

## Article

# Research on Heat Transfer Coefficient Prediction of Printed Circuit Plate Heat Exchanger Based on Deep Learning

Yi Su, Yongchen Zhao, Jingjin Wu  and Ling Zhang 

The School of Mechanical and Electrical Engineering, Hainan University, Haikou 570228, China; 22220854060003@hainanu.edu.cn (Y.S.); 22220854060010@hainanu.edu.cn (Y.Z.); jingjin.wu@hainanu.edu.cn (J.W.)  
\* Correspondence: 992715@hainanu.edu.cn

**Abstract:** The PCHE, as an efficient heat exchanger, plays a crucial role in the storage and regasification of LNG. However, among the existing studies, those that integrate this field with deep learning are scarce. Moreover, research on explainability remains insufficient. To address these gaps, this study first constructs a dataset of heat transfer coefficients ( $h$ ) through numerical simulations. Pearson correlation analysis is employed to screen out the most influential features. In terms of predictive modeling, the study compares five traditional machine learning models alongside deep learning models such as long short-term memory neural networks (LSTMs), gated recurrent units (GRUs), and Transformer. To further enhance prediction accuracy, three attention mechanisms—self-attention mechanism (SA), squeeze-and-excitation mechanism (SE), and local attention mechanism (LA)—are incorporated into the deep learning models. The experimental results demonstrate that the artificial neural network achieves the best performance among the traditional models, with a prediction accuracy for straight-path  $h$  reaching 0.891799 ( $R^2$ ). When comparing deep learning models augmented with attention mechanisms against the baseline models, both LSTM–SE in the linear flow channel and Transformer–LA in the hexagonal flow channel exhibit improved prediction accuracy. Notably, in predicting the heat transfer coefficient of the hexagonal channel, the determination coefficient ( $R^2$ ) of the Transformer–LA model reaches 0.9993, indicating excellent prediction performance. Additionally, this study introduces the SHAP interpretable analysis method to elucidate model predictions, revealing the contributions of different features to model outputs. For instance, in a straight flow channel, the hydraulic diameter ( $D_h$ ) contributes most significantly to the model output, whereas in a hexagonal flow channel, wall temperature ( $T_{inw}$ ) and heat flux ( $Q_w$ ) play more prominent roles. In conclusion, this study offers novel insights and methodologies for PCHE performance prediction by leveraging various machine learning and deep learning models enhanced with attention mechanisms and incorporating explainable analysis methods. These findings not only validate the efficacy of machine learning and deep learning in complex heat exchanger modeling but also provide critical theoretical support for engineering optimization.



Academic Editor: Francesco Zirilli

Received: 11 March 2025

Revised: 6 April 2025

Accepted: 9 April 2025

Published: 22 April 2025

**Citation:** Su, Y.; Zhao, Y.; Wu, J.; Zhang, L. Research on Heat Transfer Coefficient Prediction of Printed Circuit Plate Heat Exchanger Based on Deep Learning. *Appl. Sci.* **2025**, *15*, 4635. <https://doi.org/10.3390/app15094635>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** PCHE; liquefied natural gas; heat transfer coefficient; deep learning; attention mechanism; SHAP

## 1. Introduction

Liquefied natural gas (LNG), as a clean and efficient energy source, has played an increasingly important role in the global energy transition in recent years. With the global

emphasis on environmental protection and the optimization of energy structure, natural gas, with its advantages of low carbon emissions and high calorific value, has become one of the key energy sources to replace coal and oil. According to statistics from the Global Carbon Neutral Institute, global gas demand in 2024 increased by about 2.5% year-on-year to reach a record high of 4.200 billion cubic meters [1]. Floating liquefied natural gas storage and regasification units (FLNG SRUs) are considered a vital component of the future LNG supply chain. In this process, one of the key components is the intermediate fluid vaporizer (IFV). The IFV facilitates the conversion of LNG into gaseous natural gas by utilizing seawater or other cooling media. This process necessitates efficient and compact heat exchange equipment such as the printed circuit heat exchanger (PCHE), which is widely recognized as an ideal choice for such applications [2]. Consequently, accurately predicting the local performance of PCHEs, including heat transfer coefficients and other parameters, is essential for optimizing their design and operational efficiency [3]. The accurate prediction of these parameters not only enhances the overall performance of the equipment but also reduces operating costs and improves energy utilization efficiency. For instance, compared with the numerical simulation method, which requires approximately 9 h to achieve convergence, machine learning methods can complete the task in less than an hour [4].

The current research on traditional PCHEs predominantly centers on analyzing their flow and heat transfer characteristics through experimental studies and numerical simulations [5]. The primary aim of these approaches is to establish empirical correlations between flow dynamics and heat transfer performance, thereby enabling precise prediction and optimization of the heat exchanger's behavior during the design phase. Experimental methods involve measuring critical parameters such as fluid flow state, heat transfer efficiency, and pressure drop within the PCHE, offering valuable insights into its actual operational performance [6]. However, experimental methods are limited by high costs, time-consuming procedures, and an inability to comprehensively cover all potential operating conditions [7]. In contrast, numerical simulation methods construct mathematical models to simulate the internal flow and heat transfer processes of the PCHE, facilitating rapid analysis and optimization across a broad spectrum of operating conditions [8]. Numerical simulations not only assist in predicting the performance of PCHE under various scenarios but also provide essential guidance for equipment design [9]. Nevertheless, the accuracy of numerical simulations relies heavily on the fidelity of the flow models and the precision of boundary condition settings, which may lead to discrepancies between model predictions and real-world performance in certain cases.

Machine learning, as an emerging field, demonstrates the ability to uncover complex nonlinear relationships that traditional physical models struggle to capture by learning from extensive experimental data. Consequently, it holds substantial promise for the design and optimization of heat exchangers [10]. Particularly in multi-objective optimization, machine learning can simultaneously evaluate multiple design parameters such as thermal efficiency, pressure drop, volume, and cost, thereby providing more precise and efficient solutions for engineering applications [11]. For instance, Li et al. introduced a novel approach in their energy article by applying machine learning algorithms to PCHEs using supercritical methane as the working fluid, proposing a predictive model based on artificial neural networks (ANNs) with an  $R^2$  value of 0.9996 [12]. However, the study focused solely on ANNs, without exploring other potential models. Additionally, Zhang et al. [13] conducted numerical simulations of the thermohydraulic performance of supercritical carbon dioxide in serrated channels. Their findings indicate that reducing the serrated bending angle enhances heat transfer performance but compromises hydraulic performance. Moreover, higher heat transfer efficiency correlates with better convective

heat transfer, while a very low heat capacity ratio may lead to a slight reduction in heat transfer efficiency. The study determined that the optimal range for the bending angle is between  $110^\circ$  and  $130^\circ$  [13]. Seung Yeob Lee performed pioneering research on the thermohydraulic performance of a novel PCHE with straight channels inserted into zigzag flow paths. This research elucidated the impact of the inserted straight channels on heat transfer and pressure drop as functions of mass flow rate. It found that the heat transfer characteristics of 0.5 mm and 1 mm straight channels were very similar to those of Z-shaped channels but superior to those of wavy channels [14]. Han et al. [15] were the first to investigate the effects of different rib structures in straight channel configurations on the thermohydraulic performance of supercritical carbon dioxide during flow. They utilized the field synergy method to analyze heat transfer enhancement and identified the optimal rib structure [15]. KIM W proposed empirical correlations for cross-flow, parallel-flow, and counter-flow arrangements in PCHEs with straight channels and varying flow directions, examining the influence of flow channel dimensions, spacing, and overall size on heat transfer performance [16]. In summary, although there is a substantial body of experimental and numerical research on PCHEs, the absence of robust parameter prediction models and structural generalization remains a significant challenge. Consequently, accurately calculating the thermohydraulic performance within PCHE channels, particularly the heat transfer coefficient used to evaluate the heat transfer characteristics, remains highly challenging.

The aim of this study is to investigate the heat transfer coefficient of methane in PCHE channels with different flow path geometries using machine learning methods. First, this study introduces a novel approach for dataset acquisition (the distance division method), enabling the collection of more detailed samples. Second, for the first time, we comprehensively evaluate the potential of various deep learning models (CNN, LSTM, GRU, and Transformer) as well as the integration of specific attention mechanisms (SA, SE, and LA) into deep learning frameworks for predicting PCHE heat exchange performance. Finally, to the best of our knowledge, we conduct a more in-depth interpretability analysis to further examine the contributions of different features to model predictions. This study not only provides a new methodology for predicting PCHE performance parameters but also establishes a standardized workflow, offering valuable insights for the future design of heat exchangers.

## 2. Data Acquisition and Preprocessing

### 2.1. Geometric Model and Simulation Environment

In the liquefied natural gas process, the main focus is on the changes in methane. When the temperature and pressure of methane continuously increase, it reaches a supercritical state once they exceed a certain threshold [17]. This study conducted four sets of three-dimensional Computational Fluid Dynamics (CFD) simulations, considering straight channels with semicircular cross-sections of 1.2 mm and 1.8 mm in diameter, as well as a regular hexagonal channel. The length of the PCHE was 500 mm in all cases.

A three-dimensional numerical simulation was conducted in a horizontal channel, taking into account the influence of gravity. In the context of the gasification process within an offshore LNG Floating Storage and Regasification Unit (FSRU), the methane pressure ranges from 6 to 9 MPa, with a pseudo-critical temperature varying between 190 and 215 K. The inlet boundary condition is specified as the mass flow rate, while the outlet boundary condition is set to constant pressure. Based on the general requirements for transcritical methane in offshore FSRU PCHEs, the inlet temperature is set to 111.15 K, and the temperature variation during the simulation spans from 100 to 300 K. To simulate the actual heat transfer conditions between LNG and the intermediate medium, the upper and

lower walls are assigned constant heat flux boundary conditions, while the remaining walls are defined as adiabatic boundary conditions.

With reference to the numerical simulation of transcritical fluid flow and heat transfer, it is considered that the SST k- $\omega$  turbulence model can reflect the heat transfer characteristics of transcritical fluid by using the SIMPLE algorithm for the pressure-velocity coupling solution and the second-order upwind scheme for pressure dispersion. The momentum equation is discretized by the QUICK scheme, and the energy equation, turbulent kinetic energy equation, and turbulent dissipation rate equation are discretized by the second-order upwind scheme. When the residual value of each equation reaches  $1.0 \times 10^{-6}$  and the monitoring parameters remain unchanged, the calculation is considered to be convergent. At the same time, in order to eliminate the influence of the number of grids on the simulation results, the grid independence verification was carried out in this study. As shown in Figure 1, when the number of grids is greater than 2.1 million, the difference between temperature and pressure is small and tends to be stable, so 2.1 million grids was selected to calculate the final grid.

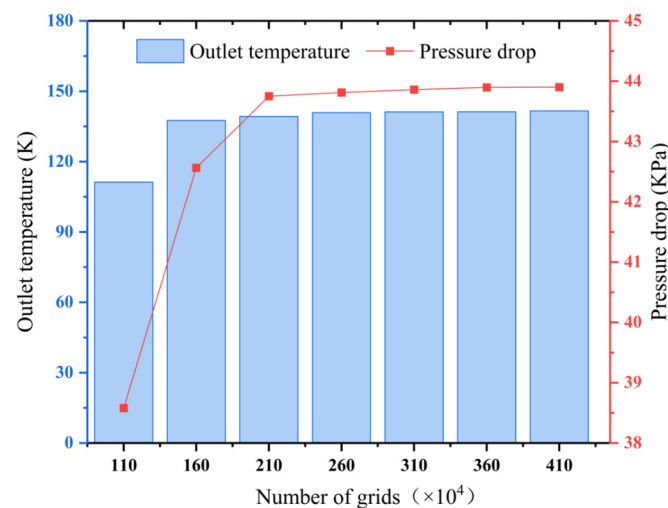


**Figure 1.** Grid independence verification.

The geometric model establishment and calculations in this study were performed using Ansys 2022R1. Feature selection, predictive model development, and other machine learning-related tasks were carried out in PyCharm 2024.3.1. The Python version utilized was 3.10, and the deep learning model was constructed based on PyTorch 1.8.0. The computational hardware included an NVIDIA GeForce RTX 4060 GPU and an Intel 14th Gen Core i7-14650HX CPU.

