ORIGINAL ARTICLE

AI-Based Prediction Module of Key Neutronic Characteristics to Optimize Loading Pattern for i-SMR with Flexible Operation

Jungseok Kwon[1] · Tongkyu Park[1] · Sung Kyun Zee[1]

Abstract

This paper proposes an AI-based module for a loading pattern (L/P) optimization algorithm applied to the i-SMR, designed for flexible operation. The AI module can be used as a surrogate model in the simulated annealing (SA) screening process, which allows for more efficient optimization. The convolution neural network (CNN) model was trained using reactor core L/Ps and corresponding core parameter values derived from a realistic core simulation code. For load-following operations, we selected core parameters such as control rod insertion depth, radial peaking factor, axial shape index, and effective multiplication factor. To calculate the objective function of an L/P during the SA process using core design codes, it takes approximately 3 s, while the AI-based module can predict the objective function within about 0.1 ms. During the prediction of selected parameters, we discovered two factors affecting prediction accuracy. First, the model exhibited a significant increase in error when trained on dataset containing negative values. Second, utilizing batch normalization (BN) layer and squeeze and excitation (SE) module, intended to improve accuracy, resulted in a decrease in performance of the model. Our study demonstrated that the CNN-based model achieves excellent prediction accuracy and has an ability to accelerate optimization algorithms by taking advantage of artificial intelligence's inherent computational speed.

Keywords Convolution neural network · Loading pattern optimization · i-SMR (innovative small modular reactor) · Flexible operation · Screening technique · Simulated annealing

Introduction

Nuclear Reactor Core Design and L/P Optimization

Nuclear reactor core design is the process of determining the design parameters of a core in a nuclear power plant considering safety, reliability and economical operation. Disciplines in a nuclear core design can be categorized into four main aspects consisting of nuclear design, thermal hydraulics design, structure/material design, and safety/reliability evaluation. To determine the optimal design parameters, there are strongly coupled calculations and iterative design procedures between these disciplines. Nuclear design work involves determining the concentration of fissile uranium isotopes in the fuel, reactivity control elements (such as control rod as neutron absorbers), and the configuration of these fuel and elements. During the process, nuclear designer calculates the power distribution and other neutronics parameters after considering core geometry, reactivity control methods and their location, core lifetime and more [1].

L/P optimization involves determining the specific locations of fuel assemblies in terms of uranium enrichment, gadolinium burnable absorber, and burnup levels. To meet safety constraints and achieve economical operation, reactor core designers have struggled to find an optimal L/P during the design and in-core fuel management stages of pressurized water reactors (PWRs). Optimization of fuel L/P is a challenging task due to its non-deterministic polynomial-time hard (NP-Hard) complexity. This difficulty grows exponentially with the number of fuel assemblies loaded in the reactor core. For instance, a two-loop PWR like 'Kori Unit 1' has 10^{153} possible loading configurations. Even when we use octant (1/8) symmetric core geometry and specific loading rules to restrict the solution space, numerous cases of potential configurations

✉ Tongkyu Park
tongkyu@fnctech.com

1 FNC Technology, Heungdeok IT Valley Bldg. 32F, 13, Heungdeok 1-ro, Giheung-gu, Yongin-si, Gyeonggi-do 16954, Republic of Korea

(approximately 10^{23}) still remained for consideration [2]. Traditionally, the fuel L/P optimization problem has been solved by core design expert's knowledge and experience to construct patterns, test them with design codes, and verify their compliance with design criteria. In an effort to automate this process, various optimization approaches have emerged, including simulated annealing (SA) [3, 4], genetic algorithm (GA) [5], and tabu search [6].
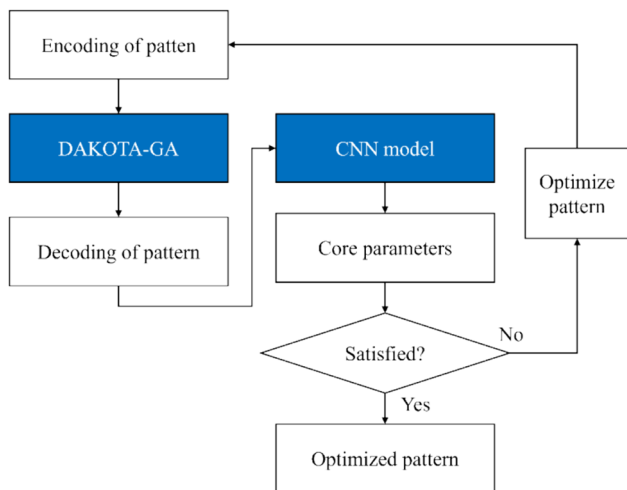

Fig. 1 Flowchart for fuel-reloading pattern optimization using DAKOTA-GA and CNN [14]

In the currently operating commercial PWRs in Korea, composition of fuel assemblies is mainly varied radially, with minimal or no variation in their composition along the axial direction. Reactivity, which is related to the power of each pin and core, can be controlled by soluble boric acid, burnable absorbers and control rods. In the commercial nuclear power plants, control rods are primarily used for urgent power adjustment or reactor shutdown. However, in the case of i-SMR, it employs a flexible operation mode within a soluble boron-free (SBF) environment in the core. There are some issues originating from flexible operation and SBF condition. First, due to this flexible operation mode, control rods have become the primary option for power adjustments in i-SMR. When the control rods are inserted downward from the top-side of the reactor, this induces an unbalanced axial power profile and power fluctuation in the upper and lower regions of the reactor [7]. Second, boric acid plays a role in controlling reactivity to suppress the power profile uniformly throughout the entire core region. However, under SBF operational conditions, burnable absorbers and control rods, which can only affect the reactivity peripherally at their locations, must substitute the role of soluble boron [8].

These factors necessitate additional consideration of the axial variation in the arrangement of fuel and burnable absorbers, unlike the current focus on the radial arrangement in existing commercial PWRs in Korea. Therefore, while it was previously sufficient to optimize L/P by
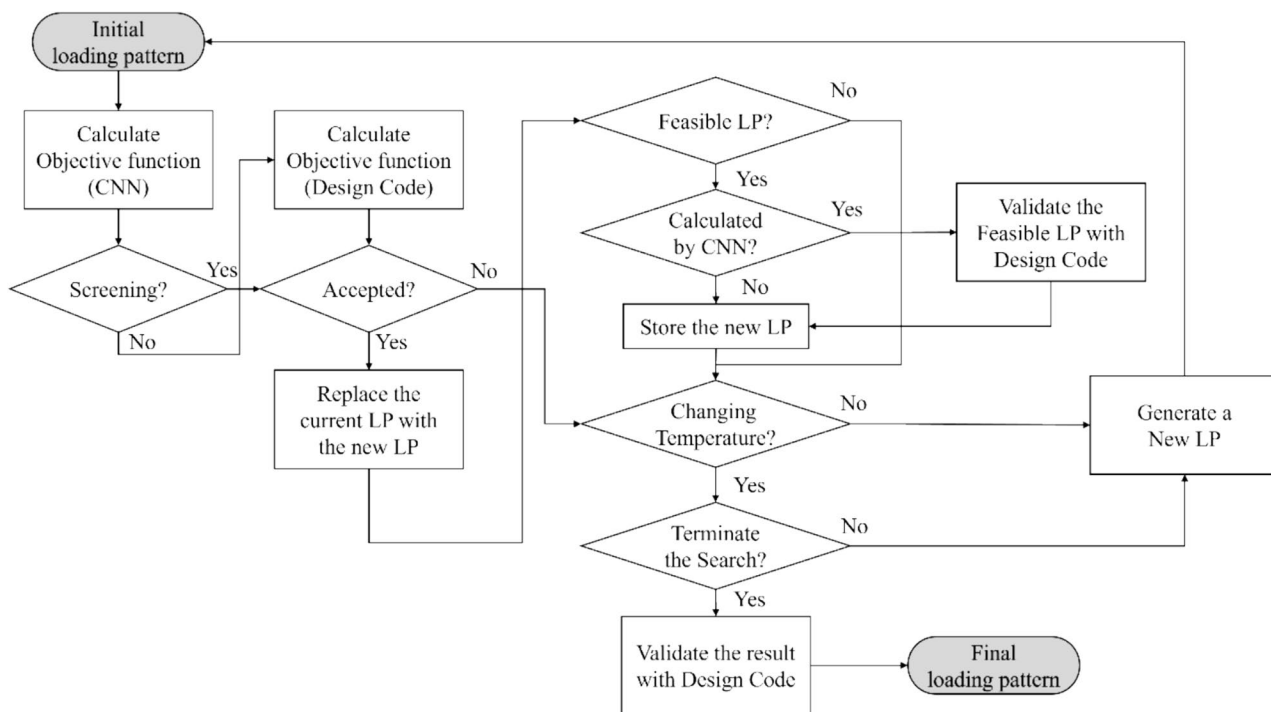

Fig. 2 Flowchart for SA algorithm with screening technique and CNN [6]

considering only the radial direction, it is now necessary to design with consideration for axial compositional variation and the control rod positions for power adjustment. This means that the dimension for analysis has expanded from 2 to 3D, vastly increasing the number of L/Ps that required to be analyzed. For example, when allowing for duplication, the number of possible configurations for arranging five types of fuel batch in an i-SMR is 5^{69}, and if 1/8 core symmetry is allowed, there are 5^{13} possible configurations. If we extend this ideally in the axial direction, there are 24 layers in the depth direction of each fuel batch, and 8 different fuel compositions can be selected in each layer. Thus, advanced techniques are necessary to expedite the L/P optimization process, despite already employing a meta-heuristics algorithm. It should be noted that due to the fabrication problem, axially only a few different zones would be practically feasible.

CNN-Based Core Parameter Prediction

A convolutional neural network (CNN)-based surrogate model has been developed to accelerate the fuel loading pattern (L/P) optimization algorithm in this study. Commonly, CNNs have been widely employed for vision applications such as image classification, object recognition, and segmentation. However, as deep learning technologies have advanced, machine learning frameworks have grown more user-friendly, making it easier for researchers without deep learning knowledge to apply machine learning techniques to their specific fields. During the 1990s, neural networks were utilized in nuclear engineering to predict reactor core parameters, but its use was limited to specific cases of L/Ps rather than being generally applicable [9]. With advancements in both CNN models [10, 11] and computational power, the model is now capable of distinguishing certain characteristics from intricated patterns in image data. Early applications of CNNs in reactor core parameter prediction focused on the eigenvalues and eigenvectors of diffusion equations such as multiplication factor (k_{eff}), neutron flux and power distributions within the core by utilizing core L/Ps [12, 13]. Subsequent studies were conducted to predict core parameters such as power peaking factor, cycle length, boron concentration at the beginning of the cycle (BOC), moderate temperature coefficient (MTC) by CNN with L/Ps as an input data and the calculation results of reactor design code as a labeled data [14].

