Nuclear Engineering and Design 439 (2025) 114111

Contents lists available at ScienceDirect

Nuclear Engineering and Design

journal homepage: www.elsevier.com/locate/nucengdes

Optimized ensemble of neural networks for the prediction of critical heat flux

Ibrahim Ahmed [a,*], Irene Gatti [a], Enrico Zio [a,b]

[a] Department of Energy, Politecnico di Milano, Milano, Italy
[b] MINES Paris-PSL University, CRC, Sophia Antipolis, France

ARTICLE INFO

Keywords:
Departure from nucleate boiling
Critical heat flux
Nuclear Reactors
Neural Networks
Ensemble models

ABSTRACT

Critical Heat Flux (CHF) is a thermal limit in boiling heat transfer, beyond which there is a substantial reduction in heat transfer efficiency. This phenomenon plays a vital role in the thermal engineering design of systems involving two-phase flow. As a result, an accurate CHF prediction is essential for both safety and performance, particularly in water-cooled nuclear reactors where thermohydraulic margins are critical. In this paper, a novel optimized ensemble of neural networks (NNs) for CHF prediction is proposed to enhance the accuracy of individual models trained separately with distinct architectures and hyperparameters settings. Two systematic procedures are presented to identify potentially optimal NN models and aggregate them into an optimal ensemble model. The proposed method is validated using experimental CHF data made available by the Working Party on Scientific Issues and Uncertainty Analysis of Reactor Systems (WPRS) Expert Group on Reactor Systems Multi-Physics (EGMUP) task force on AI and ML for Scientific Computing in Nuclear Engineering projects, promoted by the OECD/NEA. The results obtained show that the ensemble model outperforms standalone models and other state-of-the-art modelling approaches. Parametric and sensitivity analyses across various input parameters confirm the robustness of the ensemble model and its consistency with expected physical behaviors, further underlying its potential for improving CHF prediction in nuclear reactor applications.

1. Introduction

When designing thermal systems that experience two-phase flow, the concept of Critical Heat Flux (CHF) is of paramount importance. CHF is a thermal limit in boiling heat transfer, beyond which there is a drastic reduction in heat transfer efficiency, leading to a significant increase in surface temperature (Todreas and Kazimi, 2011). It is an heat flux at which a transition from nucleate boiling to film boiling (commonly known as the “boiling crisis”) occurs (Bruder et al., 2017). This phenomenon leads to a sharp deterioration in the heat transfer coefficient, resulting in system overheating, potential failure, and phase changes which can cause unpredicted behavior. For this reason, CHF is considered a critical factor in the safety analysis and technical design of efficient thermal systems, such as water-cooled nuclear reactors, steam generators and cooling devices (Chang and Baek, 2003).

During the design of power generation systems, particularly water-cooled nuclear power plants (NPPs), maximizing efficiency is essential and thermohydraulic margins, such as in relation to CHF, play a significant role in achieving safe operation and optimal performance (Chang and Baek, 2003). Specifically, CHF is an essential parameter in nuclear reactor design and it requires careful control and regulation within defined safety limits. The departure from nucleate boiling ratio (DNBR) is the ratio of expected CHF to actual operating local heat flux. For safe operation, the minimum DNBR (MDNBR) must be greater than one to provide a reliable safety margin against heat flux escalation and potential boiling crisis. Despite that the specific MDNBR value varies based on reactor design and the CHF model employed, the U.S. Nuclear Regulatory Commission (NRC) mandates a minimum MDNBR of 1.3 with respect to W-3 correlations (Todreas and Kazimi, 2011). Regarding the underlying mechanism of CHF, most researchers agree that flow conditions near the heated surface are responsible for initiating CHF (Bruder et al., 2017). Yet, the complex nature of two-phase flow interactions and a lack of general consensus on the mechanism triggering CHF (Celata et al., 1994) have impeded the development of a comprehensive physical description of the underlying phenomenon. Research over the years has, rather, resulted in a number of correlations, which can provide reasonably accurate predictions under stated conditions (Bruder et al., 2017).

* Corresponding author.
E-mail address: ibrahim.ahmed@polimi.it (I. Ahmed).

https://doi.org/10.1016/j.nucengdes.2025.114111
Received 22 November 2024; Received in revised form 14 April 2025; Accepted 28 April 2025
Available online 6 May 2025

The earliest CHF prediction models were empirical correlations (Katto, 1992; Tong, 1967; Bowring, 1973; Biasi et al., 1968) and mechanistic models (Celata et al., 1999; Kataoka et al., 2000; Whalley, 1977; Lee and Mudawar, 1988; Weisman and Pei, 1983). Empirical correlations, typically derived from experimental data, are straightforward to implement and show good agreement when applied under operating conditions similar to the experimental ones. However, these correlations are often limited in validity by their reliance on empirically estimated parameters, which restrict their applicability to the range of the original experimental data used to set up and calibrate the correlations themselves. On the contrary, mechanistic models incorporate aspects of the underlying physics such as flow pattern, bubble size and velocity, but are limited in use, given the CHF phenomenon is not yet well understood, as mentioned above. As a result, the level of accuracy required for nuclear reactor design and safety analysis cannot be achieved with mechanistic models alone and empirical parameters are introduced, whose values are estimated from design-specific data (Celata et al., 1994).

Given the complexity of CHF prediction, especially under transient conditions or non-uniform distribution of heat flux in rod bundles, empirical methods such as CHF look-up tables (LUTs) have gained prominence. The first attempt to create CHF LUT was made by Doroshchuk in 1975 (Doroshchuk et al., 1975) using a dataset of 5000 data points. This work has been continuously updated by incorporating new available experimental data. The most significant effort in this regard was led by Groeneveld et al. in 1995 (Groeneveld et al., 1996), which led to the widely adopted 2006 Groeneveld CHF LUT (Groeneveld et al., 2007). This LUT correlates CHF values over a broad range of conditions based on coolant pressure, mass flux and thermodynamic quality (Groeneveld et al., 2007). Although Groeneveld 2006 CHF LUT is based on a large dataset, offers ease of use and covers a wide range of applications, it also has limitations, including a lack of physical insight and gaps within specific regions of the data domain. Despite these limitations, the Groeneveld 2006 CHF LUT remains a reliable and widely adopted prediction model for applications in the field of nuclear engineering (Groeneveld, 2019).

Recent advances in artificial intelligence (AI) and machine learning (ML) have sparked innovative approaches for tackling complex phenomena like CHF prediction. In (Mazzola, 1997), artificial neural networks (ANNs) were used to predict empirical correlations parameters traditionally obtained through best-fit techniques. This approach preserved the advantages of analytical analysis while partially overcoming the black-box nature of direct ANN applications, which were typically restricted to correlation tuning. In (Lee et al., 2006), a backpropagation neural network (BPN) was employed to predict the MDNBR. A thermal–hydraulic (TH) code was first developed to simulate a Total Loss of Flow Accident (TLOFA), a critical scenario that produces the most adverse MDNBR. Then, the outputs from the TH code were used as inputs to auto-associative neural network (AANN) to reconstruct their refined values. Finally, the reconstructed values were used by the BPN to predict the MDNBR and the obtained results demonstrated a high accuracy and reliability of prediction. In (Jiang and Zhao, 2013), a hybrid model was developed by combining support vector regression (ν-SVR) with radial basis function networks (RBFNs) to predict CHF; the obtained results showed superior performance compared to standard SVR and empirical correlations approaches. In (He and Lee, 2018), the work on ν-SVR was extended to address sparsely distributed CHF data and obtained improved accuracy by emphasizing training near critical inflection points. In (Khalid et al., 2023), the performances of standalone ANNs, support vector machines (SVMs) and random forest (RF) were investigated. It was found that hybrid models that combined ANNs with LUTs outperformed standalone models. In (Zhao et al., 2020), a hybrid framework that integrated ML with domain knowledge (physics-informed ML) was developed. This approach extended the applicability of the model and improved both generalization capabilities and predictive performance compared to traditional models. In (Mao and Jin, 2024), ML models were combined with physical models to develop physics-informed ML (PIML) for CHF prediction, and the results showed that LUT-informed NN was the most stable and robust model. In (Zhou et al., 2024), state-of-the-art AI methods, including ANNs, convolutional neural networks (CNNs) and Transformer models, were recently employed for CHF prediction. The proposed AI-based CHF prediction model, particularly the Transformer model utilizing self-attention mechanisms, dynamically assigns importance to different parts of the input data, thereby enhancing its predictive capability. The results demonstrate that the Transformer-based AI model outperforms the LUT method and other state-of-the-art AI-based approaches, including a benchmark model from the OECD-NEA, based on a comprehensive database containing 24,579 CHF data points. Recently, in (Zubair Khalid et al., 2024), an ensemble of deep sparse autoencoders (AEs) for feature extraction was developed and coupled with a deep neural network (DNN) as a *meta*-learner for CHF prediction. This method addressed the limitations of previous models by leveraging a large dataset that covered a wide range of operating conditions, thereby significantly enhancing prediction accuracy and reliability.

In the reviewed literature, most CHF prediction models are standalone models. Although some studies have explored hybrid or ensemble approaches that combine two or more distinct models to enhance CHF prediction accuracy, there has been little investigation into the optimization of those ensemble models. Generally, combining multiple models has been shown to improve accuracy (Zubair Khalid et al., 2024). The combination of the individual models must be done in a way to avoid that *1*) the improvement in performance is achieved by an unnecessarily complex ensemble model that also risks overfitting, and *2*) a simple ensemble model is used but it achieves only a minimal improvement in performance. Therefore, the task of optimization of the ensemble model is important to balance prediction accuracy improvement with model complexity.

The present work proposes an optimized ensemble model for CHF predictions that enhances prediction accuracy and mitigates model complexity. The proposed optimized modelling approach involves developing several individual NN models systematically, each one trained and optimized separately with distinct architectures and hyperparameters settings; these individual models are, then, optimized collectively to form an optimal ensemble model. Two systematic procedures are introduced: *1*) a procedure for automatically identifying optimal NN models; and *2*) a procedure for aggregating optimized individual NN models into an optimal ensemble model. The implementation of the two procedures leads to the identification of the optimal ensemble configuration without having to rely on computationally intensive optimization techniques. The proposed method is validated using experimental CHF data published in (Groeneveld, 2019) and made available by the Working Party on Scientific Issues and Uncertainty Analysis of Reactor Systems (WPRS) Expert Group on Reactor Systems Multi-Physics (EGMUP) task force on AI and ML for Scientific Computing in Nuclear Engineering projects, promoted by the OECD/NEA (Le-Corre et al., 2024). The obtained results show the effectiveness of the optimized ensemble model, which outperforms standalone and other state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 presents the problem formulation. The proposed method for CHF prediction is described in Section 3 and the metrics used to evaluate its performance are presented in Sections 4. The case study is described in Section 5. In Section 6, the results are presented and discussed. Finally, Section 7 presents the conclusions of the work performed.

2. Problem formulation

We consider a heat transfer system of a water-cooled nuclear reactor at the design phase, in which accurate prediction of CHF is necessary for ensuring effective heat transfer. A set of laboratory measurements $\boldsymbol{y}$ depending on P relevant physical variables are available under defined

boundary conditions, at a generic observational point i:

$$\boldsymbol{y}_i = \left[y_{i,1}, \quad \cdots, \quad y_{i,j}, \quad \cdots, \quad y_{i,P} \right] \tag{1}$$

These measured physical variables include both geometrical and hydraulic variables: specifically, these are the hydraulic or equivalent tube diameter, heated length, pressure, mass flux and outlet quality. These variables affect the CHF phenomenon directly or indirectly, indicated as q_i^{CHF}, at the same observational point i.

The dataset $\boldsymbol{D} = \left[\boldsymbol{y}_i \quad q_i^{CHF} \right]_{i=1,\cdots,N}$ contains the experimental measurements collected over a specific period of time, consisting of:

- the experimental measurement matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times P}$, whose element $y_{i,j}$ is the measured physical quantity j at the observation point i, with $i = 1, \cdots, N$ and $j = 1, \cdots, P$.
- the corresponding CHF vector $\boldsymbol{q}^{CHF} \in \mathbb{R}^N$, which contains the CHF measurements q_i^{CHF} at each observation point i, with $i = 1, \cdots, N$.