## 2.2. Dataset Acquisition

The temperature variation in the transcritical methane within the PCHE channels is significant, and the property changes along the corners of the regular hexagonal channels also require evaluation. Consequently, modeling and analyzing the entire PCHE channel as a single averaged sample is unreliable. This paper adopts a more reliable differential element analysis method. Considering the characteristics of the regular hexagonal geometry, the entire channel was uniformly divided into differential element segments at 1 mm intervals. The schematic diagrams of point sampling, division by temperature difference, and division by distance are illustrated in Figure 2.

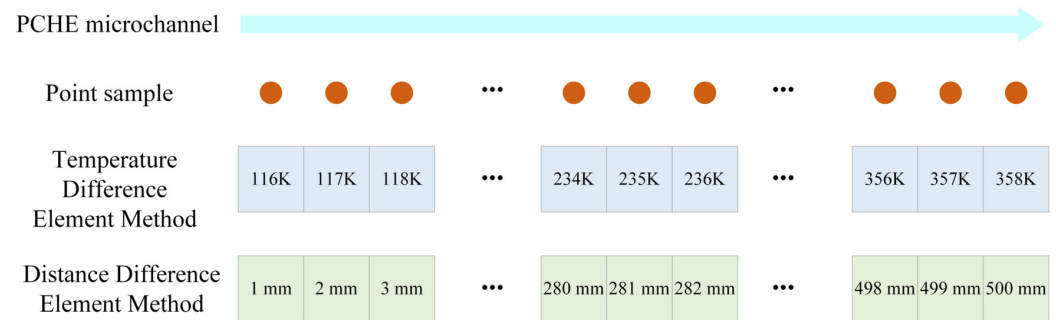


**Figure 2.** Schematic diagram of sample division.

By employing the distance division method, the sample size is 2 to 3 times larger than that with the temperature difference method. At this point, the fluid properties within each 1 mm micro-element can be approximated as constant. The parameters of each micro-element were extracted, and the heat transfer coefficient was subsequently calculated. Through numerical simulation results under four distinct operating conditions, the data for liquidation flow heat transfer inside the PCHE microchannels were segmented to generate a substantial dataset. The dataset comprised a total of 1000 micro-element data points. The dataset information is presented in Table 1. The input feature variables included inlet temperature, inlet pressure, wall heat flux density, inner wall temperature, mass flow rate, temperature difference, inlet velocity, density, dynamic viscosity, channel diameter, hydraulic diameter, Prandtl number, and Reynolds number.

**Table 1.** PCHE flow channel dataset information under four working conditions.

| Variable             | Name                                                             | Straight-Through Flow Channel Range | Hexagonal Flow Channel Range |
|----------------------|------------------------------------------------------------------|-------------------------------------|------------------------------|
| Feature variable (X) | Inlet temperature $T_{in}$ (Pa)                                  | [111.15, 377.9662]                  | [111.15, 401.14103]          |
|                      | Inlet pressure $P_{in}$ (Pa)                                     | [3.03, 11,865.28]                   | [79.26, 2412.52]             |
|                      | Wall heat flux density $q_w$ (W/m <sup>2</sup> )                 | [73,696.66, 155,707.46]             | [38,664.40, 123,604.93]      |
|                      | Inner wall surface temperature $T_{inw}$ (K)                     | [119.70, 443.74]                    | [124.51, 476.46]             |
|                      | Mass flow rate $V$ (G/kg/(m <sup>2</sup> /s))                    | [118.13, 265.39]                    | [118.13, 265.39]             |
|                      | Temperature difference $T_d$ (K)                                 | [0.1497, 2.1355]                    | [−21.44, 28.31]              |
|                      | Entrance density $v_{in}$ (m/s)                                  | [0.27, 11.45]                       | [0.27, 5.21]                 |
|                      | Entrance density $p_{in}$ (kg/m <sup>3</sup> )                   | [23.43, 432.16]                     | [32.81, 432.16]              |
|                      | Dynamic viscosity $\mu$ (kg/(m·s))                               | [0.000073, 0.015]                   | [0.000006, 0.000733]         |
|                      | Channel diameter $D$ (mm)                                        | [1.2, 1.8]                          | [1.2, 1.8]                   |
|                      | Hydraulic diameter $D_h$ (m)                                     | [0.0011, 0.001466]                  | [0.0011, 0.001466]           |
|                      | Prandtl number $Pr$                                              | [0.78, 2.95]                        | [0.00071, 0.0028]            |
|                      | Reynolds number $Re$                                             | [8.39, 4078.00]                     | [632.97, 24,610.42]          |
|                      | Convective heat transfer coefficient $h$ (W/(m <sup>2</sup> ·K)) | [781.42, 20,806.38]                 | [554.168, 9625.993]          |

The target variable is the heat transfer coefficient. The data have been open-sourced ([https://github.com/mirror52/test1/blob/main/data\\_1.2mm-1.8mm-liu-h.xlsx](https://github.com/mirror52/test1/blob/main/data_1.2mm-1.8mm-liu-h.xlsx) (accessed on 5 April 2025)), including data of four working conditions (1.2 mm diameter straight channel and hexagonal channel, 1.8 mm diameter straight channel and hexagonal channel).

The histogram of the frequency distribution generated based on the dataset information is presented in Figure 3. All data exhibit a right-skewed distribution, meaning that lower values occur more frequently, while higher values occur less frequently. The distribution characteristics of heat transfer coefficients for the straight and hexagonal flow channels are

similar; however, their specific value ranges differ. The numerical range of the heat transfer coefficient for the straight flow channel is relatively narrow, whereas that for the hexagonal flow channel is relatively broad. Specifically, the heat transfer coefficient for the straight flow channel is primarily concentrated between 0 and 8000, while for the hexagonal flow channels, it is mainly concentrated between 0 and 5000.

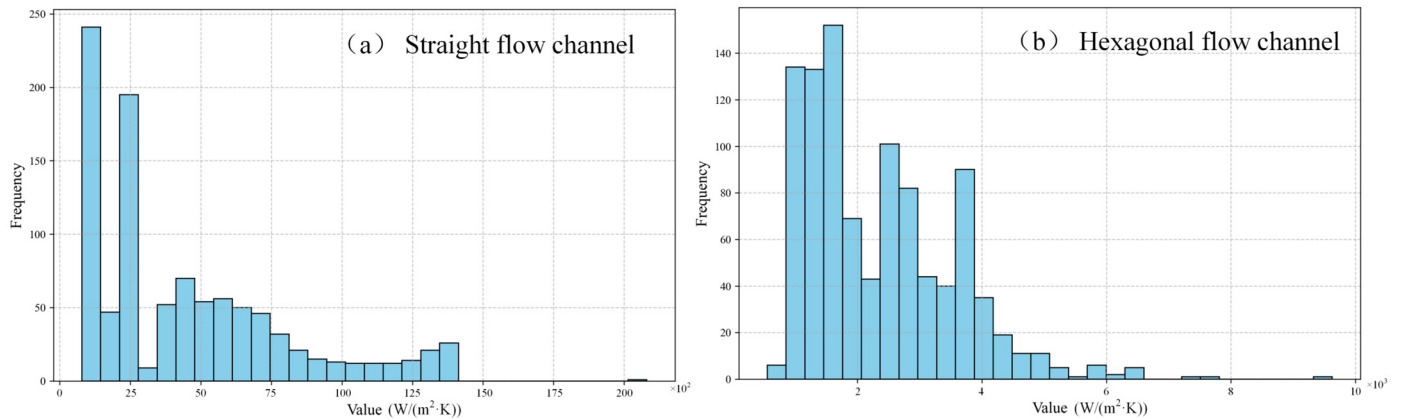


**Figure 3.** Histogram of heat transfer coefficient and pressure drop frequency distribution of different channel shapes: (a) heat transfer coefficient data statistics of straight flow channel; (b) heat transfer coefficient data statistics of hexagonal flow channel.

### 2.3. Feature Selection Method

The independent variables in Table 1 play a key role in describing the fluid behavior and thermodynamic process, and the prediction accuracy can be effectively improved through the reasonable selection and optimization of characteristics. We determined the Pearson correlation coefficient method from a large number of experiments, before which we tried RF-RFE, Brouta, and other feature selection methods. Finally, according to the experimental results, Pearson screened features were selected for follow-up study.

The Pearson correlation coefficient method is widely employed in this field as a feature selection technique. Unlike many other methods, the Pearson correlation does not rely on machine learning algorithms but instead filters features based on statistical relationships or other metrics between the input features and the target variable [18]. Due to its simplicity and computational efficiency, this filtering approach is often utilized in the early stages of feature selection to help eliminate redundant or irrelevant features. The calculation formula is as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where  $X_i$  and  $y_i$  represent the sample data points of the sum of the two variables, respectively,  $\bar{X}$  and  $\bar{y}$  are the mean values of the two variables, respectively, and the value of  $r$  is between  $-1$  and  $1$ .

### 2.4. Prediction Model

Based on the summary of the research status, it can be seen that ANN has good performance in predicting the performance of heat exchangers at present. Therefore, the ANN model was also built in this study. As a comparison, XGBOOST, CATBOOST, RF, SVM, and other models are also used for modeling and analysis. Secondly, modeling analysis of CNN and RNN series models using deep learning was carried out to explore

their predictive model potential. In the training and testing of all models, the same random seeds were set (seed = 42).

#### 2.4.1. Machine Learning Models

The artificial neural network (ANN) is a computational model that mimics the connection mode of neurons in the human brain and is widely used to address various complex pattern recognition and prediction problems [19]. It comprises a large number of neurons arranged in a hierarchical structure, where information is transmitted through weighted connections. Random forest (RF), on the other hand, generates multiple decision trees by randomly selecting samples and features, thereby reducing overfitting and enhancing generalization ability. During the generation of each tree, data subsets are extracted via the bootstrap sampling method, and random feature subsets are selected to construct decision trees [20]. The final prediction result is obtained by aggregating the outputs of multiple decision trees through voting or averaging.

XGBOOST is an efficient implementation of the boosting tree. It uses parallel processing and multiple optimization techniques, making significant improvements in speed and memory efficiency [21].

CATBOOST is a machine learning library based on gradient boosting trees introduced in 2018. The primary design objective of CATBOOST is to streamline the machine learning workflow while enhancing performance, particularly in scenarios with rich feature sets [22]. It employs the ordered boosting method, which reduces bias in the gradient boosting algorithm by randomly permuting the training data during each iteration. Additionally, it offers a variety of optimization algorithms and feature engineering tools, enabling automatic selection of the optimal feature combination.

Support Vector Regression (SVR) differs from traditional regression models in that it incorporates the principles of support vector machines and predicts outcomes by identifying an “optimal hyperplane”. Its primary objective is to minimize the deviation between predicted and actual values, and through specific techniques such as kernel tricks and regularization, the model achieves strong generalization capabilities [23].

