Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 7-8, 2026

Optimizing MISO Architectures for Severe Accident Prediction: A Statistical Candidate Selection Approach to Mitigate Error Propagation

Seok Ho Song[a], Seunghyeok Yang[b], Taewoo Kim[c], Mi Ro Seo[c] Jeong Ik Lee[a]*

[a]*Department of Nuclear and Quantum Engineering N7-1 KAIST 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea 305-338, 1812wow@kaist.ac.kr*
[b]*FNC Technology CO., LTD, Heungdeok IT Valley Bldg. 32F, 13, Heungdeok 1-ro, Giheung-gu, Yongin-si, Gyeonggi-do, 16954, Republic of Korea*
[c]*Korea Hydro and Nuclear Power Central Research Institute, 70, Yuseong-daero 1312beon-gil, Yuseong-gu, Daejeon, Republic of Korea*
**Corresponding author: jeongiklee@kaist.ac.kr*

Keywords : Severe accident, MISO architecture, Effect size, Error propagation

1. Introduction

The progression of severe accidents in nuclear power plants (NPPs) is characterized by highly nonlinear thermal-hydraulic phenomena and complex interactions among subsystems. Accurate and timely prediction of these dynamics is essential for operator decision-making and accident mitigation. While integral system codes like the Modular Accident Analysis Program (MAAP) provide detailed simulations, their substantial computational cost often limits applicability in real-time scenarios. Consequently, data-driven surrogate models employing deep learning architectures—such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—have emerged as effective alternatives, offering rapid inference capabilities without compromising essential accuracy[1].

To address the complexity of capturing diverse physical correlations, the Multi-Input Single-Output (MISO) strategy was introduced in previous studies [2]. Unlike monolithic multi-output models that attempt to predict all variables simultaneously, the MISO framework develops a dedicated model for each target variable. It has been demonstrated that this approach significantly improves prediction accuracy by allowing each model to focus on the unique dynamics of a specific parameter, thereby mitigating the interference often observed in multi-task learning environments [2]. In the context of SAMG, where variables like steam generator water levels and containment pressure exhibit vastly different temporal scales and sensitivities to operator actions, this specialized MISO approach is particularly advantageous.

Building upon this foundational concept, a hybrid architecture framework was subsequently proposed [3], where optimal network structures were selectively assigned to target variables. However, the model selection process in [3] exhibited critical limitations in its evaluation environment. To determine the best architecture for a specific variable, performance metrics were derived from simulations where all variables were simultaneously predicted using a uniform architecture. Since the evaluation relied on a fully coupled recursive system, the measured performance of a specific model was inevitably contaminated by errors propagated from other suboptimal models within the same uniform structure. Consequently, this selection method failed to isolate the intrinsic compatibility between a specific architecture and its corresponding physical variable [3].

As illustrated in figure 1, in a recursive multi-step forecasting environment, the local performance of individual models does not necessarily translate to system-level stability. The methodological flaws in previous evaluation strategies made it difficult to define the true contribution of each architecture to the overall stability, as well-suited models could be undervalued due to noise from external error interference. To address this, a more rigorous validation strategy is required—one that isolates the performance of each model-variable pair by providing ground truth values for all auxiliary variables while only the target variable is predicted recursively. By shielding the evaluation from external error propagation, pure performance metrics can be obtained to form a reliable candidate pool for system-level optimization.

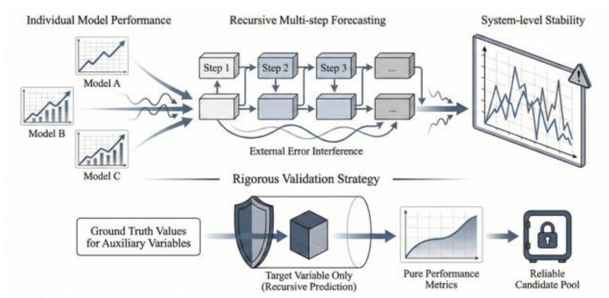

Figure 1 Mitigating Error Propagation in Recursive Forecasting: A Robust Validation Approach using Ground Truth

