



## OPEN Toward empirical correlations for estimating the specific heat capacity of nanofluids utilizing GRG, GP, GEP, and GMDH

Omid Deymi<sup>1</sup>, Fahimeh Hadavimoghaddam<sup>2</sup>, Saeid Atashrouz<sup>3</sup>✉, Dragutin Nedeljkovic<sup>4</sup>, Meftah Ali Abuswer<sup>4</sup>, Abdolhossein Hemmati-Sarapardeh<sup>5,6</sup>✉ & Ahmad Mohaddespour<sup>7</sup>✉

When nanoparticles are dispersed and stabilized in a base-fluid, the resulting nanofluid undergoes considerable changes in its thermophysical properties, which can have a substantial influence on the performance of nanofluid-flow systems. With such necessity and importance, developing a set of mathematical correlations to identify these properties in various conditions can greatly eliminate costly and time-consuming experimental tests. Hence, the current study aims to develop innovative correlations for estimating the specific heat capacity of mono-nanofluids. The accurate estimation of this crucial property can result in the development of more efficient and effective thermal systems, such as heat exchangers, solar collectors, microchannel cooling systems, etc. In this regard, four powerful soft-computing techniques were considered, including Generalized Reduced Gradient (GRG), Genetic Programming (GP), Gene Expression Programming (GEP), and Group Method of Data Handling (GMDH). These techniques were implemented on 2084 experimental data-points, corresponding to ten different kinds of nanoparticles and six different kinds of base-fluids, collected from previous research sources. Eventually, four distinct correlations with high accuracy were provided, and their outputs were compared to three correlations that had previously been published by other researchers. These novel correlations are applicable to various oxide-based mono-nanofluids for a broad range of independent variable values. The superiority of newly developed correlations was proven through various statistical and graphical error analyses. The GMDH-based correlation revealed the best performance with an Average Absolute Percent Relative Error (AAPRE) of 2.4163% and a Coefficient of Determination ( $R^2$ ) of 0.9743. At last, a leverage statistical approach was employed to identify the GMDH technique's application domain and outlier data, and also, a sensitivity analysis was carried out to clarify the degree of dependence between input and output variables.

Nanoparticles, tiny species with an average size of 1–100 nm, are classified in various aspects such as their shape (or morphology), size, and physicochemical characteristics. Aluminum (Al), silver (Ag), titanium (Ti), and other metals, as well as oxide-based or non-metallic materials like copper-oxide (CuO), titanium-dioxide (TiO<sub>2</sub>), aluminum-oxide (Al<sub>2</sub>O<sub>3</sub>), and so on, can all be found in the nanoparticles category. When nanoparticles are mixed in conventional base-fluids, they generate novel compounds known as nanofluids with distinct properties<sup>1</sup>. Water (W), ethylene glycol (EG), water-ethylene glycol mixture (W+EG), propylene glycol (PG), glycerol (GC), and engine oil (EO) are just a few examples of the wide variety of base-fluids that might be employed. Finally, nanofluids can be prepared in the forms of mono-nanofluids (e.g., Ag/W or TiO<sub>2</sub>/EG), binary hybrid nanofluids

<sup>1</sup>Department of Mechanical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. <sup>2</sup>Institute of Unconventional Oil & Gas, Northeast Petroleum University, Daqing 163318, Heilongjiang, China. <sup>3</sup>Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. <sup>4</sup>College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait. <sup>5</sup>Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. <sup>6</sup>Key Laboratory of Continental Shale Hydrocarbon Accumulation and Efficient Development, Ministry of Education, Northeast Petroleum University, Daqing 163318, China. <sup>7</sup>Department of Chemical Engineering, McGill University, Montreal, QC H3A 0C5, Canada. ✉email: saeid\_atashrouz@aut.ac.ir; s.atashrouz@gmail.com; hemmati@uk.ac.ir; aut.hemmati@gmail.com; ahmad.mohaddespour@mail.mcgill.ca

(e.g., SiO<sub>2</sub>–CuO/EG), or ternary hybrid nanofluids (e.g., Al<sub>2</sub>O<sub>3</sub>–CuO–TiO<sub>2</sub>/W), according to their expected purpose and application<sup>2</sup>.

The utilization of nanofluids for a variety of practical purposes is a popular and developing scientific field that is getting more attention day by day. Some of the most important applications of nanofluids, which can be observed in numerous research studies and publications, include the following:

- heat exchangers with different configurations (such as shell and tube, finned tube, double pipe, etc.)<sup>3</sup>,
- various types of sun-based systems (especially solar collectors and photovoltaic panels)<sup>4,5</sup>,
- cooling systems of vehicle engines and electronic components<sup>6,7</sup>,
- cooling of nuclear reactors<sup>8</sup>,
- drug-delivery and biomedical systems<sup>9</sup>,
- enhanced oil recovery (EOR) techniques<sup>10</sup>,
- common machinery processes (milling, drilling, grinding, and turning)<sup>11</sup>.

Until now, many analytical and numerical calculations have been performed on fluid-flow problems based on various kinds of nanofluids, mainly in research pertaining to the sectors of energy, electronics, and medicine. For instance, Jia et al.<sup>12</sup> carried out the numerical analysis of a glazed photovoltaic-thermal (PV/T) collector employing two kinds of coolant nanofluid (Al<sub>2</sub>O<sub>3</sub>/W and TiO<sub>2</sub>/W) with the aim of investigating its thermal and electrical performances. McCash et al.<sup>13</sup> mathematically conducted a comparative analysis for the peristaltic flow of two distinct nanofluids (Ag–Cu/W and Cu/W) inside an elliptical duct with sinusoidally advancing boundaries. Shahzad et al.<sup>14</sup> analyzed the electro-osmotic flow of a blood-based hybrid nanofluid consisting of single- and multi-walled carbon nanotubes (SWCNT and MWCNT) through a multiple-stenosed artery with permeable walls. Alghamdi et al.<sup>15</sup> mathematically modeled the two-dimensional magneto-hydrodynamics (MHD) flow based on the Cu–Al<sub>2</sub>O<sub>3</sub>/W hybrid nanofluid between two co-axial stretchable disks and numerically assessed it using the finite element method (FEM). Baig et al.<sup>16</sup> presented exact analytical solutions for the stagnation-point flow of water-dispersed MWCNTs passing across a heated stretching cylinder.

Nanofluids have exceptional properties compared to normal fluids, which makes them a preferred choice in the field of mass and heat transfer phenomena. Following the stabilization of nanoparticles in the base-fluid after mixing together, the thermophysical properties of the resultant nanofluid, such as viscosity ( $\mu$ ), thermal conductivity ( $k$ ), specific heat capacity ( $C_p$ ), and density ( $\rho$ ), can be significantly altered. Several research efforts have been conducted on this topic in order to determine what other factors affect nanofluid properties and how these properties behave under different operational conditions. The present study only looks into changes in the specific heat capacity of mono-nanofluids ( $C_{p,nf}$ ) as a function of the specific heat capacity of nanoparticles ( $C_{p,np}$ ) and base-fluids ( $C_{p,bf}$ ), temperature ( $T$ ), particle volume fraction ( $\phi_v$ ), and average particle size ( $d_{np}$ ).

Recently, many researchers have been attempting to establish the specific heat values of nanofluids in various ways, such as by performing experimental tests, applying machine learning techniques (black-box models), and developing empirical mathematical relationships (white-box models). In 2008, Vajjha and Das<sup>17</sup> released an empirical correlation for estimating the specific heat capacity of Al<sub>2</sub>O<sub>3</sub>/W+EG nanofluid as a function of temperature (313–363 K) and volume concentration (2–10 vol%). Their experimental measurements served as the basis for this cubic polynomial correlation. In addition, by comparing their measurements with estimated values achieved from the theoretical model of Pak and Cho<sup>18</sup>, they observed an acceptable level of agreement, with maximum and average deviations reported as 22% and 15%, respectively. In another investigation conducted in 2009, Vajjha and Das<sup>19</sup> created a specific heat correlation using the experimental data related to SiO<sub>2</sub>/W, ZnO/W+EG, and Al<sub>2</sub>O<sub>3</sub>/W+EG nanofluids. Their correlation took into account four variables containing specific heat capacity of both nanoparticles and base-fluids, temperature, and particle volume concentration. This correlation set a defined range for each of the first two variables, and its constants for the three types of nanofluids were different. In a subsequent study by Vajjha and Das<sup>20</sup> in 2012, they updated the correlation proposed in reference<sup>19</sup> by incorporating the experimental data of CuO/W+EG nanofluid and making the temperature term dimensionless via entering a reference temperature. In 2013, Barbés et al.<sup>21</sup> assessed the specific heat capacity of Al<sub>2</sub>O<sub>3</sub>/EG and Al<sub>2</sub>O<sub>3</sub>/W nanofluids as a function of temperature and volume fraction. The researchers also developed a simple linear regression model that suited their experimental data well. In the next year, they performed another similar research<sup>22</sup> on CuO/W and CuO/EG nanofluids. In 2015, Cabaleiro et al.<sup>23</sup> studied the specific heat changes of five distinct metal-oxide-based nanofluids at high concentrations. These nanofluids contained ZrO<sub>2</sub>, ZnO, and MgO in pure EG, as well as ZrO<sub>2</sub> and ZnO in the W+EG mixture (50:50 vol%). Finally, the researchers developed a specific heat correlation as a function of the nanoparticle and base-fluid specific heat capacities and particle volume fraction. In another research, Sekhar and Sharma<sup>24</sup>, while studying the specific heat capacity of Al<sub>2</sub>O<sub>3</sub>/W nanofluid at low concentrations (0.01–1 vol%), proposed a specific heat correlation in accordance with 81 experimental data-points of water-based Al<sub>2</sub>O<sub>3</sub>, SiO<sub>2</sub>, TiO<sub>2</sub>, and CuO nanofluids collected from other researches. This regression equation was valid in the particular range of nanofluid temperatures, nanoparticle diameters, and volume fractions. It was found that the calculated values and experimental data were well compatible with each other, such that a deviation range between –8% and +10% was obtained. Satti et al.<sup>25</sup>, after evaluating the specific heat capacities of five various nanofluids consisting of CuO, ZnO, SiO<sub>2</sub>, TiO<sub>2</sub>, and Al<sub>2</sub>O<sub>3</sub> nanoparticles suspended in a W+PG mixture (40:60 wt%), developed a correlation onto 610 measured data by employing the Minitab statistical software. Popa et al.<sup>26</sup>, according to their experimental measures on the specific heat capacity of CuO/W, Al<sub>2</sub>O<sub>3</sub>/W, and Al<sub>2</sub>O<sub>3</sub>/EG nanofluids, recalibrated the correlation developed by Vajjha and Das<sup>19</sup> and captured its new coefficients based on the analyzed nanofluids. The authors reported excellent agreement with a maximum Average Relative Error of nearly 2%. Moldoveanu and Minea<sup>27</sup>, based on their examinations measuring the specific heat capacities of TiO<sub>2</sub>/W, SiO<sub>2</sub>/W, and Al<sub>2</sub>O<sub>3</sub>/W nanofluids, modified Sekhar and Sharma's<sup>24</sup> equation

to propose another correlation. Its validity was limited to a range of room temperatures and volume fractions of less than 5%. The average deviation of the specific heat values estimated via their correlation was reported to be roughly 11% when compared to the experimental results. In 2020, Çolak et al.<sup>28</sup> presented a correlation with an average error of  $-0.005\%$  utilizing a database with 1287 data-points made up of volume fraction and temperature (as independent variables) and experimental specific heat values of Cu–Al<sub>2</sub>O<sub>3</sub>/W hybrid nanofluid. Gao et al.<sup>29</sup> proposed a correlation that was fitted by the experimental specific heat capacities related to the hybrid nanofluid of GrapheneOxide–Al<sub>2</sub>O<sub>3</sub>/W with mass fractions of 0.05–0.15 wt% at a temperature range of 20–70 °C.

Since the specific heat capacity of nanofluids relies on various conditions and factors, it is often time-consuming, expensive, and difficult to quantify this property accurately via experimental procedures. To determine the specific heat capacity values of mono-nanofluids, robust machine-learning methods were employed in our recent study<sup>30</sup>, but the current work intends to provide novel empirical correlations. The literature review clearly indicates that the number of correlations developed so far by other researchers is limited. On the other hand, old correlations are only valid and suitable for certain nanofluids and only cover a restricted range of independent variables. As a result, there is still an opportunity for advancement in this subject, and it is conceivable to provide more comprehensive correlations with fewer constraints that can be applied to an extensive collection of nanofluids.

The current study's primary aim is to provide novel and precise mathematical correlations for estimating the specific heat capacity of mono-nanofluids. It should be noted that the considered mono-nanofluids have been selected only based on the oxide-based or non-metallic nanoparticles dissolved in conventional base-fluids. For this purpose, an extensive database consisting of experimental data taken from the literature is utilized (the same one involved in our previous study<sup>30</sup>), and finally, four correlations are created using robust soft-computing procedures. To evaluate and compare the quality of the present correlations with prior ones suggested by others, various statistical and graphical error analyses are applied. The principal advantage of these new mathematical equations is that they have considerably mitigated the limitations of the earlier correlations. Therefore, they are applicable to many different oxide-based mono-nanofluids and a broad range of independent input variables.

