International Journal of Mathematical, Engineering and Management Sciences
Vol. 8, No. 5, 817-840, 2023
https://doi.org/10.33889/IJMEMS.2023.8.5.047

Modelling of Viscosity and Thermal Conductivity of Water-Based Nanofluids using Machine-Learning Techniques

Sai Ganga
School of Engineering and Technology,
BML Munjal University, Gurugram, 122413, Haryana, India.
E-mail: sai.ganga.21pd@bmu.edu.in

Ziya Uddin
School of Engineering and Technology,
BML Munjal University, Gurugram, 122413, Haryana, India.
Corresponding author: ziya.uddin@bmu.edu.in

Rishi Asthana
School of Engineering and Technology,
BML Munjal University, Gurugram, 122413, Haryana, India.
E-mail: rishi.asthana@bmu.edu.in

Hamdy Hassan
Energy Resources Engineering Department,
Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt.
E-mail: hamdyaboali@yahoo.com

Arpit Bhardwaj
School of Engineering and Technology,
BML Munjal University, Gurugram, 122413, Haryana, India.
E-mail: arpit.bhardwaj@bmu.edu.in

(Receive on May 15, 2023; Accepted on July 6, 2023)

Abstract
In this study, a variety of machine-learning algorithms are used to predict the viscosity and thermal conductivity of several water-based nanofluids. Machine learning algorithms, namely decision tree, random forest, extra tree, KNN, and polynomial regression, have been used, and their performances have been compared. The input parameters for the prediction of the thermal conductivity of nanofluids include temperature, concentration, and the thermal conductivity of nanoparticles. A three-input and a two-input model were utilized in modelling the viscosity of nanofluid. Both models considered temperature and concentration as input parameters, and additionally, the type of nanoparticle was considered for the three-input model. The order of importance of the most influential parameters in predicting both viscosity and thermal conductivity was studied. A wider range of input parameters have been considered in an open-access database. With the existing experimental data, all of the developed machine learning models exhibit reasonable agreement. Extra trees were found to provide the best results for estimating thermal conductivity, with a value of 0.9403. In predicting viscosity using a three-input model, extra trees were found to provide the best result with a value of 0.9771, and decision trees were found to provide the best results for estimating the viscosity using a two-input model with a value of 0.9678. In order to study heat transport phenomena through mathematical modelling, it is important to have an explicit mathematical expression. Therefore, the formulation of mathematical expressions for predicting viscosity and thermal conductivity has been carried out. Additionally, a comparison with the Xue and Maxwell thermal conductivity models is made to validate the results of this study, and the results are observed to be reliable.

Keywords- Nanofluid, Machine learning, Thermal conductivity, Viscosity.

1. Introduction

Nanofluids are a combination of liquid (base fluids) like water, ethylene glycol, etc, and solid nanosized particles like metals, metal oxides, nonmetals, nonmetal oxides, etc. A milestone in the direction of the invention of nanofluids is due to the work of Choi and Eastman, 1935. Nanofluids are useful in many applications, including those involving transfer of heat, due to its number of unique characteristics (Mahian et al., 2019). Since nanofluids are utilized to improve the thermophysical properties of fluids, research has mostly focused on enhancing their heat transfer ability. Understanding a nanofluid's thermophysical characteristics is essential for choosing it for a certain application. Among various properties, viscosity and thermal conductivity are two significant properties. Viscosity describes resistance in the flow of fluid and further influences different phenomena including heat transfer. Therefore, understanding how viscous nanofluids are important in many of the industrial applications (Hemmati-Sarapardeh et al., 2018). In understanding the behaviour of heat transfer, thermal conductivity has an important role (Naseri et al., 2020).

In a wide range of technological applications, there are many advantages in predicting the thermal conductivity and viscosity of nanofluids. Engineers may optimize the composition and concentration of nanoparticles to maximize heat transfer efficiency by precisely predicting the thermal conductivity of nanofluids. It will help in the improvement of the performance of cooling devices and effective thermal management in various industrial processes. Engineers can create more effective and affordable systems that use nanofluids as heat transfer fluids by knowing how they operate at different temperatures and flow rates. Therefore, precise estimates of thermal conductivity and viscosity are helpful for energy system design and optimization. Also, the development of novel materials with desired thermal properties is made easier by accurate predictions of nanofluid properties.

