Transactions of the Korean Nuclear Society Autumn Meeting
Changwon, Korea, October 30-31, 2025

Enhancing Time Resolution in Surrogate Modeling of Severe Accident Analysis Code with Attention Mechanism

Wonung Jeong [1], Sangam Khanal[2], Joongoo Jeon[3*]
[1]*Department of Quantum System Engineering, Jeonbuk National University, Jeonju, Republic of Korea*
[2]*Graduate School of Integrated Energy-AI, Jeonbuk National University, Republic of Korea*
[3]*Division of Advanced Nuclear Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea*
[*]*Corresponding author: jgjeon41@postech.ac.kr*

Keywords : severe accident, machine learning, nuclear safety

1. Introduction

The analysis of severe accidents in nuclear power plants has primarily relied on system codes such as MELCOR, MARS, and MAAP. These codes provide a comprehensive framework for analyzing plant behavior under beyond-design-basis conditions across a broad range of scenarios and initiating events. Given their ability to simulate the progression from core damage to potential containment failure, they are indispensable for safety and regulatory assessments in the nuclear industry [1].

These system codes have achieved remarkable improvements in computational efficiency—reaching near real-time performance for single accident scenarios—, But they still face limitations in scenarios involving complex sequences beyond-design-basis-Accident (BDBA), where uncertainties and long computation times remain. To overcome these challenges, recent research has focused on applying machine learning to predict accident progression. In this context, we aim to develop a Transformer-based model capable of predicting accident evolution at high time resolution, and furthermore, to investigate surrogate models that are more suitable for integration with reinforcement learning (RL) based AI agents [1, 2].

Prior surrogate modeling approaches using multilayer perceptrons (MLPs), recurrent neural networks (RNNs), and their variants have demonstrated the feasibility of predicting severe accident progression. However, because these models typically rely on relatively short time-series inputs, autoregressive predictions suffer from error accumulation. In addition, severe accident progression BDBA conditions involves highly nonlinear interactions, and current Korean regulations require severe-accident evaluations to include a 72-hour coping period, necessitating long-horizon analysis. Consequently, conventional RNN-based models exhibit clear limitations for high-resolution time-series prediction, underscoring the need for more advanced architectures [2].

To address these limitations, we propose a Transformer-based surrogate model that leverages attention mechanisms to capture long-range temporal dependencies in severe-accident time-series data. By exploiting extended temporal context to mitigate autoregressive error accumulation, the model enables high-resolution time-series prediction of accident evolution and real-time simulation. We train the model on MAAP simulation datasets for the OPR1000 reactor to predict both accident progression and the effectiveness of mitigation strategies. This approach highlights the potential of attention-based architectures to integrate with RL agents for real-time decision-making in severe-accident management.

2. Methodology

2.1 Selection of accident scenario

To select a severe accident scenario for the OPR1000 reactor, accident frequencies were quantified by calculating the frequencies of Plant Damage States (PDS) derived from Level 2 Probabilistic Safety Assessment (PSA). Among these, the Total Loss of Component Cooling Water (TLOCCW) scenario was selected based on the product of PDS frequency and the fraction of each accident progression path, prioritizing scenarios with higher combined probabilities [2].

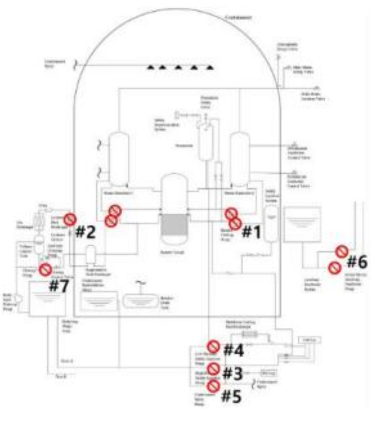

Figure 1 Detailed location of component failure on OPR1000 safety system

As shown in **Figure 1**, the TLOCCW accident leads to a loss of all component cooling water in the primary side, resulting in the failure of safety-related systems such as emergency diesel generators and essential cooling systems. Consequently, the unavailability of safety injection and auxiliary feedwater pumps necessitates a manual reactor shutdown within 10 minutes, followed by secondary-side feedwater supply and steam release for decay heat removal [1, 3].

