

Predicting thermophysical properties enhancement of metal-based phase change materials using various machine learning algorithms

Mohsen Bakouri^{a,b}, Hakim S. Sultan^c, Sarminah Samad^d, Hussein Togun^e, Marjan Goodarzi^{f,g,*}

^a Department of Medical Equipment Technology, College of Applied Medical Science, Majmaah University, Majmaah City 11952, Saudi Arabia

^b Department of Physics, College of Arts, Fezzan University, Traghan 71340, Libya

^c College of Engineering, University of Warith Al-Anbiyaa, Karbala 56001, Iraq

^d Department of Business Administration, College of Business and Administration, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

^e Department of Biomedical Engineering, College of Engineering, University of Thi-Qar, Thi-Qar 64001, Iraq

^f Department of Mechanical Engineering, Lamar University, Beaumont, TX, United States

^g Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

ARTICLE INFO

Keywords:

Thermophysical properties

Prediction

Metal-based phase change materials

Machine learning algorithms

ABSTRACT

Background: In this research, five machine learning methods are employed to create the formula and make the model for communication between the melting and solidification processes to find the latent heat and communication between the thermal conductivity and the thermal diffusivity. Paraffin wax was the PCM, while SiO₂, ZnO, Fe₂O₃, and Al₂O₃ nanoparticles were utilized as additives to enhance the thermal conductivity by considering four concentrations between 2 and 8 wt.%.

Method: The K-Nearest Neighbours (KNN), the decision tree, the stochastic gradient descent (SGD) regressor, the Support Vector Machines (SVM), and the Huber regression are the algorithms that are used in this study.

Findings: Using the Huber regression method, the formula for the latent heat of the solidification-process is created by the melting-process (by latent heat and wt.) with the polynomial form by degree 2, and the R-Squared value is 0.998, and MSE is 0.971172, and by the KNN ($K = 3$) is made the model with the linear form. The R² value is 0.999, and MSE is 0.41809, and for the latent heat of the melting-process by the solidification-process (by latent heat and wt.) with the same methods reached to best the R² for creating the formula and making the model, respectively the R² values are 0.996 and 0.999, and the MSEs are 1.482458 and 0.608658. Using the Huber regression method, reach the best result for creating the formula in the polynomial form by degree 2 for the thermal conductivity by the thermal diffusivity and wt. and create the formula with the R² value of 0.998 and MSE is 0.000033, and make the model by the SVM with the R² value of 0.999 in the linear form, and MSE is 0.000016. The same methods are used for thermal diffusivity by the thermal conductivity and wt.; the R² values for the formula and the model, respectively, are 0.997 in polynomial form (degree = 2) and 0.999 in linear form. MSEs, respectively, are 0.000020 and 0.000006.

1. Introduction

Designing approaches for storing energy, such as thermal energy, is required in a world faced with energy crises. One of the most applicable methods is the one that resorts to the latent heat hidden in materials during their phase change process. Today, these phase change materials (PCMs) have practical applications in thermal systems such as photovoltaic panels [1], heat exchangers [2], solar energy instruments [3], buildings [4], solar reservoirs [5], etc. From another viewpoint, PCMs have the ability to be utilized in the cooling processes by releasing their

heat in an infinite medium (cool air at night) under the solidifying process, then storing the heat from the target media (warm air at day) under the melting-process [6,7]. The highest efficiency will be harvested if the most appropriate PCM is applied in the thermal system. Thus, obtaining the most fitted physicochemical characteristics related to the problem condition is of great importance [8,9]. Choosing paraffin waxes as a suitable PCM makes it a high-consuming material in the heat storage application [10,11]. Mahdi et al. simulated the melting-process in a double-pipe helical coiled tube [12]. They found that the coil pitch, inlet temperature, and the velocity of the heat transfer fluid remarkably influence the PCM melting procedure. Another study by Mahdi et al. [13]

* Corresponding author.

E-mail address: mgoodarzi@lamar.edu (M. Goodarzi).

<https://doi.org/10.1016/j.jtice.2023.104934>

Received 4 March 2023; Received in revised form 21 May 2023; Accepted 22 May 2023

Available online 26 May 2023

1876-1070/© 2023 Taiwan Institute of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

Nomenclature

SVM	Support vector machine
KNN	K-nearest neighbors
SGD	Stochastic gradient descent
RBF	Radial basis function
L	Latent heat (J)
wt.	Mass fraction (%)
ML	Machine learning
MSE	Mean squared error
SiO ₂	Silica
ZnO	Zinc oxide
Fe ₂ O ₃	Iron(III) oxide
Al ₂ O ₃	Alumina
PCM	Phase-change material
MWCNT	Multi-walled carbon nanotubes
CuO	Copper(II) oxide
MLFF-ANN	A multi-layer feed-forward Artificial neural network

Greek symbols

λ	Thermal conductivity (W/ m K)
α	Thermal diffusivity (m ² /s)
ρ	Pearson coefficient

explored the melting performance of a rectangular-based heat latent heat thermal energy storage (LHTES) system. The outcomes demonstrated that natural convection is elevated by partitioning the system into eight parts, reducing the melting time by 65%.

The low rate of phase changing of the PCMs overshadows the applications of these materials. This originated from their thermal conductivity weakness [14,15]. However, plenty of work has been done on improving this drawback, which chiefly relied on adding fins to the geometry, adding nanoparticles, using metal foams within the LHTES unit, applying a magnetic field, and utilizing a composition of several PCMs. Adding nano-additives into the PCMs is an engaging, applicable method that can significantly affect thermal conductivity and other properties such as latent heat and heat capacity. Chen et al. [16] added Cu, Al₂O₃, and TiO₂ nanoparticles to PCM, to improve solar thermal storage efficiency. The observations illustrated that nanoparticles could increase the thermal features of PCM. Moreover, it was revealed that PCM containing 0.1% TiO₂ has the lowest viscosity and the highest latent heat properties. In experimental work, Li et al. [17] assessed the heat transfer and energy storage of a rectangular cavity filled with graphene-nanoplatelets-based PCM. The findings demonstrated that adding more GNP diminishes heat transfer rate as well as thermal storage. They observed that thermal conductivity enhancement obtained by GNP was insufficient to compensate for the viscosity increment of the PCM, consequently causing a decline in natural convection. Ramakrishnan et al. [18] worked on PCM-integrated cementitious composites containing carbon-based additives such as graphene nanoplatelets (GNP), carbon nanotubes (CNT), and graphite (G). The results indicated that despite the high thermal conductivity of G and GNP, the best heat transfer performance belongs to GNP-based PCM, while the G-based PCM has the lowest performance. Fikri et al. [19] measured the thermal conductivity of A70 in the presence of MWCNT and FMWCNT at different concentrations. 150.7% enhancement was achieved by adding one wt.% FMWCNT. The results also indicated that these compositions have good stability after 500 cycles. George et al. [20] tested the properties of paraffin wax containing CuO nanoparticles as well as polyaniline. They observed that latent heat could augment by 7.8% and 8.2%, respectively. To survey the melting point of nano-enhanced PCM, Kumar and Krishna [21] claimed that adding more CuO and Al₂O₃ to the PCM causes the melting point reduction, while the melting temperature

remains unchanged. Harikrishnan et al. [22] reported that adding 1 wt.% SiO₂ into the myristic acid leads to an 87.3% enhancement in thermal conductivity. Colla et al. [23] showed that carbon black improves the thermal conductivity of RT35 by 35%. Furthermore, Al₂O₃ nanoparticles enhancement was more than 10%. Liu et al. [24] studied the influence of GNP nanoparticles and RT70HC PCM on the heat performance of a heat pipe. They found that the thermal conductivity of the PCM improved up to 21.6% by 2 wt.% concentration.

On the other hand, the latent heat diminished by about 10%. Zendeheboudi et al. [25] optimized the thermal conductivity of paraffin wax containing graphene nanoparticles of different sizes. They used three concentrations of 0.1, 0.2, and 0.25 wt.% for the 25 to 75 °C temperature range. As the results indicated, temperature plays a remarkable role in the variation of thermal conductivity. They proposed that the optimum thermal conductivity of 0.275 would be achieved at the concentration of 0.186 wt.% and temperature of 69.7 °C for the average particle size of 4.5 μm.

