Refined Deep Learning-Based Time-Series Surrogate Model for the BRI2-CRIEPI Zone Fire Model with Validation against Full-Scale Test Data

Junghoon Ji[a], and Motomu Suzuki[b]
[a] Central Research Institute of Electric Power Industry, Chiba, Japan, junghoon@criepi.denken.or.jp
[b] Central Research Institute of Electric Power Industry, Kanagawa, Japan, s-motomu@criepi.denken.or.jp

Abstract: This study presents a refined deep learning–based time-series surrogate model for the BRI2-CRIEPI zone fire model using a Long Short-Term Memory (LSTM) architecture. The surrogate model was trained on a large dataset of simulations generated by the BRI2-CRIEPI model and was designed to predict the temporal evolution of key fire behavior parameters from static initial conditions. Twenty influential input parameters, representing compartment geometry, fire source characteristics, ventilation conditions, wall thermal properties including wall density, and initial conditions, were identified and used for model training. The proposed surrogate model consists of a three-layer LSTM network with 256 hidden units per layer and predicts six output variables characterizing fire behavior, including heat release rate (HRR), smoke layer temperature, smoke layer height (upper layer thickness), compartment pressure, and oxygen concentration. Notably, the model can forecast the complete time histories of these outputs without requiring time-stepped input data. To enhance training stability and convergence, a teacher forcing strategy was employed. Using a dataset of 100,000 simulated fire scenarios, the LSTM surrogate model achieved high predictive accuracy on unseen cases, with an average normalized mean absolute error (nMAE) of approximately 0.029 across all output parameters, improved from 0.042 in the previous configuration. In addition, the surrogate model was compared with full-scale experimental HRR data from mechanically ventilated single-compartment tests (PRS-SI-D1 and PRS-SI-D2) in the OECD/NEA PRISME project. Although the early growth and burning duration showed discrepancies due to unlearned variations in fire growth rate and fuel amount, the surrogate reproduced the quasi-steady HRR levels reasonably well.

1. INTRODUCTION

Fire Probabilistic Risk Assessment (Fire PRA) for nuclear power facilities relies on quantitative scenario analysis grounded in validated fire models. In practice, Fire PRA requires evaluating a large number of potential fire scenarios, which amplifies the importance of computational efficiency in fire progression calculations [1]. The characteristics of each widely used fire model are presented in Table 1. Spreadsheet-type models such as Fire Dynamics Tools (FDTs [2]) are computationally inexpensive but can be limited in reproducing time-dependent behavior under mechanical ventilation conditions that are common in nuclear facilities. Zone models such as CFAST [3] provide rapid calculations with a moderate level of fidelity, but they operate on volume-averaged quantities and typically cannot provide spatial distributions; furthermore, they may be constrained in representing complex compartment geometries. By contrast, CFD-based field models such as FDS [4] can capture three-dimensional transient fire dynamics, but their high computational cost makes routine use in Fire PRA impractical at present.

Recent advances in deep learning have enabled surrogate modeling of computationally expensive simulations, offering the potential to preserve predictive performance while dramatically lowering computation time. In fire engineering, early studies [5, 6] have explored machine learning for surrogate prediction of fire behavior and for accelerating CFD-like outputs using neural networks, yet sustained follow-up progress toward practical adoption remains limited. Motivated by these developments, our

long-term goal is to construct a deep learning surrogate for CFD-based field fire simulations to support Fire PRA workflows. However, in the early stage of this research, the primary need is to establish and validate the surrogate-modeling workflow itself—i.e., to acquire the necessary know-how in data preparation, model design, training stabilization, and evaluation. Therefore, as a staged approach, we first develop and validate the surrogate modeling methodology using the computationally efficient zone model BRI2-CRIEPI, which enables large-scale data generation, before extending the framework to CFD-based field models.

BRI2-CRIEPI is a two-layer zone fire model derived from the BRI2002 [7], lineage and enhanced for nuclear-facility Fire PRA applications. In particular, BRI2-CRIEPI includes mechanisms tailored to mechanical ventilation conditions, such as a fan-curve model representing the relationship between flow rate and pressure difference across fans, enabling more realistic prediction of pressure–flow interactions in ventilated compartments. The model can also incorporate a thermal feedback treatment to improve HRR prediction by accounting for incident heat flux to the fire source from flames and hot surroundings (e.g., smoke, walls, ceiling). In typical use, zone models represent the compartment by dividing it into an upper hot smoke layer and a lower cooler air layer. Accordingly, BRI2-CRIEPI provides time-varying predictions of layer-averaged temperatures, layer interface height (or upper-layer thickness), oxygen concentrations in each layer, and compartment pressure. These quantities are directly relevant for Fire PRA metrics such as damage times, available safe egress or suppression times, and scenario severity indicators.