It is observed that there have been several studies on various parameters affecting viscosity and thermal conductivity like temperature, shear rate, concentration of nanoparticles, nanoparticle shape, size, etc. There have been numerous studies on two of the most influential parameters namely concentration and temperature. In general, rising temperature results in a drop in viscosity and an increase in thermal conductivity, whereas rising concentration results in a rise in viscosity and thermal conductivity (Maleki et al., 2021).

It is vital to predict the viscosity and thermal conductivity of nanofluids through modelling and simulation rather than relying solely on experimental methods. Experimentation can be tedious and costly, especially when dealing with nanofluids that require specialized tools and methods. When choosing an experimental setup, researchers and engineers can realistically explore a wide range of conditions using predictive modelling before choosing the optimal experimental setup. Also, predictive modelling is an efficient method that makes design optimization easier, provides insight into nanoscale processes, and allows parameter sensitivity research.

Several research works have been carried out in its prediction using various analytical and empirical models. Many machine learning algorithms have lately attracted a lot of attention and have been widely used to assess the thermophysical characteristics of nanofluids. A comprehensive review of machine learning models on hybrid nanofluids can be found in (Maleki et al., 2021). A few of the significant works on conventional nanofluids are presented below.

1.1 Thermal Conductivity

The literature demonstrates that various machine learning techniques, such as Multi-Layer Perceptron (MLP), Artificial Neural Network (ANN), Least Square Support Vector Machine (LSSVM), Gradient

Boosting Regression (GBR), Radial Basis Function Neural Network (RBFNN), Support Vector Regression (SVR), Random Forest Regression (RFR), Decision Tree Regression (DTR), K-Nearest Neighbour (KNN), Gaussian Process Regression and Genetic Algorithm-Polynomial Neural Network (GA-PNN) were used in the prediction of the thermal conductivity of water-based and non-water-based nanofluids. Hemmat Esfe et al. (2016) used ANN to predict the Thermal Conductivity (TC) of aluminium oxide water-Ethylene Glycol (EG) based nanofluid. Different temperatures and concentrations between 20 and 60 degrees Celsius and 0-1.5% were considered. This study came to the conclusion that ANN could accurately determine the TC using these factors they had taken into account. The size of the nanoparticles, various concentration ranges, and temperatures were taken into account. It was discovered that ANN outperformed other works that were available in the literature. Afrand et al. (2016) used ANN to predict TC of water-based Fe_3O_4 nanofluids in experimental research. As input parameters, concentrations of 0.1%, 0.2%, 0.4%, 1%, 2%, 3%, and temperatures between 20°C to 55°C were used. With the experimental data, an empirical correlation was also proposed, and it was compared to the ANN model that had been developed. The accuracy of the ANN model was found to be higher than that of empirical correlation. To evaluate the TC of two oxide water nanofluids, Longo et al. (2012) employed two ANN models, one with three and the other with four input parameters. One model took into account the effect of nanoparticle size, while another model additionally took into account temperature, concentration, and the thermal conductivity of the nanoparticles. The data sample used 1%, 2% and 4% of concentration value for the nanoparticle Al_2O_3 and 1%, 2%, 4% and 6% for the nanoparticle TiO_2 at the temperature of 1°C, 10°C, 20°C, 30°C and 40°C. Both models were observed to give reliable results, even though the four-input model was performing better. In their experimental research of non-Newtonian nanofluids made up of Al_2O_3, CuO, and TiO_2, Hojjat et al. (2011) suggested a neural network model that took into account three input features: temperature, the thermal conductivity of the nanoparticles and concentration.

