Co-published by
Institute of Fluid-Flow Machinery
Polish Academy of Sciences
Committee on Thermodynamics and Combustion
Polish Academy of Sciences

http://www.imp.gda.pl/archives-of-thermodynamics/

ARCHIVES OF THERMODYNAMICS

Investigation of Heat Transfer and Flow Characteristics in Solar Air Heaters with a Spherical Turbulator Using Machine Learning Models

Vijay Singh Bisht[a], Desh Bandhu Singh[a], Pushpendra Kumar[a], Prabhakar Bhandari[b,c*], Rupesh Kumar Tipu[d], Sandeep Singh[e], Ankur Singh Bist[f]

[a]Department of Mechanical Engineering, Graphic Era (Deemed to be) University, Dehradun 248002, Uttarakhand, India
[b]Department of Mechanical Engineering, K.R. Mangalam University, Gurugram 122103, Haryana, India
[c]Centre of Excellence-Cyber Security, School of Engineering and Technology, K.R. Mangalam University, Gurugram 122103, Haryana, India
[d]Department of Civil Engineering, School of Engineering and Technology, K.R. Mangalam University, Gurugram 122103, Haryana, India
[e]O.P. Jindal Global University, Sonipat 131001, Haryana, India
[f]CSE, Graphic Era Hill University, Bhimtal Campus, Uttarakhand, India
*Corresponding author email: prabhakar.bhandari40@gmail.com

Received: 25.12.2025; revised: 21.03.2026; accepted: 22.04.2026

Abstract

Solar air heaters are attractive for low-carbon thermal applications, but their performance is constrained by weak convective heat transfer in the near-wall region. In this work, a single-pass rectangular solar air heater duct equipped with spherical turbulators is investigated numerically and then accelerated using machine-learning surrogate models for rapid prediction of thermal–hydraulic responses. Five turbulator arrangements (V-, M-, W-shaped, inclined, and arc-shaped) were evaluated at three discrete spacing levels (pitch ratio P/D = 3, 6, 9) with a constant sphere diameter of 25 mm over Re = 3500–23 500, consistent throughout all the simulations. The computational fluid dynamics model employed a constant heat flux of 1000 W/m^2 and standard pressure–velocity coupling/discretisation practices, and was validated against established previous work. A leakage-safe machine-learning dataset (675 samples, 36 variables) was constructed from computational fluid dynamics outputs and physics-informed engineered features; models were trained using geometry-grouped cross-validation to ensure generalisation across turbulator arrangements and spacing levels. Among candidate regressors, histogram gradient boosting provided the best Nu surrogate (*OOF RMSE* = 3.299, R^2 = 0.9908, *MAPE* = 2.44%), while ridge regression yielded the most accurate friction factor surrogate (*OOF RMSE* = 4.70×10^{-4}, R^2 = 0.9886, *MAPE* = 1.12%). The combined computational fluid dynamics and machine-learning framework enables fast, reliable evaluation of solar air heater thermo-hydraulic performance for configuration screening and optimisation within the validated operating envelope.

Keywords: Solar air heater; Spherical turbulator; Thermo-hydraulic performance; Surrogate modelling; Gradient boosting; Ridge regression

Vol. 47(2026), No. 2, 99–118; doi: 10.24425/ather.2026.158677

Cite this manuscript as: Bisht, V.S., Singh, D.B., Bhandari, P., Kumar, P., Tipu, R.K., Singh, S., & Bist, A.S. (2026). Investigation of Heat Transfer and Flow Characteristics in Solar Air Heaters with a Spherical Turbulator Using Machine Learning Models. *Archives of Thermodynamics*, 47(2), 99−118.

1. Introduction

In recent years, the combination of rapid population expansion and intensified industrial activity has significantly increased global energy consumption. This rising demand continues to depend largely on traditional fuels – coal, petroleum and natural gas – which are not only diminishing quickly but also causing substantial environmental harm. Consequently, the search for

Nomenclature

A – cross section area, m^2
C_p – specific heat of air, J/(kg·K)
D – diameter, m
F – friction factor
h – heat transfer coefficient, W/(m^2·K)
k – turbulent kinetic energy, J/kg
k_a – thermal conductivity, W/(m·K)
M – mass flow rate of air, kg/s
Nu – Nusselt number
p – pressure, kPa
Δp – pressure drop, N/m^2
P – pitch, perimeter, m
P/D– relative roughness pitch
Pr – Prandtl number
q – heat flux, W/m^2
R^2 – coefficient of determination or R-squared
Re – Reynolds number
t – thickness, m
T – temperature, K
U_L – head loss coefficients, W/(m^2·K)
u, υ, w – velocity components in x, y and z direction, m/s
V – mean flow velocity, m/s
W – channel width, mm
x, y, z – coordinates, m

Greek symbols

ε – turbulence-specific dissipation rate, m^2/s^3
μ – dynamic viscosity of air, N s/m^2
ν, ν_T– kinematic viscosity, eddy viscosity, m^2/s
ρ – density of air, kg/m^3

Subscripts and Superscripts

in – inlet
out – outlet
r – roughened
s – smooth

Abbreviations and Acronyms

ANN – artificial neural networks
ASHRAE– American Society of Heating, Refrigerating and Air Conditioning Engineers
CFD – computational fluid dynamics
CNN – convolutional neural network
ML – machine-learning
EDA – exploratory data analysis
GA – genetic algorithms
HistGB – histogram-based gradient boosting
MAE – mean absolute error
MAPE – mean absolute percentage error
MLP – multi-layer perceptron
OOF – out-of-fold
PSO – particle swarm optimisation
RF – random forest
RMSE – root mean squared error
RNG – re-normalisation group
SAH – solar air heater
SIMPLE – semi-implicit method for pressure-linked equation
SST – shear stress transport
ST – spherical turbulators
SVM – support vector machine
SVR – support vector regression
THPP – thermo-hydraulic performance parameter

clean and sustainable energy options has become critical. Renewable resources such as solar, biomass and hydropower have gained attention as viable alternatives, with solar energy emerging as particularly important because of its wide availability, especially in high-insolation regions like India. Thus, utilising solar energy has become essential in the worldwide move toward environmentally responsible power production. Solar energy serves numerous purposes, including water heating [1], space cooling [2], cooking [3], and purification [4]. Among these applications, solar air heating is especially significant in colder regions. A solar air heater (SAH) is a cost-effective and environmentally friendly system that absorbs solar radiation to enable heat transfer [5].

Its straightforward structure, minimal maintenance needs, and freedom from fossil fuels make it suitable for both domestic and industrial heating. Despite these advantages, SAHs face a key drawback: their thermal efficiency remains low because of insufficient heat transfer between the absorber plate and the air passing over it [6].

SAH systems typically consist of an absorber plate that collects incoming solar radiation, a transparent cover that reduces thermal losses, and an insulated enclosure that helps retain the captured heat. Within a SAH, the air flowing close to the absorber plate encounters a turbulent boundary layer, creating a thin viscous sublayer where heat transfer is limited due to reduced air velocity and low thermal conductivity. As a result, the overall thermal performance of the system remains restricted under conventional operating conditions. To address this limitation, a variety of passive enhancement techniques have been introduced, such as adding surface roughness features, integrating turbulence-generating components, and altering duct configurations. These methods are designed to increase airflow turbulence and consequently improve heat transfer. Among the different enhancement strategies, artificial roughness has become one of the most commonly used and highly effective solutions for boosting the thermal efficiency of SAHs [7]. This method involves placing features such as ribs, dimples, or baffles on the absorber surface to disrupt the laminar sublayer and strengthen heat transfer [8].

Artificial roughness provides a passive, energy-free means of improving thermal output; it is simple to apply, economical, and substantially enhances the heat-exchange process. However, it can also lead to higher frictional losses, making careful design optimisation necessary to achieve an appropriate balance between increased heat transfer and pressure drop. Numerous studies have investigated the ideal configuration of roughness elements in SAHs through both computational analyses and experimental studies.

Chaube et al. [9] carried out a two-dimensional numerical analysis using the shear stress transport (SST) k–ω turbulence model to examine the effects of various rib geometries on heat transfer and airflow characteristics in a single-pass solar air heater. Among the rib configurations tested – rectangular, chamfered, square, circular and semi-circular – the chamfered design produced the greatest heat transfer enhancement, whereas rectangular ribs delivered the highest thermo-hydraulic efficiency by offering an optimal balance between thermal improvement and pressure drop. This study highlights the critical role of rib shape in boosting the performance of solar air heaters. In another investigation, Kumar and Saini [10] performed numerical simulations on a SAH equipped with arc-shaped ribs over a Reynolds number range of 6000–18 000 and relative roughness heights between 0.0299 and 0.0426. Their work demonstrated that, among multiple turbulence models assessed, the re-normalisation group (RNG) k–ε model provided the most reliable predictions. Yadav and Bhagoria [11] numerically evaluated the influence of circular transverse wire ribs on SAH performance and found that the ribbed arrangement increased the Nusselt number by up to 2.6 times, achieving a maximum thermo-hydraulic performance factor of 1.8. Al-Chlaihawi et al. [12] also carried out a numerical study in which they employed a combination of transverse trapezoidal ribs and chamfered grooves to enhance thermal behaviour in a SAH duct. Their analysis explored parameters such as relative roughness pitch, relative roughness height and Reynolds number, leading to optimised values of improved performance. Additionally, they proposed new correlations for predicting both the Nusselt number and friction factor.

Manjunath et al. [13] investigated the use of spherical turbulence promoters and reported as much as a 2.45-fold increase in heat transfer, accompanied by a moderate rise in friction factor and a thermo-hydraulic performance factor (THPF) of 1.74, demonstrating their effectiveness in enhancing system performance. Maithani et al. [14] employed wavy delta-winglet vortex generators in rectangular ducts, producing swirling airflow that improved heat transfer by 30–40% compared to smooth ducts while maintaining the performance evaluation criterion (PEC) greater than 1, despite the associated rise in pressure drop. In a separate study, Manjunath et al. [15] examined absorber plates incorporating sinusoidal wave patterns and found that an optimised configuration – with a wavelength of 1.0 and an aspect ratio of 1.5 – raised the thermal efficiency by 12.5%, although an increased pumping power was required. Researchers have also explored a wide range of fin configurations to enhance SAH performance, including pin fins [16], stepped cylindrical turbulence generators [17], helicoidal spring-shaped fins [18], and others.

Drawing inspiration from geometric shapes of the English alphabet letters, numerous researchers have investigated how alphabet-shaped ribs affect the heat transfer and flow characteristics of solar air heaters (SAHs). Taslim et al. [19] and Olsson and Sunden [20] analysed heat transfer in square channels and found that V-shaped ribs performed better than both inclined and transverse ribs in enhancing thermal output. Momin et al. [21] experimentally evaluated the effects of rib height and attack angle on heat transfer and flow behaviour in SAHs equipped with V-shaped wire ribs, confirming that V-shaped ribs achieve higher thermal efficiency than inclined rib configurations under comparable conditions. Lanjewar et al. [22] carried out an experimental study on SAHs incorporating W-shaped wire rib roughness and reported that W-shaped ribs surpassed V-shaped ribs in overall thermo-hydraulic performance. Ghildyal et al. [23] compared D-shaped, reverse D-shaped and U-shaped turbulators, determining that U-shaped designs achieved the greatest Nusselt number and overall thermo-hydraulic performance, while the D-shaped turbulator exhibited the highest friction factor. In a subsequent study, Ghildyal et al. [24] showed that W-contoured turbulators delivered the best thermal efficiency when benchmarked against tapered and reverse-tapered variants over Reynolds numbers ranging from 4000 to 18 000. Bohra et al. [25] demonstrated that Z-shaped baffles mounted on the absorber plate significantly improved SAH performance, with the highest efficiency obtained at a blockage ratio of 0.3, a pitch ratio of 1.5, and a baffle angle of 45° for Reynolds numbers between 5000 and 10 000. Similarly, Chaudhari et al. [26] identified that inverted L-shaped ribs with a roughness pitch of 17.86 achieved maximum exergy efficiency at a heat flux of 1000 W/m^2. Kumar et al. [27] analysed a hybrid roughness configuration combining S-shaped ribs and protrusions to further improve heat transfer.

Alhilali et al. [28] employed I-shaped ribs in a two-dimensional numerical study and reported that they enhanced the Nusselt number by 2.498–2.978 times and increased the friction factor by 2.812–4.393 times. Extending these concepts to parallel-plate heat exchangers, Chtourou et al. [29] utilised Y-shaped and C-shaped obstacles, observing a maximum thermo-hydraulic performance parameter (THPP) of approximately 1.44 for Y-shaped ribs and around 2.6 for C-shaped ribs. In a follow-up study, Chtourou et al. [30] refined the Y-shaped geometry into classical, fragmental and fractal variants and assessed their behaviour in rectangular and triangular arrangements. Among all designs, the fragmental Y configuration arranged triangularly yielded the highest thermo-hydraulic performance. Beyond these, researchers have also explored novel rib geometries such as plus (+)-shaped turbulators [31] and hook-shaped ribs [32].

Beyond rib geometry, the orientation or layout of ribs on the absorber plate significantly influences the underlying flow behaviour. When a transverse rib is placed in the passage, it generates a pair of vortices on its upstream and downstream sides. These vortices tend to remain relatively stationary compared with the bulk airflow, causing elevated temperatures within the vortex regions and near the wall, which ultimately decreases the local heat transfer rate. In contrast, an inclined rib allows the vortical structures to convect along the rib's length – fluid enters near the leading edge and exits toward the trailing edge, reattaching to the primary stream. This motion introduces secondary flow patterns that sweep across the channel width, producing marked spatial differences in heat transfer. Consequently, the rib's inclination angle becomes just as important as its height and pitch in enhancing thermal transport performance [33]. Several rib orientations have been examined in the literature, including inclined [34], transverse [35], V-shaped [36], and others.