This paper develops an LSTM-based time-series surrogate for BRI2-CRIEPI that predicts full trajectories of six fire behavior variables from static initial conditions, and demonstrates its performance using 100,000 automatically generated simulations.

Table 1: Characteristics of Each Fire Model [8]

Fire model	Governing equations	Uncertainty & bias [1]	Typical computation time	Key limitations
Spreadsheet model	Empirical correlations	±32% / 1.29	< 1 s	• Limited capability to predict time-dependent behavior • Reduced prediction accuracy under mechanically ventilated conditions
Zone model	Conservation equations (excluding momentum)	±23% / 1.13	A few seconds	• Unable to predict spatial distributions • Limitations in representing analyzable compartment geometries
Field model	Navier-Stokes equations	±20% / 1.14	Several days	• High computational cost

2. PROBLEM FORMULATION AND DATA GENERATION

2.1. Surrogate modeling target and outputs

The goal is to emulate BRI2-CRIEPI transient outputs using a neural surrogate:

$$\mathbf{y}_{1:T} = \mathcal{F}_\theta(\mathbf{x}) \qquad (1)$$

where $\mathbf{x}$ is a vector of static scenario parameters (geometry, ventilation, fire source, wall properties, initial conditions), $\mathcal{F}_\theta$ denotes the neural-network surrogate with trainable parameters θ (weights and

[1] These values represent the uncertainty and bias for smoke-layer temperature under mechanically ventilated conditions [9].

biases) and $\mathbf{y}_{1:T}$ is the time series of output variables for T time steps. The surrogate predicts the following six outputs at each time step:
1) HRR, 2) upper-layer temperature, 3) upper-layer thickness (smoke layer depth), 4) compartment pressure, 5) upper-layer oxygen concentration, 6) lower-layer oxygen concentration.

The time series is sampled at 10 s intervals over a 3600 s window (i.e., 361 points including t = 0).

2.2. Input parameters and ranges

In our previous technical report [8], we developed an LSTM-based time-series surrogate model for the BRI2-CRIEPI zone fire model using a specified set of input parameters and ranges. In the present study, we refine that configuration by adding wall density as an additional input parameter. This refinement is motivated by its high sensitivity to key zone-model outputs and is reflected in Table 2, which summarizes the input parameters and their ranges used in this work.

In the BRI2-CRIEPI simulation setup used for training data generation, fan characteristics were fixed to representative values (maximum static pressure 300 Pa for supply and 500 Pa for exhaust), and duct geometry and opening-height constraints were applied (e.g., upper-edge height no closer than 0.2 m from the ceiling, minimum upper-edge height 0.4 m above the floor). A representative fuel specification (e.g., dodecane with specified heat of combustion and other properties) and a fire pan area (e.g., 0.49 m²) were used in the data-generation configuration described in Ref.[8].

Table 2: Input Parameters of the Time-Series Fire Surrogate Model

Parameter		Range / Setting
Compartment geometry	Width	2.0 ~ 50.0 m
	Depth	2.0 ~ 50.0 m
	Height, h	3.0 ~ 30.0 m
Fire source	Fire growth rate, α	0.178 kW/s² (fixed)
	Maximum heat release rate, $\dot{Q}_{\max}$	50.0 ~ 1500.0 kW
	Heat release rate, $\dot{Q}(t)$	$\min(\alpha t^2, \dot{Q}_{\max})$
	Fuel mass	20 kg (fixed)
	Fire source height	0.3 m (fixed)
	Soot yield	0.01 ~ 0.30
	Carbon residue fraction	0.0 ~ 0.30
	Moisture content	0.0 ~ 0.30
Mechanical ventilation	Air change rate	1 ~ 15 h⁻¹
	Exhaust opening upper-edge height	0.4 m ~ $(h-0.2)$m
	Supply opening upper-edge height	0.4 m ~ $(h-0.2)$m
Wall thermal properties	Emissivity	0.1 ~ 1.0
	Specific heat	0.5 ~ 1.4 kJ/kg/K
	Thermal conductivity	$1.0\times10^{-5}, 1.0\times10^{-4}$, $1.0\times10^{-3}, 1.0\times10^{-2}$ W/m/K
	Density	500 ~ 7000 kg/m³
Initial conditions	Initial indoor/outdoor temperature	10 ~ 40 °C