To overcome the limitations of deterministic selection, the MISO framework is refined in this study by shifting to a statistical candidate selection process. Recognizing that relying solely on marginal error differences in large-scale datasets can lead to the p-value fallacy [4], a rigorous effect size criterion motivated by Cohen's d [5]—specifically utilizing Glass's Δ [6]—is adopted to define a pool of practically equivalent architectures relative to the top-performing baseline [7, 8]. This approach prevents the selection of unstable models by strictly enforcing performance consistency against the best reference model. By retaining these statistically robust candidates and exploring their combinations in a fully recursive environment, a system-level architecture that minimizes systemic error propagation and maximizes global stability for safety-critical NPP applications is identified. The resulting ANN surrogate model serves as a high-speed auxiliary tool designed to assist operators by rapidly forecasting the thermal-hydraulic trend in response to specific SAMG mitigation actions.

2. Methods

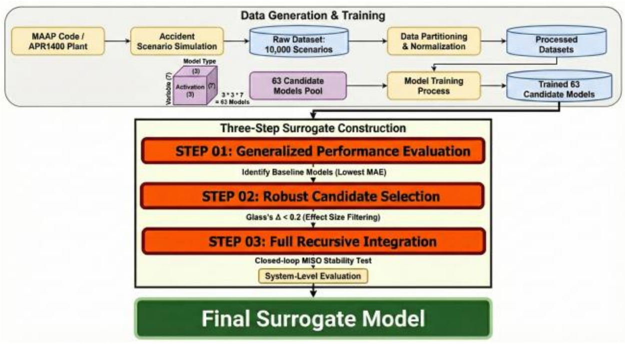

Figure 2 Overall Methods Flowchart

2.1 Accident Scenario and data generation

Table 1. Target SAMG Mitigation

#	SAMG mitigation
1	Reactor cooling system depressurization
2	Steam Generator external injection
3	Reactor cooling system external injection
4-1	Containment Spray pump activation
4-2	Emergency Containment Spray Backup System

The dataset for this study was constructed by simulating Total Loss of Feedwater (TLOFW) accidents in an APR1400 nuclear power plant using the MAAP code. A worst-case scenario was assumed where all engineering safety features are unavailable. To prevent data bias and ensure representative coverage of the accident space, the Maximin Sampling technique was applied to generate a total of 10,000 unique scenarios.

To reflect the realism of accident management, five SAMG mitigation strategies listed in Table 1 were incorporated, with each strategy set to be implemented at random intervals between 1 and 24 hours after the initiating event. This extensive temporal variability allows the AI models to sufficiently learn complex thermal-hydraulic transients, providing a robust foundation for a high-fidelity prediction framework tailored for safety-critical systems.

2.2 Training Condition

Table 2. Target Thermal-hydraulic Variable

#	Target thermal-hydraulic variable
1	Primary system pressure (PPS)
2	Cold leg temperature (Cold leg T)
3	Hot leg temperature (Hot leg T)
4	Steam generator pressure (SG P)
5	Steam generator water level (SG WL)
6	Cavity water level (CWL)
7	Containment Pressure (CTMT P)

The 10,000 generated scenarios were partitioned into a Training Set of 7,000 and a Test Set of 3,000 to ensure a robust evaluation of the models' predictive capabilities. Within the training step, a 5% subset was further isolated as a validation set to monitor for overfitting. To account for the varying physical scales of the thermal-hydraulic parameters, all input and target data were normalized using a Min-Max scaling technique. Specifically, the data were mapped to a range of [0.1, 0.9] rather than the standard [0, 1] to avoid numerical saturation and gradient vanishing issues within the recurrent framework. This study focuses on seven critical variables essential for monitoring severe accident progression, as listed in Table 2.

The training process utilized the Adam optimizer with an initial learning rate optimized for convergence stability. Mean Absolute Error (MAE) was adopted as both the loss function and the primary evaluation metric to minimize the average magnitude of prediction errors across all variables. To ensure optimal generalization, an Early Stopping mechanism was implemented with a patience of 10 epochs. This allowed the training to terminate automatically once the validation MAE ceased to improve, preventing the models from capturing idiosyncratic noise within the training data. The models were trained using a sequence length of 3-time steps to capture short-term temporal dependencies before undergoing the two-step statistical selection process.