2.2 Dataset

For this study, a TimeSeriesTransformer model was developed to predict thermal-hydraulic (TH) variables observable from the Main Control Room (MCR), using normalized MAAP simulation data corresponding to the TLOCCW scenario in the OPR1000 nuclear power plant, which was constructed based on the KAIST study [3].

The dataset construction and model architecture were optimized to reflect the characteristics of the inputs available to operators during a severe accident. Specifically, the TH variables were treated as continuous time-series signals, while component failure states and Severe Accident Management Guidance (SAMG) signals were treated as binary features. The total input features for model training are summarized in **Table 1** [3]. The attention mechanism within the Transformer architecture was optimized to effectively handle these data characteristics.

Table I Total input feature

#	(Target) thermal-hydraulic variable
1	Primary system pressure (PPS)
2	Cold leg temperature (Cold leg T)
3	Hot leg temperature (Hot leg T)
4	Reactor vessel water level (ZWV)
5	Steam generator pressure (SG P)
6	Steam generator water level (SG WL)
7	Maximum Core Exit Temperature (Max CET)
#	**Component failure**
1	Reactor coolant pump (RCP) seal LOCA
2	Letdown heat exchanger (HX)
3	High-pressure injection (HPI) pump
4	Low-pressure injection (LPI) pump
5	Containment spray system (CSS) pump
6	Motor-driven auxiliary feedwater (MDAFW) pump
7	Charging pump (CHP)
8	Refueling Water Storage Tank (RWST)
#	**SAMG mitigation**
1	Steam generator external injection
2	Reactor cooling system depressurization
3	Reactor cooling system external injection

2.2.1 RWST Feature

In previous KAIST studies, the RWST feature was shown to have a direct impact on the HPI, LPI, charging pump, and containment spray systems. [4].

2.2.2 Spikes/Peaks Feature

It was observed that features representing the characteristics of severe accidents—specifically, MAX CET and ZWV—exhibited sharp fluctuations such as spikes and peaks (Figure 2) [4]. These observations were used to confirm the distinctive characteristics of the MAX CET and ZWV data during the analysis.

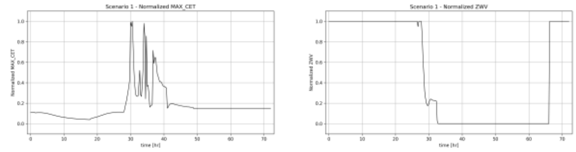

Figure 2 Time series of MAX CET and ZWV under TLOCCW scenario

2.3 Autocorrelation method

Autocorrelation analysis was conducted to examine the temporal dependencies within the dataset [5]. **Equation 1** computes the correlation between time points separated by lag k, and the confidence interval serves as a threshold for evaluating the statistical significance of the correlation coefficients.

$$\boldsymbol{Autocorrelation} = \frac{\sum_{t=k+1}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2} \qquad \textbf{Eq. 1}$$

$$\boldsymbol{Confidence\ interval} = \pm\frac{1.96}{\sqrt{T}} \approx \pm \mathbf{0.115} \qquad \textbf{Eq. 2}$$

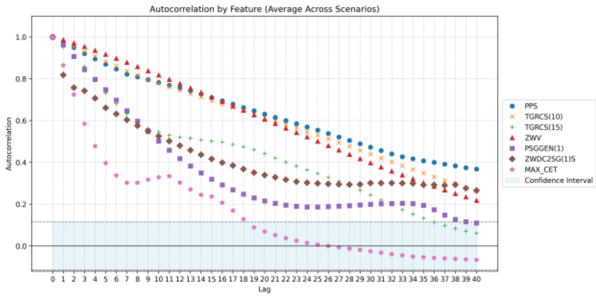

Figure 3 TH value Autocorrelation by Feature

Based on the analysis **Equation 1,2** of 72-hour datasets sampled at 15-minute intervals, it was observed that for many features, the autocorrelation values began to fall within or near the confidence interval (±0.1155) after lag 30, indicating a reduction in statistical significance **(Figure 3)**.