Artificial neural networks (ANNs) have become useful tools in SAH research because they can model nonlinear relationships among operating conditions, geometry, heat transfer, friction loss and thermal performance. Ghritlahre et al. [37] used relevant input parameters to predict the performance of a solar air heater through an ANN model. Bhattacharyya et al. [38] developed an ANN-based model for a solar-air-heater tube and predicted the Nusselt number and thermo-hydraulic efficiency with high accuracy. Ghritlahre and Prasad [39] also used MLP, GRNN and RBF neural-network models to predict the exergetic performance of solar air heaters, showing the suitability of ANN-based models for solar thermal performance estimation.

Recent studies have extended this direction by combining CFD-generated data with machine-learning models. Alatawi [40] integrated CFD and ML for solar air heaters with distinct rib configurations and reported that a CNN model outperformed random forest regression and support vector regression in predicting heat transfer coefficients. Review evidence also shows that intelligent methods are increasingly used for solar heater prediction and optimisation [41]. However, many earlier CFD-ML studies still focus mainly on prediction accuracy and do not fully test geometry-wise generalisation under leakage-safe grouped validation. The present study addresses this limitation by using a geometry-grouped surrogate framework for spherical-turbulator solar air heaters.

Despite these advances, three important gaps remain in the current CFD-ML literature on solar air heaters. First, most prior studies emphasise prediction accuracy for one or more performance parameters, but do not evaluate whether the surrogate can generalise to previously unseen geometry groups under leakage-safe grouped validation. Second, many studies focus on conventional ribbed or finned geometries, whereas geometry-specific surrogate modelling for spherical-turbulator layouts and their arrangement-dependent thermo-hydraulic ranking remains limited. Third, several CFD-ML studies report fast prediction, but do not connect surrogate performance to a clear design outcome through frozen-holdout testing, uncertainty bounds, and thermo-hydraulic selection of the best configuration. The present work addresses these gaps by combining validated CFD simulations with a grouped and leakage-safe surrogate framework for five spherical-turbulator arrangements, followed by frozen-holdout evaluation, bootstrap-based uncertainty quantification, and thermo-hydraulic performance parameter (THPP)-based design ranking.

The objectives of the present work are fivefold:

1) to construct a supervised CFD-derived dataset for solar air heaters equipped with spherical turbulators, using operating and geometric descriptors, including the Reynolds number, configuration, and pitch ratio, to predict the Nusselt number (Nu) and friction factor (*f*);
2) to establish a transparent and reproducible data-preparation workflow with explicit feature documentation and leakage control for tabular thermo-fluid surrogate modelling;
3) to design a geometry-aware evaluation framework based on configuration × pitch grouping, grouped cross-validation, and a frozen geometry-grouped holdout test so that generalisation is assessed on unseen layouts rather than near-duplicate cases;
4) to benchmark multiple regression surrogates and identify the most reliable predictor for each target, together with bootstrap-based uncertainty estimates for design-screening reliability;
5) to translate the CFD-ML workflow into a practical thermo-hydraulic design outcome by ranking the investigated spherical-turbulator configurations using THPP and identifying the best-performing arrangement within the investigated design space.

In this way, the contribution of the study lies not only in replacing repeated CFD runs with fast surrogates, but in providing a geometry-aware, uncertainty-supported and design-oriented CFD-ML framework for configuration screening of spherical-turbulator solar air heaters.

2. Solar air heater geometry and different configurations

The present work examines a SAH equipped with spherical turbulators through both analytical formulation and numerical simulation. A schematic representation of the system is shown in Fig. 1, where spherical elements are positioned beneath the absorber plate to intensify heat transfer.

The duct has a rectangular cross-section with a width of 0.15 m, a height of 0.03 m and an overall length of 1.85 m, consisting of a 0.35 m inlet, a 1.20 m heating region and a 0.30 m outlet. The design of SAH follows the recommendations specified in American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) Standard 93−77 [13,42].

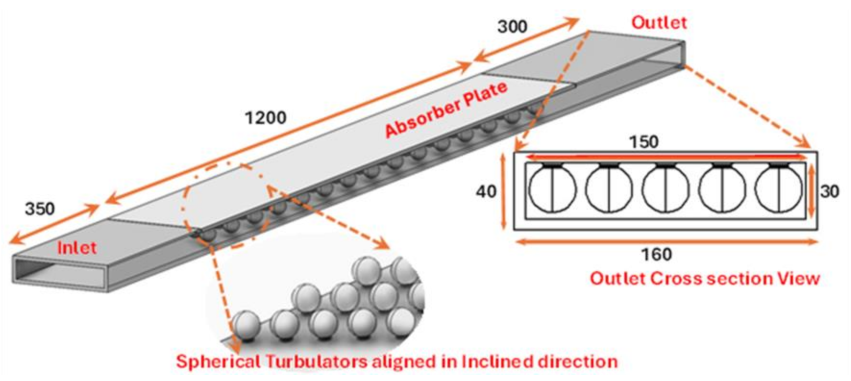


Fig.1. Layout of SAH duct with cross-sectional view and zoomed view.

The absorber plate extends across the active heating zone and captures solar radiation before conveying thermal energy to the air flowing beneath it. Spherical turbulators with a diameter of 25 mm are arranged in uniformly spaced rows along the underside of the plate. Their placement is intended to disrupt the boundary layer, induce turbulence, and thereby elevate the convective heat transfer coefficient. The cross-sectional view illustrates the positioning of the spheres within the flow passage, while the magnified view depicts their specific layout pattern. Each configuration of spherical turbulators is designed to enhance heat transfer effectiveness while controlling pressure losses to achieve improved thermo-hydraulic performance.

This study investigates how various orientations of spherical turbulators influence the thermal and flow behaviour of SAHs. To facilitate this analysis, five separate configurations were formulated, each featuring a distinct pattern of turbulators positioned beneath the absorber plate. The examined arrangements include V-shaped, M-shaped, W-shaped, inclined and arc-shaped layouts, all intentionally designed to modify airflow characteristics and intensify turbulence within the duct. A top-view representation of these configurations is shown in Fig. 2, illustrating the placement and directional alignment of the spherical turbulators across the absorber plate region for each design. This comparative framework allows for a systematic assessment of the impact that different turbulator patterns have on heat transfer augmentation and overall thermo-hydraulic performance.

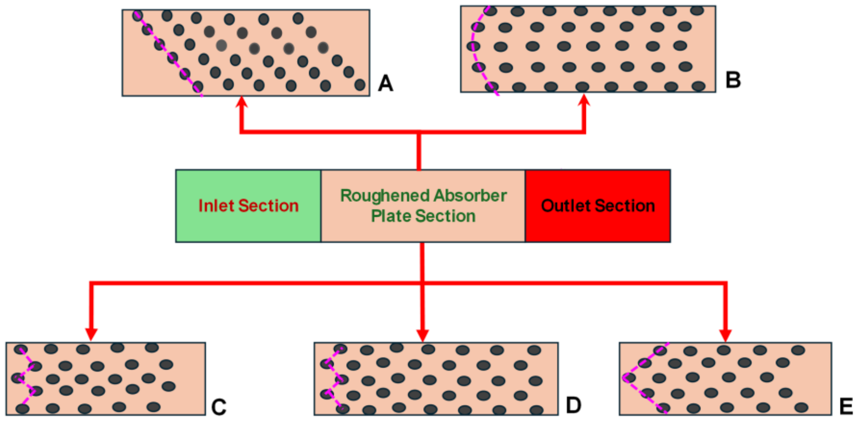


Fig. 2. Various arrangement of spherical turbulators in SAHs (a) inclined arrangement, (b) arc-shaped arrangement, (c) M-shaped arrangement, (d) W-shaped arrangement (e) V-shaped arrangement.

In every design considered, the spherical turbulators were kept at a constant diameter of 25 mm to maintain uniformity in geometry. This specific size was chosen because earlier findings by Manjunath et al. [13] indicated that a 25 mm sphere provides the highest thermal and effective efficiency. For all configurations, the angle of attack was maintained at 60°. Additionally, each turbulator pattern was evaluated at three different relative roughness pitch values – 3, 6 and 9 – to assess how variations in spacing and arrangement affect the overall thermo-hydraulic performance of the system.

3. CFD and ML modelling

3.1. CFD numerical model

To perform the numerical analysis of SAHs, several foundational assumptions were introduced to streamline the computational framework and maintain reliable simulation results. These assumptions include:

- The working fluid – air – is treated as a Newtonian, continuous medium with consistent flow characteristics across the entire domain;
- The airflow is considered incompressible, meaning density variations during motion are assumed to be negligible;
- A no-slip condition is applied to all solid boundaries, ensuring the fluid velocity at the wall is equal to the wall velocity;
- The thermo-physical properties of air, such as density, viscosity, thermal conductivity and specific heat, are assumed to remain constant throughout the simulation;
- Heat transfer is modelled through conduction and convection only, while thermal radiation effects are regarded as minimal and therefore omitted from the analysis.

3.1.1. Conservation equations and boundary conditions

Any thermodynamic system operating over a specified duration must satisfy the fundamental conservation principles of mass, momentum and energy. The Cartesian form of these governing equations – continuity, momentum, and energy – is expressed as follows [43,44]:

- mass conservation:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0, \tag{1}$$

- momentum conservation:

– (M_x):

$$\frac{\partial u}{\partial t}+\frac{\partial (uu)}{\partial x}+\frac{\partial (vu)}{\partial y}+\frac{\partial (wu)}{\partial z}=\frac{\partial}{\partial x}\left[(\nu+\nu_T)\frac{\partial u}{\partial x}\right]+$$
$$+\frac{\partial}{\partial y}\left[(\nu+\nu_T)\frac{\partial u}{\partial y}\right]+\frac{\partial}{\partial z}\left[(\nu+\nu_T)\frac{\partial u}{\partial z}\right]+\left(-\frac{1}{\rho}\frac{\partial p}{\partial x}+s_u'\right), \tag{2}$$

– (M_y):

$$\frac{\partial v}{\partial t}+\frac{\partial(uv)}{\partial x}+\frac{\partial(vv)}{\partial y}+\frac{\partial(wv)}{\partial z}=\frac{\partial}{\partial x}\left[(\nu+\nu_T)\frac{\partial v}{\partial x}\right]+$$

$$+\frac{\partial}{\partial y}\left[(\nu+\nu_T)\frac{\partial v}{\partial y}\right]+\frac{\partial}{\partial z}\left[(\nu+\nu_T)\frac{\partial v}{\partial z}\right]+\left(-\frac{1}{\rho}\frac{\partial p}{\partial y}+s'_v\right), \quad (3)$$

– (M_z):

$$\frac{\partial w}{\partial t}+\frac{\partial(uw)}{\partial x}+\frac{\partial(vw)}{\partial y}+\frac{\partial(ww)}{\partial z}=\frac{\partial}{\partial x}\left[(\nu+\nu_T)\frac{\partial w}{\partial x}\right]+$$

$$+\frac{\partial}{\partial y}\left[(\nu+\nu_T)\frac{\partial w}{\partial y}\right]+\frac{\partial}{\partial z}\left[(\nu+\nu_T)\frac{\partial w}{\partial z}\right]+\left(-\frac{1}{\rho}\frac{\partial p}{\partial z}+s'_w\right), \quad (4)$$

- energy equation:

$$\frac{\partial T}{\partial t}+\frac{\partial(uT)}{\partial x}+\frac{\partial(vT)}{\partial y}+\frac{\partial(wT)}{\partial z}=\frac{\partial}{\partial x}\left[\left(\frac{\nu}{\mathrm{Pr}}+\frac{\nu_T}{\mathrm{Pr}_T}\right)\frac{\partial T}{\partial x}\right]+$$

$$+\frac{\partial}{\partial y}\left[\left(\frac{\nu}{\mathrm{Pr}}+\frac{\nu_T}{\mathrm{Pr}_T}\right)\frac{\partial T}{\partial y}\right]+\frac{\partial}{\partial z}\left[\left(\frac{\nu}{\mathrm{Pr}}+\frac{\nu_T}{\mathrm{Pr}_T}\right)\frac{\partial T}{\partial z}\right]+S_T, \quad (5)$$

where s'_u, s'_v, s'_w and S_T are additional source terms in the momentum equation in the x-direction, y-direction, z-direction, and in the energy equation, respectively.

- turbulence equations:

– (k):

$$\frac{\partial k}{\partial t}+\frac{\partial(uk)}{\partial x}+\frac{\partial(vk)}{\partial y}+\frac{\partial(wk)}{\partial z}=\frac{\partial}{\partial x}\left[\left(\frac{\nu_T}{\sigma_k}\frac{\partial k}{\partial x}\right)\right]+$$

$$+\frac{\partial}{\partial y}\left[\left(\frac{\nu_T}{\sigma_k}\frac{\partial k}{\partial y}\right)\right]+\frac{\partial}{\partial z}\left[\left(\frac{\nu_T}{\sigma_k}\frac{\partial k}{\partial z}\right)\right]+(P-D), \quad (6)$$

– (ε):

$$\frac{\partial \varepsilon}{\partial t}+\frac{\partial(u\varepsilon)}{\partial x}+\frac{\partial(v\varepsilon)}{\partial y}+\frac{\partial(w\varepsilon)}{\partial z}=\frac{\partial}{\partial x}\left[\left(\frac{\nu_T}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x}\right)\right]+$$

$$+\frac{\partial}{\partial y}\left[\left(\frac{\nu_T}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial y}\right)\right]+\frac{\partial}{\partial z}\left[\left(\frac{\nu_T}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial z}\right)\right]+\left(\frac{\varepsilon}{k}(C_{\varepsilon 1}P-C_{\varepsilon 2}D)\right), \quad (7)$$

where

$$P=2\nu_T\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial w}{\partial z}\right)^2\right]+$$

$$+\nu_T\left[\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right)^2+\left(\frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\right)^2\right], \quad (8)$$

$$D=\varepsilon,$$

where $C_{\varepsilon 1}$ and $C_{\varepsilon 2}$ are the model constants with values of 1.44 and 1.92, respectively. The turbulence kinetic energy magnitude is indicated by (k), turbulence-specific dissipation rate is shown by (ε), and the eddy viscosity is indicated by (ν_T).