#### 2.4.2. Deep Learning Models

Convolutional neural networks (CNNs) are a type of deep learning model that imitate the working mode of the biological visual nervous system and use convolutional layers to extract local features from input data [24].

Recurrent neural networks (RNNs) are a special type of neural network model mainly used for sequence data [25]. Long short-term memory networks (LSTMs) are an improvement on traditional RNNs, aiming to solve the vanishing gradient problem of standard RNNs when processing long sequences [26]. Gated recurrent units (GRUs) are an improved version of LSTMs that simplify the structure of the LSTMs to enhance computational efficiency [27]. Transformer is a deep learning model for processing sequence data first proposed by Vaswani et al. in 2017 [28]. Unlike traditional recurrent neural networks, Transformer relies on self-attention mechanisms, enabling the model to process all elements in a sequence in parallel rather than step by step. This feature significantly improves training efficiency and enables the capture of long-range dependencies.

#### 2.5. Evaluation Indicators

In the evaluation of the regression model in this study,  $R^2$ , RMSE, and MAE are three commonly used indicators employed, each reflecting the predictive performance of the model from different perspectives [29].  $R^2$  measures the model's ability to explain the variation in the data; the closer the value is to 1, the better the model's fitting effect. RMSE measures the model's error by calculating the square root of the difference between

the predicted and actual values; the smaller the RMSE, the more accurate the model's prediction. MAE calculates the average of the absolute differences between each predicted and actual value, providing a direct reflection of the model's prediction accuracy and being less affected by extreme values compared with RMSE. This study combines the use of these three indicators to comprehensively assess the predictive ability and accuracy of the PCHE microchannel thermal hydraulic performance prediction model.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

### 3. Experimental Results and Analysis

#### 3.1. Experimental Results of Different Proportions of Dataset Division

Different dataset division methods and proportions can have a certain impact on model performance. The most commonly used division method is random division according to a certain proportion, such as 6:4, 7:3, and 8:2 [30]. This study used three proportions of 6:4, 7:3, and 8:2 to divide the dataset for the heat transfer coefficient prediction of the straight channel and the hexagonal channel. To ensure that both the training set and the test set contained all areas of the heat exchanger, every ten samples were regarded as a group, and the division was carried out within each group. After dividing all groups, the divided data were combined to form the training set and the test set.

In this study, in addition to using RF, SVM, XGBOOST, and CATBOOST, four traditional machine learning models to perform the optimal segmentation ratio experiment, ANN was also used to test the optimal dataset division ratio. From the experimental results, we found that the ratio of 7:3 was dominant in most cases. Further, this study used ANN to conduct further research to explore the optimal segmentation ratio and number of neurons under different layers and neurons. The input layer of this study consisted of 13 features, the output layer was a single target variable, and the number of neurons in the hidden layer varied within the range [25, 50, 75, 150, 200], while the number of hidden layers ranged within [1, 2, 3]. Grid search combined with ten-fold cross-validation was employed to determine the optimal parameters. The experimental results are summarized in Table 2. As shown in Table 2, the model's performance fluctuates as the ratio of the training set to the test set varies. Notably, when the number of hidden layers is set to 2, the 7:3 split achieves the best performance. Under this configuration, the  $R^2$  value reaches 0.888936, and both RMSE (1275.91) and MAE (123.90) exhibit relatively low errors. In contrast, the 8:2 and 6:4 splits yield slightly inferior results. Similarly, in Table 3, the 7:3 split again demonstrates superior performance when the number of hidden layers is 2, particularly in terms of the  $R^2$  value, which attains 0.995199, with RMSE and MAE values of 79.016 and 36.672, respectively. These results are significantly better than those obtained with other ratios.

Therefore, for the straight channel, the 7:3 split represents the optimal proportion. Under this configuration, the model achieves a higher  $R^2$  value and lower prediction errors, suggesting that this ratio enhances the model's generalization capability while effectively mitigating issues of overfitting or underfitting. Similarly, for the hexagonal channel, the 7:3 split outperforms other ratios, yielding a higher  $R^2$  value. This indicates that the division between the training set and test set significantly influences the model's learning performance and predictive accuracy.

**Table 2.** Heat transfer coefficient prediction results of the artificial neural network under different partitioning conditions (8:2, 7:3, 6:4) and different network layer numbers (1, 2, 3).

| Flow Channel Shape            | The Number of Hidden Layers | Proportion | R <sup>2</sup>           | RMSE    | MAE     |
|-------------------------------|-----------------------------|------------|--------------------------|---------|---------|
| Straight-through flow channel | 1                           | 8:2        | 0.811366 ( $\pm 0.023$ ) | 1684.46 | 220.96  |
|                               |                             | 7:3        | 0.888352 ( $\pm 0.015$ ) | 1279.26 | 212.50  |
|                               |                             | 6:4        | 0.872762 ( $\pm 0.018$ ) | 1334.39 | 206.80  |
|                               | 2                           | 8:2        | 0.833303 ( $\pm 0.021$ ) | 1583.49 | 134.61  |
|                               |                             | 7:3        | 0.888936 ( $\pm 0.016$ ) | 1275.91 | 123.90  |
|                               |                             | 6:4        | 0.880276 ( $\pm 0.017$ ) | 1294.39 | 100.16  |
|                               | 3                           | 8:2        | 0.825920 ( $\pm 0.022$ ) | 1618.17 | 140.13  |
|                               |                             | 7:3        | 0.891799 ( $\pm 0.014$ ) | 1259.36 | 128.88  |
|                               |                             | 6:4        | 0.877902 ( $\pm 0.019$ ) | 1307.16 | 95.89   |
| Hexagonal flow channel        | 1                           | 8:2        | 0.971331 ( $\pm 0.012$ ) | 177.025 | 99.994  |
|                               |                             | 7:3        | 0.972071 ( $\pm 0.011$ ) | 190.571 | 115.218 |
|                               |                             | 6:4        | 0.965544 ( $\pm 0.013$ ) | 225.748 | 119.630 |
|                               | 2                           | 8:2        | 0.992383 ( $\pm 0.008$ ) | 91.245  | 36.105  |
|                               |                             | 7:3        | 0.995199 ( $\pm 0.007$ ) | 79.016  | 36.672  |
|                               |                             | 6:4        | 0.992690 ( $\pm 0.009$ ) | 103.981 | 43.366  |
|                               | 3                           | 8:2        | 0.993313 ( $\pm 0.008$ ) | 93.246  | 41.798  |
|                               |                             | 7:3        | 0.994084 ( $\pm 0.007$ ) | 80.419  | 35.863  |
|                               |                             | 6:4        | 0.991977 ( $\pm 0.009$ ) | 108.933 | 47.613  |

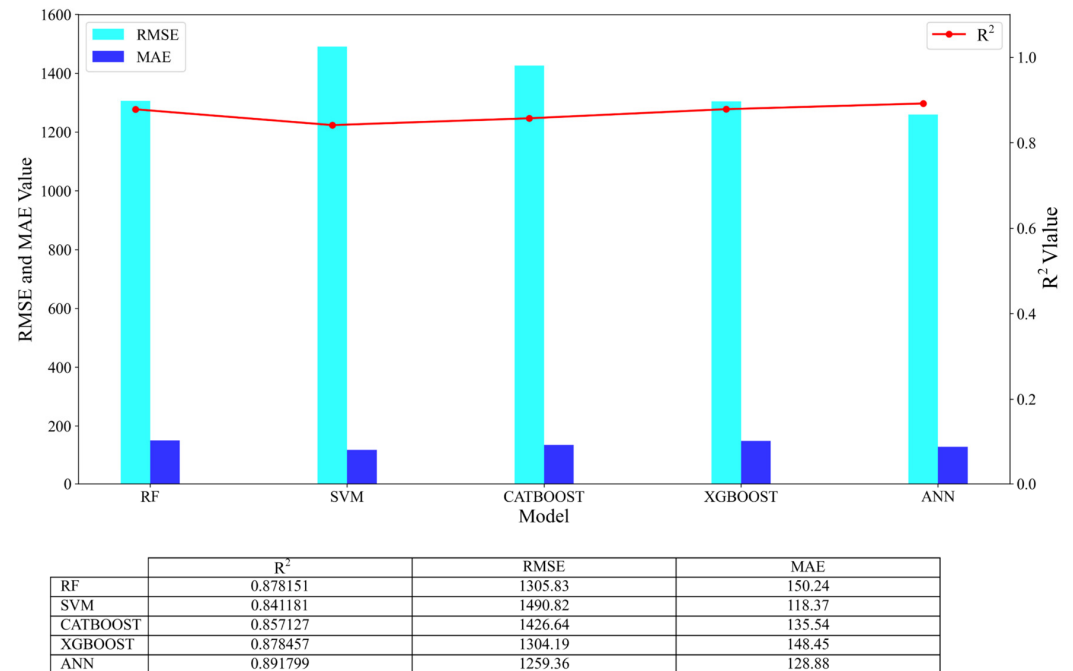
**Table 3.** The modeling results of two different flow channel models before and after the optimization feature.

| Flow Channel Shape            | Feature Selection Methods | Model    | R <sup>2</sup>           | RMSE    | MAE    |
|-------------------------------|---------------------------|----------|--------------------------|---------|--------|
| Straight-through flow channel | None                      | RF       | 0.878151 ( $\pm 0.015$ ) | 1305.83 | 150.24 |
|                               |                           | SVM      | 0.841181 ( $\pm 0.020$ ) | 1490.82 | 118.37 |
|                               |                           | CATBOOST | 0.857127 ( $\pm 0.018$ ) | 1426.64 | 135.54 |
|                               |                           | XGBOOST  | 0.878457 ( $\pm 0.016$ ) | 1304.19 | 148.45 |
|                               |                           | ANN      | 0.891799 ( $\pm 0.014$ ) | 1259.36 | 128.88 |
|                               | Pearson                   | RF       | 0.880556 ( $\pm 0.017$ ) | 1292.86 | 162.91 |
|                               |                           | SVM      | 0.857282 ( $\pm 0.019$ ) | 1413.23 | 123.74 |
|                               |                           | CATBOOST | 0.875124 ( $\pm 0.016$ ) | 1292.53 | 117.41 |
|                               |                           | XGBOOST  | 0.880421 ( $\pm 0.015$ ) | 1293.61 | 187.34 |
|                               |                           | ANN      | 0.882849 ( $\pm 0.014$ ) | 1280.41 | 144.92 |
| Hexagonal flow channel        | None                      | RF       | 0.983842 ( $\pm 0.008$ ) | 144.95  | 62.19  |
|                               |                           | SVM      | 0.988545 ( $\pm 0.007$ ) | 122.05  | 35.28  |
|                               |                           | CATBOOST | 0.991751 ( $\pm 0.006$ ) | 114.13  | 49.47  |
|                               |                           | XGBOOST  | 0.975712 ( $\pm 0.009$ ) | 177.72  | 71.64  |
|                               |                           | ANN      | 0.995199 ( $\pm 0.007$ ) | 79.02   | 36.67  |
|                               | Pearson                   | RF       | 0.985837 ( $\pm 0.008$ ) | 135.71  | 58.59  |
|                               |                           | SVM      | 0.990887 ( $\pm 0.007$ ) | 108.86  | 82.31  |
|                               |                           | CATBOOST | 0.992613 ( $\pm 0.006$ ) | 98.12   | 51.55  |
|                               |                           | XGBOOST  | 0.966728 ( $\pm 0.010$ ) | 208.04  | 95.96  |
|                               |                           | ANN      | 0.991593 ( $\pm 0.007$ ) | 104.55  | 53.81  |