The considered concentrations are 0.1%, 0.2%, 0.5%, 1.0%, 1.5%, 3.0%, and 4.0%. It was found that the obtained model and the experimental results exhibited good agreement. To determine the TC of Al_2O_3 - water nanofluids considering a range of temperatures, concentrations, and nanoparticle sizes, Mehrabi et al. (2012) employed two different models and compared their performances. Different concentration values between 0.3% and 4% and, temperature between 21°C and 71°C were considered. MLP, RBFNN, and LSSVM were developed in the prediction of TC covering more input parameters including temperature, concentration, particle size, particle TC and base fluid TC over a wider range by Hemmati-Sarapardeh et al. (2020). Committee Machine Intelligence System (CMIS) model was also developed combining the used intelligent models. Comparing CMIS to other theoretical and empirical models, it was found that it performed the best. Zhang and Xu (2020) used the same input parameters and obtained a Gaussian process regression model to predict the TC enhancement. An accurate estimation was achieved after taking into account concentration levels between 0.1% and 3% and temperatures between 20°C and 50°C. Each of these papers considered different input parameters and could predict thermal conductivity with good accuracy. The range of values of various input parameters differs based on the data set chosen. Naseri et al. (2020) developed a novel LSSVM-Improved Simulated Annealing (LSSVM-ISA) model in which a wider range of concentration and temperature is considered. Five features were selected as the input parameters. The outcomes were evaluated against other models, such as RBF-NN, KNN, and pre-existing correlation models. The observed performance of the suggested model outperforms other models. Concentration was found to be the key parameter in the sensitivity analysis that was carried out. Recently, Sharma et al. (2022) used five machine learning models- ANN, GBR, SVR, DTR, and RFR and compared their results for TiO_2-water nanofluid in predicting the TC. Four input parameters including shape, size, concentration, and temperature were chosen and it was observed from the comparison that GBR was the best algorithm. Sahin et al. (2023) predicted zeta potential and thermal conductivity of Fe_3O_4 – water nanofluid accurately using ANN.

1.2 Viscosity

For the prediction of Viscosity (VIS), machine learning algorithms including ANN, LSSVM, MLP, SVR, RFR, DTR, GPR, Extra Tree (ET), Radial Basis Function (RBF) and Multivariate Adaptive Regression Splines (MARS) have been utilized. Various attempts have been made to improve the breadth of applicability of existing models by considering more input parameters, large data sets, and modifying the algorithms. The LSSVM model, which takes into account temperature, nanoparticle size, and concentration, was used by Meybodi et al. (2015) to evaluate the viscosity of Al_2O_3 , TiO_2, CuO, and SiO_2 nanofluids based on water. Concentration values between 0.03% and 13% and temperature between 10°C and 72°C were considered in the prediction. The predictability of the model was assessed with various existing empirical models and the LSSVM model was observed to give accurate results. Yousefi et al. (2012) had used the same nanoparticles suspended in various nanofluids to predict the viscosity using diffusional neural networks. Temperature, concentration, nanoparticle size, and the physical characteristics of the base fluid are taken into account as input factors. The range of temperature and concentration considered are -34.9°C to 50.2°C and 0.0246 to 0.0431 respectively. It was concluded that the proposed model produces reliable findings in comparison to other theoretical and experimental models that are presently in use. Four LSSVM-based methods were employed by Ramezanizadeh et al. (2018), to simulate the viscosity of Al_2O_3 nanofluid based on water. The input parameters namely concentration between 0.003% and 13%, size, and temperature between 21°C and 70°C were taken into consideration. It was concluded that all the different models could give accurate results. Considerations included nanoparticle diameter, concentration with a maximum of 9.4%, and temperature between -0.15°C and 71.85°C. The conclusion reached was that multi-layer perceptron neural networks perform better than other intelligent methods. Gholizadeh et al. (2020) used RFR to determine the viscosity of numerous nanofluids for the first time. The model's performance was also compared with MLP and SVR, and the RFR model was giving the best prediction. Concentration, temperature, nanoparticle size, nanoparticle density, and base fluid viscosity are the input parameters that are used. The most influential factors were found to be the concentration and nanoparticle density. Hemmati-Sarapardeh et al. (2018) predicted the viscosity of several nanofluids using the MLP model with four different optimizations, the RBF model with two different optimizations, and the LSSVM model. Additionally, all of these models were merged into the CMIS model. The input parameters included nanoparticle size, temperature between -35°C and 80°C, and, concentration between 0% and 10%. It was concluded that CMIS outperforms all the models considered and, existing theoretical and empirical models.