Advanced Approaches of L/P Optimization

Various meta-heuristic algorithms are being adapted for optimization problems. But it is also time-consuming that

Table 1 Major design features of i-SMR [26]

Reactor type	Integral PWR
Plant capacity (number of reactor modules)	680 MWe (4)
Thermal/electrical capacity per reactor	520 Mwt/170 MWe
Reactor coolant pump	Vertical canned motor type
NSSS operating pressure	15 MPa
Core inlet/outlet coolant temperature	295.5/320.0 ℃
Fuel type/assembly array	UO_2/17×17 square pitch
Number of fuel assemblies in the core	69
Fuel enrichment	<5 w/o
Core discharge burnup	<62,000 MWD/MTU
Refueling cycle, months	24
Reactivity control (soluble boron free)	Control rod, burnable absorber rods, moderator temperature
Steam generator	Helical once-through type
Safety systems	Fully passive
Design life, years	80
Seismic design (SSE)	0.5 g

evaluating each L/P generated by the algorithm with the actual design code. While hundreds to thousands of L/Ps are assessed during this process, the computational cost of the design code itself significantly contributes to the overall assessment time. In addition, in the case of L/P optimization, it is difficult to compare the performances of each algorithm. First, their differences of performance are not significant [15, 16]. It is hard to determine the superiority of the sets of objective functions derived from each algorithm. In addition, even if some algorithms derive better objective functions, they might have slower convergence speeds. In order to improve base algorithms, several approaches have been devised: parallel meta-heuristic algorithm [17, 18], probabilistic sampling approaches [19], and integration of surrogate model into algorithms [20–22].

In some cases where surrogate model approaches are used, only the model itself is employed to derive objective functions [20, 22]. Wan et al. [22] examined the efficiency of combined model of CNN with GA. Their study was conducted on the CNP-1000 PWR core to predict critical boron concentration at specified burnup state and power peaking factors. They selected four characteristic parameters of fuel rod, namely fuel enrichment, burnup, number of burnable rods, and number of cycle burned. These parameters are used as input data for the CNN model. Each loading pattern's burnup and power peaking factors were calculated using the design code, SPARK as target data. The corresponding flowchart can be seen in Fig. 1.

Other cases use a design code after the surrogate model [21]. Several candidate L/Ps are first prepared by heuristic algorithm. Then, the AI-based surrogate model predicts their

objective functions and reduces the number of candidate L/Ps. Finally, the selected L/Ps with better objective functions are calculated using the design code. If we derive the objective functions throughout the entire procedure using a surrogate model, we can take advantage of the convergence speed. However, the inherent uncertainty of the surrogate model could result in a less optimal L/P compared to using only a design code with a heuristic algorithm.

The probabilistic sampling technique presented by Liao et al. [19] constructs a probability distribution between operations (fuel assembly exchange and rotation) and their impact on objective functions. If an operation positively improves the objective function, the probability of that operation is increased through dynamic probability adjustment, improve optimization process.

In probabilistic sampling techniques, there is also a method called the screening technique. Screening technique employs both the probabilistic approach and the surrogate model. It can take the benefits of accelerating with partially resolve the uncertainty of artificial neural networks through probabilistic screening techniques. Instead of assessing all L/Ps using the design code, the screening technique was utilized to select only the valuable ones for examination. This method employs a surrogate model to predict the range of core characteristic parameters, thereby reducing the number of cases that require design code calculation. The flowchart for the screening technique and CNN surrogate model can be seen in Fig. 2.

Unlike Liao et al. [19], this screening technique uses a predetermined distribution function for the L/Ps and their objective functions. Specifically, the distribution function is constructed using the differences between the objective functions calculated by the design code and the surrogate model for the same L/P. This distribution is then used to determine whether a transition occurs in the given SA optimization algorithm. If the values predicted by the surrogate model lie within a specific reliable range of the probability distribution, they are accepted and used in the optimization process. Otherwise, a detailed analysis using the design code is performed. Finally, if the objective function value of the new L/P's being better than the current L/P is significant, the transition is allowed; otherwise, it is not [4, 23, 24].

Overview of i-SMR with Flexible Operation

The i-SMR (innovative small modular reactors) is a small modular reactor being developed in Korea that offers significant advantages over conventional nuclear power plants in terms of safety, economic viability, and operational

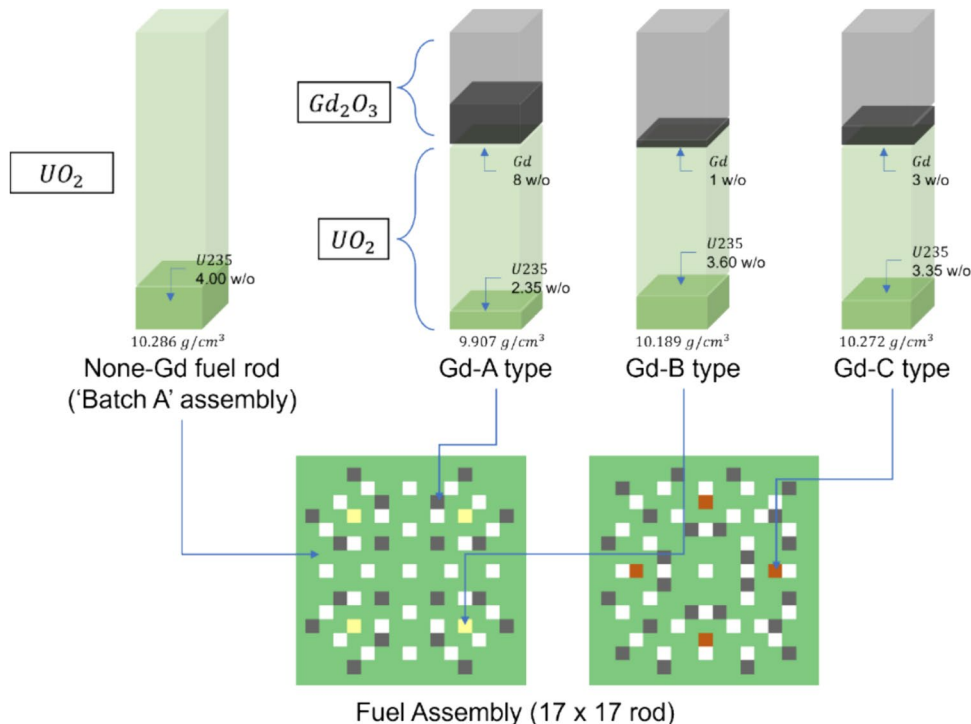**Table 2** Gadolinium burnable poison rod options

Rod type	U-235 w/o	Density [g/cm^3]	Gd w/o
None-Gd	4.00	10.2863	0
Gd-A	2.35	9.9066	8.0%
Gd-B	3.60	10.1887	1.0%
Gd-C	3.35	10.2715	3.0%

Fig. 3 Gadolinium burnable poison rod options

Fuel Assembly (17 x 17 rod)

Table 3 Categories of fuel batches

Category	Number of Gd rods
Low	0, 12
Intermediate	16, 20
High	24, 28

flexibility. It features a 520 MW thermal power output, an integrated reactor coolant system (RCS) configuration that eliminates large pipes, and a fully passive safety system. The i-SMR can be utilized for various applications beyond electricity generation, including industrial heat supply, seawater desalination, and hydrogen production. Table 1 represents the major design features of the i-SMR.

Traditionally, nuclear power plants have been utilized as baseload power sources. However, the rise of renewable energy generation, such as solar and wind power, has led to inconsistent power supply–demand depending on weather conditions, resulting in either insufficient or excessive power supply. Traditionally, nuclear power plants have been utilized as baseload power sources. However, the rise of renewable energy generation, such as solar and wind power, has led to inconsistent power supply–demand depending on weather conditions, resulting in either insufficient or excessive power supply. In power systems, carbon-based sources (such as coal and natural gas) have traditionally served as flexible generation sources. However, due to their greenhouse gas emissions, decarbonization of the power system has become necessary.

The flexible operation of nuclear power can help decrease the share of carbon-based generation while preventing the curtailment of renewable energy sources. In addition, it can help reduce the operational and maintenance (O&M) costs of the power system. By achieving flexible operation of nuclear power plants, we can improve the performance of nuclear systems by addressing issues such as reactivity fluctuations caused by xenon and enhancing the reliability of nuclear fuel and structural materials. Consequently, there is a growing demand for the flexible operation of nuclear power plants [25].

The i-SMR is developed to meet this demand with its innovative design. It employs a soluble boron-free (SBF) operation mode for load-following capability. In addition, the i-SMR features a high level of autonomous operation, which reduces operator burden during multi-module
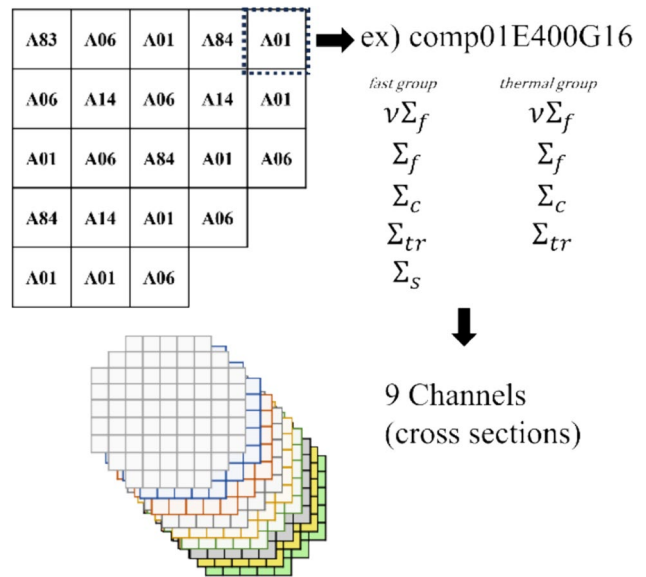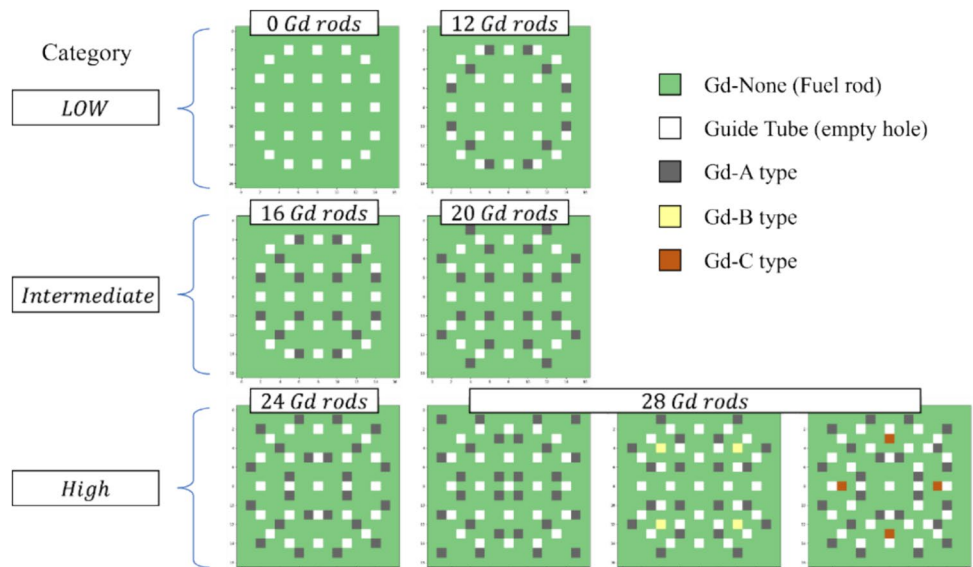