Given a new vector of measurements $\boldsymbol{y}_{test}$ taken at the current observation, the objective of the present work is to develop a data-driven model that receives in input $\boldsymbol{y}_{test}$ and predicts in output the corresponding CHF value $\hat{\boldsymbol{q}}_{test}^{CHF}$. Note that we implicitly assume the following conditions:

- there exists an unknown function $\mathsf{F} : \boldsymbol{q}^{CHF} = f(\boldsymbol{y})$, which maps the measured physical quantities in vector $\boldsymbol{y}$ to the corresponding CHF $\boldsymbol{q}^{CHF}$.
- the amount of data points N in the dataset $\boldsymbol{D}$ is sufficient to approximate the function F with satisfactory accuracy.

Given the impossibility of demonstrating these two conditions theoretically, the present work makes the empirical assumption that they hold if a prediction model $\mathsf{F}^* : \hat{\boldsymbol{q}}^{CHF} = f^*(\boldsymbol{y})$ can be developed and can estimate the true value of $\boldsymbol{q}^{CHF}$ with satisfactory performance.

3. Proposed method

The proposed ensemble model developed for CHF prediction is sketched in Fig. 1. It comprises the development of individual NN models separately trained with optimal distinct architectures and hyperparameter settings (Section 3.1) and the optimization of the ensemble of these models (Section 3.2).

3.1. Development of individual NN models

A typical representation of a neural network (NN) model is shown in Fig. 2. As illustrated in Fig. 2, a NN is made up of layers of nodes (also called neurons), organized in an input layer (which receives the input data vector $\boldsymbol{y}_i$), one or more hidden layers (which perform computations based on weights and biases) and an output layer (which provides the final prediction of CHF, $\hat{q}_i^{CHF}$).

Mathematically, a NN can be represented as a series of transformations applied to input data, whose development consists of three processes: forward propagation, loss function evaluation and backpropagation. In the forward propagation, the input is passed through the network in which, for each layer l, the output activation is computed as follows:

$$a^{l+1} = f\left(\boldsymbol{W}^l a^l + \boldsymbol{b}^l\right) \tag{2}$$

where $\boldsymbol{W}^l$ and $\boldsymbol{b}^l$ are the weight matrix and bias vector, a^l is the output activation from the previous layer l (which is the input $\boldsymbol{y}_i$ for the first hidden layer $l = 1$), and $f(\bullet)$ is the activation function (e.g. ReLU, Sigmoid and Tanh). The output activation from layer l becomes the input for the next layer $l+1$ (IBM, 2024). Finally, the target CHF prediction is computed by:

$$\hat{q}_i^{CHF} = f\left(\boldsymbol{W}^L a^L + \boldsymbol{b}^L\right) \tag{3}$$

where L is the total number of hidden layers. The parameters of the NN are set by applying the error backpropagation algorithm (Rumelhart

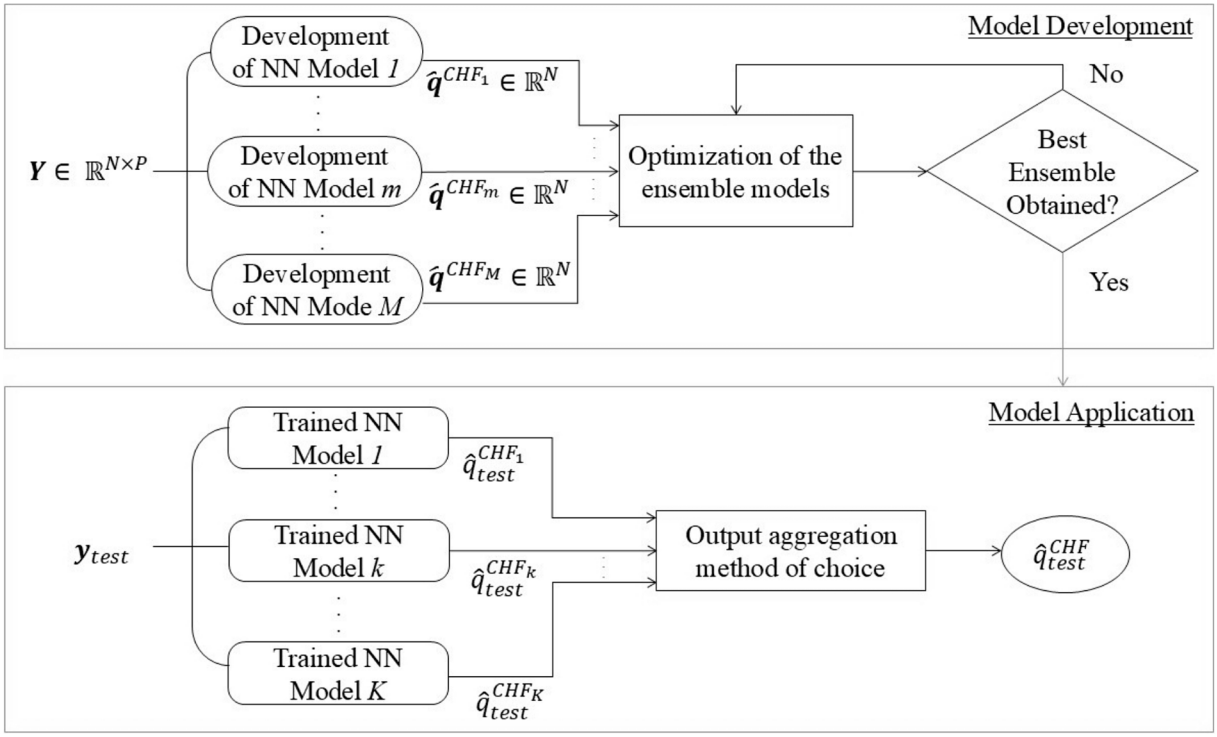

Fig. 1. Flow diagram of the proposed ensemble model for CHF prediction.

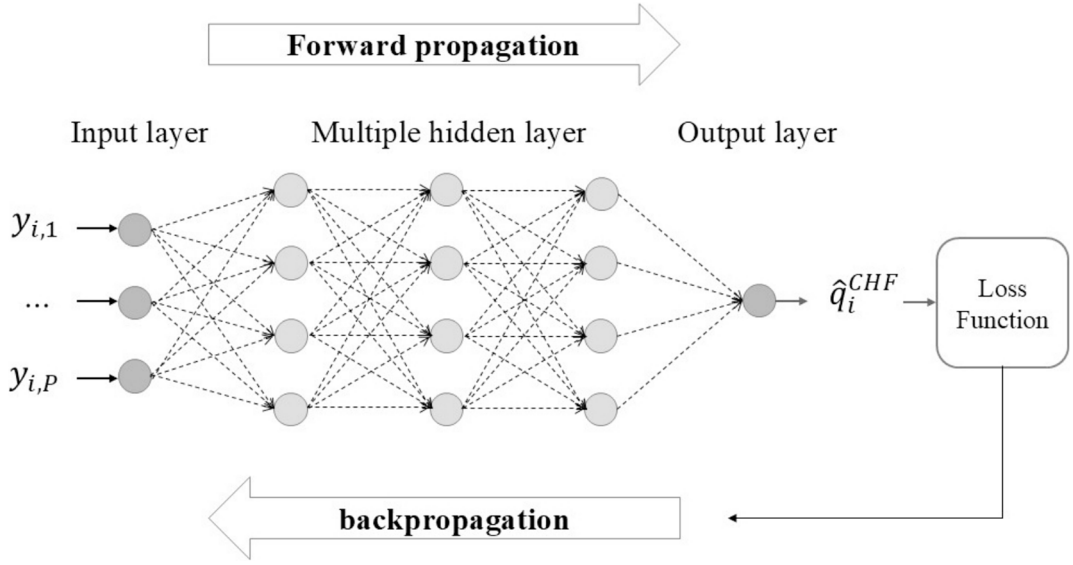

Fig. 2. A typical Neural Network Architecture.

et al., 1986), through the minimization of the mean squared error (MSE) used as loss function:

$$\mathscr{L} = MSE = \frac{1}{N}\sum_{i=1}^{N}\left(q_i^{CHF} - \hat{q}_i^{CHF}\right)^2 \tag{4}$$

where q_i^{CHF} and $\hat{q}_i^{CHF}$ are the actual experimental and predicted values of CHF, respectively.

Multiple individual NN models are here trained separately with distinct architectures and hyperparameters settings, which are optimized following the procedure presented in Fig. 3. Specifically, M different NN models are developed and trained independently during the training phase (Fig. 1, top) using the measurements $\boldsymbol{Y} \in \mathbb{R}^{N\times P}$ as input data and $\boldsymbol{q}^{CHF} \in \mathbb{R}^{N}$ as output.

Given the inherent stochastic nature of NNs, due to the random initialization of layer weights, training the same model architecture on identical data can yield multiple optimized models, each performing differently on unseen data and possessing unique strengths and limitations. To limit this randomness, and provide consistent results, a weight initialization technique is assigned to each layer, so that only the hyperparameters are influencing the different prediction results. To obtain optimal NN models, we propose a systematic, sequential procedure for identifying the best NN models, of varying layers and nodes, based on the training data. Specifically, for each defined number of hidden layers $L = 1, 2, 3, \ldots$, the proposed sequential procedure (summarized in Fig. 3) is as follows:

1) develop and optimize all potential s number of NN models with L number of hidden layers using Bayesian optimization (BO) algorithm by varying model hyperparameters, such as batch size and learning rate. BO algorithm has recently gained popularity in the field of AI and ML due to its capability of effectively identifying the global optimum of a complex function in few iterations (Wang et al., 2020);
2) select optimal NN models based on their root mean squared percentage error (RMSPE) results on validation set, and fine-tune their weights for optimal performance;
3) compare the performance of the optimized NN models with L layers from step 2) to that of the previously optimized NN models with $L-1$ layers;
4) increment the number of hidden layers by 1 and repeat steps 1) to 3) if any optimized model with L layers outperforms any of those with $L-1$ layers. Otherwise, proceed to step 5);
5) collect all M finalized individual models to be used for the optimization of the ensemble model.

Finally, each of the collected trained models generates predictions $\hat{\boldsymbol{q}}^{CHF} \in \mathbb{R}^{N}$ that are subsequently used in Section 3.2 for the optimization of the ensemble model.

3.2. Optimization of ensemble of NN models

Following the development and training of individual models, the subsequent stage is the combination of their predictions through ensemble learning, a method that aggregates the outputs of multiple models to enhance robustness and accuracy. Different models often capture unique data characteristics, and thus, their combination can mitigate the limitations of single models, resulting in a more generalized solution and reducing the likelihood of overfitting (Hansen and Salamon, 1990). To identify the optimal ensemble of the developed NN models, the results from Section 3.1 are used. To facilitate the selection of the best combination of models among the M single models with minimal computational load (and without employing specific optimization algorithms), the combinatorics theory (R. Wilson Combinatorics: A Very Short Introduction 1st ed., 2016) is applied. Specifically, the proposed systematic approach is based on the following steps.

Step 1 – Calculation of potential ensemble models.

The total number of possible combinations of two or more models from the M individual models is determined:

$$N^e = \sum_{r=2}^{M} C(M,r) \tag{5}$$

where $C(M,r) = M!/(r!(M-r)!)$ is the combination operator (binomial coefficient), i.e., the number of possible combinations of r models ($r = 2, 3, \cdots, M$) from M distinct models without regard to order and without replacement.

Step 2 – Enumeration of model combinations.

All N^e possible combinations of the M models are generated. For example, if $M = 3$, then, $N^e = 4$ from Eq. (5) and the following potential combinations are produced: $\{1,2\}$, $\{1,3\}$, $\{2,3\}$ and $\{1,2,3\}$.

Step 3 – Evaluation of ensemble performance.

