2026 International Nuclear Atlantic Conference – INAC 2026

Escola de Guerra Naval, Rio de Janeiro, RJ, August 24 - 28, 2026

ISBN: 978-65-272-2398-6

Predicting Critical Heat Flux in Water-Cooled Vertical Tubes: An Exploratory Data Analysis and Comparative Machine Learning Modeling

Aron P. V. Gomes[1] and Marcelo da Rocha Silva[2]

[1] *aronpedro@usp.br*
[2] *msrocha.ipen@gmail.com*

1. Introduction

In thermal-hydraulic safety the critical heat flux (CHF) is a fundamental limit in water-cooled nuclear reactors. It is the heat flux at which the efficient nucleate boiling regime is replaced by film boiling, causing a sharp drop in the heat transfer coefficient between the heated surface and the coolant. In a nuclear reactor, exceeding the local CHF leads to a rapid excursion in cladding temperature that can compromise fuel integrity and, in extreme cases, initiate severe accident sequences [1].

The most widely adopted CHF prediction tool in nuclear thermal-hydraulics is the Critical Heat Flux Look-Up Table (LUT) developed by Groeneveld and collaborators. The 2006 version of the LUT is structured as a three-dimensional matrix indexed by pressure (P), mass flux (G) and thermodynamic quality (X_{chf}) for a reference tube diameter of 8 mm, and provides CHF predictions with a root-mean-square error of approximately 7.10% under constant-inlet conditions when compared against the screened experimental database used in its derivation [2]. In 2019, the United States Nuclear Regulatory Commission (USNRC) consolidated this experimental database in the NUREG/KM-0011 report, which contains over 25,000 CHF data points obtained in water-cooled vertical tubes across more than 60 years of experimentation in 62 independent sources [3].

The advancement and ease of implementation of machine learning tools has motivated investigations and reviews of historically known CHF databases [4,5,6]. These approaches allow for the analysis of all experimental variables, avoiding the need for empirical correlations or interpolation grids, which are commonly used in LUT. Therefore, a systematic comparison of classical ML models trained on the full database is useful both as a benchmark for more sophisticated approaches and as a didactic foundation for understanding the predictive structure of the data.

In this context, this research presents an exploratory analysis of the NUREG/KM-0011 CHF database and a comparative evaluation of three supervised machine learning models, multiple linear regression, decision tree, and k-nearest neighbors, for CHF prediction in water-cooled vertical tubes. Model performance is assessed through five-fold cross-validation using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and coefficient of determination (R^2) as evaluation metrics.

2. Methodology

This paper proposes a comparative methodology for evaluating supervised machine learning models for CHF prediction in water-cooled vertical tubes, as illustrated in Fig. 1.

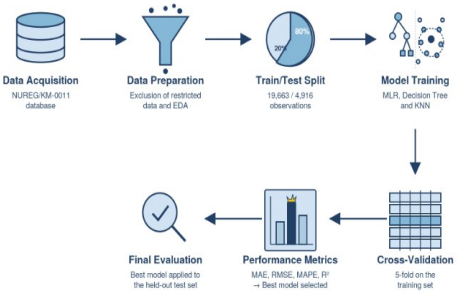


Figure 1: Overview of methodological steps for CHF prediction.

The dataset used in this study was obtained from the NUREG/KM-0011 report and comprises 25,540 experimental CHF observations from 62 independent sources, collected across more than 60 years of CHF experimentation in water-cooled vertical tubes. From the original records, 961 data points classified as restricted or proprietary were excluded, resulting in a working dataset of 24,579 observations. Each observation contains seven features — tube diameter D, heated length L, pressure P, mass flux G, critical thermodynamic quality X_{chf}, inlet enthalpy rise (ΔH_{in}) and inlet temperature T_{in} and one target variable, the critical heat flux (CHF).

The exploratory data analysis included univariate statistics, boxplot-based detection of moderate and critical outliers by the interquartile ranges method, and pairwise correlation analysis using the Pearson coefficient. No critical outliers were removed prior to modeling, as their physical plausibility could not be ruled out within the operating ranges reported in the database. The dataset was randomly split into a training set comprising 80% of the observations (19,663 points) and a held-out test set comprising the remaining 20% (4,916 points) using a fixed random seed for reproducibility. Three supervised machine learning models were investigated: multiple linear regression (MLR), decision tree (DT), and k-nearest neighbors (KNN). The MLR model was implemented in two configurations: a full model using all seven input variables and a reduced model excluding T_{in} and D, which presented redundancy with ΔH_{in} and weak Pearson correlation with the target, respectively. The decision tree was trained using all seven features with a maximum depth of nine, selected to balance predictive performance and tree interpretability. For the KNN model, input variables were standardized via z-score transformation, the euclidean distance was adopted as the dissimilarity measure, and the optimal number of neighbors was selected by grid search over the set k = {1, 3, 5, 7, 10, 15, 20, 30, 50, 75, 100}, based on cross-validation performance .