Fig. 5 L/P preprocess with macroscopic cross-sections
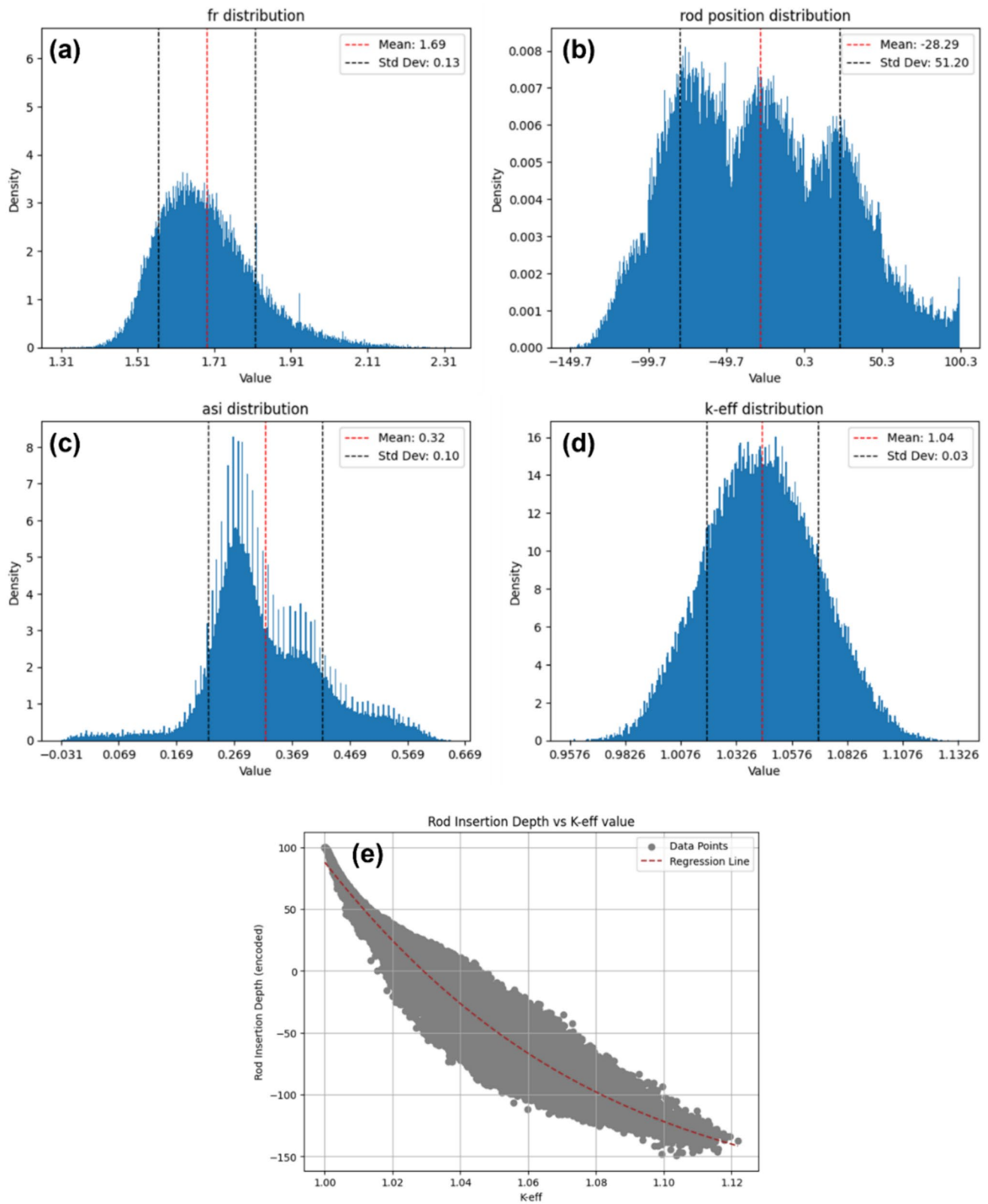
Fig. 4 Categories of fuel batches

Fig. 6 **a–d** Data distributions of Fr, rod position, ASI, *k*-eff, respectively. **e** Rod depth vs *k*-eff in the same L/P dataset
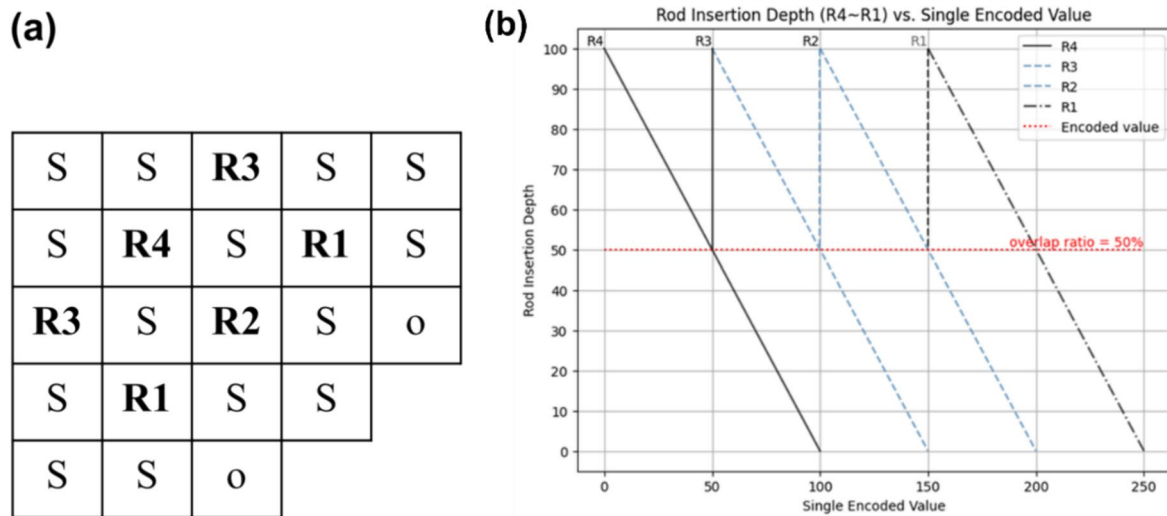
Fig. 7 **a** Control rod layout of quarter core. **b** Rod insertion strategy

Table 4 Pseudocode for control rod transformation algorithm

```
Rod_banks = [R4, R3, R2, R1]
R4-R1 varies 0 to 100
# condition 1: [0, 0, R2, R1], R1 != 100
if R1 != 100:
single_encoded_value = R1 – 150
# condition 2: [0,R3,R2,100], R2 != 100
elif R2 != 100 and R1 == 100:
single_encoded_value = R2–100
# condition 3: [R4,R3,100,100], R3 != 100
elif R3 != 100 and R2 == 100 and R1 == 100:
single_encoded_value = R3–50
# condition 4: [R4,100,100,100], R1,R2,R3 == 100
elif R3 == 100 and R2 == 100 and R1 == 100:
single_encoded_value = R4
```

operation. Furthermore, the integrated RCS configuration inherently eliminates the possibility of a large break loss-of-coolant accident (LB-LOCA), further enhancing the safety of the i-SMR [26]. Researches have been actively conducted in Korea for i-SMR that operating under load-following and SBF conditions. Researches include developing methodologies for optimizing fuel loading patterns for load-following operation [27], analyzing candidate safety system designs [28], and confirming the applicability of burnable absorbers [29].

Methodology

This study aims to predict key reactor core parameters at BOC, including power peaking factor, control rod insertion depth, k_{eff}, and Axial Shape Index (ASI). We achieve this by employing a neural network model trained on both core loading patterns and calculation results generated by the ASTRA design code. The main focus of this study is to predict these parameters efficiently while maintaining a sufficient level of precision using CNN-based model.

Research Object: i-SMR

This section illustrates the design features of the i-SMR, particularly focus on the effect of SBF condition. Here are the two key considerations for SBF operation:

(1) Increased control rod reactivity worth: unlike conventional reactors relying on a combination of BA (burnable absorbers), control rods, and dissolved boron for reactivity control, the i-SMR's boron-free core requires only control rods and burnable absorbers to manage excess reactivity [29]. This necessitates increased control rod reactivity worth, which influence reactor design and operational strategies.

(2) Enhanced dependency on burnable absorbers for power profile: in conventional reactors, dissolved boron flattens the axial and radial power profile within the core. In the SBF condition, i-SMR rely more heavily on burnable absorbers rods embedded within the fuel assemblies for this purpose. Careful consideration is required during the design stage include the quantity and placement of these absorber rods to ensure an optimal power profile, effective reactivity control, and material integrity during load-following operations [30].
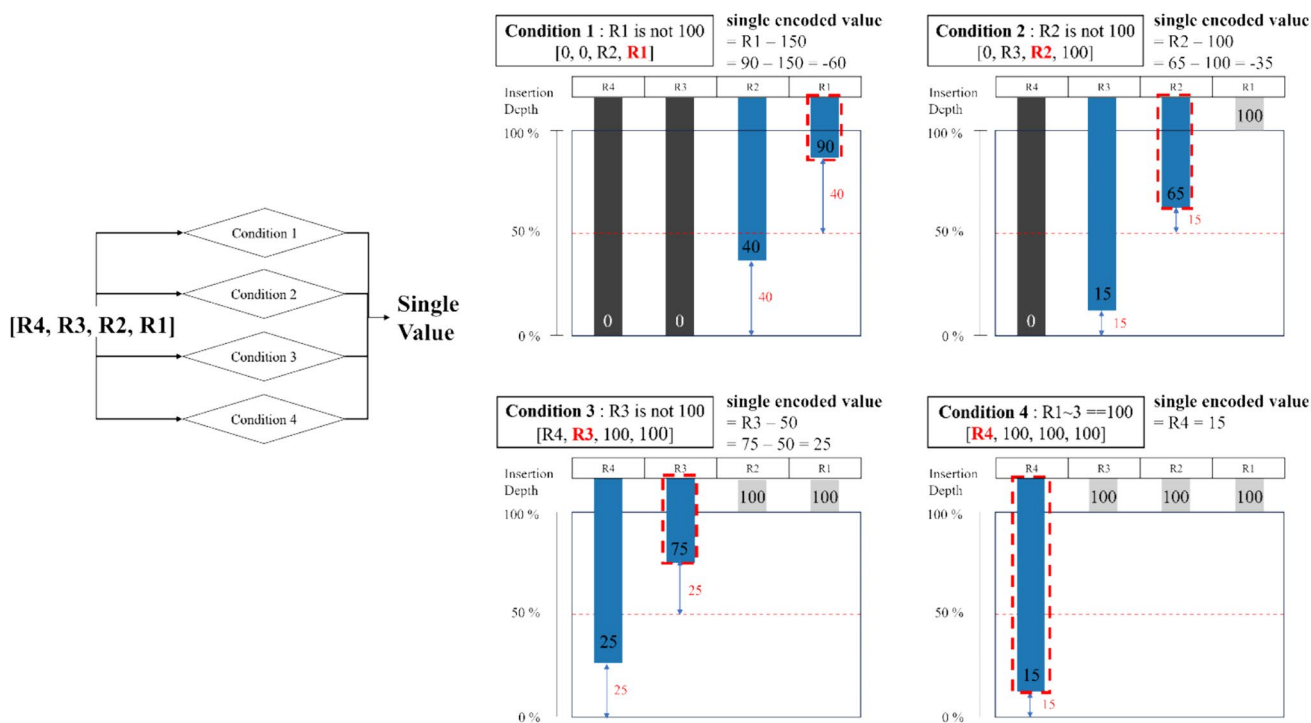
Fig. 8 Example of control rod transformation algorithm

