scientific reports

Check for updates

OPEN

Prediction of heat transfer and fluid flow performance of fly ash–Cu hybrid nanofluid using machine learning approaches

V. Vamsi Krishna[1], Ved Prakash Mishra[2], A. Jayanthiladevi[3], Thanapal Senthilkumar[4], N. Pratheeksha Hegde[5]✉ & Haiter Lenin Allasi[6]✉

Accurate prediction of thermal–hydraulic characteristics is essential for the efficient design and optimization of nanofluid (NF)-based heat transfer systems. In this study, Support Vector Regression (SVR) and Gradient Boosting (GB) models were developed to simultaneously predict the heat transfer coefficient (HTC), Nusselt number (Nu), pressure drop (ΔP), and friction factor (f) of fly ash–Cu hybrid nanofluids under turbulent forced-convection conditions. A total of 75 data samples derived from the experimental study of Kanti et al. served as the basis for training and evaluating the predictive models. This data set covering Reynolds number (Re) from 6,807 to 45,155, nanofluid concentrations from 0 to 2.0 vol.%, and temperatures of 30, 45, and 60 °C was utilized for model development and validation. Both machine-learning approaches exhibited excellent predictive capability. SVR achieved regression coefficient (R^2) values of 0.9931, 0.9929, 0.9913, and 0.9727 for HTC, Nu, ΔP, and f, respectively, with an average mean absolute percentage error (MAPE) of 3.13%. Correspondingly, the GB model produced R^2 values of 0.9934, 0.9926, 0.9660, and 0.9826, with an average MAPE of 3.62%. Five-fold cross-validation yielded mean R^2 values of 0.9590 for SVR and 0.9691 for GB, confirming strong model robustness. Feature importance analysis revealed that Reynolds number contributed 76.7% to the predictive performance, followed by nanoparticle concentration (18.9%) and temperature (4.4%). The results demonstrate that both models can accurately predict the thermal–hydraulic behavior of fly ash–Cu hybrid nanofluids, providing efficient alternatives to extensive experimental testing. These models offer a computationally efficient design-support tool for the investigated operating range; however, given the modest 75-sample dataset restricted to a single fly ash–Cu formulation, they should be regarded as interpolative within the investigated range rather than broadly generalizable.

Keywords Fly ash–Cu hybrid nanofluid, Turbulent flow, Thermal–hydraulic performance, Machine learning, Support vector regression, Gradient boosting

List of symbols

c	Hybrid nanofluid volumetric concentration (vol%)
f	Darcy–Weisbach friction factor
h/HTC	Convective heat transfer coefficient ($W \cdot m^{-2} \cdot K^{-1}$)
D	Pipe inner diameter (m)
k	Thermal conductivity of fluid ($W \cdot m^{-1} \cdot K^{-1}$)
Nu	Nusselt number (= h·D/k)
n	Number of observations
ΔP	Pressure drop across test section (Pa)
Pr	Prandtl number
R^2	Coefficient of determination

[1]Post-Doctoral Research Affiliate, Amity University, Dubai, United Arab Emirates. [2]Department of Computer Science and Engineering, Amity University, Dubai, United Arab Emirates. [3]Department of Computer Science and Engineering, BGSIT, Adichunchangiri University, Belluru, Karantaka, India. [4]AI Architect, HCLTech, Dallas, TX, USA. [5]Nitte (Deemed to be University), NMAM Institute of Technology (NMAMIT), Department of Information Science & Engineering, Nitte, India. [6]School of Engineering and Applied Sciences, Kampala International University, Box 20000, Kampala, Uganda. ✉email: pratheeksha.hegde@nitte.edu.in; haiterlenina@gmail.com

Re	Reynolds number (= ρ·u·D/μ)
RMSE	Root mean squared error (output-dep.)
T	Bulk fluid temperature (°C)
u	Mean flow velocity ($m \cdot s^{-1}$)
μ	Dynamic viscosity (mPa·s)
ρ	Fluid density ($kg \cdot m^{-3}$)

Abbreviations

CV	Cross-validation
GB	Gradient boosting regressor
MAE	Mean absolute error
MAPE	Mean absolute percentage error (%)
MSE	Mean squared error
MDI	Mean decrease in impurity (feature importance)
OAT	One-At-a-Time sensitivity analysis
PINN	Physics-informed neural network
RBF	Radial basis function
RMSE	Root mean squared error
SVR	Support vector regression
SVR-RBF	Support vector regression with radial basis function kernel

Improving heat transfer is considered to be one of the crucial factors for the efficiency of thermal systems such as solar collectors, heat exchangers, electronic coolers, and automobile radiators[1,2]. Conventional fluids such as oils, ethylene glycol (EG), and water have relatively low thermal conductivity[3]. This limitation has motivated the development of advanced heat-transfer fluids[4], including enhanced designs for solar and thermal-electric applications[5]. Nanofluids consist of a base fluid with uniformly distributed nanoparticles suspended in it, typically under 100 nm[6]. Such particle dispersion significantly improves thermal performance and convective HTC[7]. Nanotechnology's impact extends across diverse sectors, from food science[8] to energy systems, biomedical devices, and cooling technologies, while also enhancing thermal conductivity of the base fluid[9].

Hybrid nanofluids are composed of more than two types of nanoparticles dispersed in a base fluid[10]. These often show superior thermophysical properties versus mono nanofluids[11]. Kanti et al.[12] reported a 44% thermal-conductivity enhancement for graphene oxide–TiO_2 hybrid nanofluid compared to water. Vidhya et al.[13] confirmed the suitability of binary-base-fluid hybrid nanofluids for thermal-management applications, and Rafid et al.[14] found hybrid nanofluids offer superior thermal performance and transport properties over single-component nanofluids. Ramadhan et al.[15] found friction-factor deviation peaking at 3.5% for TiO_2–SiO_2/EG–water hybrid nanofluid at 1–3 vol.%. Twisted-tape inserts (twist ratio 5) at Re = 14,000 increased Nu and HTC by 36.28% and 36.3% for Al_2O_3–CuO/water hybrid nanofluid[16]. Sundar et al.[17] found hybrid nanofluids transfer heat better than mono nanofluids and base fluids; the reasons behind this improvement remain poorly understood, and further research is needed. A 0.5 vol.% MWCNT–Al_2O_3/water hybrid nanofluid produced a maximum HTC of 357.4 W/m^2K and 83.9% thermal efficiency at 1065 W/m^2 solar irradiation[18]. Al_2O_3–Fe_2O_3 hybrid nanofluid viscosity peaked at a 40:60 mixing ratio and fell by up to 87.2% with rising temperature[19]. A 1 vol.% Al_2O_3–TiO_2/water hybrid nanofluid viscosity increased by 15.8% and Nu by 70.4% versus water[20], while Al_2O_3–CuO/water hybrid nanofluid under turbulent flow improved heat transfer and lowered entropy generation by 71%[21]. Reviews on hybrid nanofluids in machining field, reports that improved heat transfer, surface finish, and tool durability[22], alongside common preparation and stabilization methods used in industrial applications[23]. Recent advances in machine learning (ML) enable accurate prediction and optimization of complex thermophysical properties without relying solely on costly experiments or computationally intensive analytical models[24]. SVR is popular for its robustness on small-to-medium nonlinear datasets[25], while GBR is an ensemble method that iteratively fits residual errors to improve accuracy[26]; both accurately predict nanofluid thermal conductivity, viscosity, and heat-transfer characteristics[27].

Both algorithms suit the nonlinear dependencies between inputs (temperature, concentration, particle loading) and outputs (thermal conductivity, heat transfer) common to these systems: SVR offers strong generalization via structural risk minimization on limited samples[25], while GBR handles noisy, complex datasets well through iterative residual correction[28]. Both outperform traditional regression and basic ML on predictive accuracy[29]. Representative applications include: Vicki et al.[19], artificial neural network (ANN) achieving > 99.99% correlation accuracy (mean squared error = 0.000036); Kanti et al.[20], XGBoost achieving $R^2 = 0.991$ (entropy generation) and 0.975 (Nu); Mathur et al.[24], soft-computing for specific heat capacity; Sajjad et al.[30], deep learning for boiling HTC on porous surfaces; Albdour et al.[31], ML for condensation heat transfer; Alade et al.[32], SVR for specific heat capacity, later extended to thermal conductivity prediction[33]; and Sharma et al.[34], a comprehensive review of ML for sustainable energy solutions.