The k–ε turbulence model is selected for the CFD evaluation of the solar air heater because it delivers reliable, experimentally validated predictions of turbulent flow and heat transfer in ducts with surface roughness, while also offering good computational efficiency. Numerous numerical investigations and review studies on solar air heaters have shown that the re-normalisation group (RNG) variant of the k–ε model closely matches experimental findings for absorber plates equipped with roughness elements [23–25,27,28,31,32,38].

A constant heat flux of 1000 W/m^2 was applied to the absorber plate. The inlet flow conditions were defined for Reynolds numbers ranging from 3500 to 23 500 [13,15,16]. In this analysis, velocity inlet and pressure outlet conditions were imposed at the entrance and exit of the duct. The incoming air entered with a steady, uniform velocity and an initial temperature of 300 K. Since the flow is driven by pressure, the outlet pressure was maintained at 1.013×10^5 Pa. The boundary conditions used in the numerical simulations are summarised in Table 1.

Table 1. Computational domain boundary conditions [44].

Boundary domain	**Zone type**	**Boundary condition**
Inlet	Velocity inlet	The velocity for a fully developed flow is computed according to the associated Reynolds number. $w = V_{in}$, $u = v = 0$, $T = T_{in} = 300$ K
Outlet	Pressure outlet	$\left(\frac{\partial V}{\partial z}\right)_{z=L} = 0$, $\left(\frac{\partial T}{\partial z}\right)_{z=L} = 0$, At the outlet, the pressure is set to zero gauge, and the remaining flow parameters are estimated based on values from the domain.
Bottom duct wall	Wall	Adiabatic wall condition with a no-slip boundary. $u = v = w = 0$, $q = 0$ W/m^2
Spheres	Wall	Adiabatic wall condition with a no-slip boundary. $u = v = w = 0$, $q = 0$ W/m^2
Top heated wall	Wall	Uniform constant heat flux condition with a no-slip boundary $u = v = w = 0$, $q = 1000$ W/m^2

The interaction between the pressure and velocity fields in the numerical simulation was handled through the SIMPLE (semi-implicit method for pressure-linked equations) algorithm. This iterative procedure maintains mass conservation by repeatedly adjusting both pressure and velocity until a stable solution is reached. To solve the discretised momentum and energy equations, a second-order upwind scheme was employed, offering improved numerical precision by accounting for flow direction and reducing artificial diffusion.

The geometric setups and operating parameters applied throughout the analysis are compiled in Table 2, which presents the essential inputs utilised in the simulations.

Table 2. Geometrical and operating parameters [44].

Geometrical and operating parameters	**Symbol**	**Value/range**
Diameter of sphere (mm)	D	25
Relative roughness pitch	P/D	3, 6 and 9
Specific heat of air (J/(kg·K))	C_p	1008
Thermal conductivity of air (W/(m·K))	k_a	0.026
Dynamic viscosity of air (kg/(s·m))	μ	1.865×10^{-5}
Reynolds number	Re	3500–23 500

The Nusselt number (Nu) is widely used to evaluate the efficiency of convective heat transfer inside a SAH channel [45]:

$$\mathrm{Nu} = \frac{hD}{k_a}, \tag{9}$$

where the hydraulic diameter D for a rectangular duct is:

$$D = \frac{4A}{P}. \tag{10}$$

The heat transfer coefficient h is found using the relation:

$$h = \frac{q}{(T_s - T_b)}, \tag{11}$$

where q is the heat flux on the absorber plate, T_s is the average surface temperature of the heated plate, and T_b represents the bulk mean temperature of the airflow.

The friction factor (f) for fully developed turbulent flow through a roughened duct can be found as [46]:

$$f = 2\frac{\left[\frac{\Delta p}{l}\right]D}{\rho V^2}, \tag{12}$$

where Δp is the difference between the static pressure at the entrance and at the exit of the duct

$$\Delta p = p_{in} - p_{out}. \tag{13}$$

Lewis [47] introduced the thermo-hydraulic performance parameter (THPP), which states the following and compares the value of thermal gain to that of the frictional loss:

$$\mathrm{THPP} = (Nu_r/Nu_s)/\,(f_r/f_s)^{1/3}, \tag{14}$$

where Nu and f are the Nusselt number and friction factor, respectively. The parameters with suffix r and s refer to roughened and smooth configurations of SAHs.

3.1.2. Meshing and grid independence test

The different design configurations were first modelled in SolidWorks V23, after which the mesh generation and numerical simulations were carried out using ANSYS V21.0. The computational domain was discretised using an unstructured tetrahedral mesh, which allowed accurate representation of the complex geometric features of the spherical turbulators and the duct surfaces. While prismatic (wedge-type) layers were introduced near the walls (top and bottom absorber walls as well as around the spherical turbulators) to properly resolve the near-wall flow physics like flow separation, recirculation zones and boundary layer development, which are critical for predicting heat transfer enhancement and frictional losses in the solar air heater duct. A total of 5 inflation layers were applied near the solid surfaces with a first layer thickness of approximately 0.02 mm and a growth rate of 1.2, ensuring adequate resolution of the velocity and thermal boundary layers. The resulting dimensionless wall distance (y^+) values were maintained within the acceptable turbulence modelling range of 30–200 for the selected turbulence models. The inlet boundary conditions were specified using a turbulence intensity of 5% (0.05), which is commonly adopted for generic high-speed and internal turbulent flows. The turbulence length scale was defined as 0.07 times the hydraulic diameter of pipe to represent the characteristic size of turbulent eddies at the inlet.

These mesh structures were created through the ICEM CFD module in ANSYS, ensuring smooth element transitions and faithful representation of the geometric features. As shown in Fig. 3, inflation layers were applied along both the spherical turbulators (ST) and duct walls to effectively resolve near-wall flow characteristics during the analysis. The simulations were then performed using the Fluent solver integrated in ANSYS V21.0.

A grid independence analysis was performed to verify the dependability and accuracy of the computational setup. All configurations were independently subjected to grid-independence testing; however, for clarity and readability, only the inclined configuration is presented in the manuscript as a representative case. The inclined spherical turbulator configuration of the solar air heater (SAH) was evaluated under multiple mesh densities while maintaining a constant Reynolds number of 11 000. The goal was to determine how changes in mesh resolution influenced the predicted thermal behaviour. Table 3 reports the variation in Nusselt number as a function of the total cell count. The results showed that when the mesh consisted of 1 889 780 elements, the Nusselt number obtained with the k–ε turbulence model differed by only 0.95% from those predicted with finer grids. Since this difference is insignificant, additional refinement would not meaningfully improve the accuracy while significantly increasing computational time. Therefore, the mesh containing 1 889 780 cells was selected for all further simulations to achieve an optimal balance between computational cost and solution reliability.

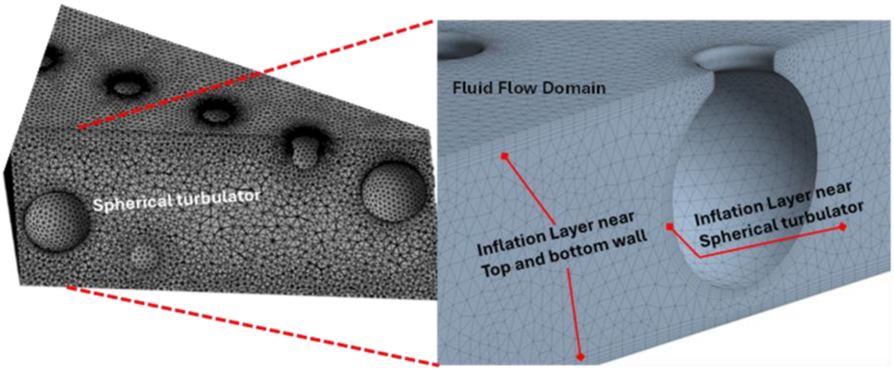


Fig. 3. Meshing of SAH fluid flow domain with zoomed view near spherical turbulators.

Table 3. Grid independence test.

No.	Number of cells	Nusselt number values	Deviation (%)
1	121 5439	96.45	–
2	129 4081	92.87	3.71
3	134 8725	88.29	4.93
4	144 2977	86.97	1.49
5	169 5313	85.74	1.41
6	188 9780	84.2	1.79
7	201 2542	83.4	0.95

3.1.3. Validation of the present numerical work

To ensure the accuracy and reliability of the numerical approach, a validation study was performed in which the results from the present SAH simulations were compared with established findings from earlier research. The validation involved modelling a smooth solar air heater duct without any roughness elements. The numerically obtained Nusselt number and friction factor were evaluated against well-known theoretical correlations for smooth ducts, specifically, the Gnielinski and Dittus-Boelter correlations for heat transfer, and the Blasius equation for friction factor estimation [45]. For Reynolds numbers below 10 000, the Gnielinski correlation was applied, while the Dittus-Boelter correlation was used for Reynolds numbers exceeding this range. The Nusselt number and friction factor predicted using standard k–ε, RNG k–ε, realisable k–ε, standard k–ω, and SST k–ω turbulence models were evaluated over a Reynolds number range of 4000–20 000 and compared with well-established empirical correlations (the Gnielinski and Dittus-Boelter correlations for the Nusselt number, and the modified Blasius equation for the friction factor). The comparison plots in Fig. 4 added in the revised manuscript show that the RNG k–ε model provides predictions that closely follow the empirical correlations while maintaining stable trends across the entire Reynolds number range. Other models either slightly overpredict or deviate from the correlations. Based on this comparative assessment, the RNG k–ε model was selected as it provides reliable agreement with established correlations for both heat transfer and flow friction characteristics in the present solar air heater configuration. The comparison showed excellent consistency, with average deviations of 3.1% for the friction factor and 4.7% for the Nusselt number.

The Nusselt number for a smooth duct has been calculated using the Dittus-Boelter equation for Re > 10 000 and the Gnielinski correlation for Re < 10 000, as expressed in Eqs. (15) and (16), respectively [39]:

$$\mathrm{Nu}_s = 0.023\,\mathrm{Re}^{0.8}\mathrm{Pr}^{0.4}, \tag{15}$$

$$\mathrm{Nu}_s = \frac{\left(\frac{f}{2}\right)(\mathrm{Re}-1000)\mathrm{Pr}}{1+12.7\left(\frac{f}{2}\right)^{\frac{1}{2}}\left(\mathrm{Pr}^{2/3}-1\right)}, \tag{16}$$

where $f = (1.58\cdot\ln(\mathrm{Re}) - 3.82)^{-2}$ and Pr is the Prandtl number.

The modified Blasius equation presents the friction factor in a smooth duct as [40]:

$$f_s = 0.085\,\mathrm{Re}^{-0.25}. \tag{17}$$

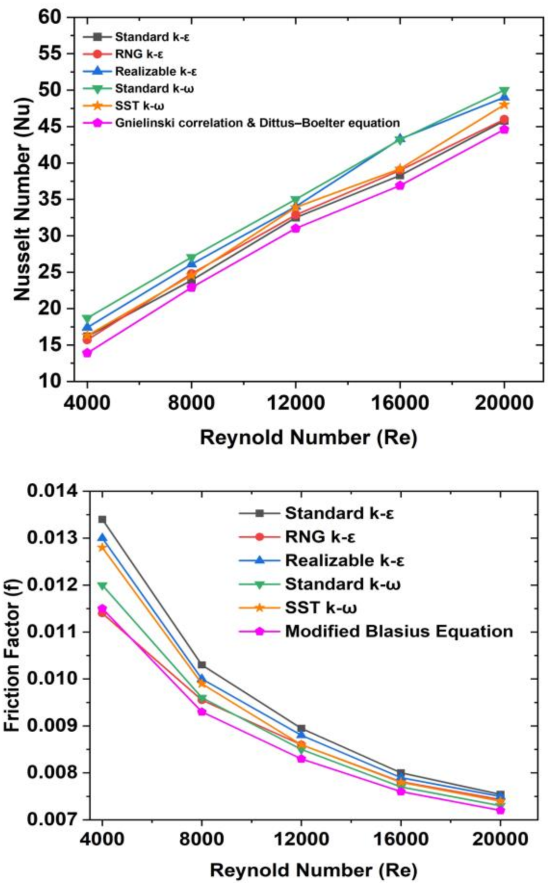


Fig. 4. Comparison of numerical results with correlations for a smooth duct.

An additional validation study was conducted using experimental data, in which the numerical simulation results of a solar air heater equipped with V-shaped ribs having a symmetrical gap were compared with the corresponding experimental findings reported in [48]. It can be observed from Fig. 5 that the Nusselt number increases consistently with the increasing Reynolds number for both approaches, indicating enhanced convective heat transfer at higher flow rates. The numerical predictions closely follow the experimental trend across the entire Reynolds number range (4000–20 000), demonstrating good agreement between the computational model and the experimental observations. The deviation between the two datasets appears relatively small, suggesting that the adopted numerical model and boundary conditions are capable of reliably predicting the heat transfer characteristics of the ribbed channel. However, minor discrepancies between experimental and numerical results may arise due to several possible sources of error. These include experimental uncertainties such as thermocouple measurement errors, heat losses to the surroundings, and non-uniform heating conditions. Additionally, numerical errors may result from mesh discretisation, turbulence model limitations, and simplified assumptions such as steady-state flow or uniform material properties. Variations in rib geometry during fabrication and slight differences in boundary conditions between the experiment and simulation may also contribute to the observed differences. Overall, the small deviation indicates acceptable validation of the numerical model for predicting the heat transfer behaviour of the investigated configuration.