Data Preparation and Preprocess

Input Data

(1) Batch type and composition

In this study, we considered 24 batch types of fresh fuel assemblies as a part of input data. Each batch type represents a fuel assembly with a unique composition and arrangement of fuel and burnable absorber rods. The enrichment of U-235 within the fuel assemblies has two options: 4 and 4.95 w/o However, at the initial cycle, only the 4 w/o Batch A type assembly will be used. In addition, each batch type includes gadolinium (Gd) burnable absorber rods in six variations: 0, 12, 16, 20, 24, and 28 rods as shown in Table 3 and Fig. 4. In
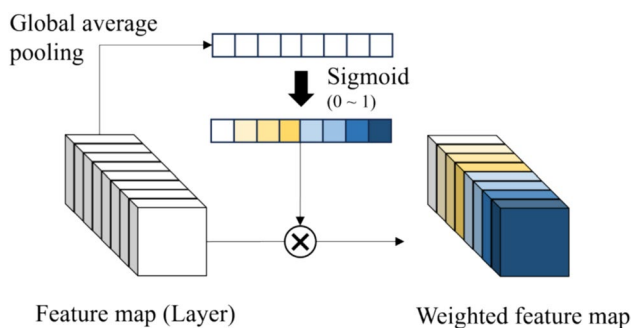

Fig. 9 SE block

addition, these Gd rods are classified into three types based on the specific combinations of UO_2 and Gd_2O_3, as shown in Table 2 and Fig. 3.

(2) Cross-section data of each batch types

Fast and thermal group macroscopic cross-section data of each batch types produced by the KNF's lattice physics design code, KARMA, and ENDF/B-VI.8 library has been used. Macroscopic cross-section consists of five types of cross-section; fission, ν-fission, capture, transport and scattering. Since the thermal group does not have scattering cross-section data, this is because only neutrons of the fast group scatters into the thermal group, while the opposite scattering (up-scattering) is negligible and is considered as 0 in the two-group cross-sections. Each batch types have nine cross-section coefficients related with their own composition.

(3) Random L/P generation

There are two components to the random L/P generation algorithm: In the first part, three fuel batches were selected from three different categories defined by the unique number of gadolinium rods without any duplicates shown in Table 3 and Fig. 4. Then, from the remaining batches, two or three more batches were selected at random.

This process generates approximately 66,000 data representing L/Ps made up of varying quantities of gadolinium rods in each fuel assembly for the consideration of power distribution flattening. In the second section, approximately
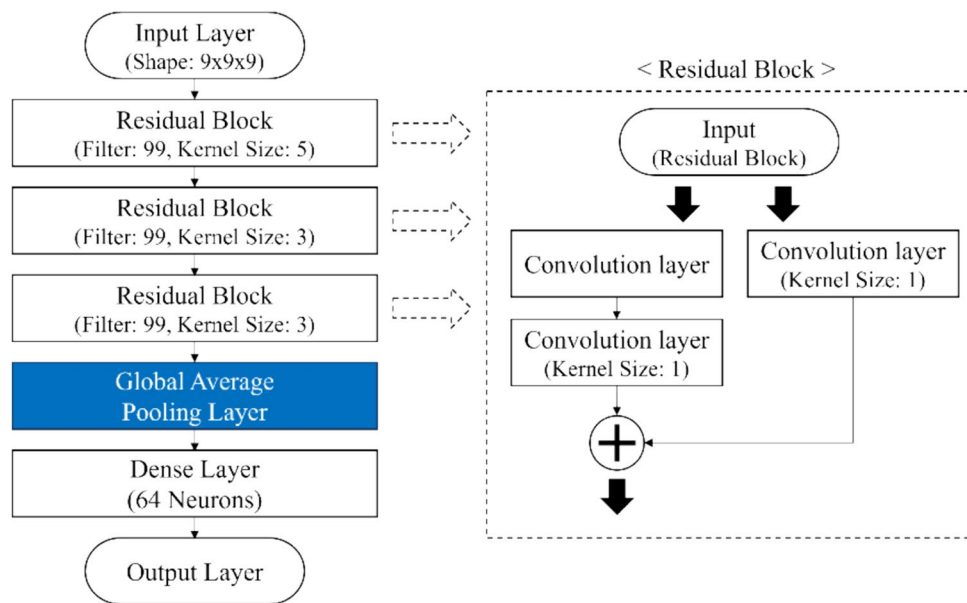
Fig. 10 CNN model

Table 5 Core parameter prediction error

	k_{eff}	Fr (%)	ASI (%)
Mean relative error	0.05%	0.80	1.02
Max relative error	0.25%	6.25	17.31
Proportion of error ≥ 3%	None	1.57	4.60
Proportion of error ≥ 5%	None	0.10	1.81

33,000 extra data points were generated after five unique fuel batches were selected to establish the L/P.

(4) Loading pattern preprocess

The randomly generated layout results were pre-processed before being used as input data for a CNN. Since raw characters could not be directly handled by neural networks, we converted them into suitable numerical data. Each batch type characters (e.g., A01, A06, and A84) were first transformed into a composition ID (see Fig. 5).

The algorithm then referenced these IDs in the ASTRA library, based on the ENDF (evaluated nuclear data file) database, and assigned the corresponding macroscopic cross-section data. In addition, empty positions without assemblies were filled with a value of zero. As a result, the original 9 × 9 2D plans were converted into 3D shapes with dimensions of 9 × 9 × 9, where the last dimension represents the nine cross-section channels.

Target Data

(1) Data distribution and features

Four parameters were chosen for analysis: Fr, ASI, control rod position, and k_{eff}. The data distributions for these parameters are presented in Fig. 6. Each data point represents the output of a specific fuel L/P (referred to as input data) simulated with the ASTRA code.

The regulating control rod banks (R1 ~ R4) have fixed locations in the core (see Fig. 7 (left)). These control rods undergo a sequential operation of insertion (R4 → R3 → R2 → R1) or withdrawal (R1 → R2 → R3 → R4) with a 50% overlap ratio. For example, when the insertion depth value of R4 falls below 50 (indicating injection) from its fully withdrawn position (100), the next control rod (R3) begins its insertion. The same sequential operation applies to R2 and R1 as well.

Compared to the actual value range of 0–100 and −150–100, the neural network sometimes predicted values outside of this range. Therefore, we chose to limit the

Table 6 Rod position prediction error

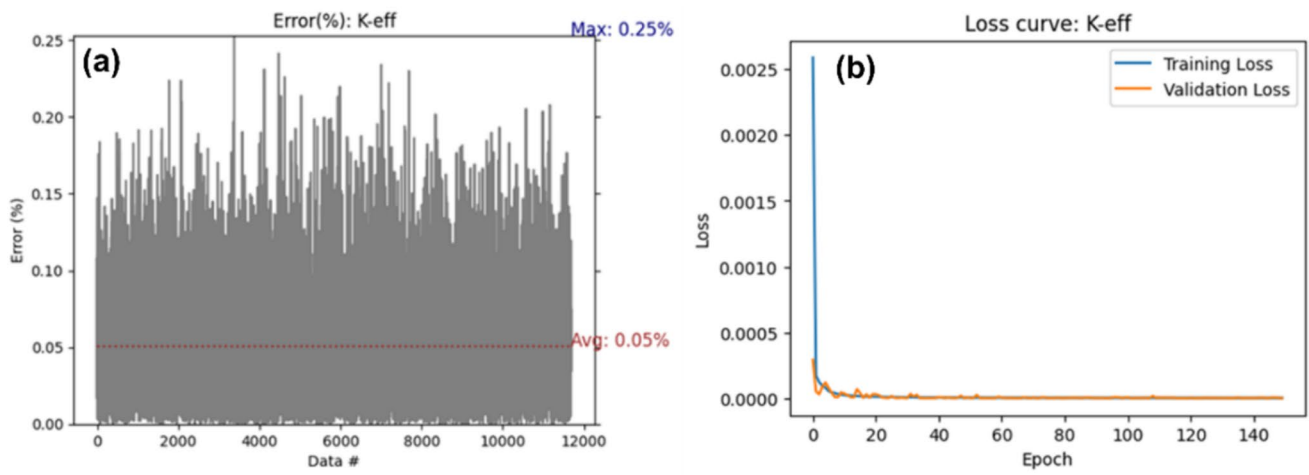|	Rod position				Encoded rod position			
	R4	R3	R2	R1	R4	R3	R2	R1
Mean absolute error	0.47	0.59	0.73	0.58	0.36	0.53	0.58	0.32
Max absolute error	11.21	7.91	6.96	6.68	11.74	7.39	7.42	6.99
Proportion of error ≥ 3 (%)	1.67	0.86	0.91	0.22	2.01	1.15	0.74	0.12
Proportion of error ≥ 5 (%)	0.30	0.04	0.02	0.01	0.39	0.05	0.02	0.01

Fig. 11 **a** Relative error and **b** training loss of multiplication factor (k_{eff})
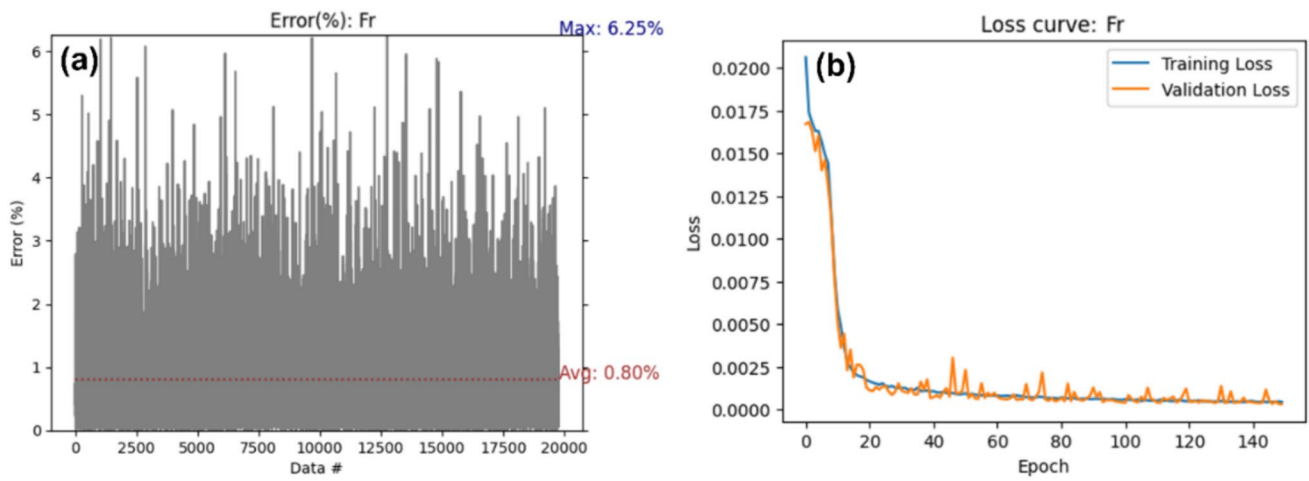