The performance metrics (Section 4) for each ensemble configuration are calculated using the following aggregation methods:

- Mean aggregation:

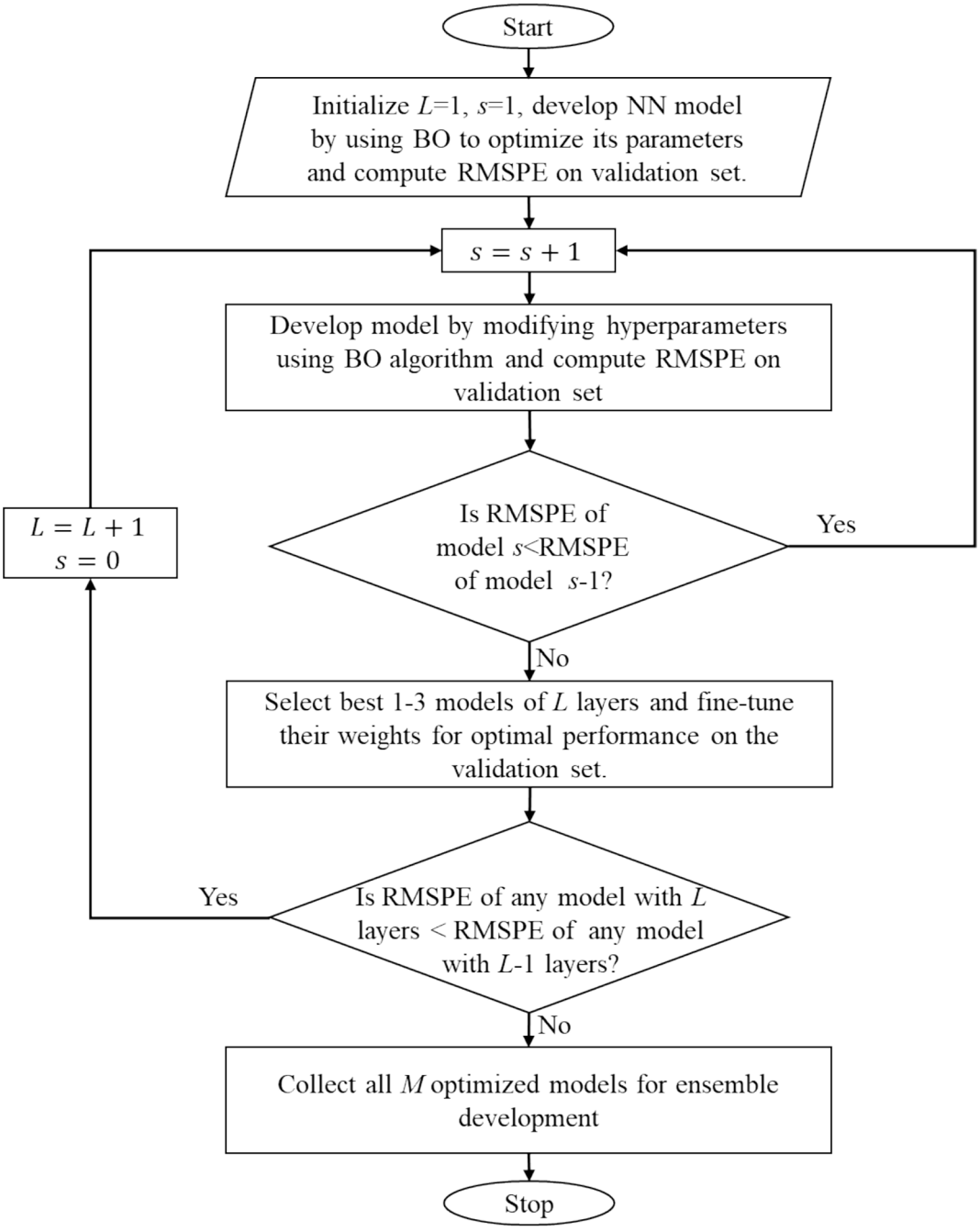

Fig. 3. Flowchart of the development of multiple individual NN models.

$$\hat{q}_i^{CHF_{mean}} = \frac{1}{K}\sum_{k=1}^{K} \hat{q}_i^{CHF_k} \tag{6}$$

where $\hat{q}_i^{CHF_k}$ is the predicted CHF value from model k ($k = 1,2,\cdots,K$) and K is the total number of models in the ensemble. This method provides a balanced prediction by averaging the over- and under-predictions.

- Median aggregation:

$$\hat{q}_i^{CHF_{median}} = \begin{cases} \hat{q}_i^{CHF_{[(K+1)/2]}} & if K is odd \\ \dfrac{\hat{q}_i^{CHF_{[K/2]}} + \hat{q}_i^{CHF_{[(K/2)+1]}}}{2} & if K is even \end{cases} \tag{7}$$

where $\hat{q}_i^{CHF_{[\bullet]}}$ is the predicted CHF value of a model within the ensemble, arranged in ascending order of the predicted CHF values. The median aggregation is chosen because it is little sensitive to outliers and resilient against extreme predictions.

Both methods are tested to determine the optimal balance between robustness and accuracy for CHF predictions.

Step 4 – Selection of the optimal ensemble.

The ensemble that achieved the best performance based on the metrics computed in Step 3 is selected. The proposed systematic approach not only enables the identification of the optimal ensemble but also reduces the computational burden with respect to conventional optimization techniques, such as genetic algorithms or random search, thereby enhancing efficiency in determining the ideal ensemble configuration. Finally, the developed ensemble of the NN models is used to provide online prediction of CHF (Fig. 1, bottom). Given a new input parameter vector $\boldsymbol{y}_{test}$ measured at the current time t, each of the NN

model k ($k = 1, 2, \cdots, K$) in the ensemble receives in input the measurement, $\boldsymbol{y}_{test}$, and provides in output the CHF value, $\widehat{q}_{test}^{CHF_k}$. Subsequently, the obtained predictions from individual models in the ensemble are aggregated into a single prediction, $\widehat{q}_{test}^{CHF}$ as the output of the ensemble model.

4. Performance metrics

The following metrics are employed to assess the performance of the proposed method on the test dataset $\boldsymbol{D}_{test} = \left[\boldsymbol{y}_i \quad q_i^{CHF}\right]_{i=1,\cdots,N^t}$ (Le-Corre et al., 2024):

1) the root mean square percentage error (RMSPE), calculated as:

$$RMSPE = 100\sqrt{\frac{1}{N^t}\sum_{i=1}^{N^t}\left(\frac{\widehat{q}_i^{CHF} - q_i^{CHF}}{q_i^{CHF}}\right)^2} \tag{8}$$

where N^t is the amount of data points in $\boldsymbol{D}_{test}$.

2) the mean absolute percentage error (MAPE), computed as:

$$MAPE = \frac{100}{N^t}\sum_{i=1}^{N^t}\left|\frac{\widehat{q}_i^{CHF} - q_i^{CHF}}{q_i^{CHF}}\right| \tag{9}$$

$Q^2$3) −error, calculated as:

$$Q^2 = \frac{\sum_{i=1}^{N^t}\left(\widehat{q}_i^{CHF} - q_i^{CHF}\right)^2}{\sum_{i=1}^{N^t}\left(\widehat{q}_i^{CHF} - \widehat{\mu}\right)^2} \tag{10}$$

where $\widehat{\mu}$ is the mean of the measured CHF values, $\boldsymbol{q}^{CHF} \in \mathbb{R}^{N^t}$ in $\boldsymbol{D}_{test}$.

4) the mean of the Predicted/Measured (P/M) ratio, computed as:

$$meanP\Big/M = \frac{1}{N^t}\sum_{i=1}^{N^t}\left(\frac{\widehat{q}_i^{CHF}}{q_i^{CHF}}\right) \tag{11}$$

5) the standard deviation of the P/M ratio, computed as:

$$stdP\Big/M = \sqrt{\frac{\sum_{i=1}^{N^t}\left(\frac{\widehat{q}_i^{CHF}}{q_i^{CHF}} - meanP\Big/M\right)^2}{N^t}} \tag{12}$$

The RMSPE is a beneficial metrics as it aligns with the scale of the data. It expresses relative error as a percentage of the observed value, aiding in comparing prediction accuracy across various data scales. This makes RMSPE particularly valuable in large datasets with variable data scales; however, RMSPE is also more sensitive to outliers, emphasizing larger errors. MAPE offers a straightforward interpretation, which represents the average percentage error directly. Q^2, a normalized error metric, compares the prediction error to the deviation from the mean, allowing for a relative assessment of prediction accuracy. Q^2 values near 0 indicate strong predictive capability of the model, whereas values larger than 1 imply that the predictive capability of the model is worse than that of a model that simply provides the mean value as the predicted outputs. The mean of the P/M ratio serves as an indicator of potential bias within the AI model. A substantial deviation from unity suggests that the model may not have attained an optimal fit. Additionally, the standard deviation of the P/M ratio quantifies the proportional error of the model, recognizing that each evaluation is performed as a multiplicative factor.

5. Case study

To validate the proposed CHF prediction model, the publicly available CHF experimental dataset, originally provided by the US Nuclear Regulatory Commission along with (Groeneveld, 2019) is used. This dataset is a collection of 59 different experiments performed in vertical water-cooled uniformly heated tubes during the past decades and made available by the WPRS-EGMUP task force on AI and ML for Scientific Computing in Nuclear Engineering projects, promoted by the OECD/NEA (Le-Corre et al., 2024). The dataset used contains 24,579 observations of five inputs and one output (CHF). The input data consists of geometric (tube diameter, heated length), measured (pressure, mass flux) and calculated (outlet quality) parameters, and it covers a wide range of system parameters values as shown in Table 1.

The data is partitioned into 80 % for model development (64 % training set and 16 % validation set) and 20 % for model evaluation (test set). Given the significant variations in the ranges of input parameters (Table 1), the dataset is scaled between 0 and 1 using MinMaxScaler to ensure that the model accounts for all parameters adequately (Bhandari, 2024):

$$y_{scaled} = \frac{y_{original} - y_{min}}{y_{max} - y_{min}} \tag{12}$$

6. Results and Discussion

The optimal architectures of the individual NN models are determined considering their performances on the validation set, following the procedure described in Section 3.1. To configure each model in terms of numbers of layers and neurons, the rectified linear unit (ReLU) activation is chosen for all the hidden layers due to its robust performance across varied applications (Sharma et al., 2020) and Softplus activation is used in the output layer to ensure non-negative CHF values. To mitigate issues like vanishing or exploding gradients, Kaiming initialization (He et al., 2015) has been applied for weights initialization, whereas biases have been initialized to zero. During training of each individual model, the number of epochs is set to a maximum value of 2000, with early stopping implemented using a patience of 25 iterations to prevent overfitting. A dropout of 0.01 is used for all the hidden layers of each model to further prevent overfitting. The Adam optimization (Kingma and Ba, 2014) with a MSE loss function is applied. Optimal hyperparameters, such as number of nodes per layer, learning rate, batch size and beta value for Softplus output activation, are determined using a BO algorithm applied to models with varying hidden layers, with RMSPE serving as the optimization function evaluated on the validation set. Table 2 presents the optimization parameters along with their respective search ranges used in the BO framework.

6.1. Results of individual NN models

Table 3 shows the optimal parameters for each of the individual NN models, with architectures limited to a maximum of eight hidden layers. The nine- or more layer-architectures are excluded due to increased error and a noisy learning curve, as they did not improve performance. In total, twelve optimized individual NN models are developed ($M = 12$), as summarized in Table 3, and their respective validation and test performances are reported in Table 4. Among these models, Model 10 achieves the lowest values of the metrics and consistently outperforms all other models across nearly all metrics. Overall, Model 10 is identified as the best individual model.

These results underscore the fact that even when models have the same number of layers, variations in the hyperparameters, such as learning rate and number of nodes, can lead to different performances

Table 1
Parameters range of CHF dataset.

	Tube Diameter	Heated Length	Pressure	Mass Flux	Outlet Quality
	[m]	[m]	[kPa]	[kg/ mˆ2/s]	–
Min Value	0.002	0.05	100	8.2	−0.497
Max Value	0.016	20	20,000	7964	0.999

Table 2
Ranges of the hyperparameters optimization for individual single models.

Optimization parameter	Range of search
Number of nodes in hidden layers	[12, 1024]
Learning rate	[0.0001, 0.01]
Batch size	[32, 512]
Beta	[1, 10]
Optimization function	RMSPE
Hidden layers dropout	0.01

on new, unseen data. Notice that Models 2 and 3, with identical number of layers but differing in learning rate and number of nodes, perform differently on the test set, with Model 2 significantly outperforming Model 3 across all metrics. Additionally, increasing the number of layers does not necessarily yield improved performance; for instance, while Model 12 (eight hidden layers) exceeds the performance of Model 3 (two hidden layers), it significantly underperforms relative to Model 4, which utilizes three hidden layers. These findings indicate that each individual model has distinct strengths and limitations, suggesting that an ensemble model combining the optimized individual models could enhance their individual strengths while minimizing weaknesses. This approach is expected to improve prediction performance by exploiting the different capabilities of the individual models within the ensemble.

