Comparative Evaluation of Data-Driven and Correlation-Guided Physics-Informed Neural Networks for Critical Heat Flux Prediction in Vertical Tubes

Eunhye Yeom [a] , Taeseok Kim[b]*
[a]*Department of Electrical and Energy Engineering, Jeju National University, Jeju, 63243, Republic of Korea*
[b]*Department of Nuclear Engineering, Jeju National University, Jeju, 63243, Republic of Korea*
**Corresponding author: tkim@jejunu.ac.kr*

***Keywords : Critical heat flux, Physics-informed neural network, Deep neural network, Empirical correlation, Thermal-hydraulics**

1. Introduction

Accurate prediction of critical heat flux (CHF) is essential for maintaining thermohydraulic safety margins in nuclear reactor systems, as CHF defines the onset of boiling crisis and departure from nucleate boiling. For decades, CHF estimation has relied on empirical correlations such as the Westinghouse W-3 correlation [1], the Katto–Ohno generalized correlation [2], and the Qu & Mudawar micro-channel correlation [3].

These correlations have been developed based on experimental observations under specific thermohydraulic conditions and demonstrate reliable performance within their calibrated domains. However, their applicability is inherently limited by regime-dependent validity ranges, and their predictive capability often degrades when applied outside these domains. Lookup-table (LUT) approaches extend coverage but remain interpolation-based and discrete in high-dimensional parameter spaces.

Recent advances in machine learning have enabled CHF to be modeled as a continuous nonlinear mapping of operating variables. Purely data-driven deep neural networks (DNNs) provide strong interpolation performance across large datasets. However, due to the absence of explicit thermohydraulic structure, DNNs may exhibit physically inconsistent behavior in sparsely sampled or extrapolation regions.

Physics-informed neural networks (PINNs) have been proposed as an alternative framework that incorporates physical knowledge into the learning process through additional constraints [4]. In CHF modeling, empirical correlations can serve as thermohydraulic priors that guide the learned response surface. While previous studies have demonstrated that such approaches can improve predictive performance, the influence of different types of embedded physical knowledge on the learning behavior of neural networks has not been systematically investigated.

In this study, three representative empirical correlations are selected to reflect distinct thermohydraulic regimes: the Westinghouse W-3 correlation representing conventional high-pressure vertical tube conditions, the Katto–Ohno correlation providing a generalized formulation across a wide range of operating conditions, and the Qu & Mudawar correlation describing micro-channel boiling regimes. This selection enables a structured investigation of how regime-specific physical knowledge influences neural network behavior.

Unlike conventional studies that focus primarily on improving predictive accuracy, this work aims to analyze how different forms of embedded physical knowledge affect the inductive bias of neural networks. In this framework, the DNN provides global approximation flexibility, while empirical correlations introduce localized structural bias. The interaction between these components allows for a systematic evaluation of how physical priors shape model behavior across different regimes.

To this end, a purely data-driven DNN is compared with three correlation-guided PINNs, each incorporating one of the selected empirical correlations. All models share identical neural architectures and differ only in loss formulation, enabling direct assessment of the structural effects induced by different physical priors.

2. Methodology

2.1 Dataset Characteristics and Regime Definition

Figure 1 presents the global distribution of the experimental CHF dataset in the thermohydraulic parameter space.

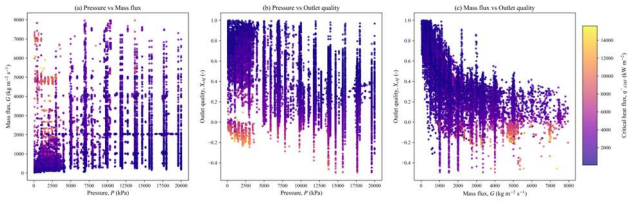

Fig. 1. Global distribution of experimental CHF data. (a) Pressure vs mass flux. (b) Pressure vs outlet quality. (c) Mass flux vs outlet quality [1].

Figures 1(a)–(c) illustrate the coverage of the dataset across pressure, mass flux, and outlet quality domains. The dataset spans a wide range of operating conditions, including both conventional vertical tube regimes and conditions approaching micro-channel flow characteristics. The distributions indicate a high density of data in moderate pressure and mass flux regions, with broader dispersion at higher operating ranges.