### Dataset specifications

The specific heat capacity of mono-nanofluids ( $C_{p,nf}$ ) can be defined as a function of average particle size ( $d_{np}$ ), particle volume fraction ( $\phi_v$ ), temperature ( $T$ ), nanoparticle specific heat capacity ( $C_{p,np}$ ), and base-fluid specific heat capacity ( $C_{p,bf}$ ). Accordingly, 2084 data-points related to the experimental values of  $C_{p,nf}$  were collected from diverse sources (19 references) available in the literature<sup>17,19,21–23,25,29,31–42</sup>. Before generating the final flawless dataset, necessary corrective actions were carried out in the data pre-processing stage, such as data cleaning, data integration, and data reduction, to ensure the establishment of more accurate estimations. Table S1 (in the Supplementary File) lists the references used to extract the required data, the characteristics related to the various types of nanoparticles and base-fluids available in the final database, as well as the ranges of input (independent) and output (dependent) variables. Additionally, Table S2 (in the Supplementary File) provides the descriptive statistics for each of the variables included in the database. It should be mentioned that the information given in these two tables was also reported in our prior article<sup>30</sup>. Moreover, box-and-whisker plots and frequency distribution histograms related to the six variables available in the database were drawn in previous work<sup>30</sup>, which are avoided here.

### Description of techniques

The modeling and prediction of complex or unknown systems' behavior employing input–output data is a common use of system identification techniques. Because of this, there is now much more interest in soft-computing techniques, which deal with data processing in uncertain and inexplicit environments. Many scientific studies have discussed the application of evolutionary algorithms as efficient soft-computing tools for system identification<sup>43–46</sup>.

Due to the data-driven nature of all soft-computing strategies, having more data leads to a more reliable and comprehensive model. In this regard, the present study covers four popular techniques for creating distinct mathematical models to estimate the specific heat capacities of mono-nanofluids. These techniques include Generalized Reduced Gradient (GRG), Genetic Programming (GP), Gene Expression Programming (GEP), and Group Method of Data Handling (GMDH), which are all covered with further details in the following subsections.

The specific algorithms adopted in this study were essentially chosen based on their suitability and ability to address the research goal, their compatibility with the quantity and quality of data-points contained within the dataset, and their own unique strengths and weaknesses. The successful usage and efficient performance of these techniques have been proven across a variety of investigations conducted in numerous fields of engineering sciences<sup>47–54</sup>.

### A: Generalized Reduced Gradient (GRG)

One of the typical strategies used to solve multi-variable problems is the generalized reduced gradient (GRG). In this technique, the decision variable is the vector of  $X(x_1, x_2, \dots, x_n)$ , and the constraints are the functions of  $g_1, g_2, \dots, g_m$ . The objective function or any constraint function can be linear or non-linear. Additionally, it's possible that the ranges can be undefined. If no constraints exist, the problem is called an unconstrained optimization problem. It should be noticed that the lower and upper limits of the variables do not act as additional constraints, but are applied separately to the given program<sup>55</sup>.

The GRG technique uses the first-order partial derivatives of each function  $g_i$  with respect to the variables  $x_i$ , which are calculated by approximation of the forward or central finite difference. The program is executed according to the simulator's initial values. If the values provided by the simulator do not satisfy all constraints

$g$ , the first step of optimization starts, and in this situation, the objective function equals the sum of deviations from the constraints plus a fraction of the problem's objective function. This optimization ends with a message that indicates the practicality or impracticality of solving the problem. It should be noted that in some cases, the impractical response is due to the limitation of the program in local minima, which can be corrected by changing the initial value and re-executing. In the next step, the complete optimization cycle is performed, and the final report is obtained<sup>54,56</sup>.

### B: Genetic Programming (GP)

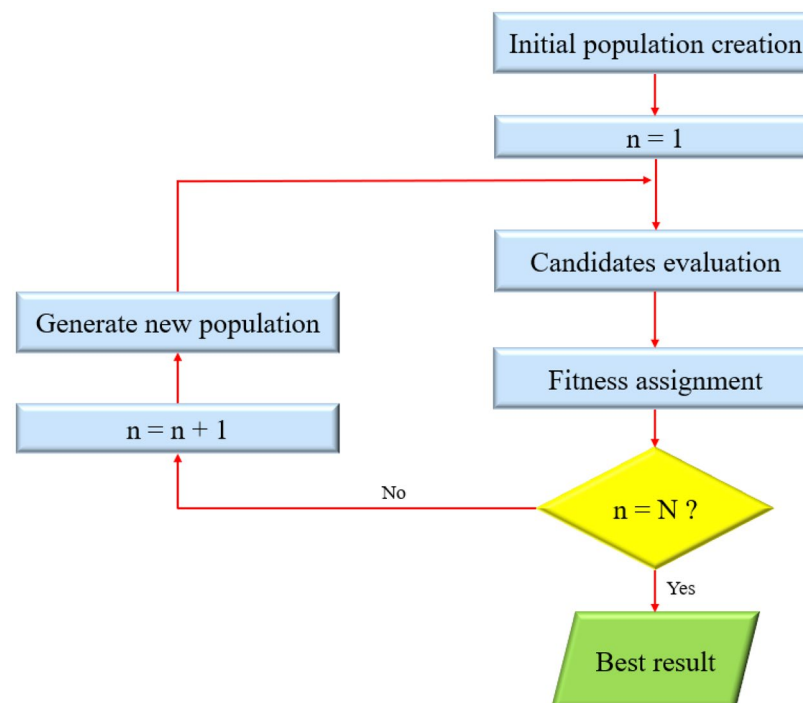
A subset of the expanding family of evolutionary algorithms, namely genetic programming (GP), involves a population of randomly generated computer programs (Fig. 1). GP applies natural evolution-inspired search principles to a variety of diverse problems, particularly parameter optimization<sup>57</sup>. The concept of genetic programming clarifies a branch of artificial intelligence (AI) study that is concerned with the evolution of computer programming codes. The evolution process operates artificially in a way that is comparable to how living organisms evolve naturally, but most of its complicated details have been summarized<sup>57</sup>.

GP is an organized procedure for getting machines to solve a problem automatically, starting with a high-level statement that must be executed. This technique is considered a systematic and domain-independent algorithm that genetically reproduces a population of programs for problem-solving<sup>58,59</sup>. Besides, GP is an AI framework that imitates natural selection to identify the best prediction. This iterative procedure chooses the most suitable descendant to pass and regenerate at each new step of the algorithm<sup>60</sup>. GP has been successfully employed for a great number of various tasks, including data-mining<sup>61</sup>, pattern recognition<sup>62</sup>, and artificial neural network (ANN) synthesis<sup>63</sup>.

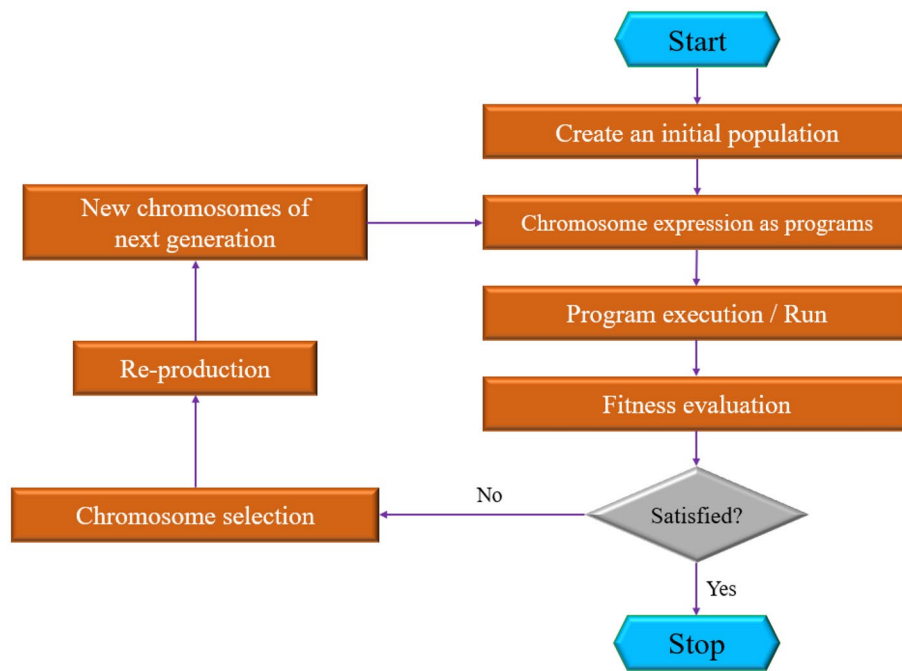
### C: Gene Expression Programming (GEP)

In 1999, *Ferreira* first presented the concept of Gene Expression Programming (GEP) as a strong soft-computing technique<sup>64,65</sup>. It is an advanced and upgraded version of GP and employs two major computational components in its structural computations for regression tasks, namely chromosomes (genotype) and expression trees (phenotype). However, the GP structure utilizes nonlinear forms of responses, namely parse trees. Actually, the chromosomes detect the selection or initial predictions for a particular problem, in which case using an interpretation strategy will allow the expression trees to provide more accurate solutions to the issues<sup>66–68</sup>. The simple flowchart related to this technique is shown in Fig. 2.

GEP, as a full-grown genotype/phenotype technique, develops computer programs encoded in fixed-length linear chromosomes. The structure of linear chromosomes makes possible the beneficial and unrestricted activity of crucial genetic operators like recombination, transposition, and mutation. Besides, GEP contributes a similar form of tree description and illustration to GP in order to retrace readily the stages accomplished by GP and discover effortlessly new boundaries established by exceeding the phenotype threshold<sup>66</sup>.



**Figure 1.** A flowchart of the GP technique.

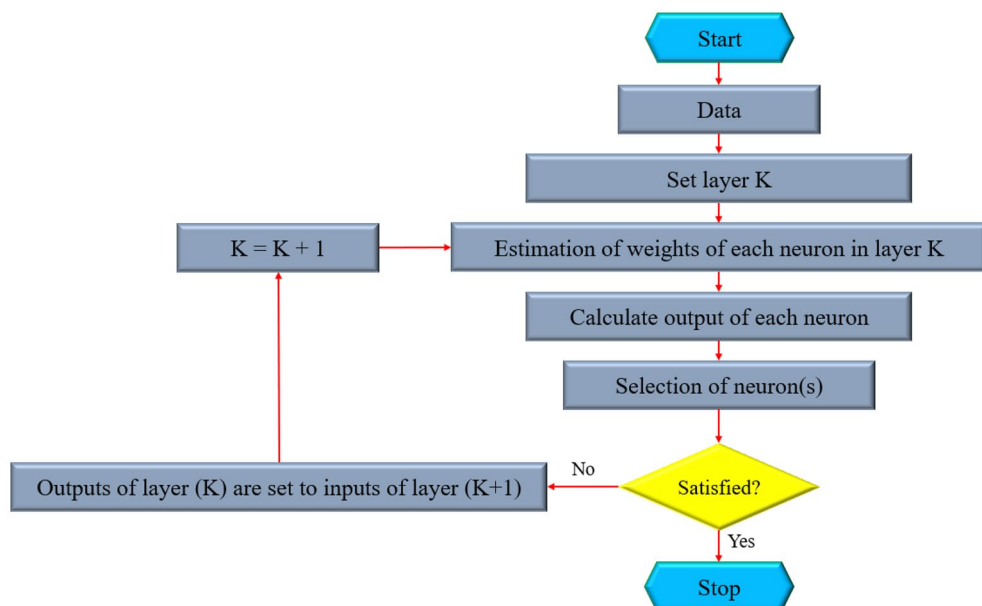


**Figure 2.** A flowchart of the GEP technique.

In recent decades, GEP, as a preferred and recognized evolutionary approach for automatically creating computer programs, has developed and advanced quickly, such that a variety of advanced GEPs have been suggested for real-world applications that are proliferating rapidly<sup>69</sup>.

#### D: Group Method of Data Handling (GMDH)

A novel computational method for managing complex nonlinear computer tasks, known as the group method of data handling (GMDH), was first developed by *Ivakhnenko*. Figure 3 illustrates a functional flowchart for the GMDH technique. This technique can describe an explicit relationship between a mathematical model's inputs and its corresponding output<sup>70,71</sup>. This technique is one of the most satisfactory groups of artificial neural networks (ANNs) and is also known as a polynomial neural network (PNN). The inner layers of the GMDH technique contain a variety of independent nodes. This technique is provided in polynomial form, where all



**Figure 3.** A flowchart of the GMDH technique.

nodes in every layer are joined in couples through a quadratic polynomial and generate new polynomial-formed nodes in the subsequent layer, in accordance with the self-organizing principle<sup>67,72</sup>.

The GMDH technique is based on employing several nodes from the middle layers. Every node of GMDH returns a value by applying a quadratic polynomial approach that incorporates the preceding node<sup>65,73</sup>. As previously pointed out, the nodes of the next layers are produced using the quadratic polynomial functions that combine the nodes already present in the prior layers. The following formula represents the procedure of the GMDH technique<sup>65,67,74</sup>:

$$y_i = c + \sum_{i=1}^w \sum_{j=1}^w \dots \sum_{k=1}^w c_{ij\dots k} x_i^n x_j^n \dots x_k^n \quad ; \quad n = 1, 2, \dots, 2^m \quad (1)$$

in which  $x_{ij\dots k}$  and  $y_i$  respectively depict input and output variables of the model,  $m$  and  $w$  respectively indicate the size of layers and the number of inputs, as well as  $c$  and  $c_{ij\dots k}$  stand for the polynomial factors.

By defining new nodal parameters (namely  $P_i$ ) for the situation of two nodes connected using a quadratic polynomial approach, the below formula is presented<sup>65</sup>:

$$P_i^{GMDH} = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad (2)$$

The least-squares method (LSM) is then used to minimize the variance between model predictions and actual data, as shown below<sup>65</sup>:

$$\delta_j^2 = \sum_{i=1}^{N_i} (y_i - P_i^{GMDH})^2 \quad ; \quad j = 1, 2, \dots, \binom{w}{2} \quad (3)$$

where  $N_i$  and  $w$  are the size of the training subset and the number of variables, respectively.