6.2. Results of ensemble of NN models

To obtain the optimal ensemble of NN models, all twelve individual models presented in Section 6.1 are included and considered in the ensemble model optimization. For clarity and ease of analysis, these models, originally labeled from 1 to 12, are re-labeled as Models *A* through *L*, respectively. Given the twelve developed models ($M = 12$), 4083 possible ensemble configurations (i.e., $N^e = 4083$) from eleven sets of combinations are identified, comprising 66 configurations with two models, 220 configurations with three models, 495 configurations with four models, 792 configurations with five models, 924 configurations with six models, 792 configurations with seven models, 495 configurations with eight models, 220 configurations with nine models, 66 configurations with ten models, 12 configurations with eleven models, and 1 configuration with all twelve models. Each ensemble configuration is subjected to the same validation set and evaluation metrics. Specifically, both RMSPE and MAPE are calculated for each configuration, and two aggregation methods (mean and median) described in Section 3.2 are tested, with only the results of the best-performing aggregation method being here presented.

Figs. 4, 5 and 6 show the performance of the four sets of model combinations, consisting of 66 configurations with two models, 66 configurations with ten models, and 12 and 1 configurations with eleven and twelve models, respectively. Specifically, Fig. 4 presents the results from the 66 ensemble configurations combining two individual models. Here, both mean and median aggregation methods yield identical outcomes, as each ensemble consists of only two models. Based on these results, the configuration $\{I, J\}$ achieves the best performance in both RMSPE (11.98) and MAPE (7.90). When comparing these results to other configurations made of three models, it turns out that most ensembles of three models outperform those of two. Similarly, the performance of most four-model ensembles (Fig. 6(a)) surpass that of the two- and three-model configurations. Overall, as the number of individual models in an ensemble increases, performance generally improves. However, despite these improvements, many ensembles with fewer models substantially outperform some that include a greater number of models (Figs. 5 and 6). For example, the ensemble configuration $\{G, H, I, J\}$ with four models significantly outperforms the configuration $\{B, C, D, E, F, G, H, I, J, K, L\}$ with eleven models across both metrics (Table 5). This finding confirms the importance of optimizing ensemble models to identify the most effective configuration rather than arbitrarily combining multiple models into an ensemble.

Table 3
Optimal parameters of the obtained individual single models.

Hidden layers	Model	Hyperparameters					
		Optimizer	Loss Function	Nodes	Beta	Learning rate	batch size
$L = 1$	Model 1	Adam	MSE	5/1024/1	6.7265	0.0008	161
$L = 2$	Model 2	Adam	MSE	5/712/695/1	1.8826	0.0038	512
	Model 3	Adam	MSE	5/857/236/1	4.1474	0.0043	300
$L = 3$	Model 4	Adam	MSE	5/563/500/326/1	5.3228	0.0001	32
	Model 5	Adam	MSE	5/1012/500/186/1	6.0064	0.0001	221
	Model 6	Adam	MSE	5/592/254/458/1	2.3616	0.0012	112
$L = 4$	Model 7	Adam	MSE	5/1024/367/76/252/1	3.7385	0.0041	62
$L = 5$	Model 8	Adam	MSE	5/303/143/322/102/256/1	4.4957	0.0009	32
$L = 6$	Model 9	Adam	MSE	5/154/426/12/252/90/92/1	4.8208	0.0012	32
	Model 10	Adam	MSE	5/123/291/397/45/117/102/1	4.4647	0.0037	352
$L = 7$	Model 11	Adam	MSE	5/468/233/86/169/12/12/85/1	5.6002	0.0034	216
$L = 8$	Model 12	Adam	MSE	5/462/173/154/19/60/104/50/14/1	2.3746	0.0082	133

Table 4
Validation and test results of the optimized individual models.

Model	Performance on Validation Set				Performance on Test Set			
	RMSPE	MAPE	Q^2	Within ± 20 % error	RMSPE	MAPE	Q^2	Within ± 20 % error
Model 1	14.50	9.98	0.0263	88.61 %	14.90	10.04	0.0258	88.24 %
Model 2	13.75	9.31	0.0255	89.80 %	13.75	9.25	0.0205	89.79 %
Model 3	15.13	10.88	0.0318	86.24 %	15.30	10.79	0.0312	86.29 %
Model 4	12.96	8.81	0.0183	90.56 %	13.70	9.05	0.0174	90.44 %
Model 5	14.47	9.50	0.0219	89.11 %	14.95	9.46	0.0200	89.36 %
Model 6	13.00	8.71	0.0232	90.51 %	13.81	8.86	0.0229	90.21 %
Model 7	13.05	9.12	0.0210	89.70 %	13.45	9.20	0.0189	89.89 %
Model 8	12.49	8.48	**0.0192**	90.79 %	13.08	8.61	0.0168	**90.76 %**
Model 9	12.48	8.43	0.0211	90.87 %	13.09	8.58	0.0196	90.62 %
Model 10	**12.45**	**8.19**	0.0240	**91.51 %**	**12.87**	**8.33**	**0.0158**	90.40 %
Model 11	12.83	8.52	0.0233	90.92 %	13.27	8.58	0.0195	90.58 %
Model 12	14.88	10.23	0.0363	87.79 %	15.28	10.19	0.0322	87.75 %

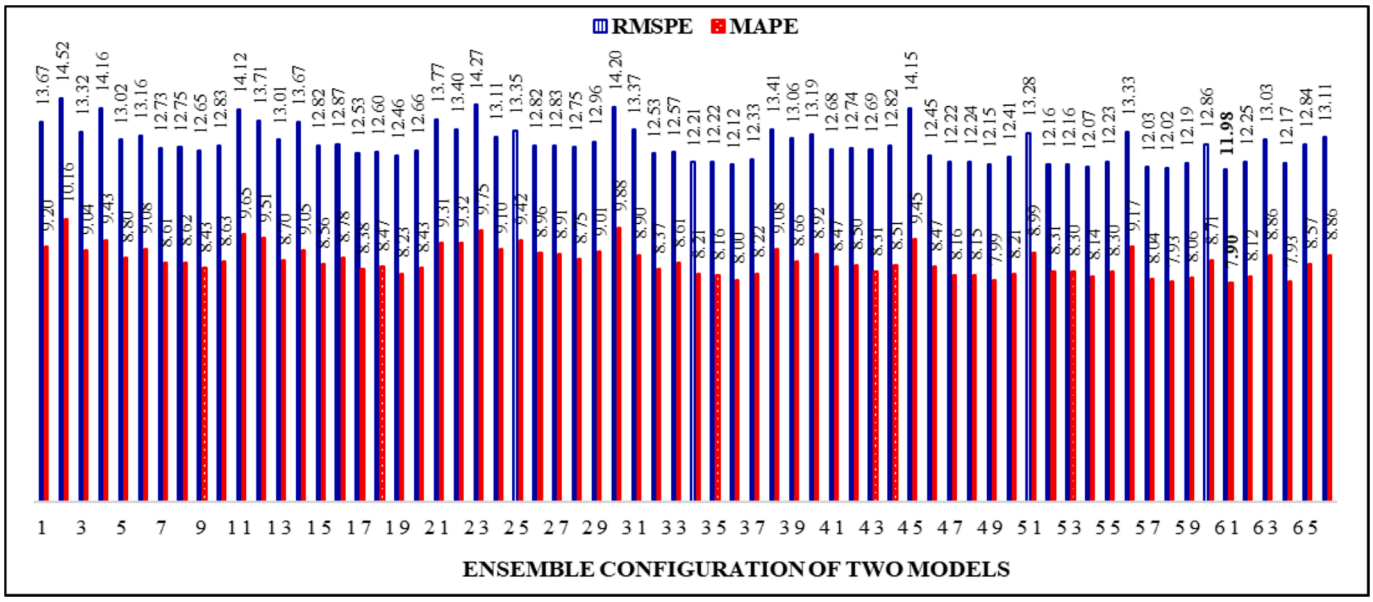

Fig. 4. Performance of the ensemble configurations of two individual models based on validation set.

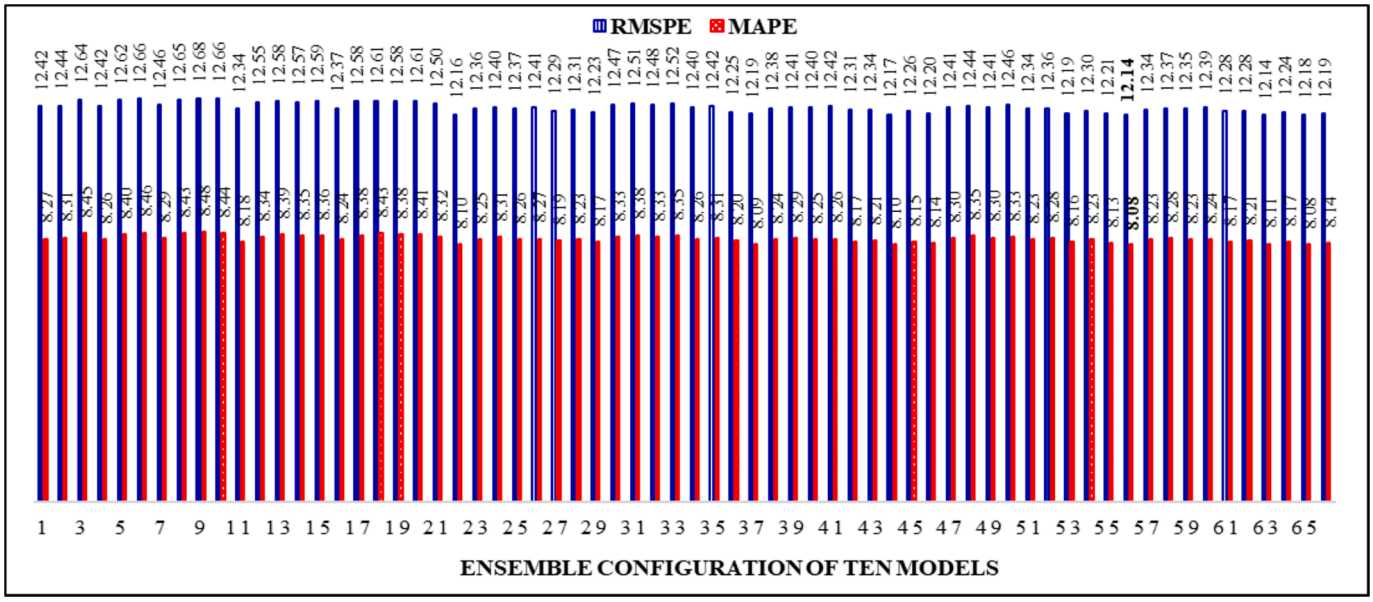

Fig. 5. Performance of the ensemble configurations of ten individual models based on validation set.

Among each set of model combinations, the best-performing ensemble model is selected. Table 5 reports the performances of the best-performing ensembles across all eleven sets of model combinations (i.e., 4083 ensemble configurations), where the selections based on both RMSPE and MAPE metrics are reported. These performances are also visualized in Fig. 7. Overall, the Ensemble Model $\{G, H, I, J\}$ (Table 5) achieves the best RMSPE (11.80) using the mean aggregation method, whereas Ensemble Model $\{H, I, J, K\}$ (Table 5) demonstrates superior performance in terms of MAPE (7.81) using the median aggregation method. Since the RMSPE metric is sensitive to outliers, where larger deviations from actual values have a greater impact on calculations, Ensemble Model $\{H, I, J, K\}$ is selected as the final optimal configuration. The proposed optimization procedure effectively identified this configuration as optimal, illustrating the benefit of systematic selection over arbitrary model combinations. As observed in Fig. 7(a) and 7(b), the performance of the ensemble model improves with an increasing number of model combinations until it reaches a plateau, after which it begins to decline. This trend demonstrates that integrating multiple NNs enhances predictive reliability by leveraging the strengths of individual models while mitigating their weaknesses. However, it also highlights the presence of diminishing returns as additional models are incorporated, underscoring the importance of strategic ensemble selection rather than indiscriminately increasing model complexity.

Fig. 8 compares the CHF prediction of the proposed ensemble model and the LUT method on the test dataset with respect to the measured experimental values. The results indicate that a higher proportion of CHF data points from the ensemble model fall within a $\pm$ 20 % error margin than those from the LUT. This indicates the effectiveness of the ensemble model in capturing input–output relationships and its robust generalization capabilities.

