

Benchmarking Reduced-Order Models of CFD Flow Fields for SMR Digital Twins

Seongmin Oh^a, Minseo Lee^b, Joongoo Jeon^{b*}

^a Department of Quantum System Engineering, Jeonbuk National University, 567, Baekje-daero, Deokjin-gu, Jeonju-si, Republic of Korea

^b Division of Advanced Nuclear Engineering, POSTECH, 77 Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Republic of Korea

* Corresponding author: jgjeon41@postech.ac.kr

*Keywords: computational fluid dynamics, reduced-order model, deep learning, small modular reactor

1. Introduction

Computational fluid dynamics (CFD) can resolve complex flow phenomena, such as those in reactor thermal-hydraulic systems, at high resolution, but its high computational cost limits its direct use for real-time prediction. To realize a virtual reactor capable of rapidly predicting the thermal-hydraulic flow behavior of an actual reactor, AI-based surrogate models are required to predict high-fidelity CFD flow fields rapidly and accurately. However, because CFD flow fields are extremely high-dimensional, directly learning and predicting the full flow field is inefficient in terms of computational cost and training efficiency.

To address this issue, data-driven reduced-order models (ROMs) compress high-dimensional CFD flow fields into low-dimensional latent spaces, allowing AI surrogate models to learn the behavior of a compact set of latent variables instead of directly predicting the entire flow field. This substantially reduces the dimensionality of the learning problem and alleviates the computational and memory requirements associated with training and inference. In particular, deep-learning-based autoencoders can represent the dominant spatial characteristics of flow fields as low-dimensional latent vectors, and when coupled with operator-learning models such as DeepONet, these latent representations can be used to predict flow fields under varying time or operating conditions [2-4].

However, if important flow characteristics, such as vortex structures, velocity gradients, and spatial correlations, are not sufficiently preserved during dimensionality reduction, the accuracy of the reconstructed flow field can be fundamentally limited even when the subsequent latent-space surrogate model performs well. Therefore, ROM is not merely a preprocessing step but a key component that can determine the accuracy and computational efficiency of the overall AI surrogate framework. Accordingly, selecting a dimensionality-reduction architecture that effectively preserves the dominant spatial characteristics of the original CFD flow field is essential.

In this study, four dimensionality-reduction models, a multilayer perceptron (MLP), a convolutional neural network (CNN), a graph convolutional network (GCN)-based graph autoencoder (GCN-GAE), and a MeshGraphNet (MGN)-based graph autoencoder

(MGN-GAE) were compared in terms of reconstruction accuracy and computational efficiency [5-7]. The objective is to identify a suitable dimensionality-reduction architecture for subsequent application to three-dimensional reactor thermal-hydraulic flow fields and coupling with latent-space-based AI surrogate models.

2. Methods

This section describes the target problem and dataset, the preprocessing into two input representations, the four autoencoder architectures, the controlled training configuration, and the evaluation methods.

2.1 Target Problem and Dataset

Two-dimensional transient CFD data representing cross-flow in the helical coil steam generator (HCSG) of the SMART reactor were used to train and evaluate the dimensionality-reduction models. As shown in Fig. 1, the computational domain consists of a staggered array of circular tubes and was designed to capture vortex shedding and flow interactions between adjacent tubes. A two-dimensional unstructured triangular mesh was employed, and transient simulations were performed using ANSYS Fluent with the SST $k-\omega$ turbulence model.

To represent different operating conditions, CFD simulations were conducted for inlet velocities from 0.10 m/s to 0.60 m/s at 0.01 m/s intervals. The 0.40 m/s case was excluded from training and used for interpolation evaluation, while an additional 0.70 m/s case outside the training range was used for extrapolation evaluation. Each operating condition contains 100 snapshots of pressure (p), x-velocity (u), and y-velocity (v).