2.3. Automated simulation campaign and dataset composition

A dedicated automation code was created to generate BRI2-CRIEPI input files, execute simulations, and collect outputs. Input parameters were randomly sampled (uniformly) within the prescribed ranges to generate a large and diverse dataset. A total of 100,000 simulation cases were produced, and each case contributed one paired dataset of $\mathbf{x}$ and $\mathbf{y}_{1:T}$.

The dataset was split into training/validation/test sets with a typical partitioning of 70% / 20% / 10%.

2.4. Preprocessing

To ensure stable learning across variables with different physical scales, input and output variables were standardized (e.g., using a StandardScaler approach) based on the training set statistics. During inference and evaluation, predictions were inverse-transformed to physical units prior to computing error metrics.
The raw simulation outputs were downsampled to a 10 s interval; missing values were handled by forward/backward fill strategies in the data loading pipeline.

3. TIME-SERIES SURROGATE MODEL

3.1. RNN/LSTM background

Recurrent neural networks (RNNs) are designed to model sequential dependencies by propagating a hidden state through time. However, conventional RNNs often suffer from vanishing gradients when learning long-term dependencies. LSTM networks address this limitation by introducing gated memory cells that regulate information flow via input, forget, and output gates, enabling more robust learning of long-range temporal relationships.

3.2. Model architecture

We implement a three-layer LSTM network with 256 hidden units per layer as the core sequence model. The surrogate is formulated as an autoregressive sequence predictor that generates $\hat{\mathbf{y}}_t$ step-by-step.
A key design choice is that the model predicts the full time series using only static initial conditions. To achieve this while preserving sequential structure, the input at each time step concatenates:

- the static scenario vector $\mathbf{x}$ (time-invariant features), and
- the previous-step output t $\mathbf{y}_{t-1}$(ground truth during teacher forcing; predicted value during free-running inference).

The LSTM input dimension is 26, corresponding to 20 static input parameters + 6 previous outputs, and the output dimension is 6.
Dropout is applied within the LSTM layers to reduce overfitting; the best-performing configuration for the final evaluation used dropout = 0.2.

3.3. Teacher Forcing strategy

Autoregressive sequence prediction can become unstable when early prediction errors propagate forward. To mitigate this, training uses a Teacher Forcing strategy: with a given probability, the ground-truth y_{t-1} is used as the next-step input instead of the model's own prediction $\hat{y}_{t-1}$. In our training setup, the Teacher Forcing ratio follows a linear schedule decreasing from 0.8 at the start of training to 0.0 by the end, thereby transitioning from guided learning to fully self-conditioned prediction. Alternative schedules (exponential, inverse-sigmoid) are also supported in the codebase.

3.4. Training configuration and optimization

Training minimizes a mean squared error loss over all time steps and outputs:

$$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T} \| \hat{\mathbf{y}}_t - \mathbf{y}_t \|_2^2 \qquad (2)$$

The optimization uses the Adam optimizer with an initial learning rate of 0.001 and L2 regularization (weight decay) of 10^{-5}. Training runs up to 100 epochs with:

- Early stopping if validation loss does not improve for 10 epochs (minimum improvement threshold 0.001),
- ReduceLROnPlateau learning-rate scheduling: if validation loss stagnates for 5 epochs, reduce learning rate by a factor of 0.5,
- Gradient clipping with max norm 1.0,
- Batch size 25.

The training environment uses PyTorch 2.4.0, CUDA Toolkit 12.1.0, cuDNN 9.2.1, and a workstation equipped with an Intel Xeon Silver 4416+ CPU, 128 GB RAM, and an NVIDIA RTX A5000 GPU. Total training time for the 100,000-case dataset was approximately one week.

3.5. Evaluation metrics

We evaluate predictive performance using MSE and MAE computed on inverse-transformed outputs in physical units. While MSE is sensitive to large errors, MAE provides a more interpretable measure in original units.