Precise prediction of physicochemical characteristics of nano-enhanced PCMs is essential since the heat transfer behavior of a thermal system containing PCMs is significantly hinged on those estimated characteristics. During melting and solidification, the characteristics have mostly non-linear changes, leading to an intricate problem. Machine learning (ML) can be helpful in these types of problems because it has the ability to learn the complex relationship between data and then estimate data accurately in new conditions [26–28]. Jirandeh et al. [29] implemented ANN, empirical correlations, and ant colony optimization (ACO_R) for modeling the thermophysical properties of NPCM slurry. The ANN-ACO_R model is formed by two, six, and two neurons in the input, hidden, and output layers, respectively. PCM concentration and temperature were the input. They found that the ACOR algorithm and ANN combination are compatible with experimental work. Chen et al. [30] considered Gaussian process regression (GPR) strategies such as four different kernel functions (KFs) to estimate thermal conductivity nano-enhanced PCM containing n-octadecane and oxide nanoparticles. GRP consisted of the matern KF whose R² training and testing values were 1. Sensitivity analysis revealed that the temperature of nano-enhanced PCM is the most efficient input parameter. Parashar et al. [31] experimentally examined the thermal conductivity and viscosity of MXene-doped [MMIM][DMP] ionic liquid for concentrations from 0.05 wt.% to 0.2 wt.% and temperatures from 19 °C to 60 °C. Thermal conductivity and viscosity were augmented by 45% and 14.5%, respectively. The results of multiple linear regression (MLR) models were poor. They proposed two feedforward multilayer perceptron models. The optimum models had 4 and 10 neurons for thermal conductivity and viscosity models, respectively, with high precision in experimental data prediction. Karaagac et al. [32] experimented with the drying performance of a PV/T in the presence of Al₂O₃/paraffin wax. The modeling section used ANN and SVM ML methods to predict critical parameters. Both algorithms had precise estimations, though they preferred ANN rather than SVM. Urresti et al. [33] carried out a dynamic MLP-ANN to evaluate the behavior of PCMs, which was used in a building envelope application. The results showed high accuracy in the prediction of experimental data. They indicated that due to the essence of the high variation of the problem, the training of ANN is problematic, even with applying the Bayesian algorithm. They found that the time-consuming training process is inevitable since adding more experimental data is the only resolution. Al-Waeli et al. [34] analyzed the thermal performance of a PVT considering cooling methods of nano-enhanced PCM and SiC/water nanofluid. MLP and SOFM were employed for ANN. Ambient temperature and solar irradiance were considered as input, and output was the power of PVT. The ANN model included R² and MSE equal to 0.8742 and 0.0223, respectively. In a comprehensive study, Jalilantabar [35] used different ML methods, such as MARS, CART, and ANN, to predict the thermal conductivity of nano-enhanced PCMs. This study used a comprehensive database corresponding to the thermal conductivity of nano-enhanced PCMs,

Table 1
Thermophysical features of PCM [39].

	Melting point (°C)	Specific heat (J/g °C)	Density (gr/cm ³)	Dynamic viscosity (mPa s)	Kinematic viscosity (mm ² /s)	Latent heat (kJ/kg)
Solid paraffin	53–57	2.384	0.895	–	–	184.48
Liquid paraffin	–	–	0.867	135	52.2	240

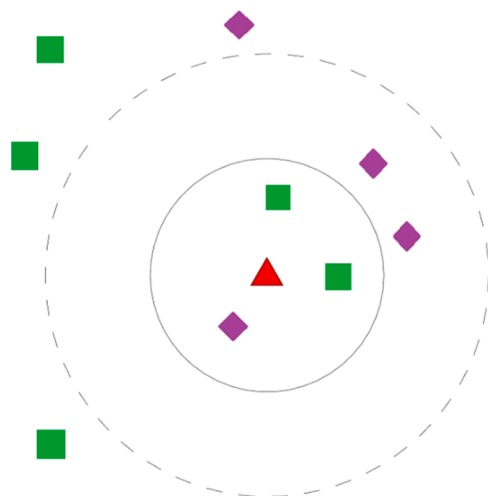


Fig. 1. The KNN method structure, K is 3 inside the inner circle, and K is 5 inside the outer circle.

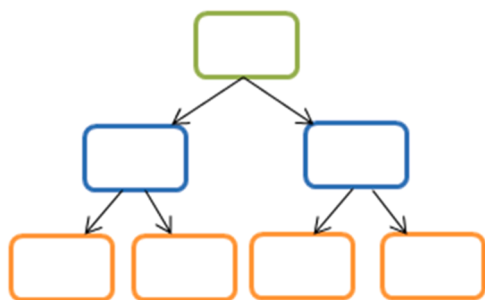


Fig. 2. A sample tree with three levels (depth = 3).

including eleven nanoparticles used in eleven different PCMs containing 911 data. For the target output, input consisted of temperature, concentration, phase, and thermal conductivity of particles and materials. R^2 was calculated for MARS, ANN, and CART; their values were 0.93, 0.96, and 0.93. Bhamare et al. [36] applied ML methods to PCM building envelope data. They utilized five ML methods, such as XDBR, RFR, CarBoost Regression, ETR, and GBR as well as one ANN. They implemented the MKR index to evaluate 500 data used in ML; it was the dependent variable, while density, specific heat, melting temperature, thermal conductivity, and latent heat were independent. The results showed that the Gradient boosting regression method has the highest prediction accuracy. Motahar and Sadri [37] estimated the thermal conductivity of nano-enhanced PCMs of n-octadecane with CuO, Al₂O₃, SiO₂, and TiO₂ nanoparticles. The inputs were temperature, solid mass fraction, and thermal conductivity. The precision of MLFF-ANN analysis was 0.9999 for R^2 , 0.00013512 for MSE, and 0.002458 for AAD. Abu-Hamdeh et al. [38] utilized ANN and RSM methods to predict the viscosity of MWCNT-based PCM between 5 and 65 °C for weight volume fractions 0.005% to 5%. The R^2 for the best polynomial was 0.988. The results of non-linear regression indicated that the viscosity estimation by ANN, including 3 and 13 neurons in the input and hidden layers, is more accurate, so the R^2 value was calculated to be 0.998.

As mentioned formerly, introducing nanoparticles into a PCM gives the ability to enhance heat-storing identifications. Progress in technology leads systems to work automatically by using artificial intelligence. Using machine learning (ML) found scientific attention in the heat storage fields because these tools help save costs and time lost by numerical and experimental methods. Therefore, the present work aims to illustrate different ML models' applicability in assessing a nano-enhanced PCM's thermophysical properties. For this purpose, thermal conductivities of four different nano-enhanced PCMs containing SiO₂, ZnO, Fe₂O₃, and Al₂O₃ nanoparticles were used as the target, which should be estimated. The inputs were considered temperature, capacity, and latent heat. Five ML methods, such as the SVM, including Kernels, RBF, and Poly, the SGD Regression, the KNN, the decision trees, and the Huber regression, were utilized for modeling. The consequences of the mentioned methods were analyzed, their accuracy was calculated, and finally, the best model was proposed. The innovative procedure and valuable results may help researchers further estimation in the heat storage systems.

2. Problem statement

The experimental data originated from the conducted study by Babapoor et al. [39]. They carried out experiments on paraffin wax containing various nanoparticles such as SiO₂, ZnO, Fe₂O₃, and Al₂O₃, at 11 and 20 nm diameters and measured the three most important parameters of heat storage studies like thermal conductivity, heat capacity, and latent heat. Sodium dodecyl sulfate and cetyltrimethylammonium bromide (CTAB) were used as PCM, whose thermo-physical features are listed in Table 1. The experiments were performed for nanoparticle mass fractions of 0%, 2%, 4%, 6%, and 8%. Using machine learning algorithms can be solved complex problems. In this research, the latent heat in the solidification-process and melting-process, thermal conductivity, and thermal diffusivity have been studied, and each subject created the formula and the model.

3. Methodology

This study uses five supervised machine-learning methods to create formulas and models. Using the Huber regression, the SGDregressor, and the SVM by the linear kernel are used to create the formulas because all methods cannot reach the formula but can reach the model, and also the KNN, the decision tree, and the SVM by RBF and poly kernel are used to create the models.

3.1. K-Nearest neighbors

Machine learning has several methods, and the KNN [40] is a non-parametric algorithm that is one of the supervised methods. This algorithm can be used in data analysis and pattern recognition because it is easy to implement and has good results. Non-parametric algorithms are a set of statistical techniques used for numerical analysis that does not have a strong hypothesis about the form of the mapping function. The KNN can be applied for classification and regression issues. The KNN uses distance to analyze the nearest neighbor points, which generally is Euclidean distance but can be used the Manhattan and the Minkowski distances, etc. Fig. 1 is shown the KNN structure and the Euclidean distance formula as follows [15]:

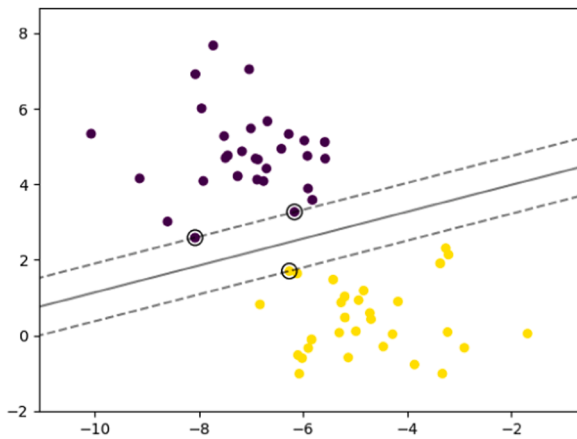


Fig. 3. The optimal hyperplane with the SVM for separation between two classes.