Although tremendous strides have been made in the study of nanofluids, there are still some drawbacks. Previous research has largely relied on experimental correlations, generally applicable only to given operating conditions, or limited to general prediction. Moreover, the use of ML and data-driven predictive approaches to study the behavior of Fly Ash–Cu hybrid nanofluids has not been broadly studied. There are also not many comparative studies conducted among various ML models to predict the thermophysical properties and heat-transfer under different temperature and concentration conditions. Kanti et al.[1] in their work, researched the heat and flow behaviour of a low-cost hybrid nanofluid (Fly Ash–Cu) in turbulent flow regime ranging from 6800 to 45,200 Reynolds numbers under a heat flux of 7962 W/m^2. This study conducted by them provided detailed experimental data of heat-transfer and friction-factor characteristics, demonstrating the Fly Ash–Cu hybrid nanofluids' potential for thermal management applications. Following this, the current study extends

the analysis by incorporating ML techniques on the experimental data to predict the behaviours corresponding to the heat-transfer and friction-factor. This article presents the implementation of ML models using SVR and GBR for the prediction of thermophysical and heat transfer properties of the Fly Ash–Cu hybrid nanofluids. To identify the most accurate prediction approach, the study systematically evaluates the predictive performance of several ML models. By using data-driven modeling, the dependence on extensive experimental testing can be reduced, which helps in saving time and cost. In addition, the use of low-cost fly ash with highly conductive copper nanoparticles contributes toward the design of sustainable and energy-efficient thermal systems. The novelty of present work, distinct from prior single-output/single-algorithm nanofluid-ML studies (e.g.,[12,20,24]), lies in: (i) simultaneous four-output prediction (HTC, Nu, ΔP, f) not previously modelled with ML; (ii) a head-to-head SVR-vs-GB comparison showing neither dominates; and (iii) benchmarking against classical correlations (Dittus–Boelter, Petukhov, Blasius).

Validation of experimental setup

All data presented in this section were obtained by experimental work of Kanti et al.[1]; this section is included only to establish the physical consistency and validity of the underlying dataset used for the ML models developed in Section "Machine learning methodology" and does not represent new experimental work performed by the present authors. Prior to testing the fly ash–Cu hybrid nanofluid, the experimental setup was verified using distilled water[1]. Verification of the experimentally obtained Nu and friction factor (f) was carried out against the Dittus–Boelter correlation[35], Petukhov[36], and Blasius correlations[37] over the turbulent range[1]. The maximum deviations were within ± 6.2% for Nu and ± 4.8% for f, which falls within the ± 8% accuracy range commonly reported for nanofluid forced-convection studies[38], proving the consistency of the experimental facility. The Nu was plotted against the Reynolds number for distilled water at 30 °C, as shown in Fig. 1a, and compared to the Dittus–Boelter correlation (Eq. 1). The experimental results demonstrate the anticipated behaviour of Nu with Reynolds number as expected, and a good correlation with the theoretical result is observed. The experimentally determined friction factor is compared with Petukhov (Eq. 2) and Blasius (Eq. 3) correlations in Fig. 1b. Pressure-drop measurements and flow characterization are validated by the measured values that closely follow both correlations over the investigated range.

Dittus –Boelter equation[35]

$$Nu = 0.023\mathrm{Re}^{0.8}\,\mathrm{Pr}^{0.4} \tag{1}$$

Petukhov equation[36]

$$\mathrm{f} = (0.79\,\ln(\mathrm{Re}) - 1.64)^{-2} \tag{2}$$

Blasius equation[37]

$$\mathrm{f} = \frac{0.316}{Re^{0.25}} \tag{3}$$

Machine learning methodology

Dataset description and exploratory data analysis

The dataset was derived from controlled experimental measurements on fly ash–Cu hybrid nanofluid flowing through a circular pipe under turbulent flow conditions. There were 75 experimental observations obtained in the study by systematically varying the three independent parameters: concentration (0.0, 0.5, 1.0, 1.5,

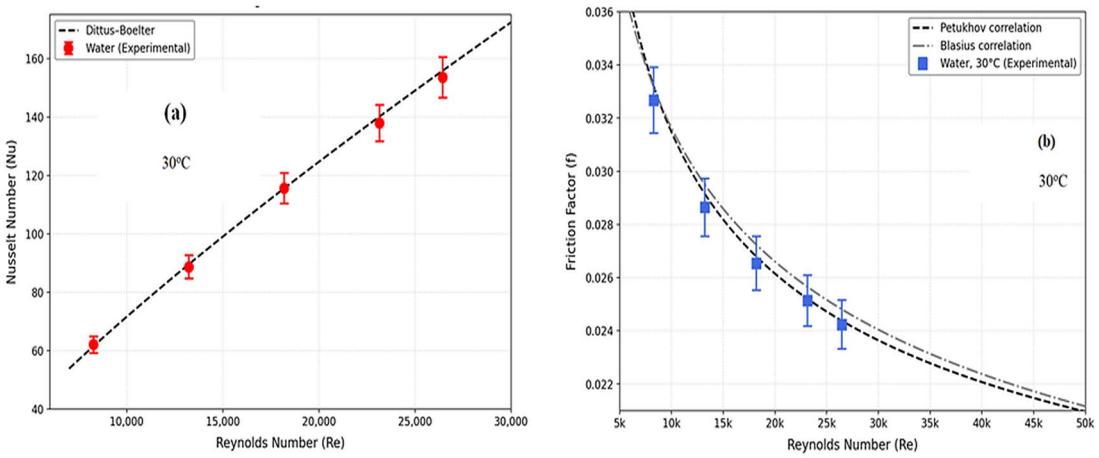


Fig. 1. Validation of experimental (**a**) Nu and (**b**) friction factor data for distilled water against the standard correlations

Variable	Min	Max	Mean	Std Dev	Units
Concentration	0.0000	2.0000	1.0000	0.7119	vol. %
Temperature	30.00	60.00	45.00	12.33	°C
Reynolds number	6,807	45,155	21,035	9,372	—
HTC	2,399.9	15,923.6	7,894.8	3,102.1	$W{\cdot}m^{-2}{\cdot}K^{-1}$
Nusselt number	62.034	306.960	165.234	56.626	—
Pressure drop	228.0	2,240.0	1,111.8	649.4	Pa
Friction factor	0.02165	0.03642	0.02740	0.00348	—

Table 1. Descriptive statistics of the experimental dataset.

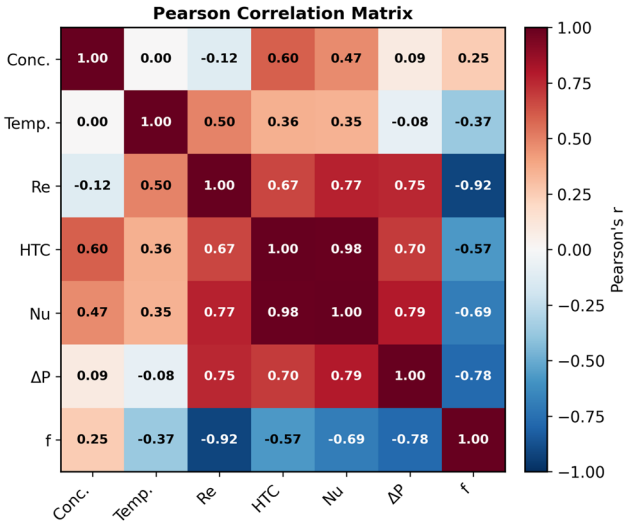


Fig. 2. Pearson correlation matrix of the experimental variables

and 2.0 vol%), bulk fluid temperature (30, 45, and 60 °C), and Reynolds number (Re ≈ 6,807 to 45,155). The four thermal–hydraulic response variables used in the recordings were: HTC ($W{\cdot}m^{-2}$ K^{-1}), Nusselt number (Nu), pressure drop (ΔP, Pa), and friction factor (f)[24]. The 75 observations form a complete factorial design (5 concentrations × 3 temperatures × 5 Reynolds settings) covering the full envelope of Kanti et al.[1], not a random sample. Representativeness means full coverage of this envelope; extrapolation beyond these bounds, or to other systems/geometries, is outside the models' validated domain.

Descriptive statistics

The range, central tendency and dispersion of each experimental variable are summarized in Table 1. The Reynolds number range is 6807 to 45,155; the distribution is slightly positively skewed (skewness = + 0.46) and the mean Re (21,035) is greater than the median (19,437). The HTC ranges from 2,399.9 to 15,923.6 $W{\cdot}m^{-2}{\cdot}K^{-1}$; friction factor from 0.02165 to 0.03642; and pressure drop from 228 to 2,240 Pa. It can be seen that the large standard deviations both of HTC and ΔP compared to their means show high sensitivity to the operating parameters[39].

It can be seen from the Pearson correlation matrix (Fig. 2) that the Reynolds number has the highest correlation with HTC (r = 0.67) and Nu (r = 0.77). The HTC (r = 0.60) is a significant positive correlation for concentration, confirming the improvement of the Cu-bearing hybrid particle phase thermal conductivity. The correlation coefficient between the friction factor and the Reynolds number is very negative (r = − 0.92); it is in good agreement with the predictions of Petukhov and Blasius.

The ML pipeline follows a structured, reproducible workflow encompassing data preprocessing, model development, training, evaluation, and comparative analysis. The overall methodology is summarized in a schematic fashion in Fig. 3 with a sequence from the raw experimental data to z-score standardization, partitioning the data into train and test sets, training parallel models (SVR-RBF and GB), evaluation against the held-out test data, and comparative analysis. The pipeline was implemented in Python 3.11 using scikit-learn 1.3.2, with NumPy 1.26.2[40] and pandas 2.1.3[41] for numerical and tabular data handling, consistent with similar Python-based ML pipelines developed for nanofluid thermo-hydraulic prediction[42].