2.3 Model Architecture and Candidate Pool

To systematically identify the optimal predictive structure for each thermal-hydraulic variable, this study employs a hierarchical Candidate Pool generation process. As illustrated in Figure 3, the model space is constructed by cross-combining three distinct neural network with three specialized output activation strategies. This combinatorial approach results in a total of 63 unique model instances (7 Variables * 3 Model types * 3 Activations) that form the initial training pool for the step 1

2.3.1 Model Types

Figure 3 details the internal architecture of the three model types used in this study. Each architecture is designed to capture specific temporal characteristics of severe accident data:

CNN (Convolutional Neural Network): As shown in the first model type block, this model utilizes a 1D-convolutional layer to extract local temporal features from the input sequence. The feature map is then normalized via batch normalization and activated by ReLU to ensure training stability before being mapped to the output by a fully connected layer.

LSTM (Long Short-Term Memory): The second block illustrates the sequence modeling approach. It employs two stacked LSTM layers to capture long-term dependencies. layer normalization and dropout are interposed between the recurrent layers to prevent overfitting and mitigate the vanishing gradient problem.

COMB (Hybrid CNN-LSTM): The third block represents a hybrid architecture. It sequentially integrates a convolutional block for spatial feature extraction and a LSTM for temporal context learning, leveraging the strengths of both preceding models to capture global accident dynamics.

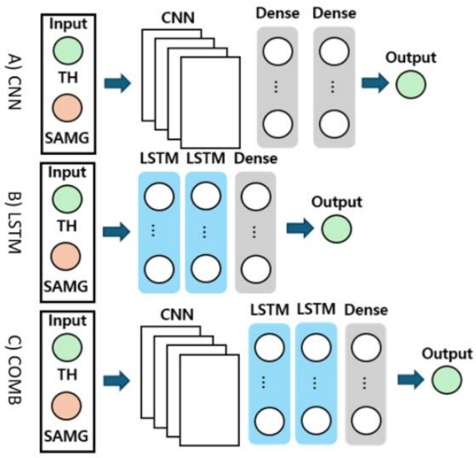

Figure 3 Diagram of Three Model Types

2.3.2 Output Activation Strategies

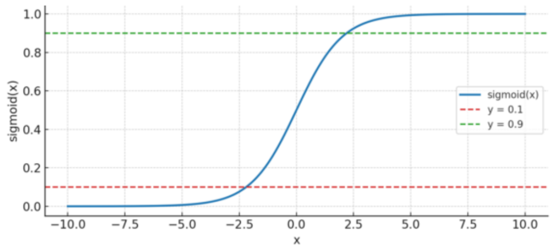

Figure 4 Activation Function of Default Model

To ensure physical consistency and prevent numerical instability during recursive forecasting, each model type is paired with one of the three output strategies:

Default: Utilizes a standard sigmoid (σ) activation to squash the output, followed by a hard clamp to the range [0.1, 0.9], as depicted by the S-curve in the figure 4.

Linear: Applies a raw linear projection without activation during training to allow unrestricted gradient flow, with clamping applied strictly during the inference step.

Scaled Sigmoid: Implements a custom mapping defined by:

$$\text{out} = 0.1 + 0.8 \times \sigma(x)$$

where, x represents the logits—the raw, unnormalized output produced by the final fully connected layer. By applying the sigmoid function σ(·) to these logits and scaling the result, this strategy naturally bounds the output within the safe range (0.1, 0.9) while maintaining differentiability across the entire domain.

2.4 Three-Step Process to Construct Surrogate Model

To ensure the reliability of safety-critical systems, a deterministic selection of a single model based on marginal error differences is insufficient. Therefore, a Three-Step Process is proposed to identify a robust candidate pool that is practically equivalent to the optimal model.

2.4.1 Step 1: Generalized Performance Evaluation

In the first step, the generalized performance of all 63 candidate models generated in Section 2.3 is evaluated. To isolate the intrinsic predictive capability of each architecture, a hybrid evaluation scheme is employed on the test set. For the target variable, the model operates in an auto-regressive, while ground truth values (MAAP data) are provided for all other input features.

