifically examine the mass distribution of the fission products, FPY(A), because the available data are abundant compared to the independent yields, FPY(Z, A, m), where Z stands for the charge number, A is the mass number m denotes the isomeric state. The architecture shown in Figure 1 enables learning of FPY data and learning their FPY errors supplementarily while assisting each other.

The overall loss function for training the proposed model comprises two terms: the loss function $Loss_{FPY}$ for the FPY prediction task and the loss function $Loss_{ERROR}$ for the FPY error prediction task.

$$L_{\mathrm{total}} = \alpha * Loss_{FPY} + (1 - \alpha) * Loss_{ERROR}. \tag{6}$$

Parameter α is a tuning parameter that adjusts the weight balance between the FPY prediction and the FPY error prediction. We determined $\alpha = 0.9$ as its optimal value through grid search over the set

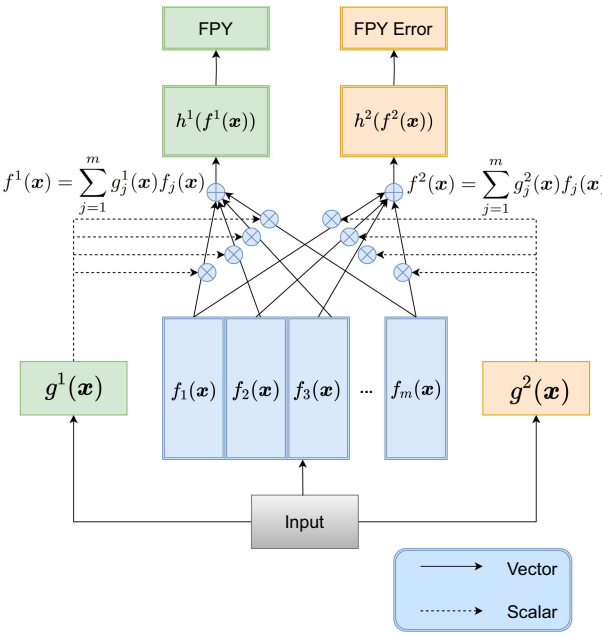

Figure 1. Network architecture of the proposed method.

$\{0.5, 0.6, 0.7, 0.8, 0.9\}$ to maximize only the FPY prediction because FPY prediction is more important than the FPY error prediction.

The first problem is which loss functions this study employs. The most popular loss function is the following mean squared error (MSE),

$$\mathrm{MSE}^k = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^k - \hat{y}_i^k\right)^2, \tag{7}$$

where $k = 1$ signifies the FPY prediction task, and where $k = 2$ signifies the FPY error prediction task. Also, y_i^k represents the true value of the i-th data point for task k; $\hat{y}_i^k$ stands for the predicted value of the i-th data point for k th task. That is to say, the MSE^k (5) signifies the MSE of k-th task for the dataset (the sample size n) within each batch. Although the important challenge of this study is the prediction of the peak data of FPY data, as described previously, the MSE function is unable to catch the peak data characteristics sufficiently.

Next, we propose an innovative loss function for FPY data to predict the peak data effectively, instead of MSE^1. The proposed loss function, which we designate as a 'weighted loss function,' is designed to allow the model to examine peak data specifically, ultimately improving the predictive capacity. It is noteworthy that the proposed loss function is only used for the prediction of FPY, not for the prediction of FPY error. The proposed 'weighted loss function' is defined as presented below.

$$weightedLoss = \frac{1}{n}\sum_{i=1}^{n} L_i^{\mathrm{W}}, \tag{8}$$

where

$$L_i^{\mathrm{W}} = \begin{cases} \beta * (y_i - \hat{y}_i^{\mathrm{W}})^2, & normalize(y_i) < r \\ w(y_i) * (y_i - \hat{y}_i^{\mathrm{W}})^2, & normalize(y_i) \geq r \end{cases}, \tag{9}$$

$$\text{and} \quad w(y_i) = 1 + normalize(y_i). \tag{10}$$

Equation (6) represents the total loss value of the FPY prediction for the dataset (of sample size n) within each batch. Also, L_i^{W} denotes the prediction error for the i-th data point within a batch, as obtained using the proposed method, where y_i represents the true value of the i-th data point. Also, $\hat{y}_i^{\mathrm{W}}$ represents the proposed method's prediction result for the i-th data point. Here, superscript W denotes that the value is obtained from the weighted loss function. To prevent an excessively large grid search over (α, β, r) in the proposed multi-task model, the tuning parameters (β, r) of the weighted loss were determined in advance using a single-task deep neural network (DNN) with $\alpha = 1$, where only the FPY prediction term is optimized. Specifically, (β, r) is selected from 25 possible combinations, defined as the Cartesian product $(\beta, r) \in \{0.001, 0.005, 0.01, 0.02, 0.05\} \times \{0.5, 0.6, 0.7, 0.8, 0.9\}$, where β and r are tuning parameters. Therefore, the selected (β, r) are then fixed when training the proposed multi-task model. Also, $w(y_i^{\mathrm{W}})$ stands for the weight coefficients; $normalize(y_i)$ denotes the standardized score of y_i with mean 0.0 and standard deviation 1.0. In Equation 10, $w(y_i)$ functions to assign higher weights to larger normalized values of y_i to force the model to fit a larger value data of FPY better.

When $normalize(y_i)$ is larger than tuning parameter r, the input vector $\mathbf{x_i}$ is called 'peak data.' When $normalize(y_i)$ is smaller than the tuning parameter r, the input vector $\mathbf{x_i}$ is designated as 'non-peak data.' By assigning an extremely small value to β, the weight of non-peak data is diminished. It increases the model's learning sensitivity to peak data. Additionally, optimizing the value of $w(y_i)$ adjusts the peak data weight according to the training data FPY value. Consequently, as the FPY value of the training data increases, the corresponding weight coefficients also increase. This strategic weighting forces the model learning to prioritize the prediction performance of the peak data.

3. Results and discussion

3.1. Comparison of prediction errors

To evaluate the peak data prediction accuracies of the multi-task DNN methods, the proposed multi-task DNN methods were compared with the previous BNN method and a standard single-task DNN.

3.1.1. Datasets

To evaluate the effectiveness of the proposed methods, we use four distinct sets of training and testing data from JENDL-5 and ensure that the data in the test set are not included in the training set, as described below

Data Set 1: Training Set: ^{236,237,238}U, ^{242,244,245,246,248}Cm, ^{254,255,256}Fm, ^{239,241,242}Pu, ^{249,250,252}Cf, ^{241,242,243}Am, ^{237,238}Np, ^{231}Pa, ^{253,254}Es, ^{227,229,232}Th, and ^{254,255,256}Fm. Testing Set: ^{235}U.

Data Set 2: Training Set: ^{235,236,237}U, ^{242,244,245,246,248}Cm, ^{254,255,256}Fm, ^{239,241,242}Pu, ^{249,250,252}Cf, ^{241,242,243}Am, ^{237,238}Np, ^{231}Pa, ^{253,254}Es, ^{227,229,232}Th, and ^{254,255,256}Fm. Testing Set: ^{238}U.

Data Set 3: Training Set: ^{233,234,235,236,237,238}U, ^{242,244,245,246,248}Cm, ^{254,255,256}Fm, ^{238,240,241,242}Pu, ^{249,250,252}Cf, ^{241,242,243}Am, ^{237,238}Np, ^{231}Pa, ^{253,254}Es, ^{227,229,232}Th, and ^{254,255,256}Fm. Testing Set: ^{239}Pu.

Data Set 4: Training Set: ^{233,234,235,236,237,238}U, ^{242,244,245,246,248}Cm, ^{254,255,256}Fm, ^{238,239,240,242}Pu, ^{249,250,252}Cf, ^{241,242,243}Am, ^{237,238}Np, ^{231}Pa, ^{253,254}Es, ^{227,229,232}Th, and ^{254,255,256}Fm. Testing Set: ^{241}Pu.

The testing sets ^{235}U, ^{238}U, ^{239}Pu, and ^{241}Pu were selected because they are major actinides in the JENDL library for which abundant and reliable experimental FPY data are available. These nuclides are also of primary importance in practical reactor analyses, including light-water and fast reactors, and therefore serve as representative benchmarks for evaluating predictive performance. In contrast, other nuclides in JENDL often have more limited experimental information, and the predictive accuracy for such nuclides may be affected by insufficient training data.

To examine the influence of data augmentation (DA) on the prediction performance, two control settings were prepared: one using DA and the other without DA. In the case without DA, the training dataset contained 6,039 samples in total. These samples were composed of 80% of the JENDL-5 FPY entries (5,382 data points) together with additional experimental datasets consisting of 286 [36], 185 [37], and 189 [38] data points taken from previous studies. The remaining 20% of the JENDL-5 FPY data (1,071 samples) were reserved for validation. When the DA procedure was applied, the number of training samples increased to 16,713, whereas the validation dataset was kept identical to that used in the non-DA case.

3.1.2. Experimental settings

For the experiments, the network input vector $\boldsymbol{x_i}$ comprises four values $x_i = (Z_i, N_i, A_i, E_i)$, where Z_i signifies the charge number, N_i represents the neutron number of the fission nucleus, A_i represents the mass number of the fission fragment, and E_i denotes the excitation energy of the compound nucleus.