To compare errors across outputs with different scales, we use a normalized MAE (nMAE):

$$\mathrm{nMAE} = \frac{\mathrm{MAE}}{\sigma_{\mathrm{train}}} \qquad (3)$$

where $\sigma_{\mathrm{train}}$is the standard deviation of the corresponding output in the training set. A single summary indicator, average nMAE, is computed by averaging nMAE equally over the six outputs.

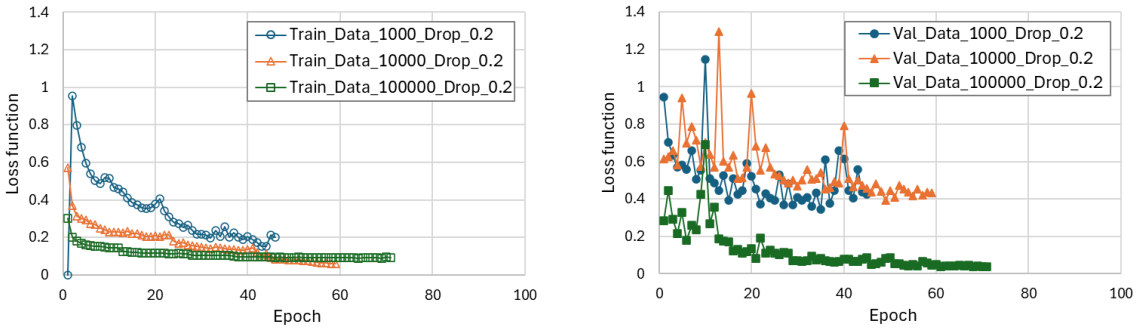
4. RESULTS

4.1. Training stability: effects of dataset size and dropout

To evaluate training robustness under the refined input configuration (i.e., including wall density), we conducted multiple training runs across different dataset sizes and dropout values. The refined model increases the static input dimensionality (20 static parameters, resulting in an LSTM input dimension of 26 when combined with six previous outputs), which can make optimization more sensitive to data availability and regularization choices.

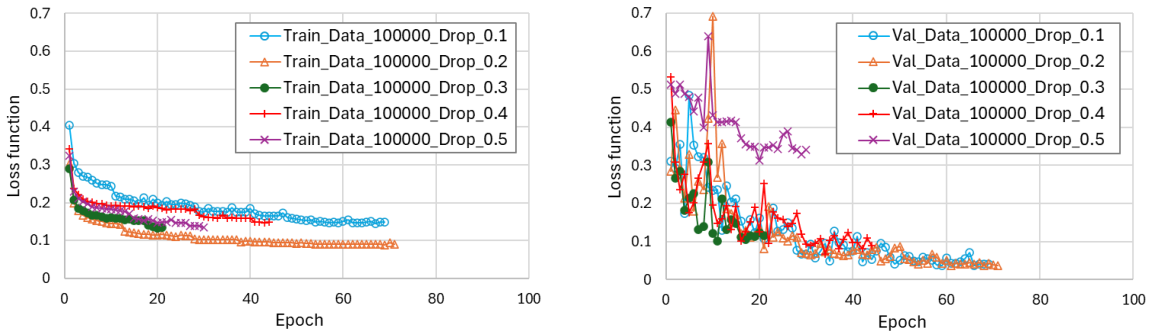
Figure 1 compares learning curves for dataset sizes of 1,000, 10,000, and 100,000 cases under a fixed dropout of 0.2. The validation loss shows substantial fluctuations for 1,000 and 10,000 cases, including intermittent spikes, and it does not exhibit a clearly stable downward trend. In contrast, with 100,000 cases the validation loss decreases more smoothly and reaches a markedly lower level, indicating improved convergence and generalization under the refined input set. These results suggest that sufficiently large training data are required to stably learn the expanded input–output mapping when wall density is included.

Figure 2 summarizes the impact of dropout (0.1 ~ 0.5) using 100,000 training cases. Overall, moderate dropout improves training stability and validation behavior, while overly large dropout tends to hinder convergence. Based on the combined observations from Figures 1 and 2, we selected 100,000 training cases with dropout = 0.2 as the final configuration for subsequent evaluations.



(a) Train data (b) Validation data

Figure 1: Effects of Dataset Size for Training Stability



(a) Train data (b) Validation data

Figure 2: Effects of Dropout for Training Stability

4.2. Predictive accuracy on unseen test cases

Figure 3 presents representative examples of the predicted temporal variation of heat release rate (HRR) for randomly selected unseen test cases. Overall, the LSTM surrogate closely follows the target HRR trajectories, reproducing the main phases of compartment-fire evolution, including the initial rise, the quasi-steady burning period, and the decay as fuel becomes depleted.