Fig. 12 **a** Relative error and **b** training loss of radial peaking factor (Fr)
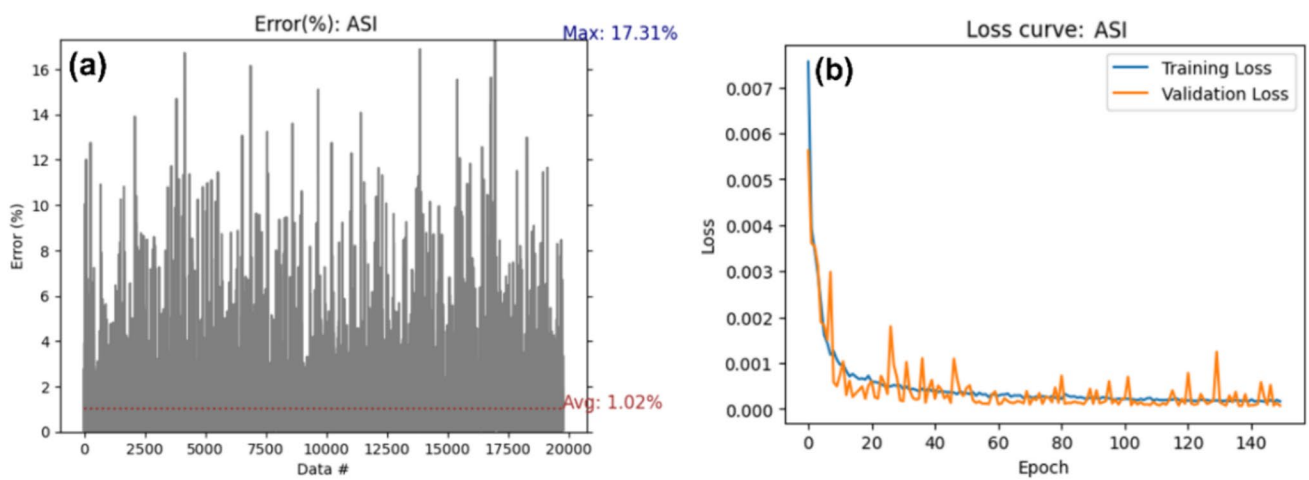

Fig. 13 **a** Relative error and **b** training loss of axial shape index (ASI)
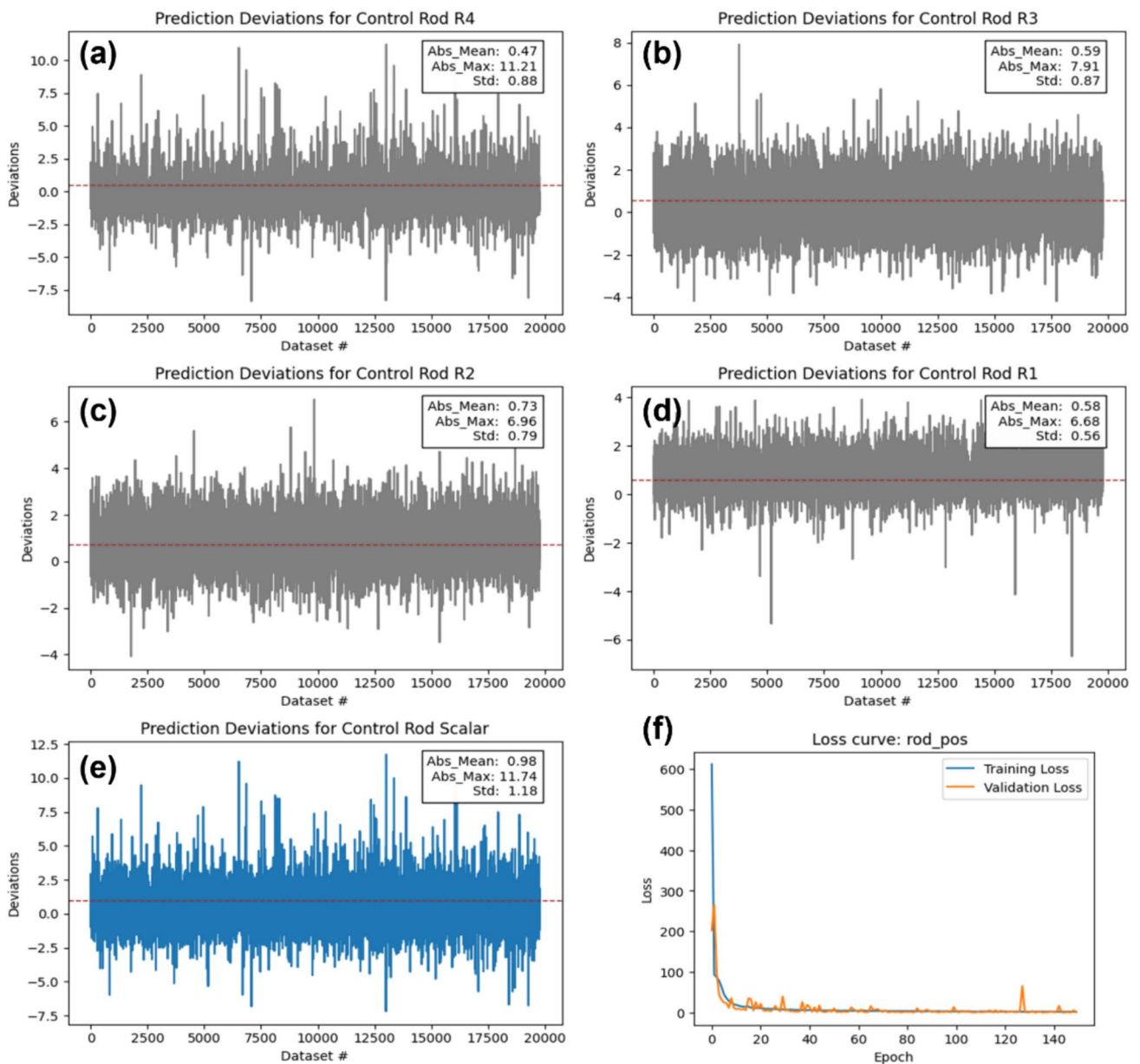
Fig. 14 **a–d** Absolute error plots of control rod position R4–R1. **e** Absolute error plot of encoded control rod position (single integer value). **f** Training loss plot

Table 7 Device specification

CPU processor	12th Gen Intel(R) Core(TM) i7-1255U 1.70 GHz
RAM	32.0 GB
OS	Windows 11 Pro (23H2 version)
System Type	64-bit operating system, ×64-based processor

range of existence using overlap between control rods and encoding and decoding.

(2) Multi-variable regression issue: control rod insertion depth

Our initial approach was to predict all four target variables (R4-R1) simultaneously using a neural network regression model. The results showed significant errors, so we initially judged that multiple-output prediction was the main cause, because it is commonly known that multiple-output regression typically yields lower performance compared to single-output task [31]. To circumvent this issue, we devised an encoding process (see Table 4 and

Table 8 Comparison of absolute errors with and without BN layer (rod position)

Rod bank	Batch normalize layer		Non-batch normalize layer	
	Absolute mean	Absolute max	Absolute mean	Absolute max
R4	1.64	30.02	0.62	14.16
R3	3.85	23.68	0.95	9.80
R2	6.05	27.93	1.05	9.80
R1	3.30	25.80	0.50	7.29
Total average	**3.71**	**26.86**	**0.78**	**10.26**

Fig. 8) based on the overlapping insertion sequence of the control rods. By adjusting rules, four of each control rod insertion depth values (R4-R1) could be encoded into one single integer value.

(3) Negative value issue: ASI, control rod insertion depth

In this study, the relative error (between predicted and labeled values) increased considerably when training data included both positive and negative values for ASI and control rod insertion depth. Even after controlling for other factors such as model structure, hyperparameters, and utilizing the same dataset, relatively large errors were observed in many cases during analysis. To address this issue, we applied a value-shifting step to both control rod insertion depth and ASI data. The initial range for encoded control rod insertion depth data was -150 to 100 (see Fig. 3). We shifted all values by adding a constant of 150, resulting in a new range of 0–250. Similarly, ASI data inherently range from -0.3 to 0.3. To avoid potential problems caused by negative values during training, we added 0.3 to all ASI values, shifting the range to 0 to 0.6 and then we subtracted 0.3 from predicted values.

Neural Network Architecture

Module

(1) Residual block module

We employed residual block as the main module of our network. It consisted of several convolution layers with different kernel sizes. Due to the limited dimensions of the core L/P ($9 \times 9 \times 9$, which are smaller than typical image sizes such as 128×128 or 256×256 pixels), we set the convolution window stride value to 1. In addition, we employed 'padding' option to maintain the width (horizontal shape) until the pooling layer of the network. When data enter the residual block, it is simultaneously processed by the first convolutional layer and the shortcut layer. The output of the first layer is then passed into the second layer. Meanwhile, the shortcut layer handles the original input independently. Following that, the outputs of the second layer and the shortcut layer are combined element-wise, pixel by pixel. Finally, this module generates the output.

(2) Candidate methods: SE module, batch normalize layer

We evaluated using the Squeeze-and-Excitation (SE) [32] module and the batch normalize layer to improve the accuracy of the network's predictions. In prior studies on different types of reactor cores, we found that these strategies were efficient in reducing errors between predictions and labeled data across multiple core parameters. The SE module is designed to multiply weights for particular channels, allowing the network to focus on important channels within each layer. These channels are located in the last dimension of the layer's output tensor. For instance, in images sized $256 \times 256 \times 3$, this dimension comprises the 3 RGB channels. The process is summarized in Fig. 9.

Batch normalization is a typical technique for mitigating overfitting, gradient vanishing, and accelerating training speed. It works by preventing internal covariate shifts from one layer to another. Internal covariate refers to the problem of learning instability caused by changes in input distribution in each layer. To address this issue, batch normalization layer adjusts the mean of inputs to zero and standardizes the variance to one, thereby facilitating stable learning across each layer of the network [33].

Neural Network Model

(1) Model architecture

The model architecture consists of three key parts: residual blocks, a global pooling layer, and a dense layer. All residual blocks used the same filter size of 99. This configuration was chosen after experimenting with various combinations of filter sizes (36, 72, 99, 108, 144), number of residual blocks (2–5), and kernel sizes (e.g., (5, 2, 2), (5, 3, 3), (7, 2, 2), (3, 3, 3)). We also explored different training parameters. The optimal performing model is presented in Fig. 10.