For BNNs, FPY prediction requires marginalization over the posterior probability distribution of the network parameters. This marginalization involves an intractable integral that cannot be evaluated analytically. Therefore, we employ the Markov Chain Monte Carlo (MCMC) method to approximate the posterior probability distribution, according to reports of earlier study [26]. By sampling from the posterior distribution, the MCMC-based BNN enables not only point prediction of FPY values but also uncertainty quantification, such as the estimation of confidence intervals for the predicted FPYs.

For learning BNNs, we employ the MCMC method with two hidden layers, each containing 16 neurons, according to reports of earlier study [26]. In this study, an MCMC-based BNN is implemented in NumPyro library, which is a probabilistic programming framework. Posterior inference was performed using the No-U-Turn Sampler (NUTS) [39], an adaptive variant of Hamiltonian Monte Carlo (HMC) [40], as implemented in NumPyro. After discarding the burn-in samples, posterior samples of the network parameters are used to compute the predictive mean and confidence intervals of FPY.

This configuration was justified because Wang et al. [26] demonstrated that increasing the number of hidden layers up to seven did not improve the prediction performance and demonstrated that the best results were obtained with a shallow architecture consisting of two hidden layers, each with neurons. Additionally, BNN is unable to incorporate the weighted loss function (7) because it employs not the minimum loss function estimates but the MCMC estimates.

For learning DNNs, we construct 16 neurons per layer, as with BNN. The number of hidden layer 10 was ascertained by grid search from $\{5, 10, 15\}$.

For learning multi-task DNNs, we employ 16 neurons per layer and 10 hidden layers, as with DNNs for control under the same experimental conditions. Consequently, the multi-task DNN includes both $f_j(\boldsymbol{x})$ and $h^k(\boldsymbol{x})$, each comprising five hidden layers with 16 neurons per layer. In this study, the number of experts was selected by grid search over $(m = 1, 2, 3, 4, 5)$, using the validation performance of FPY prediction as the selection criterion. Among these candidates, $(m = 3)$ provided the best overall validation performance while avoiding unnecessary model complexity.

To improve the accuracy of FPY prediction further, we incorporate the odd–even effect into the input. In the mass distribution of fission products, the odd–even effect is a prominent phenomenon that originates from the nucleon pairing effect and from differences in nuclear binding energy. During the fission process, nuclei with even mass numbers are generally more stable and exhibit higher yields than those with odd mass numbers, producing a characteristic jagged pattern in the fission product yield distribution [6]. Reports of recent studies [25,26] have described that the odd–even effect helps to capture the local variations in the yield distribution better and improves the predictive accuracy of FPY models. To account for the odd–even effect, we specifically introduce the parity indicator O_i as an additional input feature as follows: when A_i is even, $O_i = 1$, representing the even-mass enhancement effect; when A_i is odd, $O_i = 0$. For non-peak data, the value of O_i is set to 0.5 in order to reduce the parity contrast in regions where the odd–even effect is less pronounced. It is noteworthy that the value of variable O_i is not a physical observable but a binary encoding of the mass-number A_i parity.

As described in Section 2.2, the classification of peak and non-peak data is determined according to Equation 9 based on the normalized FPY value. Specifically, data points satisfying $normalize(y_i) \geq r$ are classified as peak data, whereas those with $normalize(y_i) < r$ are classified as non-peak data. Accordingly, the value of O_i is assigned automatically based on this classification.

3.1.3. Performance comparison among BNN, DNN, and the proposed method

We compare the prediction errors of the FPY values and the FPY error values among BNN, DNN, and multi-task DNN for both peak and non-peak data. For this study, we use data from JENDL-5 [2], which includes both evaluated FPY values and FPY error data. Particularly, the FPY error is estimated via error propagation to obtain the experimental error, considering the covariance matrix of the variables necessary for estimating the FPY value [27]. The FPY and the associated FPY error in JENDL are evaluated based on Equations (6) and (7) in [27], respectively. Our multi-task learning method learns these evaluation results as training data. Equations (6) and (7) share a common second term. In this case, the proposed method is expected to increase the prediction accuracies of both the FPY data and the FPY error because the architecture enables learning the FPY and learning the FPY error supplementarily, with mutual assistance. To ensure this advantage, we train the BNN and the DNN using the FPY data and the FPY error independently to compare their predictions with those of the multi-task DNN. Particularly, this advantage can be evaluated directly by comparing the performances of the proposed method and (single task) DNNs, which learn the FPY data and the FPY error data independently, because it can be controlled under the same experimental conditions. Specifically, we compare the prediction performances of BNN, DNN using the MSE loss function ('DNN (MSE)'), DNN using the weighted loss function ('DNN (Weighted loss function)'), DNN using the MSE loss function with odd–even values ('DNN (MSE) + odd–even'), DNN using the weighted loss function with odd–even values ('DNN (Weighted loss function) + odd–even'), Multi-task DNN using the MSE ('Multi-task DNN (MSE)'), Multi-task DNN using the weighted loss function ('Multi-task DNN (Weighted loss function)'), Multi-task DNN using the MSE with odd–even values ('Multi-task DNN (MSE) + odd–even'), and Multi-task DNN using the weighted loss function with odd–even values ('Multi-task DNN (Weighted loss function) + odd–even'). Because a BNN estimates the Gaussian distribution to approximate the data distribution, predictive evaluation based only on the expected value estimates of FPY without using CIs might impede the BNN. Therefore, it is noteworthy that this experiment is limited to evaluating point estimates related to the prediction of the detailed peak structure. Table 1 presents the validation errors for both FPY and FPY errors for each method at energies of 2.53×10^{-8} MeV (thermal neutron energy), 0.5 MeV, and 14 MeV.

In Table 1, 'energy' represents the excitation energy in MeV. The 'average total χ^2', 'average peak χ^2', and "average non-peak χ^2"respectively stand for the mean values obtained under various experimental conditions for total χ^2, peak χ^2, and non-peak χ^2. Total χ^2 is defined as $\sum_i [t_i - f(x_i)]^2/n$, where t_i signifies the

Table 1. Average validation errors $\chi^2(10^{-6})$ of various models under three energy conditions.

		average			FPY error		
Methods	Energy	Total χ^2	Peak χ^2	Non-peak χ^2	Average total χ^2	Average peak χ^2	Average non-peak χ^2
BNN	2.53×10^{-8}	4.66	6.74	4.03	10.45	33.79	3.34
DNN (MSE)	2.53×10^{-8}	7.09	8.30	6.73	7.47	19.14	3.91
DNN (Weighted loss function)	2.53×10^{-8}	9.51	7.18	10.22	7.47	19.14	3.91
DNN (MSE) + odd–even	2.53×10^{-8}	4.58	9.36	3.13	6.38	18.59	2.65
DNN (Weighted loss function)+oddeven	2.53×10^{-8}	4.52	9.26	**3.08**	6.38	18.59	2.65
Multi-task DNN (MSE)	2.53×10^{-8}	4.46	6.28	3.90	6.81	17.47	3.56
Multi-task DNN (Weighted loss function)	2.53×10^{-8}	5.00	5.31	4.90	7.05	18.82	3.46
Multi-task DNN (MSE) + odd–even	2.53×10^{-8}	4.43	5.53	5.87	4.20	11.36	2.01
Multi-task DNN (Weighted loss function) + odd–even	2.53×10^{-8}	**3.36**	**4.26**	**3.08**	**3.75**	**9.87**	**1.89**
BNN	0.5	7.49	15.70	4.99	9.13	29.5	2.92
DNN (MSE)	0.5	7.19	15.53	4.80	4.70	**9.44**	3.10
DNN (Weighted loss function)	0.5	8.26	14.52	6.50	4.70	**9.44**	3.10
DNN (MSE) + odd–even	0.5	6.53	16.59	2.97	4.41	10.48	2.56
DNN (Weighted loss function) + odd–even	0.5	5.95	13.97	3.65	4.41	10.48	2.56
Multi-task DNN (MSE)	0.5	6.94	17.65	3.67	4.06	12.25	1.23
Multi-task DNN (Weighted loss function)	0.5	6.26	13.98	3.91	5.05	15.40	1.89
Multi-task DNN (MSE) + odd–even	0.5	6.53	14.44	4.12	**3.16**	9.57	1.20
Multi-task DNN (Weighted loss function) + odd–even	0.5	**5.50**	**11.60**	**3.63**	3.21	10.10	**1.11**
BNN	14	15.03	39.90	7.45	9.10	17.06	6.67
DNN (MSE)	14	10.02	22.17	6.98	4.99	8.07	4.06
DNN (Weighted loss function)	14	11.02	19.62	8.41	4.99	8.07	4.06
DNN (MSE) + odd–even	14	8.74	17.58	6.05	6.37	12.14	4.41
DNN (Weighted loss function) + odd–even	14	8.24	15.65	5.97	6.37	12.14	4.41
Multi-task DNN (MSE)	14	9.57	21.05	6.07	3.50	**5.27**	2.95
Multi-task DNN (Weighted loss function)	14	8.06	20.77	**4.19**	2.30	5.88	3.82
Multi-task DNN (MSE) + odd–even	14	9.77	23.27	5.65	**2.12**	5.76	**1.01**
Multi-task DNN (Weighted loss function) + odd–even	14	**7.35**	**13.08**	5.60	2.48	5.59	1.53

i-th ground truth value of the test set. The error χ^2 value ('peak χ^2') corresponds only to the peak structure, whereas the error χ^2 value ('non-peak χ^2') corresponds only to the non-peak structure.