The same CFD dataset was used for the MLP, CNN, GCN-GAE, and MGN-GAE models. The MLP and graph-based models directly used nodal data from the unstructured mesh, whereas the CNN required interpolation onto a structured grid.

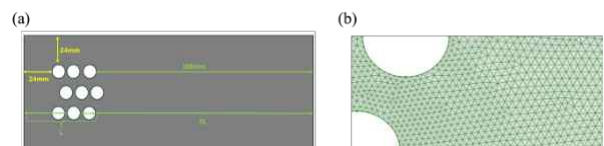


Fig. 1. (a) Geometry and (b) unstructured mesh of the HCSG cross-flow domain.

The 50 training cases thus provide 5,000 snapshots in total, which were split into 4,000 for training and 1,000 for validation.

2.2 Data Preprocessing

Two preprocessing paths were applied to feed the same CFD data to the four models. Since convolution presupposes a uniform structured mesh, the unstructured field was interpolated onto a structured mesh for the CNN, whereas the MLP and the two graph autoencoders use the nodal values of the unstructured mesh directly. Consequently, the information loss due to interpolation exists only for the CNN and not for the other three models. This asymmetry arises inherently from how each architecture handles CFD data and must be interpreted as part of the comparison results.

2.2.1 Structured-Mesh Conversion

Considering both interpolation accuracy and computational efficiency, the structured-grid resolution for the CNN was set to approximately 3.5 times the number of nodes in the original unstructured mesh (Fig. 2).

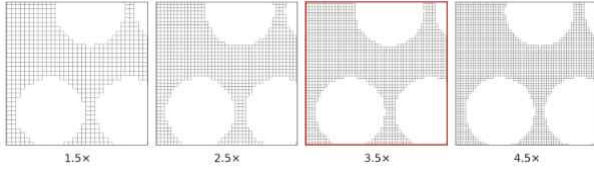


Fig. 2. Structured grids near a coil at 1.5 $\times$, 2.5 $\times$, 3.5 $\times$ (adopted, red), and 4.5 $\times$ resolution.

The unstructured CFD data were interpolated onto the structured grid using inverse distance weighting (IDW). To improve interpolation accuracy, the interpolation parameters were optimized by considering spatial variations in the flow field, particularly near wall regions. With the optimized parameters, the coefficients of determination (R^2) were 0.99595, 0.99617, and 0.99462 for pressure, x-velocity, and y-velocity, respectively. Grid points corresponding to non-fluid regions, such as the tube interiors, were excluded from training and evaluation.

2.2.2 Unstructured-Node and Graph Representations

The MLP and the two graph autoencoders use the 28,877 nodes of the original unstructured mesh directly. Because no structured-grid interpolation is required, no interpolation-induced information loss occurs in this preprocessing path, and no masking is needed because no fictitious grid points are generated inside the solids. The storage volume is also reduced to 28% of the structured-mesh training set.

The graph was constructed by connecting each node to its four nearest neighbors and symmetrizing, giving

130,360 directed edges, and each edge carries a three-dimensional feature composed of the relative position and distance of the connected nodes.

2.3 Autoencoder Architectures

The overall architectures of the four models are shown in Fig. 3, with the latent dimension fixed at 256 for a consistent comparison.

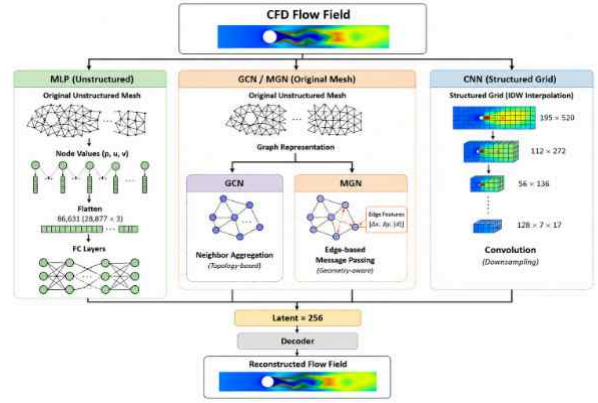


Fig. 3. Architectures of the four autoencoders (latent dimension 256).