A notable discrepancy is observed for HRR predictions near the initial time: in some cases, the surrogate outputs a non-zero HRR at t = 0 s, which violates the physical expectation that HRR should start from zero at ignition. This issue suggesting that the data-driven model does not automatically internalize the constraint "HRR(t = 0 s) = 0 kW" without additional enforcement. A practical mitigation is to impose a code-level constraint so that the output at t = 0 s is forced to zero for variables that must satisfy such initial conditions. In future work, this can be strengthened by incorporating physics-informed penalties in the loss function or by designing the output parameterization so that the HRR trajectory is guaranteed to satisfy the initial condition by construction (e.g., predicting increments from zero).

Table 3 summarizes quantitative test performance for the selected configuration (dataset size = 100,000; dropout = 0.2) and compares the baseline results reported in Ref [8] with those of the refined model developed in this study. The metrics (MSE, MAE, and nMAE) are computed on inverse-transformed outputs in physical units, and nMAE is used to enable comparisons across variables with different scales. Overall, the refined model achieves lower nMAE across all six outputs, with the largest improvement observed for upper-layer temperature (0.083 → 0.039). Improvements are also obtained for HRR (0.023 → 0.019), upper-layer thickness (0.049 → 0.044), pressure (0.017 → 0.014), upper-layer O2 (0.041 → 0.027), and lower-layer O2 (0.040 → 0.033). Averaged equally over the six outputs, the refined model attains an average nMAE of approximately 0.029, improved from 0.042 in the previous model.

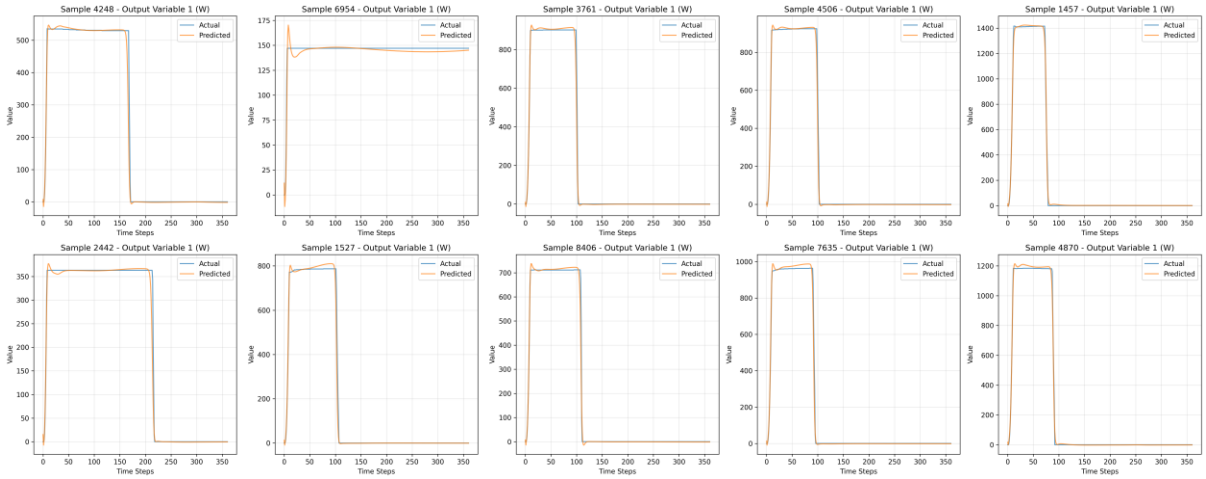


Figure 3: Predicted Temporal Variation of Heat Release Rate

Table 3: Test Performance by Output Variable (Dataset Size = 100,000, Dropout = 0.2)