Although certain features maintained meaningful correlations at longer lags, increasing the time series length leads to higher input dimensionality and the risk of noise propagation. Considering feature-wide consistency, overfitting prevention, computational efficiency, and domain-specific interpretability, a lag of 20 ~ 30 was selected as the optimal range. Consequently, the sequence length was fixed at 30, ensuring that only meaningful historical data were used as model inputs.

3. Transformer model architecture

3.1 Decoder transformer.

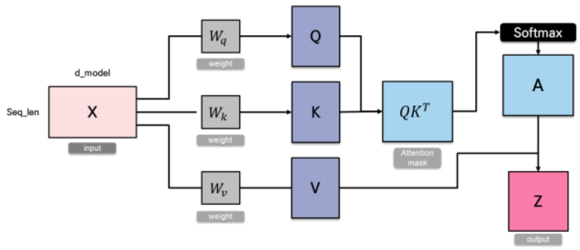

Figure 4 Scale-dot product (Kim. L. 2025)

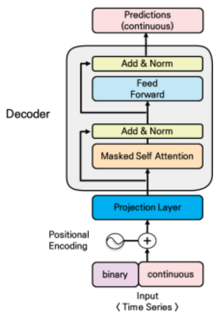

Figure 5 Decoder transformer architecture

As illustrated in **Figure 4**, the Transformer architecture addresses the long-term dependency problem inherent in traditional RNNs models by utilizing an attention mechanism that computes the dot product of Key, Query, and Value matrices [6]. In the context of severe accident progression over a 72-hour period with 15-minute resolution, the self-attention mechanism is particularly effective in capturing complex and nonlinear temporal relationships within the data.

Furthermore, as shown in **Figure 5**, the decoder architecture embeds both binary and continuous input features into a unified representation and uses them as inputs to predict only the TH values.

3.2 Cross-attention transformer

To optimize the attention-based model architecture, the characteristics of TH values, component failure states, and SAMG signals were explicitly considered. In this configuration, TH variables were treated as continuous inputs, while component failure and SAMG features were treated as binary signals. Based on this distinction, a Cross-Attention Transformer was constructed to incorporate feature-type awareness. The latter was designed to more effectively integrate heterogeneous input modalities and enhance prediction accuracy under severe accident conditions.

As illustrated in **Figure 6** and **Figure 7**, the Cross-Attention Transformer explicitly separates continuous TH variables from binary component failure and SAMG signals and integrates them through a cross-attention mechanism to improve predictive performance.

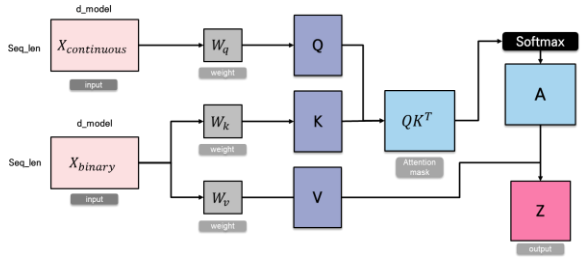

Figure 6 Cross scale-dot product

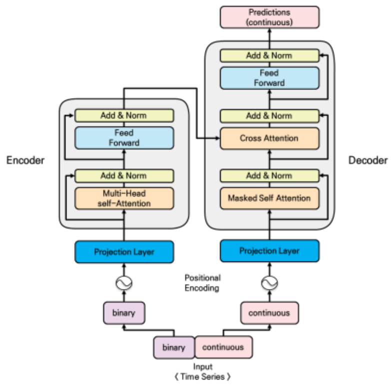

Figure 7 Cross attention transformer architecture

4. Training and Testing Method

4.1 Hyperparameter and Training Configuration

In this study, all Transformer models were trained with the same set of hyperparameters to ensure fair comparisons across different input features and time resolutions. Specifically, the model dimension was fixed at $\boldsymbol{D_{model} = 64}$, the number of attention heads at $\boldsymbol{n_{head} = 4}$, and the number of layers at $\boldsymbol{N_{head} = 8}$. By maintaining consistent architecture, the effects of varying data characteristics and temporal resolutions on model performance could be isolated and fairly evaluated.