Similarly, the 'FPY error average total χ^2', 'FPY error average peak χ^2', and "FPY error average non-peak χ^2"respectively represent the mean values obtained under various experimental conditions for FPY error total χ^2, FPY error peak χ^2, and FPY error non-peak χ^2. Here, the FPY error total χ^2 represents the total FPY error value. FPY error peak χ^2 denotes the FPY error value for the peak structure. In addition, FPY error non-peak χ^2 stands for the FPY error value for the non-peak structure.

It is apparent from Table 1 that at either energy (2.53×10^{-8} MeV, 0.5 MeV or 14 MeV), the multi-task DNN with the weighted loss function and the odd–even effect provides the best FPY prediction accuracies for average total χ^2, with the lowest average prediction errors: better than BNN and DNN. Although the average errors (total errors χ^2) for the FPY error prediction of the multi-task DNNs also exhibit the lowest values, the optimal loss function (MSE or weighted loss function) depends on the nucleus and its excitation energy. The reason is that the loss function weight α in (4) was optimized to maximize only the FPY prediction. In addition, the average errors (total errors χ^2) for FPY error prediction of the Multi-task DNN (Weighted loss function)+odd–even model is only slightly inferior to those of the best prediction model. Consequently, when we specifically examine FPY prediction, the experimental results suggest that the Multi-task DNN (Weighted loss function) + odd–even model is recommended to predict unknown FPY values.

We present all error values in Tables 2–5 to clarify the superiority of the proposed methods. In Tables 2–5, the term 'nuclei' denotes atomic nuclei. 'Energy' denotes the excitation energy in units of mega electron volts. The terms 'total χ^2', 'peak χ^2', 'non-peak χ^2', 'FPY error total χ^2', 'FPY error peak χ^2', and "FPY error non-peak χ^2"are the same as those in Table 1. The definitions of 'total χ^2', 'peak χ^2', 'non-peak χ^2', 'FPY error total χ^2', 'FPY error peak χ^2', and "FPY error non-peak χ^2"remain consistent those of earlier descriptions.

Findings presented in Tables 2–5 demonstrate that the Multi-task DNN methods provide lower χ^2 values for predictions of FPY compared to BNN and DNN in most cases. The reason is that MMoE for the simultaneous learning of both FPY and FPY errors improves the prediction accuracies because the FPY error value is related to the FPY value. They affect one another, which increases their learning accuracies.

Table 2 and Table 3 signify that the DNN with MSE loss function provides lower values of both total χ^2 and peak χ^2 than BNN does for ^{235}U and ^{238}U at both 0.5 MeV and 14 MeV. However, for ^{239}Pu at 0.5 MeV in Table 4, as well as ^{241}Pu at both 0.5 MeV and 2.53×10^{-8} MeV in Table 5, the DNN with MSE loss function provides higher values of both total χ^2 and peak χ^2 compared to the BNN.

For the prediction of the FPY error, except for ^{235}U at 14 MeV, the FPY error total χ^2 and the FPY error peak χ^2 values of the DNN are lower than those of the BNN.

In summary, the DNN demonstrates comparable predictive capability to the BNN for FPY. In contrast, it demonstrates an advantage over the BNN in predicting FPY error, achieving consistently higher accuracy.

Furthermore, Tables 2–5 indicate that the Multi-task DNN with MSE loss function provides lower values of total χ^2 than DNN with the MSE loss function does for ^{239}Pu at either 0.5 MeV or 14 MeV, ^{241}Pu at 0.5 MeV or 2.53×10^{-8}MeV, or for ^{238}U at 14 MeV. Although the χ^2 values for ^{235}U (0.5 MeV and 14 MeV) and ^{238}U (0.5 MeV) are higher for the multi-task DNN with the MSE loss function compared to the DNN with the MSE loss function, they are still lower than those of the BNN.

Table 2. Validation errors $\chi^2(10^{-6})$ for ^{235}U obtained using various methods.

Methods	Nuclei	Energy	Total χ^2	Peak χ^2	Non-peak χ^2	FPY error		
						Total χ^2	Peak χ^2	Non-peak χ^2
BNN	^{235}U	0.5	9.33	24.30	4.77	6.30	16.20	3.28
DNN (MSE)	^{235}U	0.5	6.53	18.49	2.89	3.01	**8.54**	1.32
DNN (Weighted loss function)	^{235}U	0.5	8.40	18.24	5.39	3.01	**8.54**	1.32
DNN (MSE) + oddeven	^{235}U	0.5	8.09	24.62	3.05	4.75	16.60	1.14
DNN (Weighted loss function) + oddeven	^{235}U	0.5	6.64	19.59	2.69	4.75	16.60	1.14
Multi-task DNN (MSE)	^{235}U	0.5	7.30	17.85	4.09	**2.69**	9.19	0.76
Multi-task DNN (Weighted loss function)	^{235}U	0.5	5.91	16.67	2.63	3.50	12.56	0.74
Multi-task DNN (MSE) + odd–even	^{235}U	0.5	7.00	17.16	3.90	2.77	9.98	**0.57**
Multi-task DNN (Weighted loss function) + odd–even	^{235}U	0.5	**5.71**	**16.58**	**2.39**	4.05	15.12	0.68
BNN	^{235}U	14	20.70	62.96	7.81	6.83	12.75	5.03
DNN (MSE)	^{235}U	14	5.66	14.02	3.12	2.82	6.46	1.71
DNN (Weighted loss function)	^{235}U	14	5.46	11.73	3.56	2.82	6.46	1.71
DNN (MSE) + odd–even	^{235}U	14	5.47	12.75	3.25	6.38	18.59	2.65
DNN (Weighted loss function) + odd–even	^{235}U	14	6.15	12.04	4.35	6.38	18.59	2.65
Multi-task DNN (MSE)	^{235}U	14	8.91	20.08	5.50	2.37	6.68	**1.05**
Multi-task DNN (Weighted loss function)	^{235}U	14	5.96	11.44	4.29	2.63	7.04	1.28
Multi-task DNN (MSE) + odd–even	^{235}U	14	12.11	28.53	7.10	2.19	5.21	1.27
Multi-task DNN (Weighted loss function) + odd–even	^{235}U	14	**3.44**	**8.32**	**1.95**	**1.78**	**3.56**	1.23

Table 3. Validation errors $\chi^2(10^{-6})$ for ^{238}U obtained using various methods.

Methods	Nuclei	Energy	total χ^2	Peak χ^2	Non-peak χ^2	FPY error		
						Total χ^2	Peak χ^2	Non-peak χ^2
BNN	^{238}U	0.5	11.33	24.06	7.45	9.99	35.94	2.08
DNN (MSE)	^{238}U	0.5	9.98	21.09	6.60	8.01	11.28	7.02
DNN (Weighted loss function)	^{238}U	0.5	15.87	21.16	14.26	8.01	11.28	7.02
DNN (MSE) + odd–even	^{238}U	0.5	7.45	23.84	2.46	4.92	10.97	3.07
DNN (Weighted loss function) + odd–even	^{238}U	0.5	8.80	19.75	5.46	4.92	10.97	3.07
Multi-task DNN (MSE)	^{238}U	0.5	10.48	30.67	**4.33**	7.27	23.28	2.39
Multi-task DNN (Weighted loss function)	^{238}U	0.5	9.78	22.01	6.06	9.48	29.66	3.33
Multi-task DNN (MSE) + odd–even	^{238}U	0.5	9.21	24.84	4.45	2.55	8.43	0.76
Multi-task DNN (Weighted loss function) + odd–even	^{238}U	0.5	**8.53**	**16.92**	5.97	**1.82**	**5.71**	**0.64**
BNN	^{238}U	14	15.48	36.42	9.10	10.75	20.30	7.83
DNN (MSE)	^{238}U	14	15.65	31.19	10.91	7.61	11.83	6.32
DNN (Weighted loss function)	^{238}U	14	18.03	27.12	15.26	7.61	11.83	6.32
DNN (MSE) + odd–even	^{238}U	14	13.37	26.96	9.23	9.29	13.96	7.87
DNN (Weighted loss function) + odd–even	^{238}U	14	11.48	22.80	8.02	9.29	13.96	7.87
Multi-task DNN (MSE)	^{238}U	14	12.79	26.88	8.50	6.03	**4.27**	6.57
Multi-task DNN (Weighted loss function)	^{238}U	14	**6.41**	**14.47**	**3.95**	**2.19**	4.54	1.47
Multi-task DNN (MSE) + odd–even	^{238}U	14	11.19	25.96	6.69	2.63	7.58	**1.12**
Multi-task DNN (Weighted loss function) + odd–even	^{238}U	14	13.53	19.23	11.79	3.44	6.21	2.60

Table 4. Validation errors $\chi^2(10^{-6})$ for ^{239}Pu obtained using various methods.