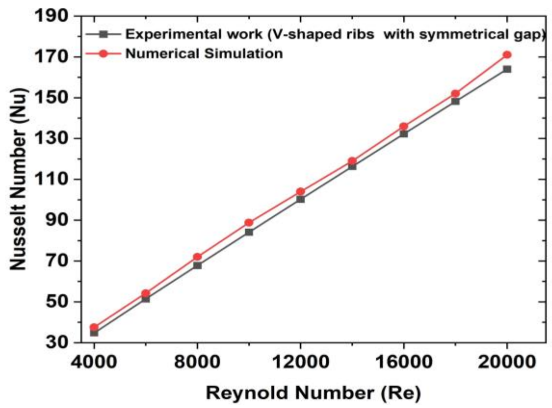


Fig. 5. Validation of the CFD model through comparison with experimental data for V-shaped rib configuration.

3.2. ML surrogate modelling

High-fidelity CFD provides reliable thermo-hydraulic performance predictions for solar air heaters (SAHs) equipped with spherical turbulators; however, repeated CFD evaluations become computationally expensive when configuration screening, sensitivity analyses, or optimisation are required. To enable rapid evaluation over the investigated design space, a supervised machine-learning (ML) surrogate model was developed to predict the Nusselt number (Nu) and friction factor (f) directly from the governing operating and geometric descriptors (the Reynolds number and turbulator arrangement, together with the dimensionless spacing/ratio parameter used in the numerical campaign). The surrogate is trained on the CFD-derived dataset spanning over the range of Re = 3500–23 500 and the three discrete levels of the dimensionless spacing/ratio parameter (3, 6 and 9), consistent with the simulation campaign.

For clarity, the ML workflow can be understood as four sequential stages: dataset assembly from CFD outputs, cleaning and feature construction, leakage-safe grouped model development, and final reliability assessment on a frozen holdout set. First, CFD cases were compiled into a tabular dataset in which each row corresponds to one simulated operating condition and geometry combination. Second, the raw descriptors were cleaned, standardised, and expanded into engineered features that reflect known thermo-fluid scaling behaviour. Third, the processed data were split by geometry identity so that similar cases from the same configuration-pitch pair did not appear in both training and evaluation subsets. Finally, the selected surrogate models were assessed using grouped cross-validation and a frozen holdout test with uncertainty estimation. A schematic overview of this workflow is shown in Fig. 6.

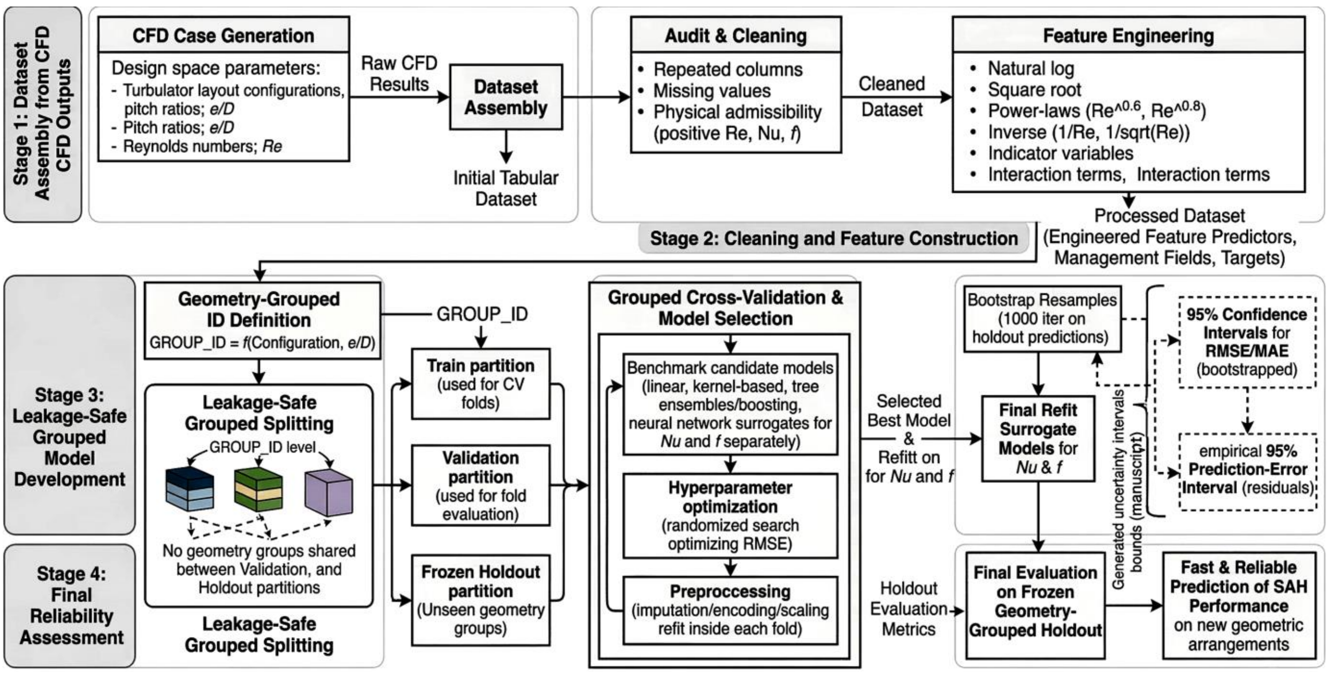


Fig. 6. Schematic overview of the CFD-derived surrogate-modelling workflow.

The ML workflow follows reporting for data-driven surrogates: (i) transparent dataset audit and physically consistent cleaning, (ii) leakage-safe train-test splitting using geometry-level grouping, and (iii) exploratory data analysis (EDA) to verify coverage of the operating envelope and to interpret dominant trends prior to model training, as shown in Fig. 7.

In contrast to many earlier CFD-ML studies that mainly report pointwise prediction accuracy, the present surrogate workflow is structured around geometry-aware generalisation and design screening. The methodological contribution here is the combination of grouped geometry-level splitting, explicit leakage control, frozen-holdout validation, uncertainty estimation, and post-prediction thermo-hydraulic ranking for spherical-turbulator configurations. This makes the surrogate framework more directly relevant for engineering selection tasks than a purely accuracy-oriented regression exercise.

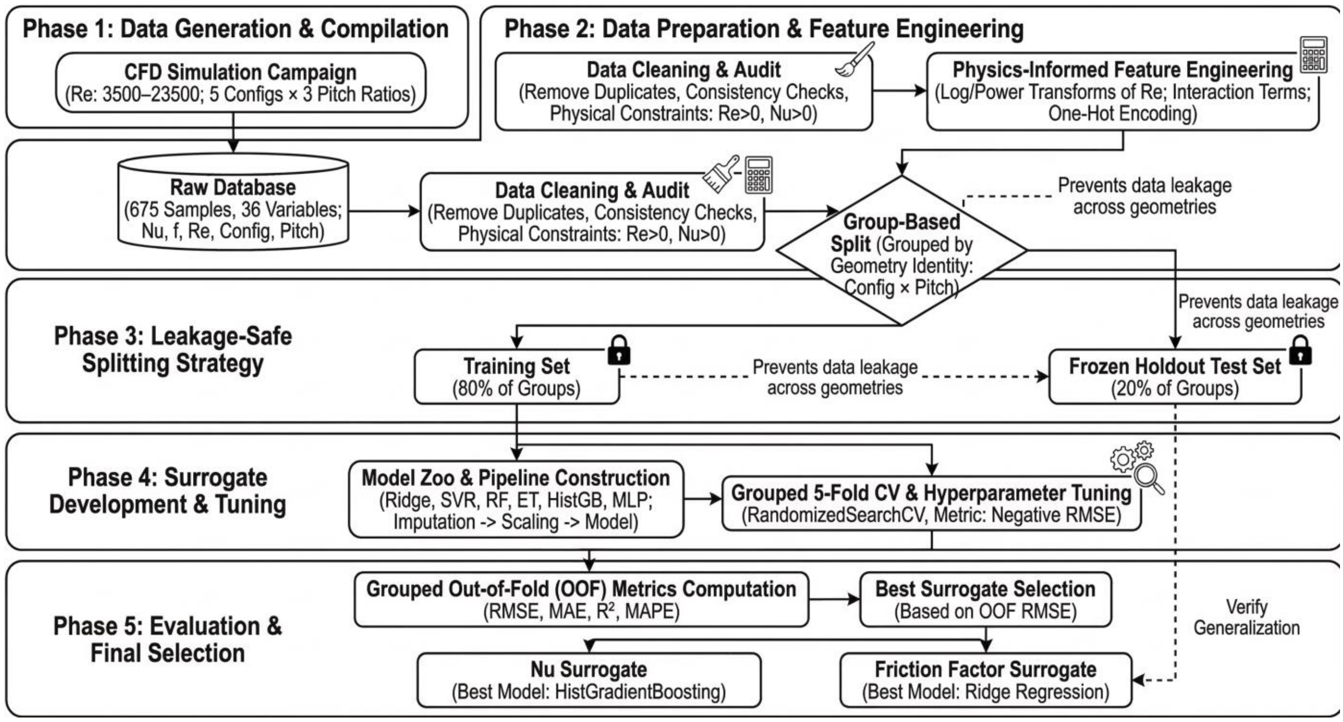


Fig. 7. Overall ML approach flowchart.

3.2.1. Dataset description

The compiled dataset contains 675 samples and 36 columns, including two supervised targets (Nu and friction factor), categorical descriptors of the turbulator arrangement (configuration, 5 levels: V-, M-, W-shaped, inclined, arc-shaped), and a dimensionless spacing/ratio parameter stored as *P_over_D* (three levels: 3, 6, 9). These five configurations correspond to the five spherical-turbulator arrangements investigated in the CFD study.

To improve learnability on tabular models, the raw physical inputs were augmented with physics-informed feature engineering, including logarithmic, inverse, square-root and power-law transforms of Reynolds number, together with interaction terms involving the spacing ratio and indicator variables for configuration class. The complete feature scheme used for ML, including raw variables, engineered variables, units, ranges and transformation definitions, is documented in Table 4. To avoid leakage, only pre-target operating and geometric descriptors were used as predictors; the response variables (Nu and friction factor), thermo-hydraulic performance metrics derived from them, and split-management fields were excluded from model inputs.

Table 4 distinguishes clearly between raw design inputs, engineered predictors, split-management fields, target variables, and derived thermo-hydraulic metrics. Variables such as Nu, friction factor and THPP were never supplied to the ML models as predictors. Likewise, GROUP_ID and SPLIT were used only for grouped evaluation and frozen holdout construction, not for training. This separation was enforced to eliminate inadvertent target or partition leakage.

In practical terms, preprocessing was performed in a simple sequence. The compiled CFD table was first checked for repeated columns, inconsistent names, and missing target values. After this audit, categorical labels for configuration were standardised so that each layout appeared under a single consistent name. Numerical inputs were then screened for physical admissibility, retaining only rows with positive Reynolds number, positive Nusselt number and positive friction factor. Next, the raw design descriptors were transformed into a richer predictor set using interpretable mathematical operations such as logarithm, inverse, square-root and power-law terms, together with configuration indicators and selected interaction terms. These engineered variables were created only from pre-target quantities and were intended to help the regression models represent known nonlinear thermo-fluid behaviour more effectively.

A conservative audit-and-cleaning protocol was applied to ensure that the surrogate is trained on physically meaningful and internally consistent samples. Because aggregated datasets may occasionally contain repeated fields (e.g., duplicate “Re” columns from concatenation), columns were standardised by trimming/normalising names, and then checked for duplication. Perfectly duplicated columns (value-by-value identical) were removed to prevent redundant signals and unstable parameter estimation in linear models. When multiple Reynolds-number columns existed, a single canonical Re was selected based on numeric type and completeness. Furthermore, since the task is supervised regression, any sample missing either target (Nu or friction factor) was excluded. In the final compiled dataset used here, the targets are fully populated (675/675 non-null), which avoids bias from imputation on the dependent variables. Samples violating basic physical admissibility were removed using non-aggressive guards: Re > 0, Nu > 0, friction factor > 0. These checks are lightweight but important: they prevent accidental inclusion of corrupted rows (e.g. sign errors) that can distort loss surfaces and inflate error metrics.

Table 4. Summary of variables used for ML surrogate modelling (targets, raw inputs and engineered features).