The four autoencoder architectures were constructed with a common latent dimension of 256. The CNN-based autoencoder compresses structured-grid inputs through convolutional layers and reconstructs them using transposed convolutions. The MLP-based autoencoder serves as a baseline by directly flattening the unstructured nodal data without explicitly considering spatial topology. The GCN-GAE aggregates neighboring nodal information through graph convolutions, whereas the MGN-GAE employs edge-conditioned message passing that incorporates relative position and distance as geometric edge features [7]. In all models, the encoded features are mapped to the same 256-dimensional latent space for a consistent comparison.

2.4 Training and Evaluation

To ensure a fair comparison, all models were trained using the same dataset, latent dimension, optimizer, and overall training strategy. All three flow variables, pressure (p), x-velocity (u), and y-velocity (v), were included in training and reconstruction. Pressure and x-velocity were selected as representative variables for quantitative evaluation using the channel-wise coefficient of determination (R^2) averaged over all 100 test frames.

3. Results and Discussion

3.1 Training Convergence

The training and validation loss curves of the four models are shown in Fig. 4.

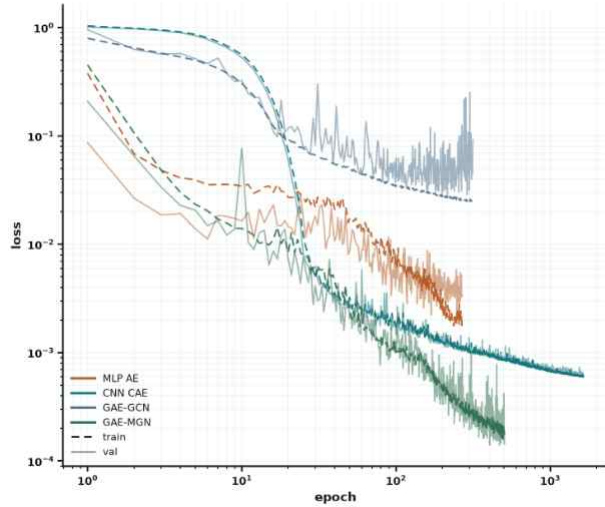


Fig. 4. Training and validation loss curves.

The MGN achieved the lowest validation loss but required the largest training cost, whereas the CNN reached the second-lowest loss with substantially lower computational cost. The GCN converged at a higher and more oscillatory loss level, while the MLP required a much larger parameter count without a corresponding improvement in reconstruction accuracy.

3.2 Qualitative Comparison

The qualitative comparison targets the instantaneous flow field at the 800th time step. This instant lies well past the initial transient, where the vortex shedding in the tube wakes has settled into a periodic regime, so the reconstruction characteristics of each model can be compared for a fully developed wake structure.

The velocity magnitude distributions for the interpolation and extrapolation conditions are shown together in Fig. 5. Under both conditions the MLP, CNN, and MGN reproduce the accelerated passages between tubes and the vortex structures developing in the wake to a degree visually indistinguishable from the reference. The GCN, in contrast, exhibits high-frequency speckle noise across the entire domain and smeared velocity gradients near the tube bundle. Because the flow field is dominated by its large-scale structure, however, the differences among the three models other than the GCN are not discernible in this representation; the differences emerge only in the error fields.