Later, the following general matrix will be created after obtaining the constants of Eq. (2)<sup>65</sup>:

$$Y = A^T X \quad \text{or} \quad A^T = YX^T (XX^T)^{-1} \quad (4)$$

where  $X = [x_1, x_2, \dots, x_w]$ ,  $Y = [y_1, y_2, \dots, y_w]$ , and  $A = [a_0, a_1, \dots, a_5]$ .

The model constants are attained over the training step. According to the below criterion, the testing subset is utilized to identify the optimum combination of two independent variables<sup>70</sup>:

$$\delta_j^2 = \sum_{i=1+N_i}^N (y_i - P_i^{GMDH})^2 < \varepsilon \quad ; \quad j = 1, 2, \dots, \binom{w}{2} \quad (5)$$

## Mathematical models for calculation of nanofluids' specific heat capacity

### Old correlations presented in the previous studies

Two primary theoretical models have been established to approximate the specific heat capacity values of nanofluids. The first one was developed by Pak and Cho<sup>18</sup> in 1998 based on the ideal gases' mixing theory according to Eq. (6). Another one was made by Xuan and Roetzel<sup>75</sup> in 2000 on the basis of the thermal equilibrium of nanoparticles and base-fluid according to Eq. (7).

$$C_{P,nf} = \phi_v C_{P,np} + (1 - \phi_v) C_{P,bf} \quad (6)$$

$$C_{P,nf} = \frac{\phi_v \rho_{np} C_{P,np} + (1 - \phi_v) \rho_{bf} C_{P,bf}}{\phi_v \rho_{np} + (1 - \phi_v) \rho_{bf}} \quad (7)$$

Besides these two models, several other empirical correlations have been generated by various researchers in recent studies. Since the full description of these studies was detailed in the literature review part of the present work, only the developed correlations and their validity conditions are reported here:

- Vajjha and Das<sup>17</sup>:

$$C_{P,nf} = A \cdot \phi_v + B \cdot \phi_v \cdot T + C \cdot \phi_v \cdot T^2 + D \cdot \phi_v \cdot T^3 \quad (8)$$

where  $T$  is in Kelvin and the coefficients  $A$ ,  $B$ ,  $C$ , and  $D$  are quartic polynomial functions of  $\phi_v$ . By applying Eq. (8), the specific heat capacity of  $\text{Al}_2\text{O}_3$  nanoparticles dissolved in a 40:60 (wt%) water-ethylene glycol mixture can be calculated at volume concentrations of 2–10 vol% and a temperature range of 313–363 K.

- Vajjha and Das<sup>19</sup>:

$$\frac{C_{P,nf}}{C_{P,bf}} = \frac{(A \cdot T) + B \cdot \left( \frac{C_{P,np}}{C_{P,bf}} \right)}{C + \phi_v} \quad (9)$$

- Vajjha and Das<sup>20</sup>:

$$\frac{C_{P,nf}}{C_{P,bf}} = \frac{A' \cdot \left(\frac{T}{T_0}\right) + B \cdot \left(\frac{C_{P,np}}{C_{P,bf}}\right)}{C + \phi_v} \quad (10)$$

Equations (9) and (10) are applicable across temperatures of  $315 < T(K) < 363$  as well as volume fractions of  $0 < \phi_v(\%) \leq 0.07$  for nanofluids containing ZnO and  $0 < \phi_v(\%) \leq 0.1$  for nanofluids containing SiO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub>, all dispersed in a mixture of water and ethylene glycol (40:60 wt%). Regarding the type of nanoparticles, Table 1 presents the constants of these two equations and the error values reported by the aforementioned authors.

- Sekhar and Sharma<sup>24</sup>:

$$\frac{C_{P,nf}}{C_{P,bf}} = 0.8429 \left(1 + \frac{T_{nf}}{50}\right)^{-0.3037} \left(1 + \frac{d_{np}}{50}\right)^{0.4167} \left(1 + \frac{\phi_v}{100}\right)^{2.272} \quad (11)$$

Equation (11) is valid for water-based nanofluids comprising CuO, SiO<sub>2</sub>, TiO<sub>2</sub>, and Al<sub>2</sub>O<sub>3</sub> and in the ranges of  $20 < T_{nf}(\text{°C}) < 50$ ,  $15 < d_{np}(\text{nm}) < 50$ , and  $0.01 < \phi_v(\%) < 4.00$ . The error range of this correlation was reported between -8% and +10%.

- Çolak et al.<sup>28</sup>:

$$C_{P,nf} = (A \cdot T + B)(1 + C \cdot \phi_v^D) \quad (12)$$

where  $T$  is temperature (°C) and  $\phi_v$  is volume fraction (vol%). This correlation, which has been developed for Cu–Al<sub>2</sub>O<sub>3</sub>/W hybrid nanofluid, has constants and error values shown in Table 2.

- Gao et al.<sup>29</sup>:

$$C_{P,nf} = A + B \cdot \phi_m + C \cdot T + D \cdot \phi_m^2 + E \cdot \phi_m \cdot T + F \cdot T^2 \quad (13)$$

Equation (13) has been generated based on experimental data of GrapheneOxide–Al<sub>2</sub>O<sub>3</sub>/W hybrid nanofluids, in which  $T$  and  $\phi_m$  are temperature (°C) and mass fraction (wt%), respectively. Table 3 lists the constants and the error value related to this equation.

## New correlations developed in the present study

### (a) The GRG-based correlation:

Using the *solver* tool available in the *Data* tab of *Excel* software, various mathematical relationships with different numbers of constants can be created through the implementation of the GRG technique. For this purpose, a correlation in the form of Eq. (14) was developed with the aim of minimizing the average absolute percent relative error (AAPRE) regarding input and output variables (for the total dataset). According to the average and mode values related to the nanoparticle size and temperature variables, available in the descriptive statistics presented in Table S2 (in the Supplementary File), the reference values for this equation were considered to be  $d_0 = 50$  nm and  $T_0 = 300$  K.

Nanoparticle	A	A'	B	C	Maximum error (%)	Average absolute error (%)
ZnO	0.0004604	0.12569	0.9855	0.2990	4.4	2.70
SiO <sub>2</sub>	0.0017690	0.48294	1.1937	0.8021	3.1	1.50
Al <sub>2</sub> O <sub>3</sub>	0.0008911	0.24327	0.5179	0.4250	5.0	2.28

**Table 1.** Characteristics of correlations (9) and (10) provided by Vajjha and Das<sup>19,20</sup>.

A	B	C	D	Maximum error	Average error (%)
0.0002322	4.168	-0.01175	0.6207	0.031	-0.005

**Table 2.** Characteristics of correlation (12) provided by Çolak et al.<sup>28</sup>.

A	B	C	D	E	F	Maximum error (%)	R <sup>2</sup>
3.918	-218.3	$6.596 \times 10^{-3}$	$3.185 \times 10^5$	5.278	$-5.35 \times 10^{-5}$	0.86	0.983

**Table 3.** Characteristics of correlation (13) provided by Gao et al.<sup>29</sup>.

$$C_{P,nf}^{GRG} = C_{P,bf} \cdot \left[ a_0 + a_1 \cdot \left( \frac{d_{np}}{d_s} \right)^{a_2} + a_3 \cdot \left( \frac{\phi_v}{100} \right)^{a_4} + a_5 \cdot \left( \frac{T}{T_s} \right)^{a_6} + a_7 \cdot \left( \frac{C_{P,np}}{C_{P,bf}} \right)^{a_8} \right] \quad (14)$$

where  $a_0 = -1.459532$ ,  $a_1 = 1.191867$ ,  $a_2 = -0.044737$ ,  $a_3 = -1.889018$ ,  $a_4 = 1.014473$ ,  $a_5 = 1.172350$ ,  $a_6 = 0.102400$ ,  $a_7 = 0.202981$ ,  $a_8 = 0.960041$

The correlations obtained using meta-heuristic techniques are as follows.

(b) *The GP-based correlation:*

$$C_{P,nf}^{GP} = b_0 + b_1 \cdot \left\{ b_2 \cdot \phi_v - [b_3 + \exp(b_4 \cdot C_{P,bf})] \cdot \log \left[ \begin{array}{l} b_5 + b_6 \cdot T + \left( b_7 \cdot \frac{T}{d_{np}} + b_8 \cdot \frac{1}{\phi_v} \right) \cdot C_{P,np} \\ - \exp(b_9 \cdot C_{P,bf}) \end{array} \right] \right\} \quad (15)$$

where  $b_0 = -2.124005$ ,  $b_1 = -0.071654$ ,  $b_2 = 0.786437$ ,  $b_3 = 8.280478$ ,  $b_4 = 0.474517$ ,  $b_5 = 9.110290$ ,  $b_6 = 0.814494$ ,  $b_7 = 1.277965$ ,  $b_8 = 0.989456$ ,  $b_9 = 0.772331$

(c) *The GEP-based correlation:*

$$C_{P,nf}^{GEP} = c_0 - \left[ (c_1 \cdot \phi_v + c_2 \cdot C_{P,np}) - (c_3 + c_4 \cdot C_{P,np}) \cdot \left( c_5 - c_6 \cdot \frac{d_{np} \cdot \phi_v}{T} \cdot C_{P,bf} \right) \right] \cdot C_{P,bf} - (c_7 - c_8 \cdot T) \cdot C_{P,bf}^2 \quad (16)$$

where  $c_0 = 0.327155$ ,  $c_1 = 2.615432 \times 10^{-3}$ ,  $c_2 = 2.191448 \times 10^{-3}$ ,  $c_3 = 0.0559709$ ,  $c_4 = 5.341911 \times 10^{-3}$ ,  $c_5 = 12.798509$ ,  $c_6 = 0.555788$ ,  $c_7 = 0.0123802$ ,  $c_8 = 1.3806004 \times 10^{-4}$

(d) *The GMDH-based correlation:*

$$C_{P,nf}^{GMDH} = 0.045762 + 0.798159 d_0 + 0.170415 d_1 + 1.046540 d_0 d_1 - 0.614356 d_0^2 - 0.426339 d_1^2 \quad (17)$$

where

$$d_0 = -8.53143 + 5.2795 d_2 - 0.804102 d_2^2 + 0.0791396 d_2 C_{P,bf} + 0.638003 C_{P,bf} \quad (17a)$$

$$d_1 = -0.645007 - 0.83654 d_3 - 0.601428 d_3^2 + 1.64147 d_3 C_{P,bf} + 2.13504 C_{P,bf} - 1.07153 C_{P,bf}^2 \quad (17b)$$

$$d_2 = 4.02018 - 0.0248309 d_{np} - 2.45377 \times 10^{-3} d_{np} \phi_v + 2.32343 \times 10^{-4} d_{np}^2 + 6.5419 \times 10^{-3} \phi_v^2 \quad (17c)$$

$$d_3 = 1.81744 + 1.19136 d_4 - 1.44623 d_5 - 0.0777568 d_4 d_5 + 0.290077 d_5^2 \quad (17d)$$

$$d_4 = 2.208 - 0.0779644 \phi_v - 0.372357 C_{P,bf} + 2.96343 \times 10^{-3} \phi_v C_{P,bf} + 1.60252 \times 10^{-3} \phi_v^2 + 0.195263 C_{P,bf}^2 \quad (17e)$$

$$d_5 = -8.27289 + 0.0301767 d_{np} + 0.0604513 T - 2.14838 \times 10^{-4} d_{np} T + 2.83985 \times 10^{-4} d_{np}^2 - 6.68282 \times 10^{-5} T^2 \quad (17f)$$

## Results and discussion

### Evaluation and comparison of the correlations

In the current research, four new correlations—Eqs. (14) to (17)—were attained with the aim of estimating the specific heat capacity of mono-nanofluids by executing the GRG, GP, GEP, and GMDH techniques. On the other hand, three previously proposed Eqs. (6), (11), and (12) were selected and implemented on the entire employed database. Then, in order to compare the estimated outcomes of all seven equations with their corresponding experimental values and to evaluate the performance and accuracy of those equations, the following statistical parameters for train and test subsets (including 1667 and 417 data-points, respectively) and the total database (with 2084 data-points) were computed and declared according to Tables 4 and 5.