While Figures 1(a)–(c) describe the geometric distribution of input variables, they do not directly reflect the thermohydraulic behavior of CHF. To provide physical insight into the dataset, Figure 2 presents the dependence of CHF on outlet quality under varying pressure and mass flux conditions.

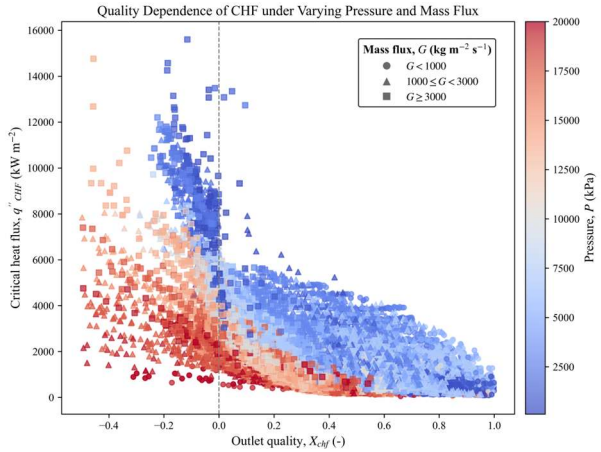

Fig. 2. CHF dependence on outlet quality under varying pressure and mass flux conditions. Color indicates pressure, and marker types represent mass flux ranges [1].

Figure 2 reveals a strong dependence of CHF on outlet quality, characterized by a decreasing trend with increasing quality. In addition, coupled effects of pressure and mass flux are observed, where higher pressure generally reduces CHF, while higher mass flux tends to increase CHF. These trends are consistent with known thermohydraulic behavior and highlight the nonlinear interactions among operating variables.

Based on these observations, two representative regimes are defined for structured evaluation. The tube regime corresponds to conventional high-pressure vertical tube conditions where correlations such as W-3 are applicable. The micro-channel regime represents operating conditions consistent with the Qu & Mudawar correlation domain, characterized by relatively high mass flux and distinct scaling behavior.

This regime-based classification enables systematic analysis of how different embedded physical priors influence model behavior across thermohydraulic conditions.

2.2 Empirical Correlations

Three empirical models are implemented as standalone benchmarks and as structural priors within the PINN framework: the Westinghouse W-3 empirical correlation, the Katto–Ohno generalized CHF correlation, and the Qu & Mudawar micro-channel CHF correlation.

Each correlation represents distinct thermohydraulic regimes and underlying physical assumptions. The W-3 correlation is developed for high-pressure vertical tube conditions and reflects strong pressure-dependent CHF behavior. The Katto–Ohno correlation provides a generalized formulation applicable over a wide range of operating conditions, capturing broader trends across multiple regimes. The Qu & Mudawar correlation is specifically designed for micro-channel flows and incorporates dominant mechanisms at small hydraulic diameters.

This selection enables a systematic investigation of how different types of embedded physical knowledge influence the inductive bias of neural networks and their regime-dependent behavior.

2.3 Purely Data-Driven Deep Neural Network

The baseline model is a fully connected purely data-driven deep neural network (DNN), which serves as a reference for global nonlinear regression across the entire dataset. The architecture remains identical across all experiments to ensure fair comparison.

The network is trained to minimize the mean squared error between predicted and experimental CHF values without incorporating any explicit physical constraints. As a result, the DNN prioritizes fitting the dominant data distribution, achieving strong interpolation performance. However, due to the absence of structural guidance, the model may exhibit physically inconsistent behavior in sparsely sampled or extrapolation regions.

2.4 Correlation-Guided PINN Framework

Figure 3 presents the architecture of the correlation-guided PINN.

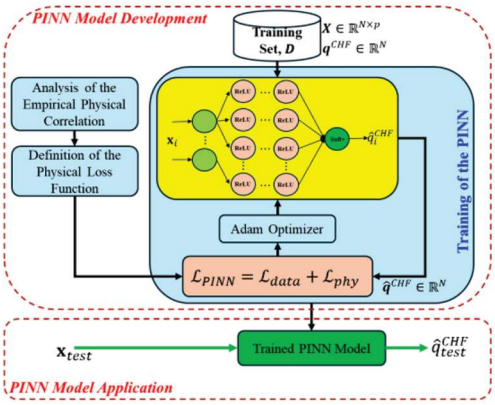

Fig. 3. Architecture of the correlation-guided physics-informed neural network (PINN) framework. The total loss consists of data loss and gradient-based physics constraint derived from an embedded empirical correlation [5].