Using the same data set, Shateri et al. (2020) carried out a comparative analysis using various models in the assessment of VIS. The goal was to apply a DT model, RF model, and ET model, and a deeper MLP network to enhance the efficacy and accuracy of the model used by Hemmati-Sarapardeh et al. (2018). The maximum concentration and temperature considered are 10% and 80°C respectively. The viscosity of a TiO_2 water-based nanofluid was estimated by Ahmadi et al. (2020a) utilizing the LSSVM, MLP-ANN, RBF-ANN and ANFIS algorithms. In terms of performance, LSSVM outperformed all other approaches. The input parameters used were concentration, temperature, and diameter of the nanoparticle. It was concluded that each factor had a direct impact on predicting the viscosity. Considering density and size of nanoparticle, concentration, temperature, and base fluid viscosity as input parameters, a DNN model was proposed by Changdar et al. (2020). Al_2O_3, CuO, SiO_2, TiO_2, Ag and Fe_2O_3 were the chosen nanoparticles.

It was observed that the proposed models outdo the existing models and overcome the limitations of conventional models in predicting viscosity. It was also concluded that concentration, temperature, and density were the most important parameters to determine viscosity. For the purpose of predicting the viscosity of various nanofluids, Alade et al. (2020) employed ANN and BSVR. The influence of various input factors, including concentrations varying from 0% to 9%, temperatures varying from -35.029°C to 71.201°C, nanoparticle size, density, and viscosity of base fluids, has been investigated. It was concluded

that the base fluid's viscosity might be ignored when predicting the viscosity of nanofluids, potentially increasing the model's accuracy. It was determined that BSVR produced superior outcomes to ANN. When predicting viscosity for a silver/water nanofluid, Ahmadi et al. (2019) used the ANN-MLP, MARS, and Multivariate Polynomial Regression (MPR) techniques. When comparing input characteristics such as temperature, concentration, and nanoparticle size, it was found that temperature was the most important factor. All the methods were observed to give good predictive accuracy and ANN-MLP was observed to perform better as compared to other models. In predicting the viscosity of CuO water-based nanofluids, a variety of machine learning techniques, namely MPR, MARS, ANN-MLP, Group Method of Data Handling (GMDH), and M5-tree, were employed Ahmadi et al. (2020b). Nanoparticle size, concentration, and temperature have been taken into consideration as input factors. Concentration was found to be highly significant in this investigation. It was concluded that in comparison ANN-MLP was having a better predictive ability. Using temperature and concentration as input factors, Kumar and Kavitha (2021) utilized MLP and Gaussian process technique to determine the viscosity of an Al_2O_3 nanofluid based on water. The developed model is observed to be efficient and accurate. Bhaumik et al. (2023) created a physics-aided deep learning technique for the prediction of the viscosity of Al_2O_3, SiO_2, TiO_2, CuO nanoparticles, which is said to overcome the limitations of existing deep learning models. The base fluid's density, temperature, size, concentration, and viscosity were chosen as independent variables. The concentration parameter in the study was found to be important, and the model produced accurate findings. Dai et al. (2023) predicted the viscosity of SiO_2 combined with ethylene glycol using GPR and was observed to be reliable method for using in related applications.

Using machine learning methods, Durgam and Kadam (2021) studied the viscosity and thermal conductivity of nanofluids. Thermal conductivity is calculated by taking into account the temperature, thermal conductivity, concentration, nanoparticle diameter, and thermal conductivity of the base fluid. The input variables for viscosity include molecular weight, nanoparticle diameter and concentration, base fluid viscosity, and nanofluid temperature. It was concluded that thermal conductivity was better predicted by linear regression and ANN, and viscosity was better predicted by ANN.

From the literature review, it is observed that the models- RF, DT, ET, and MPR, are not extensively studied when it comes to predicting the viscosity and thermal conductivity of nanofluids. Also, all of the existing models give accurate results for particular nanoparticles across a specific range of temperature and concentration. It is important to note that, when we have to study the heat transfer phenomena of nanofluids, we mathematically model the physical scenario into a set of differential equations and solve them. And most of those equations require explicit mathematical expressions for modelling the thermal conductivity and viscosity of nanofluids. Predictions given by most machine learning algorithms don't give an explicit mathematical expression that can be used for further analysis. But the non-linear regression used in this study can be used for the same purpose.