Fig. 9 presents plots of the predicted-to-measured CHF ratio (P/M) against various input parameters for both the proposed ensemble model

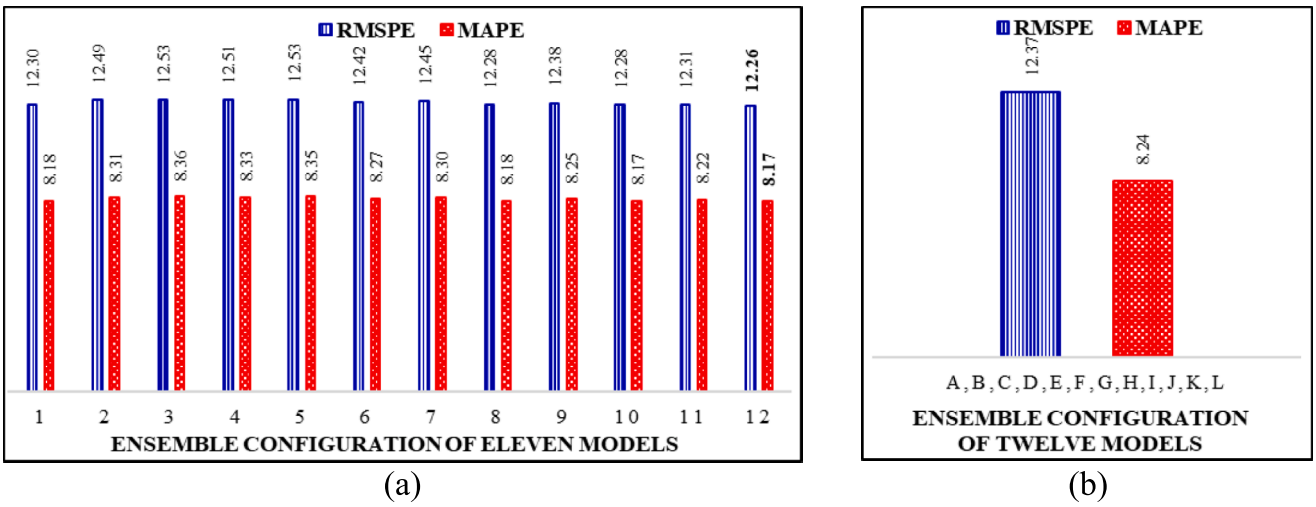

Fig. 6. Performance of the ensemble configurations of: (a) eleven and (b) all twelve individual models based on validation set.

Table 5
Performance of the best-performing ensembles across all eleven sets of combinations based on validation set.

Configuration	Best-preforming ensemble based on RMSPE				Best-preforming ensemble based on MAPE			
	RMSPE	MAPE	Aggregation method	Ensemble model	RMSPE	MAPE	Aggregation Method	Ensemble model
Configuration of 2 models	11.98	7.90	Median	I,J	11.98	7.90	Median	I,J
Configuration of 3 models	11.92	8.02	Median	G,I,J	12.04	7.91	Median	I,J,K
Configuration of 4 models	**11.80**	7.88	Mean	**G,H,I,J**	11.89	**7.81**	Median	**H,I,J,K**
Configuration of 5 models	11.81	7.88	Mean	F,G,H,I,J	11.92	7.85	Median	D,H,I,J,K
Configuration of 6 models	11.82	7.83	Median	F,G,H,I,J,K	11.88	7.82	Median	D,F,H,I,J,K
Configuration of 7 models	11.85	7.86	Median	D,F,G,H,I,J,K	11.85	7.86	Median	D,F,G,H,I,J,K
Configuration of 8 models	11.93	7.97	Mean	C,D,F,G,H,I,J,K	11.94	7.91	Median	B,D,F,G,H,I,J,K
Configuration of 9 models	12.03	8.01	Median	B,C,D,F,G,H,I,J,K	12.03	8.01	Mean	B,C,D,F,G,H,I,J,K
Configuration of 10 models	12.14	8.08	Median	B,C,D,E,F,G,H,I,J,K	12.14	8.08	Mean	B,C,D,E,F,G,H,I,J,K
Configuration of 11 models	12.26	8.17	Median	B,C,D,E,F,G,H,I,J,K,L	12.26	8.17	Mean	B,C,D,E,F,G,H,I,J,K,L
Configuration of 12 models	12.37	8.24	Median	A,B,C,D,E,F,G,H,I,J,K,L	12.37	8.24	Mean	A,B,C,D,E,F,G,H,I,J,K,L

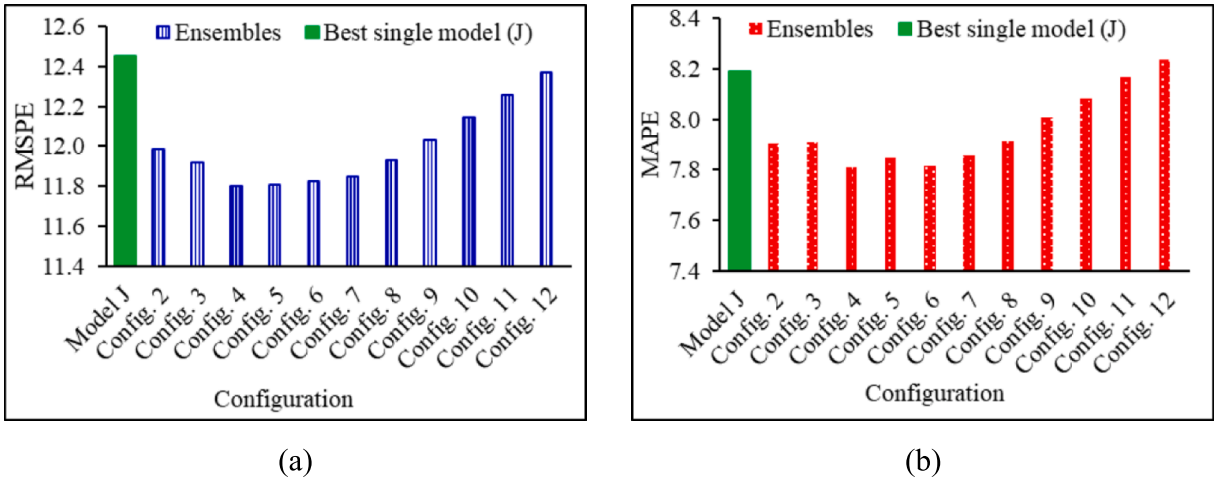

Fig. 7. Performance of the best-performing ensembles across all eleven sets of combinations based on validation set: (a) RMSPE and (b) MAPE.

and LUT. An ideal P/M ratio close to 1 across the test set indicates desirable predictive accuracy; P/M ratios greater than 1 suggest overestimation, whereas ratio values less than 1 indicate underestimation. As shown in Fig. 9, the proposed ensemble model significantly enhances CHF prediction accuracy across multiple key input parameters, including tube diameter, heated length, mass flux, pressure and outlet quality. These results demonstrate the improved generalization capabilities of the proposed ensemble method, highlighting its potential for broader applications within the thermal engineering field.

The performance of the proposed ensemble method has been verified by its comparison to those of three state-of-the-art prediction methods based on LUT, Random Forest (RF) and Support Vector Regression (SVR), respectively. Furthermore, since there is no significant difference in performance among the following ensemble models: $\{G,H,I,J\}$, $\{H,I,J,K\}$, $\{D,H,I,J,K\}$, $\{F,G,H,I,J\}$, $\{D,F,H,I,J,K\}$, $\{F,G,H,I,J,K\}$, and $\{D,F,G,H,I,J,K\}$ on the validation set results (Table 5), we evaluate

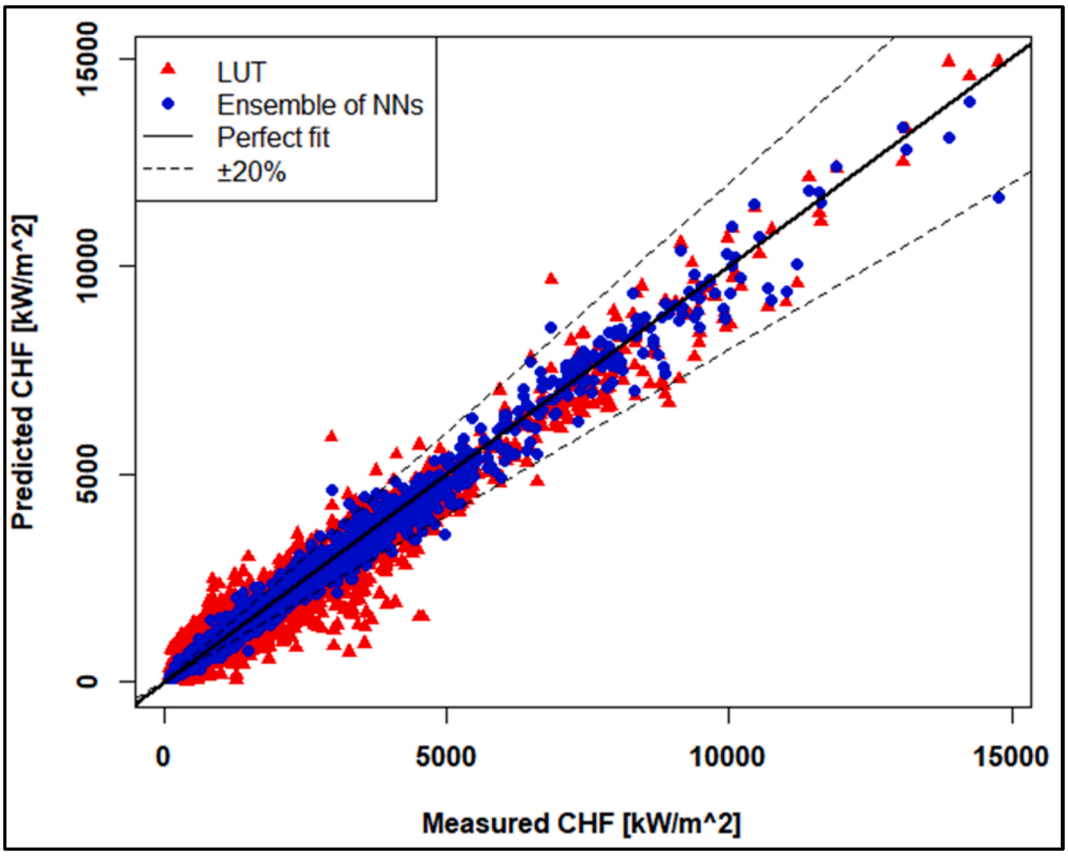

Fig. 8. Predicted vs. measured CHF on test data for the optimal ensemble model and LUT.

those models on the test set and compare their performance. Both RF and SVR are trained and optimized using the same 80 % training set used to develop the proposed method. Table 6 reports the values obtained by the different methods for the performance metrics. The performance metrics indicate that all the NN models (both best individual and ensemble of NN models) outperform all other methods, with the ensemble models showing superior performance across all metrics: the ensemble models not only achieves the lowest RMSPE and MAPE values but also indicates the smallest value of Q^2-error (nearly 0), indicating strong predictive capability of the ensemble model. Notice that there is no significant difference in the performance of the ensemble models $\{G, H, I, J\}$ and $\{H, I, J, K\}$, both demonstrating superior performance compared to other ensemble models. However, the ensemble model $\{H, I, J, K\}$ improves the prediction accuracy with 91.81 % of the predicted data points falling within a 20 % error band, value that is better than those of the other models.

To further evaluate the performance of the proposed ensemble method, the entire CHF database (24579 datapoints) has been used to compare its effectiveness against recently published state-of-the-art AI-based CHF prediction methods in (Zhou et al., 2024), including the benchmark NN model in (Le-Corre et al., 2024) and the Convolutional Neural Network (CNN) and transformer models presented in (Zhou et al., 2024). The use of the entire dataset ensures a fair comparison between the proposed ensemble method and other NN-based models in (Zhou et al., 2024), as these models were also evaluated on the same entire CHF dataset. Table 7 reports the performance results across the entire CHF database. As shown in Table 7, the best-performing ensemble model $\{H, I, J, K\}$ consistently surpasses other NN-based models in nearly all performance metrics, achieving an RMSPE of 10.43, MAPE of 6.76, mean P/M of 1.007 and std P/M of 0.104. The only exception is the Q^2 metric (1.46 %), where the transformer model (the best-performing model in (Zhou et al., 2024) excels with a Q^2 value of 1.26 %. This indicates an exceptionally close match between the predictions of the ensemble model and the actual measurements. Specifically, the predictions of the ensemble model are, on average, only 0.7 % higher than the actual values, indicating a minor tendency towards overprediction. Additionally, the lowest Std. P/M value achieved by the ensemble model indicates relatively low variability in its predictions compared to actual values, reinforcing its stability and reliability. These findings demonstrate that combining multiple NNs enhances predictive reliability by leveraging the strengths of individual models while mitigating their weaknesses.