The PINN structure employs the same neural backbone as the DNN but incorporates an augmented loss formulation to embed physical knowledge. The total loss is defined as:

$$L_{total} = L_{data} + \lambda L_{physi}$$

where L_{data} represents the data-driven loss and $L_{physics}$ represents a physics-based constraint derived from the selected empirical correlation. The weighting parameter λ controls the relative importance of the physics constraint and data-driven learning.

Instead of enforcing direct value matching between the model prediction and the empirical correlation, a partial-derivative-based constraint is applied. Following previous studies [5,6], the gradients of the predicted CHF with respect to selected input variables are aligned with those derived from the embedded correlation. This approach has been shown to provide more stable and physically consistent learning, particularly in capturing trend behavior.

This formulation allows the model to retain the global flexibility of data-driven learning while introducing localized structural guidance from empirical correlations. As a result, the embedded correlation acts as a structural prior that influences the inductive bias of the neural network.

3. Results

3.1 Regime-Specific Behavior: Tube Regime

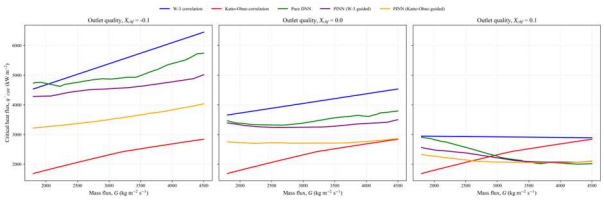

Fig. 4. Model predictions in the tube regime at representative outlet quality conditions (X_{chf} = −0.1, 0.0, 0.1) as functions of mass flux. Comparison between empirical correlations (W-3, Katto–Ohno), purely data-driven DNN, and correlation-guided PINNs.

Figure 4 presents model predictions in the tube regime at representative outlet quality conditions (X_{chf} = −0.1, 0.0, 0.1) as functions of mass flux.

The Westinghouse W-3 correlation exhibits a strong monotonic increase of CHF with mass flux, reflecting its pressure-dependent scaling characteristics. In contrast, the Katto–Ohno correlation shows a comparatively milder slope, indicating a different sensitivity to mass flux under the same thermohydraulic conditions. These differences highlight the distinct structural assumptions embedded in each empirical model.

The purely data-driven DNN produces smooth interpolation trends that closely follow the dominant data distribution. However, its behavior does not consistently reflect the structural trends implied by the empirical correlations, particularly in regions where data are sparse.

The correlation-guided PINNs demonstrate intermediate behavior between purely empirical and purely data-driven models. The W-3-guided PINN preserves the increasing trend of CHF with mass flux while moderating the steepness observed in the standalone correlation. Similarly, the Katto–Ohno-guided PINN reflects the milder scaling behavior of the corresponding correlation while maintaining smoothness consistent with data-driven learning.

Although the PINNs exhibit slightly higher prediction errors compared to the DNN, they provide improved structural consistency by incorporating correlation-based physical guidance. These results indicate that the embedded correlation functions as a structural prior that directly influences the inductive bias of the neural network in a regime-dependent manner.

3.2 Regime-Specific Behavior: Micro-Channel Regime

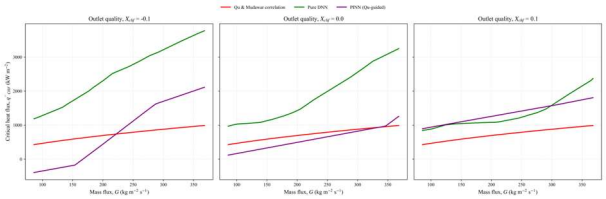

Fig. 5. Model predictions in the micro-channel regime corresponding to the Qu & Mudawar correlation domain. CHF variation with mass flux at representative outlet quality conditions (X_{chf} = −0.1, 0.0, 0.1). Comparison between empirical correlation, purely data-driven DNN, and correlation-guided PINN.

Figure 5 presents model predictions in the micro-channel regime corresponding to the operating domain of the Qu & Mudawar correlation.

The standalone Qu & Mudawar correlation shows moderate scaling of CHF with mass flux, reflecting the underlying physics of micro-channel boiling. In contrast, the purely data-driven DNN exhibits a stronger dependence on mass flux, driven by global patterns in the dataset rather than regime-specific behavior.