1. *Average Percent Relative Error (APRE)* This parameter determines the relative deviation of estimated data-points from the associated experimental values and is computed as:

$$APRE = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(C_{P,nf,i})_{exp} - (C_{P,nf,i})_{est}}{(C_{P,nf,i})_{exp}} \times 100 \right] \quad (18)$$



Equation	APRE (%)	AAPRE (%)	RMSE	SD	R <sup>2</sup>
Equation (6): Pak and Cho <sup>18</sup>	- 8.1331	8.2695	0.3009	0.1048	0.9183
Equation (11): Sekhar and Sharma <sup>24</sup>	- 8.1348	16.1876	0.6092	0.2109	0.3045
Equation (12): Çolak et al. <sup>28</sup>	- 33.0155	33.1611	1.0837	0.4250	0.0692

**Table 4.** The statistical parameters of old correlations presented in the previous studies.

Technique	Dataset	APRE (%)	AAPRE (%)	RMSE	SD	R <sup>2</sup>
GRG [Eq. (14)]	Train	- 0.0893	2.9009	0.1346	0.0417	0.9549
	Test	- 0.2619	3.1357	0.1436	0.0455	0.9513
	Total	- 0.1239	2.9479	0.1365	0.0425	0.9542
GP [Eq. (15)]	Train	- 1.0456	3.0771	0.1314	0.0428	0.9588
	Test	- 0.8034	2.7447	0.1180	0.0376	0.9676
	Total	- 0.9971	3.0106	0.1288	0.0418	0.9607
GEP [Eq. (16)]	Train	- 0.5012	2.4082	0.1102	0.0362	0.9702
	Test	- 0.2091	2.1603	0.0935	0.0295	0.9790
	Total	- 0.4428	2.3586	0.1071	0.0349	0.9720
GMDH [Eq. (17)]	Train	- 0.1050	2.3974	0.1007	0.0334	0.9750
	Test	- 0.1351	2.4915	0.1065	0.0351	0.9716
	Total	- 0.1110	2.4163	0.1019	0.0338	0.9743

**Table 5.** The statistical parameters of new correlations developed in the current study.

2. *Average Absolute Percent Relative Error (AAPRE)* This is calculated similarly to APRE with the exception that the errors' absolute values are considered when determining the final values of errors:

$$AAPRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(C_{P,nf,i})_{exp.} - (C_{P,nf,i})_{est.}}{(C_{P,nf,i})_{exp.}} \times 100 \right| \quad (19)$$

3. *Root Mean Square Error (RMSE)* This parameter indicates the level of data scattering around the zero point and is computed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(C_{P,nf,i})_{exp.} - (C_{P,nf,i})_{est.}]^2} \quad (20)$$

4. *Standard Deviation (SD)* This parameter represents the amount of data scattering from the average value and is calculated as:

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left[ \frac{(C_{P,nf,i})_{exp.} - (C_{P,nf,i})_{est.}}{(C_{P,nf,i})_{exp.}} \right]^2} \quad (21)$$

5. *Coefficient of Determination (R<sup>2</sup>)* This parameter evaluates how well the experimental and estimated values match, as follows:

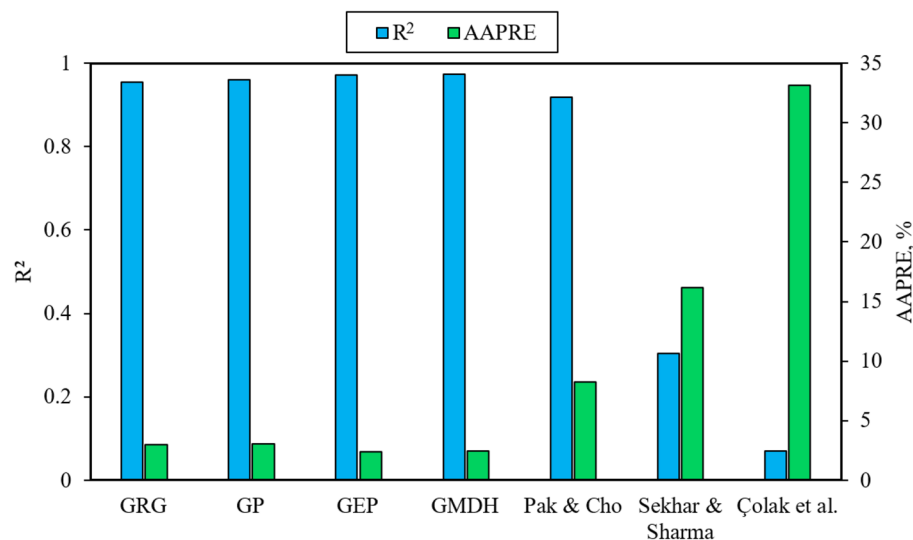
$$R^2 = 1 - \frac{\sum_{i=1}^n [(C_{P,nf,i})_{exp.} - (C_{P,nf,i})_{est.}]^2}{\sum_{i=1}^n [(C_{P,nf,i})_{exp.} - C_{P,nf,ave}]^2}, \quad \text{where } C_{P,nf,ave} = \frac{1}{n} \sum_{i=1}^n (C_{P,nf,i})_{exp.} \quad (22)$$

As seen in Table 4, Eq. (6) can provide relatively good estimations of the specific heat capacity for mono-nanofluids available in the database. Such a consequence has also been mentioned in the research of Vajjha and Das<sup>17,20</sup> and Murshed<sup>31</sup>, but it contradicts the findings of Zhou and Ni<sup>76</sup> and Zhou et al.<sup>77</sup>. However, Eq. (11) has not shown a good performance due to the parameter-related limitations implied in Ref.<sup>24</sup> and the broad range of the variables existing in the current work. In addition, a poor and unacceptable performance is observed from Eq. (12) because it was obtained based on Cu–Al<sub>2</sub>O<sub>3</sub>/W hybrid nanofluids, while the database used in the present study only comprises mono-nanofluids. This result is owing to the dissimilarity of stability levels and thermophysical properties found in mono and hybrid nanofluids caused by their structure-related differences.

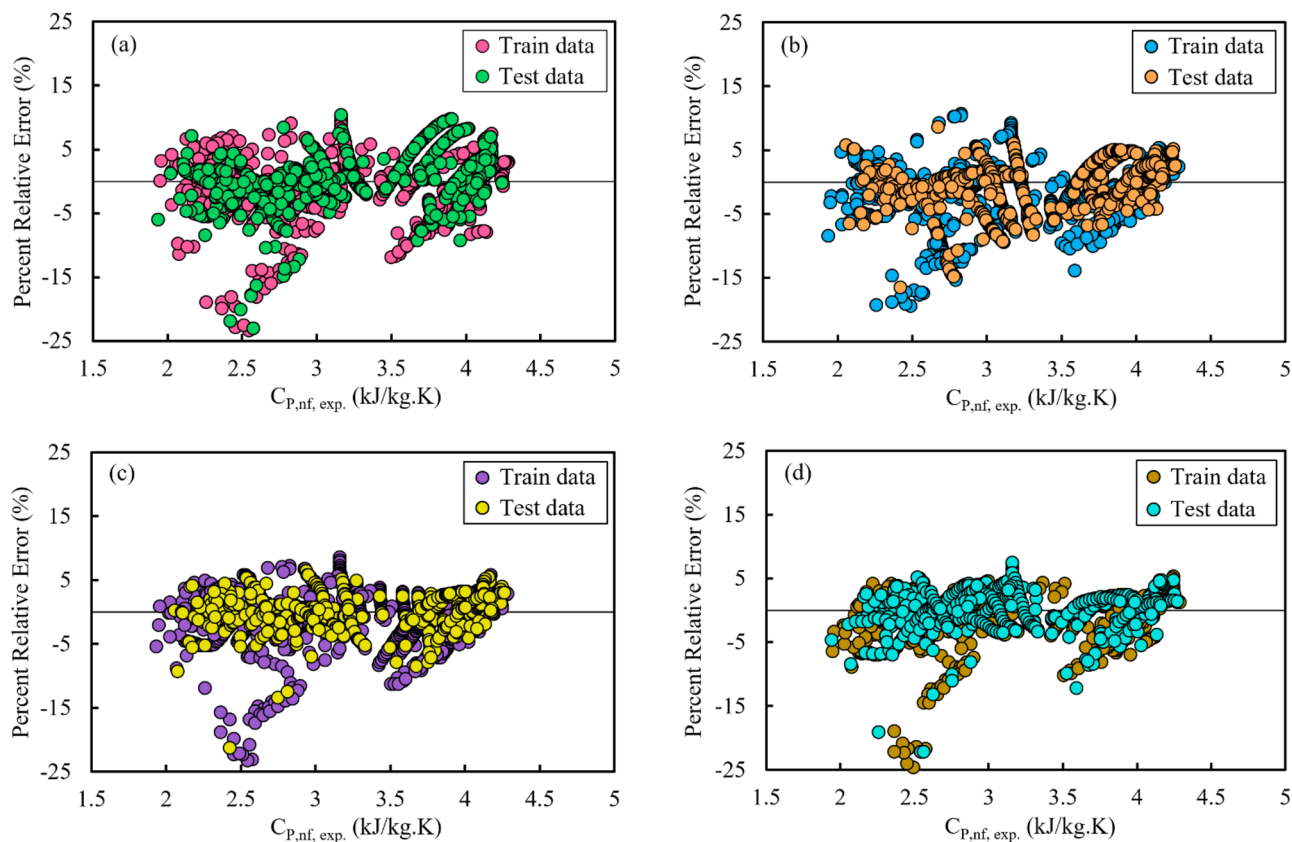
If the results of four novel correlations developed in the current study are compared with the estimations of the other three equations, it can be understood that these four correlations have very good accuracy in estimating

the specific heat capacity of mono-nanofluids and have achieved acceptable error rates (AAPRE < 3.2% and  $R^2 > 0.95$ ). This issue is also depicted in Fig. 4 as a graphical analysis of the AAPRE and  $R^2$  parameters. When carefully observed, the best performance is found for correlations obtained using the GMDH and GEP techniques.

Applying graphical error analysis is a suitable and useful tool for checking the performance of predictive models, especially when multiple various models are to be compared side by side. Figure 5 displays percent relative errors related to the estimated values of  $C_{p,nf}$  obtained from the GRG, GP, GEP, and GMDH techniques for the train and test subsets versus the experimental values of  $C_{p,nf}$ . In this graph, the data-points are scattered



**Figure 4.** Graphical comparison of AAPRE and  $R^2$  values for the introduced correlations.



**Figure 5.** Percent relative error distribution against experimental values of  $C_{p,nf}$  for different correlations: (a) GRG, (b) GP, (c) GEP, and (d) GMDH.

along a line with zero error to shed light on the error pattern of estimation techniques. A large accumulation of data-points around this line indicates that a technique is highly accurate.

As illustrated in Fig. 5, a large percentage of the experimental data-points are overestimated by all techniques. In other words, a majority of the estimated values are greater than the experimental values, and therefore, the relative errors are negative numbers [see Eq. (18)]. However, Fig. 5d demonstrates that the GMDH technique overestimates a relatively small fraction of data-points (low to medium values of  $C_{p,nf}$ ) and performs a credible task in evenly estimating the other data-points. Although this technique would be suitable for estimating just medium- and large-range values of  $C_{p,nf}$ , it still has almost an error tendency and needs to be utilized carefully.

According to Fig. 5d, the relative errors related to the  $C_{p,nf}$  values estimated by the GMDH technique are located near the line with zero error (within the error range of  $-24.6\%$  to  $+7.5\%$ ). This technique has made more accurate and reliable estimates than those of the other techniques.

In order to undertake a more thorough analysis regarding the accuracy and precision of the suggested correlations, another visual evaluation was performed by utilizing a particular kind of scatter chart known as the cross-plot. In this method, the estimated data-points from a correlation are put against the experimental data and across a  $45^\circ$  line (known as a unit-slope line) passing the plot's origin. Hence, the effectiveness of each correlation can be judged by how close the trend is to the  $45^\circ$  line. The cross-plots related to the developed correlations for  $C_{p,nf}$  estimation applying the GRG, GP, GEP, and GMDH techniques are presented in Fig. 6.

According to Fig. 6, a fairly packed accumulation of data-points along the unit-slope line is observed, particularly for the GMDH technique, whose outputs are in respectable agreement with the experimental data. Furthermore, the GRG technique has higher data deviations from the unit-slope line than the other techniques.

In Fig. 7, the group-error distribution diagrams in terms of five input variables are exhibited for three old and four novel correlations. This method first involves classifying the input or independent variables into a certain number of ranges (or categories) according to the scope of their changes, after which the estimation error values for the target variable are determined and plotted in each range.

As depicted by Fig. 7a regarding the “average particle size” parameter, the lowest AAPRE values correspond to the range of 52–66 nm for the GRG, GP, and GEP techniques and the range of 10–24 nm for the GMDH technique. As shown in Fig. 7b regarding the “particle volume fraction” parameter, the lowest AAPRE values are achieved in the range of 4–6 vol% for the GRG and GEP techniques and the range of 8–10 vol% for the GP, GEP, and GMDH techniques. As can be seen in Fig. 7c concerning the “temperature” parameter, the lowest AAPRE values are attained in the range of 373.15–413.15 K for the GRG and GMDH techniques, the range of 293.15–333.15 K for the GP technique, and the range of 333.15–373.15 K for the GEP technique. As indicated in Fig. 7d regarding the “nanoparticle specific heat” parameter, the lowest AAPRE values are seen in the range of 0.536–0.692 kJ/kg.K (mainly  $\text{Si}_3\text{N}_4$ , TiN, CuO, and  $\text{TiO}_2$ ) for the GRG technique, the range of 0.692–0.848 kJ/kg.K (mainly  $\text{Al}_2\text{O}_3$  and  $\text{SiO}_2$ ) for the GEP technique, and the range of 1.004–1.160 kJ/kg.K (mainly MgO) for the GP and GMDH techniques. As can be observed in Fig. 7e concerning the “base-fluid specific heat” parameter, the lowest AAPRE values are found for all four techniques of the GRG, GP, GEP, and GMDH in the range of 2.210–2.632 kJ/kg.K (mainly EthyleneGlycol). It is worth noting that with a growth in the heat capacity values of the base-fluids, the AAPRE values computed for the correlation belonging to Çolak et al.<sup>28</sup> [i.e., Eq. (12)] have decreased.

Figure 8 illustrates a graphical analysis of cumulative frequency for evaluating the performance of the different correlations mentioned in the present study. According to this plot, the cumulative frequency of all data-points is depicted versus the percent relative errors to measure the number of data-points that the correlations can reliably estimate. From Fig. 8, it can be inferred that the GMDH technique surpasses the other techniques by successfully and effectively estimating more than 90% of total data-points with a relative error of  $-10\%$  to  $+5\%$ . Furthermore, the benefit and priority of the new correlations developed in this study are clearly visible compared to the other correlations.