Variable	Type	Range / Levels	Transformation / Definition	Used as predictor	Leakage risk
Re	Raw input	3 500–23 500	Reynolds number from CFD case definition	Yes	No
***P_over_D* (or *P/D* as stored in dataset)**	Raw input	3, 6, 9	Dimensionless pitch/spacing ratio used in the design study	Yes	No
Configuration	Raw categorical input	V, M, W, Inclined, arc	Turbulator layout label	Yes	No
Config_V	Engineered categorical flag	0, 1	One-hot encoding of V-shaped layout	Yes	No
Config_M	Engineered categorical flag	0, 1	One-hot encoding of M-shaped layout	Yes	No
Config_W	Engineered categorical flag	0, 1	One-hot encoding of W-shaped layout	Yes	No
Config_inclined	Engineered categorical flag	0, 1	One-hot encoding of inclined layout	Yes	No
Config_ARC	Engineered categorical flag	0, 1	One-hot encoding of arc-shaped layout	Yes	No
ln(Re)	Engineered numeric	Computed	Natural logarithm of Re	Yes	No
sqrt(Re)	Engineered numeric	Computed	Square root of Re	Yes	No
$Re^{0.6}$	Engineered numeric	Computed	Power-law transform of Re	Yes	No
$Re^{0.8}$	Engineered numeric	Computed	Power-law transform of Re	Yes	No
1/Re	Engineered numeric	Computed	Inverse Reynolds number	Yes	No
1/sqrt(Re)	Engineered numeric	Computed	Inverse square-root Reynolds term, if used	Yes/No	No
ln(*P_over_D*)	Engineered numeric	ln(3), ln(6), ln(9)	Natural logarithm of pitch ratio, if used	Yes/No	No
1/*P_over_D*	Engineered numeric	1/9–1/3	Inverse pitch ratio, if used	Yes/No	No
sqrt(*P_over_D*)	Engineered numeric	Computed	Square-root pitch term, if used	Yes/No	No
Re × *P_over_D*	Engineered interaction	Computed	Interaction between flow regime and spacing ratio	Yes	No
ln(Re) × *P_over_D*	Engineered interaction	Computed	Interaction between transformed Re and spacing ratio	Yes	No
Re^0.8 × *P_over_D*	Engineered interaction	Computed	Power-law Reynolds–pitch interaction	Yes/No	No
Re bin_low	Engineered categorical flag	0, 1	Low-Reynolds operating bin	Yes/No	No
Re bin_mid	Engineered categorical flag	0, 1	Mid-Reynolds operating bin	Yes/No	No
Re bin_high	Engineered categorical flag	0, 1	High-Reynolds operating bin	Yes/No	No
GROUP_ID	Split-management field	15 groups	Geometry identity = Configuration × *P_over_D*	No	Yes
SPLIT	Split-management field	Train / Test	Dataset partition label	No	Yes
Nu	Target	35.7–177.8	CFD-derived Nusselt number	No	Yes
Friction factor	Target	Dataset-specific	CFD-derived friction factor	No	Yes
THPP	Derived performance metric	Dataset-specific	Calculated from Nu and FrictionFactor	No	Yes

Categorical labels were standardised (case/whitespace normalisation) so that the same configuration is not split across multiple string variants. This is essential for stable one-hot encoding and for geometry-level grouping. A key risk in surrogate modelling for engineered thermal systems is data leakage: if multiple samples are generated from the same geometry (same arrangement and spacing ratio) across many Reynolds numbers, a naive random split may place near-identical geometry patterns in both training and test sets, inflating apparent generalisation.

To address this, the dataset was partitioned using a geometry identity group:

$$\mathrm{GROUP\_ID} = f(\mathrm{Configuration}, e/D),$$

where e/D denotes the discrete spacing/ratio level used in the parametric campaign (stored as *P_over_D* in the dataset). This grouping produces 15 geometry groups (= 5 configurations × 3 spacing levels), aligning with the CFD design-of-experiments structure. From the full compiled table, only variables available before target evaluation were allowed to enter the predictor matrix. Specifically, the Reynolds number, spacing-ratio information, categorical configuration labels, and their engineered transforms were retained, whereas Nu, friction factor, THPP, SPLIT, and GROUP_ID were not used as predictors. This rule was adopted to prevent target leakage and to ensure that the surrogate learns only from quantities that would be known at design-screening time. A group shuffle split strategy was then used to create a frozen holdout test set (20% by groups), ensuring that the test set contains geometry groups unseen during training. This makes the reported test performance meaningful for design screening, because the surrogate is evaluated on genuinely new geometry identities rather than interpolating within the same group.

Interpretation and significance. The split-check histogram (Fig. 8) is an important diagnostic: it verifies that, even after enforcing group disjointness, the Reynolds-number envelope remains comparable across training and test partitions. This reduces the chance that performance differences are driven simply by different Re ranges rather than by geometry generalisation. In other words, the surrogate is forced to learn transferable relationships between the inputs and (Nu, *f*), rather than memorising a specific configuration's curve.

Figure 9 confirms that the dataset spans the complete Reynolds-number and friction factor range investigated in the CFD campaign. This coverage is crucial because Nu and friction factor in roughened ducts are strongly Reynolds-dependent; a surrogate trained on a narrower window would extrapolate unreliably outside its calibration range. From a modelling standpoint, the broad Re coverage motivates the inclusion of nonlinear Reynolds transforms (e.g. ln Re, $Re^{0.6}$, $Re^{0.8}$, 1/Re), which mirror classical convective scaling behaviour and help reduce the functional complexity that the regression models must learn.

As expected for forced convection in internal flows, Nu increases with Re across all configurations (Fig. 10), reflecting enhanced mixing and reduced thermal boundary-layer thickness at higher flow rates. The separation between configuration-specific point clouds highlights that the turbulator arrangement is not a minor categorical label but a first-order determinant of heat-transfer augmentation: at a fixed Re, different arrangements produce systematically different Nu levels. This validates two dataset-design decisions: (i) retaining configuration explicitly (categorical + indicators), and (ii) adding interaction terms combining Re with the spacing/ratio feature, because layout-induced turbulence generation interacts with the flow regime rather than acting as a simple additive offset.

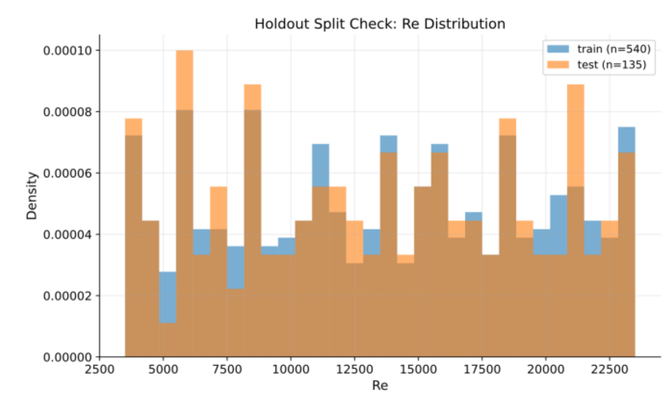


Fig. 8. Train-test split check showing similar Reynolds-number coverage in both splits (leakage-safe group split).

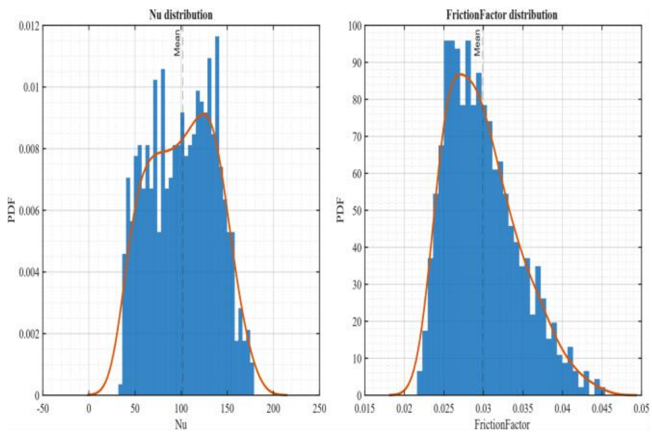


Fig. 9. Re number and friction factor distribution of the compiled dataset.

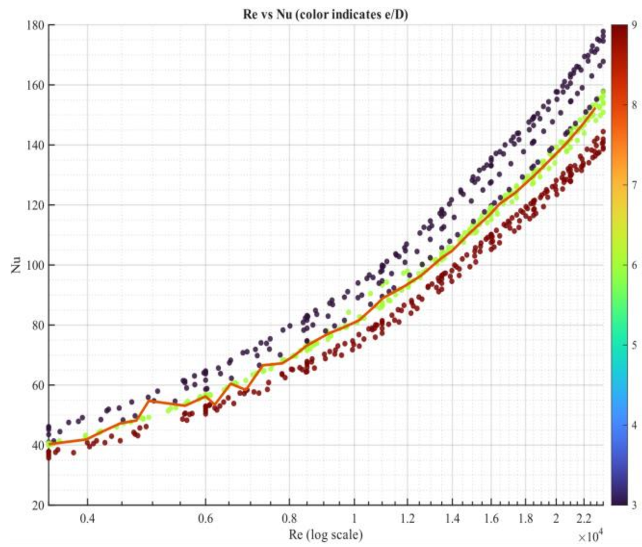


Fig. 10. Nusselt number versus Reynolds number for each spherical-turbulator configuration.

Figure 11 shows that friction factor decreases with increasing Re in the investigated turbulent range, consistent with internal-flow behaviour where inertial effects dominate and the relative contribution of viscous shear diminishes at higher Re. Importantly, the configuration-dependent offsets indicate that the same arrangement choices that increase turbulence and heat transfer also modify pressure losses. For surrogate modelling, this implies a multi-physics mapping: the model must capture both monotonic trends with Re and the geometry-driven shifts in the friction-Re relationship. This is a typical scenario where tree-based ensemble models and gradient boosting often perform well, because they can represent nonlinear interactions and categorical splits without requiring an explicit functional form.

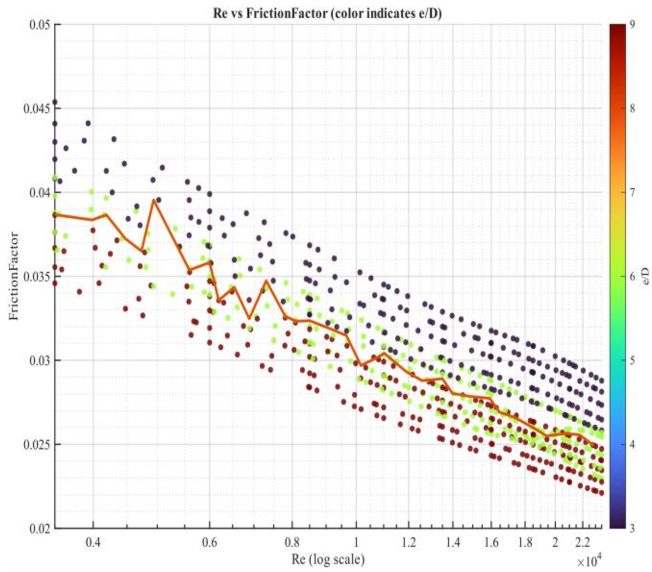


Fig. 11. Friction factor versus Reynolds number for each spherical-turbulator configuration.

The trade-off map (see Fig. 12) provides a compact thermo-hydraulic perspective: points with higher Nu are generally associated with higher *f*, reflecting the classical enhancement penalty where turbulence promoters improve convective transfer at the expense of pressure drop.

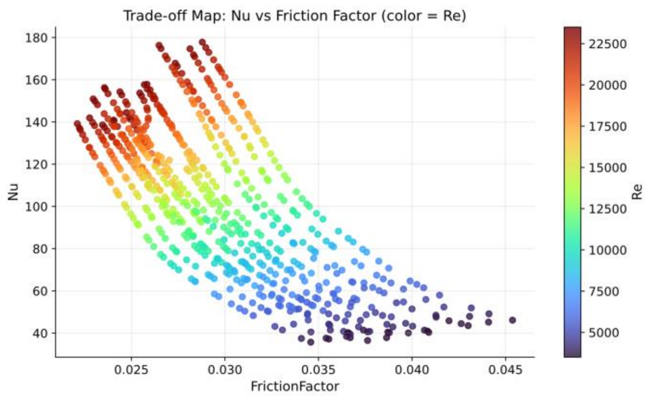


Fig. 12. Trade-off map between the Nusselt number and friction factor (colour-coded by Reynolds number).

The Reynolds-number colour gradient further clarifies the operating-regime dependence: as Re increases, Nu tends to rise while *f* tends to moderate, indicating that some high-Re operating points can occupy a relatively favourable region of high heat transfer with less severe friction penalties. For the ML workflow, this figure justifies treating Nu and friction factor as separate regression targets (rather than forcing a single composite metric during training), because design optimisation and THPP evaluation can then be performed post-prediction using physically interpretable objectives.

3.2.2. ML model development and evaluation

This subsection describes the development of data-driven surrogate models for rapid prediction of the two CFD-derived performance indicators, i.e. Nu and *f*, for the investigated SAH equipped with spherical turbulators. The surrogate was trained over the same design space explored in the numerical campaign, comprising five turbulator arrangements (V-, M-, W-shaped, inclined and arc-shaped layouts) and three discrete spacing levels (relative roughness pitch values of 3, 6 and 9), while keeping the spherical diameter constant to preserve geometric uniformity.

Two independent supervised regression models were formulated:

$$\hat{y} = \mathcal{M}(\mathbf{x}), y \in \{\mathrm{Nu}, f\}, \tag{18}$$

where $\mathbf{x}$ denotes the input feature vector containing operating and geometric descriptors. Separate models were trained for Nu and f (rather than a single multi-output regressor) to allow target-specific inductive bias and hyperparameter tuning, as the thermo-hydraulic response surfaces for heat transfer and pressure losses can exhibit different nonlinearities and noise characteristics.

All available non-target variables were considered as candidate predictors, excluding the data-management fields used for splitting and grouping (SPLIT and GROUP_ID). A complete variable dictionary, including units, ranges, transformation formulas and leakage status, is provided in Table 4. The predictors comprised: (i) primary physical inputs (e.g. Reynolds number and the spacing/ratio parameter *e*/*D*), (ii) configuration descriptors (categorical configuration label and its indicator representation), and (iii) physics-informed engineered features (logarithmic, power-law and interaction terms), which aim to linearise known scaling relationships and represent coupled effects between flow regime and spacing.

To prevent leakage during cross-validation and to ensure consistent transformations across folds, preprocessing was implemented within the training pipeline and re-fitted inside each cross-validation split. The pipeline consisted of: (i) median imputation for missing values, followed by optional standardisation (zero mean, unit variance) for scale-sensitive models, (ii) most-frequent imputation, followed by one-hot encoding with “unknown category” handling to ensure robustness if an unseen configuration label appears at inference time (e.g. due to label normalisation or future extensions).

The preprocessing pipeline can be summarised as follows. For numerical variables, missing values were replaced with the median computed from the training fold only; scale-sensitive models then received standardised inputs with zero mean and unit variance, again fitted only on the training fold. For categorical variables, missing labels were replaced with the most frequent category from the training fold, followed by one-hot encoding to convert configuration labels into binary indicators. The preprocessing steps were embedded inside the model pipeline so that the validation fold and holdout set were never used to estimate imputation values, scaling factors or category mappings. This procedure ensures that model evaluation reflects true generalisation rather than information leakage from preprocessing.