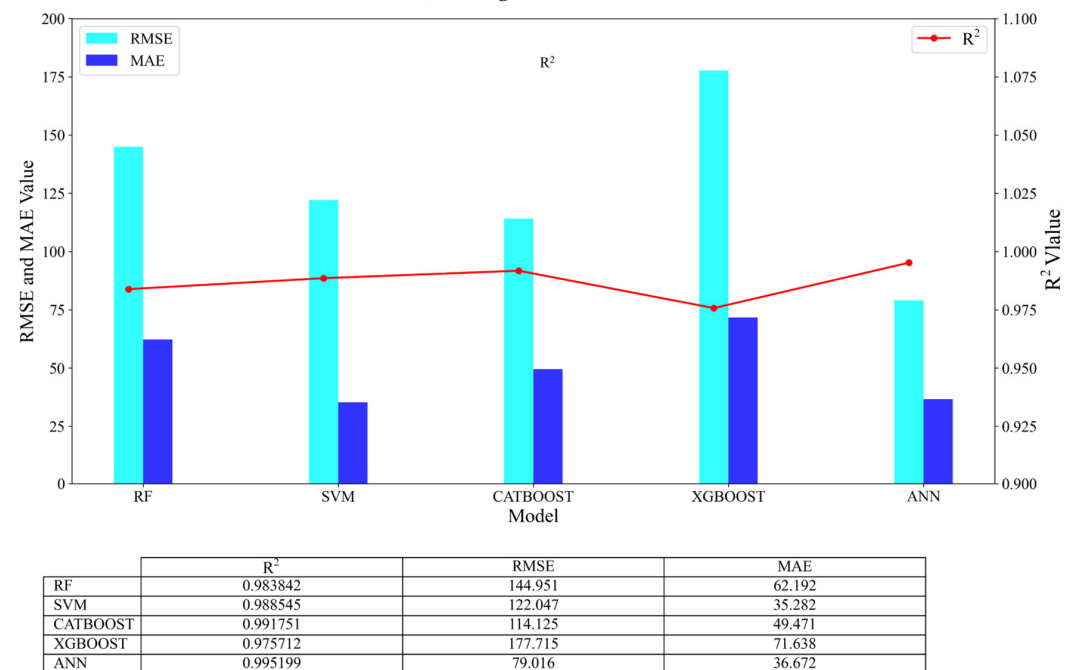
### 3.2. Results of Full Feature Modeling

Figure 4 presents the predicted heat transfer coefficients from different models under two channel shapes. Figure 4a shows the performance of different models in the straight channel modeling. By comparing the RMSE, MAE, and R<sup>2</sup> metrics, the following conclusions can be drawn: In the straight channel modeling, the RMSE values of the RF and SVM models are relatively high, at 1305.83 and 1490.82, respectively, indicating that these two models have relatively large prediction errors. The ANN model has the lowest RMSE value, at 1259.36, demonstrating better prediction accuracy. In the straight channel modeling, the ANN model has the highest R<sup>2</sup> value, at 0.891799, indicating that this model

has the best data fitting degree. In contrast, the  $R^2$  value of the SVM model is the lowest, at 0.841181, showing relatively poor fitting performance. Considering both the error and fitting degree comprehensively, the ANN model performs best in the straight channel modeling, followed by the CATBOOST model, with an  $R^2$  value of 0.857127 and an RMSE value of 1426.64, demonstrating a relatively good balance of performance.



(a) Straight flow channel



(b) Hexagonal flow channel

**Figure 4.** Modeling results of two flow channel working conditions with different models (The segmentation ratio was 7:3, and  $R^2$ , RMSE, and MAE were used as evaluation indicators to evaluate all traditional machine learning models).

Figure 4b shows the performance of different models in the modeling of hexagonal channels. By comparing the RMSE, MAE, and  $R^2$  indicators, the following conclusions can be drawn: In the modeling of hexagonal channels, the XGBOOST model has the highest RMSE value, 177.715, indicating a relatively large prediction error. In contrast, the ANN model has the lowest RMSE value, 79.016, demonstrating the best prediction accuracy. In the modeling of hexagonal channels, the ANN model has the highest  $R^2$  value, 0.995199, indicating the best fit to the data. Relatively speaking, the XGBOOST model has the lowest  $R^2$  value, 0.975712, showing a relatively poor fitting effect. Considering both error and fitting degree comprehensively, the ANN model performs best in the modeling of hexagonal channels, followed by the CATBOOST model, which has an  $R^2$  value of 0.991751 and an RMSE of 114.145, demonstrating a relatively good balance of performance.

### 3.3. Feature Selection Set Modeling Results

The Pearson correlation coefficient was used to calculate the correlation between the features and between features and the target value, as shown in Figure 5. Based on experience, a Pearson correlation coefficient greater than 0.5 is considered to have a high correlation with the target. Therefore, all features with a correlation greater than 0.5 with the convective heat transfer coefficient  $h$  were selected. Thus,  $D$ ,  $V_q$ ,  $Pin$ ,  $Q_w$ ,  $RE$ , and  $Dh$  were chosen as the sensitive input features for the straight channel. There was a strong correlation between  $T_{in}$  and  $h$  ( $-0.73$ ),  $Pin$  and  $h$  ( $0.75$ ),  $T_{inw}$  and  $h$  ( $-0.78$ ),  $V_{in}$  and  $h$  ( $-0.59$ ),  $pin$  and  $h$  ( $0.75$ ), and  $PR$  and  $h$  ( $0.78$ ). Therefore,  $T_{in}$ ,  $Pin$ ,  $T_{inw}$ ,  $V_{in}$ ,  $pin$ , and  $PR$  were selected as the sensitive input features for the hexagonal channel.

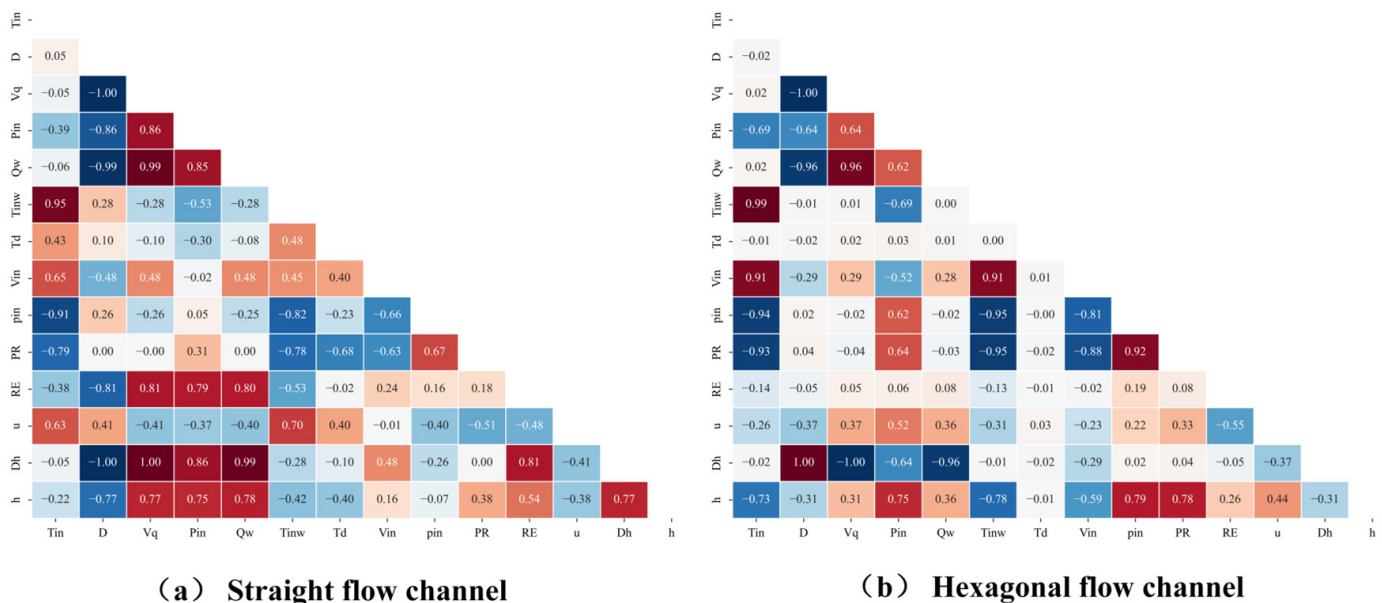


**Figure 5.** Pearson correlation coefficient heat map: (a) straight flow channel; (b) hexagonal flow channel.

The results of modeling the performance of five models in predicting the heat transfer coefficient under different channel shapes and feature selection methods are summarized in Table 3. Under the straight channel condition, although all models performed relatively well, the ANN model achieved the highest  $R^2$  value (0.891799) and the lowest RMSE (1259.36), indicating its superior fitting ability and prediction accuracy for this channel shape. When the channel shape changed to a hexagonal channel, the performance of all models significantly improved, with  $R^2$  values generally approaching or exceeding 0.99. Specifically, the optimal ANN model achieved the highest  $R^2$  value (0.995199) and the

lowest RMSE (79.02), demonstrating that the hexagonal channel has a substantial positive impact on enhancing the accuracy of the heat transfer coefficient prediction.

Among the two channel shapes, the improvement effect of Pearson on most models was relatively limited. In particular, for the straight channel, the  $R^2$  values and error metrics of some models either did not show significant improvement or even deteriorated after feature selection. In contrast, for the hexagonal channel, although the model performance remained strong post-feature selection, certain models (e.g., SVM and CATBOOST) continued to exhibit high prediction accuracy. This suggests that feature selection still holds a certain optimization value in complex structures.

### 3.4. Deep Learning Modeling Results

#### 3.4.1. Basic Deep Learning Modeling Results

Deep learning models, such as CNNs, possess a large number of hyperparameters that directly influence model performance, including the learning rate and batch size. Through systematic search, the optimal combination of these hyperparameters can be identified to enhance the model's performance on the validation or test set. However, hyperparameter optimization typically demands substantial computational resources. By prioritizing the search for commonly used parameters with the greatest impact on model performance, the most effective results can be achieved within limited resource constraints. Therefore, based on fundamental modeling experience and the literature references, the range of hyperparameters utilized in the deep learning model of this study is presented in Table 4.

**Table 4.** Hyperparameter range of each deep learning model.

| Model                                    | Hyperparameter Range                                                                                                |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| CNN                                      | batch_size = [8, 16, 32, 64, 128, 256]<br>learning_rate = [0.01, 0.001, 0.0001]<br>Dropout_rate = [0.01, 0.05, 0.1] |
|                                          | batch_size = [8, 16, 32, 64, 128, 256]<br>learning_rate = [0.01, 0.001, 0.0001]<br>dropout_rate = [0.01, 0.05, 0.1] |
| GRU, LSTM, BiLSTM,<br>BiGRU, Transformer | hidden_dim = [16, 32, 64]<br>num_layers = [1, 2, 3]                                                                 |

After the grid search, the modeling results of six deep learning models are shown in Figure 6. The first column represents  $R^2$ , the second column represents RMSE, and the third column represents MAE. The first row is for the straight channel, and the second row is for the hexagonal channel. For the straight channel, LSTM shows the highest prediction accuracy, with an  $R^2$  of 0.89317, RMSE of 1219.79, and MAE of 118.88. Next are the BiGRU and BiLSTM models, both with  $R^2$  values exceeding 0.8923, and their fitting effects are also good. In the prediction results of the hexagonal channel, the fitting effect of LSTM is still relatively good, but the BiGRU and BiLSTM models maintain the highest accuracy results, with  $R^2$  values of 0.99844 and 0.99835, respectively. The Transformer and CNN models perform relatively poorly, which may be due to certain limitations in handling complex spatial relationships and sequence features. To improve the performance of these models, it may be necessary to introduce an attention mechanism to enhance the models' ability to capture key features and thereby improve prediction accuracy.