### Outliers detection and applicability domain of a technique

A popular and effective method known as the “leverage statistical approach” is applied to find probable outliers (or unexpectedly aberrant data) and designate the applicability domain of the developed correlations<sup>78–80</sup>. For this purpose, this method generates a graph known as the “Williams plot” by determining two parameters, the Hat Matrix ( $H$ ) and the Standardized Residuals ( $SR$ ), utilizing the Eqs. (23) and (24)<sup>80–82</sup>:

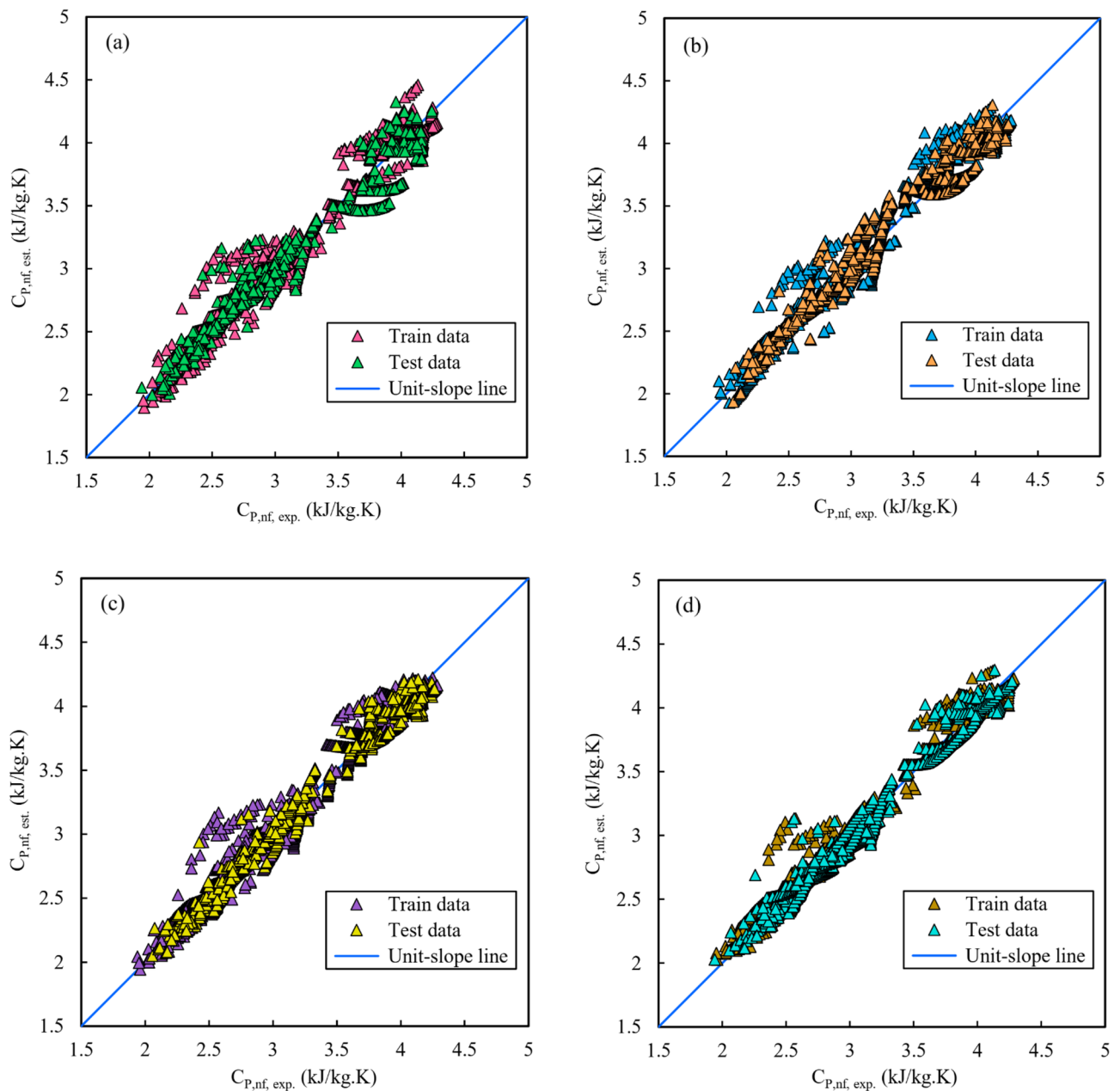
$$H = X(X^T X)^{-1} X^T \quad (23)$$

where the letter  $T$  indicates the transpose operator.  $X$  is a  $N_t \times N_i$  matrix, where  $N_t$  and  $N_i$  depict the number of all data-points and technique inputs, respectively.

$$SR_i = \frac{e_i}{RMSE \sqrt{1 - H_{ii}}} \quad (24)$$

where  $RMSE$  refers to the root mean square error of the used technique,  $e_i$  denotes the discrepancy of the  $i$ -th estimated data-point from its associated experimental value, and the Hat Indices  $H_{ii}$  represent the diagonal components of the  $H$  matrix (i.e.,  $H_{ii} = \text{diag}(H)$ ).

Williams plot visually displays three important zones associated with the existence of valid data, out-of-leverage data, and suspected data (outliers) by revealing the relation between Standardized Residuals and Hat Indices. Valid data-points can be found in the ranges  $-3 \leq SR \leq +3$  and  $0 \leq H \leq H^*$ . The leverage values equal to  $-3$  and  $+3$  for the parameter  $SR$  are called the cut-off metrics. The parameter  $H^*$ , also referred to as the critical leverage or warning leverage, has the value  $3(N_i + 1)/N_t$ . The data-points that belong to the ranges  $-3 \leq SR \leq +3$  and  $H > H^*$  are



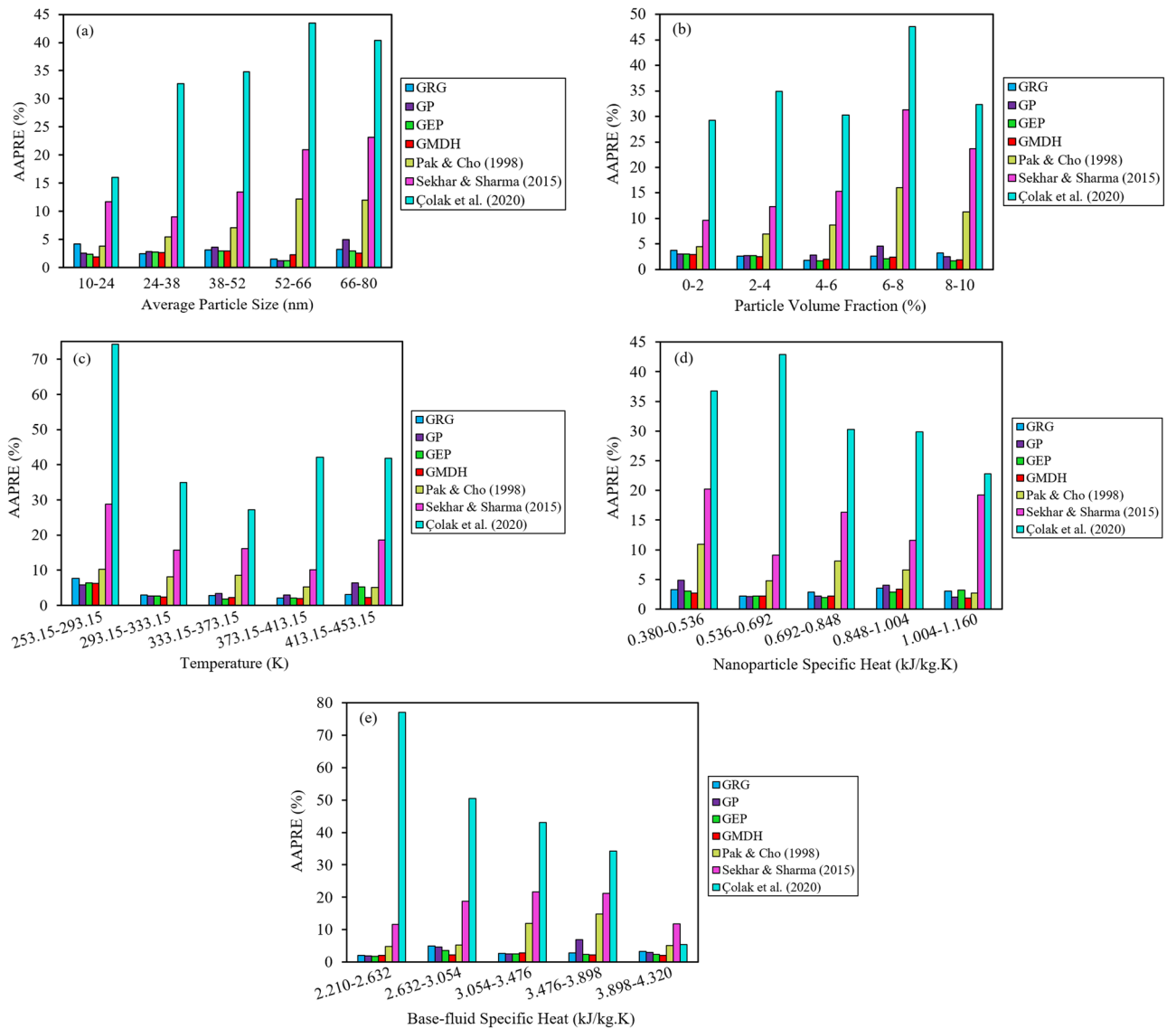
**Figure 6.** Cross-plots of the correlations developed for  $C_{P,nf}$  estimation: (a) GRG, (b) GP, (c) GEP, and (d) GMDH.

recognized as “good high leverage” points, which are outside the applicability domain of the technique by which a correlation has been generated. The rest of the data-points with  $SR < -3$  or  $SR > +3$  (independent of  $H$  values) are classified as suspected data or outliers. Due to their negative effect on the slope and intercept of the constructed regression line, they are considered a risk to reliable modeling and are also called “bad high leverage” points<sup>78–83</sup>.

Figure 9 exhibits the Williams plot for the best-proposed technique (i.e., GMDH) based on the total dataset employed in the current study. This graph shows that the GMDH technique detects 1.68% of the 2084 data-points (35 red triangular marks) as probable outliers. Furthermore, just 0.62% of the 2084 data-points (13 blue square marks) are found to be outside the applicability domain of the GMDH technique. As a result, due to the existence of a substantial portion of valid data-points (green circular marks) in the ranges  $-3 \leq SR \leq +3$  and  $0 \leq H \leq 0.0086$ , the accuracy and reliability of this technique are confirmed again.

### Sensitivity analysis

Performing sensitivity analysis can include several purposes, such as detecting errors existing in the model’s structure, determining the grade of relation between input and output variables of a model, and refining a model by discarding inputs that have little or no effect on its output. The relevancy factor ( $r$ ) is a parameter that is used



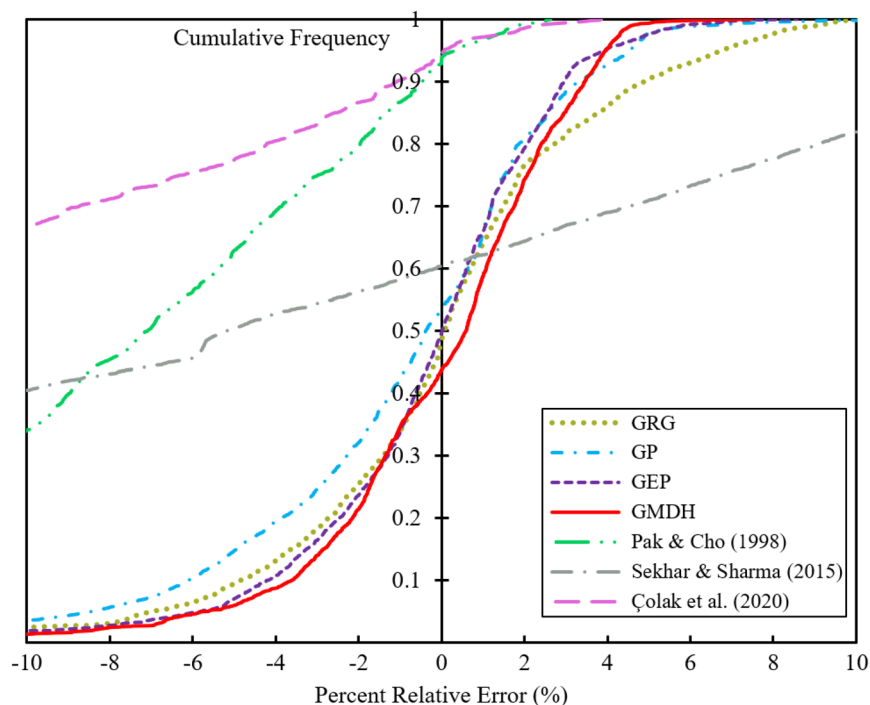
**Figure 7.** AAPRE distribution in different ranges of input variables for the suggested correlations.

to assess how much every input variable influences the output of a model. This factor's value can vary from  $-1$  to  $+1$ , and the greater its absolute value for an input variable, the higher its impact on the model's output<sup>30</sup>.