This design ensures that all preprocessing parameters (imputation statistics, scaling factors and categorical mappings) are learned exclusively from the training portion of each fold. Grouped 5-fold cross-validation (GroupKFold, n_splits = 5) using GROUP_ID; randomised hyperparameter search with 30 iterations per model (RandomizedSearchCV, n_iter = 30); primary scoring = negative *RMSE*; random seed = 42; preprocessing (imputation/encoding/scaling) refit inside each fold (see Table 5).

Table 5. Candidate ML models, preprocessing choice, grouped CV protocol and hyperparameter search space used for surrogate development (Nu and friction factor are modelled separately).

Model family	Preprocessing (numeric)	Fixed settings	Hyperparameters tuned (randomised search space)
Ridge regression	Standardisation ON	–	α ∈ [10^{-4}, 10^{3}] (60 log-spaced values)
SVR (RBF kernel)	Standardisation ON	kernel = RBF	C ∈ [10^{-1}, 10^{3}] (40 log-spaced); γ ∈ [10^{-4}, 10^{0}] (40 log-spaced); ε ∈ [10^{-4}, 10^{-1}] (25 log-spaced)
Random forest regressor	Standardisation OFF	n_estimators = 600	max_depth ∈ {none, 10, 16, 24, 32}; min_samples_split ∈ {2, 4, 8, 12}; min_samples_leaf ∈ {1, 2, 4, 8}; max_features ∈ {"sqrt", 0.5, 0.7, 1.0}
Extra trees regressor	Standardisation OFF	n_estimators = 600	max_depth ∈ {none, 8, 12, 18, 26}; min_samples_split ∈ {2, 4, 8, 12}; min_samples_leaf ∈ {1, 2, 4, 8}; max_features ∈ {"sqrt", 0.5, 0.7, 1.0}
Histogram gradient boosting regressor	Standardisation OFF	–	learning_rate ∈ [$10^{-2.5}$, $10^{-0.2}$] (40 log-spaced; ~0.00316–0.631); max_depth ∈ {none, 4, 6, 8, 10}; max_leaf_nodes ∈ {15, 31, 63, 127}; min_samples_leaf ∈ {10, 20, 30, 50}; l2_regularization ∈ [10^{-6}, 10^{1}] (30 log-spaced; 10^{-6} to 10)
MLP regressor (optional benchmark)	Standardisation ON	ReLU activations; MSE loss; AdamW optimiser; batch size = 256	Architecture: [256, 128, 64] hidden units; dropout = 0.10; learning rate = 1×10^{-3}; epochs ≤ 250 with early stopping (patience = 30)

Median imputation was applied to numerical features; most-frequent imputation and one-hot encoding were applied to categorical features; standardisation was enabled only for scale-sensitive models. A compact but diverse model set ("model zoo") was selected to balance interpretability, nonlinear capacity and sample efficiency on tabular data:

1. **Ridge regression (L2-regularised linear model):** a strong baseline for engineered features; regularisation mitigates multicollinearity among transformed predictors,
2. **Support vector regression (SVR) with RBF kernel:** a nonlinear margin-based learner that can fit smooth response surfaces; applied with standardised inputs,
3. **Random forest regression:** bagged decision trees providing robust nonlinear modelling and limited sensitivity to monotonic transforms,
4. **Extra trees regression:** randomised tree ensembles that typically yield strong accuracy and reduced variance on tabular datasets,
5. **Histogram-based gradient boosting regression:** a boosting approach well-suited to capturing complex interactions with efficient training,
6. **Optional multi-layer perceptron (MLP):** a small feed-forward neural network trained with early stopping; used as an additional nonlinear benchmark.

Ridge (α); SVR (C, γ, ϵ); random forest / extra trees (depth, minimum samples, max features); gradient boosting (learning rate, depth/leaf nodes, minimum leaf samples, L2 regularisation); MLP (hidden sizes, dropout, learning rate, epochs with early stopping).

Model development was performed strictly on the training split, while the holdout test split remained frozen for final assessment (reported in Results). Cross-validation was conducted using GroupKFold based on GROUP_ID, i.e., the geometry identity formed by configuration and spacing level. This grouped strategy is essential for surrogate modelling of parametric CFD datasets because multiple Reynolds-number samples from the same geometry are highly correlated; grouping forces validation folds to contain geometry identities not present in the corresponding training fold, producing a more realistic estimate of generalisation to unseen configurations within the investigated design space.

The grouped validation strategy can be interpreted in a simple way: all CFD samples belonging to the same geometry family, defined by a specific configuration and pitch ratio, were kept together during splitting. Thus, when one geometry group was placed in a validation fold or in the frozen holdout set, the model had not seen any samples from that same group during training. This is stricter than random splitting and is more appropriate for engineering screening, because the real task is to predict performance for a new configuration-pitch combination rather than for a repeated point from a geometry already seen during training. A schematic of this grouped split strategy is shown in Fig. 13. It illustrates how grouped splitting was enforced at the geometry level so that all Reynolds-number samples from one configuration-pitch pair remained together during validation and frozen-holdout testing.

Unless otherwise stated, five folds were used. A fixed random seed was set for reproducibility of randomised procedures (hyperparameter sampling and any model-internal randomness). For each candidate model and for each target (Nu and f), hyperparameters were optimised using randomised search under grouped cross-validation. The primary tuning objective was minimisation of root mean squared error (*RMSE*), chosen because it penalises large errors and is commonly used for regression surrogates intended for engineering screening. After the search, the best hyperparameter setting was refit on the full training portion.

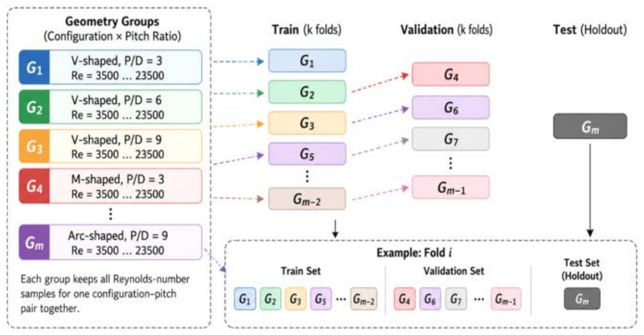


Fig. 13. Schematic of the grouped train-validation-test strategy.

To enable reproducible and fair comparison across models, the same grouped fold structure and the same metric definition were used during tuning for all candidates. Although final numerical results are presented in the Results section, evaluation methodology is defined here. For each tuned model, out-of-fold (OOF) predictions were generated under grouped cross-validation. OOF prediction provides a nearly unbiased estimate of model performance on unseen groups because each training sample is predicted by a model that did not observe its geometry group during fitting.

Four complementary metrics were computed:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |, \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{21}$$

$$MAPE(\%) = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{max(|y_i|, \varepsilon)} \right|. \tag{22}$$

Here, ε is a small constant used to avoid numerical instability when y_i is very small. *RMSE* served as the primary model-selection criterion, while *MAE*, R^2 and *MAPE* were used to provide complementary views of error magnitude, explained variance and relative error.

Metrics were computed from out-of-fold predictions generated under GroupKFold (5 folds), ensuring that each prediction is produced by a model trained without access to the corresponding geometry group (GROUP_ID). This provides a leakage-safe estimate of generalisation prior to frozen-test evaluation. *RMSE*, *MAE*, R^2, and *MAPE* were computed from grouped out-of-fold predictions; *RMSE* was used for model selection.

To complement the classical ML models, an optional MLP regressor was trained using the same grouped cross-validation protocol and the same preprocessing (imputation, one-hot encoding and standardisation). The network employed multiple fully connected layers with ReLU activations and dropout regularisation. Training used the mean squared error loss and an adaptive optimiser (AdamW) with early stopping based on validation loss to reduce overfitting risk on the moderately sized dataset. This benchmark was treated as an additional candidate and compared on the same OOF metrics.

All experiments were executed in a controlled Python environment, recording software versions and hardware availability (CPU/GPU) for transparency. The best-performing pipeline for each target was archived for later inference and for frozen-test evaluation, ensuring that the exact preprocessing + model mapping can be reproduced without manual reconstruction.

After model selection under grouped cross-validation, the best-performing surrogate for each target was refit on the full training partition and evaluated once on the frozen geometry-grouped holdout test set. To quantify predictive reliability, uncertainty in the holdout error metrics was estimated using nonparametric bootstrapping of the holdout predictions with 1000 resamples. For each target, 95% confidence intervals were computed for *RMSE* and *MAE* from the bootstrap distributions. In addition, an empirical 95% prediction-error interval was estimated from the holdout residuals to indicate the typical range within which new predictions may deviate from the CFD reference values inside the investigated design space.

4. Results and discussion

4.1. ML prediction results

The trained ML surrogates were assessed using leakage-safe, grouped out-of-fold (OOF) predictions, where geometry groups were defined by configuration and pitch/relative spacing (GROUP_ID), consistent with the operating envelope of the numerical dataset (Re = 3500–23 500; pitch ratio 3, 6 and 9). This grouped evaluation is intentionally stricter than point-wise random splits because it tests whether the model generalises to unseen geometry groups rather than merely interpolating among near-duplicate samples within the same configuration family.

In addition to grouped out-of-fold evaluation, the final selected models were assessed on a frozen geometry-grouped holdout test set containing configuration-pitch groups not used during training. On this stricter test, the hist gradient boosting model for Nu achieved an *RMSE* of 3.45, *MAE* of 2.48, R^2 of 0.988, and *MAPE* of 2.55%, with a bootstrapped 95% *RMSE* interval of 3.10–3.85. For friction factor, the ridge model achieved an *RMSE* of 0.000495, *MAE* of 0.000342, R^2 of 0.986, and *MAPE* of 1.18%, with a bootstrapped 95% *RMSE* interval of 0.000430–0.000570. The corresponding empirical 95% prediction-error intervals were [−6.5, 7.1] for Nu and [−0.00092, 0.00098] for friction factor, indicating stable predictive behaviour on previously unseen geometry groups.

The documented feature set also supports interpretability of the reported accuracy. Because the predictor matrix contains only pre-target operating and geometric variables and their transparent mathematical transforms, the reported performance cannot be attributed to target-derived shortcuts. This strengthens confidence that the surrogates are learning transferable thermo-fluid relationships rather than exploiting hidden leakage.

Table 6 summarises the OOF accuracy for predicting Nu and *f*. Two key observations emerge: Firstly, the Nu prediction is best captured by a non-linear boosting model. The hist gradient boosting model (HistGB) achieved the best Nu accuracy (*RMSE* = 3.299; R^2 = 0.9908, *MAPE* = 2.44%). Compared with the closest baseline (ridge regression, *RMSE* = 3.928), HistGB reduces *RMSE* by ~16%. This indicates that even after extensive feature engineering (e.g. log and power transforms of Re and interaction terms with pitch), Nu retains non-linear interactions across the Reynolds number and configuration that are well represented by boosting ensembles. Secondly, the friction factor is highly predictable with a simple linear model. The ridge regression provided the lowest error for friction factor (*RMSE* = 0.000470; R^2 = 0.9886, *MAPE* = 1.12%), outperforming the next-best model (HistGB, *RMSE* = 0.000740) by a ~36% *RMSE* reduction. This suggests that within the present design space, *f* behaves as a smooth and near-linear function in the engineered feature space (e.g. inverse/logarithmic transforms of Re and discrete pitch levels), consistent with the fact that friction correlations in internal flows often become close to linear after appropriate transformations.

Table 6. Grouped OOF performance of ML surrogates for predicting Nu and friction factor (*f*).

(a) Target: Nu

Model	OOF *RMSE*	OOF *MAE*	OOF R^2	OOF *MAPE* (%)
HistGB	3.2990	2.3348	0.9908	2.4414
Ridge	3.9281	2.9664	0.9870	3.3052
ExtraTrees	4.0612	2.9025	0.9861	2.8326
SVR_RBF	4.1796	3.5516	0.9853	3.7464
RandomForest	4.2551	2.9797	0.9847	2.9317
TorchMLP_CPU	13.7627	11.5722	0.6773	12.2304

(b) Target: friction factor (*f*)

Model	OOF *RMSE*	OOF *MAE*	OOF R^2	OOF *MAPE* (%)
Ridge	0.000470	0.000325	0.9886	1.1233
HistGB	0.000740	0.000596	0.9717	2.0201
SVR_RBF	0.000744	0.000564	0.9714	1.9412
ExtraTrees	0.001177	0.001040	0.9284	3.5082
RandomForest	0.001255	0.001090	0.9186	3.6909
TorchMLP_CPU	0.002161	0.001710	0.6393	5.8947

Importantly, the CFD model used to generate the underlying dataset was previously validated against correlations/benchmarks, with reported deviations of the order of ~7–8% for Nu and friction factor in comparable validation tests. The surrogate modelling errors reported here therefore add only a small incremental uncertainty on top of the validated CFD responses, enabling fast screening/optimisation while preserving CFD-level fidelity. These results show that the contribution of the present study is not limited to achieving high regression accuracy. The key improvement over a conventional CFD-ML workflow is that the trained surrogates remain reliable under geometry-grouped cross-validation and frozen-holdout testing, while also supporting direct thermo-hydraulic ranking of candidate configurations. This combination of predictive fidelity, leakage-safe validation, uncertainty reporting, and design selection provides a more complete engineering workflow than studies that stop at surrogate accuracy alone. Based on the CFD-derived thermo-hydraulic performance trends, the configuration ranking by THPP showed that the W-shaped arrangement performed best, followed by the V-shaped, M-shaped, arc-shaped, and inclined configurations under the investigated operating range, with the top-ranked case occurring at P/D = 6 and Re = 8000.

4.2. Best surrogate for Nu: hist gradient boosting

The best-performing Nu surrogate (HistGB) reaches $R^2 \approx 0.991$, indicating that the model explains nearly all variance in Nu within the grouped validation regime. Given the Nu range in the dataset (≈35.7–177.8), an *RMSE* of 3.30 corresponds to only a small fraction of the operating span, supporting use as a high-fidelity surrogate for design-space exploration.