The models are ranked based on the Mean Absolute Error (MAE). The architecture achieving the lowest MAE for each variable is designated as the baseline model, serving as the reference standard for the subsequent statistical selection.

2.4.2 Step 2: Candidate Selection (Glass's Δ)

In large-scale datasets, relying on p-values can lead to the p-value fallacy, where negligible differences are deemed statistically significant. To address this, Effect Size metrics are commonly used to quantify the magnitude of difference. While Cohen’s d is the standard metric, it utilizes a pooled standard deviation (s_{pooled}), which averages the variances of both the baseline and the candidate model.

However, in the context of safety-critical forecasting, Cohen’s d presents a limitation: if a candidate model exhibits excessive variance (instability), the pooled standard deviation increases, artificially reducing the effect size value. This could inadvertently allow an unstable, high-variance model to pass the selection criteria.

To overcome this, Glass’s Δ is adopted in this study. Unlike Cohen’s d, Glass’s Δ utilizes the standard deviation of the baseline model only. This approach treats the best-performing model as the standard for stability. The formula is defined as:

$$\text{Glass's } \Delta = \frac{MAE_{cadidate} - MAE_{baseline}}{s_{baseline}}$$

Where, s_{baseline} is the standard deviation of the Baseline Model identified in the first step evaluation. By fixing the denominator to the baseline’s stability, any candidate model with a larger error or instability is strictly penalized. Candidates are retained only if they satisfy the rigorous threshold of $\Delta < 0.2$, ensuring that the selected pool consists solely of models that are practically indistinguishable from the top performer in terms of both accuracy and stability.

2.4.3 Step 3: Full Recursive Integration

The final step integrates the statistically robust candidates selected in Step 2 into the complete MISO framework to construct a global surrogate model. Unlike the component-level evaluation in Step 1, this step operates in a full recursive forecasting environment where no ground truth data is provided for the input features. Instead, the selected candidates are coupled in a closed loop, where the output of one model serves as the input for others for subsequent time steps. This process evaluates the ability of the combined architectures to emulate system-level evolution over long horizons. By testing all valid combinations of the selected candidates, the study identifies the specific architecture that minimizes error accumulation and maintains trajectory stability, thereby validating the surrogate model’s reliability for safety-critical monitoring.

3. Results & Discussions

3.1 General Performance and Candidate Selection

The general performance of the initial model pool is evaluated in step 1. As defined in the methodology, the model space consists of 7 target variables, each assigned a localized pool of 9 candidate architectures (3 model types × 3 activation strategies). This hierarchical structure results in a total of 63 unique model instances (7 × 9) for the entire MISO framework. Under the Step 1 evaluation (auto-regressive target with ground truth features), each variable's 9-model pool was ranked to identify the baseline model (lowest MAE). As shown in Figure 5, the Glass’s Δ criterion with a threshold of $\Delta < 0.2$ was applied to each localized pool to determine the statistically equivalent candidates. The results are summarized in Table 3.

The performance analysis of the 63 total model configurations reveals significant structural trends tailored to specific thermal-hydraulic phenomena. The LSTM model emerged as the most effective architecture for capturing the complex dynamics of NPP transients, securing the baseline position in six out of seven target variables, including cold leg temperature, hot leg temperature, and containment pressure. This dominance confirms that the long-term sequential dependencies inherent in severe accident progressions are more effectively modeled by recurrent memory cells than by the local feature extraction typical of standalone CNN architectures. However, the COMB (CNN-LSTM) hybrid model proved uniquely superior for primary system pressure, achieving the lowest baseline MAE of 0.0004. This suggests that variables characterized by rapid, high-frequency fluctuations benefit from a multi-stage approach that integrates spatial-temporal feature extraction before sequence modeling.

In terms of output activation strategies, the efficacy of each method was closely linked to training stability and physical range enforcement. The scaled sigmoid and default (sigmoid + clamp) strategies collectively accounted for the majority of top-performing models, providing a necessary balance between differentiability and strict bounding within normalized limits. While the linear activation provides an unrestricted gradient flow that can theoretically accelerate convergence, it was identified as the optimal performer only for the cavity water level, which exhibits a relatively simple monotonic integration behavior. In most other variables, the linear strategy exhibited higher variance in error distribution, resulting in significant penalties under the Glass’s Δ criterion and subsequent exclusion from the candidate pool. These findings indicate that for safety-critical forecasting, bounded activation functions not only ensure physical validity but also contribute to the statistical stability required for reliable long-term predictions.