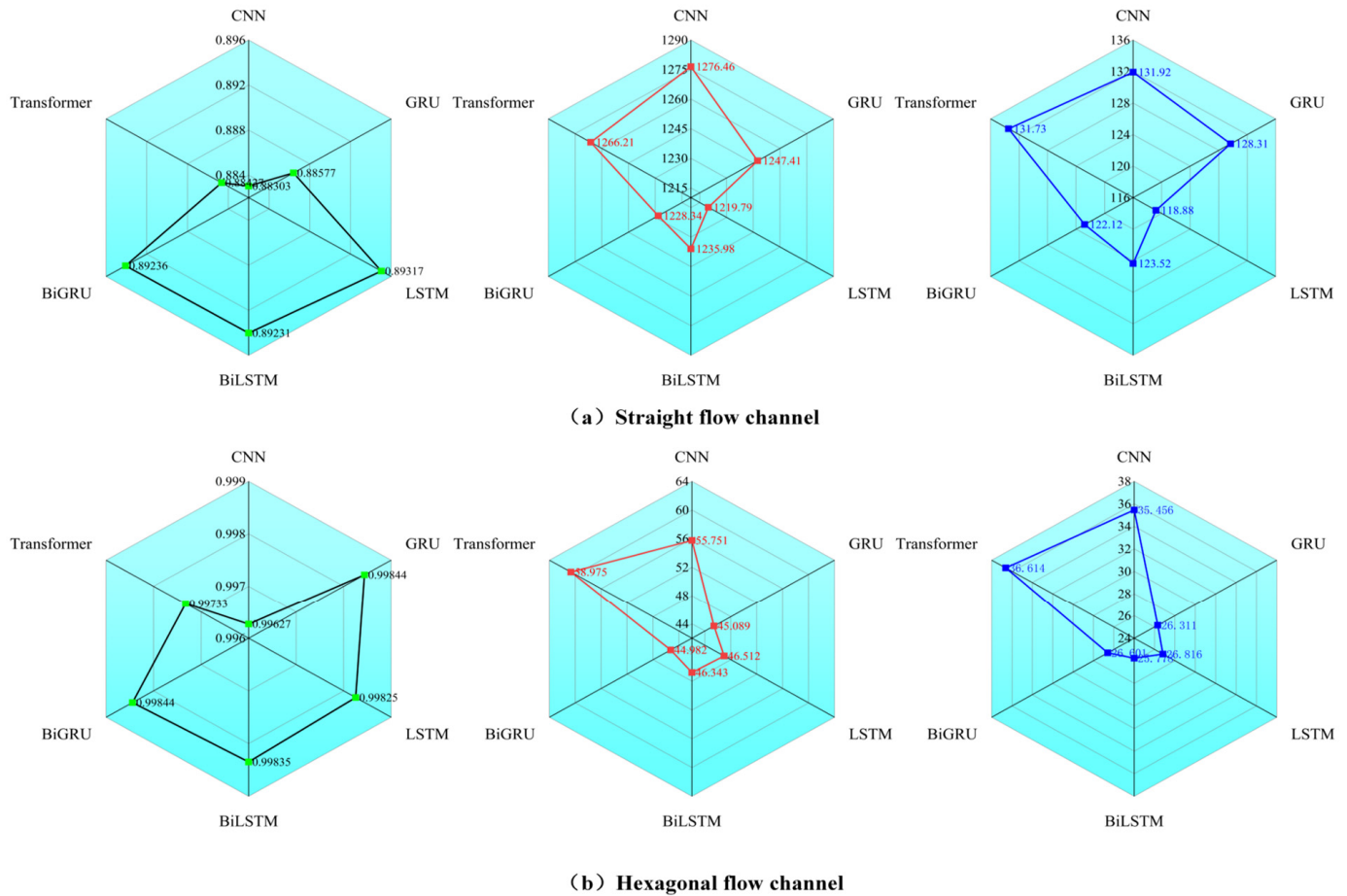


**Figure 6.** Modeling results of deep learning model. (From left to right, they are  $R^2$ , RMSE, and MAE.) (a) Straight flow channel; (b) hexagonal flow channel.

### 3.4.2. Modeling Results of Deep Learning Models Based on Attention Mechanism

To investigate the impact of different attention mechanisms on model performance, this study incorporated three attention mechanisms: self-attention (SA), squeeze-and-excitation (SE) attention, and local attention (LA). The benchmark models were based on three distinct types of deep learning architectures: convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and Transformers.

#### (1) SA Attention Mechanism

Self-attention (SA) is utilized to capture the dependencies among elements within a sequence [31]. Its core concept involves calculating the correlation weights between each element and all other elements in the sequence and dynamically assigning attention weights to each element, thereby integrating global information. The schematic diagram of SA is presented in Figure 7.

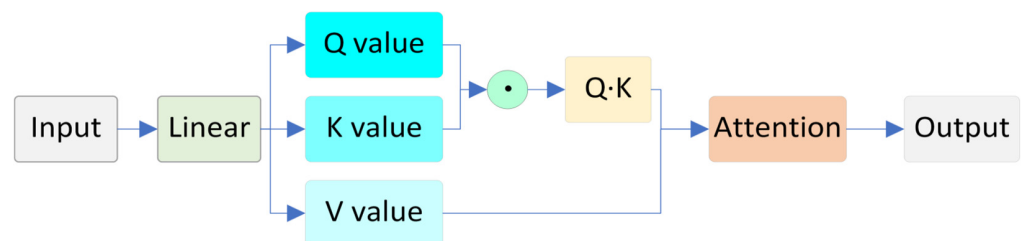


**Figure 7.** Schematic diagram of SA attention mechanism.

First, each element of the input sequence is generated through three linear transformations:

$$Q = XW^Q \quad (5)$$

$$K = XW^K \quad (6)$$

$$V = XW^V \quad (7)$$

Among them,  $X$  is the input matrix, and  $Q$ ,  $K$ , and  $V$  are the query vector, key vector, and value vector, respectively.  $W^Q$ ,  $W^K$ , and  $W^V$  are the learnable parameter matrices.

Then, calculate the dot product of  $Q$  and  $K$ , and normalize it through Softmax after scaling:

$$Attention(Q, K, V) = Softmax\left(\frac{QK}{\sqrt{d_k}}\right)V \quad (8)$$

Here,  $d_k$  represents the dimension of the key vector, which is used for scaling to prevent gradient explosion. Softmax normalizes the weights into a probability distribution. Then, the value vectors are summed up through weighted averaging:

$$Z = Attention(Q, K, V) \quad (9)$$

$Z$  represents the output of the self-attention layer.

## (2) SE Attention Mechanism

The squeeze-and-excitation (SE) attention mechanism is a module designed to enhance the feature representation capability of convolutional neural networks that comprises two key steps: squeeze and excitation [32]. Its core concept involves explicitly modeling inter-channel dependencies and dynamically recalibrating the weights of each channel, thereby amplifying important features while suppressing less relevant ones. A schematic illustration of the SE attention mechanism is presented in Figure 8.

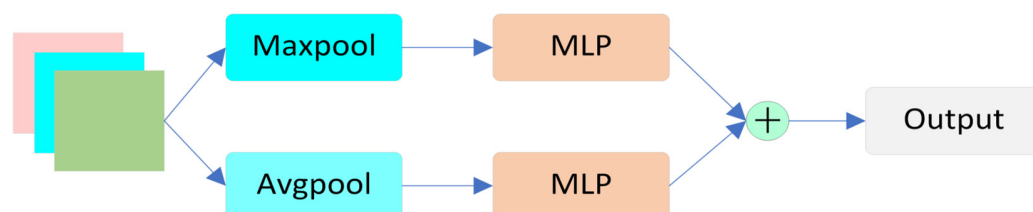


**Figure 8.** Schematic diagram of SE attention mechanism.

The squeeze operation performs global average pooling (GAP) on the input feature map  $X$  (with a shape of  $H \times W \times C$ ), compressing the spatial information of each channel into a scalar:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (10)$$

Here,  $x_c(i, j)$  is the feature value of the  $c$ th channel at position  $(i, j)$ ;  $z_c$  is the global feature descriptor of the  $c$ th channel; and  $z$  is the output vector with a shape of  $1 \times 1 \times C$ .

Then, the excitation operation learns the nonlinear relationships among channels through two fully connected layers (FC) to generate channel weights:

$$s = \sigma(W_2 \delta(W_1 z)) \quad (11)$$

Here,  $W_1$  and  $W_2$  are the weight matrices of two fully connected layers,  $\delta$  is the ReLU activation function,  $\sigma$  is the Sigmoid activation function, the weights are normalized to  $[0, 1]$ , and  $s$  is the output vector with a shape of  $1 \times 1 \times C$ .

Finally, the feature recalibration operation multiplies the learned channel weights  $s$  with the original feature map  $X$ , channel by channel, to obtain the enhanced feature map:

$$\hat{x}_c = s_c \cdot x_c \quad (12)$$

Here,  $x_c$  is the enhanced feature map of the  $c$ -th channel, and  $s_c$  is the weight of the  $c$ -th channel.

### (3) LA Attention Mechanism

The LA mechanism is a mechanism that computes attention weights only for local regions in sequence data or image data [33]. Unlike global attention, self-attention mechanisms, and squeeze-and-excitation operations, local attention focuses only on a subset of the input sequence, thereby reducing computational complexity and improving efficiency. The network structure of the LA attention mechanism is shown in Figure 9.

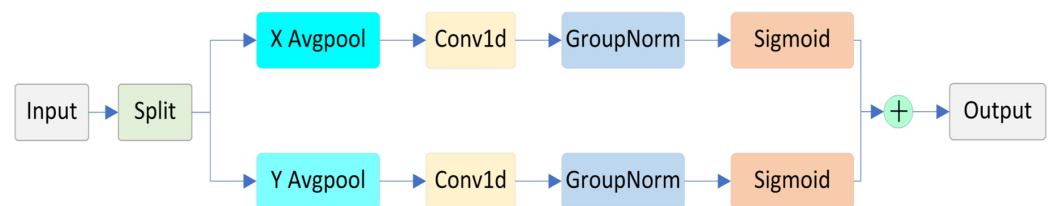


**Figure 9.** Schematic diagram of the LA attention mechanism.

Firstly, LA defines a local window. For each position  $i$  in the sequence, a local window  $[i - D, i + D]$  is defined, where  $D$  is the window radius. The position  $j$  within the window satisfies the following:

$$j \in [i - D, i + D] \quad (13)$$

Then, calculate the attention weights by computing the dot product of the query  $Q_i$  and the key  $K_j$  within the local window and normalizing it through Softmax:

$$\alpha_{ij} = \text{Softmax}\left(\frac{Q_i K_j}{\sqrt{d_k}}\right) \quad (14)$$

Here,  $Q_i$  is the query vector at position  $i$ ,  $K_j$  is the key vector at position  $j$ , and  $d_k$  is the dimension of the key vector, which is used for scaling.