The following are some of this study's primary objectives: employing several machine learning algorithms for modelling and estimating the viscosity and thermal conductivity of different water-based nanofluids and comparing their performance. Formulation of a mathematical expression to predict the viscosity and thermal conductivity. The input parameters, including concentration and temperature, have been given a wider range of consideration, and relative importance has been determined for different input parameters. Also, a comparison with the existing thermal conductivity model of Xue and Maxwell (Zhang et al., 2021) is carried out to validate the results obtained in this study.

2. Data-Set

The data-set has been consolidated from various nanofluid-related experimental data taken from a database of thermophysical properties of nanofluids (Mondejar et al., 2021). The database includes experimental data-sets collected from various works of literature. In this study, 523 and 302 experimental data samples (Abdollahi et al., 2018; Agarwal et al., 2017; Azmi et al., 2012; Barbés et al., 2013; Buonomo et al., 2015; Colla et al., 2015; Ferrouillat et al., 2013; Ghodsinezhad et al., 2016; Ho et al., 2010; Kim et al., 2007; Kumar et al., 2016; Madhesh et al., 2016; McCants et al., 2009; Pak and Cho, 2007; Pantzali et al., 2009; Pastoriza-Gallego et al., 2009; Patil et al., 2018; Paul et al., 2013; Ramalingam et al., 2016; Saxena et al., 2016; Suresh et al., 2016; Tso and Chao, 2015; Zhu et al., 2011) were taken for the model development of viscosity and thermal conductivity of nanofluids, respectively, which have been extracted from the database. Data includes various nanofluids namely Copper Oxide/water, Aluminium oxide/water, Titanium dioxide/water, Zinc oxide/water, Silver/water, Silicon dioxide/water, and Cerium oxide/water. A basic statistical description of the used data sample is given in Tables 1 and 2. Additionally, the experimental data set's dynamic viscosity and thermal conductivity values have been normalized by dividing with the respective viscosity and thermal conductivity of the base fluid at room temperature, and data pre-processing has been carried out. The concentration value given in the data-set as percentage was divided by 100 before the data pre-processing. A sample of the used data for viscosity and thermal conductivity are given in Tables 3 and 4, respectively.

Table 1. Statistical details of the data set for predicting the viscosity.

Parameters	Minimum	Maximum	Mean	Standard Deviation
T (K)	278.15	348.15	308.943480	16.502806
ϕ (%)	0.000221	22.951200	3.140485	4.961272
VIS	0.379393	2.076172	0.882566	0.289070

Table 2. Statistical details of the data set for predicting the thermal conductivity.

Parameters	Minimum	Maximum	Mean	Standard Deviation
T (K)	278.15	363.08	316.631362	20.626168
ϕ (%)	0.002	5	1.687386	2.002608
TC of nanoparticle (W/(m*K))	8.3	406	80.12839	125.394741
TC	0.970922	1.290107	1.362416	1.261757

Table 3. Sample data for predicting the viscosity.

Temp(K)	Nanoparticle	Concentration (%)	VIS
328.15	Copper Oxide	5	0.696386
313.15	Copper Oxide	5	0.971945
305.04	Aluminium oxide	1.34	1.116387
324.67	Aluminium oxide	1.34	0.822621
334.98	Aluminium oxide	1.34	0.713368
305.04	Aluminium oxide	1.34	1.116387
324.67	Aluminium oxide	1.34	0.822621
334.98	Aluminium oxide	1.34	0.713368
345.03	Aluminium oxide	1.34	0.634014
⋮	⋮	⋮	⋮

Table 4. Sample data for predicting the thermal conductivity.

Temp (K)	TC of nanoparticle (W/(m*K))	Concentration (%)	TC
278.15	32	2	1.007479
283.15	32	2	1.024222
288.15	32	2	1.036313
293.15	32	2	1.040034
308.15	32	2	1.051195
278.15	32	4	1.064217
287.15	32	4	1.080959
293.15	32	4	1.096771
298.15	32	4	1.081889
⋮	⋮	⋮	⋮

3. Brief Description of used Methodologies

In this study, the viscosity and thermal conductivity of water-based nanofluids are predicted and modelled using five machine learning algorithms: the decision tree regressor, random forest regressor, extra tree regressor, multivariate polynomial regression (MPR), and k-nearest neighbour regressor (KNN).