6.3. Sensitivity analysis

To further validate the proposed method in terms of its consistency with the expected physical behavior of the system, a sensitivity analysis is performed using the slicing approach from the NRC CHF dataset as described in (Le-Corre et al., 2024). Here, regions with varying values of diameter, heated length, pressure, mass flux and outlet quality have been identified, whereas other parameters have been held constant. From these regions, two slices for each varying parameter have been selected, as shown in Table 8. These “sliced” datasets allow for demonstration of the physical behavior of the CHF prediction model across the parameter space, enhancing confidence in the predictive capabilities of the model and in preventing overfitting. In addition to the “sliced” datasets, the corresponding experimental data points are extracted using the same parameter ranges as those considered in (Le-Corre et al., 2024). Fig. 10 through Fig. 14 show the results of the predictions for the ten “sliced” datasets by the ensemble model and the LUT, illustrating the physical behaviors of the models.

In Fig. 10(a) and (b), discrepancies are observed between the CHF predictions of the ensemble model and the LUT approach, particularly for tube diameters smaller than 0.008 m. It is important to note that the

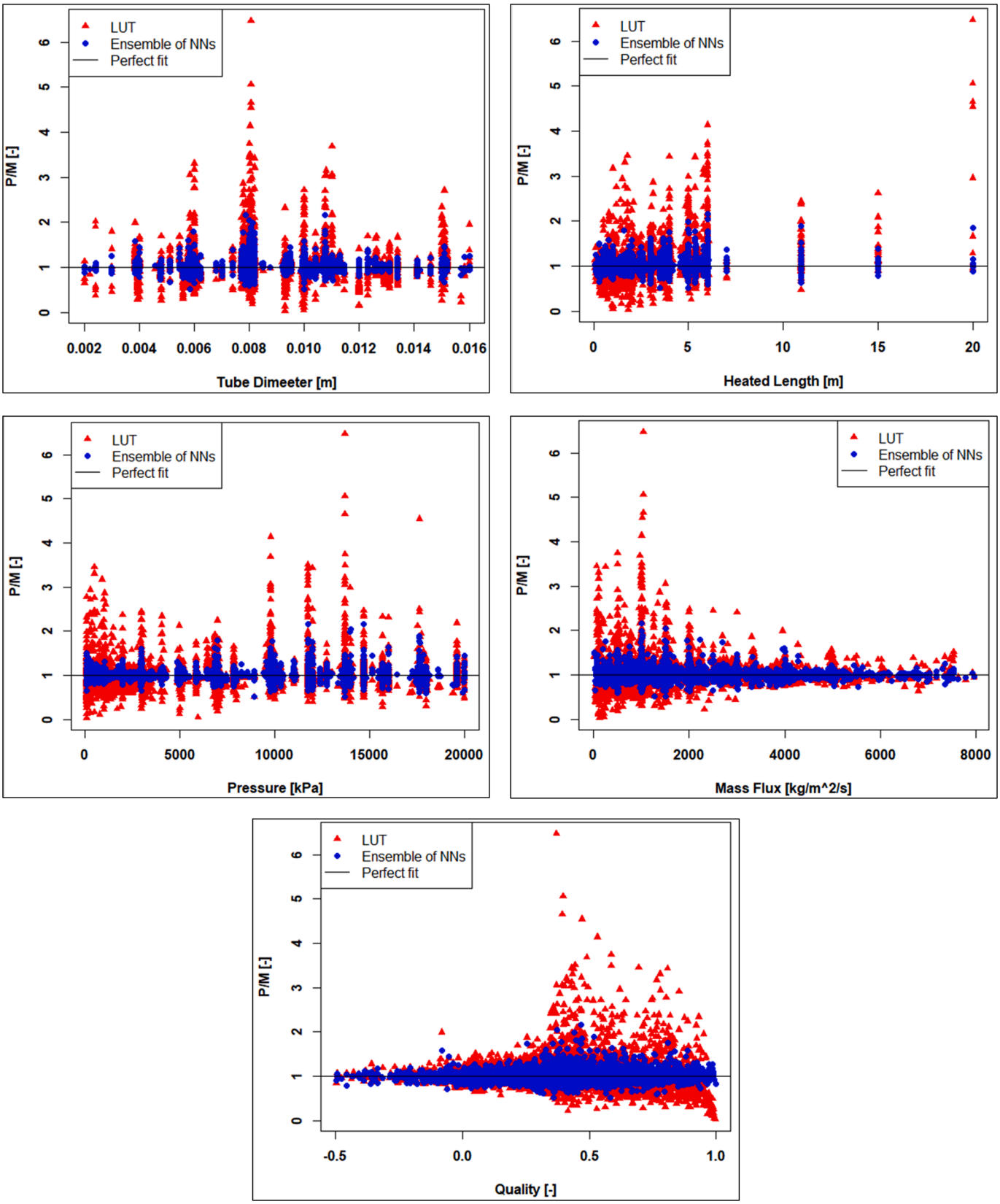

Fig. 9. Predicted-to-measured CHF ratio distribution against input parameters.

Table 6
Comparison of the accuracy of CHF prediction models on the test set.

Model	Performance on Test Set					
	RMSPE	MAPE	Q2	Within ± 20 % error	Mean P/M	Std. P/M
LUT	36.19	20.52	6.08 %	68.71 %	1.042	0.359
RF	13.47	8.79	1.98 %	90.23 %	1.013	0.134
SVR	20.75	12.28	3.79 %	85.31 %	1.024	0.206
Best Single NN Model $\{J\}$	12.87	8.33	1.58 %	90.40 %	1.013	0.128
Ensemble of NN $\{G,H,I,J\}$	**12.32**	8.02	**1.48 %**	91.54 %	1.012	0.123
Ensemble of NN $\{H,I,J,K\}$	12.42	**7.92**	1.50 %	**91.81 %**	**1.011**	**0.122**
Ensemble of NN $\{D,H,I,J,K\}$	12.43	7.95	1.48 %	91.39 %	1.014	0.123
Ensemble of NN $\{F,G,H,I,J\}$	12.40	8.01	1.53 %	91.64 %	1.012	0.123
Ensemble of NN $\{D,F,H,I,J,K\}$	12.48	7.97	1.53 %	91.70 %	1.013	0.124
Ensemble of NN $\{F,G,H,I,J,K\}$	12.37	7.95	1.55 %	91.66 %	1.012	0.123
Ensemble of NN $\{D,F,G,H,I,J,K\}$	12.41	7.99	1.52 %	91.80 %	1.012	0.124

Table 7
Comparison of the accuracy of proposed method with methods in (Zhou et al., 2024) for CHF database.

Model	Performance on Whole CHF Database (24579 datapoints)					
	RMSPE	MAPE	Q2	Within ± 20 % error	Mean P/M	Std. P/M
LUT	36.30	19.77	6.31 %	68.90 %	1.032	0.362
Benchmark NN (Le-Corre et al., 2024)	12.60	8.03	2.20 %	–	1.010	0.115
NN (Zhou et al., 2024)	20.40	11.20	1.98 %	–	1.034	0.201
CNN (Zhou et al., 2024)	16.80	9.72	1.36 %	–	1.027	0.166
Transformer (Zhou et al., 2024)	12.30	7.22	**1.26 %**	–	1.008	0.122
Best Single NN Model $\{J\}$	10.60	6.87	1.46 %	94.22 %	1.008	0.106
Ensemble of NN $\{G,H,I,J\}$	10.55	6.95	1.44 %	94.21 %	1.009	0.105
Ensemble of NN $\{H,I,J,K\}$	**10.43**	**6.76**	1.46 %	**94.38 %**	**1.007**	**0.104**
Ensemble of NN $\{D,H,I,J,K\}$	10.55	6.83	1.40 %	94.32 %	1.010	0.105
Ensemble of NN $\{F,G,H,I,J\}$	10.62	6.91	1.48 %	94.23 %	1.008	0.106
Ensemble of NN $\{D,F,H,I,J,K\}$	10.61	6.83	1.44 %	94.31 %	1.009	0.106
Ensemble of NN $\{F,G,H,I,J,K\}$	10.52	6.82	1.48 %	94.37 %	**1.007**	0.105
Ensemble of NN $\{D,F,G,H,I,J,K\}$	10.63	6.89	1.43 %	94.31 %	1.008	0.106

Table 8
Ranges of slice datasets (Le-Corre et al., 2024).

Slice #	Diameter [m]	Heated Length [m]	Pressure [kPa]	Mass Flux [kg/m^2/s]	Outlet Quality [-]
1	**0 – 0.016**	6.000	14,701	998.5	0.391
2	**0 – 0.016**	6.000	9807	1003.3	0.529
3	0.00801	**0 – 20**	9806	1000.0	0.587
4	0.00811	**0 – 20**	2009	752.2	0.756
5	0.00800	0.998	**0 – 20,000**	2040.2	0.140
6	0.01340	3.658	**0 – 20,000**	2040.2	0.378
7	0.00800	1.570	12,750	**0 – 8000**	0.144
8	0.01000	4.966	16,000	**0 – 8000**	0.343
9	0.00814	1.943	9831	1519.5	**−0.5 – 1.0**
10	0.00800	0.997	17,650	2002.7	**−0.5 – 1.0**

LUT CHF predictions were generated using the empirical formula provided in (Groeneveld et al., 2007). According to the LUT equations, CHF decreases monotonically with increasing diameter, which is evident in Fig. 10. However, from a theoretical perspective, the simplified energy balance for CHF expressed in (Hall and Mudawar, 2000) shows that an increase in diameter (with the other parameters remaining constant) requires an increase in surface heat flux to satisfy energy conservation, which, in turn, leads to a higher CHF. The ensemble model captures this relationship, with its CHF predictions increasing with diameter, aligning with physical principles and experimental data. Furthermore, the ensemble model demonstrates better agreement with the experimental trends and higher predictive accuracy compared to the LUT model, highlighting its ability to model complex CHF behaviors.

Fig. 11(a) and (b) present the impact of heated length on CHF. While the LUT model assumes a linear and constant CHF response to heated length (reflecting its omission of influence of the heated length (Groeneveld et al., 2007), the proposed ensemble model predicts a nonlinear trend, with CHF decreasing as heated length increases up to approximately 10 m, beyond which the curve flattens. This suggests that the influence of the heated length on CHF decreases with increasing length. The observed trend is consistent with energy distribution principles (Kinoshita et al., 2000): as the heated length increases, localized heat flux is spread over a larger area, reducing the local heat flux intensity and, thus, the CHF. The proposed ensemble model captures this physical relationship more accurately than the LUT approach.

In Fig. 12(a) and (b), both ensemble and LUT models successfully reproduce the nonlinear relationship between CHF and system pressure, particularly in regions supported by experimental data. CHF initially increases and, then, decreases with rising pressure. This behavior stems from the pressure-dependent variation in enthalpy values (Chun et al., 2001), which influences the outlet quality through a nonlinear mechanism embedded within the energy equations. Although both models replicate this trend, the ensemble model shows closer agreement with experimental observations.

Fig. 13(a) and (b) show the CHF physical behavior with respect to mass flux. Both ensemble and LUT models predict a linear increase in CHF with mass flux, reflecting the direct proportionality between surface heat flux and mass flux. This linear relationship is well captured by both methods; however, the ensemble model demonstrates better alignment with experimental data, indicating its improved ability to model the response of CHF to variations in mass flux.

Finally, Fig. 14(a) and (b) show the effect of outlet quality on CHF. The ensemble and LUT models both capture the decreasing trend of CHF with increasing outlet quality, a behavior supported by experimental data (Fig. 14). It is important to note that at high outlet quality values, increased vapor content leads to the formation of an insulating vapor blanket on the heating surface, thereby reducing heat transfer efficiency (Yang et al., 2025), a key mechanism underlying CHF reduction. Notice that the ensemble model predictions closely follow experimental trends, further validating its physical interpretability.