Methods	Nuclei	Energy	Total χ^2	Peak χ^2	Non-peak χ^2	FPY error		
						Total χ^2	Peak χ^2	Non-peak χ^2
BNN	^{239}Pu	0.5	5.58	7.99	4.84	10.78	36.64	2.89
DNN (MSE)	^{239}Pu	0.5	6.65	12.35	4.92	3.62	6.93	2.61
DNN (Weighted loss function)	^{239}Pu	0.5	4.30	8.83	2.91	3.62	6.93	2.61
DNN (MSE) + oddeven	^{239}Pu	0.5	4.41	8.11	3.28	4.35	**2.95**	4.77
DNN (Weighted loss function) + oddeven	^{239}Pu	0.5	4.22	7.74	3.14	4.35	**2.95**	4.77
Multi-task DNN (MSE)	^{239}Pu	0.5	5.13	7.65	4.36	**3.37**	9.49	**0.15**
Multi-task DNN (Weighted loss function)	^{239}Pu	0.5	**2.82**	**7.13**	**1.50**	3.87	10.91	1.73
Multi-task DNN (MSE) + oddeven	^{239}Pu	0.5	4.37	7.60	3.39	4.02	12.01	1.58
Multi-task DNN (Weighted loss function) + oddeven	^{239}Pu	0.5	4.25	7.57	3.24	3.91	11.90	1.47
BNN	^{239}Pu	14	8.91	20.32	5.44	9.71	18.12	7.15
DNN (MSE)	^{239}Pu	14	8.75	21.31	4.91	4.56	5.91	4.15
DNN (Weighted loss function)	^{239}Pu	14	9.58	20.02	6.40	4.56	5.91	4.15
DNN (MSE) + oddeven	^{239}Pu	14	7.38	13.02	5.66	3.44	5.86	2.70
DNN (Weighted loss function) + oddeven	^{239}Pu	14	7.08	12.12	5.54	3.44	5.86	2.70
Multi-task DNN (MSE)	^{239}Pu	14	7.00	16.2	4.20	2.09	4.86	1.24
Multi-task DNN (Weighted loss function)	^{239}Pu	14	11.82	36.41	4.32	2.09	6.07	8.70
Multi-task DNN (MSE) + oddeven	^{239}Pu	14	6.00	15.31	3.17	**1.53**	**4.50**	**0.63**
Multi-task DNN (Weighted loss function) + oddeven	^{239}Pu	14	**5.07**	**11.68**	**3.07**	2.23	7.01	0.77

Table 5. Validation errors $\chi^2(10^{-6})$ for ^{241}Pu obtained using various methods.

Methods	Nuclei	Energy	total χ^2	Peak χ^2	Non-peak χ^2	FPY error		
						Total χ^2	Peak χ^2	Non-peak χ^2
BNN	^{241}Pu	0.5	3.73	6.46	2.90	9.46	29.27	3.42
DNN (MSE)	^{241}Pu	0.5	5.60	8.20	4.81	4.15	13.01	1.44
DNN (Weighted loss function)	^{241}Pu	0.5	4.47	7.84	3.44	4.15	13.01	1.44
DNN (MSE) + oddeven	^{241}Pu	0.5	4.18	7.77	3.09	3.61	11.39	**1.25**
DNN (Weighted loss function) + oddeven	^{241}Pu	0.5	4.14	6.80	3.32	3.61	11.39	**1.25**
Multi-task DNN (MSE)	^{241}Pu	0.5	4.83	14.42	**1.90**	**2.89**	**7.04**	1.63
Multi-task DNN (Weighted loss function)	^{241}Pu	0.5	6.52	10.09	5.43	3.34	8.47	1.77
Multi-task DNN (MSE) + oddeven	^{241}Pu	0.5	5.53	8.16	4.72	3.28	7.85	1.89
Multi-task DNN (Weighted loss function) + oddeven	^{241}Pu	0.5	**3.49**	**5.31**	2.93	3.06	7.68	1.65
BNN	^{241}Pu	2.53×10^{-8}	4.66	6.74	4.03	10.45	33.79	3.34
DNN (MSE)	^{241}Pu	2.53×10^{-8}	7.09	8.30	6.73	7.47	19.14	3.91
DNN (Weighted loss function)	^{241}Pu	2.53×10^{-8}	9.51	7.18	10.22	7.47	19.14	3.91
DNN (MSE) + oddeven	^{241}Pu	2.53×10^{-8}	4.58	9.36	3.13	6.38	18.59	2.65
DNN (Weighted loss function) + oddeven	^{241}Pu	2.53×10^{-8}	4.52	9.26	3.08	6.38	18.59	2.65
Multi-task DNN (MSE)	^{241}Pu	2.53×10^{-8}	4.46	6.28	3.90	6.81	17.47	3.56
Multi-task DNN (Weighted loss function)	^{241}Pu	2.53×10^{-8}	5.00	5.31	4.90	7.05	18.82	3.46
Multi-task DNN (MSE) + oddeven	^{241}Pu	2.53×10^{-8}	4.43	5.53	5.87	4.20	11.36	2.01
Multi-task DNN (Weighted loss function) + oddeven	^{241}Pu	2.53×10^{-8}	**3.36**	**4.26**	**3.08**	**3.75**	**9.87**	**1.89**

For the prediction of FPY error, the values of FPY error total χ^2 from the multi-task DNN with the MSE loss function are lower than those of the DNN with the MSE loss function, except for ^{238}U at 0.5 MeV.

Additionally, differences in FPY error predictions among the multi-task DNN methods are slight because the loss functions for the FPY error prediction task of multi-task DNN are the same. It is noteworthy that negative predictions of FPY occur frequently in BNN methods, with a proportion of 17.52%. In contrast, multi-task DNN methods demonstrate high predictive accuracy, producing negative predictions of FPY at a lower rate of 8.29%. Even without additional measures, such as incorporating extra data or modifying activation functions, negative predictions in multi-task DNN methods are fewer than those obtained using BNN methods.

3.1.4. Performance comparison of the proposed method with the weighted loss function

In this subsection, we validate the weighted loss function effectiveness by comparing the error values obtained for 'Multi-task DNN (MSE)' and 'Multi-task DNN (Weighted loss function)' and between 'DNN (MSE)' and 'DNN (Weighted loss function).'

For DNN, Tables 2–5 show that, except for ^{238}U at 0.5 MeV, the peak χ^2 values of DNN with the weighted loss function are lower than those of DNN with MSE across all energy levels (0.5 MeV, 14 MeV, and 2.53×10^{-8}MeV).

For Multi-task DNN, Tables 2–5 indicate that, in the case of 0.5 MeV, the peak χ^2 values of Multi-task DNN with the weighted loss function are lower than those of Multi-task DNN with MSE for ^{235}U, ^{238}U, ^{239}Pu, and ^{241}Pu, particularly for ^{238}U. This result represents marked improvement in the peak prediction accuracy using the weighted loss function. Additionally, except for ^{241}Pu, the total χ^2 values for ^{235}U, ^{238}U, ^{239}Pu are lower than those of multi-task DNNs with MSE. For FPY error prediction, the performances of the weighted loss function and MSE are not markedly different.

Findings presented in Tables 2–5 signify that, in the cases of 14 and 2.53×10^{-8}MeV, the peak χ^2 values of multi-task DNNs with the weighted loss function are lower than those of multi-task DNNs with MSE for ^{235}U, ^{238}U, and ^{241}Pu. However, both the peak χ^2 value and the total χ^2 value of the multi-task DNN with the weighted loss function are higher than those of the multi-task DNN with MSE for ^{239}Pu. For the FPY error prediction, similarly to the results obtained for 0.5 MeV, total χ^2 values with the weighted loss function are higher than those with MSE, except for ^{238}U.

3.1.5. Performance comparison of the proposed method with the odd–even effect

Finally, in this subsection, we demonstrate that the odd–even effect parity indicator O_i is effective for enhancing the FPY prediction accuracy.

For DNN, Tables 2–5 show that, except for the cases of ^{235}U at 0.5 MeV (MSE loss function) and 14 MeV (weighted loss function), the total χ^2 values of methods with the odd–even effect are consistently lower than those without the odd–even effect.

For multi-task DNNs, at 0.5 MeV for ^{235}U, ^{238}U, and ^{241}Pu, at 14 MeV for ^{238}U and ^{239}Pu, and at 2.53×10^{-8} MeV for ^{241}Pu, the methods with the odd–even effect consistently exhibit lower peak χ^2 and total χ^2 values compared to those without the odd–even effect, irrespective of whether the loss function is the weighted loss function or MSE. This finding signifies that the introduction of the odd–even effect is effective in most cases.

However, for ^{239}Pu at 0.5 MeV and ^{238}U at 14 MeV, the multi-task DNNs with the weighted loss function, and the odd–even effect exhibit higher peak χ^2 and total χ^2 values than the multi-task DNNs with the weighted loss function but without the odd–even effect. The reason for this difference is that the peak structures of the FPY are smooth. Details underlying the reason for that finding are presented later using figures for the prediction of FPY.

Similarly, from Tables 2 to 5, it is readily apparent that DNNs with the odd–even effect do not consistently exhibit higher or lower FPY error total χ^2 values and FPY error peak χ^2 values compared to those without the odd–even effect. This result is consistent with the fact that the literature has confirmed clear correlation between the odd–even effect and FPY error data [6].