Data pre-processing and dataset partitioning

All input features (concentration, temperature, Reynolds number) were subjected to Z-score standardization (StandardScaler), transforming each variable to unit variance and zero mean. Z-score standardization is

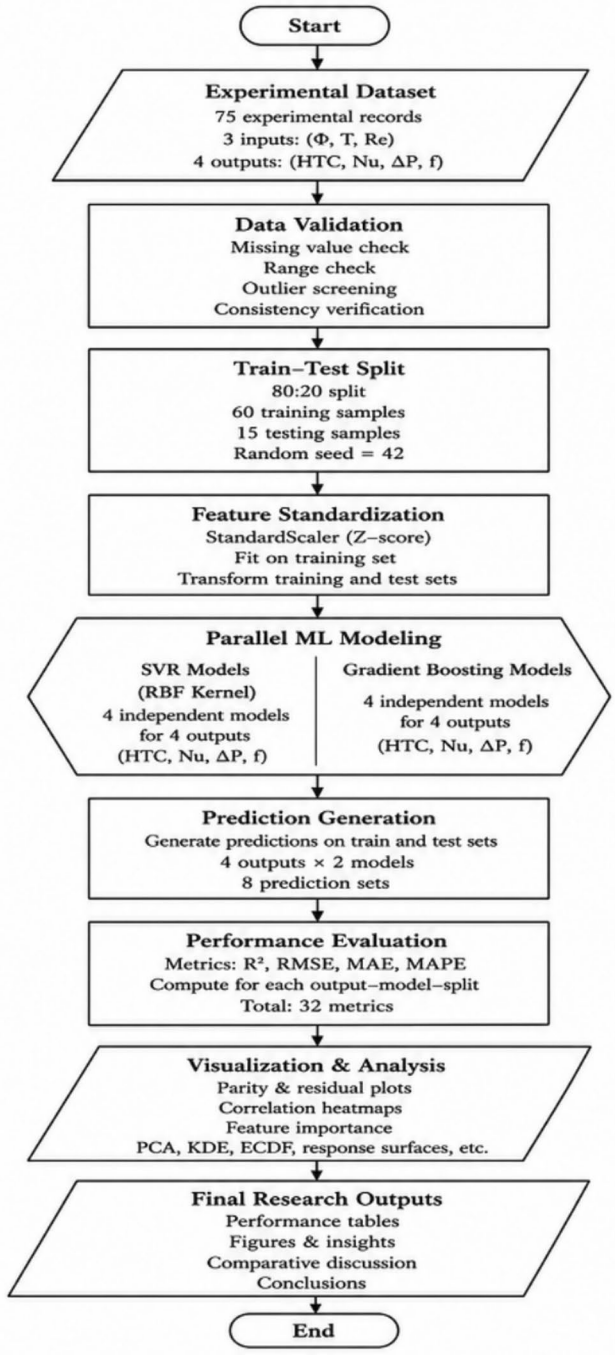


Fig. 3. Workflow of the machine learning model for predicting thermal–hydraulic properties of fly ash–Cu hybrid nanofluid

particularly well-suited to SVR with Radial Basis Function (RBF) kernel—which is sensitive to feature scale—and benefits GB by ensuring consistent gradient behaviour across all three inputs. To ensure a completely data-leakage-free assessment pipeline, all scalers were fitted solely on the training subset and applied without refitting to the held-out test set. Targets were not scaled, as neither kernel nor tree-based methods require target normalization for regression.

To ensure complete repeatability, our adopted dataset consisting of 75-samples was divided into 80% training (n = 60) and 20% testing (n = 15) subsets via random sampling, with a fixed random seed of 42[23]. The 15-sample test set provides limited statistical representativeness, partially mitigated by the five-fold cross-validation in Section "Cross-validation stability analysis", wherein every observation serves as a validation sample exactly once. Overfitting risk on this small dataset is mitigated through: (i) SVR's structural risk minimization principle, which limits model complexity via the regularization parameter C; (ii) GB's subsample fraction (0.8) and max_depth constraints, which act as stochastic regularization analogous to dropout; and (iii) five-fold cross-validation,

Partition	Samples (n)	Fraction (%)	Purpose
Training	60	80	Model Fitting and Hyper parameter Optimization
Testing	15	20	Out-of-sample performance evaluation
Total	75	100	Full experimental dataset

Table 2. Dataset partitioning scheme.

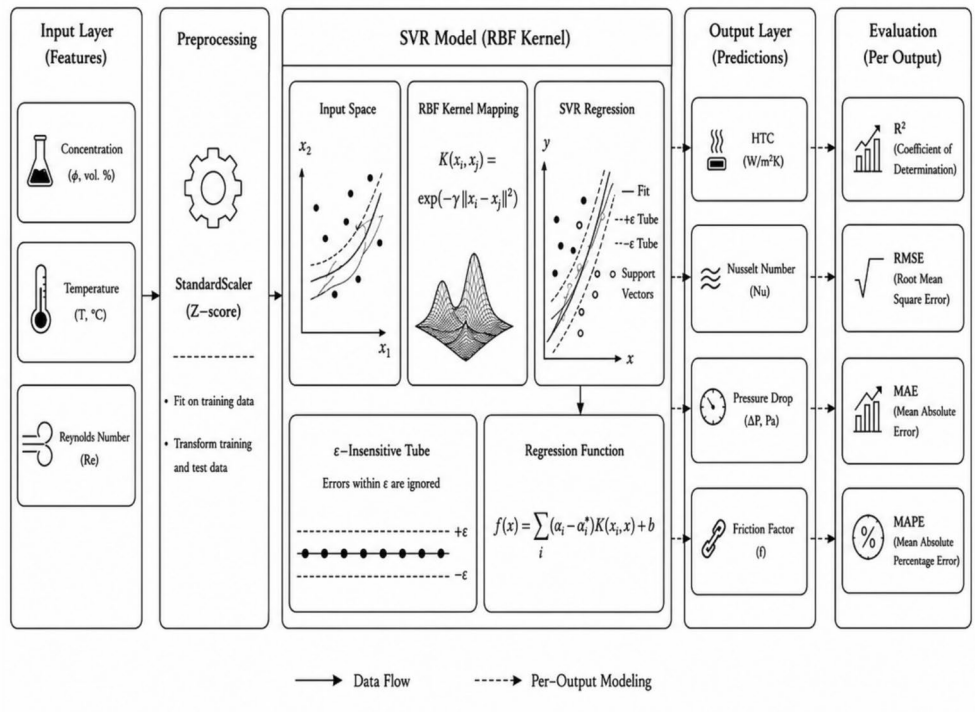


Fig. 4. Schematic representation of the SVR model with RBF kernel showing the regression architecture and prediction mechanism

which offers a consistent estimate of generalization performance beyond a single hold-out split. With the fixed seed, the row corresponding to 0 vol%, 30 °C, Reynolds number = 8,264 (HTC = 2,399.9; Nu = 62.034) is placed in the test set; both values lie below the training-subset minimum bounds (HTC min = 2,736.0; Nu min = 68.6), constituting mild extrapolation at the lower boundary. No information leakage occurred during hyperparameter tuning: grid search for both SVR and GB was performed using cross-validation restricted to the 60-sample training subset only, and the 15-sample held-out test set was accessed exactly once, after the final configuration was fixed, to compute the reported test-set metrics. Table 2 illustrates the dataset partitioning scheme.

Support vector regression (SVR) model

SVR adapts the Support Vector Machine framework to handle continuous regression problems. Rather than seeking a hyperplane that maximizes class separation, SVR works by finding a function that keeps prediction errors within a defined tolerance ε, while keeping the model as simple as possible through the Structural Risk Minimization principle[43]. SVR is particularly attractive for small experimental datasets because its sparse kernel representation depends only on a subset of training points (the support vectors), inherently limiting overfitting[27]. The RBF kernel is employed here owing to its demonstrated effectiveness for nonlinear thermal–hydraulic regression problems[42].

Kernel architecture and training

The SVR model architecture is illustrated in Fig. 4. The RBF kernel maps each standardized input vector into features of high dimension through Eq. (4):

$$K\left(x_i, x_j\right) = \exp\left(\gamma \left\| x_i - x_j \right\|^2\right) \tag{4}$$

where, parameter γ denotes the kernel width controlling the spread of the kernel.

The regression function is given by

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{5}$$

where α_i and α_i^* refers to dual Lagrange multipliers and parameter b denotes the bias term. Four independent SVR models were trained—one per output (HTC, Nu, ΔP, f). Hyperparameter values (C, γ, ε) were selected via grid search over a predefined candidate set, retaining the configuration yielding minimum cross-validated root mean squared error (RMSE) on the training set. Table 3 presents the optimal configuration for each output.

Evaluation metrics

Predictive performance was computed via four standard statistical evaluation metrics. The R^2 representing the coefficient of determination is defined as[12]:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{6}$$

The root mean square error (RMSE) can be obtained from[12]:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \tag{7}$$

The mean absolute error (MAE) is expressed as[12]:

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \tag{8}$$

The mean absolute percentage error (MAPE) is defined as[12]:

$$MAPE = \frac{100}{n} \sum \frac{|y_i - \hat{y}_i|}{|\hat{y}_i|} \tag{9}$$

Parameters $\boldsymbol{y_i}$ and $\hat{\boldsymbol{y}}$ represent the measured and predicted results, respectively, while the mean of the experimental values is represented by $\bar{\boldsymbol{y}}$ and the number of observations is represented by the parameter n.