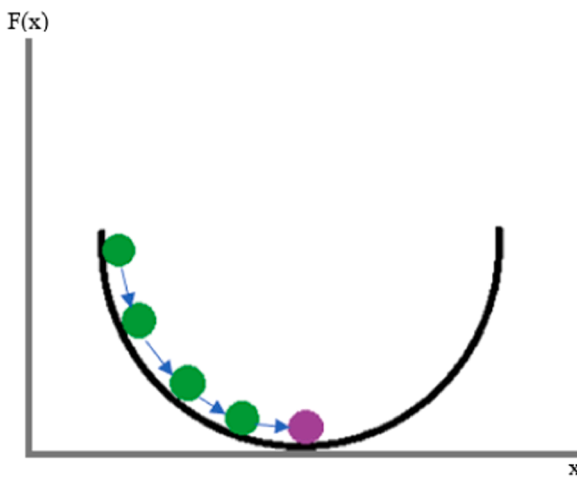


Fig. 4. Structure of SGD, the purple circle, is shown the best point which can be arrived.

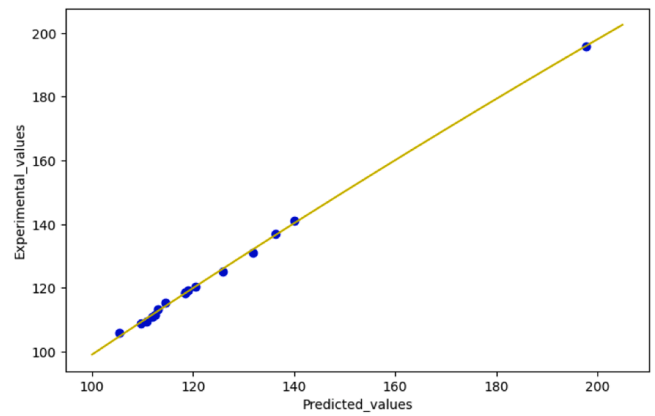
$$F(X, Y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad (1)$$

3.2. Decision tree

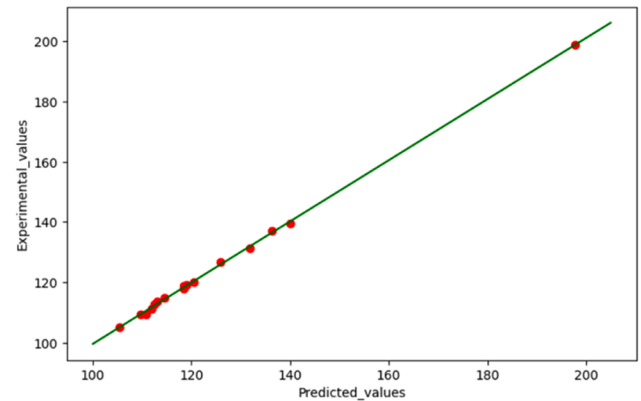
A series of if-else rules with multiple depths create the decision tree method. The tree includes a condition at each node; if that condition is appropriate, so is moved to a lower depth till the tree reaches the best situation. The decision tree is expected of supervised algorithms, one of the methods used in regression and classification [41]. Fig. 2 shows a sample tree:

3.3. Huber regression

The inputs by the numerical values can be solved using the regression algorithms. Using data to discover their relationship is the basis of regression algorithms. When outlier values exist in data inputs, drawing a line using regression methods must be biased, so if this does not happen, the outlier makes a worse prediction result. The Huber regression [42] is a type of robust regression method that is robust to outliers in training. The Huber regression method applies a linear loss to samples classified as outliers, so the Huber regression method works differently than the ridge regression method. If the absolute error of a sample is lesser than a certain threshold, therefore that sample is classified as an



(a) Huber regression with the polynomial (degree =2)



(b) KNN (K=3) with the linear form

Fig. 5. Regression plots.

inlier [43].

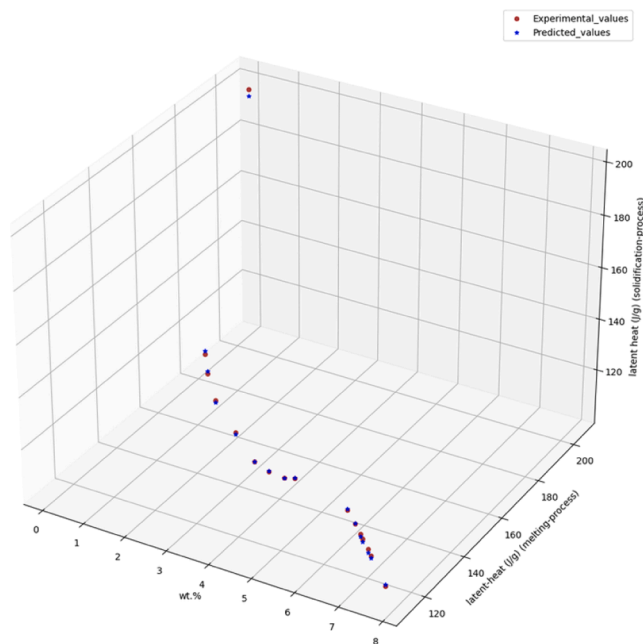
3.4. Support vector machines

The SVM method is one of the famous and influential machine learning methods, and using the SVM can solve linear [44] or non-linear regression and can be used for exploring outlier data in input data and classification. The SVM is a kind of supervised learning method that provides the optimal hyperplane to classify input data in various categories by the training data label [45]. The SVR [46] is support vector regression, used for regression issues with different kernels that can be trained and made the model. Several kernels like linear, poly and RBF (Radial basis function) can be used. Fig. 3 is shown the structure of the SVM:

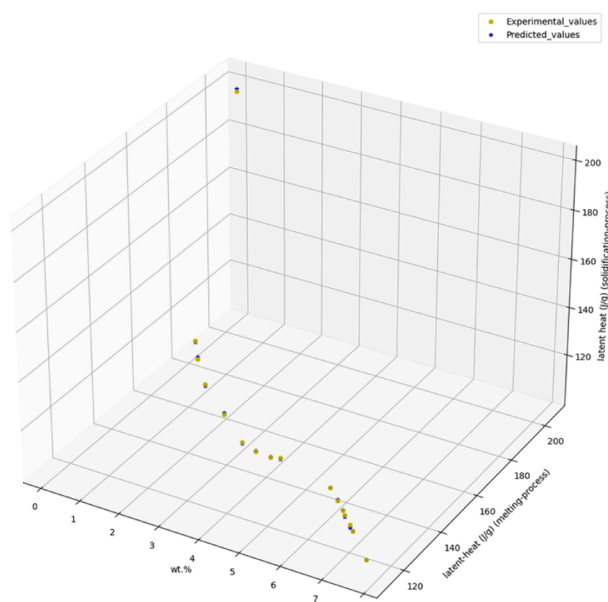
3.5. SGD regressor

The SGD [47] is an iterative method that begins from a random point on a function and stepwise moves on a Gradient to arrive at the lowest point of the cost function. The SGD is a popular and powerful optimization technique in machine learning and deep learning. The SGD provides the algorithm to reach the minimum value of a cost function with iterative loops. However, this algorithm has a learning rate, and this issue can be helped to reach the best point of the cost function. The SGD Regressor [48] is solved by numerical inputs. Fig. 4 shows shown that the SGD algorithm how to works as follows:

Fig. 4 shows that when we arrive at the purple circle, the results are reached the optimized point, so the made model has the highest accuracy, but the purple circle is the ideal point. In the real world, if we reach this point, the model results have high accuracy.



(a) Huber regression



(b) The KNN (K = 3)

Fig. 6. 3-D representation result.

4. Results and discussion

The main goal of machine learning is to allow the computer to learn the training process without human intervention automatically, and these algorithms are able to solve problems related to the set target with the least amount of error and the fastest possible time. Machine learning algorithms are divided into two main branches, supervised and unsupervised. Supervised algorithms are used labels on training data in the learning process, and in unsupervised algorithms, the training data has not been labeled in the learning process. In this research used five supervised ML methods: the KNN, the SVM, the decision tree, the Huber regression, and the SGD regressor. In addition, this research has studied the linear and the polynomial forms.

Polynomial regression is a kind of regression that uses a polynomial equation with different degrees to numerical data and a curve

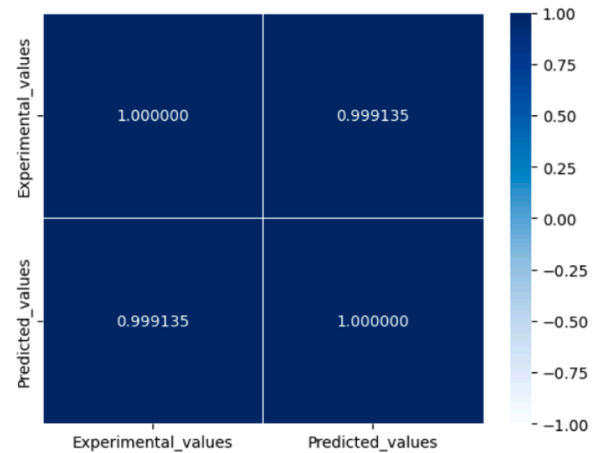
(a) Huber regression ($\rho = 0.999135$)(b) KNN (K=3) ($\rho = 0.999575$)

Fig. 7. Heat map of the Pearson correlation coefficient.

relationship among the dependent and independent variables. The polynomial regression algorithm is practical when the results of the linear form do not best result for solving the problems. Then, the MSE and the R^2 value can represent the precision of the models and the formulas obtained by the machine learning methods.