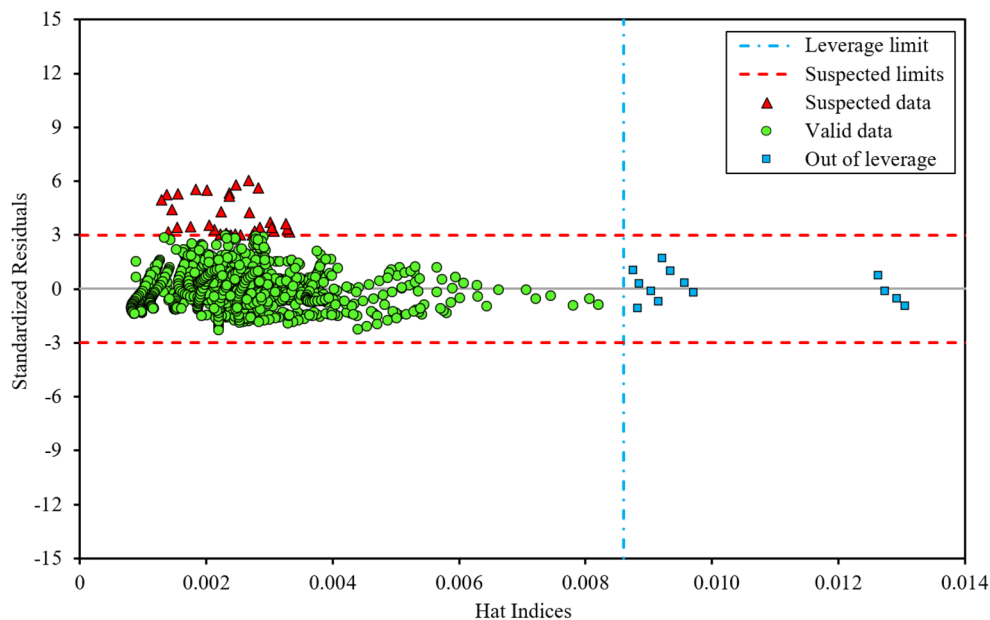
One of the ways to determine the relevancy factors is to use the *Correlation* option in the *Data Analysis* tools available in the *Data* tab of *Excel* software. The relevancy factors were calculated for every one of the techniques proposed in the present research, and similar results were discovered. Figure 10 represents the relevancy factors derived for each input variable according to the estimations of all techniques. Obviously, the parameters of temperature ( $T$ ) as well as specific heat capacities of nanoparticle ( $C_{p,np}$ ) and base-fluid ( $C_{p,bf}$ ) have a direct relationship with nanofluids' specific heat capacity ( $C_{p,nf}$ ), whereas two parameters of average particle size ( $d_{np}$ ) and volume fraction ( $\phi_v$ ) do not. Additionally, this figure indicates that the two parameters  $C_{p,bf}$  and  $d_{np}$  (with  $r$  factors of about  $+0.94$  and  $-0.42$ , respectively) have significant impacts on estimating the target variable  $C_{p,nf}$ . This issue was also discovered in the studies of Jamei et al.<sup>84,85</sup>, who investigated the specific heat capacity of various metal-oxide-based and carbon-based nanofluids dispersed in different base-fluids through the application of powerful machine learning techniques.

### Analyses of statistical trends

Based on the relevancy factors outlined in the prior section, the inverse influence of average particle size ( $d_{np}$ ) and volume fraction ( $\phi_v$ ) on the mono-nanofluids' specific heat capacity ( $C_{p,nf}$ ) was revealed. In addition, three other variables comprising temperature ( $T$ ), nanoparticle specific heat capacity ( $C_{p,np}$ ), and base-fluid specific heat capacity ( $C_{p,bf}$ ) had a direct proportional influence on the target variable  $C_{p,nf}$ . The graphical trend analysis shown in Fig. 11 makes this issue quite evident. Here, only some experimental data and their corresponding values estimated by the best technique (i.e., GMDH) are provided.



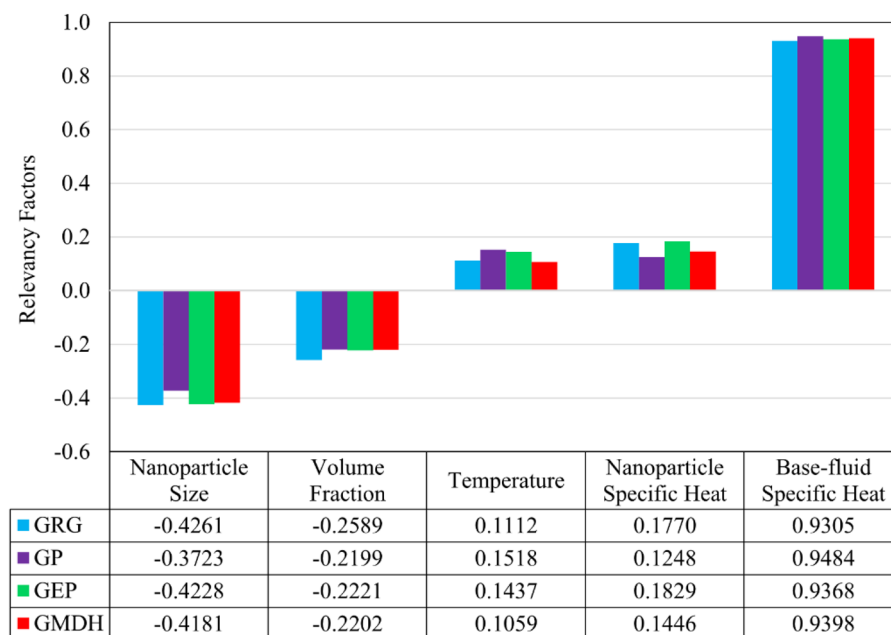
**Figure 8.** Cumulative frequency against percent relative error for suggested correlations.



**Figure 9.** Williams plot related to the correlation developed by the GMDH technique.

Figure 11a depicts the decreasing trend of mono-nanofluid heat capacity ( $C_{p,nf}$ ) with growing nanoparticle size ( $d_{np}$ ). This impact was also found in the investigations of Zhang et al.<sup>86</sup>, Angayarkanni et al.<sup>87</sup>, Xiong et al.<sup>88</sup>, Wang et al.<sup>89</sup>, and Novotny et al.<sup>90</sup>. The underlying cause of this matter can be assigned to the fact that smaller nanoparticles may have greater surface-to-volume ratios and atomic thermal vibrational energies<sup>86,88</sup>. However, it was discovered that the nanoparticle size had a negligible effect, according to the study conducted by Żyła et al.<sup>38</sup> on the specific heat capacity of nanofluids comprising AlN, TiN, and Si<sub>3</sub>N<sub>4</sub> dispersed in ethylene glycol. Such a finding was also reported in Satti et al.'s<sup>25</sup> research.

Based on the experimental results of Murshed<sup>31</sup>, Fig. 11b depicts that the values of mono-nanofluid heat capacity ( $C_{p,nf}$ ) reduce as the volume fraction of nanoparticles ( $\phi_v$ ) increases. Furthermore, Satti et al.<sup>25</sup> found that for small volume fractions (varying from 0.5% to 1.5%), the specific heat values of their examined nanofluids



**Figure 10.** Relevancy factors between input and output variables.

changed relatively little. In another study related to Raud et al.<sup>91</sup>, the specific heat values of TiO<sub>2</sub>/W and Al<sub>2</sub>O<sub>3</sub>/W nanofluids dropped linearly with an increase in the mass fraction of the utilized nanoparticles. The investigations of Raja et al.<sup>92</sup> also disclosed that the specific heat capacity of nanofluids containing copper, silver, and aluminum nanoparticles suspended in water or ethylene glycol reduced as the nanoparticles' concentration increased. The reason is that the solid nanoparticles have lower specific heat capacities in comparison with the base-fluids. Therefore, the more nanoparticles are dissolved in a certain amount of base-fluid, the more the specific heat capacity of the resultant nanofluid will decrease<sup>17,93</sup>. Similar outcomes were also attained in this field by Maghrabie et al.<sup>94</sup>, Moldoveanu and Minea<sup>27</sup>, Verma et al.<sup>35</sup>, and Zhou and Ni<sup>76</sup>.

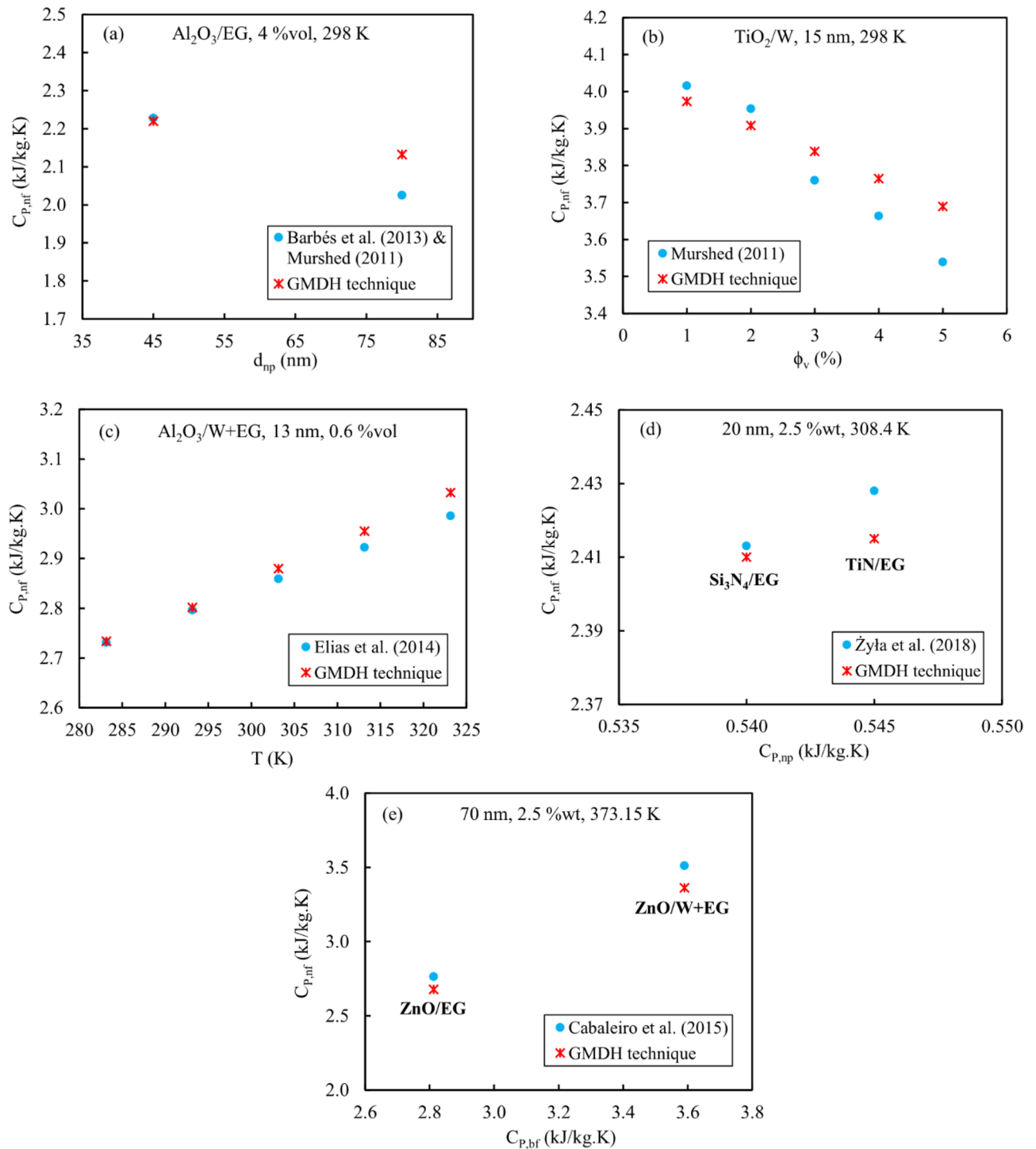
In Fig. 11c, it is demonstrated how the specific heat value of the mono-nanofluids ( $C_{p,nf}$ ) varies with temperature ( $T$ ) in such a way that there is a direct relationship between these two variables. As the temperature of the nanofluid rises, its ability to absorb and store thermal energy (without undergoing a phase transition) and as a result, the specific heat values increase. This behavior has been clearly noted in numerous studies, including Gao et al.<sup>29</sup>, Wole-Osho et al.<sup>39</sup>, Raud et al.<sup>91</sup>, Popa et al.<sup>26</sup>, Angayarkanni et al.<sup>87</sup>, Elias et al.<sup>33</sup>, Barbés et al.<sup>21,22</sup>, Heyhat et al.<sup>95</sup>, and Vajjha and Das<sup>17,19,20</sup>.

Figures 11d and e illustrate the trend of mono-nanofluids' specific heat changes against the specific heat capacities of nanoparticles and base-fluids, respectively. The direct relationship of  $C_{p,nf}$  variations with  $C_{p,np}$  and  $C_{p,bf}$  is presented for some experimental data. According to Fig. 11d, TiN-containing nanofluids have a greater specific heat capacity than those containing Si<sub>3</sub>N<sub>4</sub> (taken from the experimental data of Żyła et al.<sup>38</sup>) due to the higher specific heat value of TiN nanoparticles, when considering a certain base-fluid (in this case, ethylene glycol) and keeping other parameters constant. Such an effect was also reported in the experimental study of Vijayakumar et al.<sup>37</sup> regarding the difference in specific heat values of CuO/W and Al<sub>2</sub>O<sub>3</sub>/W nanofluids. As seen in Fig. 11e concerning the experimental data of Cabaleiro et al.<sup>23</sup>, since the specific heat value of the water-ethylene glycol mixture is greater than that of the ethylene glycol alone, the specific heat capacity of nanofluids comprising ZnO suspended in the W+EG mixture has been raised. Moreover, Akilu et al.<sup>36</sup> came to a similar conclusion by evaluating the specific heat changes of SiO<sub>2</sub> nanoparticles dissolved in three different base-fluids (ethylene glycol, glycerol, and ethylene glycol+glycerol). According to Fig. 11e, the base-fluid type has a considerable effect on the specific heat capacity of a nanofluid. As previously declared, the relevancy factor corresponding to the variable  $C_{p,bf}$  had the highest value (+0.94).

## Conclusions

Four soft-computing techniques were implemented on a database including 2084 data-points taken from 19 experimental studies in order to estimate the specific heat capacity of diverse mono-nanofluids ( $C_{p,nf}$ ) as a function of five input variables ( $d_{np}$ ,  $\varphi_v$ ,  $T$ ,  $C_{p,np}$ , and  $C_{p,bf}$ ). Then, four distinct high-accuracy correlations were developed, and numerous statistical and graphical error analyses were also conducted. The performance of these novel correlations was compared to the outcomes of three earlier correlations published by other researchers. The main advantage of these novel empirical correlations is that they have significantly reduced the constraints of prior correlations. As a result, they are adaptable to various oxide-based mono-nanofluids for a broad range of independent variable values.