3.1.6. Visual performance comparison among BNN, DNN, and the proposed method

Figures 2 and 3 respectively depict the predictions of FPY for ^{235} U at 0.5 MeV and 14 MeV using BNN, DNN, and the proposed method.

In the respective figures, the horizontal axes show the mass numbers, whereas the vertical axes show the FPY values. Black dots represent the true values of FPY data from JENDL-5 ('JENDL FPY value'). Blue lines represent the predictions of the FPY values using 'BNN,' 'DNN' and 'Multi-task DNN.' The purple shaded regions represent the predictions of FPY errors using 'BNN,' 'DNN' and 'Multi-task DNN.' The red shaded regions represent the true value of the FPY error data from JENDL-5 ('JENDL FPY error').

Figure 2 demonstrates that the predictions of FPY for ^{235} U obtained by the multi-task DNN methods in

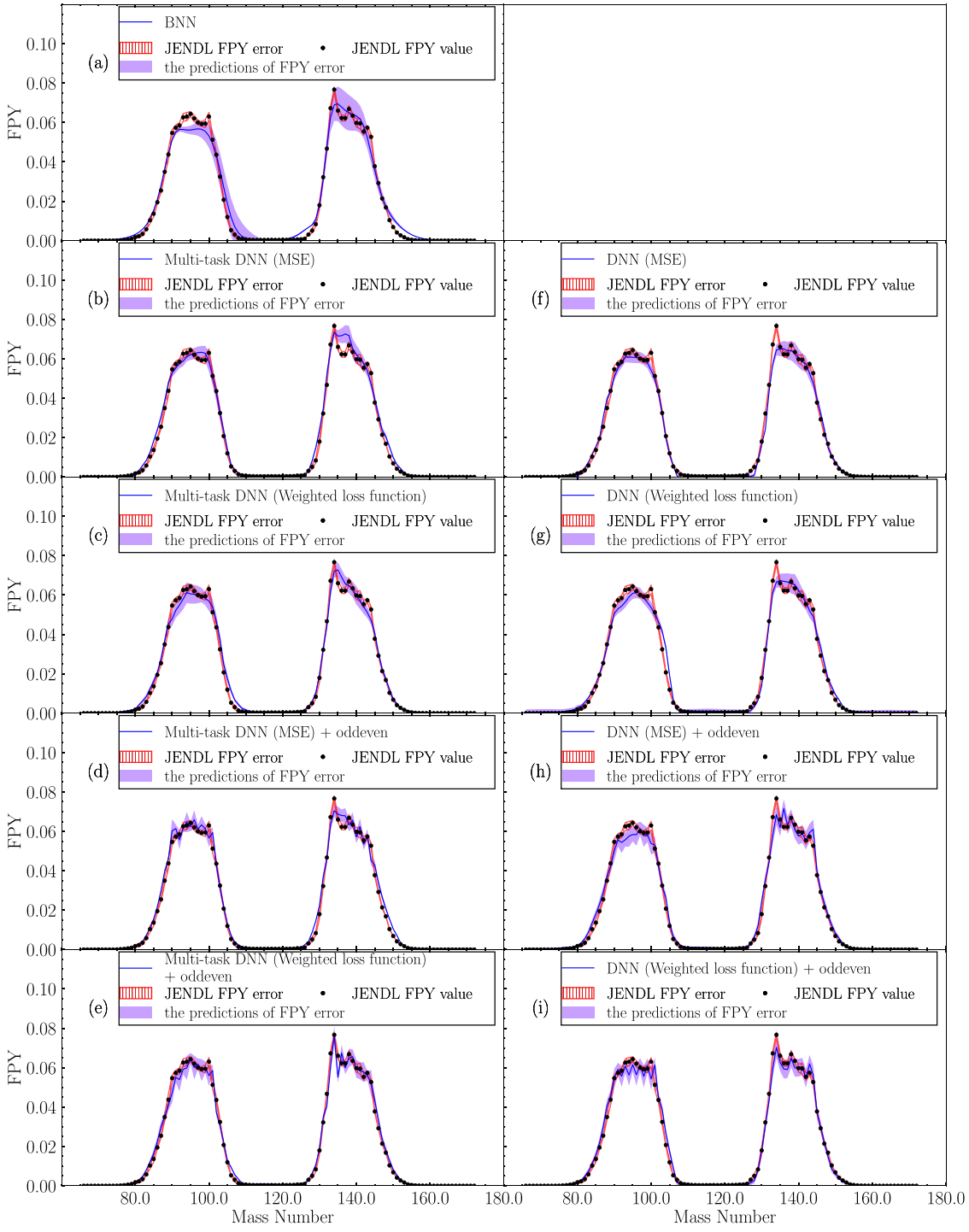

Figure 2. FPY prediction of ^{235}U at 0.5 MeV using various methods.

Figure 2(b–e)are closer to the true FPY values from JENDL-5 than the BNN results shown in Figure 2(a), particularly in the peak region. The proposed multi-task DNN improves predictive accuracy because it directly and simultaneously learns the evaluated nuclear data FPY error together with FPY value, exploiting their complementary relationship during optimization. Among these multi-task DNN methods, Figure 2(e), which incorporates both the weighted loss function and the odd–even effect, provides the true value of FPY data from JENDL-5 in all methods. Specifically, in the actinoid region, the multi-task DNN method using the weighted loss function and the odd–even effect reproduces the characteristic double-humped FPY structure influenced by the spherical magic number at $A = 132$ and the deformed magic number at $A = 142$, with fine oscillatory structures arising from the odd–even effect.

Figure 3 presents that, even at 14 MeV, the multi-task DNN with the weighted loss function and the odd–even effect still exhibits the closest predictions of FPY to the true value of FPY data from JENDL-5. However, for all of these methods, the FPY error predictions are larger in the case of 14 MeV than for

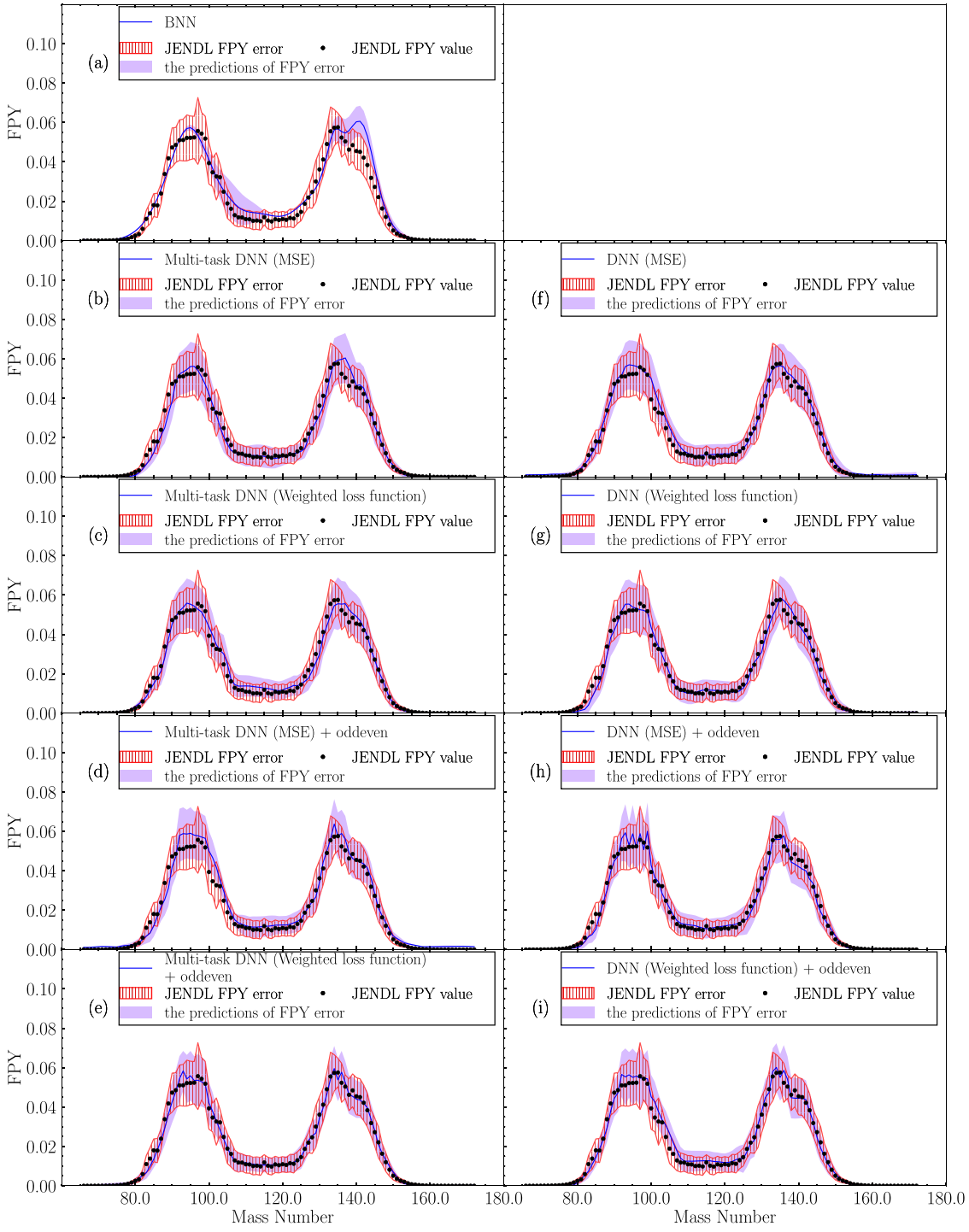

Figure 3. FPY prediction of ^{235}U at 14 MeV using various methods.