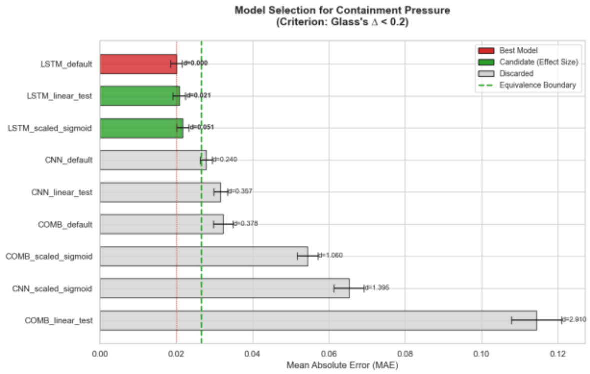

Figure 5 Glass's Δ of Containment Pressure

Table 3 Best Models and Statistically Equivalent Candidate Pools (Δ < 0.2)

Target Variable	Best Model (Baseline)	Candidate Pool (Statistically Equivalent Models)	Total (N)
PPS	COMB-Scaled Sigmoid	LSTM-Scaled Sigmoid, COMB-Default, LSTM-Linear, COMB-Linear	5
Cold Leg T	LSTM-Scaled Sigmoid	LSTM-Default, COMB-Scaled Sigmoid	3
Hot Leg T	LSTM-Default	COMB-Scaled Sigmoid, LSTM-Scaled Sigmoid	3
SG P	LSTM-Default	LSTM-Scaled Sigmoid, LSTM-Linear	3
SG WL	LSTM-Scaled Sigmoid	LSTM-Default, LSTM-Linear	3
CWL	LSTM-Linear	LSTM-Scaled Sigmoid	2
CTMT P	LSTM-Default	LSTM-Linear, LSTM-Scaled Sigmoid	3

3.2 Full Recursive Integration and Final Architecture

The MISO framework integrates the statistically robust candidates identified in step 2 into a unified system. A total of 2,430 ($5 \times 3^5 \times 2 = 2430$)-representing every possible combination of the candidate models listed for each variable in Table 3-unique architectural combinations were constructed and subjected to a full recursive simulation, where the system operates without ground truth feedback. The distribution of the Total Average MAE across all combinations is illustrated in Figure 6.

A critical finding from the recursive simulation is the significant performance discrepancy between local and global optimization. The configuration tagged Hybrid_tag_0000000, which represents the component-wise optimal baseline (the assembly of the best-performing models for each individual variable identified in step 1), achieved a Total MAE of 0.0309.

Unexpectedly, this locally optimized assembly ranked 857th out of 2,430 potential architectures. This empirical evidence demonstrates that simply assembling the best individual components does not yield the optimal system. The error propagation dynamics between variables—where a slightly less accurate but more stable model in one subsystem might stabilize the input for another—play a decisive role in recursive forecasting. This finding justifies the computational expense of the Step 3 exhaustive search, as it identified superior configurations that would have been missed by a deterministic selection approach.

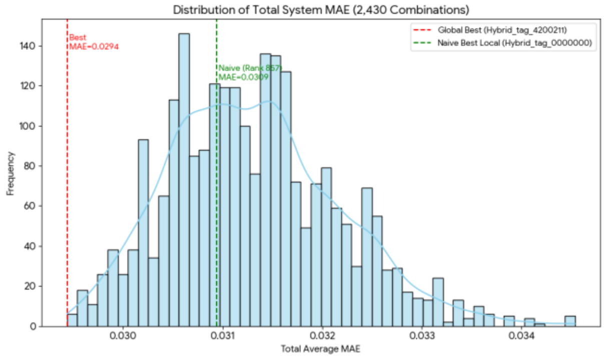

Figure 6 Distribution of Global MAE

Table 4 Final MISO Architecture

Target variable	Selected Model
Primary system pressure	COMB-Linear
Cold leg temperature	COMB-Scaled Sigmoid
Hot leg temperature	LSTM-Default
Steam generator pressure	LSTM-Default
Steam generator water level	LSTM-Linear
Cavity water level	LSTM-Scaled Sigmoid
Containment Pressure	LSTM-Linear