3.1 Decision Tree

A decision tree is a supervised machine learning algorithm that creates a tree-structured regression or classification model. It separates the dataset into smaller subsets while gradually building the corresponding decision trees. A tree having internal nodes, leaf nodes, branches, and a root node is the final result. The root node is the primary node that yields the best predictor, while the interior nodes stand in for the data set's features, the branches for the decision-making processes, and the leaf nodes for the final outcome. The goal of the decision tree is to keep the tree as small as possible. For that, variables are chosen in such a way that data can be divided into uniform groups in the best possible way so that the data entropy is minimized in the next branch. Entropy describes the degree of dispersion of data between different classes and the reduction in entropy is the information gain. Simply put, each level of data should be more uniform than the data at the previous level. So, now the idea is to choose a greedy algorithm that reduces the entropy at different levels of the decision tree. One such algorithm is proposed by Quinlan (1986) namely Dichotomizer (ID3) which is a core algorithm for building decision trees. In case of regression problems, information gain is replaced with standard deviation reduction and the idea remains same. Standard deviation reduction is related to the reduction in standard deviation after the splitting of the data set on an attribute. The mathematical formula of the standard deviation reduction is as follows:

$$SDR(Y|X) = S(Y) - S(Y|X) \quad (1)$$

where, Y is the target variable, X is a specific attribute of the data set, $SDR(Y|X)$ is the standard deviation reduction of Y given the information of $X, S(Y)$ is the standard deviation of $Y, S(Y|X)$ is the standard deviation of Y given the information of X.

$$S(Y|X) = \sum_{c \in X} P(c)S(c) \quad (2)$$

where, $P(c)$ is the probability of each value of the independent variables and $S(c)$ is the standard deviation of Y based on those values of the independent variables. The $S(Y|X)$ is calculated for all independent variables in the data set, and that independent variable with largest $SDR(Y|X)$ is used in splitting of the tree. However, this algorithm is prone to over fitting. The crucial processes in creating a decision tree are splitting, stopping, and pruning in order to prevent over fitting (Song and Lu, 2015). The pseudo algorithm of the method which gives the overview of algorithm is given below:

Start
(i) Input dataset for training
(ii) Select the best feature for splitting the dataset such that data entropy is least
(iii) Create a decision node with the selected feature
(iv) By making use of the selected feature, split the dataset into subsets
(v) For each and every subset:
    If the subset is pure
        Make a leaf node that contains the predicted value of output
    Else
        Repeat the process and create subtrees on the subset
        Attach the obtained subtree to the decision node
(vi) Output the decision node
Stop.

3.2 Random Forest

Another supervised learning approach called random forest creates several decision trees as the basis for regression or classification models. For each of the subsets obtained from the original dataset using row and feature sampling, individual decision trees are formed. After several decision trees are formed the final prediction is based on the majority votes of that class predicted by different decision trees, for classification problems. And the average value of the predictions by different decision trees, for regression problems. A random forest is constructed based on the bagging principle. Breiman (2001) developed the random forest algorithm. It unifies the technique of bagging with the decision tree forest method proposed by Ho (1995). Breiman (2001) gives the definition of random forest classifier as follows:

***Definition* 3.1** "A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, ...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x" (Breiman, 2001).

One of the primary ensemble techniques used in machine learning is bagging. The idea of using ensemble learning is that multiple learning models improve the overall result and produce one optimal model. Bagging is basically bootstrap aggregation. Bootstrapping randomly samples a subset of a data set over a specified number of iterations and a specified number of variables. And all these results give a stronger prediction as they are averaged together. So, the core idea of using bagging is to reduce the variance in the individual model, here the decision tree, to avoid overfitting. We are not bothered if the decision tree chosen overfits a data set when bagging with a decision tree. Because, as we take a set of high variance and low bias models and join them with bagging, we get a model with less variance and less bias. As we add more trees, random forest does not tend to overfit; rather, a limiting value of generalization error is created (Breiman, 2001). The pseudo-algorithm of the method, which gives an overview of the algorithm, is given below:

Start
(i) Input dataset for training
(ii) Create bootstrapped dataset
(iii) Making use of the bootstrapped dataset and only a random subset of features at each step, create a decision tree.
(iv) Repeat step 2-3 until desired number of decision trees are formed
(v) Input the test dataset
(vi) Consider each datapoint from the test data

(vii) For each and every decision trees formed, predict the output
(viii) Final prediction is the mean of the predictions from all decision trees
(ix) Repeat step 6-8 for every datapoints
(x) Output the test data prediction
Stop.

3.3 Extra Tree

Extremely Randomized Trees (Extra Trees) is a supervised learning algorithm that develops regression or classification models in the form of multiple decision trees. The only significant conceptual difference between it and the random forest method is how the decision trees are built within the forest. Both algorithms are intended to provide an ideal model. The extra tree uses the complete original sample to construct each decision tree, as opposed to the subsampling done by a random forest. Another difference lies in the splitting of the nodes. Extra Tree chooses a random split, whereas Random Forest chooses the best split. So, the extra tree is comparatively computationally efficient. Also, after choosing a split, both the random forest and extra tree choose the best of all the subsets of features. Hence, extra trees add randomization, but they also optimize. The bias/variance analysis of the model can be found in Geurts et al. (2006).

3.4 Multivariate Polynomial Regression

Multivariate Polynomial Regression is a regression procedure that explains the linear or non-linear relationship seen between the independent and dependent variables using an nth-degree polynomial. The general equation of multivariate polynomial regression of order two is given by:

$$y = a_0 + a_1x_1 + a_2x_2 + a_{11}x_1^2 + a_{22}x_2^2 + a_{12}x_1x_2 \quad (3)$$

where, a_1, a_2 are linear parameters that effect linearly, a_{11}, a_{22} parameters that has quadratic effect and a_{12} is interaction effect parameter. This can be generalized to a higher order. These unknown parameters have to be estimated during the training process such that the loss function is minimized. The loss function computes the degree to which a predicted value differs from its actual value. Depending on the task, several loss functions, such as mean absolute error, mean squared error, mean squared logarithmic error, mean bias error, etc., may be used. Now, to obtain optimal parameters, the loss function has to be minimized, and which is done with the help of optimizers. There are different optimizers, that optimize the parameters during the training process. Stochastic gradient descent is the most popular and common optimizer used in machine learning. The hyperparameter in polynomial regression is the degree of the polynomial.

3.5 K-Nearest Neighbour

Another non-parametric supervised learning algorithm used for both classification and regression problems are K-Nearest Neighbour. KNN was first proposed by Fix and Hodges (1951), which was later broadened by Cover and Hart (1967). The KNN algorithm stores the dataset while training for both the regression and classification problems, and the new input data is categorized into the most similar category among the available categories i.e., the input includes the k closest training data. The output in the classification problem is based on the majority vote from the available neighbours, i.e., the new data is assigned to that category whose values are most frequently occurring. Similar ideas are used in the case of regression, but the average value of the k nearest neighbours is considered the output. The basic methodology of the KNN algorithm is to evaluate the distance to determine which data points are closer to the new input value. In both classification and regression problems, various notions of distance are utilized. The KNN algorithm assumes that all the data points are nearby in the geometric sense. There are various ways to measure distance, a few of them are Euclidean distance, Manhattan distance, Minkowski distance, and Hamming distance. The hamming distance is used for categorical data, and the first three are used for continuous data.

Consider $x = (x_1, x_2, \dots, x_k)$ and $y = (y_1, y_2, \dots, y_k)$, Euclidean distance, Manhattan distance and Minkowski distance between x and y are respectively given by,

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (4)$$

$$\sum_{i=1}^{k}|x_i - y_i| \quad (5)$$

$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^p\right)^{1/p} \quad (6)$$

where, i, k and p are integers.