0.5 MeV because the covariance matrix evaluated by Tsubakihara et al. [27] is not available at 14 MeV. In this case, the FPY error is estimated as the square root of the sum of squared individual errors, resulting in larger predicted errors. By contrast, in Figure 2, the FPY error for 0.5 MeV is estimated using the covariance matrix, resulting in small predicted errors.

It can be observed that the predictions of FPY using Multi-task DNN with the weighted loss function in Figures 4(a) and 5(a) are sufficiently close to the true value of FPY data from JENDL-5. However, after employing the odd–even effect, the predictions of FPY at peak structures in Figures 4(b) and 5(b) get farther away from the true value of FPY data from JENDL-5. Additionally, one can observe that the shaded region of the predictions of FPY error in Figure 5(a) overlaps more with the shaded region of the true value of FPY error data from JENDL-5. These phenomena might occur because of the extremely smooth peak structures of the FPY for ^{239} Pu at 0.5 MeV and ^{238} U at 14 MeV, where the jagged features might not be readily apparent. Therefore, the odd–

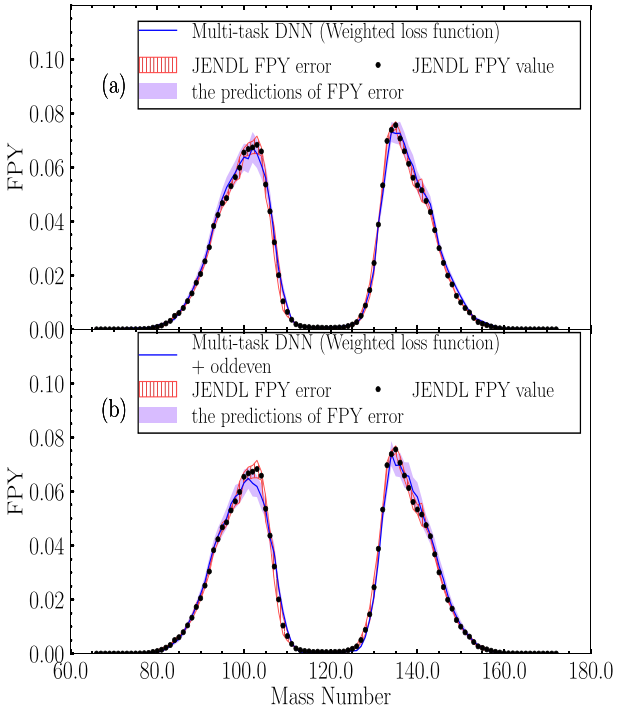

Figure 4. FPY prediction of ^{239}Pu at 0.5 MeV using a weighted loss function and odd–even.

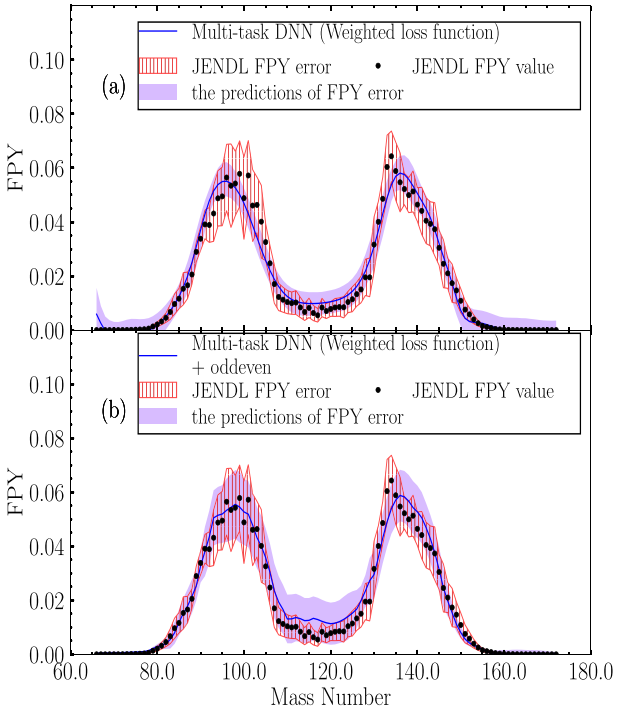

Figure 5. FPY prediction of ^{238}U at 14 MeV using a weighted loss function and odd–even.

even effect is ineffective at increasing the learning accuracies for these data.

Finally, these predictions from the Multi-task DNN with the weighted loss function and the odd–even effect align closely with known physical principles [2,41]. The global double-humped structure widens because of the rise in the temperature of the fissioning nucleus, leading to decreased yield at the peaks and increased yield in the tails. Simultaneously, the fine structures become smoother because peak yields in the fine structure decrease, while the yields in the valleys increase. Additionally, the peak of the heavier fission fragment shifts to smaller A values (to the left) because of an increase in the prompt neutron yield from the heavier fragments. When the neutron energy exceeds several mega electron volts and multi-chance fission occurs, the contributions from fissioning nuclei with mass numbers reduced by 1, 2, or more become considerable, causing the peak of the lighter fission fragment to shift further leftwards. Specifically, the widening of the global double-humped structure and the smoothing of fine structures correspond to the thermal effects and neutron emission dynamics of the fissioning nucleus. The shifts in the peaks of the lighter and heavier fission fragments are also consistent with changes induced by increased prompt neutron production and multi-chance fission. Our results demonstrate that the predictions of the multi-task DNN with the weighted loss function and the odd–even effect not only reproduce experimental trends but also align closely with these established physical mechanisms. The predictions of the FPY error reflect the reliability of our knowledge about this quantity and serve as a basis for evaluating the theoretical model validity. From an engineering perspective, it provides an estimate of the uncertainty in the amount of radioactive nuclides produced in nuclear reactors, offering crucially important guidance for the design feasibility of reprocessing and geological disposal facilities.

3.2. Evaluation of FPY at different incident energies

Our proposed methods are motivated primarily by the need to predict incomplete experimentally obtained FPY data using information from completed assessments of nuclei. Figure 6 presents predictions of FPY and FPY errors using the multi-task DNN with the weighted loss function and the odd–even effect at various energies.

For this experiment, we expand the training dataset to include ^{235} U from JENDL-5. Additionally, we introduce supplementary data generated from the Gaussian model [42] because the training data for the unknown energies are insufficient. The Gaussian model represents the fission product mass-yield distribution as a superposition of five Gaussian functions, each corresponding to characteristic components such as the light and heavy peaks and the symmetric component. The fission product mass-yield distribution provides an analytical approximation of the mass distribution and has been widely used in nuclear data evaluations. These supplementary data augment the training data and function similarly to a prior

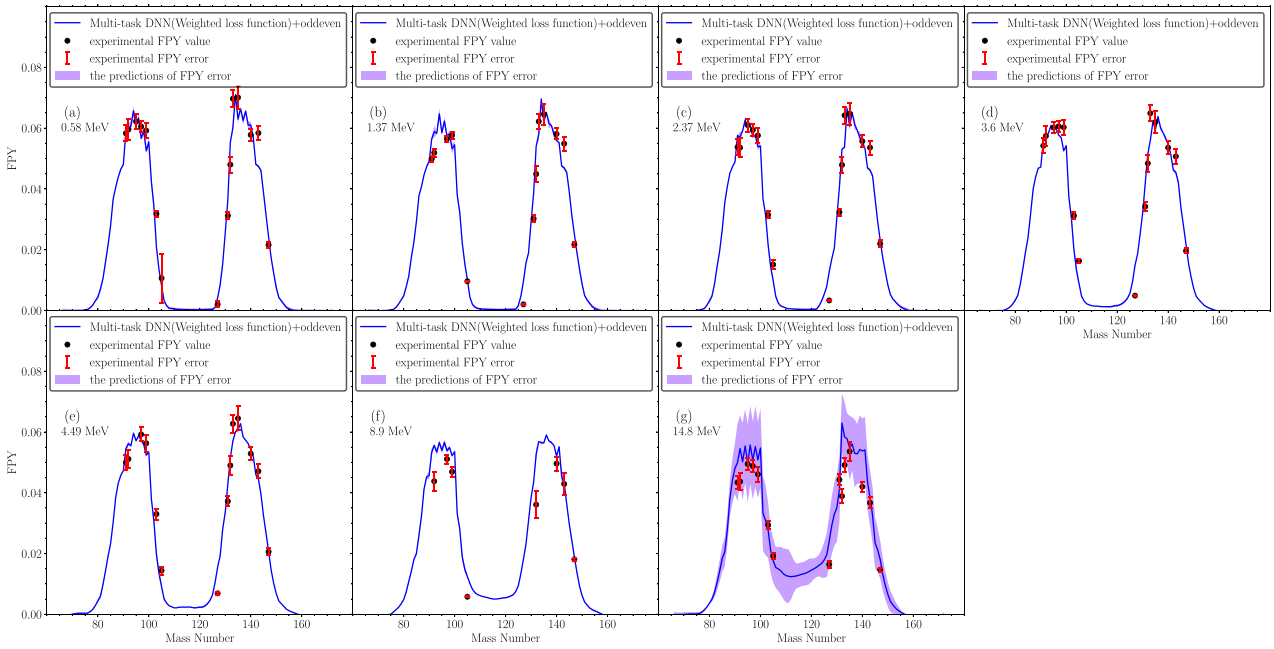

Figure 6. Compiled evaluations of FPY and FPY error for ^{235}U at different incident energies using Multi-task DNN(Weighted loss function + odd–even).

distribution of Bayesian inference. In this study, we employ this model to generate supplementary FPY data at selected excitation energies, thereby compensating for the scarcity of experimental data in the intermediate energy region. It should be noted that in this study, the Katakura Gaussian model was utilized solely to provide supplementary analytical FPY values for the intermediate energy region, where experimental information is scarce. Because this phenomenological model lacks corresponding uncertainty information, the FPY error values for these supplementary data were assigned as zero within the training dataset. Consequently, no correlations were introduced into these analytical data.