The loss function used for training was Mean Squared Error (MSE), defined as **Equation 3**:

$$\boldsymbol{L_{MSE} = \frac{1}{N_{TH}} \sum_{i=1}^{N_{TH}} (y_{MAAP,i} - \hat{y}_{pred,i})^2} \qquad \textbf{Eq. (3)}$$

$$\text{where } N_{TH} = \text{number of TH variables}$$
$$\hat{y}_{pred,i} = \text{TH variable } i \text{ predicted by surrogate}$$

$$y_{MAAP,i} = TH\ variable\ i\ predicted\ by\ MAAP$$

For optimization, the AdamW (Adaptive Moment Estimation with Weight Decay) algorithm was adopted. This method adjusts the learning rate adaptively using momentum, updating the first and second moment estimates to stabilize convergence and improve generalization.

4.2 Experimental Cases

A feasibility study was conducted using two datasets: one with the baseline of 17 input features and another including the RWST feature, totaling 18 features. Three experimental cases were designed, as summarized in **Table 2**:

Base Case: A Decoder-only Transformer model trained on 1-hour interval time-series data. Case 1 and Case 2 differ only in the length of historical sequences used for prediction. Input features were normalized to the range [0.2 ~ 0.8].

RWST Case: A Decoder-only Transformer model trained on 15-minute interval time-series data with the RWST feature included. Input features were normalized to the range [0 ~ 1.0].

Cross Case: A Cross-Attention Transformer model trained on 15-minute interval data. Input features were also normalized to the range [0 ~ 1.0].

Table II Experimental Case

Case	Base Case 1	Base Case 2	RWST Case	Cross Case
Interval	**1h**	**1h**	**15min**	**15min**
Seq_len	**3**	**7**	**30**	**30**
Input size	17	17	18	18
Total Para	203k	203k	203k	935k

Table III Accuracy and Loss by Case

	Acc (val.)	train_loss	Val. loss
Base Case	0.91857	0.0004221	0.0004067
Base Case 2	0.91940	0.0003962	0.0003843
RWST Case	0.93051	**0.0002716**	**0.0002727**
Cross Case	**0.99029**	0.0003374	0.0003048

5. Results

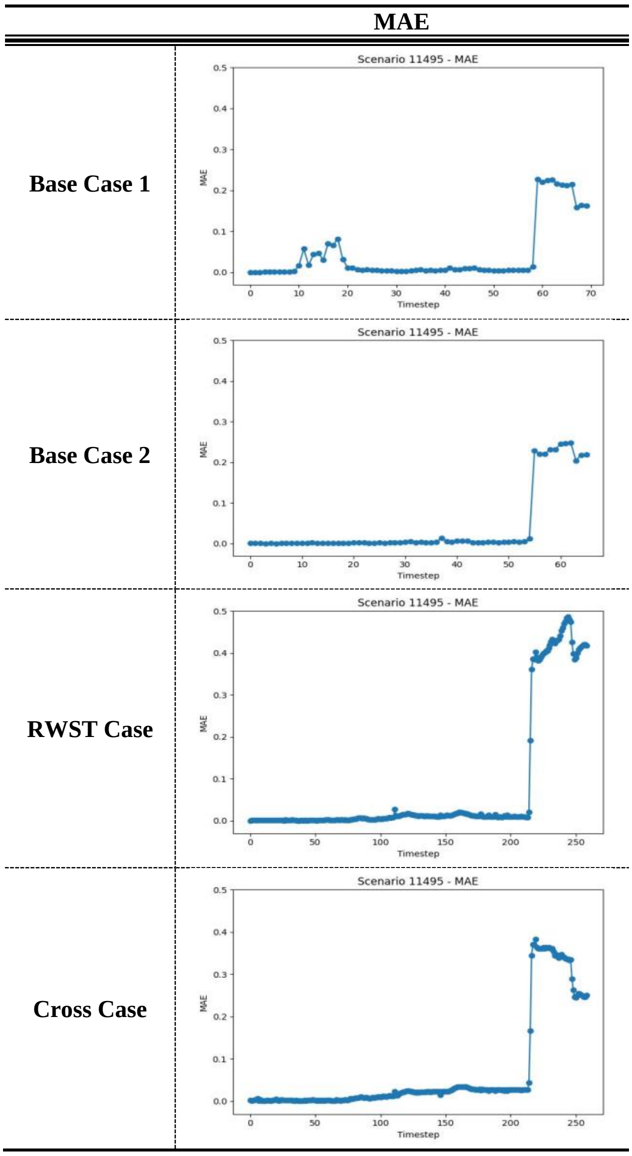

Figure 8 Prediction Results by Case (Solid Line: True / Dashed Line: Prediction)

5.1 Results and Discussion

The analysis shows that in Base Case 1 and 2, the train and validation losses as well as MAE decreased as the time-series length increased, confirming the benefit of longer input sequences. Moreover, when comparing the RWST Case and the Cross Case, the overall train and validation losses were lower with the simpler decoder model, but the MAE beyond approximately 220 steps was smaller in the Cross Case.