Predictive performance

On the held-out test set, the SVR model achieves R^2 values across all four outputs as follows. Pressure drop, HTC and Nu exhibits $R^2 = 0.9913$, 0.9931, and 0.9929, respectively. The lowest R^2 is 0.9727, for friction factor, which has a relatively narrow dynamic range. All outputs have MAPE values below 4.5%. The Dittus–Boelter correlation is applied to the same 15-point test set, and gives $R^2 = -0.070$ and MAPE = 28.19% for Nu, effectively proving that the data-driven models with concentration effects have a definite advantage. For friction factor, the Blasius correlation gave the $R^2 = 0.9752$ and the MAPE = 1.75%. Parity plots are shown for all four outputs in Fig. 5a–d, and actual versus predicted are plotted for them.

Residual analysis

The HTC residuals are plotted against the SVR-predicted HTC values in Fig. 6. HTC is chosen as the representative output due to being the widest range in absolute terms. There is no apparent monotonic drift or heteroscedasticity along the zero axis in the residuals. Qualitatively similar results were obtained from analogous diagnostics of Nu, ΔP, and f. The residual histogram (Fig. 7) is close to a normal distribution centered on zero. The amount of positive skew is slight, around + 0.001, which indicates that the model may slightly under-predict the number of HTC at high values of the training data set.

Gradient boosting regressor (GB) model

The GB is an ensemble learning approach that iteratively fits shallow regression trees to pseudo-residuals of the fitted current ensemble to create a good predictive model[44]. Unlike SVR, GB iteratively constructs the model to reveal highly non-linear, non-stationary relationships without needing to specify a kernel[45], and also has a built-in feature importance mechanism that gives direct insight into the contributions of the predictors[46]. Four independent GB models were trained — one per thermal–hydraulic output.

Output	Regularisation	Length-scale	Insensitivity	Kernel	Rationale
HTC	High	Auto	Narrow	RBF	High C fits wide HTC range; narrow ε for precision
Nusselt Number	Moderate	Auto	Auto	RBF	Moderate C; scale-based γ for smooth surface
Pressure Drop	High	Auto	Narrow	RBF	High C captures steep Re–ΔP power-law; narrow ε
Friction Factor	Low	Auto	Narrow	RBF	Low C: narrow dynamic range → prefer smooth model

Table 3. SVR Hyper parameter configuration by output variable.

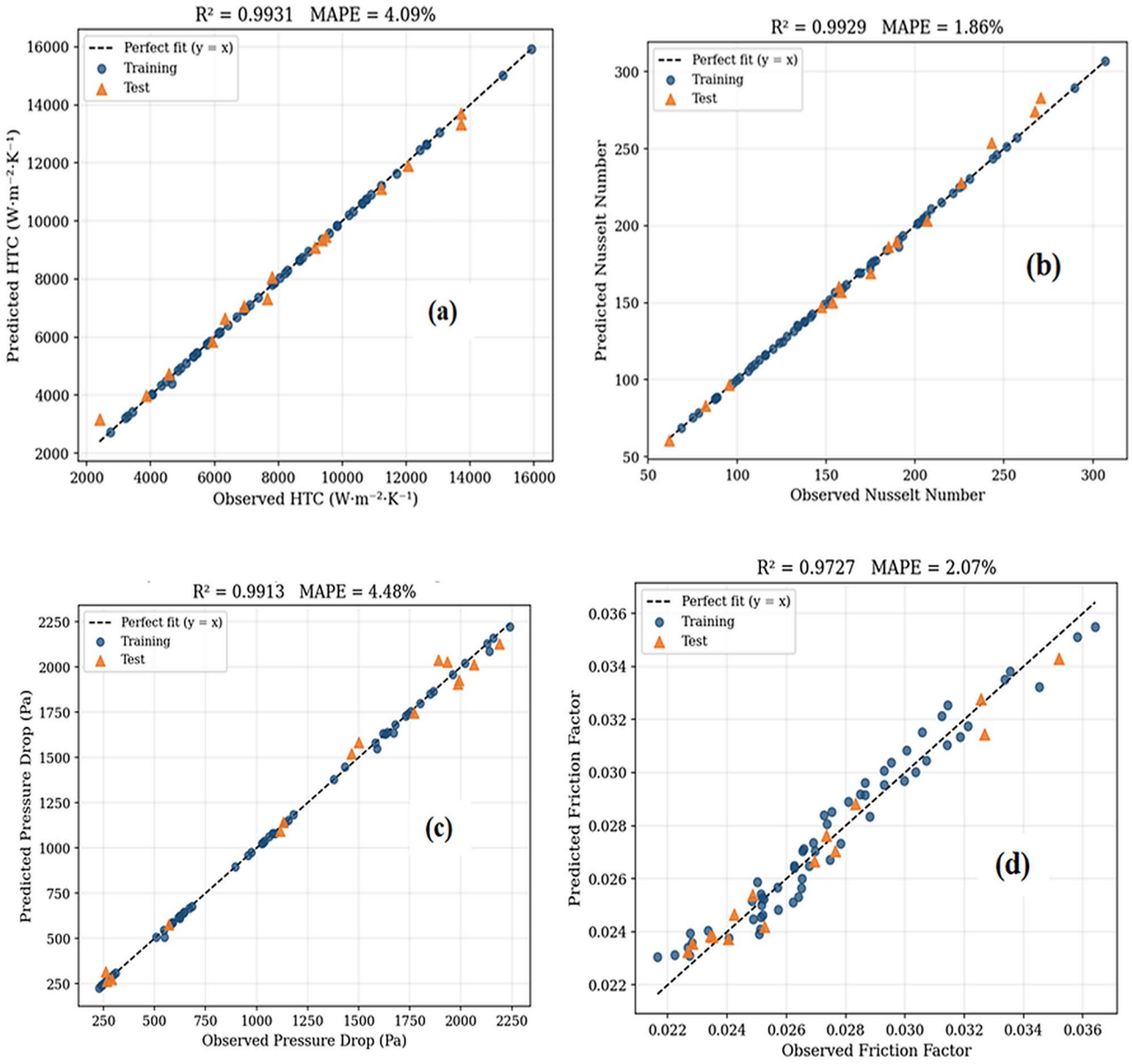


Fig. 5. (**a**–**d**) SVR parity plots for (**a**) heat transfer coefficient, (**b**) Nusselt Number, (**c**) pressure drop, and (**d**) friction factor.

Algorithm architecture and mathematical formulation

The Gradient Boosting (GB) architecture is illustrated in Fig. 8. The model begins with an initial estimator $\mathbf{F_0(x)}$, corresponding to the mean response for squared-error loss. At each boosting iteration **m**, pseudo-residuals are computed as:

$$r_{im} = \left[\frac{\partial L\left(y_i, F(x_i)\right)}{\partial F(x_i)}\right]_{F=F_{m-1}} \tag{10}$$

A regression tree $\mathbf{h_m(x)}$ is trained on the pseudo-residuals, and the ensemble is then updated as according to Eq. (11):

$$F_m(x) = F_{m-1}(x) + \nu h_m(x) \tag{11}$$

where the learning rate is denoted by ν. Finally, after M boosting iterations, the prediction model is written as:

$$F_m(x) = F_0(x) + \nu \sum_{m=1}^{M} h_m(x) \tag{12}$$

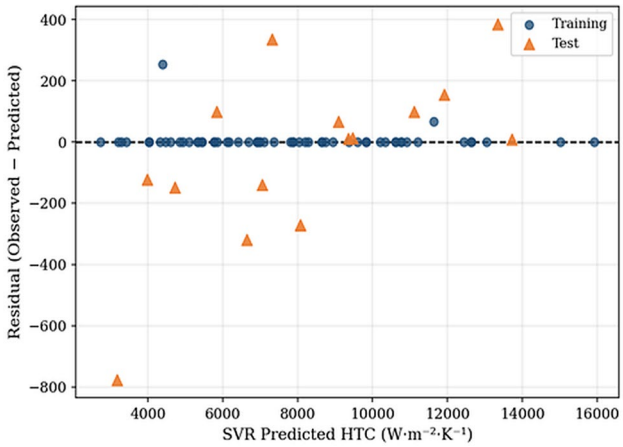


Fig. 6. Residual distribution (observed – predicted) versus predicted heat transfer coefficient for the SVR Model

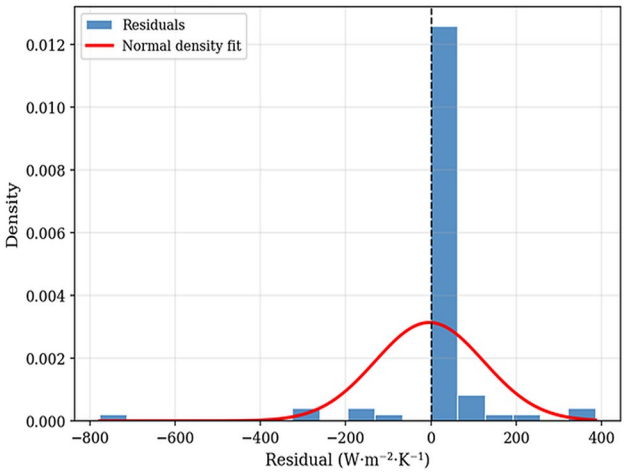


Fig. 7. Distribution of SVR residuals for heat transfer coefficient with superimposed normal density fit.

Subsampling the training observations in each iteration acts as a regularization technique which helps prevent overfitting and enhances the generalization of the model.