While the DNN captures a smooth interpolation trend, it does not explicitly preserve the qualitative structural characteristics associated with micro-channel CHF behavior. This discrepancy becomes more evident at higher mass flux conditions, where deviations from the expected trend can be observed.

The Qu-guided PINN demonstrates intermediate behavior, aligning more closely with the empirical correlation while maintaining smooth transitions across the input space. Compared to the standalone correlation, the PINN reduces extreme deviations and improves consistency with the broader dataset.

As in the tube regime, the PINN shows slightly higher prediction errors compared to the DNN but better preserves physically meaningful trends. This highlights the role of the embedded correlation as a structural constraint that guides the model toward regime-consistent behavior while maintaining data-driven flexibility.

3.3 Global Parity Analysis

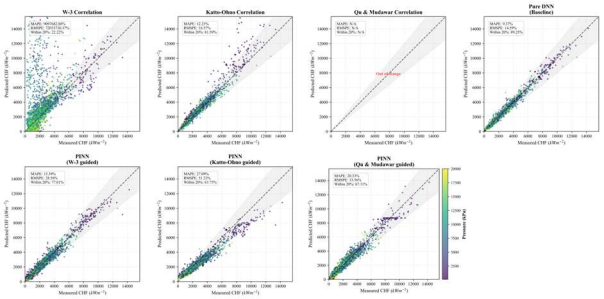

Fig. 6. Global predicted-versus-experimental parity plots for all models across the full dataset. Dashed lines indicate ±20% error bounds.

Figure 6 presents global predicted-versus-experimental parity plots for all models across the full dataset. The standalone empirical correlations exhibit substantial dispersion when evaluated globally, reflecting their limited applicability outside their original calibration domains. In particular, the W-3 correlation shows significant deviations due to its strong sensitivity to specific input conditions, while the Qu & Mudawar correlation is not applicable across the full dataset.

The purely data-driven DNN demonstrates the best overall predictive performance, with most predictions falling within the ±20% error band. The correlation-guided PINNs achieve comparable but slightly lower accuracy than the DNN due to the imposed physical constraints. However, they exhibit improved structural consistency by reducing extreme deviations observed in standalone correlations and maintaining physically meaningful trends. The W-3-guided and Qu-guided PINNs show stable behavior across a wide range of conditions, while the Katto–Ohno-guided PINN reflects moderate sensitivity influenced by the generalized nature of the embedded correlation.

Table I: Quantitative performance metrics (MAPE and RMSPE) for all models

Model	MAPE (%)	RMSPE (%)
W-3 Correlation	9097682.88	72033730.37
Katto–Ohno Correlation	12.23	16.57
Qu & Mudawar Correlation	—	—
Pure DNN (Baseline)	9.37	14.59
PINN (W-3 guided)	15.39	28.58
PINN (Katto–Ohno guided)	27.09	51.23
PINN (Qu guided)	20.33	33.56

To quantitatively evaluate these observations, model performance is further assessed using mean absolute percentage error (MAPE) and root mean square percentage error (RMSPE), as summarized in Table I [7].

The purely data-driven DNN achieves the lowest prediction error, with a MAPE of 9.37% and an RMSPE of 14.59%, confirming its strong capability in capturing global nonlinear relationships. Among the standalone empirical correlations, the Katto–Ohno correlation shows reasonable performance (MAPE of 12.23% and RMSPE of 16.57%). Conversely, the W-3 correlation exhibits extremely large errors due to numerical instability outside its validity range, and the Qu & Mudawar correlation fails to provide valid predictions for a significant portion of the dataset, resulting in undefined error values.

The correlation-guided PINNs exhibit higher errors compared to the DNN, with MAPE values ranging from 15.39% to 27.09%, depending on the embedded correlation. However, a critical finding is that unlike the standalone correlations, the PINNs maintain stable and finite predictions across all conditions, effectively avoiding the severe numerical divergence observed in the empirical models.

These results further support the existence of a necessary trade-off between strict predictive accuracy and physical consistency. While the DNN minimizes global prediction error, the PINNs introduce structured behavior through embedded correlations and maintain robust predictive capabilities even when the underlying empirical models become completely unstable.

4. Discussion

The results presented in this study provide insight into how different types of embedded empirical correlations influence the behavior of physics-informed neural networks in CHF prediction.