According to the outcomes of each innovative and robust correlation developed in this research, the utilized techniques can be rated based on their accuracy as GMDH > GEP > GP > GRG. The results revealed



**Figure 11.** Variations of mono-nanofluids' specific heat capacity with any input parameters.

that the correlation achieved using the GMDH technique is the most effective and appropriate one (with AAPRE = 2.4163% and  $R^2 = 0.9743$ ). However, the GEP technique has also exhibited acceptable behavior that is comparable. By applying the leverage statistical approach for the GMDH technique and distinguishing valid data, out-of-leverage data, and suspected data (outliers), the reliability and accuracy of this technique were confirmed. To recognize the relation between input and output variables available in the entire dataset, a sensitivity analysis was performed and the relevancy factors were determined for the best technique (GMDH). It was found that average nanoparticle size and base-fluid specific heat capacity are significant and influencing factors in establishing the specific heat values of mono-nanofluids. Ultimately, the variation trend of the target variable ( $C_{p,nf}$ ) for each of the input variables was graphically evaluated using the experimental data obtained from the previous studies and the estimated data generated by the GMDH technique. The findings of this study agree



with many experimental measurements and computer-based surveys in calculating the specific heat capacity of mono-nanofluids.

## Data availability

All data have been gathered from the literature. All references used for extracting the required data have been cited in the text. However, the data will be available from the corresponding author upon reasonable request.

Received: 8 July 2023; Accepted: 12 November 2023

Published online: 25 November 2023

## References

- Khan, I., Saeed, K. & Khan, I. Nanoparticles: Properties, applications and toxicities. *Arab. J. Chem.* **12**(7), 908–931. <https://doi.org/10.1016/j.arabjc.2017.05.011> (2019).
- Adun, H., Wole-Osho, I., Okonkwo, E. C., Kavaz, D. & Dagbasi, M. A critical review of specific heat capacity of hybrid nanofluids for thermal energy applications. *J. Mol. Liq.* **340**, 116890. <https://doi.org/10.1016/j.molliq.2021.116890> (2021).
- Menni, Y., Chamkha, A. J. & Ameer, H. Advances of nanofluids in heat exchangers: A review. *Heat Transf.* **49**(8), 4321–4349. <https://doi.org/10.1002/htj.21829> (2020).
- Xiong, Q. *et al.* State-of-the-art review of nanofluids in solar collectors: A review based on the type of the dispersed nanoparticles. *J. Clean. Prod.* **310**, 127528. <https://doi.org/10.1016/j.jclepro.2021.127528> (2021).
- Hemmat Esfe, M., Kamyab, M. H. & Valadkhani, M. Application of nanofluids and fluids in photovoltaic thermal system: An updated review. *Solar Energy* **199**, 796–818. <https://doi.org/10.1016/j.solener.2020.01.015> (2020).
- Xian, H. W., Sidik, N. A. C. & Najafi, G. Recent state of nanofluid in automobile cooling systems. *J. Therm. Anal. Calorim.* **135**(2), 981–1008. <https://doi.org/10.1007/s10973-018-7477-3> (2019).
- Balaji, T., Selvam, C. & Mohan Lal, D. A review on electronics cooling using nanofluids. in *IOP Conference Series: Materials Science and Engineering, International Conference on Advances in Renewable and Sustainable Energy Systems (ICARSES 2020)*, vol. 1130, 012007 (2021). <https://doi.org/10.1088/1757-899X/1130/1/012007>.
- Siddiqui, O. K., Shams, A. & Al-Athel, K. A comprehensive review on the use of nanoparticles in nuclear power plants. *Arab. J. Sci. Eng.* **1**, 1–25. <https://doi.org/10.1007/s13369-023-07954-y> (2023).
- Sheikhpour, M., Arabi, M., Kasaean, A., RoknRabei, A. & Taherian, Z. Role of nanofluids in drug delivery and biomedical technology: Methods and applications. *Nanotechnol. Sci. Appl.* **13**, 47–59. <https://doi.org/10.2147/nsa.s260374> (2020).
- Sun, Y. *et al.* Properties of nanofluids and their applications in enhanced oil recovery: A comprehensive review. *Energy Fuels* **34**(2), 1202–1218. <https://doi.org/10.1021/acs.energyfuels.9b03501> (2020).
- Hemmat Esfe, M., Bahiraei, M. & Mir, A. Application of conventional and hybrid nanofluids in different machining processes: A critical review. *Adv. Colloid Interface Sci.* **282**, 102199. <https://doi.org/10.1016/j.cis.2020.102199> (2020).
- Jia, Y., Ran, F., Zhu, C. & Fang, G. Numerical analysis of photovoltaic-thermal collector using nanofluid as a coolant. *Solar Energy* **196**, 625–636. <https://doi.org/10.1016/j.solener.2019.12.069> (2020).
- McCash, L. B., Akhtar, S., Nadeem, S. & Saleem, S. Entropy analysis of the peristaltic flow of hybrid nanofluid inside an elliptic duct with sinusoidally advancing boundaries. *Entropy* **23**(6), 732 (2021).
- Shahzad, M. H., Awan, A. U., Akhtar, S. & Nadeem, S. Entropy and stability analysis on blood flow with nanoparticles through a stenosed artery having permeable walls. *Sci. Progress* **105**(2), 1–34. <https://doi.org/10.1177/00368504221096000> (2022).
- Alghamdi, M., Akbar, N. S., Hussain, M. F., Akhtar, S. & Muhammad, T. Thermodynamic study of hybrid nanofluid to explore synergistic effects of multiple ferromagnetic nanoparticles in co-axial disks for magnetized fluid. *Tribol. Int.* **188**, 108867. <https://doi.org/10.1016/j.triboint.2023.108867> (2023).
- Baig, M. N. J. *et al.* Exact analytical solutions of stagnation point flow over a heated stretching cylinder: A phase flow nanofluid model. *Chin. J. Phys.* **86**, 1–11. <https://doi.org/10.1016/j.cjph.2023.03.017> (2023).
- Vajjha, R. S. & Das, D. K. Measurements of specific heat and density of Al<sub>2</sub>O<sub>3</sub> nanofluid. in *American Institute of Physics Conference Proceedings*, vol. 1063, no. 1 (2008). <https://doi.org/10.1063/1.3027181>.
- Pak, B. C. & Cho, Y. I. Hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. *Exp. Heat Transf.* **11**(2), 151–170. <https://doi.org/10.1080/08916159808946559> (1998).
- Vajjha, R. S. & Das, D. K. Specific heat measurement of three nanofluids and development of new correlations. *ASME J. Heat Mass Transf.* **131**(7), 071601. <https://doi.org/10.1115/1.3090813> (2009).
- Vajjha, R. S. & Das, D. K. A review and analysis on influence of temperature and concentration of nanofluids on thermophysical properties, heat transfer and pumping power. *Int. J. Heat Mass Transf.* **55**(15–16), 4063–4078. <https://doi.org/10.1016/j.ijheatmasstransfer.2012.03.048> (2012).
- Barbés, B. *et al.* Thermal conductivity and specific heat capacity measurements of Al<sub>2</sub>O<sub>3</sub> nanofluids. *J. Therm. Anal. Calorim.* **111**, 1615–1625. <https://doi.org/10.1007/s10973-012-2534-9> (2013).
- Barbés, B., Páramo, R., Blanco, E. & Casanova, C. Thermal conductivity and specific heat capacity measurements of CuO nanofluids. *J. Therm. Anal. Calorim.* **115**, 1883–1891. <https://doi.org/10.1007/s10973-013-3518-0> (2014).
- Cabaleiro, D., Gracia-Fernández, C., Legido, J. L. & Lugo, L. Specific heat of metal oxide nanofluids at high concentrations for heat transfer. *Int. J. Heat Mass Transf.* **88**, 872–879. <https://doi.org/10.1016/j.ijheatmasstransfer.2015.04.107> (2015).
- Sekhar, Y. R. & Sharma, K. V. Study of viscosity and specific heat capacity characteristics of water-based Al<sub>2</sub>O<sub>3</sub> nanofluids at low particle concentrations. *J. Exp. Nanosci.* **10**(2), 86–102. <https://doi.org/10.1080/17458080.2013.796595> (2015).
- Satti, J. R., Das, D. K. & Ray, D. Specific heat measurements of five different propylene glycol based nanofluids and development of a new correlation. *Int. J. Heat Mass Transf.* **94**, 343–353. <https://doi.org/10.1016/j.ijheatmasstransfer.2015.11.065> (2016).
- Popa, C. V., Nguyen, C. T. & Gherasim, I. New specific heat data for Al<sub>2</sub>O<sub>3</sub> and CuO nanoparticles in suspension in water and ethylene glycol. *Int. J. Therm. Sci.* **111**, 108–115. <https://doi.org/10.1016/j.ijthermalsci.2016.08.016> (2017).
- Moldoveanu, G. M. & Minea, A. A. Specific heat experimental tests of simple and hybrid oxide-water nanofluids: Proposing new correlation. *J. Mol. Liq.* **279**, 299–305. <https://doi.org/10.1016/j.molliq.2019.01.137> (2019).
- Çolak, A. B., Yıldız, O., Bayrak, M. & Tezekici, B. S. Experimental study for predicting the specific heat of water based Cu-Al<sub>2</sub>O<sub>3</sub> hybrid nanofluid using artificial neural network and proposing new correlation. *Int. J. Energy Res.* **44**(9), 7198–7215. <https://doi.org/10.1002/er.5417> (2020).
- Gao, Y. *et al.* Experimental investigation of specific heat of aqueous graphene oxide Al<sub>2</sub>O<sub>3</sub> hybrid nanofluid. *Therm. Sci.* **25**(1), 515–525. <https://doi.org/10.2298/TSC190404381G> (2021).
- Deymi, O. *et al.* Employing ensemble learning techniques for modeling nanofluids' specific heat capacity. *Int. Commun. Heat Mass Transf.* **143**, 106684. <https://doi.org/10.1016/j.icheatmasstransfer.2023.106684> (2023).
- Murshed, S. M. S. Determination of effective specific heat of nanofluids. *J. Exp. Nanosci.* **6**(5), 539–546. <https://doi.org/10.1080/17458080.2010.498838> (2011).