Finally, perform a weighted summation operation. Within the local window, conduct a weighted summation of the values  $V_j$  to obtain the output at position  $i$ :

$$z_c = \sum_{j=i-D}^{i+D} \alpha_{ij} V_j \quad (15)$$

The attention mechanism was used to enhance three typical deep learning models, namely, the CNN model, the LSTM model in RNN, and the Transformer model, which is more effective at processing sequence data. The modeling results are shown in Table 5. The results indicate that the combination of different models and attention mechanisms has a significant impact on task performance. Among CNN, LSTM, and Transformer, the models combined with the SE attention mechanism all showed performance improvements. For instance, the  $R^2$  value of LSTM-SE reaches 0.9387, and the RMSE and MAE were reduced to 1102.43 and 105.94, respectively, significantly outperforming other attention mechanisms. From the results of CNN-SE and Transformer-SE, it was also found that the performance of the models improved after adding SE, indicating that the SE mechanism can effectively enhance feature

representation by explicitly modeling the inter-channel dependencies. Additionally, the SA mechanism brought performance improvements in both LSTM and Transformer, but it performed slightly worse in CNN. For example, the  $R^2$  value of LSTM-SA was 0.9034, which is better than the basic LSTM model. However, the  $R^2$  value of CNN-SA slightly decreased (0.8735), suggesting that the SA mechanism may have limited improvement on the feature extraction ability of CNN. Meanwhile, it was found that the local attention mechanism LA combined with each model performed poorly in the prediction of the heat transfer coefficient on the straight road. For instance, the  $R^2$  value of LSTM-LA was only 0.8254, and the RMSE and MAE increased to 1608.54 and 211.49, respectively. This indicates that the local attention mechanism may not be able to fully capture the global dependencies, leading to a decline in performance.

**Table 5.** Prediction results of the heat transfer coefficient of the straight flow channel based on deep learning with attention mechanism.

| Model       | $R^2$    | RMSE    | MAE    |
|-------------|----------|---------|--------|
| CNN         | 0.883029 | 1276.46 | 131.92 |
| +SA         | 0.873534 | 1353.64 | 151.79 |
| +SE         | 0.883679 | 1285.78 | 135.95 |
| +LA         | 0.843451 | 1451.12 | 186.87 |
| LSTM        | 0.893166 | 1219.79 | 118.88 |
| +SA         | 0.903425 | 1208.32 | 113.48 |
| +SE         | 0.938732 | 1102.43 | 105.94 |
| +LA         | 0.825413 | 1608.54 | 211.49 |
| Transformer | 0.883673 | 1266.21 | 131.73 |
| +SA         | 0.886291 | 1268.49 | 149.13 |
| +SE         | 0.914283 | 1155.07 | 128.27 |
| +LA         | 0.833451 | 1480.45 | 175.39 |

The LSTM-SE with the best performance has 3 num\_layers, 32 hidden\_dim, 0.1 dropout\_rate, and 16 batch\_size.

For the hexagonal flow channel, the same attention embedding operation was performed, and the results are summarized in Table 6. The CNN model achieved an  $R^2$  value of 0.996270, an RMSE of 55.751, and an MAE of 35.456. After incorporating the attention mechanism, the  $R^2$  values of the CNN-SA, CNN-SE, and CNN-LA models slightly increased, while the RMSE and MAE values also rose, suggesting that the attention mechanism has a limited positive impact on the performance of the CNN model. Among these, the LA attention mechanism demonstrated a more noticeable improvement in the CNN model. Meanwhile, it was observed that for the LSTM and Transformer models, the introduction of the attention mechanism resulted in a decline in model performance. However, it was also noted that the LA attention mechanism contributed to enhancing the accuracy of both models, particularly for the Transformer model, where the accuracy significantly improved, with the  $R^2$  increasing from 0.992415 to 0.999257. This indicates that the LA attention mechanism exhibits strong compatibility with the Transformer model.

**Table 6.** Prediction results of the heat transfer coefficient of the hexagonal flow channel based on deep learning with attention mechanism.

| Model       | R <sup>2</sup> | RMSE    | MAE    |
|-------------|----------------|---------|--------|
| CNN         | 0.996270       | 55.751  | 35.456 |
| +SA         | 0.996367       | 68.737  | 40.859 |
| +SE         | 0.996446       | 62.744  | 38.157 |
| +LA         | 0.997125       | 60.147  | 36.806 |
| LSTM        | 0.998251       | 46.512  | 26.816 |
| +SA         | 0.987895       | 146.027 | 97.554 |
| +SE         | 0.983603       | 161.526 | 98.774 |
| +LA         | 0.997325       | 58.975  | 36.614 |
| Transformer | 0.992415       | 97.613  | 39.735 |
| +SA         | 0.989617       | 131.531 | 68.774 |
| +SE         | 0.998861       | 56.820  | 35.138 |
| +LA         | 0.999257       | 39.561  | 26.005 |

The Transformer–LA with the best performance has 2 num\_layers, 64 hidden\_dim, 0.1 dropout\_rate, 0.001 learning\_rate, and 8 batch\_size.

### 3.5. Explainable Analysis

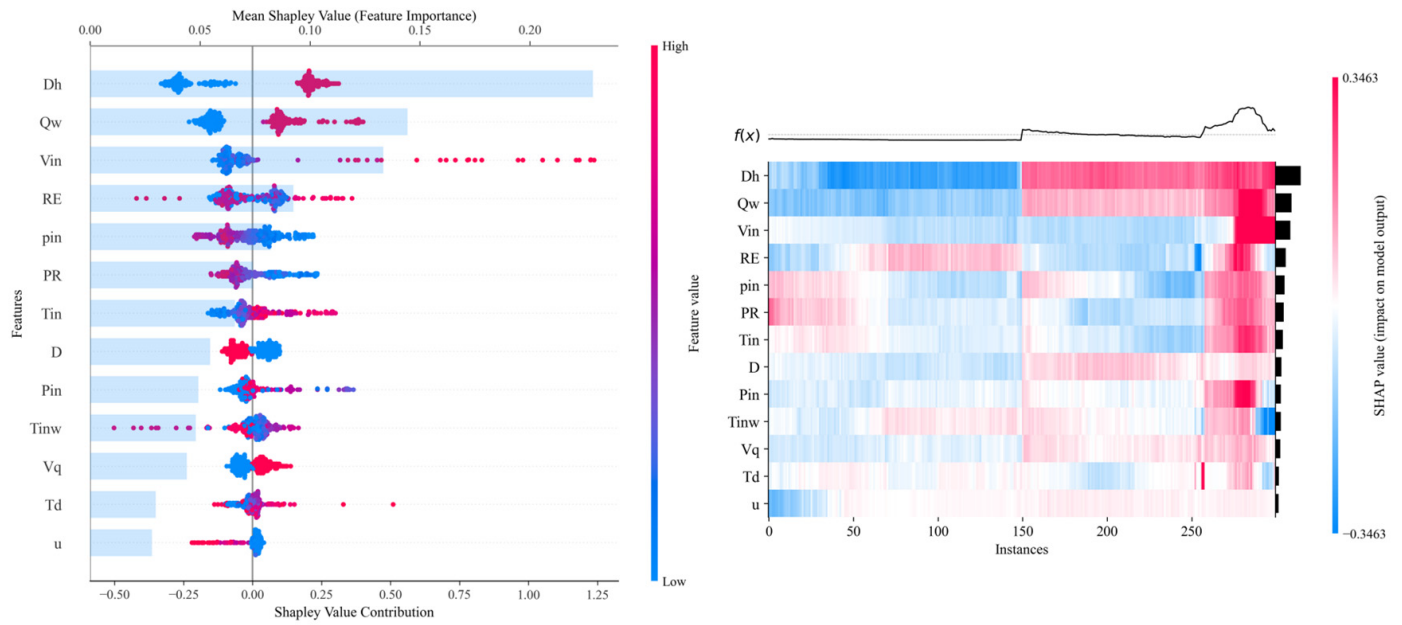
SHAP (SHapley Additive exPlanations) is a technique utilized to interpret model predictions that is grounded in the Shapley value from game theory, which quantifies the contribution of each feature to the model output. For deep learning models, Deep SHAP serves as an interpretation tool [34]. Deep SHAP is a SHAP explainer specifically tailored for deep learning models, integrating the principles of SHAP with the architecture of deep neural networks. By incorporating feature importance analysis into deep learning models, Deep SHAP effectively quantifies the contribution of each feature to the model's prediction outcomes. Based on the core concept of the SHAP algorithm, Deep SHAP measures feature contributions by occluding (or setting to zero) specific features for a given input and analyzing the activation process layer by layer within the neural network. Specifically, the shap. DeepExplainer module was employed for interpreting deep learning models.

#### 3.5.1. Global Interpretation

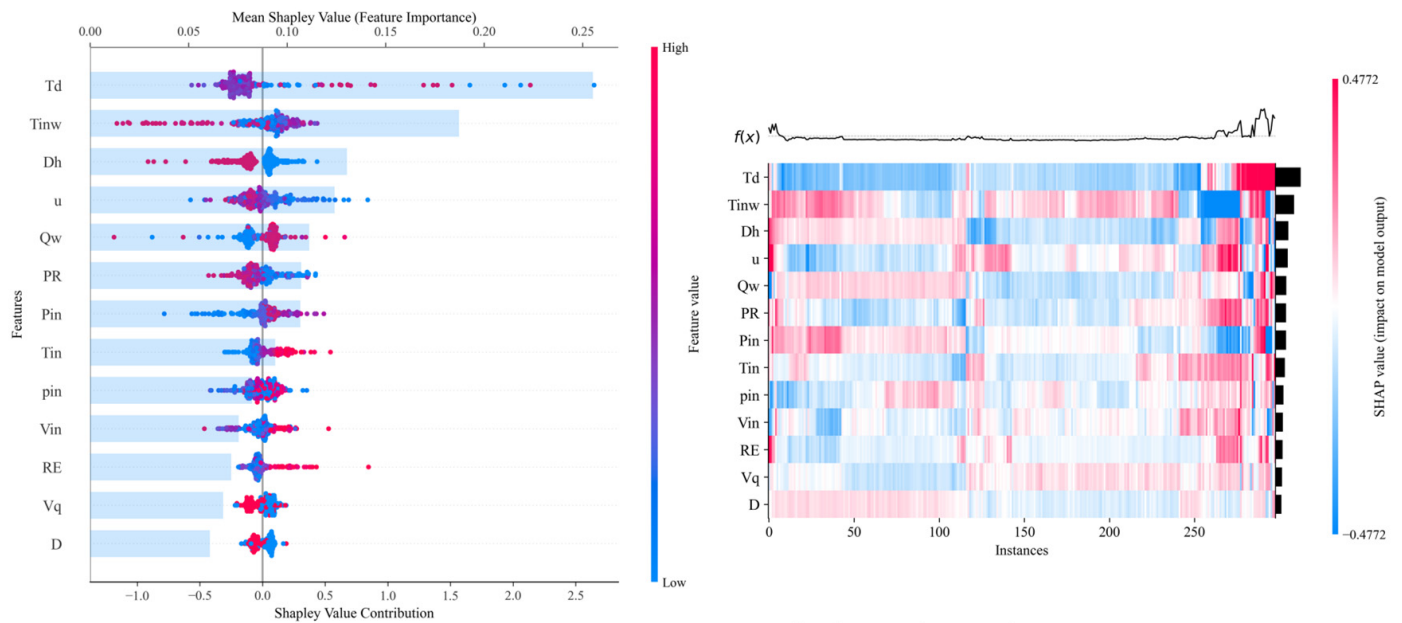
Global interpretation analysis was performed on the Transformer–SE model in the straight channel ( $R^2 = 0.914$ ) and the Transformer–LA model in the hexagonal channel ( $R^2 = 0.999$ ). SHAP statistics plots and heatmaps were generated for all features, as presented in Figure 10. The analysis of the two channel shapes—namely, the straight channel (Figure 10a) and the hexagonal channel (Figure 10b)—revealed substantial differences in feature importance and their contributions to the model output.