The residual block performs a convolutional operation to extract features from the input data. Next, a global average pooling layer converts the output from these blocks into a single-dimensional (1D) vector. This vector contains the most important features. Finally, a dense layer processes this 1D vector, and the output layer generates the model's prediction.
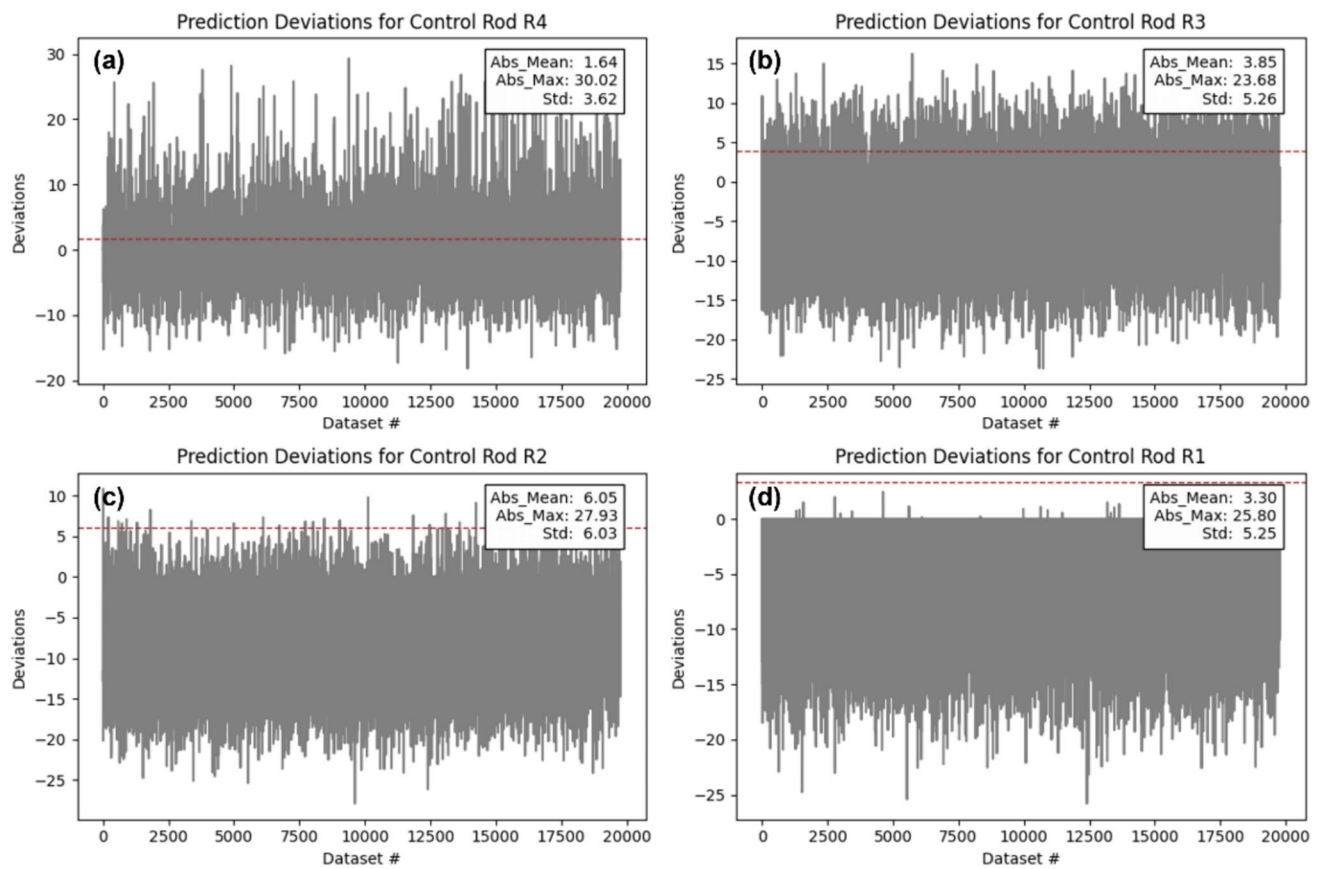
(2) Training parameters and validation method

Fig. 15 a–d Absolute error plots for the control rod insertion depths of R4–R1 (using BN layer)

Results

Prediction Error

Overall prediction errors are presented in Tables 5 and 6. 'Proportion of Error ≥ 3%' and 'Proportion of Error ≥ 5%' indicate the proportion of data points exceeding 3% and 5%

In terms of training parameters, we used the 'ReLU' activation function and the Adam optimizer to train the model for 150 epochs (learning rate = 0.001). To avoid potential overfitting issues, we investigated using Dropout layers and L1/L2 regularization. However, because these techniques resulted in higher prediction errors and slower convergence speeds, we decided to exclude them from the model.

We used the K-fold cross-validation process to evaluate both overfitting of the model and to assess the presence of dataset imbalance. During the K-fold procedure, the dataset was divided into five subsets. The model was then trained five times, each using a different subset as the testing set and the remaining data as the training set.

of errors in the entire prediction dataset. The relative errors and loss of the target core parameters are shown in Figs. 11, 12, 13 and loss of control rod positions are shown in Fig. 14.

(1) Core parameters: k_{eff}, Fr, ASI

It is noted that when predicting k_{eff}, we use a different dataset unlike the other parameters. Since k_{eff} and rod position search employ different calculation modes in the ASTRA code, we obtained a total of 50,000 data points from the ASTRA results.

(2) Rod insertion depth

Prediction errors of 'rod position' and 'encoded rod position' can be seen in Table 6 and Fig. 14. The error values in the 'Encoded Rod' column of Table 6 were obtained by calculating the errors after restoring the single encoded value into R4 ~ R1, which was used to plot Fig. 14e. Except for the encoding process, other conditions were the same during the training. It can be observed that the prediction error for the rod position is lower when it is encoded as a single value compared to when it is not encoded.
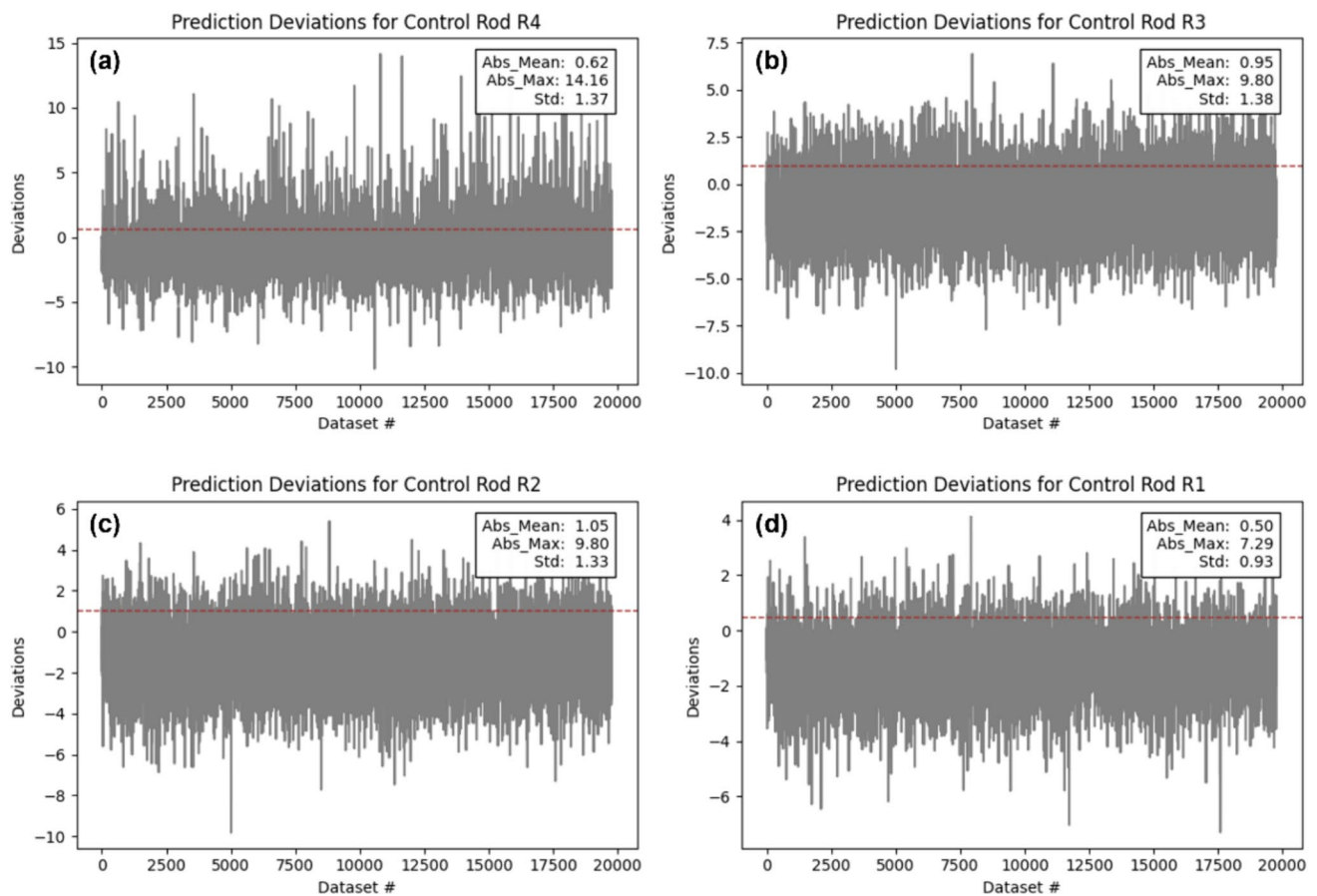
Fig. 16 **a**–**d** Absolute error plots for the control rod insertion depths of R4–R1 (non-BN layer)

Table 9 Comparison of absolute errors with and without SE block (rod position)

Rod bank	SE block		Non-SE block	
	Absolute mean	Absolute max	Absolute mean	Absolute max
R4	1.37	31.71	0.82	19.27
R3	2.33	17.30	1.21	15.00
R2	2.43	16.72	1.20	10.45
R1	1.04	10.73	0.51	7.96
Total average	**1.79**	**19.11**	**0.94**	**13.17**

Computational Time

Computational time was measured in the following environment without GPU acceleration. Device specification used in this study can be seen in Table 7.

For 16,559 cases, the computational time per L/P of BOC using the ASTRA code averaged 2.980 s, with a minimum time of 2.476 s, a maximum time of 10.853 s, and a standard deviation of 0.448. Regarding the prediction time for variables using CNN surrogate model, the time taken by the model to predict 9884 L/P cases was measured. The average time was 0.101 ms per each L/P and total time was 1.002 s.