4.1. Solidification-process by the melting-process

For creating the formula, the model is compared all of the results in linear and polynomial form by five algorithms (the decision tree, the Huber regression, the KNN, the SVM, and the SGD regressor) which are used in all parts will be used. The Huber regression had the best R^2 , which means this algorithm has the best results for creating the formula for predicting the latent heat of the solidification-process by the melting-process (by latent heat and wt.). The input numbers for the train and test are 64 and 16. The R^2 values in the polynomial form (degree = 2) are 0.998, and in the linear form for this method are 0.986. MSEs for polynomial form by degree 2 and linear form, respectively, are 0.971172 and 6.544063. Now, these results can be compared with other algorithms' results, for the KNN (K = 3) in the polynomial (degree = 2) and linear form is 0.999 and MSEs respectively are 0.456690 and 0.418095, so this method can be made the model with the linear form because the MSE is less than polynomial form for this issue. The values of R^2 of the SVM with the different kernels, such as linear is 0.985, poly is 0.955 (degree=3) and 0.932 (degree=2), and RBF in the linear and polynomial forms are 0.6, 0.424, the decision tree is 0.998 in the linear, and the polynomial form and the SGD regressor in the linear and polynomial form are respectively 0.966, 0.991.

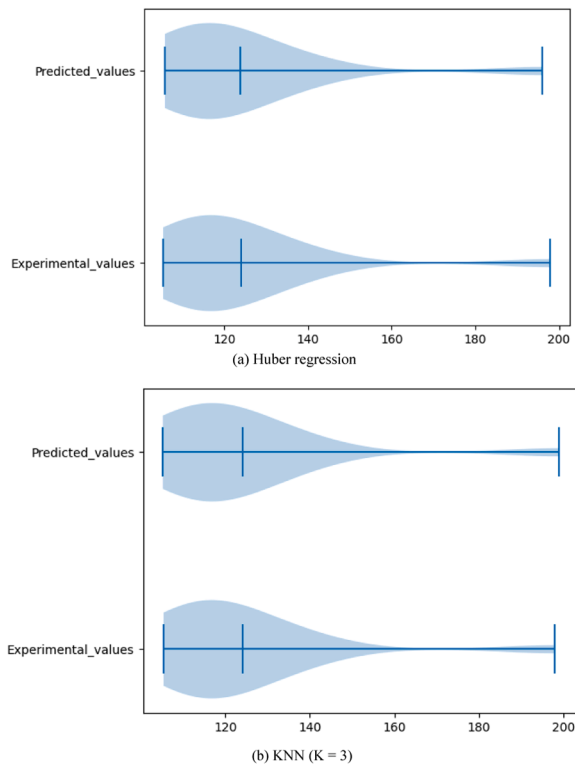


Fig. 8. Data distribution and data density by the violin plot.

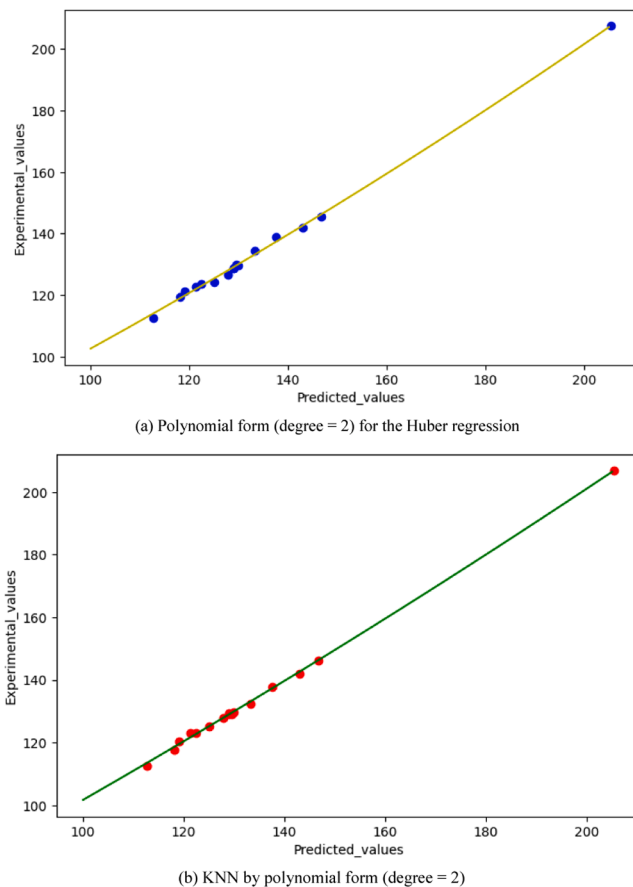


Fig. 9. Regression plots.

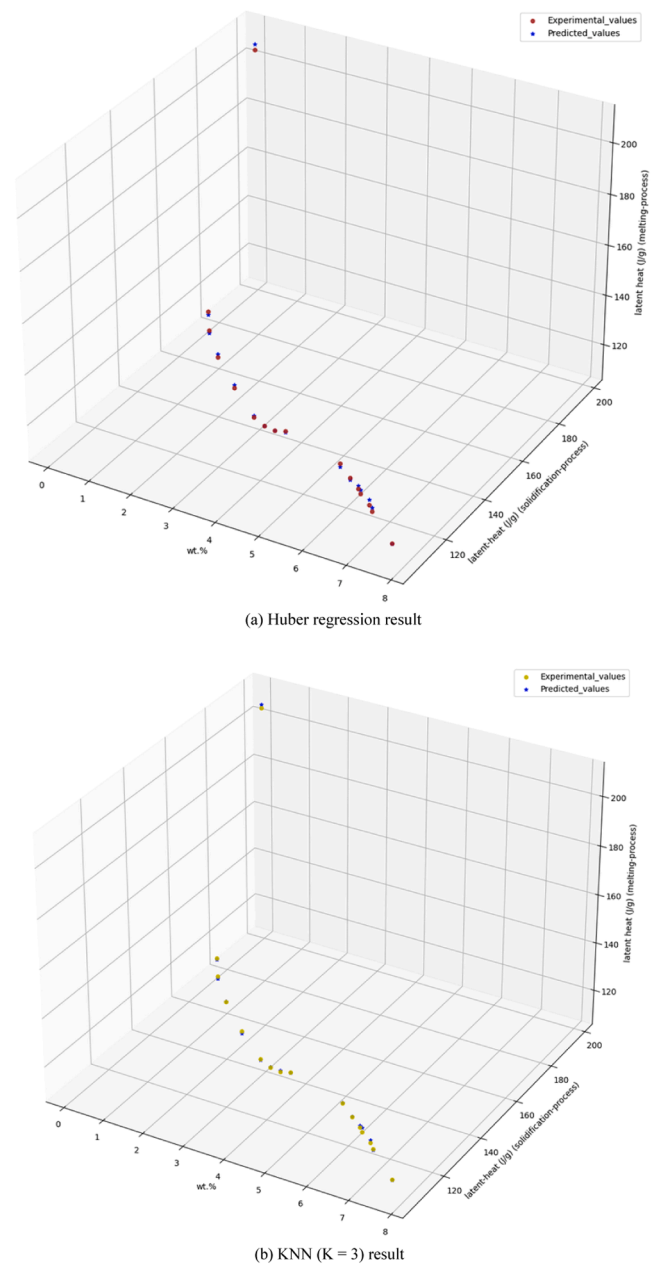


Fig. 10. 3-D representation.

Sometimes the input data need to apply data normalization until some algorithms on that input data represent good results. For example, the SGD regressor needs this method to improve the results in this part. Data normalization is an algorithm in which data characteristics are structured to improve the consistency of different entities in the data model. When data normalization is done best, the input data will be structured, which is useful in the results of some algorithms like the SGD regressor.