Hyperparameter configuration
Key hyperparameters governing GB performance are: (i) n_estimators M (model capacity); (ii) learning rate ν (shrinkage); (iii) max_depth (tree complexity); and (iv) subsample fraction (stochastic regularization). Hyperparameter values were selected via grid search restricted to the 60-sample training subset, using the same cross-validation procedure as for SVR (see Section "Kernel architecture and training"). Table 4 presents the optimal configuration for each output.

Predictive performance of the GB model
The GB model achieves R^2 values in the range 0.9660 to 0.9934 across all four outputs. GB achieves the highest R^2 for HTC (0.9934) and friction factor (0.9826), while SVR achieves higher R^2 for Nu (0.9929 vs. 0.9926) and pressure drop (0.9913 vs. 0.9660). MAPE ranges from 1.43% (friction factor) to 6.10% (pressure drop). Figure 9a–d present actual-versus-predicted scatter diagrams for all four GB outputs.

Comparative analysis of models

A comparative analysis assessed the relative predictive performance, stability, and practical trade-offs of the SVR and GB models across all four thermal–hydraulic outputs.

Model performance comparison
A detailed side-by-side comparison of R^2, RMSE, MAE, and MAPE on the held-out test set is shown in Table 5. There is no universally superior model in the four outputs; SVR is superior for Nu and ΔP; GB is superior

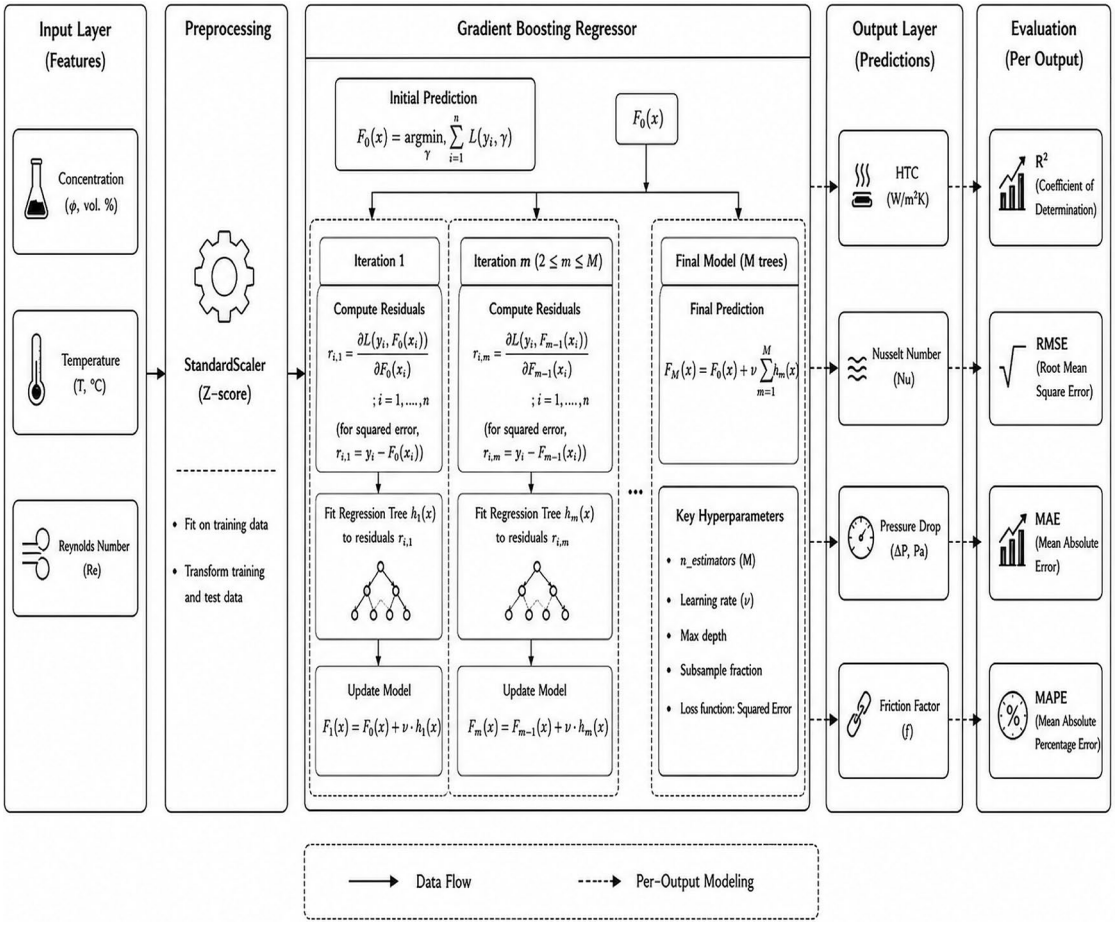


Fig. 8. Schematic representation of the Gradient Boosting Regressor architecture and iterative learning workflow

Output	n_est	Learn. rate	Max depth	Subsample	Rationale
HTC	High	Low	Deeper	Partial	Deeper tree captures concentration–Re interaction in HTC
Nusselt Number	High	Moderate	Standard	Partial	Standard depth; higher lr for well-structured Nu surface
Pressure Drop	High	Moderate	Standard	Partial	High n_est. for accurate ΔP across wide Re range
Friction Factor	Moderate	Low	Standard	Partial	Few trees sufficient: smooth monotonic f–Re relationship

Table 4. GB hyper parameter configuration by output variable.

for HTC and friction factor. The most striking difference is for pressure drop where SVR obtained $R^2 = 0.9913$ compared to GB's 0.9660, with the SVR RMSE (64.16 Pa) being about half that of GB's (126.61 Pa) showing that the RBF kernel has the ability to model the smooth Re–ΔP power-law relationship. No confidence intervals are presented for individual metrics in Table 5 because the held-out test set contains only 15 samples, which does not give a strong indication of the variation in performance; cross-validation results in Section "Cross-validation stability analysis" give a stronger indication of the variation in performance. Figure 10a,b present grouped bar charts comparing R^2 and MAPE, for both models, respectively.

To further verify that model performance is not an artefact of the small held-out test set, Table 6 reports training, test, and full-dataset (n = 75) metrics for both models across all four outputs.

Cross-validation stability analysis

A standard five-fold cross-validation (CV) procedure was conducted on the full 75-observation dataset[1]. Figure 11 presents the mean R^2 score per fold for each model. The GB model demonstrates more consistent cross-validated R^2 scores (mean = 0.9691, σ = 0.0194, range: 0.9321–0.9876) than the SVR model (mean = 0.9590, σ = 0.0263, range: 0.9223–0.9875). Formal statistical comparison of the five-fold pairs is inadmissible: folds share overlapping training sets and the minimum achievable two-tailed p-value at n = 5 is 0.0625. The CV results

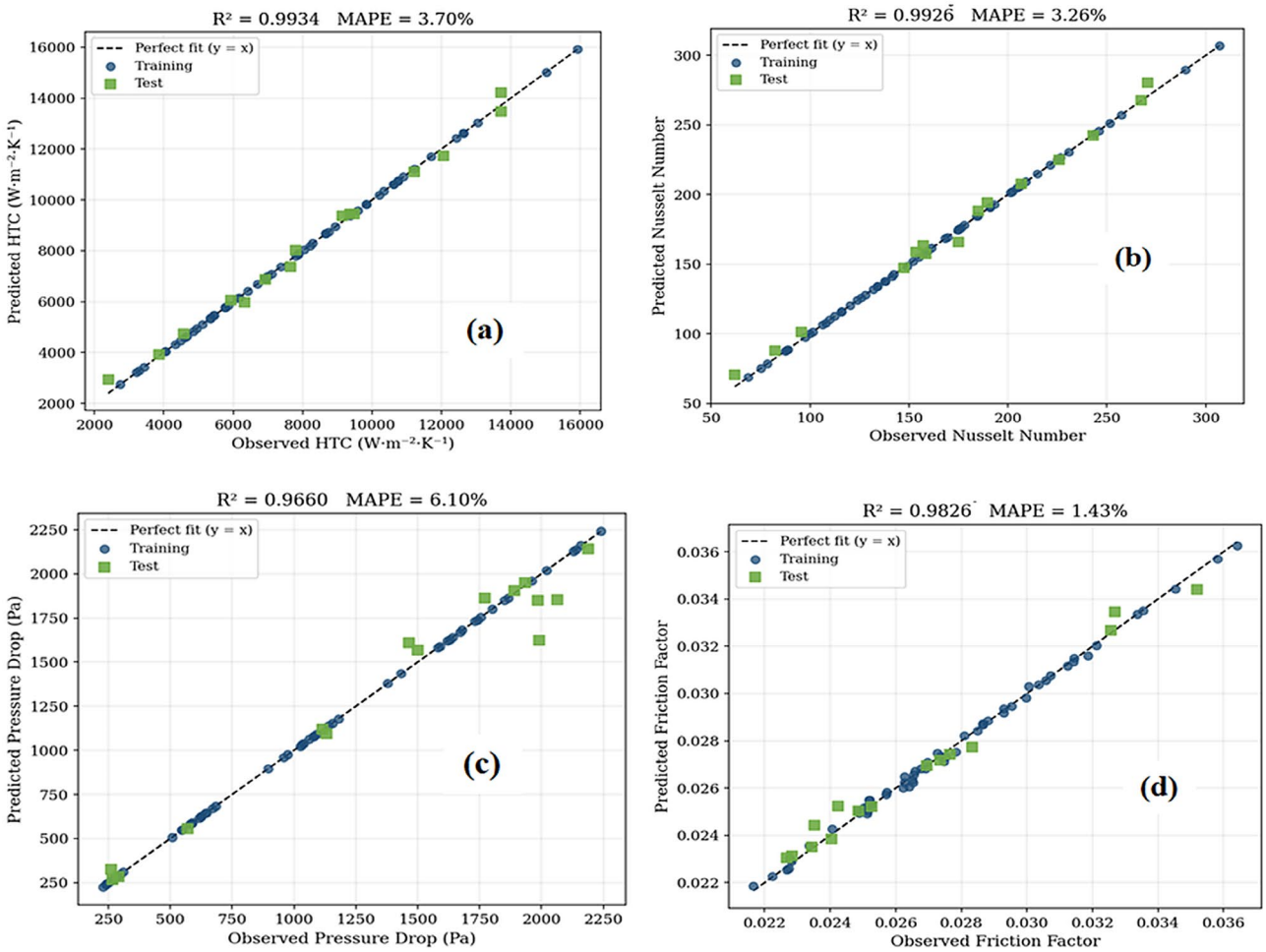


Fig. 9. (**a**–**d**) GB Parity Plots for (**a**) Heat Transfer Coefficient, (**b**) Nusselt Number, (**c**) Pressure Drop, and (**d**) Friction Factor.