Firstly, from the SHAP statistics plot (left), it can be observed that for the straight channel, “Dh” contributes most significantly to the model output. Other features, such as “RE”, “Qw”, “Tinw”, “PR”, “D”, and “pin”, also exhibit notable contributions but to a lesser degree compared with “Dh”. In contrast, for the hexagonal channel, “Tinw” emerges as the most important feature, followed by “Qw”, “PR”, and “pin”. The importance of features like “D”, “Dh”, and “Td” is relatively low.

From the SHAP heatmap (right), it can be observed that the model output ( $f(x)$ ) of the heat transfer coefficient for the straight channel steadily increases as the number of instances grows, indicating a strong correlation with the feature values. In contrast, the model output ( $f(x)$ ) of the heat transfer coefficient for the hexagonal channel exhibits a more fluctuating pattern, suggesting a more complex relationship between the features and the output.



(a) Straight Flow Channel



(b) Hexagonal Flow Channel

**Figure 10.** Statistical diagram of SHAP. (The blue background on the left represents feature contribution. Each point represents a sample, and the color represents whether the corresponding feature on that sample contributes positively or negatively. In the figure on the right, the horizontal coordinate is the sample serial number, and the black column on the right represents the feature contribution degree.) (a) Straight flow channel; (b) hexagonal flow channel.

### 3.5.2. Local Interpretation

The SHAP local interpreter was employed to perform local interpretation on four samples, including continuous samples from two types of flow channels. The SHAP waterfall plots are presented in Figure 11. It can be observed that there are substantial differences in feature importance between the straight channel and the hexagonal channel within a single sample. In the two samples of the straight channel, the features RE and Dh exhibit the greatest impact on the model output, whereas in the hexagonal channel, the

features  $u$  and  $Tinw$  demonstrate a more significant influence. Additionally, even within the same type of flow channel, feature importance varies across different samples. For instance, in the straight channel, features  $RE$  and  $Dh$  have a greater impact on sample 75, while their influence is relatively smaller on sample 76, with an increased contribution from feature  $Tinw$ . Moreover, the direction of feature contributions to the model output also differs across the different flow channels and samples. For example, feature  $RE$  predominantly contributes negatively in the straight channel but positively in the hexagonal channel.

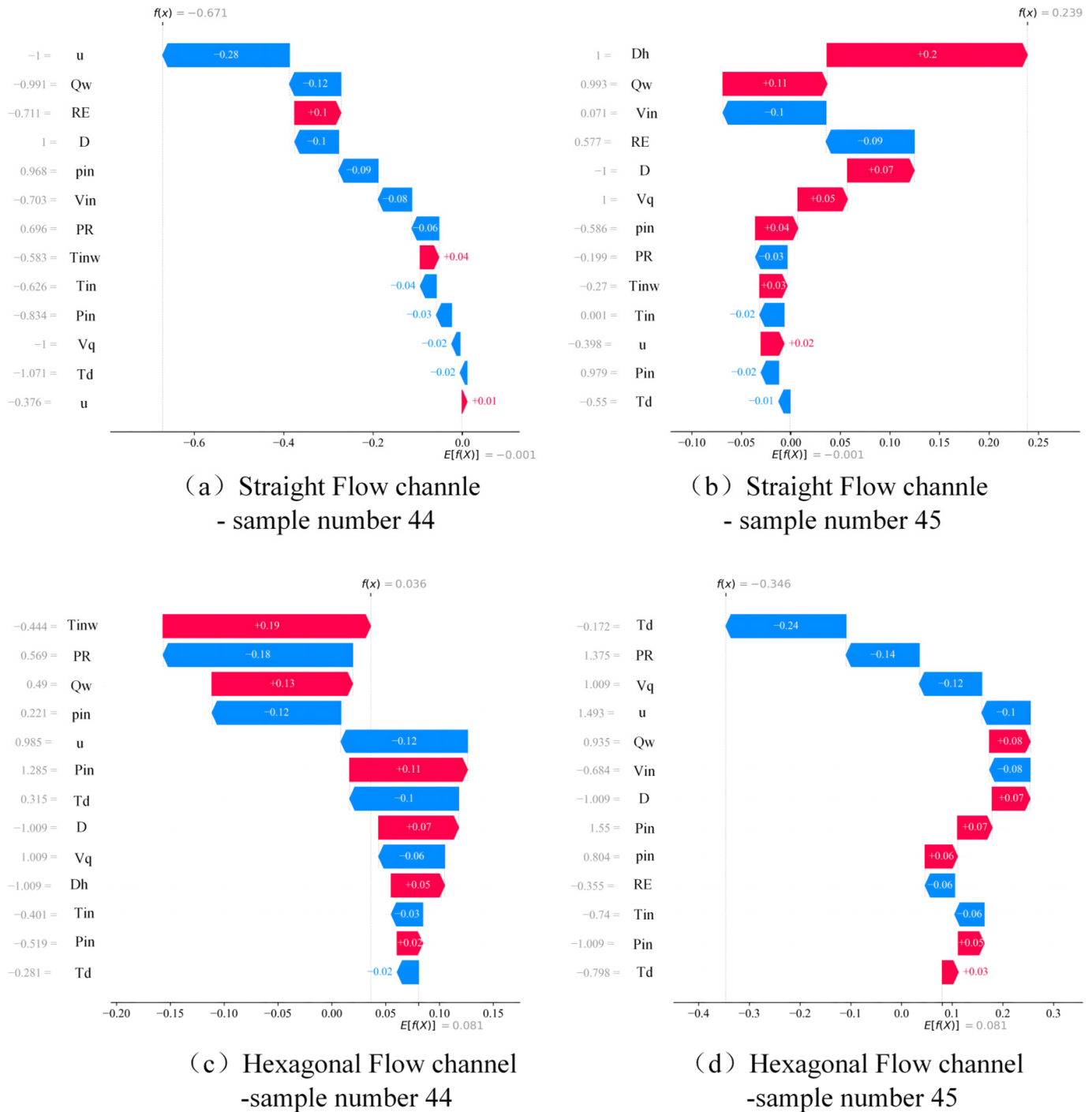


**Figure 11.** SHAP local interpretation results of four samples ( $X$  represents the SHAP value, and  $E[f(x)]$  represents the expectation of all samples  $f(x)$ ): (a) Straight flow channel, sample number 44; (b) straight flow channel, sample number 45; (c) hexagonal flow channel, sample number 44; (d) hexagonal flow channel, sample number 46.

From the results of both global and local explanations, in practical applications, the appropriate features should be selected for optimization based on the specific type of flow channel and sample characteristics to accurately predict the heat transfer coefficient. For straight flow channels, Dh, RE, Qw, Tin, PR, and D should be given priority; for hexagonal flow channels, Tinw, Qw, PR, and pin should be emphasized.

#### 4. Discussion

This study successfully constructed a high-precision PCHE heat transfer coefficient prediction model by systematically comparing traditional machine learning and deep learning models and integrating attention mechanisms and explainability analysis methods, providing a new methodological framework for complex heat exchanger performance modeling. The research results show that the deep learning model based on the attention mechanism has a significantly higher prediction accuracy ( $R^2 \geq 0.999$ ) in hexagonal channels than traditional empirical correlations, verifying the potential of deep learning in capturing nonlinear thermohydraulic relationships. Additionally, SHAP analysis revealed the dynamic contribution differences of key features under different channel types, providing an explainable physical basis for engineering optimization. Compared with existing studies, the innovation of this work lies in the following: (1) the first introduction of self-attention mechanisms and Transformer into PCHE performance prediction, breaking the traditional model's reliance on local sequence features; (2) the construction of a high-frequency dataset through the distance micro-element analysis method, solving the limitations of traditional averaging methods in complex channels; (3) a systematic evaluation of the impact of feature selection on model generalization ability, providing a standardized process reference for data-driven heat exchanger design. Finally, the numerical simulation stage took about 12 h, and the training and verification of each set of parameters in deep learning was counted in minutes, and we used NVIDIA GeForce RTX 4060.

It can also be observed from the results in Tables 5 and 6 that when the same model is applied to predict the PCHE heat transfer coefficient, the prediction performance for the DC channels and hexagonal channels differs significantly. On hexagonal flow channels, even the least effective XGBOOST model achieves high accuracy ( $R^2 \geq 0.96$ ), while the best-performing Transformer-LA model attains an accuracy exceeding 0.999. In contrast, for DC channels, the prediction accuracy is generally below 0.9. The explanation provided in this study is that hexagonal channels exhibit relatively complex flow patterns and temperature field changes, such as the formation of local vortices and enhanced thermal convection, which increases data diversity and complexity, enabling models to fully leverage the features within the flow channel. For DC channels, the flow pattern is overly simplistic, and the data exhibit strong linearity, leading to model overfitting and reduced prediction accuracy [35]. Additionally, we found that SE performed better on straight flow channels, whereas LA demonstrated superior performance on hexagonal flow channels when attention mechanisms were incorporated. This indicates that the SE mechanism effectively enhances feature representation by explicitly modeling inter-channel dependencies, while LA focuses on local information in the complex variations in hexagonal flow channels, resulting in differing performances.

According to SHAP analysis, two independent variables—wall heat flux density (Qw) and Prandtl number (PR)—make significant contributions to the heat transfer characteristics of both flow channel shapes. Qw serves as a quantitative measure of energy transport in PCHE, and variations in its value directly influence the steepness of the temperature gradient in the boundary layer and the enhancement effect of turbulent pulsation on heat transport in the near-wall region [36]. For PR, it reflects the relative strengths of momentum and heat diffusion. When PR is large, momentum diffusion dominates, leading

to a relatively thick flow boundary layer and slower heat transfer. Conversely, when PR is small, heat diffusion prevails, resulting in a relatively thick thermal boundary layer and slower momentum transfer [37]. These differences in behavior significantly impact the heat transfer process, enabling machine learning and deep learning models to effectively distinguish the heat transfer characteristics of different fluids based on these features.

Although this study has achieved significant results, there are still certain limitations that require further improvement in future research. Firstly, the scale and diversity of the dataset are restricted. The dataset used in this study primarily originates from numerical simulation results under specific operating conditions, with a relatively limited number of samples and predominantly focused on specific channel shapes and sizes. In the future, the dataset can be expanded by incorporating data from diverse operating conditions, various channel configurations, and extreme scenarios to enhance the model's generalization capability. Secondly, the interpretability of the model requires further enhancement. While the SHAP method provides a degree of interpretability analysis, the understanding of the internal mechanisms of deep learning models remains limited. In the future, additional interpretability techniques, such as feature importance analysis and activation function visualization, can be integrated to further elucidate the internal workings of the model. Additionally, this study primarily focuses on predicting heat transfer coefficients but lacks in-depth exploration of other critical thermodynamic performance parameters, such as pressure drop and thermal efficiency. Future research can broaden its scope to predict and optimize these parameters for a comprehensive evaluation of PCHE performance. Therefore, future research will focus on the following four aspects: (1) Expanding the dataset to include multiple working fluids, varied channel configurations, and extreme operating conditions to improve the engineering versatility of the model; (2) Overcoming the interpretability limitations by integrating physics-informed neural networks (PINNs); (3) Investigating the combination of multimodal attention mechanisms and graph neural networks (GNN) to more accurately represent the coupling effects between channel topology and local flow characteristics; (4) Exploring alternative deep learning models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs) to further enhance prediction performance.