For this experiment, we expand the training dataset to include ^{235}U from JENDL-5. Additionally, we introduce supplementary data generated from the Gaussian model [42] because the training data for the unknown energies are insufficient. These supplementary data augment the training data and function similarly to a prior distribution of Bayesian inference. Nevertheless, the FPY error values for these supplementary data are set to zero because it is difficult to estimate the FPY error for the supplementary data generated by the Gaussian model. Although this simplification enables the model to learn the structural features of the mass-yield distribution, it may lead to an underestimation of FPY error in energy regions where experimental covariance information is unavailable.

To identify the supplementary dataset that achieves optimal performance for the FPY prediction, we employ ^{235}U at 0.5 MeV from JENDL-5 as the test dataset to evaluate the effects of different supplementary datasets on prediction accuracy.

The energy range of 0–14 MeV is divided into multiple parts at equal intervals. The division points are then used as excitation energies to generate supplementary data based on the Gaussian model. Details are provided in Table 6, where the 'Number of divisions' represents the number of equal divisions of the 0 to 14 MeV range, 'quantity' denotes the amount of supplementary data, and 'energy' represents the excitation energies of supplementary data in MeV. The meanings of "total χ^2"and "peak χ^2"are consistent with those in Tables 1–5. The case in which the number of divisions is marked as – indicates that no supplementary data are included in the training dataset. This case is treated as the baseline.

Table 6. Validation errors $\chi^2(10^{-6})$ of various supplementary datasets for ^{235}U at 0.5 MeV.

Number of divisions	Quantity	Energy	Total χ^2	peak χ^2
–	0	–	5.34	12.67
2	107	7.00	4.51	8.97
3	214	4.67, 9.33	4.27	9.59
4	321	3.50, 7.00, 10.50	4.06	10.73
5	428	2.80, 5.60, 8.40, 11.20	3.07	6.82
6	535	2.33, 4.67, 7.00, 9.33, 11.67	**2.06**	**6.37**
7	642	2.00, 4.00, 6.00, 8.00, 10.00, 12.00	3.26	9.71
8	749	1.75, 3.50, 5.25, 7.00, 8.75, 10.50, 12.25	3.42	9.93
9	856	1.56, 3.11, 4.67, 6.22, 7.78, 9.33, 10.89, 12.44	4.13	9.34

As shown in Table 6, the values of "total χ^2"and "peak χ^2"achieve their minima at six divisions. Consequently, we incorporate 535 supplementary data points from the six-division scenario into the training dataset to predict incomplete experimental FPY data.

Figure 6 depicts the predictions of FPY and FPY errors using the Multi-task DNN with the weighted loss function and the odd–even effect at various energies: (a) 0.58 MeV, (b) 1.37 MeV, (c) 2.37 MeV, (d) 3.6 MeV, (e) 4.49 MeV, (f) 8.9 MeV, and (g) 14.8 MeV. In Figure 6, the horizontal axis shows the mass number, whereas the vertical axis shows the FPY value. The blue lines represent the predictions of the FPY values. The black dots represent the limited experimental value of FPY data ('experimental FPY value') [36]. The red bars represent the limited experimental value of FPY error data ('experimental FPY error') [36]. The purple shaded regions represent the predictions of FPY error.

Figure 6 presents that the multi-task DNN with the weighted loss function and the odd–even effect exhibits the close predictions of FPY to the experimentally obtained value of FPY data [36]. Additionally, the predictions of the FPY error in Figure 6 are aligned closely with the experimental value of the FPY error data [36].

In Figure 6, an observable trend emerges: as the excitation energy increases, the fission yield in the peak region decreases gradually, whereas the valley region (around mass numbers 110–120) exhibits an upward trend. This trend aligns with the energy dependence pattern of ^{235} U [2,41]. Furthermore, irrespective of variations in the incident energy, the highest point of the right peak structure remains around mass number 134, which is consistent with results obtained from earlier studies [36,43]. These observed features demonstrate the capability of a multi-task DNN with the weighted loss function and the odd–even effect to learn and predict energy-dependent FPY effectively.

On the other hand, the FPY error values exhibit energy-dependent behavior in Figure 6. The FPY error values are small in subfigure panels (a)–(f) because these predicted energies are close to those of the supplementary datasets. Since the supplementary training data were generated using the Gaussian model and their FPY error values were set to zero, the predicted FPY error values in these panels tend to be small. In contrast, subfigure panel (g), corresponding to 14.8 MeV, shows considerably larger FPY error values. This is reasonable because 14.8 MeV is closest to 14 MeV in JENDL-5, for which evaluated FPY error values based on the covariance matrix by Tsubakihara et al. [27] are available. Therefore, at 14.8 MeV, the multi-task DNN with the weighted loss function and the odd–even effect reflects the influence of non-zero FPY error training data, whereas at other energies the FPY error predictions are suppressed due to the absence of reliable error information in the supplementary datasets.

4. Summary

We have introduced an application of a multi-task DNN to learn and predict FPY and FPY errors of actinide nuclei. Multi-task DNN architecture enables learning of FPY data and learning their FPY errors supplementarily while assisting each other. Therefore, the proposed method enhanced the prediction accuracy of FPY. Simultaneously learning both FPY and FPY errors captures their common features more effectively. This innovative approach incorporates a novel loss function and the odd–even effect to improve the accuracy of predicting peak-shaped FPY data considerably while estimating and evaluating the FPY error. Our study has demonstrated the effectiveness of the proposed loss function and the odd–even effect for improving the predictive capabilities of the multi-task DNN, yielding satisfactory outcomes related to the peak-shaped distribution, predicting FPY data, FPY error data simultaneously, and energy dependence of FPY.

Our research has highlighted the enhancement of multi-task DNN learning and predictive abilities for FPY and FPY error to predict the peak-shaped distribution accurately, emphasizing the need for continued refinement of the loss functions. As future work, we expect to improve loss functions of the multi-task DNN approach to improve the peak structure prediction performance.

Because a BNN estimates the Gaussian distribution to approximate the data distribution, predictive evaluation based only on the expected value estimates of FPY without using CIs might be an impediment to the BNN. Alternatively, although this study did not use the CIs of the BNN in the experiments, it can be inferred that the BNN has high predictive accuracy if evaluated based on whether the confidence interval includes the true value. Consequently, it is noteworthy that the comparison results for the proposed method and BNN in this study specifically emphasize prediction of the detailed structure of peaks. Accordingly, it is not claimed for this study that the proposed method is superior to BNN for predicting FPY data because the prediction objectives of BNN and those of the proposed method differ. Using different methods depending on the purpose is important.

A limitation of the present study is that validation was conducted only for major actinides (^{235}U, ^{238}U, ^{239}Pu, and ^{241}Pu), for which abundant and reliably evaluated FPY data are available. The generalization capability of the proposed method to other nuclides

remains to be systematically verified using sufficient experimental datasets. Future work will extend validation to additional nuclides as more comprehensive experimental data become available, in order to assess the robustness and generalization performance of the proposed multi-task framework.

In addition, the supplementary FPY data generated from the Katakura Gaussian model do not include the corresponding FPY error values. Therefore, the FPY error of the supplementary data was set to zero for the training data. Although this simplification enables the model to learn the structural features of the mass-yield distribution, it may lead to an underestimation of the FPY error in energy regions where experimental covariance information is unavailable.

In the future work, we plan to establish a simulation-based framework to estimate FPY errors for excitation energies where experimental covariance data are insufficient. Incorporating physically consistent uncertainty estimation into the supplementary dataset is expected to improve the reliability and predictive robustness of the proposed multi-task model.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Ministry of Education, Culture, Sports, Science, and Technology MEXT Innovative Nuclear Research and Development Program Grant Number JPMXD0222682620, entitled “Fission product yields predicted by machine learning technique at unmeasured energies and its influence on reactor physics assessment,” entrusted to the Tokyo Institute of Technology.