	R^2		RMSE		MAE		MAPE%	
Output variable	**SVR**	**GB**	**SVR**	**GB**	**SVR**	**GB**	**SVR**	**GB**
HTC ($W \cdot m^{-2} \cdot K^{-1}$)	0.9931	0.9934	276.34	270.34	196.72	222.12	4.09	3.70
Nusselt number	0.9929	0.9926	5.144	5.252	3.619	4.158	1.86	3.26
Pressure drop (Pa)	0.9913	0.9660	64.16	126.61	51.38	81.59	4.48	6.10
Friction factor	0.9727	0.9826	6.28×10^{-4}	4.60×10^{-4}	5.55×10^{-4}	3.73×10^{-4}	2.07	1.43

Table 5. Model performance comparison: SVR vs GB (held-out test set, n = 15).

are reported as directional indicators of consistency. Table 7 provides a numerical summary of five-fold cross-validation performance comparison between both models.

Feature importance analysis

To access feature importance, the work adopted built-in impurity-based (MDI) importance scores of the fitted GB models for each output, averaged across all four response variables[33]. Figure 12 presents the mean GB feature importance. As anticipated, the Reynolds number is the major contributor, with a total mean importance of 76.7%, followed by nanoparticle concentration at 18.9%, with bulk temperature the smallest contributor at 4.4%. These estimates are only approximate and are missing the interaction term, nor do they represent global sensitivity if readers need the interaction-sensitive version of importance scores, then they should use the decomposition methods of Sobol'. This matches the physical picture: Re governs turbulence intensity and boundary-layer thinning, the main driver of HTC, Nu, and pressure drop; concentration acts secondarily via thermal conductivity and micro-convection; temperature contributes least given the narrow 30–60 °C range. This, and the gap versus Dittus–Boelter ($R^2 = -0.070$ for Nu), shows the models capture concentration-

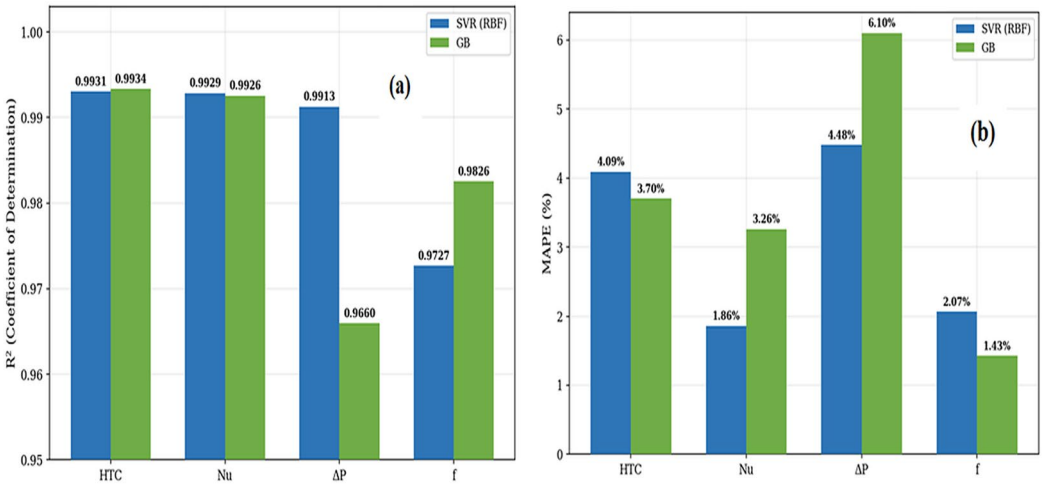


Fig. 10. Comparison of SVR and GB Model Performance Across all four Output Variables based on (**a**) R^2 and (**b**) MAPE (%)

Output	Subset	SVR R^2	SVR MAPE	GB R^2	GB MAPE
HTC	Train	0.9999	0.10%	1.0000	0.14%
HTC	Test (n = 15)	0.9931	4.09%	0.9934	3.70%
HTC	Total (n = 75)	0.9983	0.90%	0.9985	0.85%
Nu	Train	0.9998	0.22%	1.0000	0.07%
Nu	Test (n = 15)	0.9929	1.86%	0.9926	3.26%
Nu	Total (n = 75)	0.9982	0.55%	0.9981	0.74%
ΔP	Train	0.9996	0.37%	1.0000	0.15%
ΔP	Test (n = 15)	0.9913	4.48%	0.9660	6.10%
ΔP	Total (n = 75)	0.9978	1.19%	0.9923	1.34%
f	Train	0.9549	2.29%	0.9977	0.50%
f	Test (n = 15)	0.9727	2.07%	0.9826	1.43%
f	Total (n = 75)	0.9596	2.25%	0.9941	0.68%

Table 6. Performance of SVR and GB models on the training, test, and full datasets (n = 75) for different outputs. Note: Friction factor RMSE and MAE values are dimensionless. Mean MAPE (across all four outputs): SVR = 3.13%; GB = 3.62%

dependence classical correlations miss. Our R^2 range (0.966–0.993) is comparable to related ML-on-nanofluid studies (Kanti et al.[12,20,21]). Permutation importance on the test set confirms the MDI ranking: Re dominant (SVR 73.1%, GB 80.5%), followed by concentration (13.9%/13.4%) and temperature (13.0%/6.1%). A local power-law fit of GB-predicted Nu–Re (c = 1.0 vol%, T = 45 °C) gives exponent 0.66 vs. Dittus–Boelter's 0.8, a flatter learned relationship, consistent with ensemble-learning practice for nanofluid property prediction[47]. Similar data-driven modeling has produced empirical correlations for nanofluid specific heat capacity[48] and neural-network-based density models[49], with interpretable white-box alternatives also proposed for density prediction[50]. Applicability domain: validated only for concentration 0–2.0 vol%, temperature 30–60 °C, Re ≈ 6,807–45,155, and the single fly ash–Cu system/geometry of Kanti et al.[1]; predictions outside these bounds require additional data and retraining. As data-driven models, physical constraints (e.g., monotonicity of Nu with Re) aren't enforced by construction — a physics-informed model (see Future Scope) is the natural extension.

Conclusions

The present investigation demonstrates the effectiveness of machine-learning methods for predicting the coupled thermal and hydraulic behavior of fly ash–Cu hybrid nanofluids. Analysis of the developed models showed that Reynolds number was the dominant controlling parameter, contributing 76.7% to the overall prediction capability, whereas nanoparticle concentration and temperature contributed 18.9% and 4.4%, respectively. Although both algorithms achieved satisfactory performance, their predictive strengths differed across the target variables, detailed below. During five-fold cross-validation, GB provided more consistent predictive behavior, attaining a mean R^2 of 0.9691 with a standard deviation of 0.0194. Based primarily on total-dataset metrics (Table 6, consistent with Table 5): SVR is more accurate for Nu (R^2 = 0.9982 vs. 0.9981) and ΔP (0.9978 vs. 0.9923); GB is more accurate for HTC (0.9985 vs. 0.9983) and f (0.9941 vs. 0.9596). Model choice should thus