The analysis identified the configuration Hybrid_tag_4200211 as the Global Optimum. This architecture achieved the Total Average MAE of 0.0294, significantly outperforming the baseline assembly. The detailed composition of the final MISO architecture is presented in Table 4. The final architecture represents a synergistic integration, combining the long-term stability of LSTM-based models for the secondary and ex-vessel systems with the responsive hybrid features of COMB-based models for the primary loop. This strategic mix allows the system to maintain global stability and high accuracy throughout the entire accident progression timeline.

3.2 Sensitivity Analysis of Global System Performance

To elucidate the drivers of global system stability beyond the single optimal architecture, a comprehensive sensitivity analysis was conducted across the entire solution space of 2,430 combinations. This analysis aims to quantify the impact of individual model selection on the Total Average MAE, thereby identifying which variables serve as critical keystones for system reliability and which allow for architectural flexibility.

Figure 7 illustrates the distribution of the global system error corresponding to the specific model choices for each target variable. In this visualization, the vertical position of the boxplots represents the global system MAE, while the spread within each box indicates the variance caused by the combinations of the other six variables. As shown in the figure, the impact of model selection is highly asymmetrical across the thermal-hydraulic parameters. The steam generator pressure emerged as the most critical determinant of global performance. The boxplots for this variable show a distinct separation, where the selection of the LSTM-Default model consistently shifts the entire error distribution downwards, whereas other architectures lead to significantly higher systemic errors.

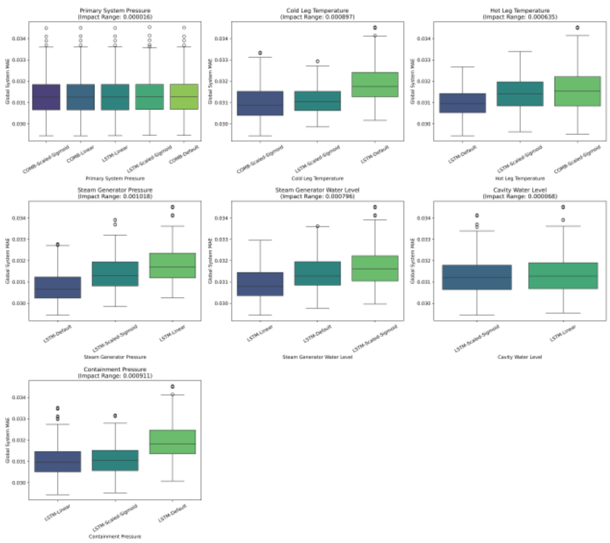

Figure 7 Impact of Model Selection on Global MAE

This high dependency is further shown on Figure 8, which ranks the variables based on their sensitivity range—defined as the difference between the maximum and minimum mean Global MAE yielded by the candidate models. Specifically, altering the candidate model for the steam generator pressure causes the most drastic fluctuations in the overall system error (a sensitivity range of ≈0.0010), with containment pressure and cold leg temperature exhibiting similar critical impacts. This reveals a clear cause-and-effect relationship: the system's global performance is not equally dependent on all variables. Instead, overall stability is strictly dictated by a few highly influential parameters, where a suboptimal model choice locally will disproportionately degrade the accuracy of the entire MISO framework.

From a data-driven perspective, suboptimal model choices for these high-sensitivity parameters introduce significant numerical deviations that propagate recursively through the feature vectors. This recursive error amplification degrades the entire system's trajectory regardless of the accuracy of other components. Consequently, this result validates the necessity of the proposed combinatorial search strategy (Step 3), as a deterministic or localized selection approach would fail to account for these dominant statistical dependencies inherent in the coupled multivariate forecasting.