References

[1] Bernstein L, Brown D, Hurst A, et al. Nuclear data needs and capabilities for applications https://arxiv.org/abs/1511.07772

[2] Iwamoto O, Iwamoto N, Kunieda S, et al. Japanese evaluated nuclear data library ver. 5: jENDL-5. J Nucl Sci Tech. 2023;60(1):1–60. doi: 10.1080/00223131.2022.2141903

[3] Ge Z, Zhao Z, Xia H, et al. The updated version of Chinese evaluated nuclear data library (CENDL-3.1). J Korean Phys Soc. 2011;59(2):1052–1056.

[4] Chadwick MB, Herman M, Obložinský P, et al. Endf/b-vii. 1 nuclear data for science and technology: cross sections, covariances, fission product yields and decay data. Nucl Data Sheets. 2011;112(12):2887–2996. doi: 10.1016/j.nds.2011.11.002

[5] OECD Nuclear Energy Agency. The JEFF-3.3 nuclear data library. 2017. https://www.oecd-nea.org/dbdata/jeff/

[6] Schmidt KH, Jurado B. Review on the progress in nuclear fission – experimental methods and theoretical descriptions. Rep Prog Phys. 2018;81(10):106301. doi: 10.1088/1361-6633/aacfa7

[7] Bulgac A, Magierski P, Roche K, et al. Induced fission of 240Pu within a real-time microscopic framework. Phys Rev Lett. 2016;116(12):122504.

[8] Regnier D, Dubray N, Schunck N, et al. Fission fragment charge and mass distributions in Pu 239 (n, f) in the adiabatic nuclear energy density functional theory. Phys Rev C. 2016;93(5):054611. doi: 10.1103/PhysRevC.93.054611

[9] Ivanyuk FA, Ishizuka C, Chiba S. Five-dimensional Langevin approach to fission of atomic nuclei. Phys Rev C. 2024 Mar;109(3):034602. Available from: https://link.aps.org/doi/10.1103/PhysRevC.109.034602

[10] Verriere M, Mumpower MR. Improvements to the macroscopic-microscopic approach of nuclear fission. Phys Rev C. 2021 Mar;103(3):034617. Available from: https://link.aps.org/doi/10.1103/PhysRevC.103.034617

[11] Albertsson M, Carlsson BG, Døssing T, et al. Calculated fission-fragment mass yields and average total kinetic energies of heavy and superheavy nuclei. Eur Phys J A. 2020;56(2):46. doi: 10.1140/epja/s10050-020-00036-9

[12] Fujio K, Okumura S, Ishizuka C, et al. Connection of four-dimensional Langevin model and Hauser–Feshbach theory to describe statistical decay of fission fragments. J Nucl Sci Tech. 2024;61(1):84–97. doi: 10.1080/00223131.2023.2273470

[13] Neufcourt L, Cao Y, Giuliani SA, et al. Quantified limits of the nuclear landscape. Phys Rev C. 2020;101(4):044307. doi: 10.1103/PhysRevC.101.044307

[14] Utama R, Piekarewicz J, Prosper HB. Nuclear mass predictions for the crustal composition of neutron stars: a Bayesian neural network approach. Phys Rev C. 2016;93(1):014311. doi: 10.1103/PhysRevC.93.014311

[15] Niu ZM, Liang HZ. Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects. Phys Lett B. 2018;778:48–53. doi: 10.1016/j.physletb.2018.01.002

[16] Wu D, Bai CL, Sagawa H, et al. Calculation of nuclear charge radii with a trained feed-forward neural network. Phys Rev C. 2020;102(5):054323. doi: 10.1103/PhysRevC.102.054323

[17] Utama R, Chen WC, Piekarewicz J. Nuclear charge radii: density functional theory meets Bayesian neural networks. J Phys G Nucl Part Phys. 2016;43(11):114002. doi: 10.1088/0954-3899/43/11/114002

[18] Ma Y, Su C, Liu J, et al. Predictions of nuclear charge radii and physical interpretations based on the naive Bayesian probability classifier. Phys Rev C. 2020;101(1):014304. doi: 10.1103/PhysRevC.101.014304

[19] Ma CW, Peng D, Wei HL, et al. Isotopic cross-sections in proton induced spallation reactions based on the Bayesian neural network method. Chin Phys C. 2020;44(1):014104. doi: 10.1088/1674-1137/44/1/014104

[20] Lovell A, Nunes FM, Catacora-Rios M, et al. Recent advances in the quantification of uncertainties in reaction theory. J Phys G Nucl Part Phys. 2020;48(1):014001. doi: 10.1088/1361-6471/abba72

[21] Neudecker D, Cabellos O, Clark AR, et al. Informing nuclear physics via machine learning methods with

differential and integral experiments. Phys Rev C. 2021;104(3):034611. doi: 10.1103/PhysRevC.104.034611
[22] Wang Y, Li F, Li Q, et al. Finding signatures of the nuclear symmetry energy in heavy-ion collisions with deep learning. Phys Lett B. 2021;822:136669. doi: 10.1016/j.physletb.2021.136669
[23] Bedaque P, Boehnlein A, Cromaz M, et al. AI for nuclear physics. Eur Phys J A. 2021;57(3):1–27. doi: 10.1140/epja/s10050-020-00290-x
[24] Qiao CY, Pei JC, Wang ZA, et al. Bayesian evaluation of charge yields of fission fragments of U 239. Phys Rev C. 2021;103(3):034621. doi: 10.1103/PhysRevC.103.034621
[25] Wang ZA, Pei J, Liu Y, et al. Bayesian evaluation of incomplete fission yields. Phys Rev Lett. 2019;123(12):122501. doi: 10.1103/PhysRevLett.123.122501
[26] Wang ZA, Pei J. Optimizing multilayer Bayesian neural networks for evaluation of fission yields. Phys Rev C. 2021;104(6):064608. doi: 10.1103/PhysRevC.104.064608
[27] Tsubakihara K, Okumura S, Ishizuka C, et al. Evaluation of fission product yields and associated covariance matrices. J Nucl Sci Tech. 2021;58(2):151–165. doi: 10.1080/00223131.2020.1813643
[28] Bai J, Niu Z, Sun B, et al. The description of giant dipole resonance key parameters with multitask neural networks. Phys Lett B. 2021;815:136147. doi: 10.1016/j.physletb.2021.136147
[29] Wu X, Lu Y, Zhao P. Multi-task learning on nuclear masses and separation energies with the kernel ridge regression. Phys Lett B. 2022;834:137394. doi: 10.1016/j.physletb.2022.137394
[30] Hizawa N, Hagino K. Nonempirical shape dynamics of heavy nuclei with multitask deep learning. Phys Rev C. 2024;109(1):014312. doi: 10.1103/PhysRevC.109.014312
[31] Ma J, Zhao Z, Yi X, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK; 2018. p. 1930–1939.
[32] Zhang Y, Xin Y, Liu Z, et al. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE. Reliab Eng System Saf. 2022;220:108263. doi: 10.1016/j.ress.2021.108263
[33] Chen Y, Ren L, Xia H, et al. A compound fault diagnosis method based on multi-task learning with multi-gate mixture-of-experts. In: 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). Changsha, China: IEEE; 2022. p. 281–285.
[34] Guan T, Peng J, Liang J. Spatial–temporal graph multi-gate mixture-of-expert model for traffic prediction. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). Bilbao, Spain: IEEE; 2023. p. 36–41.
[35] Mills RW. A new UK fission yield evaluation ukfy3.7. In: EPJ Web of Conferences. Bruges, Belgium: EDP Sciences; 2017. Vol. 146. p. 04008.
[36] Gooden ME, Arnold C, Becker J, et al. Energy dependence of fission product yields from 235U, 238U and 239Pu for incident neutron energies between 0.5 and 14.8 MeV. Nucl Data Sheets. 2016;131:319–356. doi: 10.1016/j.nds.2015.12.006
[37] Naik H, Mulik V, Prajapati P, et al. Mass distribution in the quasi-mono-energetic neutron-induced fission of 238U. Nucl Phys A. 2013;913:185–205. Available from: https://www.sciencedirect.com/science/article/pii/S0375947413005861
[38] Naik H, Mukerji S, Crasta R, et al. Measurement of fission product yields in the quasi-mono-energetic neutron-induced fission of 238u. Nucl Phys A. 2015;941:16–37. Available from: https://www.sciencedirect.com/science/article/pii/S0375947415001244
[39] Hoffman MD, Gelman A. The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15(1):1593–1623.
[40] Duane S, Kennedy AD, Pendleton BJ, et al. Hybrid Monte Carlo. Phys Lett B. 1987;195(2):216–222. doi: 10.1016/0370-2693(87)91197-X
[41] Brosa U, Grossmann S, Müller A. Nuclear scission. Phys Rep. 1990;197(4):167–262. doi: 10.1016/0370-1573(90)90114-H
[42] Katakura J. A systematics of fission product mass yields with 5 Gaussian functions. JAERI-Research 2003-004. Jpn At Energy Res Inst; 2003. 36 p.
[43] Tonchev AP, Silano JA, Hagmann C, et al. Toward short-lived and energy-dependent fission product yields from neutron-induced fission. In: EPJ Web of Conferences. Sanary-sur-Mer, France: EDP Sciences; 2020. Vol. 239. p. 03001.