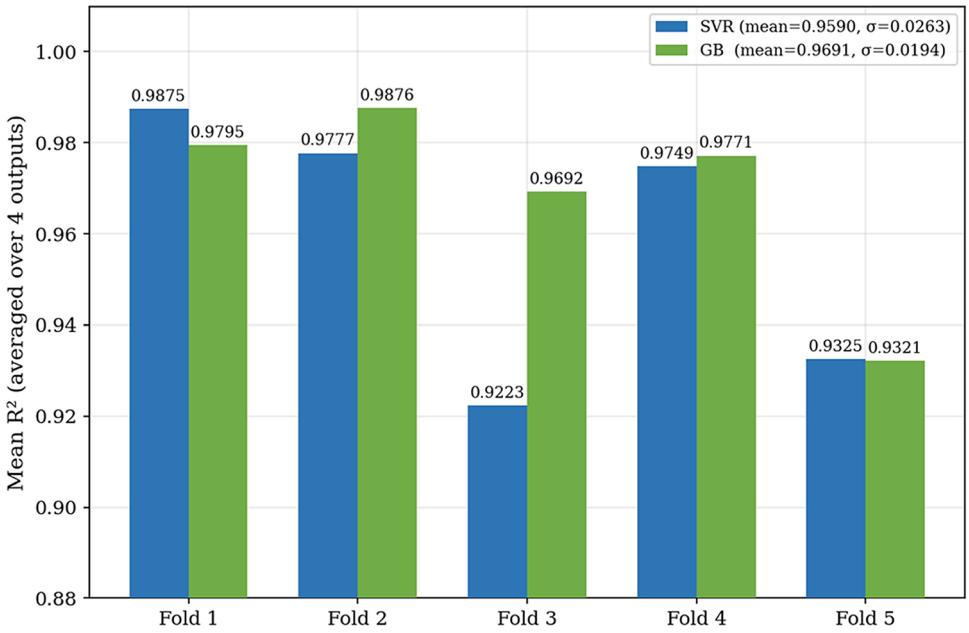


Fig. 11. Five-fold cross-validation: mean R^2 comparison of SVR and GB models

Model	CV Mean R^2	CV Std Dev	CV Min R^2	CV Max R^2
SVR (RBF)	0.9590	0.0263	0.9223	0.9875
GB	0.9691	0.0194	0.9321	0.9876

Table 7. Five-fold cross-validation results: SVR vs GB performance comparison.

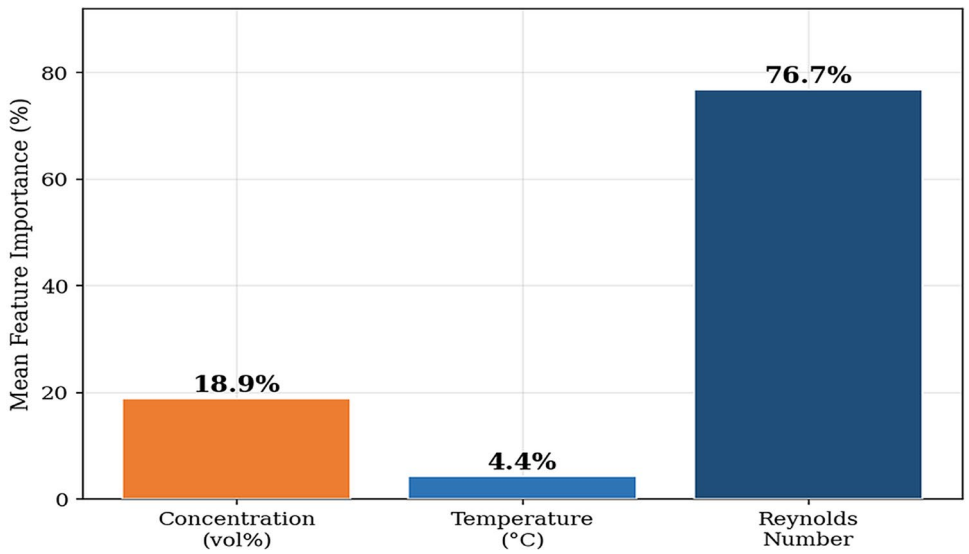


Fig. 12. Gradient boosting feature importance

depend on the target variable, not a single recommendation. The developed models provide a practical, low-cost framework for rapid performance assessment of hybrid nanofluids, and could be embedded as lightweight surrogates within a design-optimization loop or a real-time monitoring system for thermal systems operating within this validated range.

Limitations

Several limitations should be acknowledged. The models were trained and evaluated exclusively on the 75-sample dataset of Kanti et al.[1]—a complete factorial design covering concentration (0–2.0 vol%), temperature (30–60 °C), and Reynolds number (≈ 6,807–45,155) for a single fly ash–Cu system and geometry — with no independent or externally published dataset available for further validation; predictions outside this envelope require additional data and retraining. As purely data-driven models, physical constraints (e.g., monotonicity of Nu with Re) are not enforced by construction. The five-fold cross-validation results should be read as directional indicators of consistency rather than formal statistical comparisons, since the minimum achievable two-tailed p-value at n = 5 is 0.0625. Full graphical partial-dependence visualizations, beyond the power-law exponent check and permutation importance already reported in Section "Feature importance analysis", were not included in this revision.

Future scope of work

Going forward, expanding the experimental dataset to cover a broader range of hybrid nanofluid compositions, nanoparticle mixing ratios, and particle sizes would improve model reliability and applicability. Comparative studies involving Artificial Neural Networks, Random Forests, XGBoost, COMBI models, and Gaussian Process Regression are worth taking up to establish stronger predictive benchmarks, and benchmarking against recently published ensemble learning and gradient boosting frameworks for nanofluid property prediction would further strengthen this assessment. Physics-Informed Neural Networks also present a promising direction, as they incorporate governing physical laws directly into the learning process, making them better suited for reliable predictions beyond the experimentally validated operating range — which remains a key limitation of the present models.

Data availability

The databases used during the current study are available from the corresponding author or reasonable request.

Received: 3 July 2026; Accepted: 10 August 2026
Published online: 20 August 2026

References

1. Kanti, P. K., Sharma, K. V., Said, Z. & Gupta, M. Experimental investigation on thermo-hydraulic performance of water-based fly ash–Cu hybrid nanofluid flow in a pipe at various inlet fluid temperatures. *Int. Commun. Heat Mass Transf.* **124**, 105238. https://doi.org/10.1016/j.icheatmasstransfer.2021.105238 (2021).
2. Said, Z. et al. Sustainable thermal solutions: Enhancing heat transfer with turbulators and nanofluids. *Adv. Energy Sustain. Res.* **6**, 2400335. https://doi.org/10.1002/aesr.202400335 (2025).
3. Ganvir, R. B., Walke, P. V. & Kriplani, V. M. Heat transfer characteristics in nanofluid: A review. *Renew. Sustain. Energy Rev.* **75**, 451–460. https://doi.org/10.1016/j.rser.2016.11.010 (2017).
4. Bacha, H. B., Ullah, N., Hamid, A. & Shah, N. A. A comprehensive review on nanofluids: Synthesis, cutting-edge applications, and future prospects. *Int. J. Thermofluids* **22**, 100595. https://doi.org/10.1016/j.ijft.2024.100595 (2024).
5. Bose, J. R., Manova, S., Asirvatham, L. G. & Wongwises, S. Comprehensive case study on heat transfer enhancement using micro pore metal foams: From solar collectors to thermo electric generator applications. *Case Stud. Therm. Eng.* **27**, 101333. https://doi.org/10.1016/j.csite.2021.101333 (2021).
6. Choi, S. U. S. & Eastman, J. A. Enhancing thermal conductivity of fluids with nanoparticles. *ASME International Mechanical Engineering Congress & Exposition* 99–106 (1955).
7. Rahman, M. A. et al. Review on nanofluids: Preparation, properties, stability, and thermal performance augmentation in heat transfer applications. *ACS Omega* https://doi.org/10.1021/acsomega.4c03279 (2024).
8. Pushparaj, W. C. et al. Balasubramanian Nano- from nature to nurture: a comprehensive review on facets, trends, perspectives, and sustainability of nanotechnology in the food sector. *Energy* **240**, 122732. https://doi.org/10.1016/j.energy.2021.122732 (2022).
9. Shafiq, A., Iqbal, M. A., Sindhu, T. N., Iqbal, H. & Iqbal, H. Synthesis, thermophysical behavior, and environmental implications of nanofluids: A comprehensive review. *Results Eng.* **28**, 107763. https://doi.org/10.1016/j.rineng.2025.107763 (2025).
10. Kalsi, S., Kumar, S., Kumar, A., Alam, T. & Sharma, A. Yadav A review on hybrid nanofluids for heat transfer: advancements, synthesis, challenges and applications. *Discov. Appl. Sci.* **7**(7), 698. https://doi.org/10.1007/s42452-025-07141-8 (2025).
11. Tekir, M. Experimental study on the thermal performance of hybrid nanofluid in a compact plate heat exchanger under the influence of a magnetic field. *Case Stud. Therm. Eng.* **69**, 106031. https://doi.org/10.1016/j.csite.2025.106031 (2025).
12. Kanti, P. K. et al. Experimental and explainable machine learning approach on thermal conductivity and viscosity of water based graphene oxide based mono and hybrid nanofluids. *Sci. Rep.* **14**, 30967. https://doi.org/10.1038/s41598-024-81955-1 (2024).
13. Vidhya, R., Balakrishnan, T. & Kumar, B. S. Investigation on thermophysical properties and heat transfer performance of heat pipe charged with binary mixture based ZnO-MgO hybrid nanofluids. *Mater. Today Proc.* **37**, 3423–3433 (2021).
14. Rafid, M. et al. Augmentation of heat exchanger performance with hybrid nanofluids: Identifying research gaps and future indications - A review. *Int. Commun. Heat Mass Transf.* **155**, 107537. https://doi.org/10.1016/j.icheatmasstransfer.2024.107537 (2024).
15. Ramadhan, A. I., Azmi, W. H., Mamat, R. & Hamid, K. A. Experimental and numerical study of heat transfer and friction factor of plain tube with hybrid nanofluids. *Case Stud. Therm. Eng.* **22**, 100782. https://doi.org/10.1016/j.csite.2020.100782 (2020).
16. Assaf, Y. H. et al. Investigation of heat transfer performance in heat exchangers using hybrid nanofluids and twisted tape inserts with fixed special rings. *Sci. Rep.* **15**, 18450. https://doi.org/10.1038/s41598-025-02135-3 (2025).
17. Syam Sundar, L., Sharma, K. V., Singh, M. K. & Sousa, A. C. M. Hybrid nanofluids preparation, thermal properties, heat transfer and friction factor: A review. *Renew. Sustain. Energy Rev.* **68**, 185–198. https://doi.org/10.1016/j.rser.2016.09.108 (2017).
18. Kanti, P. K. et al. Experimental investigation of solar collector efficiency enhancement using surfactant stabilized MWCNT and Al_2O_3 hybrid nanofluid. *Sci. Rep.* https://doi.org/10.1038/s41598-026-42016-x (2026).
19. Vicki Wanatasanappan, V. et al. Viscosity and rheological behavior of Al2O3-Fe2O3/water-EG based hybrid nanofluid: A new correlation based on mixture ratio. *J. Mol. Liq.* **375**, 121365. https://doi.org/10.1016/j.molliq.2023.121365 (2023).
20. Kanti, P. K. et al. Thermal performance, entropy generation, and machine learning insights of Al_2O_3-TiO_2 hybrid nanofluids in turbulent flow. *Sci. Rep.* **15**, 16035. https://doi.org/10.1038/s41598-025-93749-0 (2025).
21. Kanti, P. K. et al. Bayesian-optimized machine learning and experimental study of Al_2O_3-CuO hybrid nanofluid thermal performance in turbulent circular tube flow. *Sci. Rep.* **15**, 38717. https://doi.org/10.1038/s41598-025-23785-3 (2025).