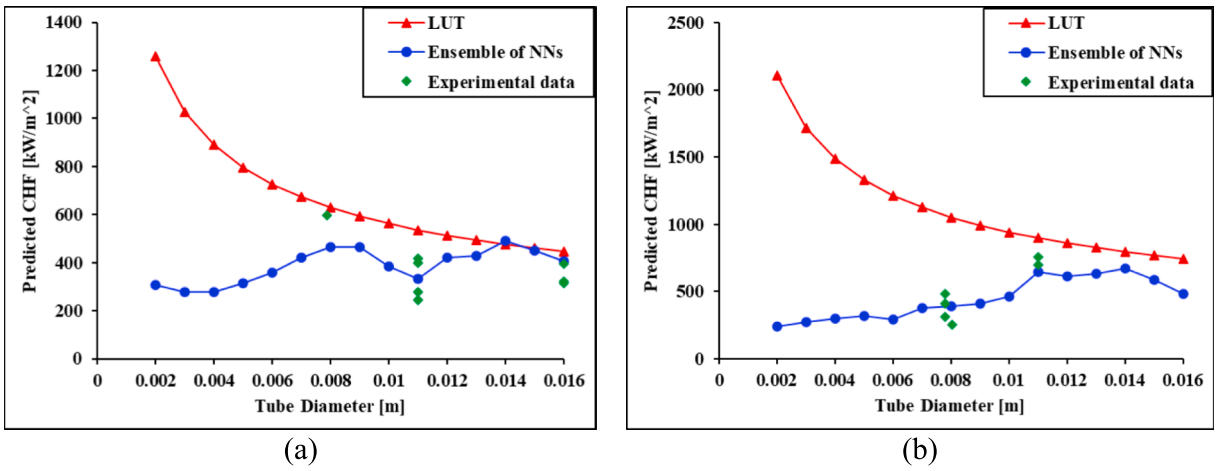

Fig. 10. Physical behaviors of CHF prediction from ensemble of NNs and LUT against tube diameter for: (a) Slice 1 and (b) Slice 2, and corresponding NRC CHF data points.

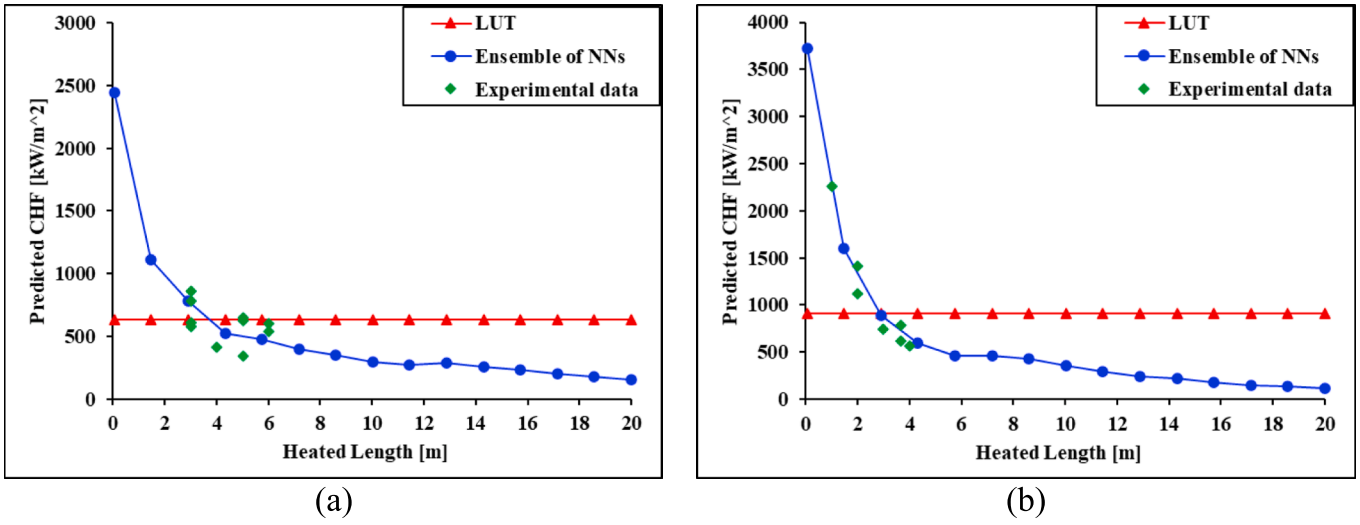

Fig. 11. Physical behaviors of CHF prediction from ensemble of NNs and LUT against heated length for: (a) Slice 3 and (b) Slice 4, and corresponding NRC CHF data points.

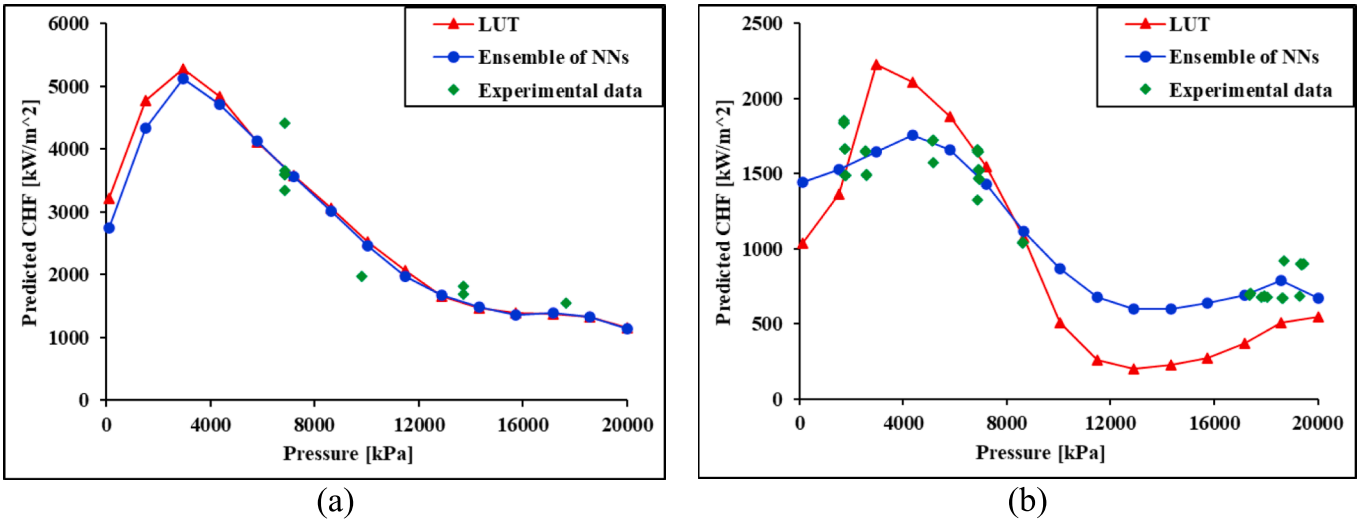

Fig. 12. Physical behaviors of CHF prediction from ensemble of NNs and LUT against pressure for: (a) Slice 5 and (b) Slice 6, and corresponding NRC CHF data points.

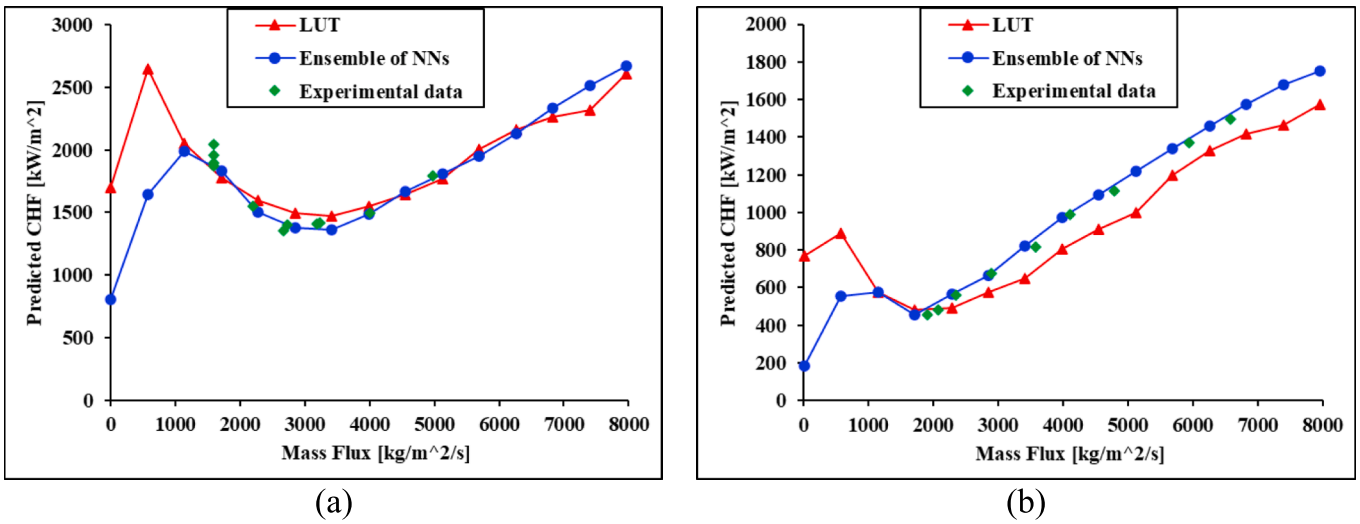

Fig. 13. Physical behaviors of CHF prediction from ensemble of NNs and LUT against mass flux for: (a) Slice 7 and (b) Slice 8, and corresponding NRC CHF data points.

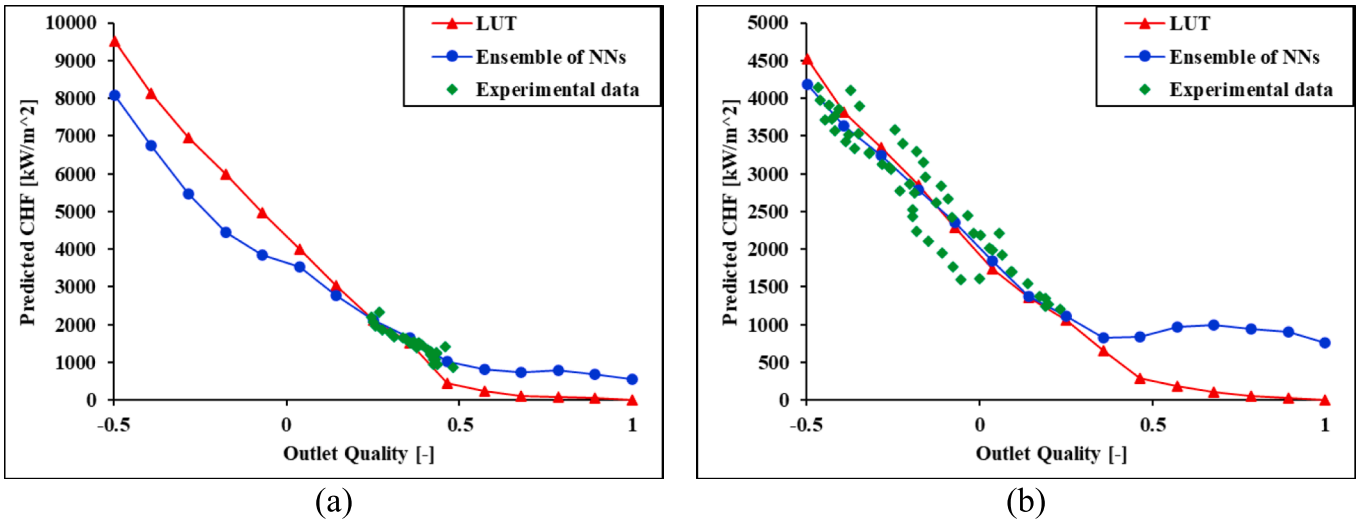

Fig. 14. Physical behaviors of CHF prediction from ensemble of NNs and LUT against outlet quality for: (a) Slice 9 and (b) Slice 10, and corresponding NRC CHF data points.

In summary, the proposed ensemble model yield CHF trends that are generally consistent with those derived from the LUT. Deviations in physical behavior, particularly regarding the effects of tube diameter (Fig. 10) and heated length (Fig. 11), can be attributed to the limitations of LUT, such as its oversimplified treatment of diameter as a strictly decreasing factor and its omission of the influence of heated length. The proposed ensemble model demonstrates enhanced physical consistency and predictive accuracy, underscoring its suitability for modeling complex thermohydraulic phenomena.

7. Conclusions

This paper presents a new method for CHF prediction in vertical tubes based on an optimized ensemble of NN models to enhance predictive accuracy and generalization capabilities. First, multiple individual NN models are developed, each optimally trained with distinct architectures and hyperparameters settings. These individual models are, then, combined to form an optimized ensemble model. Two systematic procedures are introduced: *1*) a procedure for automatically identifying optimal NN models; and *2*) a procedure for aggregating optimized individual NN models into an optimal ensemble model. The proposed procedures allow the identification of the optimal ensemble configuration without relying on computationally intensive optimization techniques. The proposed procedures can be applied to any given training data. Due to its dynamic nature, particularly in terms of varying hyperparameters, the proposed approach is expected to identify optimal NN models specific to the data under consideration. It is important to note that the representativeness of the training data across all conditions is a common challenge in AI/ML methodologies, as these models learn only the underlying structure of the provided data.