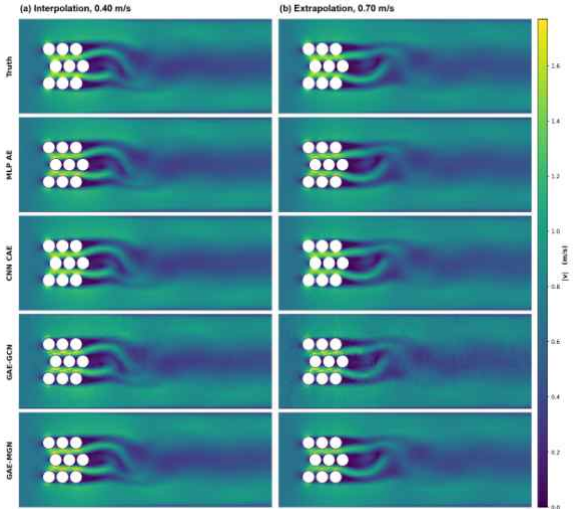


Fig. 5. Velocity magnitude fields for (a) interpolation and (b) extrapolation.

The velocity-magnitude error fields for the two conditions are presented in Fig. 6. Under interpolation, the MGN error is the lowest, almost indistinguishable from the background, whereas the GCN shows the largest error as broadband noise, a behavior consistent with over-smoothing in repeated neighborhood aggregation [8]. Under extrapolation, the CNN error becomes the lowest, while the GCN noise amplifies to near the upper end of the color range.

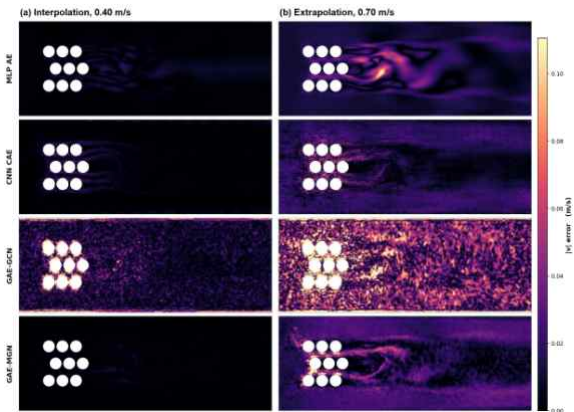


Fig. 6. Velocity magnitude error fields for (a) interpolation and (b) extrapolation.

3.3 Compression Accuracy under Interpolation and Extrapolation

Although all three flow variables (p , u , and v) were included in the reconstruction objective, the quantitative comparison presented here focuses on pressure and x -velocity as representative physical variables. The channel-wise coefficients of determination (R^2), computed over all 100 test frames, are presented in Table I.

Table I. Channel-wise R^2 over all 100 test frames

Model	p interp.	p extrap.	u interp.	u extrap.
MLP	0.99895	0.99308	0.99853	0.99390
CNN	0.99936	0.99688	0.99896	0.99412
GCN	0.98787	0.97802	0.95314	0.92251
MGN	0.99989	0.99491	0.99982	0.98991

Under the interpolation condition, the MGN is the most accurate for both reported variables, with R^2 above 0.9998, while the GCN falls substantially behind, its x-velocity R^2 dropping to the 0.95 level. Under the extrapolation condition, the ranking reverses: the CNN becomes the most accurate for both variables and the MGN falls behind it, consistent with the error distributions in Fig. 6.

This reversal shows that the architecture that reconstructs most precisely inside the training distribution does not necessarily generalize best beyond it. Since a ROM may be deployed at previously unseen operating conditions, extrapolation robustness should be weighed alongside interpolation accuracy.

3.4 Accuracy and Computational Cost

Along the two axes of accuracy and computational cost, the MGN is the most accurate at interpolation but carries the largest training cost, whereas the CNN offers the most favorable balance, achieving the best extrapolation accuracy at the expense of the structured-mesh interpolation error. The MLP gains no accuracy from its far larger parameter count, and the GCN offers no advantage on either axis. For the three-dimensional full-scale application, the GCN is excluded for its accuracy and the MLP for its parameter count growing with the number of points; the CNN and the MGN remain as candidates, whose respective limitations (interpolation error and message-passing cost) will be evaluated in a subsequent study.