The formula is created by the Huber regression in polynomial form and shown in Eq. (2) as follows:

$$L_{\text{solidification-process}} = 0.1192857wt. + 1.2490157L + 0.3815848wt.^2 - 0.046255(wt. \times L) - 0.0014302L^2 + 0.03093872 \quad (2)$$

Using several different plots, such as the regression plot, the heat map, and the violin plot can describe the result of each algorithm. The regression plot shows the relationship between the experimental and

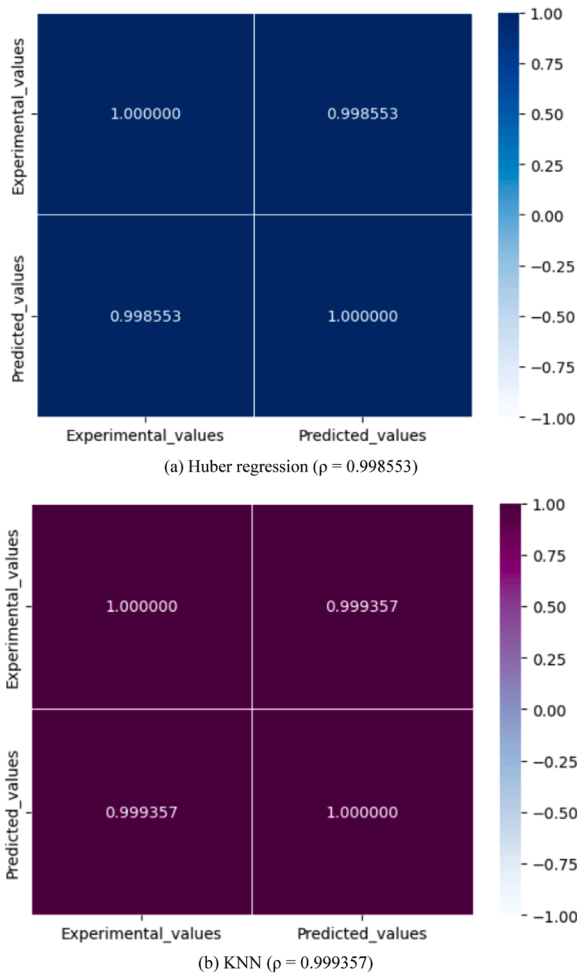


Fig. 11. Representation of heat map for the Pearson correlation coefficient.

predicted numbers in linear and polynomial forms. The plots of the Huber regression and the KNN are shown for this part because these methods had the best results. Fig. 5 shows the regression plots for the latent heat of the solidification process. Therefore, all figures (from Figs. 5 to 8) are as follows:

The 3-D plot can be used for the representation of the inputs and output results, and this plot is handy for comparing the experimental and predicted numbers. Axes x and y are shown the input values, and axis z shows the output results. With 3-D model can be shown better this result, Fig. 6 shows this issue:

The Pearson coefficient is the algorithm for representing the communication between two variables. The limit of the Pearson coefficient is between -1 to 1 ; this method can determine the direction and the strength of linear communication between two variables. The Pearson coefficient value is near 1 , there is a positive correlation in communication among the two variables. If the value of the Pearson coefficient is near -1 , there is a negative correlation in communication among the two variables. The heat map plot can describe the Pearson coefficient. Fig. 7 shows the heat map for the Huber regression and the KNN:

The secondary diagonal in the heat map shows the coefficient of Pearson correlation among the experimental and predicted values.

The violin plot is used to visualize the numerical data distribution and the density of numerical data. This is like the box plot, but the violin plot shows data density [49]. The violin plot combines a density plot and a box plot to show the shape of the numerical data distribution, giving it a better visual quality than the box plot. Fig. 8 is shown the violin plot for the solidification process:

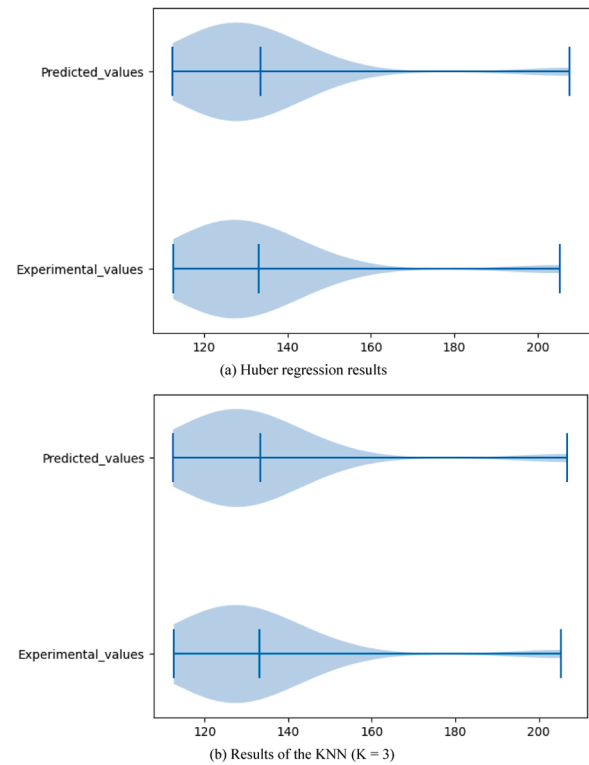


Fig. 12. Data distribution and data density by violin plot.

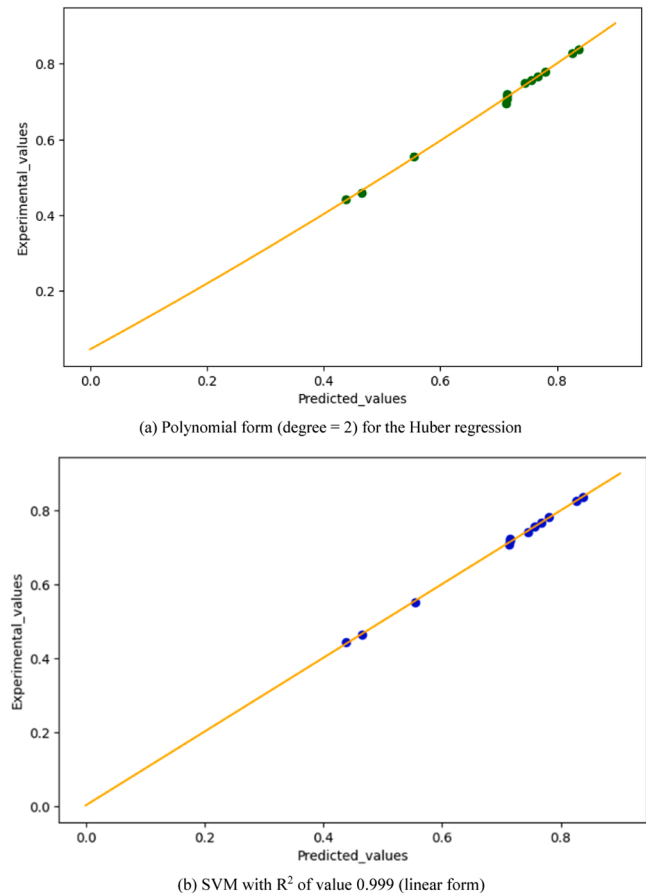
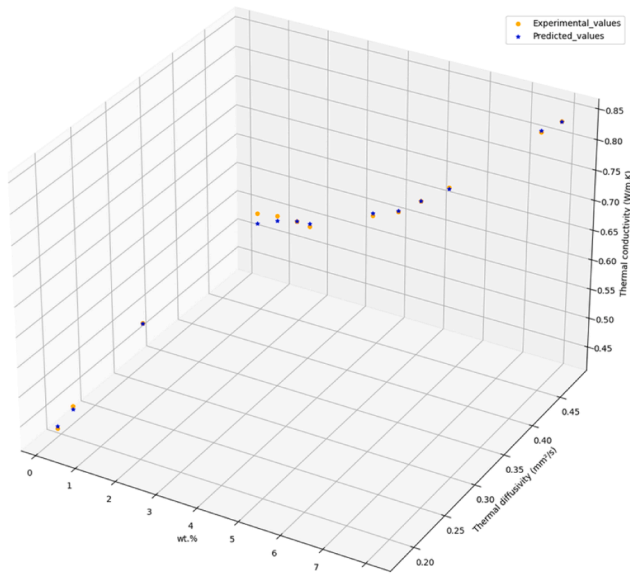
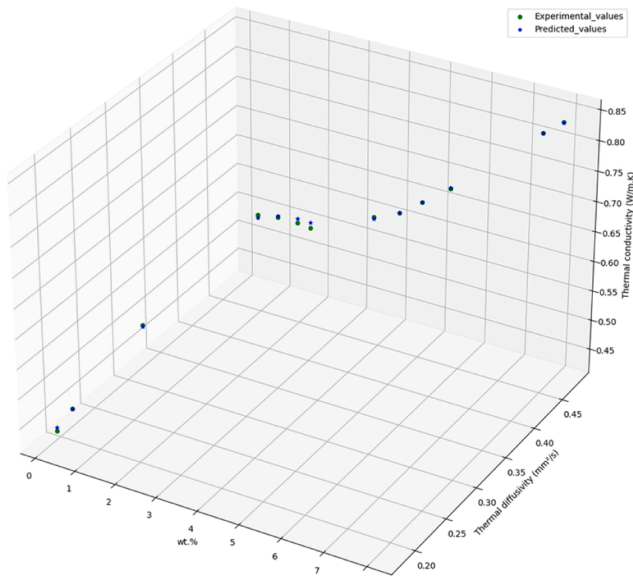


Fig. 13. Regression plots for thermal conductivity.



(a) Huber regression



(b) SVM by RBF kernel

Fig. 14. 3-D display for thermal conductivity.