32. Teng, T. P. & Hung, Y. H. Estimation and experimental study of the density and specific heat for alumina nanofluid. *J. Exp. Nanosci.* **9**(7), 707–718. <https://doi.org/10.1080/17458080.2012.696219> (2014).
33. Elias, M. M. *et al.* Experimental investigation on the thermo-physical properties of Al<sub>2</sub>O<sub>3</sub> nanoparticles suspended in car radiator coolant. *Int. Commun. Heat Mass Transf.* **54**, 48–53. <https://doi.org/10.1016/j.icheatmasstransfer.2014.03.005> (2014).
34. Kadhim, Z. K., Kassim, M. S. & Abdul Hassan, A. Y. Effect of MgO nanofluid on heat transfer characteristics for integral finned tube heat exchanger. *Int. J. Mech. Eng. Technol.* **7**(2), 11–24 (2016).
35. Verma, S. K., Tiwari, A. K. & Chauhan, D. S. Experimental evaluation of flat plate solar collector using nanofluids. *Energy Convers. Manag.* **134**, 103–115. <https://doi.org/10.1016/j.enconman.2016.12.037> (2017).
36. Akilu, S., Baheta, A. T., Sharma, K. V. & Said, M. A. Experimental determination of nanofluid specific heat with SiO<sub>2</sub> nanoparticles in different base-fluids. in *4th International Conference on the Advancement of Materials and Nanotechnology (ICAMN IV 2016)*, vol. 1877, no. 1, 090001. (2017). <https://doi.org/10.1063/1.4999896>.
37. Vijayakumar, M., Navaneethkrishnan, P., Kumaresan, G. & Kamatchi, R. A study on heat transfer characteristics of inclined copper sintered wick heat pipe using surfactant free CuO and Al<sub>2</sub>O<sub>3</sub> nanofluids. *J. Taiwan Inst. Chem. Eng.* **81**, 190–198. <https://doi.org/10.1016/j.jtice.2017.10.032> (2017).
38. Žyła, G., Vallejo, J. P. & Lugo, L. Isobaric heat capacity and density of ethylene glycol-based nanofluids containing various nitride nanoparticle types: An experimental study. *J. Mol. Liq.* **261**, 530–539. <https://doi.org/10.1016/j.molliq.2018.04.012> (2018).
39. Wole-Osho, I., Okonkwo, E. C., Kavaz, D. & Abbasoglu, S. An experimental investigation into the effect of particle mixture ratio on specific heat capacity and dynamic viscosity of Al<sub>2</sub>O<sub>3</sub>-ZnO hybrid nanofluids. *Powder Technol.* **363**, 699–716. <https://doi.org/10.1016/j.powtec.2020.01.015> (2020).
40. Choudhary, S., Sachdeva, A. & Kumar, P. Investigation of the stability of MgO nanofluid and its effect on the thermal performance of flat plate solar collector. *Renew. Energy* **147**(1), 1801–1814. <https://doi.org/10.1016/j.renene.2019.09.126> (2020).
41. Akram, N. *et al.* Experimental investigations of the performance of a flat-plate solar collector using carbon and metal oxides based nanofluids. *Energy* **227**, 120452. <https://doi.org/10.1016/j.energy.2021.120452> (2021).
42. Salameh, T. *et al.* Fuzzy modeling and particle swarm optimization of Al<sub>2</sub>O<sub>3</sub>/SiO<sub>2</sub> nanofluid. *Int. J. Thermofluids* **10**, 100084. <https://doi.org/10.1016/j.ijft.2021.100084> (2021).
43. Kristinsson, K. & Dumont, G. A. System identification and control using genetic algorithms. *IEEE Trans. Syst. Man Cybern.* **22**(5), 1033–1046. <https://doi.org/10.1109/21.179842> (1992).
44. Rodríguez-Vázquez, K. Multiobjective evolutionary algorithms in non-linear system identification. Ph.D. thesis, Department of Automatic Control and Systems Engineering, The University of Sheffield (1999).
45. Sanchez, E., Shibata, T. & Asker Zadeh, L. *Genetic Algorithms and Fuzzy Logic Systems: Soft-computing Perspectives* (World Scientific, 1997).
46. Söderström, T. & Stoica, P. Instrumental variable methods for system identification. *Circuits Syst. Signal Process.* **21**, 1–9. <https://doi.org/10.1007/BF01211647> (2002).
47. Gupta, D., Hazarika, B. B., Berlin, M., Sharma, U. M. & Mishra, K. Artificial intelligence for suspended sediment load prediction: A review. *Environ. Earth Sci.* **80**(9), 346. <https://doi.org/10.1007/s12665-021-09625-3> (2021).
48. Loni, R., Askari Asli-Ardeh, E., Ghobadian, B., Ahmadi, M. H. & Bellos, E. GMDH modeling and experimental investigation of thermal performance enhancement of hemispherical cavity receiver using MWCNT/oil nanofluid. *Solar Energy* **171**, 790–803. <https://doi.org/10.1016/j.solener.2018.07.003> (2018).
49. Vakili, M. & Salehi, S. A. A review of recent developments in the application of machine learning in solar thermal collector modeling. *Environ. Sci. Pollut. Res.* **30**(2), 2406–2439. <https://doi.org/10.1007/s11356-022-24044-y> (2023).
50. Hemmat Esfe, M., Tatar, A., Hassani Ahangar, M. R. & Rostamian, H. A comparison of performance of several artificial intelligence methods for predicting the dynamic viscosity of TiO<sub>2</sub>/SAE 50 nano-lubricant. *Physica E* **96**, 85–93. <https://doi.org/10.1016/j.physe.2017.08.019> (2018).
51. Rostami, A., Arabloo, M. & Ebadi, H. Genetic programming (GP) approach for prediction of supercritical CO<sub>2</sub> thermal conductivity. *Chem. Eng. Res. Des.* **122**, 164–175. <https://doi.org/10.1016/j.cherd.2017.02.028> (2017).
52. Algdamsi, H. *et al.* Consistent black oil PVT table modification using the generalized reduced gradient method and constrained cubic spline for variable bubble-point simulation. in *International Petroleum Technology Conference* (2020). <https://doi.org/10.2523/IPTC-20278-Abstract>.
53. Osman, A. I. A. *et al.* Past, present and perspective methodology for groundwater modeling-based machine learning approaches. *Arch. Comput. Methods Eng.* **29**(6), 3843–3859. <https://doi.org/10.1007/s11831-022-09715-w> (2022).
54. TajikJamal-Abadi, M. & Zamzamin, A. H. Optimization of thermal conductivity of Al<sub>2</sub>O<sub>3</sub> nanofluid by using ANN and GRG methods. *Int. J. Nanosci. Nanotechnol.* **9**(4), 177–184 (2013).
55. Lasdon, L. S., Fox, R. L. & Ratner, M. W. Nonlinear optimization using the generalized reduced gradient method. *Rev. Franç. d'Autom. Inform. Recherche Opér.* **8**(3), 73–103 (1974).
56. Lasdon, L. S., Waren, A. D., Jain, A. & Ratner, M. Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Trans. Math. Softw.* **4**(1), 34–50. <https://doi.org/10.1145/355769.355773> (1978).
57. Banzhaf, W. Artificial intelligence: Genetic programming. *Int. Encyclop. Soc. Behav. Sci.* **1**, 789–792. <https://doi.org/10.1016/B0-08-043076-7/00557-X> (2001).
58. Poli, R. & Koza, J. Genetic programming, in *Search methodologies: Introductory tutorials in optimization and decision support techniques*, 2nd ed., 143–185 (Springer, 2014).
59. Whigham, P. A. Grammatically-based genetic programming. *Proc. Workshop Genet. Program. Theory Real World Appl.* **16**(3), 33–41 (1995).
60. Spector, L., Langdon, W. B., O'Reilly, U. M. & Angeline, P. J. *Advances in Genetic Programming* (MIT Press, 1999).
61. Wong, M. L., Leung, K. S. & Cheng, J. C. Y. Discovering knowledge from noisy databases using genetic programming. *J. Am. Soc. Inf. Sci.* **51**(9), 870–881. [https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:9%3C870::AID-AS190%3E3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-4571(2000)51:9%3C870::AID-AS190%3E3.0.CO;2-R) (2000).
62. Tackett, W. A. Genetic programming for feature discovery and image discrimination. in *Proceedings of the 5th International Conference on Genetic Algorithms*, 303–311 (1993).
63. Gruau, F. Neural network synthesis using cellular encoding and the genetic algorithm. Ph.D. thesis, l'Ecole Normale Supérieure de Lyon (1994).
64. Ferreira, C. Gene expression programming: A new adaptive algorithm for solving problems. *Compl. Syst.* **13**(2), 87–129. <https://doi.org/10.48550/arXiv.cs/0102027> (2001).
65. Menad, N. A. *et al.* Modeling temperature dependency of oil-water relative permeability in thermal enhanced oil recovery processes using group method of data handling and gene expression programming. *Eng. Appl. Comput. Fluid Mech.* **13**(1), 724–743. <https://doi.org/10.1080/19942060.2019.1639549> (2019).
66. Ferreira, C. Gene expression programming in problem solving, in *Soft Computing and Industry: Recent Applications*, 635–653 (Springer, 2002).
67. Rostami, A., Hemmati-Sarapardeh, A. & Mohammadi, A. H. Estimating n-tetradecane/bitumen mixture viscosity in solvent-assisted oil recovery process using GEP and GMDH modeling approaches. *Pet. Sci. Technol.* **37**(14), 1640–1647. <https://doi.org/10.1080/10916466.2018.1531885> (2019).

68. Rostami, A., Kamari, A., Panacharoensawad, E. & Hashemi, A. New empirical correlations for determination of minimum miscibility pressure (MMP) during N<sub>2</sub>-contaminated lean gas flooding. *J. Taiwan Inst. Chem. Eng.* **91**, 369–382. <https://doi.org/10.1016/j.jtice.2018.05.048> (2018).
69. Zhong, J., Feng, L. & Ong, Y. S. Gene expression programming: A survey. *IEEE Comput. Intell. Mag.* **12**(3), 54–72. <https://doi.org/10.1109/MCI.2017.2708618> (2017).
70. Hemmati-Sarapardeh, A. *et al.* Modeling natural gas compressibility factor using a hybrid group method of data handling. *Eng. Appl. Comput. Fluid Mech.* **14**(1), 27–37. <https://doi.org/10.1080/19942060.2019.1679668> (2020).
71. Hemmati-Sarapardeh, A., Dabir, B., Ahmadi, M., Mohammadi, A. H. & Husein, M. M. Modelling asphaltene precipitation titration data: A committee of machines and a group method of data handling. *Can. J. Chem. Eng.* **97**(2), 431–441. <https://doi.org/10.1002/cjce.23254> (2019).
72. Huang, W. *et al.* Application of modified GMDH network for CO<sub>2</sub>-oil minimum miscibility pressure prediction. *Energy Sources A* **42**(16), 2049–2062. <https://doi.org/10.1080/15567036.2019.1604908> (2020).
73. Rostami, A. *et al.* Modeling heat capacity of ionic liquids using group method of data handling: A hybrid and structure-based approach. *Int. J. Heat Mass Transf.* **129**, 7–17. <https://doi.org/10.1016/j.ijheatmasstransfer.2018.09.057> (2019).
74. Mahdaviara, M., Menad, N. A., Ghazanfari, M. H. & Hemmati-Sarapardeh, A. Modeling relative permeability of gas condensate reservoirs: Advanced computational frameworks. *J. Pet. Sci. Eng.* **189**, 106929. <https://doi.org/10.1016/j.petrol.2020.106929> (2020).
75. Xuan, Y. & Roetzel, W. Conceptions for heat transfer correlation of nanofluids. *Int. J. Heat Mass Transf.* **43**(19), 3701–3707. [https://doi.org/10.1016/S0017-9310\(99\)00369-5](https://doi.org/10.1016/S0017-9310(99)00369-5) (2000).
76. Zhou, S. Q. & Ni, R. Measurement of the specific heat capacity of water-based Al<sub>2</sub>O<sub>3</sub> nanofluid. *Appl. Phys. Lett.* **92**(9), 093123. <https://doi.org/10.1063/1.2890431> (2008).
77. Zhou, L. P., Wang, B. X., Peng, X. F., Du, X. Z. & Yang, Y. P. On the specific heat capacity of CuO nanofluid. *Adv. Mech. Eng.* **2**, 172085. <https://doi.org/10.1155/2010/172085> (2010).
78. Goodall, C. R. Computation using the QR decomposition. *Handb. Stat. Compu. Stat.* **9**, 467–508 (1993).
79. Rousseeuw, P. J. & Leroy, A. M. *Robust Regression and Outlier Detection* (Wiley, 2005).
80. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Combin. Sci.* **26**(5), 694–701. <https://doi.org/10.1002/qsar.200610151> (2007).
81. Roy, K., Kar, S. & Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemometr. Intell. Lab. Syst.* **145**, 22–29. <https://doi.org/10.1016/j.chemolab.2015.04.013> (2015).
82. Hemmati-Sarapardeh, A., Larestani, A., Menad, N. A. & Hajirezaie, S. *Applications of Artificial Intelligence Techniques in the Petroleum Industry* (Gulf Professional Publishing, 2020).
83. Mohammadi, A. H., EslamiManesh, A., Gharagheizi, F. & Richon, D. A novel method for evaluation of asphaltene precipitation titration data. *Chem. Eng. Sci.* **78**, 181–185. <https://doi.org/10.1016/j.ces.2012.05.009> (2012).
84. Jamei, M. *et al.* On the specific heat capacity estimation of metal oxide-based nanofluid for energy perspective: A comprehensive assessment of data analysis techniques. *Int. Commun. Heat Mass Transf.* **123**, 105217. <https://doi.org/10.1016/j.icheatmasstransfer.2021.105217> (2021).
85. Jamei, M., AhmadianFar, I., Olumegbon, I. A., Karbasi, M. & Asadi, A. On the assessment of specific heat capacity of nanofluids for solar energy applications: Application of Gaussian process regression (GPR) approach. *J. Energy Storage* **33**, 102067. <https://doi.org/10.1016/j.est.2020.102067> (2021).
86. Zhang, L., Zhang, A., Jing, Y., Qu, P. & Wu, Z. Effect of particle size on the heat transfer performance of SiO<sub>2</sub>-Water nanofluids. *J. Phys. Chem. C* **125**(24), 13590–13600. <https://doi.org/10.1021/acs.jpcc.1c02014> (2021).
87. Angayarkanni, S. A., Sunny, V. & Philip, J. Effect of nanoparticle size, morphology and concentration on specific heat capacity and thermal conductivity of nanofluids. *J. Nanofluids* **4**(3), 302–309. <https://doi.org/10.1166/jon.2015.1167> (2015).
88. Xiong, S. *et al.* Universal relation for size dependent thermodynamic properties of metallic nanoparticles. *Phys. Chem. Chem. Phys.* **13**(22), 10652–10660. <https://doi.org/10.1039/C0CP90161J> (2011).
89. Wang, B. X., Zhou, L. P. & Peng, X. F. Surface and size effects on the specific heat capacity of nanoparticles. *Int. J. Thermophys.* **27**, 139–151. <https://doi.org/10.1007/s10765-006-0022-9> (2006).
90. Novotny, V., Meincke, P. P. M. & Watson, J. H. P. Effect of size and surface on the specific heat of small lead particles. *Phys. Rev. Lett.* **28**(14), 901. <https://doi.org/10.1103/PhysRevLett.28.901> (1972).
91. Raud, R., Hosterman, B., Diana, A., Steinberg, T. A. & Will, G. Experimental study of the interactivity, specific heat, and latent heat of fusion of water-based nanofluids. *Appl. Therm. Eng.* **117**, 164–168. <https://doi.org/10.1016/j.applthermaleng.2017.02.033> (2017).
92. Raja, R. A., Sunil, J. & Maheswaran, R. Estimation of thermo-physical properties of nanofluids using theoretical correlations. *Int. J. Appl. Eng. Res.* **13**(10), 7950–7953 (2018).
93. Gamal, M., Radwan, M. S., Elgizawy, I. G. & Shedid, M. H. Experimental studies on thermophysical properties of ethylene glycol/water-based MgO nanofluids. *J. Phys.* **2299**(1), 012022. <https://doi.org/10.1088/1742-6596/2299/1/012022> (2022).
94. Maghrabie, H. M., Attalla, M. & Mohsen, A. A. Performance assessment of a shell and helically coiled tube heat exchanger with variable orientations utilizing different nanofluids. *Appl. Therm. Eng.* **182**, 116013. <https://doi.org/10.1016/j.applthermaleng.2020.116013> (2021).
95. Heyhat, M. M., Kowsary, F., Rashidi, A. M., Momenpour, M. H. & Amrollahi, A. Experimental investigation of laminar convective heat transfer and pressure drop of water-based Al<sub>2</sub>O<sub>3</sub> nanofluids in fully-developed flow regime. *Exp. Therm. Fluid Sci.* **44**, 483–489. <https://doi.org/10.1016/j.expthermflusci.2012.08.009> (2013).

## Author contributions

O.D.: writing-original draft, visualization, data curation. F.H.: writing-original draft, visualization, data curation, software. S.A.: supervision, conceptualization, methodology, reviewing and editing, investigation. D.N.: methodology, conceptualization, visualization. M.A.A.: methodology, investigation, data curation. A.H.S.: supervision, conceptualization, reviewing and editing, methodology, investigation. A.M.: supervision, conceptualization, reviewing and editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47327-x>.

**Correspondence** and requests for materials should be addressed to S.A., A.H.-S. or A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023