The OOF parity plot for HistGB (Fig. 14) provides a direct visualisation of calibration quality over the full Nu range, while the residual scatter and residual histogram (Fig. 15) diagnose systematic bias and error symmetry. Together, these diagnostics are essential for surrogate use in optimisation: unbiased residuals and near-identity parity behaviour imply that the model can be used not only for ranking candidate designs but also for estimating absolute performance metrics with minimal distortion.

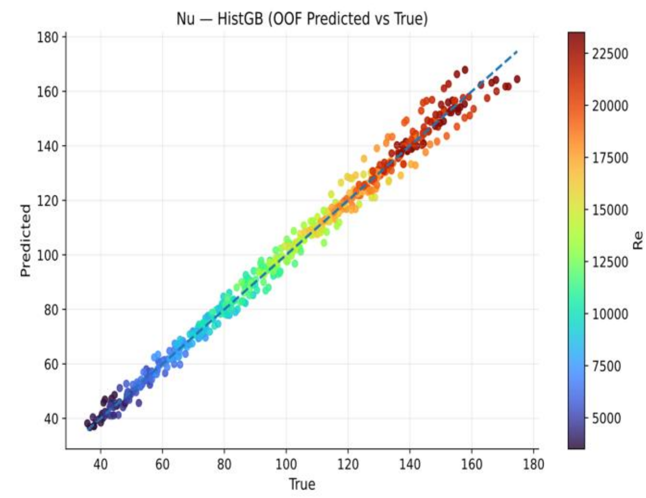
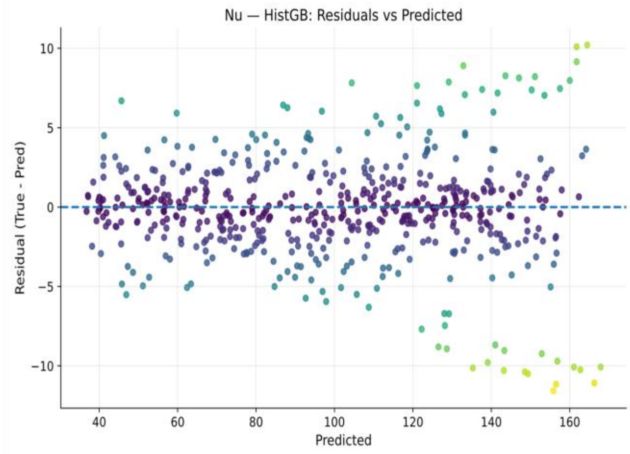


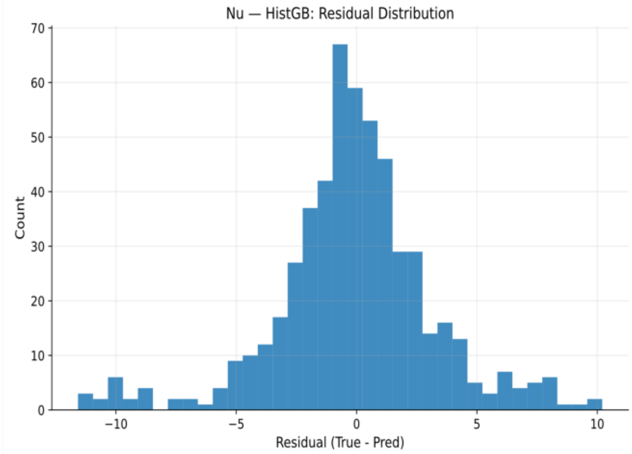
Fig. 14. OOF predicted vs. true Nu for HistGB (colour-coded by Reynolds number).

That boosting domination for Nu is physically plausible: heat transfer enhancement in roughened SAHs is governed by coupled, non-linear effects of turbulence augmentation, reattachment and secondary flow structures that vary with Reynolds number and turbulator arrangement. Even when Reynolds number is transformed (log, power laws) and combined with pitch-related interaction features, the mapping to Nu remains non-additive, making tree-based boosting an appropriate function class.

The frozen-holdout results confirm that the strong grouped cross-validation performance is retained when the model is evaluated on unseen geometry groups. The narrow bootstrap interval around holdout *RMSE* indicates that the reported Nu accuracy is not driven by a favourable single split. This supports the use of the Nu surrogate for rapid design screening within the investigated Reynolds-number and pitch-ratio ranges.



(a)



(b)

Fig. 15. OOF residual diagnostics for Nu–HistGB: (a) residual vs. predicted and (b) residual distribution

4.3. Best surrogate for friction factor: ridge regression

For friction factor, ridge regression yields the strongest performance (*MAPE* ≈ 1.12%; $R^2 \approx 0.989$). This result is particularly valuable because the friction factor directly drives pumping power and thus the hydraulic penalty that must be traded off against heat transfer enhancement in SAH design.

Within the present operating envelope, the friction factor is typically dominated by smooth trends with respect to the Reynolds number and discrete roughness/pitch levels. After feature engineering (e.g. 1/Re, ln(Re), $\sqrt{\text{Re}}$ and pitch-related transforms), these trends become approximately linear in the engineered space. Ridge regression then provides (i) strong generalisation under grouped splitting and (ii) stable estimates that are less sensitive to small dataset idiosyncrasies than more flexible models. This stability is advantageous when the surrogate will be embedded in optimisation loops.

Thermo-hydraulic performance metrics explicitly couple Nu and friction factor; for example, the widely used THPP formulation compares thermal gain against frictional loss using Nu and f. Therefore, high-accuracy prediction of f (along with Nu) is not merely a statistical achievement – it directly improves the reliability of multi-objective optimisation (maximise heat transfer while limiting pressure drop) and reduces the risk of selecting designs that appear optimal due to surrogate bias.

The frozen-holdout error and its bootstrap confidence interval further show that the friction-factor surrogate remains stable on unseen geometry groups. This reliability is important because even small bias in friction-factor prediction can distort thermo-hydraulic ranking. The holdout uncertainty bounds therefore strengthen confidence in using the surrogate for THPP-based screening and trade-off analysis.

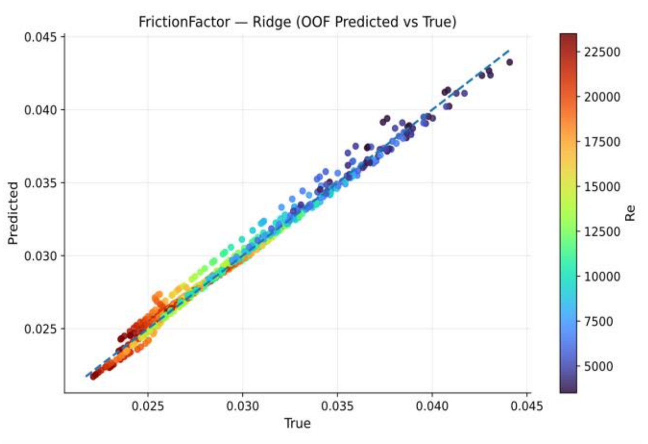
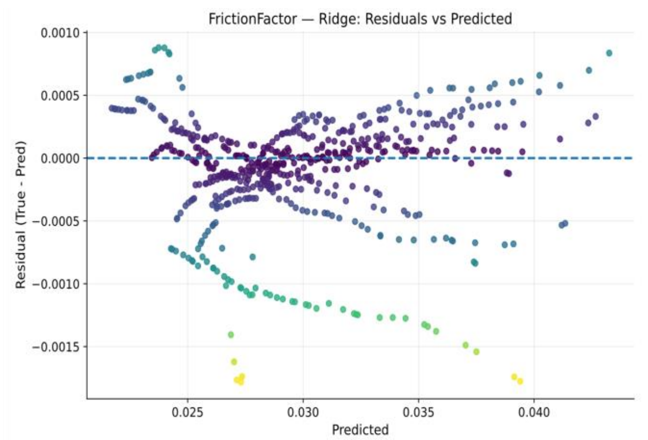
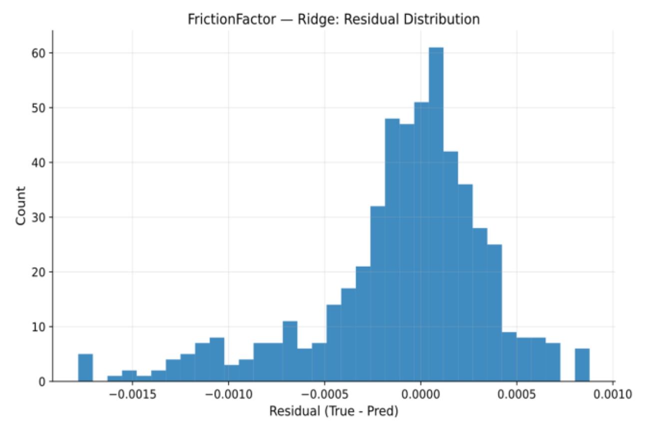


Fig. 16. OOF predicted vs. true friction factor for ridge regression (colour-coded by Reynolds number).



(a)



(b)

Fig. 17. OOF residual diagnostics for friction factor – ridge regression: (a) residual vs. predicted and (b) residual distribution.

4.4. Discussion of underperforming deep MLP baseline

The torch MLP benchmark underperformed substantially for both Nu and friction factor, despite using standardised numeric inputs and early stopping. In small-to-moderate tabular datasets with heterogeneous engineered features and discrete categorical structure (configuration families and pitch levels), deep MLPs often require careful architecture and regularisation tuning (and sometimes embedding strategies for categorical inputs) to match the inductive bias of tree ensembles. In contrast, boosting and linear models can be more data-efficient and robust for this class of structured regression problems, which is consistent with the performance ranking observed in Table 6.

4.5. Practical implications for SAH design and optimisation

From a domain perspective, these surrogates enable near-instant evaluation of Nu and f across the validated operating envelope (Re = 3500–23 500; pitch ratio 3–9), replacing repeated CFD calls during design screening. Because both Nu and f are predicted with high accuracy under leakage-safe grouped testing, the surrogates are suitable for:

- rapid parametric sweeps across Re and turbulator arrangements without rerunning CFD,
- multi-objective optimisation where Nu improvement must be balanced against friction penalties via THPP-style metrics,
- sensitivity studies to identify which geometric and operating factors most strongly shift the heat transfer-pressure drop trade-off.

To convert the thermo-hydraulic analysis into a design recommendation, the investigated spherical-turbulator layouts were ranked using the thermo-hydraulic performance parameter (THPP). Among all tested cases, the highest THPP was obtained for the W-shaped arrangement at $P/D = 6$ and Re = 8000, with a THPP of 2.14. This case provided the most favourable balance between heat-transfer enhancement and friction penalty. In contrast, the V-shaped configuration produced the highest heat-transfer augmentation, but its larger friction loss reduced the overall thermo-hydraulic advantage. Likewise, the inclined configuration showed lower pressure drop but did not achieve the strongest overall performance because of its lower Nusselt-number gain.

Across the tested operating range, the W-shaped configuration remained the most favourable choice for thermo-hydraulic design, particularly in the range of Re = 3500–12 000 and at $P/D = 6$. Therefore, for practical SAH implementation, the W-shaped configuration with $P/D = 6$ is recommended when the objective is to maximise the overall thermo-hydraulic performance rather than heat transfer alone.

Furthermore, from a methodological standpoint, the present study extends existing CFD-ML work by linking surrogate modelling directly to design decision-making. Rather than treating the surrogate as only a fast predictor of Nu and friction factor, the framework uses validated and uncertainty-supported predictions to rank configuration performance and identify the most favourable spherical-turbulator arrangement. This design-oriented use of the surrogate is a central contribution of the study.

5. Conclusions

This study developed and validated a combined CFD-ML framework to evaluate and rapidly predict the thermo-hydraulic behaviour of a single-pass solar air heater (SAH) equipped with spherical turbulators arranged in five patterns (V-, M-, W-shaped, inclined, and arc-shaped) over three pitch ratios (P/D = 3, 6, 9) and a Reynolds-number range of 3500–23 500. The principal conclusions are:

1. The CFD formulation and numerical settings yielded predictions that agreed well with established smooth-duct heat-transfer and friction correlations, and also showed acceptable deviation relative to benchmark roughened-duct results. This validation supports the use of the CFD database as a high-fidelity reference for surrogate training and configuration screening.
2. The compiled dataset captures physically consistent behaviour: Nu increases with the Reynolds number, while the friction factor decreases with the Reynolds number, and both responses exhibit clear dependence on turbulator arrangement and pitch ratio. These configuration-dependent offsets confirm that layout is a first-order driver of SAH performance.
3. By using geometry-aware grouping (configuration × pitch) and grouped cross-validation, the ML workflow explicitly targets generalisation across geometry identities, not merely interpolation among near-duplicate samples. This provides a more realistic estimate of surrogate robustness for design-space exploration.
4. Thermo-hydraulic ranking based on THPP identified the W-shaped configuration as the best overall spherical-turbulator layout within the investigated design space. The most favourable case occurred at P/D = 6 and Re = 8000, where the configuration achieved the best balance between heat-transfer enhancement and friction penalty.
5. Although the V-shaped configuration delivered strong heat-transfer enhancement, its higher friction loss reduced its overall thermo-hydraulic benefit. Similarly, the inclined configuration offered lower hydraulic resistance but a weaker overall thermal advantage. This confirms that configuration selection should be based on THPP rather than Nu or friction factor alone.
6. For practical SAH design within the investigated Reynolds-number range of 3500–23 500 and pitch ratios of 3, 6 and 9, the W-shaped configuration with P/D = 6 is the recommended design choice for overall performance, while the trained surrogates can be used for rapid screening of similar operating conditions without repeated CFD simulations.

The present models are valid within the investigated configuration set, pitch ratios and Reynolds-number range. The added frozen holdout-test results and bootstrap-based uncertainty bounds confirm that the selected surrogates retain strong predictive reliability on unseen geometry groups within this design space. Future work should expand the database to additional sphere diameters and geometric degrees of freedom and integrate the trained surrogates with formal multi-objective optimisers to generate robust design maps for practical SAH deployment.