## 5. Conclusions

This study primarily investigated the prediction of methane transcritical heat transfer coefficients under various channel shapes using machine learning approaches. Through a comprehensive analysis of data preprocessing, modeling outcomes of different models, feature selection results, and deep learning modeling performance, the following conclusions were drawn:

(1) Different dataset division ratios significantly influence model performance. By comparing the ratios of 6:4, 7:3, and 8:2, it was found that the 7:3 ratio yielded the best performance in predicting heat transfer coefficients for both straight channels and hexagonal channels, achieving higher  $R^2$  values and lower errors.

(2) Among the modeling results of the different models, the artificial neural network (ANN) model performed outstandingly in both channel shapes, whether in full-feature modeling or after feature selection, demonstrating excellent  $R^2$  and MAE values. (In the straight flow channel,  $R^2$  is 0.882849; in the hexagonal flow channel,  $R^2$  is 0.991593.) In contrast, other models such as random forest (RF), support vector machine (SVM), and XGBOOST were slightly inferior in prediction accuracy and fitting degree.

(3) Among deep learning models, the LSTM and Transformer models demonstrated strong performance in predicting heat transfer coefficients for both channel types. Particularly, models incorporating attention mechanisms, such as LSTM-SE (achieving the highest  $R^2$

value of 0.938732 for the straight flow channel) and Transformer-LA (achieving the highest  $R^2$  value of 0.999257 for the hexagonal flow channel), exhibited substantial performance improvements in both straight and hexagonal channels. This suggests that attention mechanisms can effectively enhance the model's capacity to capture key features, thereby improving prediction accuracy.

(4) The SHAP analysis results indicated that the contributions of different channel types and sample features to model outputs vary significantly. In straight channels, features such as “Dh”, “RE”, and “Qw” have a greater impact on model outputs, while in hexagonal channels, features like “Tinw”, “Qw”, and “PR” are more important. Local explanation results further revealed that the importance of features varies among different samples, emphasizing the need to select the appropriate features for optimization based on the specific channel types and sample features in practical applications.

In summary, this study proposes a data acquisition approach to obtain more extensive datasets; furthermore, we explore a variety of machine learning and deep learning models to validate the potential of deep learning models with attention mechanisms. Additionally, SHAP-based interpretable analysis offers valuable insights into understanding model predictions, thereby facilitating further model optimization and feature engineering. Meanwhile, this study highlights the primary direction for future research in this field, namely, integrating physical information with neural network models to achieve more accurate predictions of flow characteristics or exchange characteristics. The findings of this study provide a critical theoretical foundation and practical guidance for predicting methane transcritical heat transfer coefficients.

**Author Contributions:** Conceptualization, Y.S. and L.Z.; methodology, Y.S., Y.Z. and L.Z.; validation, Y.S. and L.Z.; formal analysis, Y.S.; investigation, Y.S. and J.W.; data curation, Y.S., Y.Z. and J.W.; writing—original draft preparation, Y.S.; writing—review and editing, J.W.; visualization, Y.S. and J.W.; supervision, Y.Z. and L.Z.; project administration, L.Z.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the major science and technology project of Hainan Province (ZDYF2023GXJS148). At the same time, We would like to express my sincere thanks to the professors who provided guidance for this research.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The code used in this study can be obtained from the corresponding authors. Dataset from [https://github.com/mirror52/test1/blob/main/data\\_1.2mm-1.8mm-liu-h.xlsx](https://github.com/mirror52/test1/blob/main/data_1.2mm-1.8mm-liu-h.xlsx) (accessed on 5 April 2025).

**Acknowledgments:** The authors want to thank the editor and anonymous reviewers for their valuable suggestions for improving this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Xin, F.; Ma, T.; Chen, Y.T.; Wang, Q. Study on chemical spray etching of stainless steel for printed circuit heat exchanger channels. *Nucl. Eng. Des.* **2019**, *341*, 91–99. [CrossRef]
2. Chai, L.; Tassou, S.A. A review of printed circuit heat exchangers for helium and supercritical CO<sub>2</sub> Brayton cycles. *Therm. Sci. Eng. Prog.* **2020**, *18*, 100543. [CrossRef]
3. Sun, F.; Xie, G.N.; Li, S.L. An artificial-neural-network based prediction of heat transfer behaviors for in-tube supercritical CO<sub>2</sub> flow. *Appl. Soft Comput.* **2021**, *102*, 107110. [CrossRef]
4. Wen, Z.X.; Wu, J.L.; Wang, S.S.; Cheng, J.Q.; Li, Q. Numerical study and machine learning on local flow and heat transfer characteristics of supercritical carbon dioxide mixtures in a sinusoidal wavy channel PCHE. *Int. J. Heat Mass Transf.* **2024**, *223*, 125278. [CrossRef]

5. Li, H.Z.; Zhang, Y.F.; Zhang, L.X.; Yao, M.; Kruizenga, A.; Anderson, M. PDF-Based modeling on the turbulent convection heat transfer of supercritical CO<sub>2</sub> in the printed circuit heat exchangers for the supercritical CO<sub>2</sub> Brayton cycle. *Int. J. Heat Mass Transf.* **2016**, *98*, 204–218. [CrossRef]
6. Meshram, A.; Jaiswal, A.K.; Khivsara, S.D.; Ortega, J.D.; Ho, C.; Bapat, R.; Dutta, P. Modeling and analysis of a printed circuit heat exchanger for supercritical CO<sub>2</sub> power cycle applications. *Appl. Therm. Eng.* **2016**, *109*, 861–870. [CrossRef]
7. Pesteei, S.M.; Mehrabi, M. Modeling of convection heat transfer of supercritical carbon dioxide in a vertical tube at low Reynolds numbers using artificial neural network. *Int. Commun. Heat Mass Transf.* **2010**, *37*, 901–906. [CrossRef]
8. Yin, D.D.; Zhou, Y.L.; Guo, X.T.; Wang, D. Numerical analysis of wavy PCHes in supercritical CO<sub>2</sub>/propane mixture Brayton cycle. *Appl. Therm. Eng.* **2023**, *235*, 121346. [CrossRef]
9. Longo, G.A.; Mancin, S.; Righetti, G.; Zilio, C.; Ceccato, R.; Salmaso, L. Machine learning approach for predicting refrigerant two-phase pressure drop inside Brazed Plate Heat Exchangers (BPHE). *Int. J. Heat Mass Transf.* **2020**, *163*, 120450. [CrossRef]
10. Tian, Z.; Gu, B.; Yang, L.; Liu, F. Performance prediction for a parallel flow condenser based on artificial neural network. *Appl. Therm. Eng.* **2014**, *63*, 459–467. [CrossRef]
11. Wu, H.W.; Bagherzadeh, S.A.; D’Orazio, A.; Habibollahi, N.; Karimipour, A.; Goodarzi, M.; Bach, Q.V. Present a new multi objective optimization statistical Pareto frontier method composed of artificial neural network and multi objective genetic algorithm to improve the pipe flow hydrodynamic and thermal properties such as pressure drop and heat transfer coefficient for non-Newtonian binary fluids. *Phys. A Stat. Mech. Its Appl.* **2019**, *535*, 122409.
12. Li, Q.; Zhan, Q.; Yu, S.; Sun, J.; Cai, W. Study on thermal-hydraulic performance of printed circuit heat exchangers with supercritical methane based on machine learning methods. *Energy* **2023**, *282*, 128711. [CrossRef]
13. Zhang, H.; Guo, J.; Huai, X.; Cheng, K.; Cui, X. Studies on the thermal-hydraulic performance of zigzag channel with supercritical pressure CO<sub>2</sub>. *J. Supercrit. Fluids* **2019**, *148*, 104–115. [CrossRef]
14. Lee, S.Y.; Park, B.G.; Chung, J.T. Numerical studies on thermal hydraulic performance of zigzag-type printed circuit heat exchanger with inserted straight channels. *Appl. Therm. Eng.* **2017**, *123*, 1434–1443. [CrossRef]
15. Han, Z.X.; Guo, J.F.; Huai, X.L. Theoretical analysis of a novel PCHE with enhanced rib structures for high-power supercritical CO<sub>2</sub> Brayton cycle system based on solar energy. *Energy* **2023**, *270*, 126928. [CrossRef]
16. Kim, W.; Baik, Y.; Jeon, S.; Jeon, D.; Byon, C. A mathematical correlation for predicting the thermal performance of cross, parallel, and counterflow PCHes. *Int. J. Heat Mass Transf.* **2017**, *106*, 1294–1302. [CrossRef]
17. Animah, I.; Shafiee, M. Application of risk analysis in the liquefied natural gas (LNG) sector: An overview. *J. Loss Prev. Process Ind.* **2020**, *63*, 103980. [CrossRef]
18. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J. Pearson correlation coefficient. *Noise Reduct. Speech Process.* **2009**, *2*, 1–4.
19. Pei, S.; Wang, J.; Tian, C.; Li, X.; Guo, B.; Guo, J.; Yao, Y. Assist-as-Needed Controller of a Rehabilitation Exoskeleton for Upper-Limb Natural Movements. *Appl. Sci.* **2025**, *15*, 2644. [CrossRef]
20. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
21. Chen, T.; Guestrin, C. XGBOOST: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
22. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CATBOOST: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. Available online: <https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html> (accessed on 11 March 2025).
23. Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Support vector regression. *Effic. Learn. Mach. Theor. Concepts Appl. Eng. Syst. Des.* **2015**, 67–80. Available online: [https://link.springer.com/chapter/10.1007/978-1-4302-5990-9\\_4](https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4) (accessed on 11 March 2025).
24. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
25. Medsker, L.R.; Jain, L. Recurrent neural networks. *Des. Appl.* **2001**, *5*, 64–67.
26. Hochreiter, S. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
27. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; pp. 1597–1600.
28. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]
29. Lim, B.; Zohren, S. Time-Series forecasting with deep learning: A survey. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200209. [CrossRef]
30. Li, J.; Lin, Y.; Gui, Z.; Wang, P. Inception–Attention–BiLSTM Hybrid Network: A Novel Approach for Shear Wave Velocity Prediction Utilizing Well Logging. *Appl. Sci.* **2025**, *15*, 2345. [CrossRef]
31. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 7354–7363.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.

33. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [[CrossRef](#)]
34. Huang, Z.; Bai, Y.; Liu, H.; Lin, Y. Multiscale Feature Modeling and Interpretability Analysis of the SHAP Method for Predicting the Lifespan of Landslide Dams. *Appl. Sci.* **2025**, *15*, 2305. [[CrossRef](#)]
35. Jiang, T.; Li, M.J.; Yang, J.Q. Research on optimization of structural parameters for airfoil fin PCHE based on machine learning. *Appl. Therm. Eng.* **2023**, *229*, 120498. [[CrossRef](#)]
36. Yang, Y.; Li, H.; Xie, B.; Zhang, L.; Zhang, Y. Experimental study of the flow and heat transfer performance of a PCHE with rhombic fin channels. *Energy Convers. Manag.* **2022**, *254*, 115137. [[CrossRef](#)]
37. Wen, J.; Ma, H.; Xu, G.; Dong, B.; Liu, Z.; Zhuang, L. Experimental and numerical study on the thermal and hydraulic characteristics of an improved PCHE with high-viscosity fluid. *Appl. Therm. Eng.* **2025**, *271*, 126345. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.