Output	Previous model [8]			Refined model		
	MSE	MAE	nMAE	MSE	MAE	nMAE
HRR	3.54×10^{2}	7.3×10^{0}	2.3×10^{-2}	5.20×10^{2}	6.15×10^{0}	1.92×10^{-2}
Upper-layer temperature	3.17×10^{1}	1.97×10^{0}	8.3×10^{-2}	6.39×10^{0}	9.25×10^{-1}	3.90×10^{-2}
Upper-layer thickness	9.6×10^{-1}	2.9×10^{-1}	4.9×10^{-2}	8.3×10^{-1}	2.59×10^{-1}	4.42×10^{-2}
Pressure	1.05×10^{5}	6.93×10^{0}	1.7×10^{-2}	1.58×10^{5}	5.77×10^{0}	1.39×10^{-2}
Upper-layer O_2	1.58×10^{-6}	5.3×10^{-4}	4.1×10^{-2}	1.27×10^{-6}	3.55×10^{-4}	2.71×10^{-2}
Lower-layer O_2	2.26×10^{-6}	1.8×10^{-4}	4×10^{-2}	1.90×10^{-6}	1.50×10^{-4}	3.29×10^{-2}

5. CASE STUDY: Application to Large-Scale Mechanical Ventilation Fire Test

To examine whether the proposed time-series surrogate is applicable beyond purely simulation-generated scenarios, we validate the model against large-scale experimental data from the OECD/NEA PRISME project [10] under mechanical ventilation conditions. Two mechanically ventilated single-compartment tests are considered: PRS-SI-D1 and PRS-SI-D2. The tests were conducted in a 5.0 m × 6.0 m × 4.0 m compartment using a TPH pool fire (fire area = 0.4 m²). The primary difference between the two tests is the ventilation rate: 4.7 h^{-1} (560 m^3/h) for PRS-SI-D1 and 8.4 h^{-1} (1020 m^3/h) for PRS-SI-D2. The upper-edge heights of both the supply and exhaust openings are 3.8 m above the floor, and the opening dimensions are 0.2 m × 0.2 m.

Figure 4 compares the experimental HRR time histories with the corresponding predictions from the original zone model (BRI2-CRIEPI) and from the proposed LSTM surrogate. For both tests, the experimental HRR exhibits a rapid initial rise followed by a quasi-steady burning regime and a subsequent decay phase, which is a typical pattern of ventilation-controlled fires.

While the surrogate captures the overall shape of the HRR evolution, it systematically overpredicts HRR in the quasi-steady regime. In PRS-SI-D1, the surrogate prediction remains noticeably higher than the experimental plateau, whereas BRI2-CRIEPI shows closer agreement over a longer duration. In PRS-SI-D2, both predictions capture the increased HRR level associated with the higher ventilation rate, though discrepancies emerge in the decay phase, where the experimental HRR decreases earlier than predicted, indicating differences in burnout timing.

These discrepancies stem from limitations in the training dataset. The fire growth rate α and fuel mass were treated as fixed values, restricting the model's ability to capture variations in early HRR growth and burnout duration. Furthermore, the training data combines relatively large compartment sizes with moderate HRR values, leaving ventilation-controlled fire conditions—such as those in the PRISME experiments—underrepresented. Consequently, the surrogate does not sufficiently learn the transition to ventilation-limited burning, leading to systematic HRR overestimation under weak ventilation conditions.

Overall, while the surrogate reproduces the general temporal trends of compartment fires, its performance is limited in regimes underrepresented in the training data. Improving generalization will require expanding the dataset to cover a broader range of fire scenarios, including ventilation-limited burning, with variations in fire growth rate, and fuel load.