4.2. Create a formula and model for the melting-process by the solidification-process

In this part, the Huber regression had the best R^2 value so that this algorithm can be created the formula melting-process latent heat. The input numbers for training methods and testing are 64 and 16. The R^2 value of the Huber regression in linear form and polynomial form are respectively 0.986 (MSE = 5.819890) and 0.996 (MSE = 1.482458). Eq. (3) shows this polynomial formula for the latent heat of the melting-process by the solidification-process (by latent heat and wt.) as follows:

$$L_{\text{melting-process}} = -0.0181686wt. + 0.7050063L - 0.48419447wt.^2 + 0.05912871(wt. \times L) + 0.00171252L^2 + 0.0072421 \quad (3)$$

The latent heat of the melting-process by the solidification-process (by latent heat and wt.) best model is obtained using the KNN ($K = 3$) in polynomial form degree 2, so the R^2 value in linear form is 0.998 (MSE = 0.720995) and in polynomial form is 0.999 (MSE = 0.608658).

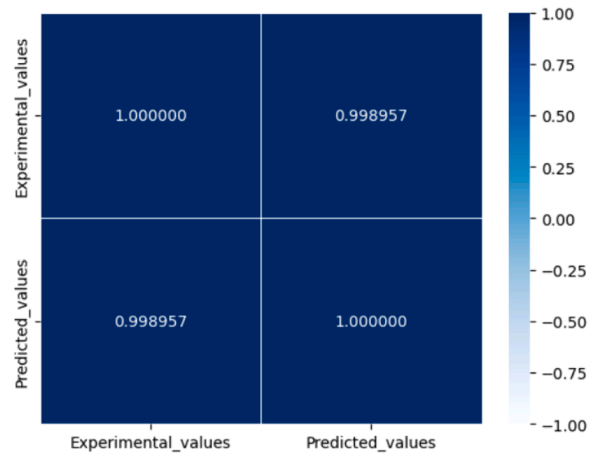
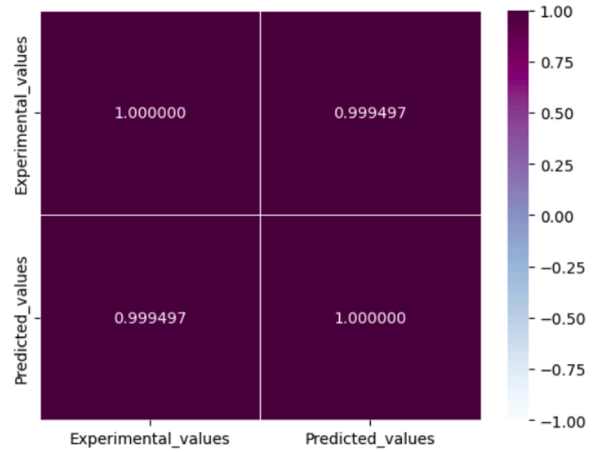
(a) Huber regression ($\rho = 0.998957$)(b) SVM by RBF kernel ($\rho = 0.999497$)

Fig. 15. Correlation for thermal conductivity.

Other results of algorithms can be compared with these algorithms, and this issue causes better clarity for inference. The R^2 value of the SGD regressor in the linear and polynomial form, respectively, are 0.966 and 0.979, the value of the R^2 for the decision tree in linear and polynomial form is 0.994, the R^2 value of the SVM in linear and polynomial form by RBF kernel are 0.608 and 0.531, and the R^2 values of the SVM by linear and poly kernel are respectively 0.985, 0.947(degree=3), 0.966 (degree=2).

All of the figures, such as Fig. 9 for regression plots, Fig. 10 for 3-D plots, Fig. 11 for heat maps, and Fig. 12 for violin plots, for the two best algorithms (the Huber regression and the KNN), are as follows:

4.3. Thermal conductivity

This part of the research looks for a connection between thermal conductivity by the thermal diffusivity and wt. The input numbers for training algorithms and tests in this part and part of the thermal diffusivity are 50 and 13. The Huber regression finds the best communication, so this method created a formula. The value of R^2 is 0.998 for the polynomial (MSE = 0.000033) form (the linear form is 0.908 and MSE = 0.001387). Eq. (4) is shown the polynomial regression formulas:

$$\lambda = 0.1365771 \times wt. - 1.6691358\alpha - 0.0008271wt.^2 - 0.3360971(wt. \times \alpha) + 5.8913872\alpha^2 + 0.5343976 \quad (4)$$

The SVM reaches the best model by RBF kernel (epsilon = 0.0001, gamma = scale) which the R^2 is 0.999 in linear form, and this value in the polynomial form is 0.956 and MSEs respectively are 0.000016 and

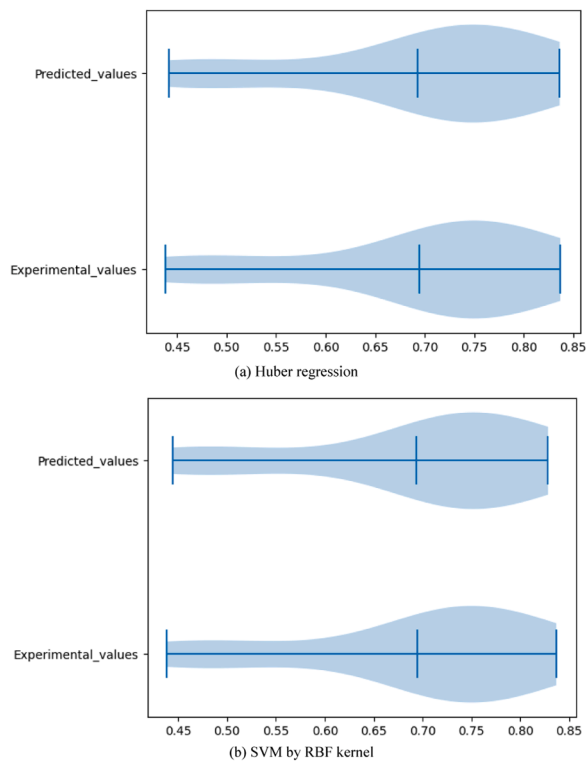


Fig. 16. Distribution of data and density of data for thermal conductivity.

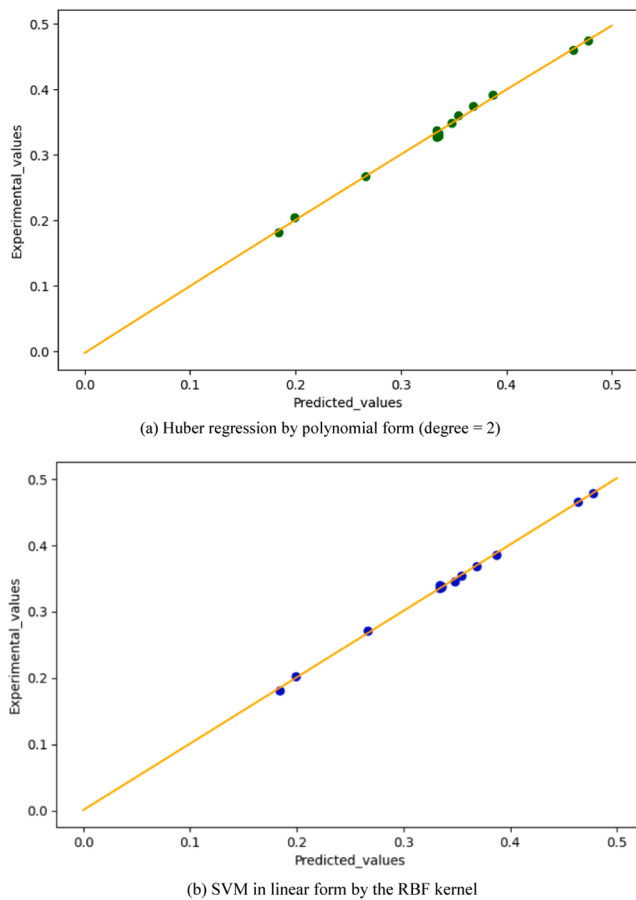
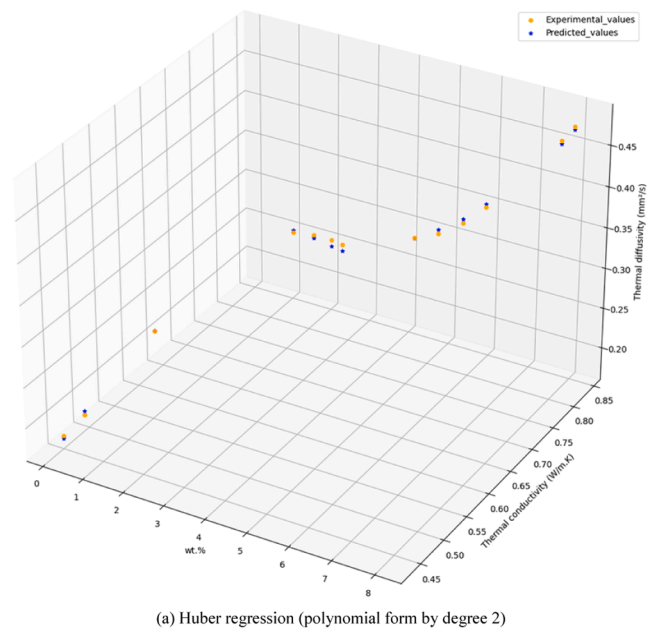
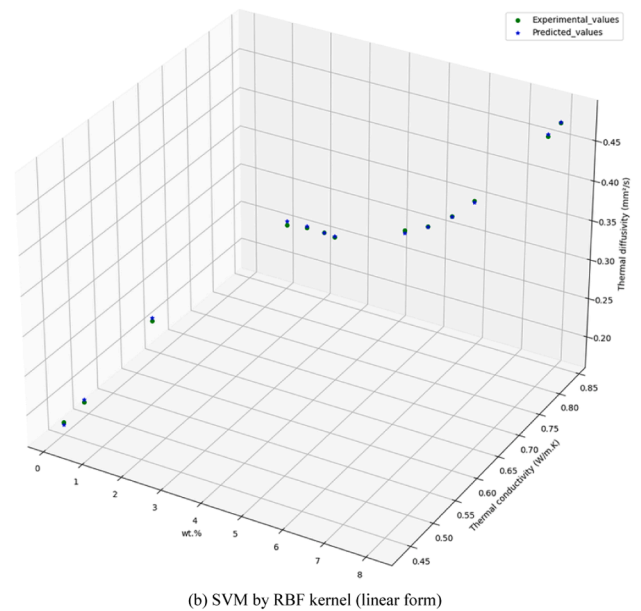


Fig. 17. Regression plots for thermal diffusivity.