The proposed method is validated using experimental CHF data from the WPRS-EGMUP task force on AI and ML for Scientific Computing in Nuclear Engineering projects, promoted by the OECD/NEA. The obtained results show that the proposed method offers superior accuracy in CHF prediction, achieving lower errors across all metrics considered (MAPE = 6.76, RMSPE = 10.43, mean P/M = 1.007, std P/M = 0.104) when compared to conventional methods such as the LUT approach (MAPE = 19.77, RMSPE = 36.30, mean P/M = 1.032, std P/M = 0.362) and other NN-based models like CNN (MAPE = 9.72, RMSPE = 16.80, mean P/M = 1.027, std P/M = 0.166) and Transformer RF (MAPE =

7.22, RMSPE = 12.30, mean P/M = 1.008, std P/M = 0.122). The optimized ensemble model achieves a coverage of 94.38 % of predicted data points within a ± 20 % error margin, marking an enhanced improvement over conventional techniques. Parametric and sensitivity analyses further confirm the strong generalization capability of the proposed method and its consistency with expected physical behaviors, thereby reinforcing the effectiveness of the proposed method for CHF prediction. Although in the case considered the best single NN model performs comparably to conventional RF (a tree-based approach) and SVR methods, the ensembles of NNs still achieve a (small) performance improvement. Besides the small enhancement, the ensemble approaches provide several key advantages, in general terms. They enhance generalization by reducing overfitting, improve robustness to noisy data and refine feature representation without extensive engineering. Additionally, they can provide greater flexibility by integrating hybrid models and adapting to evolving challenges. Ensembles of NNs also can scale efficiently with large datasets and can leverage model diversity for more balanced predictions. In synthesis, although the accuracy improvement in our case is modest, the ensemble NN approach offers superior robustness, adaptability and scalability, positioning it as a more reliable solution for CHF prediction.

Future research will focus on embedding physical laws within the NN architecture, for the development of physics-informed NNs which could further enhance the robustness and reliability of the model and increase the precision of the learning algorithm by incorporating fundamental physical principles into the NN training process.

CRediT authorship contribution statement

Ibrahim Ahmed: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Irene Gatti:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Enrico Zio:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Todreas, N.E., Kazimi, M., 2011. Nuclear Systems Volume I 0 ed. 2011 CRC Press 10.1201/b14887.

Bruder, M., Bloch, G., Sattelmayer, T., 2017. Critical Heat Flux in Flow Boiling—Review of the Current Understanding and Experimental Approaches. *Heat Transf. Eng.*, 38 (3), 347–360. https://doi.org/10.1080/01457632.2016.1189274.

Celata, G.P., Cumo, M., Mariani, A., Simoncini, M., Zummo, G., 1994. Rationalization of existing mechanistic models for the prediction of water subcooled flow boiling critical heat flux. *Int. J. Heat Mass Transf.*, 37, 347–360. https://doi.org/10.1016/0017-9310(94)90035-3.

Katto, Y., 1992. A prediction model of subcooled water flow boiling CHF for pressure in the range 0.1–20 MPa. *Int. J. Heat Mass Transf.*, 35 (5), 1115–1123. https://doi.org/10.1016/0017-9310(92)90172-O.

Tong, L.S., 1967. Prediction of departure from nucleate boiling for an axially non-uniform heat flux distribution. *J. Nucl. Energy* 21 (3), 241–248. https://doi.org/10.1016/S0022-3107(67)90054-8.

Bowring, R.W., 1973. WSC-2: a subchannel dryout correlation for water-cooled clusters over the pressure range 3.4-15.9 MPA (500-2300 PSIA). Atomic Energy Establishment Winfrith.

Biasi, L., Clerici, G.C., Tozzi, A., Sala, R., 1968. Extension of A.R.S. correlation to burnout prediction with non-uniform heating. *J. Nucl. Energy* 22 (12), 705–716. https://doi.org/10.1016/0022-3107(68)90044-0.

Celata, G.P., Cumo, M., Katto, Y., Mariani, A., 1999. Prediction of the critical heat flux in water subcooled flow boiling using a new mechanistic approach. *Int. J. Heat Mass Transf.*, 42 (8), 1457–1466. https://doi.org/10.1016/S0017-9310(98)00286-5.

Kataoka, I., Ishii, M., Nakayama, A., 2000. Entrainment and desposition rates of droplets in annular two-phase flow. *Int. J. Heat Mass Transf.*, 43 (9), 1573–1589. https://doi.org/10.1016/S0017-9310(99)00236-7.

Whalley, P.B., 1977. The calculation of dryout in a rod bundle. *Int. J. Multiph. Flow* 3 (6), 501–515. https://doi.org/10.1016/0301-9322(77)90026-X.

Lee, C.H., Mudawar, I., 1988. A mechanistic critical heat flux model for subcooled flow boiling based on local bulk flow conditions. *Int. J. Multiph. Flow* 14 (6), 711–728. https://doi.org/10.1016/0301-9322(88)90070-5.

Weisman, J., Pei, B.S., 1983. Prediction of critical heat flux in flow boiling at low qualities. *Int. J. Heat Mass Transf.*, 26 (10), 1463–1477. https://doi.org/10.1016/S0017-9310(83)80047-7.

Chang, S., Baek, W.-P., 2003. Understanding, predicting, and enhancing critical heat flux. In: The 10th Int. Topical Meeting on Nuc. Reactor Thermal Hydraulics (NURETH-10). October 5–9, 2003. Seoul, Korea.

Doroshchuk, V.E., Levitan, I.L., Lantzman, F.P., 1975. Investigation into Burnout in Uniformly Heated Tubes. ASME. Publication vol. 75-WA/HT-22, 1975.

Groeneveld, D.C., 2019. Critical heat flux data used to generate the 2006 Groeneveld critical heat flux lookup tables. United States Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, DC.

Mazzola, A., 1997. Integrating artificial neural networks and empirical correlations for the prediction of water-subcooled critical heat flux. *Rev. Générale Therm.*, 36 (11), 799–806. https://doi.org/10.1016/S0035-3159(97)87750-1.

Groeneveld, D.C., Shan, J.Q., Vasić, A.Z., Leung, L.K.H., Durmayaz, A., Yang, J., Cheng, S.C., anase, A.T, 2007. The 2006 CHF look-up table. Nucl. Eng. Des. 237 (15–17), 1909–1922. https://doi.org/10.1016/j.nucengdes.2007.02.014.

Lee, M. R., Jeong, H.-J., Choi, Y. J., Gatton, T. M., 2006. "A Nuclear Power Plant Expert System Using Artificial Neural Networks", in *Advances in Neural Networks - ISNN 2006*, vol. 3972, Wang, J., Yi, Z., Zurada, J. M., Lu, B.-L., Yin, H., Eds., in Lecture Notes in Computer Science, vol. 3972, Berlin, Heidelberg: Springer, Berlin Heidelberg, 2006, pp. 1239–1245. doi: 10.1007/11760023_180.

Jiang, B.T., Zhao, F.Y., 2013. Combination of support vector regression and artificial neural networks for prediction of critical heat flux. *Int. J. Heat Mass Transf.*, 62, 481–494. https://doi.org/10.1016/j.ijheatmasstransfer.2013.03.025.

He, M., Lee, Y., 2018. Application of machine learning for prediction of critical heat flux: Support vector machine for data-driven CHF look-up table construction based on sparingly distributed training data points. *Nucl. Eng. Des.*, 338, 189–198. https://doi.org/10.1016/j.nucengdes.2018.08.005.

Khalid, R.Z., Ullah, A., Khan, A., Khan, A., Inayat, M.H., 2023. Comparison of Standalone and Hybrid Machine Learning Models for Prediction of Critical Heat Flux in Vertical Tubes. *Energies* 16 (7), 3182. https://doi.org/10.3390/en16073182.

Zhao, X., Shirvan, K., Salko, R.K., Guo, F., 2020. On the prediction of critical heat flux using a physics-informed machine learning-aided framework. *Appl. Therm. Eng.*, 164, 114540. https://doi.org/10.1016/j.applthermaleng.2019.114540.

Mao, C., Jin, Y., 2024. Uncertainty quantification study of the physics-informed machine learning models for critical heat flux prediction. *Prog. Nucl. Energy* 170, 105097. https://doi.org/10.1016/j.pnucene.2024.105097.

Zhou, W., Miwa, S., Wang, H., Okamoto, K., 2024. Assessment of the state-of-the-art AI methods for critical heat flux prediction. *Int. J. Heat Mass Transf.*, 158, 107844. https://doi.org/10.1016/j.icheatmasstransfer.2024.107844.

Zubair Khalid, R., Ahmed, I., Ullah, A., Zio, E., Khan, A., 2024. Enhancing accuracy of prediction of critical heat flux in Circular channels by ensemble of deep sparse autoencoders and deep neural Networks, *Nucl. Eng. Des.*, vol. 429, p. 113587, Dec. 2024, doi: 10.1016/j.nucengdes.2024.113587.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536. https://doi.org/10.1038/323533a0.

Wang, L., Dernoncourt, F., Bui, T., 2020. Bayesian Optimization for Selecting Efficient Machine Learning Models. *ArXiv:2008.00386*. https://doi.org/10.48550/arXiv.2008.00386.

Hansen, L.K., Salamon, P., 1990. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12 (10), 993–1001. https://doi.org/10.1109/34.58871.

Le-Corre, J-M., Delipei, G., Wu, X., Zhao, X., 2024. "Benchmark on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering. Phase 1: Critical Heat Flux Exercise Specifications," vol. OECD Publishing, Paris, no. NEA Working Papers, 2024.

IBM, 2024. "What is a Neural Network? | IBM.", Accessed: Jun. 17, 2024. [Online]. Available: https://www.ibm.com/topics/neural-networks.

R. Wilson Combinatorics: A Very Short Introduction 1st ed. 2016 Oxford University Press 10.1093/actrade/9780198723493.001.0001.

Bhandari, A., 2024. "Feature Scaling: Engineering, Normalization, and Standardization (Updated 2024)," Analytics Vidhya. Accessed: Jun. 20, 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/.

Sharma, S., Sharma, S., Athaiya, A., 2020. ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *Int. J. Eng. Appl. Sci. Technol.*, 04 (12), 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054.

He, K., Zhang, X., Ren, S., Sun, J., 2015. "Delving Deep into Rectifiers. Surpassing Human-Level Performance on ImageNet Classification," arXiv. https://doi.org/10.48550/ARXIV.1502.01852.

Kingma, D.P., Ba, J., 2014. "Adam. A Method for Stochastic Optimization," arXiv. https://doi.org/10.48550/ARXIV.1412.6980.

Groeneveld, D.C., Leung, L.K.H., Kirillov, P.L., Bobkov, V.P., Smogalev, I.P., Vinogradov, V.N., Huang, X.C., Royer, E., 1996. The 1995 look-up table for critical

heat flux in tubes. Nucl. Eng. Des. 163 (1–2), 1–23. https://doi.org/10.1016/0029-5493(95)01154-4.
Hall, D.D., Mudawar, I., 2000. Critical heat flux (CHF) for water flow in tubes—I. Compilation and assessment of world CHF data. *Int. J. Heat Mass Transf.*, 43 (14), 2573–2604.
Kinoshita, H., Yoshida, T., Nariai, H., Inasaka, F., 2000. Effect of heated length on the critical heat flux of subcooled flow boiling. Part 1: Observation of bubbles and slug length at atmospheric pressure. *Heat Transfer Research* 29 (2), 132–143.
Chun, S.-Y., Chung, H.-J., Moon, S.-K., Yang, S.-K., Chung, M.-K., Schoesse, T., Aritomi, M., 2001. Effect of pressure on critical heat flux in uniformly heated vertical annulus under low flow conditions. *Nucl. Eng. Des.*, 203 (2–3), 159–174.
Yang, J., Yong, H., Kim, S., Park, I.W., Lee, Y.-G., Vadlamudi, S.R.G., Park, H.S., 2025. Critical heat flux dependence on surface orientation and bubble dynamics in pool boiling over silicon and silicon dioxide surfaces. *International Journal of Energy Research* 6413134.