In addition, the RWST Case indicated the feasibility of applying a high time resolution of 15-minute intervals, suggesting its potential for more fine-grained prediction tasks.

Moreover, the model employing the Cross-Attention Transformer architecture achieved the highest validation accuracy of 0.99029 (Table III). However, it also resulted in an increased number of parameters (approximately 935,000), highlighting a limitation in terms of computational cost. Additionally, error spikes observed in specific time segments (Figure 2) may be attributed to dataset characteristics or architectural constraints, suggesting the need for further investigation.

6. Conclusions and Further Works

Future work will focus on enhancing model interpretability by incorporating SHAP-based Explainable AI methods to quantitatively assess the importance of key features. These insights will be utilized to guide further architectural optimization.

In addition, we plan to develop a high-resolution surrogate model for predicting severe accident progression based on MAAP simulations of the Loss of Feed Water (LOFW) scenario in the APR1400 nuclear power plant. Finally, by applying the Decoder Transformer framework to this extended dataset, we aim to explore simultaneous optimization of both model architecture and input representation, ultimately achieving improvements in both predictive accuracy and computational efficiency.

Acknowledgment

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

REFERENCES

[1] W. Jeong, Y. Ko, J. Jeon, "Time Series Transformer for Forecasting the Progression of Severe Accidents in Nuclear Reactors," Proceedings of the Applied Artificial Intelligence Conference (AAICON), Daejeon, Korea, Feb. 2025.

[2] Korea Advanced Institute of Science and Technology (KAIST), “Feasibility study on the application of artificial intelligence for severe accident management: Final report” [in Korean], Korea Hydro & Nuclear Power Co., Daejeon, Korea, May 2023.

[3] S. H. Song, S. Joo, Y. Lee, M. R. Seo, J. I. Lee*, "Evaluation of Multi-Input Single-Output ANN Models for Thermal-Hydraulic Predictions in Nuclear Severe Accidents: Branched vs. Non-Branched Structures," Transactions of the Korean Nuclear Society Spring Meeting, Jeju, Korea, May 22–23, 2025.

[4] S. Joo, S. H. Song, Y. Lee, J. I. Lee*, *"Feasibility study of applying an explainable AI (XAI) model for an accelerated prediction of severe accident progression,"* Transactions of the Korean Nuclear Society Spring Meeting, Jeju, Korea, May 2024

[5] J. Joo, S. H. Song, Y. Lee, J. Young Bae, K. Song, M. R. Seo, Sung J. Kim, J. I. Lee, *"Surrogate model for predicting severe accident progression in nuclear power plant using deep learning methods and rolling-window forecast,"* Annals of Nuclear Energy, vol. 208, Article 110816, Dec. 2024.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.

[7] S. Joo, S. H. Song, Y. Lee, J. I. Lee, S. J. Kim, *"Accelerated prediction of severe accident progression: Sensitivity of deep neural network performance to time resolution,"* Transactions of the Korean Nuclear Society Autumn Meeting, Gyeongju, Korea, Oct. 2023.

[8] S. Khanal, S. Baral, and J. Jeon, "Comparison of CNN-based deep learning architectures for unsteady CFD acceleration on small datasets," Nuclear Engineering and Technology, vol. 57, no. 10, p. 103703, 2025. doi: https://doi.org/10.1016/j.net.2025.103703.

[9] J. Shin, C. Kim, S. Yang, M. Lee, S. J. Kim, and J. Jeon*, "Node Assigned physics-informed neural networks for thermal-hydraulic system simulation: CVH/FL module," under review, arXiv preprint arXiv:2504.16447, 2025.