Discussions

Comparison with Other Studies

We conducted predictions for k_{eff}, Fr, ASI and control rod position. Prior studies focused on predicting k_{eff}, Fr, boron concentration, and cycle length. However, the i-SMR operates in a flexible mode within a soluble boron-free core environment. Consequently, factors such as control rod positions and ASI are more important than other parameters. Unlike
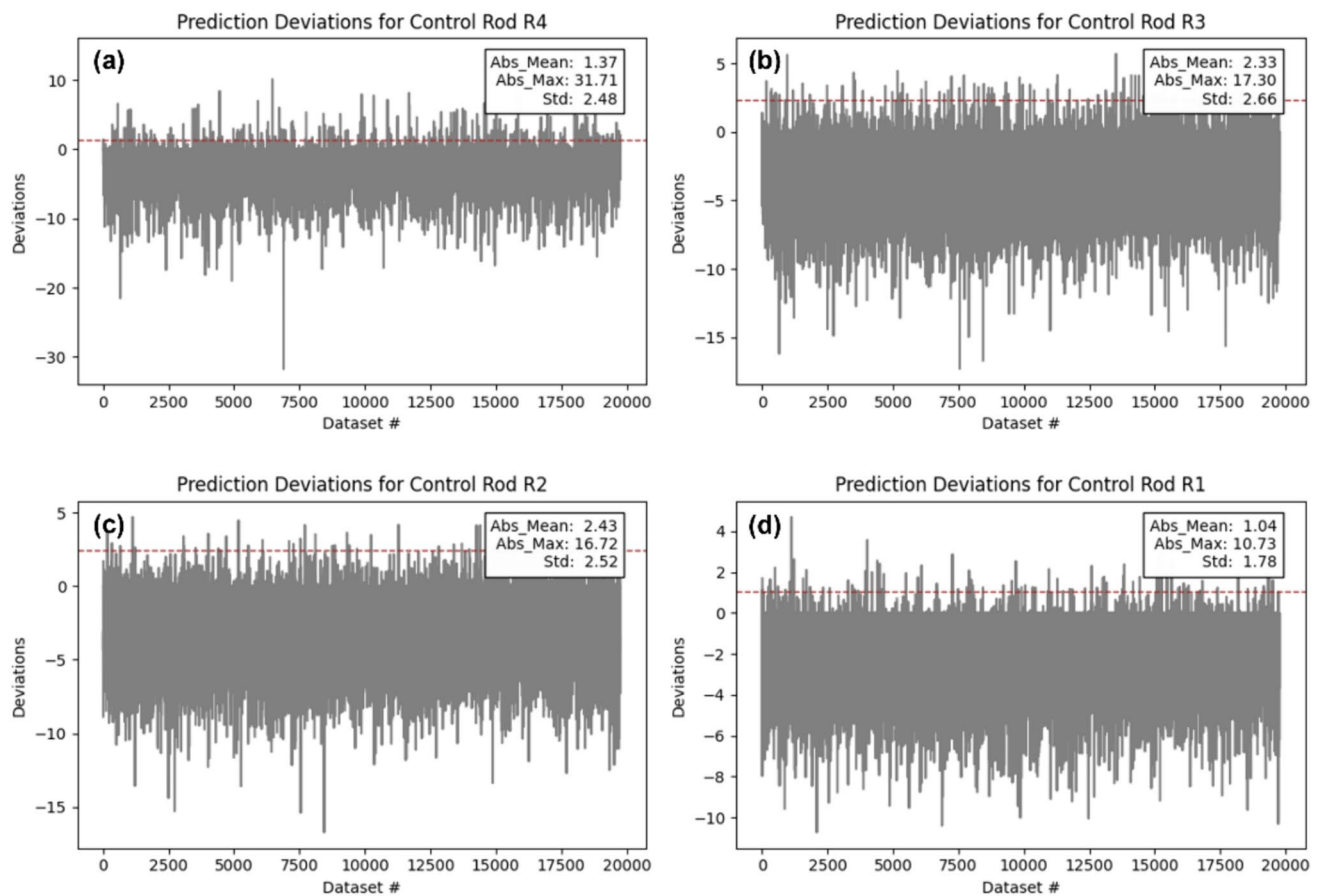
Fig. 17 **a–d** Absolute error plots for the control rod insertion depths of R4–R1 (using SE block)

prior research, our study aimed to verify whether these specific variables could be predicted effectively.

About prior AI-based surrogate models, Li et al. [20] utilized a simple fully connected neural network (FCNN) as a surrogate model, incorporating fuel burnup cycles as input. But the specifics of additional input data were not detailed in their study. This model was used to predict key parameters such as k_{eff} and power peak factor (PPF). Meanwhile, Yamamoto [21] employed FCNN to determine Fr and cycle length, achieving approximately 3–4% error for Fr and 1–2% for cycle length predictions. Wan et al. [22] used uranium enrichment in fuel assemblies and the number of burnable absorbers as inputs. Although specific computer specifications were not specified, individual evaluation times for loading models averaged 0.0005 s across 8000 test datasets. Our study distinguishes itself from prior studies by utilizing macroscopic cross-section of each fuel assembly, rather than using enrichment and number of burnable absorbers.

In addition, we identified data and model factors that pose obstacles to improving accuracy. Regarding the data, datasets containing negative values exhibited high errors. Concerning the model structure, we found that two commonly used techniques to improve CNN performance, namely SE modules and BN layers, negatively impacted prediction accuracy.

About computational speed, based on BOC, it takes 0.1 ms per individual L/P, which is over 1000 times faster than the design code that takes an average of 3 s. This speed could be further enhanced with a GPU-based system.

Limitations of the Current Model and Expectations for Practical Usage

Limitations of this study include the relatively high error in ASI prediction. Furthermore, the model only makes predictions for Beginning of Cycle (BOC) rather than the entire core cycle. Currently, the ASI prediction shows a high maximum error of 17.31%. ASI, measures the difference between power generated in the lower and upper halves of the core, divided by their sum. However, in this study, we only provided 2D data to the model. We expect that the model's performance could be improved by incorporating additional axial information. For instance, if we give control rod depth data as an additional input, the model could utilize 3D information using a combination of 2D and 1D data.
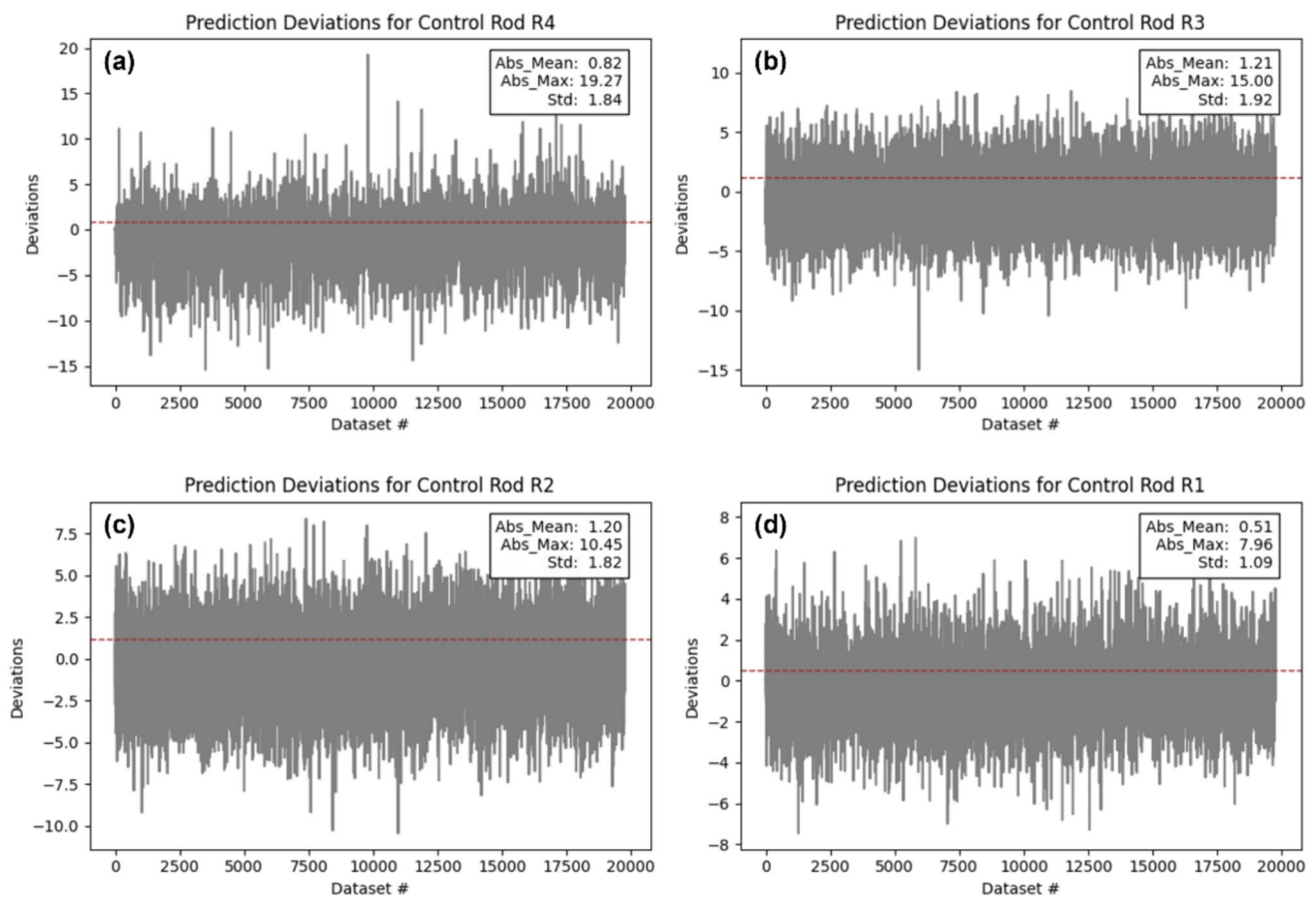
Fig. 18 a–d Absolute error plots for the control rod insertion depths of R4–R1 (non-SE block)

Regarding the core cycle, there is the issue related to the

Table 10 Comparison of relative errors caused by negative values

	Value range	Average error (%)	Maximum error (%)
Control rod (R4–R1)	−150~100	14.86	121.58
	0~250	4.22	46.73
ASI	−0.3~0.3	5.94	187.63
	0~0.6	2.22	37.86

depletion of fissile material in the core, which is related to the power of the core. As the fuel burns up, there is a problem of reactivity decrease. It is necessary to improve the model to predict the core characteristics' changes due to fuel depletion and to analyze the core throughout the entire cycle. For this, there is a method using 3D convolution layers that process continuous images like video. If we consider the core loading pattern as an image and depletion as a time dimension, it should be possible to account for fuel depletion using this 3D convolution technique.

This prediction module will be used in the core loading optimization module. As mentioned in the introduction,

flexible operation requires consideration of axial composition changes. In addition to the 5^{13} possibilities based on 1/8 symmetry core condition, we also consider composition changes along the axial direction for 24 layers. Even if we simplify these 24 layers into just 3–5 regions, the number of cases requiring analysis increases dramatically. If we use only the surrogate model without the core design code, we expect to fully benefit from the computational speed advantage during the optimization process. However, when considering the design code as the absolute standard, the surrogate model has a relatively low prediction accuracy of about 95% on average. Consequently, the optimal LP derived using the surrogate model would likely be inferior to that obtained from the optimization process using the core design code. We anticipate that by utilizing screening techniques, we can achieve a solution closer to the optimal LP while still partially benefiting from the enhanced computational speed.