22. Singh, B. & Sood, S. Hybrid nanofluids preparation, thermo-physical properties, and applications: A Review. *Hybrid Adv.* **6**, 100192. https://doi.org/10.1016/j.hybadv.2024.100192 (2024).
23. Manikanta, J. E. et al. A review on hybrid nanofluids: Preparation methods, thermo physical properties and applications. *Iran. J. Sci. Technol. Trans. Mech. Eng.* **49**, 67–79. https://doi.org/10.1007/s40997-024-00772-z (2025).
24. Mathur, P. et al. Comprehensive framework of machine learning and deep learning architectures with metaheuristic optimization for high-fidelity prediction of nanofluid specific heat capacity. *Sci. Rep.* **15**, 45705. https://doi.org/10.1038/s41598-025-28268-z (2025).
25. Karimipour, A. et al. A novel nonlinear regression model of SVR as a substitute for ANN to predict conductivity of MWCNT-CuO/water hybrid nanofluid based on empirical data. *Physica A* **521**, 89–97. https://doi.org/10.1016/j.physa.2019.01.055 (2019).
26. Sheikhmohammadi, A., Khakzad, P., Rasolevandi, T. & Azarpira, H. Leveraging artificial intelligence models (GBR, SVR, and GA) for efficient chromium reduction via UV/trichlorophenol/sulfite reaction. *Results Eng.* **26**, 104599. https://doi.org/10.1016/j.rineng.2025.104599 (2025).
27. Syam Sundar, L., Mewada, H. K. & Sangaraju, S. Experimental and ensemble machine learning analyses of heat transfer, friction factor and thermal performance factor of rGO/water nanofluids in a tube. *Int. J. Thermofluids* **21**, 100557. https://doi.org/10.1016/j.ijft.2023.100557 (2024).
28. Shekhar, S. K. et al. Predicting nanofluid density in ethylene glycol-based oxide nanoparticles using machine learning approach: GBR–GSO models. *J. Therm. Anal. Calorim.* **150**, 4441–4458. https://doi.org/10.1007/s10973-024-13928-5 (2025).
29. Alisha, A., Aneja, M. & Goel, S. Recent progress in machine learning and deep learning models for heat transfer problems. *J. Eng. Appl. Sci.* **73**, 72. https://doi.org/10.1186/s44147-026-00903-0 (2026).
30. Sajjad, U. et al. A deep learning method for estimating the boiling heat transfer coefficient of porous surfaces. *J. Therm. Anal. Calorim.* **145**, 1911–1923 (2021).
31. Albdour, S. A., Addad, Y., Rabbani, S. & Afgan, I. Machine learning-driven approach for predicting the condensation heat transfer coefficient (HTC) in the presence of non-condensable gases. *Int. J. Heat Fluid Flow* **106**, 109330 (2024).
32. Alade, I. O., Abd Rahman, M. A. & Saleh, T. A. Modeling and prediction of the specific heat capacity of Al_2O_3/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Struct. Nano-Objects* **17**, 103–111 (2019).
33. Alade, I. O., Oyehan, T. A., Popoola, I. K., Olatunji, S. O. & Bagudu, A. Modeling thermal conductivity enhancement of metal and metallic oxide nanofluids using support vector regression. *Adv. Powder Technol.* **29**(1), 157–167 (2018).
34. Said, Z. & Sharma, P. (Eds.) *Machine Learning for Sustainable Energy Solutions* First Edition (Wiley, 2025). https://doi.org/10.1002/9781394267439.
35. Dittus, F. W. & Boelter, L. M. K. Heat transfer in automobile radiators of the tubular type, university of california publications on. *Engineering* **2**, 443–461 (1930).
36. Petukhov, B. S. Heat transfer and friction in turbulent pipe flow with variable physical properties. *Adv. Heat Transf.* **6**, 503–564. https://doi.org/10.1016/S0065-2717(08)70153-9 (1970).
37. Blasius, H. Das Aehnlichkeitsgesetz bei Reibungsvorgängen in Flüssigkeiten. *Mitteilungen über Forschungsarbeiten auf dem Gebiete des Ingenieurwesens* 131 (1913).
38. Ahirwar, B. K. & Kumar, A. Experimental investigation for heat transfer performance of CuO-water nanofluid in a double pipe heat exchanger. *J. Therm. Anal. Calorim.* **149**, 4133–4151. https://doi.org/10.1007/s10973-024-12947-6 (2024).
39. Mohammed, H. I. et al. The role of nanotechnology and artificial intelligence in optimizing thermal energy systems. *Appl. Energy* **400**, 126576. https://doi.org/10.1016/j.apenergy.2025.126576 (2025).
40. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362. https://doi.org/10.1038/s41586-020-2649-2 (2020).
41. McKinney, W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference (SciPy 2010)* 56–61 (2010). https://doi.org/10.25080/Majora-92bf1922-00a
42. Miandoab, A. R., Bagherzadeh, S. A. & Isfahani, A. H. M. Python-based machine learning estimation of thermo-hydraulic performance along varying nanoparticle shape, nanofluid and tube configuration. *Adv. Eng. Softw.* **197**, 103762. https://doi.org/10.1016/j.advengsoft.2024.103762 (2024).
43. Eskandari, E., Alimoradi, H., Pourbagian, M. & Shams, M. Enhanced predictive modeling of Nusselt number in boiler tubes: Numerical simulations and machine learning for water and SiO_2/water. *J. Therm. Anal. Calorim.* **149**, 10119–10148. https://doi.org/10.1007/s10973-024-13409-9 (2024).
44. Kumar, K. P. et al. Effect of particle size on thermodynamics of Al_2O_3 nanofluid: XGBoost and gradient boosting regression. *Int. J. Therm. Sci.* **197**, 108825 (2024).
45. Godasiaei, S. & Chamkha, A. Machine learning insights for molten salt heat exchangers. *Numer. Heat Transfer A* **86**, 5047–5064 (2025).
46. Masalha, I., Alahmer, A. & Badran, O. Proactive thermal management of photovoltaic systems using nanofluid cooling and machine learning. *J. Therm. Anal. Calorim.* **150**, 17529–17570 (2025).
47. Deymi, O. et al. Employing ensemble learning techniques for modeling nanofluids' specific heat capacity. *Int. Commun. Heat Mass Transfer* **143**, 106684. https://doi.org/10.1016/j.icheatmasstransfer.2023.106684 (2023).
48. Hadavimoghaddam, F., Atashrouz, S., Nedeljkovic, D., Abuswer, M. A. & Hemmati-Sarapardeh, A. Toward empirical correlations for estimating the specific heat capacity of nanofluids utilizing GRG, GP, GEP, and GMDH. *Sci. Rep.* **13**, 20763. https://doi.org/10.1038/s41598-023-47327-x (2023).
49. Deymi, O. et al. On the evaluation of mono-nanofluids' density using a radial basis function neural network optimized by evolutionary algorithms. *Therm. Sci. Eng. Prog.* **53**, 102750. https://doi.org/10.1016/j.tsep.2024.102750 (2024).
50. Deymi, O. et al. Innovative mathematical correlations for estimating mono-nanofluids' density: Insights from white-box machine learning. *Results Phys.* **73**, 108248. https://doi.org/10.1016/j.rinp.2025.108248 (2025).

Author contributions

VKV, VPM: Conceptualization, Methodology, Writing – review, JA, ST: Supervision, Project administration, PHN: Methodology, Investigation. HLA: Methodology, Supervision, Validation, Review & editing.

Funding

No funding received.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.P.H. or H.L.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026