From Data Assimilation to Machine Learning: Surrogate Modeling from Limited Experimental Data

Nguyen Huu Tiep[1*], Jae-Yong Lee[1], Hae-Yong Jeong[1]

[1]Department of Quantum and Nuclear Engineering, Sejong University, Seoul, Republic of Korea
[*]E-mail: tiepngh@sejong.ac.kr

Keywords: reflood, quenching time, peak cladding temperature, deep neural network, experimental data scarcity.

Accurate estimation of quenching time (QT) and peak cladding temperature (PCT) during reflood transients remains a significant challenge due to the complex nonlinear behavior and multiple boiling heat transfer regimes involved. Although recent machine learning approaches have demonstrated encouraging predictive capability, most have relied on synthetic datasets generated by system thermal-hydraulic codes, causing the resulting models to inherit the assumptions and biases embedded in those simulations instead of learning directly from experimental observations. To address this limitation, this study presents a deep neural network (DNN) framework trained solely on the limited experimental data available from the complete FLECHT SEASET and RBHT reflood test programs. Given the impossibility of generating additional experimental measurements, a systematic multi-facility data fusion strategy is introduced to combine heterogeneous datasets into a unified machine learning database, providing a generalizable approach for data-driven thermal-hydraulic research under severe data scarcity. To maximize learning efficiency from the limited data, a tailored DNN architecture is designed, with its hyperparameters optimized using CASOH, a hybrid sampling-based hyperparameter optimization framework. The optimized model is validated against experimental measurements and benchmarked with predictions from the SPACE and MARS system thermal-hydraulic codes, achieving a maximum coefficient of determination (R^2) of 0.98 for critical safety-related variables. The results indicate that highly accurate prediction is feasible despite the limited availability of experimental data and demonstrate that models trained directly on experimental observations can mitigate the simulation-induced bias commonly found in code-trained machine learning models.