Model performance was assessed through stratified five-fold cross-validation on the training set, which provides a robust estimate of generalization error while mitigating the influence of a single train-test partition. Four standard regression metrics were computed, as defined in Eqs. (1) to (4):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|q_{i\,CHF,\exp} - q_{i\,CHF,pred}\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(q_{i\,CHF,\exp} - q_{i\,CHF,pred}\right)^2} \tag{2}$$

$$MAPE = \frac{100\,\%}{N}\sum_{i=1}^{N}\left|\frac{q_{i\,CHF,\exp} - q_{i\,CHF,pred}}{q_{i\,CHF,\exp}}\right| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(q_{i\,CHF,\exp} - q_{i\,CHF,pred}\right)^2}{\sum_{i=1}^{N}\left(q_{i\,CHF,\exp} - q_{avg\,CHF,\exp}\right)^2} \tag{4}$$

Where $q_{i\,CHF,exp}$ is the experimentally measured critical heat flux , $q_{i\,CHF,pred}$ is the corresponding value predicted by the model for the i-th observation and N is the number of observations in the evaluation set and $q_{avg\,CHF,exp}$ is the average of the experimentally measured values . The MAE and RMSE quantify the average prediction error in absolute terms (kW/m²), with RMSE being more sensitive to large deviations; the MAPE expresses the relative error as a percentage, enabling comparison across different magnitude ranges; and the coefficient of determination R^2 measures the proportion of the variance in the experimental data explained by the model, ranging from 0 to 1. Based on the cross-validation results, the best-performing model was selected and subsequently applied to the held-out test set to confirm the cross-validation estimates and to verify the absence of overfitting [7].

3. Results and Discussion

The exploratory data analysis revealed that the experimental CHF values span a wide range, from approximately 50 to over 16,000 kW/m², reflecting the broad diversity of operating conditions compiled in the database. The Pearson correlation analysis indicated that the critical thermodynamic quality (X_{chf}), heated length (L) and the mass flux (G) presented the strongest linear associations with CHF, while the inlet temperature (T_{in}) and inlet enthalpy rise (ΔH_{in}) exhibited a strong mutual correlation, confirming their physical redundancy. A negative correlation between CHF and pressure ($r = -0.25$) was observed, which is discussed below in light of the underlying physical mechanisms.

The performance of the three supervised models, evaluated through five-fold cross-validation on the training set, is summarized in Table I. The k-nearest neighbors (KNN) model achieved the best results across all four metrics, substantially outperforming the decision tree and the multiple linear regression configurations. The poor performance of the linear models ($R^2 \approx 0.70$) confirms that the relationship between the operating variables and CHF is strongly nonlinear, as expected from the physics of the boiling crisis. The decision tree captured part of this nonlinearity ($R^2 = 0.791$) but remained considerably less accurate than the instance-based approach.

Table I: Cross-validation performance of the three supervised models.

Model	MAE(kW/m²)	MAPE (%)	RMSE (kW/m²)	R^2
KNN (k=50)	171	11.4	320	0.962
Decision Tree	528	50.2	721	0.791
MLR (full, 7 var.)	602	66.4	865	0.700
MLR (5 var.)	603	66.6	866	0.699

Among the values evaluated in the grid search, k = 50 was selected as it provided the best balance between predictive accuracy and generalization, yielding a train-to-cross-validation RMSE ratio of approximately 1.14, indicative of low overfitting. The selected model was subsequently applied to the held-out test set, which was not used during training or model selection. As shown in Table II, the test-set performance closely reproduced the cross-validation estimates, with deviations below 6% across all metrics, confirming the generalization capability of the model and the absence of significant overfitting.

Table II: Performance of the selected KNN model.

Metric	Cross-Validation	Test Set	Difference (%)
MAE(kW/m²)	171	163	-4.3
MAPE (%)	11.4	10.7	-5.3
RMSE (kW/m²)	320	328	+2.4
R^2	0.962	0.965	+0.3

The final test-set MAPE of 10.7% is of the same order of magnitude as the root-mean-square error of approximately 7.10% reported for the 2006 CHF Look-Up Table under constant-inlet conditions [2]. The LUT was derived from a screened subset of the experimental database, in which approximately 25% of the originally available points were removed based on physical criteria such as diameter range, heated-length-to-diameter ratio, heat-balance consistency, and slice-method outlier identification [2]. Although the NUREG/KM-0011 database already incorporates most of this filtering, the screening procedure was not fully reproduced in the present work. The present database is therefore broadly equivalent to, but not strictly identical with, the subset used to construct the LUT, which supports the validity of the accuracy comparison while justifying the caution in its interpretation. The fact that a simple instance-based model trained on the full, less-filtered dataset achieves a comparable order of accuracy highlights the potential of data-driven approaches as complementary tools to empirical look-up tables.