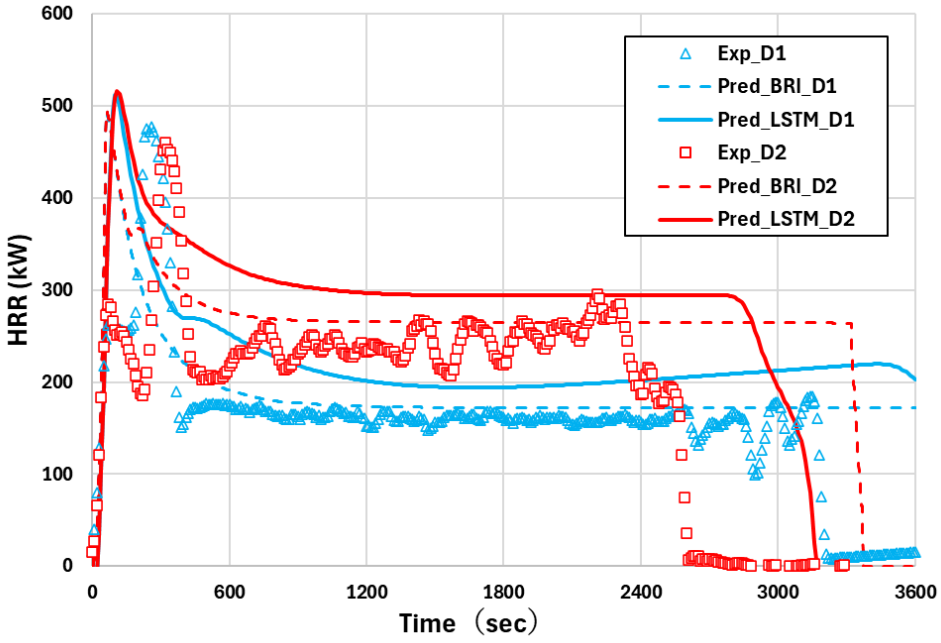


Figure 4: Comparison of Predicted HRR with Experimental Data

6. CONCLUSIONS

This study developed a deep learning–based time-series surrogate model for the BRI2-CRIEPI zone fire model using an LSTM architecture. By training on 100,000 automatically generated simulation cases, the surrogate learns to predict the complete time histories of six key fire behavior variables (HRR, smoke-layer temperature, smoke-layer thickness, pressure, and oxygen concentrations) from static scenario inputs.

The proposed model employs a three-layer LSTM with dropout regularization and a Teacher Forcing strategy, combined with standard optimization techniques to stabilize training. For unseen simulation-based test cases, the model achieved strong predictive performance, with an average normalized mean absolute error (nMAE) of approximately 0.029, showing a clear improvement over the previous configuration (0.042).

Validation against full-scale experimental data from the OECD/NEA PRISME project demonstrates that the surrogate can reproduce the overall temporal behavior of compartment fires, including key trends associated with mechanical ventilation. However, the model exhibits systematic overprediction

of HRR in the quasi-steady regime, particularly under ventilation-controlled conditions. In addition, discrepancies in the early growth and burnout phases are observed.

These limitations are attributed primarily to the characteristics of the training data. In particular, the use of fixed fire growth rate and fuel mass restricts the model's ability to capture transient variability, while ventilation-controlled fire conditions are underrepresented in the dataset due to the combination of large compartment sizes and moderate HRR levels. As a result, the surrogate does not sufficiently learn the transition to ventilation-limited burning.

Overall, the developed surrogate provides an efficient tool for approximating fire dynamics in Fire PRA applications, but its predictive performance remains sensitive to the coverage of the training data. Future work should focus on expanding the dataset to better represent relevant fire regimes, particularly ventilation-controlled conditions, and incorporating variations in fire growth rate, and fuel load to improve generalization.

References

[1] EPRI/U.S. NRC-RES, "*Fire PRA Methodology for Nuclear Power Facilities*", NUREG/CR-6850, Vol. 2: Detailed Methodology, Final Report, (2005).
[2] U.S. NRC., "*Fire Dynamics Tools (FDTs)*", NUREG-1805, (2004).
[3] Peacock, R. D., McGrattan, K. B., Forney, G. P., & Reneke, P. A. "*CFAST – Consolidated Model of Fire and Smoke Transport (Version 7)*", NIST Technical Note 1889, (2021).
[4] McGrattan, K. B., McDermott, R., Weinschenk, C. G., Forney, G. P., *et al*., "*Fire Dynamics Simulator Technical Reference Guide*", NIST Special Publication 1018, latest edition (FDS 6.11.0), National Institute of Standards and Technology, (2025).
[5] Hodges, J. L., "*Predicting Large Domain Multi-Physics Fire Behavior Using Artificial Neural Networks*", PhD Thesis, (2018).
[6] Hodges, J. L., *et al*., "*Compartment fire predictions using transpose convolutional neural networks*". Fire Safety Journal, 108, (2019).
[7] Tanaka, T., *et al*., "*BRI2002: Two Layer Zone Smoke Transport Model*", Fire Science and Technology, 23(1), (2004).
[8] Ji, J., and Suzuki, M., "*Development of an Innovative Fire Model Using Deep Learning Technology (Part 2)- A Time Series Surrogate Model of the Zone Model BRI2-CRIEPI*", CRIEPI Research Report NR25005, (2026). (in Japanese)
[9] U.S. NRC. Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications. NUREG-1824 Supplement 1, (2016).
[10] OECD/NEA, "*Final Report of the PRISME Project*", NEA/CSNI/R(2017)14, 2018.