4. Conclusions

In this study, four autoencoder architectures based on MLP, CNN, GCN, and MeshGraphNet were compared for dimensionality reduction of reactor thermal-hydraulic flow fields. The latent dimension and overall training conditions were kept consistent across the models. Reconstruction performance was evaluated using two-dimensional cross-flow data from the SMART HCSG at 0.40 m/s for interpolation and 0.70 m/s for extrapolation.

Under the interpolation condition, the MGN-GAE achieved the highest reconstruction accuracy for both pressure and x-velocity, although it required the largest training cost. Under the extrapolation condition, however, the CNN achieved the highest accuracy for

both variables. These results indicate that the dimensionality-reduction model with the best reconstruction performance within the training range does not necessarily provide the best performance outside the training range. Therefore, interpolation accuracy, extrapolation performance, and computational efficiency should all be considered when selecting a dimensionality-reduction model for ROM applications.

In addition to reconstruction accuracy, pressure drop and vorticity were evaluated to assess the preservation of thermal-hydraulic characteristics after dimensionality reduction. These quantities represent the overall flow loss and local flow structures, respectively. The CNN showed the highest accuracy for pressure drop, with errors of approximately 0.1% for interpolation and 4.4% for extrapolation. The MGN-GAE showed high vorticity reconstruction accuracy, with coefficients of determination of 0.9994 and 0.9788 for the interpolation and extrapolation cases, respectively. These results show that the dimensionality-reduction models can reproduce not only the flow fields with high accuracy but also important physical flow characteristics, suggesting that the preservation of these characteristics should also be considered when evaluating reduced-order models.

These results also demonstrate that the choice of dimensionality-reduction architecture can directly affect the accuracy and computational efficiency of the subsequent latent-space-based AI surrogate model. Because the overall performance of the surrogate prediction framework depends on how accurately and efficiently high-dimensional CFD flow fields are represented in a low-dimensional latent space, selecting an appropriate ROM architecture is an important prerequisite for constructing an efficient reactor prediction system.

Future work will apply the selected CNN and MGN-GAE architectures to three-dimensional reactor CFD data and couple the resulting low-dimensional latent representations with operator-learning-based AI surrogate models. In the three-dimensional application, heat-transfer characteristics will also be considered by including the temperature field in addition to pressure and velocity. The models will be evaluated not only in terms of reconstruction accuracy but also by examining the preservation of physically relevant quantities such as mass conservation, pressure drop, vortex structures, and heat-transfer characteristics. Through this approach, a reduced-order framework for rapid and accurate prediction of both flow and heat-transfer behavior in latent space will be developed, with the ultimate goal of assessing its applicability to reactor digital-twin systems.

Acknowledgement

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIT) (No. GTL24031-000) and by the National Research Foundation of Korea (NRF)

grant funded by the Korea government (MSIT) (RS-2026-25475350).

REFERENCES

- [1] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine Learning for Fluid Mechanics," *Annual Review of Fluid Mechanics* 52 (2020): 477–508.
- [2] T. Murata, K. Fukami, and K. Fukagata, "Nonlinear mode decomposition with convolutional neural networks for fluid dynamics," *Journal of Fluid Mechanics* 882 (2020): A13.
- [3] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Machine Intelligence* 3 (2021): 218–229.
- [4] K. Kontolati, S. Goswami, G. E. Karniadakis, and M. D. Shields, "Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems," *Nature Communications* 15 (2024): 5101.
- [5] 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* 57, no. 10 (2025): 103703.
- [6] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *International Conference on Learning Representations* (2017).
- [7] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, "Learning Mesh-Based Simulation with Graph Networks," *International Conference on Learning Representations* (2021).
- [8] Q. Li, Z. Han, and X.-M. Wu, "Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning," *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (2018).
- [9] K. K. Kim, W. Lee, S. Choi, H. R. Kim, and J. Ha, "SMART: The First Licensed Advanced Integral Reactor," *Journal of Energy and Power Engineering* 8 (2014).