(a) Huber regression (polynomial form by degree 2)



(b) SVM by RBF kernel (linear form)

Fig. 18. 3-D display for thermal diffusivity.

0.000667. If using a data normalization method like the standard scaler can be reached 0.999 to R^2 values in the polynomial form (MSE = 0.000020), so the linear form has less MSE than the polynomial form and the linear using for the model. The values of the R^2 for other kernels (linear and poly) are respectively 0.838 and 0.635 (degree = 2), and the R^2 values for the SGD regressor, and the KNN and decision tree in the linear form are 0.927, 0.998, and 0.996; in the polynomial form are respectively 0.973, 0.998, and 0.996.

All figures from Fig. 13 to 16 represent regression plot (Fig. 13), 3-D plot (Fig. 14), heat map with the Pearson coefficient (Fig. 15), and violin plot (Fig. 16) for thermal conductivity, respectively are as follows:

4.4. Thermal diffusivity

The relationship for thermal diffusivity by the thermal conductivity and wt., using The Huber regression and the SVM with RBF (gamma = scale, epsilon = 0.001) [15] kernel, are suitable for creating the formula

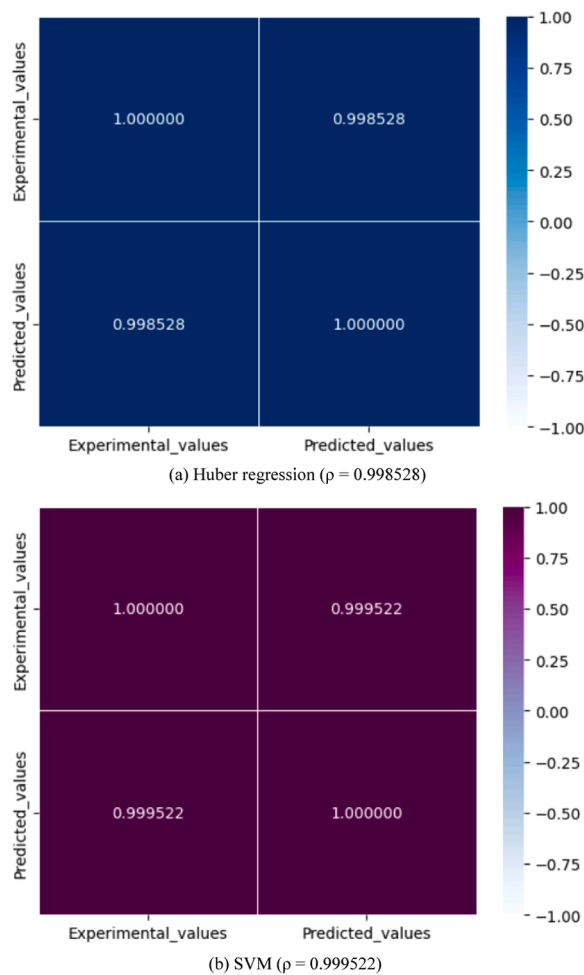


Fig. 19. Heat map to represent correlation for thermal diffusivity.

and making the model because these methods have the highest R^2 values between all algorithms. Therefore, the values of R^2 for these methods (Huber regression and SVM with RBF) are respectively 0.997 and 0.953 (SVM with standard scaler is 0.999) in polynomial in degree 2 and MSEs are 0.000020 and 0.000314 (with standard scaler is 0.000008), and 0.922 and 0.999 in linear form (MSEs = 0.000521 and 0.000006), so is used the SVM with the linear form for the model (gamma = scale, epsilon = 0.001) for the model. The SVM with poly and linear kernels, respectively, are 0.746 (degree=3), 0.866 (degree=2), and 0.584 [40].

The values of R^2 for the KNN, the decision tree, and the SGD regressor are 0.995, 0.996, and 0.923 in linear and 0.996, 0.996, 0.961 in linear form.

The formula of thermal diffusivity by the thermal conductivity and wt. is represented in Eq. (5) as follows:

$$\alpha = -0.1264724wt. + 3.1338634\lambda + 0.0029241wt.^2 + 0.1371137(wt. \times \lambda) - 2.090133\lambda^2 - 0.7796860 \quad (5)$$

The thermal diffusivity of the evaluation figures is shown in Figs. 17 to 20. The figures represent a regression plot (Fig. 17), a 3-D plot (Fig. 18), a heat map with the Pearson coefficient (Fig. 19), and a violin plot (Fig. 20):

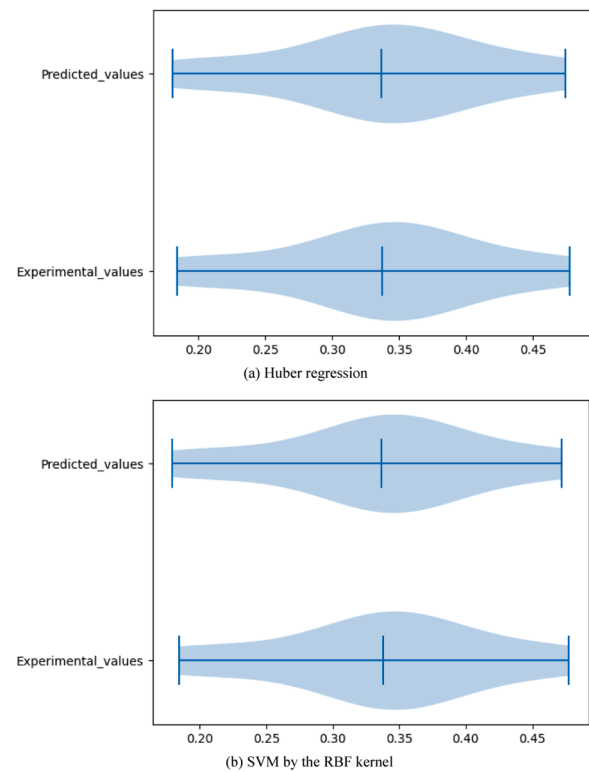


Fig. 20. Distribution of data and density of data by violin plot for thermal diffusivity.

5. Conclusion

The ML methods were applied to experimental data derived from literature. The data encompassed three parameters of thermal conductivity, heat capacity, and latent heat of nano-enhanced PCMs marked up with paraffin wax and four nanoparticles of SiO₂, ZnO, Fe₂O₃, and Al₂O₃. The parameters were measured at 0%, 2%, 4%, 6%, and 8% mass concentrations. Machine learning methods can be used for creating regression algorithms. With this issue is calculated the numerical data input, create the formula, and make the model for solving these problems. Using five machine learning methods to create the formula and make the model for the relationship between the solidification-process and the melting-process and the relationship between the thermal conductivity and the thermal diffusivity in this study. In this study, the Huber regression, the SGD regressor, the KNN, the SVM, the decision tree methods, and linear and polynomial forms are computed for the relationship between the solidification-process and the melting-process and the relationship between the thermal conductivity and the thermal diffusivity. By the Huber regression algorithm is created the formula for the latent heat of the solidification- process by the melting-process (by latent heat and wt.) with polynomial form, and the value of R^2 is 0.998,

and the MSE is 0.971172, and using the KNN ($K = 3$) is made the model with the linear form, and the value of R^2 is 0.999, and MSE is 0.41809. The Huber regression method has the best result for the latent heat of the melting-process by solidification-process (by latent heat and wt.), which is created a polynomial formula with R^2 value 0.996, and by the KNN ($K = 3$) is made the model with R^2 0.999 in polynomial form and MSEs are

1.482458 and 0.608658. In this research, the Huber regression method has the best output values for creating the formula for each part, so this method is the best method for making the formula for the thermal conductivity by the thermal diffusivity and wt. in polynomial form, and the R^2 value is 0.998 MSE is 0.000033, and make the model by the SVM (by RBF kernel), thus the R^2 value is 0.999 in the linear form, and MSE is 0.000016. The Huber regression method and the SVM (by RBF kernel) are used for thermal diffusivity by the thermal conductivity and wt. The R^2 values are 0.997 and 0.999 (in the polynomial by degree 2 and the linear forms). MSEs, respectively, are 0.000020 and 0.000006. Even though input data were limited but the precision of output results is at the highest level.