The comparison across tube and micro-channel regimes indicates that the choice of correlation directly affects the structural response of the model. In the tube regime, correlations such as W-3 and Katto–Ohno introduce distinct scaling behaviors with respect to mass flux, and these characteristics are reflected in the corresponding PINN predictions. Similarly, in the micro-channel regime, the Qu & Mudawar correlation guides the model toward trend behavior consistent with micro-scale boiling physics.

At the same time, the results demonstrate that the effectiveness of such physical guidance depends strongly on the compatibility between the embedded correlation and the underlying data regime. When a correlation is applied outside its validity domain, the imposed physics constraint does not necessarily improve predictive accuracy and may instead introduce deviations from the dominant data distribution. This behavior is particularly evident in the micro-channel analysis, where the global data-driven DNN captures broader trends, while the Qu-guided PINN aligns more closely with the correlation but exhibits reduced global accuracy.

These observations suggest that, within the PINN framework, embedded physical knowledge does not universally enhance model performance. Rather, it acts as a structural bias that can either improve or degrade prediction quality depending on its consistency with the data domain. However, it is noteworthy that the PINN

architecture demonstrates inherent robustness; even when standalone empirical correlations fail to converge or diverge numerically outside their validity domains, the corresponding PINNs maintain stable predictions by leveraging data-driven regularization. In this sense, an inappropriate physical constraint may function as an additional source of bias, but the neural network structure prevents catastrophic numerical instability.

Therefore, the results highlight that the selection of empirical correlations should be treated as a critical design choice in physics-informed learning. The findings indicate that physically informed models require not only accurate data representation but also careful alignment between the chosen physical priors and the operating regime of interest.

From this perspective, the present study can be interpreted as a preliminary step toward more advanced frameworks in which multiple correlations are selectively integrated based on their applicable regimes. Such approaches may enable adaptive weighting of physical constraints, allowing the model to leverage appropriate physical knowledge while reducing the influence of incompatible correlations.

5. Conclusion

This study presented a comparative evaluation of a purely data-driven deep neural network and correlation-guided physics-informed neural networks for CHF prediction across multiple thermohydraulic regimes.

While the purely data-driven DNN achieved the highest global predictive accuracy, the PINN models introduced structured, regime-specific physical trends through embedded empirical correlations.

A key finding of this work is that the incorporation of physical knowledge acts as a strong structural bias rather than a guaranteed performance enhancement. The effectiveness of physics-informed learning depends on the compatibility between the embedded physical prior and the data distribution. When applied outside their validity domains, empirical correlations introduce incompatible biases that reduce predictive accuracy.

Nevertheless, the PINN framework demonstrates notable robustness by preventing the severe numerical divergence and calculation failures often observed in standalone empirical correlations under extreme conditions.

Ultimately, this study shows that empirical correlations should not be indiscriminately incorporated into neural networks. It provides a foundation for future PINN frameworks that selectively integrate multiple correlations with regime-dependent weighting, ensuring that physical knowledge is applied only where it is appropriate.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by Korea government (MSIT) (RS-2025-02310831)

REFERENCES

[1] Groeneveld, D.C., Critical Heat Flux Data Used to Generate the 2006 Groeneveld Lookup Tables, NUREG/KM-0011, U.S. Nuclear Regulatory Commission, 2019.
[2] Katto, Y., Ohno, H., An improved version of the generalized correlation of critical heat flux for the forced convective boiling in uniformly heated vertical tubes, International Journal of Heat and Mass Transfer, 27 (9) (1984) 1641–1648.
[3] Qu, W., Mudawar, I., Measurement and correlation of critical heat flux in two-phase micro-channel heat sinks, International Journal of Heat and Mass Transfer, 47 (2004) 2045–2059.
[4] Raissi, M., Perdikaris, P., Karniadakis, G.E., Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 378 (2019) 686–707.
[5] Ahmed, I., Gatti, I., Zio, E., Prediction of Critical Heat Flux in Vertical Tubes by Physics-informed Neural Networks, Proceedings of the 35th European Safety and Reliability Conference (ESREL SRA-E 2025), 2025.
[6] Ahmed, I., Gatti, I., Zio, E., Optimized ensemble of neural networks for the prediction of critical heat flux, Nuclear Engineering and Design, 439 (2025) 114111.
[7] OECD Nuclear Energy Agency, Benchmark on Artificial Intelligence and Machine Learning for CHF, NEA/WKP(2023)1, 2023.