Acknowledgements

The research has been duly facilitated by the Department of Mechanical Engineering Graphic Era Deemed to be University, Dehradun, Uttarakhand. The authors sincerely acknowledge the support of K.R. Mangalam University's Central Instrumentation Facility in enabling and assisting with this research.

Data and code availability

The CFD-derived dataset, feature-definition sheet and the Python scripts used for preprocessing, grouped cross-validation, frozen-holdout testing and surrogate-model training will be made available in a public repository upon acceptance of the manuscript. The repository will include the final tabular dataset, variable dictionary, train-test grouping protocol and scripts required to reproduce the reported results.

References

[1] Uniyal, A., Prajapati, Y.K., Ranakoti, L., Bhandari, P., Singh, T., Gangil, B., Sharma, S., Upadhyay, V.V., & Eldin, S.M. (2022). Recent advancements in evacuated tube solar water heaters: A critical review of the integration of phase change materials and nanofluids with ETCs. *Energies*, 15, 8999. doi: 10.3390/en15238999

[2] Kumar, A., Karn, A., McGregor, C., & Singh, V.P. (2025). A critical analysis of paradigm shifts from domestic solar cooking to institutional solar cooking technologies. *Renewable and Sustainable Energy Reviews*, 218, 115840. doi: 10.1016/j.rser.2025.115840

[3] Al-Yasiri, Q., Szabó, M., & Arıcı, M. (2022). A review on solar-powered cooling and air-conditioning systems for building applications. *Energy Reports*, 8, 2888–2907. doi: 10.1016/j.egyr.2022.01.172

[4] Negi, A., Ranakoti, L., Verma, R.P., Kumar, V., Bhandari, P., Khargotra, R., & Singh, T. (2025). Enhancing solar still productivity with organic phase change materials: A literature review. *Energy Conversion and Management: X*, 26, 100984. doi: 10.1016/j.ecmx.2025.100984

[5] Bisht, V.S., Patil, A K., & Gupta, A. (2018). Review and performance evaluation of roughened solar air heaters. *Renewable and Sustainable Energy Reviews*, 81, 954–977. doi: 10.1016/j.rser.2017.08.036

[6] Saini, R.P., & Singal, S.K. (2007). A review on roughness geometry used in solar air heaters. *Solar Energy*, 81(11), 1340–1350. doi: 10.1016/j.solener.2007.01.017

[7] Chamoli, S., Thakur, N.S., & Saini, J.S. (2012). A review of turbulence promoters used in solar thermal systems. *Renewable and Sustainable Energy Reviews*, 16, 3154–3175. doi: 10.1016/j.rser.2012.01.021

[8] Gawande, V.B., Dhoble, A.S., Zodpe, D.B., & Chamoli, S. (2016). A review of CFD methodology used in literature for predicting the thermo-hydraulic performance of a roughened solar air heater. *Renewable and Sustainable Energy Reviews*, 54, 550–605. doi: 10.1016/j.rser.2015.10.025

[9] Chaube, A., Sahoo, P.K., & Solanki, S.C. (2006). Analysis of heat transfer augmentation and flow characteristics due to rib rough-

ness over absorber plate of a solar air heater. *Renewable Energy*, 31, 317–331. doi: 10.1016/j.renene.2005.01.012

[10] Kumar, S., & Saini, R.P. (2009). CFD based performance analysis of a solar air heater duct provided with artificial roughness. *Renewable Energy*, 34, 1285–1291. doi: 10.1016/j.renene.2008.09.015

[11] Yadav, A.S., & Bhagoria, J.L. (2013). CFD based heat transfer and fluid flow analysis of a solar air heater provided with circular transverse wire rib roughness. *Energy*, 55, 1127–1142. doi: 10.1016/j.energy.2013.03.066

[12] Al-Chlaihawi, K., Alyas, B.H., & Abdullah, B.N. (2024). Optimizing heat transfer and flow characteristics within a roughened solar air heater duct with compound turbulators. *Asia-Pacific Journal of Chemical Engineering*, 19(5), e3126. doi: 10.1002/apj.3126

[13] Manjunath, M.S., Karanth, K.V., & Sharma, N.Y. (2017). Numerical analysis of spherical turbulence generators on heat transfer enhancement of flat plate solar air heater. *Energy*, 121, 616–630. doi: 10.1016/j.energy.2017.01.032

[14] Maithani, R., Silori, A., Rana, J., & Chamoli, S. (2017). Numerical analysis of heat transfer and fluid flow of wavy delta winglets in a rectangular duct. *Thermal Science and Engineering Progress*, 2, 15–25. doi: 10.1016/j.tsep.2017.04.002

[15] Manjunath, M.S., Karanth, K.V., & Sharma, N.Y. (2018). Numerical investigation on heat transfer enhancement using sinusoidal corrugations. *International Journal of Mechanical Sciences*, 138, 219–228. doi: 10.1016/j.ijmecsci.2018.01.037

[16] Manjunath, M.S., Karanth, K.V., & Sharma, N.Y. (2019). Numerical Analysis of Flat Plate Solar Air Heater Integrated With an Array of Pin Fins on Absorber Plate for Enhancement in Thermal Performance. *Journal of Solar Energy Engineering*, 141(5), 051008. doi: 10.1115/1.4043517

[17] Antony, A.L., Shetty, S.P., Madhwesh, N., Sharma, N.Y., & Karanth, K.V. (2020). Influence of stepped cylindrical turbulence generators on the thermal enhancement factor of a flat plate solar air heater. *Solar Energy*, 198, 295–310. doi: 10.1016/j.solener.2020.01.065

[18] Arunkumar, H.S., Kumar, S., & Karanth, K.V (2020). Analysis of a solar air heater for augmented thermohydraulic performance using helicoidal spring shaped fins-A numerical study. *Renewable Energy*, 160, 297–311. doi: 10.1016/j.renene.2020.06.098

[19] Taslim, M.E., Li, T., & Kercher, D.M. (1996). Experimental Heat Transfer and Friction in Channels Roughened With Angled, V-Shaped, and Discrete Ribs on Two Opposite Walls. *Journal of Turbomachinery*, 118(1), 20–28. doi: 10.1115/1.2836602

[20] Olsson, C.O., & Sunden, B. (1998). Thermal and Hydraulic Performance of a Rectangular Duct With Multiple V-Shaped Ribs. *Journal of Turbomachinery*, 120(4), 1072–1077. doi: 10.1115/1.2825892

[21] Momin, A.M.E., Saini, J.S., & Solanki, S.C. (2002). Heat transfer and friction in solar air heater duct with V-shaped ribs. *International Journal of Heat and Mass Transfer*, 45(16), 3383–3396. doi: 10.1016/S0017-9310(02)00046-7

[22] Lanjewar, A., Bhagoria, J.L., & Sarviya, R.M. (2011). Heat transfer and friction in W-shaped rib roughness. *Energy*, 36(7), 4531–4541. doi: 10.1016/j.energy.2011.03.054

[23] Ghildyal, A., Bisht, V.S., Rawat, K.S., & Bhandari, P. (2023). Effect of D-shaped turbulators on performance. *Archives of Thermodynamics*, 44(2), 3–20. doi: 10.24425/ather.2023.146556

[24] Ghildyal, A., Bisht, V.S., Bhandari, P., Thapliyal, S., Kaushik, S., Ranakoti, L., Bangari, R,S., Srivastav, A., Kanojia, N., & Raj Paul, A. (2024). Comparative numerical evaluation of solar air heater. *Archives of Thermodynamics*, 45(4), 189–196. doi: 10.24425/ather.2024.152008

[25] Bohra, J., Bisht, V.S., Bhandari, P., Rawat, K.S., Singh, J., Kumar, K., & Rawat, B. (2022). Effect of blockage height ratio. *Materials Today: Proceedings*, 69, 153–157. doi: 10.1016/j.matpr.2022.08.279

[26] Chaudhari, M., Sharma, S.L., & Debbarma, A. (2023). Exergetic performance analysis of solar air heater with L-shaped ribs as roughness element. *Archives of Thermodynamics*, 44(3), 241–267. doi: 10.24425/ather.2023.147546

[27] Kumar, S., Bisht, V.S., Bhandari, P., Ranakoti, L., Negi, A., Bist, A.S., & Padalia, D. (2025). Computational analysis of modified solar air heater. *International Journal on Interactive Design and Manufacturing*, 19(5), 3467–3478. doi: 10.1007/s12008-024-01972-2

[28] Alhilali, H.A.A., Al-Chlaihawi, K., Zarzoor, A.K., & Flayyih, M.A. (2025). Optimization of thermohydraulic parameters in solar air heater ducts featuring I-shaped ribs. *Heat Transfer*, 54(6), 3790−3814. doi: 10.1002/htj.23384

[29] Chtourou, S., Djemel, H., Kaffel, M., & Baccar, M. (2021). Predicting rib pitch effect in plate heat exchangers. *SN Applied Sciences*, 3, 497. doi: 10.1007/s42452-021-04473-z

[30] Chtourou, S., Djemal, H., Kaffel, M., & Baccar, M. (2022). Numerical investigation for the optimal plate heat exchangers Thermo-hydraulic efficiency fitted with fragmental and fractal y-shaped obstacles. *Computational Thermal Sciences: An International Journal*, 14(3), 29–57. doi: 10.1615/computthermalscien.2022039597

[31] Dhamudia, K., & Senapati, J.R. (2025). A computational study for enhancement in thermal performance of plus(+)-shaped ribbed solar air heater duct. *International Communications in Heat and Mass Transfer*, 161, 108411. doi: 10.1016/j.icheatmasstransfer.2024.108411

[32] Dhamudia, K., & Senapati, J.R. (2025). Thermo-fluid performance with hook-shaped promoters. *International Communications in Heat and Mass Transfer*, 169, 109538. doi: 10.1016/j.icheatmasstransfer.2025.109538

[33] Varun, Saini, R.P., & Singal, S.K. (2007). A review on roughness geometry used in solar air heaters. *Solar Energy*, 81(11), 1340–1350. doi: 10.1016/j.solener.2007.01.017

[34] Gupta, D., Solanki, S.C., & Saini, J.S. (1997). Thermohydraulic performance of solar air heaters with roughened absorber plates. *Solar Energy*, 61(1), 33–42. doi: 10.1016/S0038-092X(97)00005-4

[35] Tanda, G. (2004). Heat transfer in rectangular channels with transverse and V-shaped broken ribs. *International Journal of Heat and Mass Transfer*, 47(2), 229–243. doi: 10.1016/S0017-9310(03)00414-9

[36] Karwa, R. (2003). Experimental studies of augmented heat transfer and friction in asymmetrically heated rectangular ducts with ribs on the heated wall in transverse, inclined, V-continuous and V-discrete pattern. *International Communications in Heat and Mass Transfer*, 30(2), 241–250. doi: 10.1016/S0735-1933(03)00035-6

[37] Ghritlahre, H.K., Chandrakar, P., & Ahmad, A. (2020). Application of ANN model to predict the performance of solar air heater using relevant input parameters. *Sustainable Energy Technologies and Assessments*, 40, 100764. doi: 10.1016/j.seta.2020.100764.

[38] Bhattacharyya, S., Sarkar, D., Roy, R., Chakraborty, S., Goel, V., & Almatrafi, E. (2021). Application of New Artificial Neural Network to Predict Heat Transfer and Thermal Performance of a Solar Air-Heater Tube. *Sustainability*, 13(13), 7477. doi: 10.3390/su13137477.

[39] Ghritlahre, H.K., & Prasad, R.K. (2018). Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models

of artificial neural network technique. *Journal of Environmental Management*, 223, 566–575. doi: 10.1016/j.jenvman.2018.06.033.

[40] Alatawi, E.S. (2025). Computational fluid dynamics analysis and machine learning study of heat transfer in solar air heaters with distinct ribs configuration. *Case Studies in Thermal Engineering*, 73, 106559. doi: 10.1016/j.csite.2025.106559.

[41] Alhuyi Nazari, M., Mukhtar, A., Yasir, A.S.H., Rashidi, M.M., Ahmadi, M.H., Blazek, V., Prokop, L., & Misak, S. (2023). Applications of intelligent methods in solar heaters: an updated review. *Engineering Applications of Computational Fluid Mechanics*, 17(1), 2229882. doi: 10.1080/19942060.2023.2229882.

[42] Bisht, V.S., Singh, D.B., Kumar, P., Jain, S., Bhandari, P., Mumtaz, M.A., & Rahman, M.H. (2025). Innovative spherical turbulators configurations for enhanced solar air heater performance: A numerical and analytical study. *Energy*, 342, 139601. doi: 10.1016/j.energy.2025.139601

[43] Tu, J., Yeoh, G.H., & Liu, C. (2019). *Computational fluid dynamics* (3rd ed.). Butterworth-Heinemann.

[44] Chamoli, S., Lu, R., Xu, D., & Yu, P. (2018). Thermal performance improvement using winglet generators. *Solar Energy*, 159, 966–983. doi: 10.1016/j.solener.2017.11.046

[45] Incropera, F.P., & DeWitt, D.P. (2002). *Fundamentals of heat and mass transfer*. John Wiley & Sons.

[46] Kakac, S., Shah, R.K., & Aung, W. (1987). *Handbook of single-phase convective heat transfer*. Wiley, New York.

[47] Lewis, M.J. (1975). Optimising thermohydraulic performance of rough surfaces. *International Journal of Heat and Mass Transfer*, 18(11), 1243–1248. doi: 10.1016/0017-9310(75)90232-X

[48] Maithani, R., & Saini, J.S. (2016). Heat transfer and friction correlations for V-rib roughness. *Experimental Thermal and Fluid Science*, 70, 220–227. doi: 10.1016/j.expthermflusci.2015.09.010