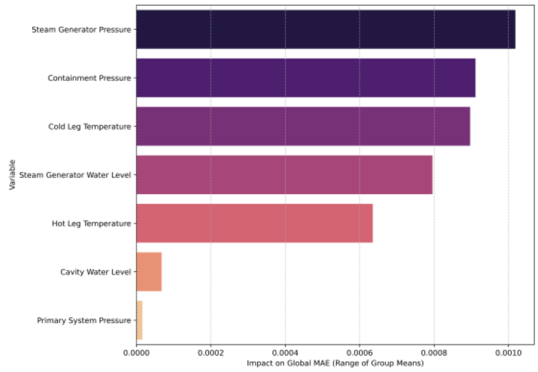

Figure 8 Global MAE Sensitivity Ranking

Conversely, the Primary system pressure exhibited negligible sensitivity, with an impact range of approximately 1.6×10^{-5}. As observed in Figure 7, the boxplots for all five candidate models for this variable are nearly identical in terms of their medians and interquartile ranges. Rather than being tied to complex accident phenomenology, this phenomenon is primarily due to the inherent statistical simplicity of the variable's data profile within the dataset. The pressure data exhibits a straightforward, monotonic decay pattern that lacks the severe, high-frequency nonlinear bifurcations seen in secondary-side variables. Because this represents a relatively low-difficulty forecasting task, most candidate models—regardless of their internal architectural complexity—successfully captured this dominant trend with high accuracy. Consequently, the choice of model for this specific variable introduces almost no variance to the global system error.

4. Conclusions & Further Works

This study proposed a robust optimization framework for MISO-based severe accident forecasting by implementing a tree-step statistical selection using Glass's Δ and a subsequent global recursive integration. The empirical results demonstrated that a naive combination of individually optimal models failed to achieve system-level stability, ranking only 857th out of 2,430 potential configurations. In contrast, the globally optimized architecture significantly minimized the total system error by accounting for the asymmetrical sensitivity of keystone variables, such as steam generator pressure, versus flexible components like primary system pressure. Therefore, a comprehensive combinatorial search is strictly necessary to identify the optimal heterogeneous architecture that effectively mitigates recursive error propagation in coupled multivariate forecasting systems, as local optimization alone cannot resolve the complex inter-variable dependencies.

Future research will extend this framework by applying it to accident scenarios with fundamentally different statistical distributions, specifically loss of coolant accidents (LOCA) and station black out (SBO). Furthermore, the analysis will be deepened from the current global error perspective to a granular cross-variable sensitivity assessment. By quantifying how architectural changes in specific variables directly influence the forecasting accuracy of the other target variables, a directed graph of error propagation is aimed to be constructed, enabling a more precise and explainable optimization of the coupled multivariate forecasting framework.

ACKNOWLEDGMENT

This work was supported by the K-Cloud project of KOREA HYDRO & NUCLEAR POWER CO., LTD (No. 2024-Tech-08)

REFERENCES

[1] Lee, Yeonha, et al. "Surrogate model for predicting severe accident progression in nuclear power plant using deep learning methods and Rolling-Window forecast." Annals of Nuclear Energy 208 (2024): 110816.
[2] Joo, Semin. et al "Improving accuracy of a machine learning approach for forecasting severe accidents." 14th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics, Operation, and Safety, 28 Aug. 2
[3] Song, Seok Ho, et al. "Advancing the MISO Strategy: Hybrid Architectures for Thermal-Hydraulic Forecasting in Severe Accidents."
[4] Lin, Mingfeng, Henry C. Lucas Jr, and Galit Shmueli. "Research commentary—too big to fail: large samples and the p-value problem." Information systems research 24.4 (2013): 906-917.
[5] Cohen, Jacob. Statistical power analysis for the behavioral sciences. routledge, 2013.
[6] Glass, Gene V. "Primary, secondary, and meta-analysis of research." Educational researcher 5.10 (1976): 3-8.
[7] Arcuri, Andrea, and Lionel Briand. "A practical guide for using statistical tests to assess randomized algorithms in software engineering." Proceedings of the 33rd international conference on software engineering. 2011.
[8] Rajput, Daniyal, Wei-Jen Wang, and Chun-Chuan Chen. "Evaluation of a decided sample size in machine learning applications." BMC bioinformatics 24.1 (2023): 48.