Error-Inducing Factors

This section discusses the factors that influence the errors discovered during model improvement. These issues can be divided into two categories: errors caused by neural network
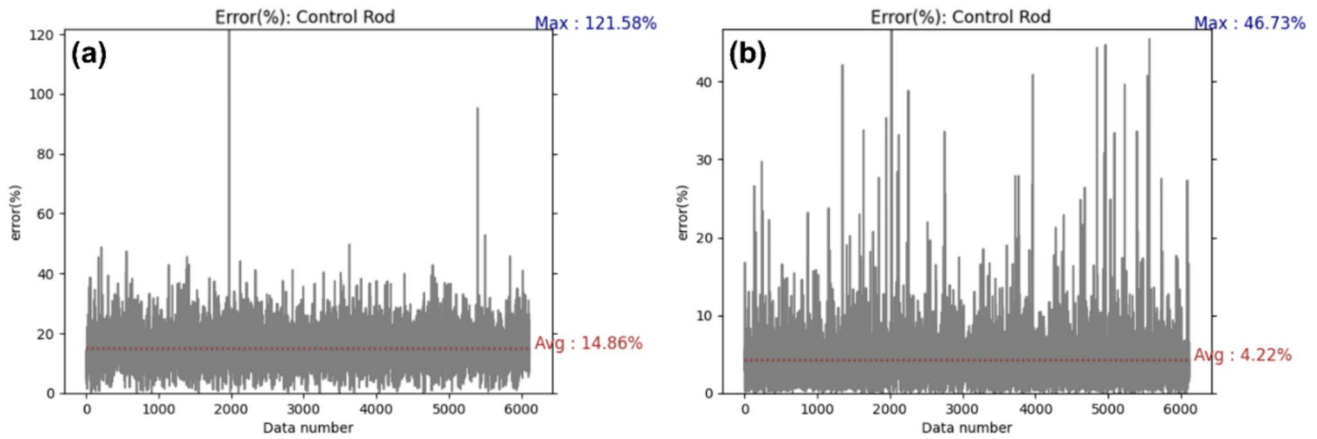
Fig. 19 Relative error of control rod insertion depth. Training the model using dataset **a** with negative values and **b** without negative values
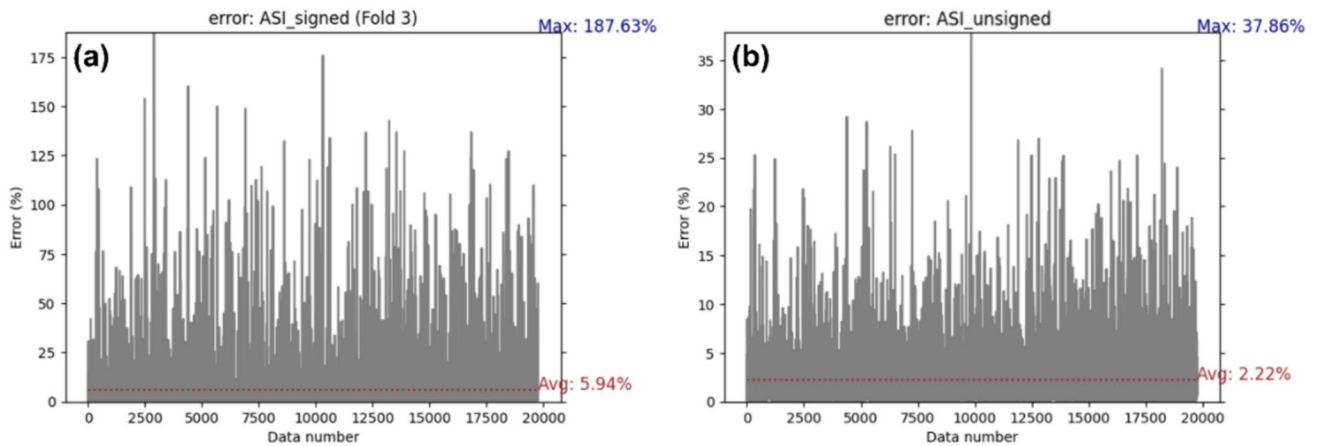

Fig. 20 Relative error of ASI. Training the model using dataset **a** with negative values and **b** without negative values

training procedures (layers, modules, activation functions, etc.) and errors caused by the training data itself. Instead of evaluating all core parameters, we focus only on the two that consistently provide the largest prediction errors: control rod position and ASI.

(1) Batch normalize layer

We compared the errors with and without using the BN layer. Table 8 shows the relative errors of the encoded single values of rod insertion depth. We observed that batch normalization resulted in larger mean and maximum errors than non-BN. In addition, the variance of the predicted values during training was also substantial. Figures 15 and 16 show plots of the absolute errors of values decoded from a single value into R4–R1 values.

(2) SE block

We compared the errors with and without the SE module, and the results are shown in Table 9. Regarding the control rod insertion depth, we discovered that utilizing the SE module resulted in higher average and maximum errors

than not using it. In addition, employing the SE module increased computation time due to the additional parameters for assigning weights to the model's layers, as well as the adjustment process. Due to the downsides of increased error and training speed, it was not used in our model. Figures 17 and 18 show the error plot based on the presence or absence of the SE block.

(3) Error caused by negative values

During our research, we found a considerable rise in errors when processing two core parameters (ASI and Rod insertion depth) with negative values. When encoding the control rod values for four banks (R4–R1) into a single value, the range of converted values is from -150 to 100. In addition, the range of ASI values lies between -0.3 and 0.3. Training the model directly with these data resulted in many occurrences with maximum relative error values exceeding 100%. To investigate this issue, we trained the model under identical settings but with different data. Table 10 provides a comprehensive comparison of relative errors caused by

negative values for both the Control Rod (R4–R1) and ASI parameters, clearly demonstrating the significant reduction in errors achieved by shifting the value into non-negative ranges. The error plot for the control rod is shown in Fig. 19, and the error plot for ASI is presented in Fig. 20.

Conclusion

This study confirmed the prediction of core characteristic parameters for a given loading pattern arrangement at BOC. While the average computation time of the ASTRA design code for one case, including control rod search, was about 3 s for each L/P, the CNN model performed predictions for 9884 cases within 1 s. If we integrated the model with the SA algorithm that contains a screening technique, the optimization process would benefit from a remarkable increase in computational speed while ensuring accuracy.

The study also predicted variables related to axial direction and flexible operation, such as control rod positions and ASI. Unlike previous studies that used enrichment values for prediction, this research demonstrated the possibility of using cross-section data as input. In addition, the study identified error-inducing factors within the CNN-based prediction methodology. Overall, this study demonstrated that the surrogate model, with its benefit of fast computation time, may be used as part of our SA algorithm's screening technique to speed up the optimization process.

Data availability Data used for the model's input was calculated using the ASTRA design code from the designed fuel composition. To maintain data confidentiality, the data supporting the findings of this study are available only in a redacted form upon request. However, the data predicted by the neural network, as well as the target data, are available from the corresponding authors upon a reasonable request.

Declarations

Conflict of interest There is no conflict of interest.

References

1. J.J. Duderstadt, L.J. Hamilton, *Nuclear Reactor Analysis* (John Wiley & Sons, New York, 1976), pp.447–464
2. H.G. Kim, S.H. Chang, B.H. Lee, Nucl. Sci. Eng. **115**, 152 (1993)
3. J.G. Stevens, K.S. Smith, K.R. Rempe, T.J. Downar, Nucl. Sci. Eng. **121**, 67 (1995)
4. T. Park, H.G. Joo, C.H. Kim, Nucl. Sci. Eng. **162**, 134 (2009)
5. G.T. Parks, T. Geoffrey, Nucl. Sci. Eng. **124**, 178 (1996)
6. C. Lin, J. Yang, K. Lin, Z. Wang, Nucl. Sci. Eng. **129**, 61 (1998)
7. J.J. Ingremeau, M. Cordiez, EPJ Nucl. Sci. Technol. **1**, 11 (2015)
8. J. Mart, A. Klein, A. Soldatov, Nucl. Technol. **188**(1), 8–19 (2014)
9. H.G. Kim, S.H. Chang, B.H. Lee, Nucl. Sci. Eng. **113**, 70 (1993)
10. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Proc. IEEE **86**, 2278 (1998)
11. K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 2016.
12. Q. Zhang, "A deep learning model for solving the eigenvalue of the diffusion problem of 2-D reactor core", *Proceedings of the Reactor Physics Asia 2019 (RPHA19)* Osaka, Japan Dec. 2–3 2019.
13. Q. Zhang, J. Zhang, L. Liang, Z. Li, T. Zhang, EPJ Web Conf. **247**, 8 (2021)
14. T. Park, K. B. Park, B. H. Ha, AI를 활용한 원자로 노심 최적 설계의 사례 (in Korean). *Proceedings of the Korea Society for Industrial Systems Conference*, Busan, June. 2–3, 2023.
15. M. Abdel-Basset, L. Abdel-Fatah, A. K. Sangaiah, *Intelligent Data-Centric Systems*, Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications, 185–231 (2018)
16. J.L. Francois, J.J. Ortiz-Servin, C. Martin-del-Campo, A. Castillo, J. Esquivel-Estrada, Ann. Nucl. Energy **51**, 189–195 (2013)
17. R. Hays, P. Turinsky, Prog. Nucl. Energy **53**(6), 600–606 (2011)
18. A. Norouzi, M. Aghaie, A.R. Mohamadi Fard, A. Zolfaghari, A. Minuchehr, Ann. Nucl. Energy **60**, 308–315 (2013)
19. H. Liao, Y. Hu, Q. Li, Y. Yu, S. Huang, F. Chen, H. Xiang, Ann. Nucl. Energy **171**, 109008 (2022)
20. Z. Li, J. Huang, J. Wang, M. Ding, Ann. Nucl. Energy **165**, 108685 (2022)
21. A. Yamamoto, Nucl. Technol. **144**(1), 63–75 (2003)
22. C. Wan, K. Lei, Y. Li, Ann. Nucl. Energy **171**, 109028 (2022)
23. T. Park, C.H. Kim, H.C. Lee, H.K. Joo, *Screening Technique for Loading Pattern Optimization by Simulated Annealing* (ANS Winter Meeting and Nuclear Technology Expo, Washington, 2005), pp.13–17
24. T. Park, H. C. Lee, H. K. Joo, C. H. Kim, Improvement of Screening Efficiency in Loading Pattern Optimization by Simulated Annealing, *ANS Annual Meeting*, Boston, USA, Jun.24–28, 2007.
25. J.D. Jenkins, Z. Zhou, R. Ponciroli, R.B. Vilim, F. Ganda, F. Sisternes, A. Botterud, Appl. Energy **222**, 872–884 (2018)
26. H.O. Kang, B.J. Lee, S.G. Lim, Nucl. Eng. Des. **419**, 112966 (2024)
27. K. B. Park, B. H. Ha, B. M. Ahn, T. Park, S. K. Zee, S. Y. Oh, Fuel Loading Pattern Optimization of a Load-Follow Operating SMR using A-Genre_LP. *Transactions of the Korean Nuclear Society Autumn Meeting*, Changwon, Korea, October 20–21, 2022.
28. H.J. An, J.H. Park, C.H. Song, J.I. Lee, Y.H. Kim, S.J. Kim, Nucl. Eng. Technol. **56**, 949 (2024)
29. J. S. Kim, B. H. Cho, S. G. Hong, Applicability Evaluation of Enriched Gadolinium as a Burnable Absorber in Assembly Level for Boron-Free i-SMR core. *Transactions of the Korean Nuclear Society Spring Meeting*, Jeju, Korea, May 19–20, 2022.
30. J.S. Kim, T.S. Jung, J.I. Yoon, Nucl. Eng. Technol. **56**(8), 3144–3154 (2024).
31. H. Borchani, G. Varando, C. Bielza, P. Larranaga, Wiley Interdisciplin. Rev.: Data Min. Knowl. Discov. **5**, 216 (2015)
32. J. Hu, L. Shen, S. Gang, Squeeze-and-excitation networks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* Salt Lake City, Utah. 18–22 June, 2018.
33. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of*

the 32nd International Conference on Machine Learning, PMLR 37:448–456, 2015.