Declaration of Competing Interest

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

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research - Majmaah University, for supporting this research work under project number R-2023-446.

References

- [1] Sarafraz M, et al. Experimental investigation on thermal performance of a PV/T-PCM (photovoltaic/thermal) system cooling with a PCM and nanofluid. *Energies* 2019;12(13):2572.
- [2] Punniakodi BMS, et al. Numerical study on melting of phase change material in a horizontal container using multi heat transfer tubes. *J Taiwan Inst Chem Eng* 2022; 131:104214.
- [3] Alazwari MA, Algarni M, Safaei MR. Effects of various types of nanomaterials on PCM melting process in a thermal energy storage system for solar cooling application using CFD and MCMC methods. *Int J Heat Mass Transf* 2022;195: 123204.
- [4] Arumugam P, Ramalingam V, Vellaichamy P. Effective PCM, insulation, natural and/or night ventilation techniques to enhance the thermal performance of buildings located in various climates—a review. *Energy Build* 2022;111840.
- [5] Safaei MR, Goshayeshi HR, Chaer I. Solar still efficiency enhancement by using graphene oxide/paraffin nano-PCM. *Energies* 2019;12(10):2002.
- [6] Fini AT, Fattahi A, Musavi S. Machine learning prediction and multiobjective optimization for cooling enhancement of a plate battery using the chaotic water-microencapsulated PCM fluid flows. *J Taiwan Inst Chem Eng* 2023;104680.
- [7] Nguyen-Thoi T, et al. Analysis on the heat storage unit through a Y-shaped fin for solidification of NEPCM. *J Mol Liq* 2019;292:111378.
- [8] Chizoo E, et al. Adaptive neuro-fuzzy inference system-genetic algorithm versus response surface methodology-desirability function algorithm modelling and optimization of biodiesel synthesis from waste chicken fat. *J Taiwan Inst Chem Eng* 2022;136:104389.
- [9] Tili I, et al. Macroscopic modeling for convection of Hybrid nanofluid with magnetic effects. *Phys A: Stat Mech Appl* 2019;534:122136.
- [10] Gao Y, et al. Thermophysical properties prediction of carbon-based nano-enhanced phase change material's using various machine learning methods. *J Taiwan Inst Chem Eng* 2023;104662.
- [11] Sheikholeslami M. Numerical modeling of nano enhanced PCM solidification in an enclosure with metallic fin. *J Mol Liq* 2018;259:424–38.
- [12] Mahdi MS, et al. Improved PCM melting in a thermal energy storage system of double-pipe helical-coil tube. *Energy Convers Manag* 2020;203:112238.
- [13] Mahdi MS, et al. Natural convection improvement of PCM melting in partition latent heat energy storage: numerical study with experimental validation. *Int Commun Heat Mass Transf* 2021;126:105463.
- [14] Goodarzi M, et al. Applying Bayesian Markov chain Monte Carlo (MCMC) modeling to predict the melting behavior of phase change materials. *J Energy Storage* 2022; 45:103570.
- [15] Bhatti M, Ellahi R. Numerical investigation of non-Darcian nanofluid flow across a stretchy elastic medium with velocity and thermal slips. *Numer Heat Transf, Part B: Fundam* 2023:1–21.
- [16] Chen H, et al. Experimental study on characteristics of a nano-enhanced phase change material slurry for low temperature solar energy collection. *Solar Energy Mater Solar Cells* 2020;212:110513.
- [17] Li Z, et al. Experimental investigation of heat storage and heat transfer rates during melting of nano-enhanced phase change materials (NePCM) in a differentially-heated rectangular cavity. *J Therm Sci* 2020;29:503–11.
- [18] Ramakrishnan S, Wang X, Sanjayan J. Effects of various carbon additives on the thermal storage performance of form-stable PCM integrated cementitious composites. *Appl Therm Eng* 2019;148:491–501.
- [19] Fikri MA, et al. Thermal conductivity, reliability, and stability assessment of phase change material (PCM) doped with functionalized multi-wall carbon nanotubes (FMWCNTs). *J Energy Storage* 2022;50:104676.
- [20] George M, et al. A novel polyaniline (PANI)/paraffin wax nano composite phase change material: superior transition heat storage capacity, thermal conductivity and thermal reliability. *Solar Energy* 2020;204:448–58.
- [21] Sandeep Kumar M, Murali Krishna V. Experimental investigation on performance of hybrid PCM's on addition of nano particles in thermal energy storage. *Mater Today: Proc* 2019;17:271–6.
- [22] Harikrishnan S, et al. Improved performance of a newly prepared nano-enhanced phase change material for solar energy storage. *J Mech Sci Technol* 2017;31(10): 4903–10.
- [23] Colla L, et al. Nano-PCMs for enhanced energy storage and passive cooling applications. *Appl Therm Eng* 2017;110:584–9.
- [24] Liu Y, et al. Characteristics of thermal storage heat pipe charged with graphene nanoplatelets enhanced organic phase change material. *Energy Convers Manag* 2022;267:115902.
- [25] Zendeheboudi A, et al. Optimization of thermal conductivity of NanoPCM-based graphene by response surface methodology. *J Adv Res Fluid Mech Therm Sci* 2020; 75(3):108–25.
- [26] Yavari M, Maihami R, Esmaeili M. Ambulance dispatching and relocation problem considering overcrowding of emergency departments. *IIEE Trans Healthc Syst Eng* 2022;12(4):263–74.
- [27] Shakoor DM, Maihami V, Maihami R. A machine learning recommender system based on collaborative filtering using Gaussian mixture model clustering. *Math Methods Appl Sci* 2021.
- [28] Zohoori S, et al. An AIS-Based approach for measuring waterway resiliency: a case study of Houston ship channel. *Marit Policy Manag* 2022:1–21.
- [29] Hashemi Jirandeh MR, et al. Intelligent modeling of rheological and thermophysical properties of nanoencapsulated PCM slurry. *Heat Transf* 2020;49 (4):2080–102.
- [30] Chen T-C, et al. Insights into the estimation of the enhanced thermal conductivity of phase change material-containing oxide nanoparticles using Gaussian process regression method. *Int J Chem Eng* 2022. 2022.
- [31] Parashar N, et al. ANN modeling of thermal conductivity and viscosity of MXene-based aqueous ionanofluid. *Int J Thermophys* 2021;42:1–24.
- [32] Karaagaç MO, et al. Experimental analysis of CPV/T solar dryer with nano-enhanced PCM and prediction of drying parameters using ANN and SVM algorithms. *Solar Energy* 2021;218:57–67.
- [33] Urresti A, Campos-Celador A, Sala J. Dynamic neural networks to analyze the behavior of phase change materials embedded in building envelopes. *Appl Therm Eng* 2019;158:113783.
- [34] Al-Waeli AH, et al. Mathematical and neural network modeling for predicting and analyzing of nanofluid-nano PCM photovoltaic thermal systems performance. *Renew Energy* 2020;145:963–80.
- [35] Jaliliantabar F. Thermal conductivity prediction of nano enhanced phase change materials: a comparative machine learning approach. *J Energy Storage* 2022;46: 103633.
- [36] Bhamare DK, et al. A machine learning and deep learning based approach to predict the thermal performance of phase change material integrated building envelope. *Build Environ* 2021;199:107927.
- [37] Motahar S, Sadri S. Applying artificial neural networks to predict the enhanced thermal conductivity of a phase change material with dispersed oxide nanoparticles. *Int J Energy Res* 2021;45(10):15092–109.
- [38] Abu-Hamdeh NH, Golmohammadzadeh A, Karimipour A. Performing regression-based methods on viscosity of nano-enhanced PCM-Using ANN and RSM. *J Mater Res Technol* 2021;10:1184–94.
- [39] Babapoor A, Karimi G. Thermal properties measurement and heat storage analysis of paraffinnanoparticles composites phase change material: comparison and optimization. *Appl Therm Eng* 2015;90:945–51.
- [40] So A., S.W., Nagy Z., The applied artificial intelligence workshop: start working with AI today, to build games, design decision trees, and train your own machine learning models 2020.
- [41] Harrington P. Machine learning in action . 20 balding road. Shelter Island: Manning Publications Co; 2012.
- [42] Feng Y, Wu Q. A statistical learning assessment of Huber regression. *Journal of Approximation Theory* 2022;273:105660.
- [43] https://scikit-learn.org/stable/modules/linear_model.html#huber-regression. Scikit learn. Available online: (accessed on 05/26/, 2022).
- [44] A. G., Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems: " O'Reilly Media, Inc". 2019.
- [45] Haykin S. Neural networks and learning machines. New Jersey, USA: Prentice Hall; 2008. edit.
- [46] <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>, 2020.
- [47] Deisenroth MP, Faisal AA, Ong CS. Mathematics for machine learning. Cambridge University Press; 2020.
- [48] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html, 2020.
- [49] Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. *Am Stat* 1998;52(2):181–4.