The moderate negative correlation between CHF and pressure observed in the exploratory analysis reflects the well-known non-monotonic dependence of CHF on operating pressure. At low pressures, increasing pressure raises CHF due to reduced bubble size and increased bubble departure frequency; beyond a maximum near 3 MPa for water, CHF decreases as the vapor density rises, the latent heat of vaporization declines, and the surface tension diminishes, all of which facilitate the

formation of an insulating vapor layer on the heated surface. Since the database is dominated by pressures relevant to reactor operation, the aggregated correlation is negative [8].

Finally, an analysis of the residuals of the selected model indicated a tendency to slightly underestimate CHF at high values. From a thermal-hydraulic safety perspective, this behavior is conservative: a model that systematically predicts CHF below the true value yields a conservative estimate of the Departure from Nucleate Boiling Ratio, contributing to a safer margin in reactor design rather than eroding it [1].

4. Conclusions

The results showed that the k-nearest neighbors model (k = 50) clearly outperformed the other approaches, achieving a coefficient of determination of 0.962 in cross-validation and 0.965 on the test set, with a test-set mean absolute percentage error of 10.7%. The close agreement between cross-validation and test-set performance confirmed the generalization capability of the model and the absence of significant overfitting. The poor performance of the linear models confirmed the strongly nonlinear nature of the relationship between the operating variables and CHF, while the instance-based approach proved well-suited to capturing the local structure of the experimental data. Notably, the accuracy obtained was of the same order of magnitude as that reported for the 2006 CHF Look-Up Table, even though the present model was trained on the full, less rigorously screened database.

One limitation noted for the training of the models is the lack of a considerable volume of data for high CHF values, which hinders the model for more general use. For future work, it is recommended to use physics-informed machine learning to improve the prediction of CHF at points that still require experimental information [6]. Overall, the results indicate that supervised machine learning models, even in their simpler forms, constitute a promising complementary tool to empirical look-up tables for critical heat flux prediction in nuclear thermal-hydraulics.

Acknowledgements

The authors gratefully acknowledge Prof. Renan Favarão da Silva for his guidance on the application of data science techniques. The authors also wish to thank Admiral André Luis Ferreira Marques for the many insights that greatly enriched this research.

References

[1] N. E. Todreas and M. S. Kazimi, *Nuclear Systems Volume I: Thermal Hydraulic Fundamentals*, 1st ed., Hemisphere Publishing Corporation, New York, USA (1989).
[2] D. C. Groeneveld, J. Q. Shan, A. Z. Vasic, L. K. H. Leung, A. Durmayaz, J. Yang, S. C. Cheng and A. Tanase, “The 2006 CHF look-up table,” *Nuclear Engineering and Design*, vol. 237, pp. 1909–1922 (2007).
[3] U.S. Nuclear Regulatory Commission, "Critical Heat Flux Data Used to Generate the 2006 Groeneveld Lookup Tables," NUREG/KM-0011, Washington, USA (2019).
[4] H. Liu, H. Jiang, D. Chen, J. Qin, Y. Wang and M. Liu, “Critical heat flux prediction for annular channel through application of machine learning techniques,” *International Communications in Heat and Mass Transfer*, vol. 167, 109279 (2025).
[5] S. Yang, X. Li, B. Ren, L. Yang, C. Chen, Q. Lu and Z. Wei, “A transfer learning framework for critical heat flux prediction pretrained on the 2006 lookup table,” *Applied Thermal Engineering*, vol. 287, 129451 (2026).
[6] L. Corre, J.-Marie, et al., "Benchmark on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering. Phase 1: Critical Heat Flux Exercise Specifications," *NEA Working Papers*, OECD Publishing, Paris, France (2024).
[7] G. James, D. Witten, T. Hastie and R. Tibshirani, An Introduction to Statistical Learning with Applications in R, 2nd ed., Springer, New York, USA (2021).
[8] M. S. Greenwood, The Critical Heat Flux at High Pressure and Low Mass Flux Prototypic of Proposed Light Water Small Modular Reactors, Ph.D. dissertation, University of Wisconsin–Madison, Madison, USA (2015).