The pseudo algorithm of the method which gives the overview of algorithm is given below:
Start
(i) Input the training dataset
(ii) Input the test dataset
(iii) Consider each datapoint from the test data
(iv) Calculate the distances to between the test datapoint to all the points in the training data
(v) Arrange the distances in increasing order
(vi) Based on the ordered distances, choose the k closest neighbors
(vii) Evaluate the average value of the k nearest neighbours
(viii) Evaluated average is the predicted output for the considered datapoint
(ix) Repeat steps 3-8 for every datapoints
(x) Output the test data prediction
Stop.

4. Results and Discussion

As previously indicated, it appears that the random forest, decision tree, extra tree, and multivariate polynomial regression (MPR) have not been extensively investigated in terms of determining the viscosity and thermal conductivity of nanofluids. Hence, we have considered these models in this study. Also, decision trees and ensembles of decision trees, such as random forests and extra trees, are effective machine learning techniques because the best split is found by going through each feature and selecting among them. This behaviour aids in overcoming certain machine learning model's poor performance in a particular range of input features (Shateri et al., 2020). The results from these models were also compared with those predicted by the KNN model. All five machine learning models have been trained using the Scikit-learn machine learning library available for the Python programming language. Data pre-processing has been done prior to training the data set. In order to prevent sampling bias, 30% of the data were used to test the trained model and 70% were used for training. Additionally, cross-validation has been employed to prevent over fitting. We have several options for the values of the hyper parameters that may be selected for each model. The model was tuned while taking into consideration several hyper parameters using Grid search, a fundamental hyper parameter tuning technique. However, compared to utilizing the default values, there was only a slight improvement in accuracy when using the tuned hyper parameters. Therefore, the default parameters in the Scikit-learn library were used for random forest, decision tree, extra tree and KNN for the prediction of viscosity. In the case of the predicting thermal conductivity, for the KNN algorithm, choosing the nearest neighbour equal to 1 gave the best result. For the random forest, decision tree, and extra tree, default parameters were chosen. Nevertheless, in the case of the MPR model, when Lasso regularization and polynomial degree greater than five were both taken into account, the MPR model showed a better improvement in accuracy. Therefore, with the polynomial transformation of the input features and LASSO regularization, the polynomial regression model is developed. All of the cross-coupled terms have been included, and the polynomial's degree has been set to five. Using a random forest model,

permutation feature importance analysis has been done to determine the relative importance of the input features. The idea of permutation feature importance analysis is as follows: Features are permuted, and the significance is determined by computing the increase in error, in the model's predicted output. A feature is significant if changing its values causes the error to rise; otherwise, it is unimportant if the error remains the same. Numerous statistical score metrics and scatter plots were implemented for examining and comparing the performance of the models. A statistical measure called the coefficient of determination (R^2) was determined by averaging the R^2 results from cross validation. Other criteria for evaluation used were Mean Squared Error (MSE) and Mean Absolute Error (MAE).

4.1 Viscosity Model

In predicting viscosity, five different machine learning algorithms were implemented on the data set in two different cases over a wider range of concentrations and temperatures.

4.1.1 Case 1 (3-input model)

In the first case, three input parameters including concentration, temperature, and type of nanoparticles were considered. Considered nanofluids include Copper Oxide/water, Aluminium oxide/water, Titanium dioxide/water, Zinc oxide/water, Silver/water, Silicon dioxide/water, and Cerium oxide/water. Since the type of nanoparticles is string variables, corresponding nanoparticles were given the indices from 1 to 7. The most governing parameters in predicting viscosity with 3 input models are found in the order $T, \phi,$ and type of nanoparticles (Figure 1). It can be observed that the influence of the type of nanoparticle is comparatively less, so we consider a second case excluding the type of nanoparticle as an input parameter. Different evaluation criteria including coefficient of determination (R^2), Mean squared error (MSE) and mean absolute error (MAE) are used to assess the performance of each machine learning algorithm, and the details are summarized in Table 5.

4.1.2 Case 2 (2-input model)

In the second instance, temperature and concentration were the only two input variables taken into account. In this case, the most governing parameters are found in the order T and ϕ (Figure 2). Different evaluation criteria including coefficient of determination (R^2), Mean squared error (MSE) and mean absolute error (MAE) are used to assess the performance of each machine learning algorithm, and the details are summarized in Table 5.

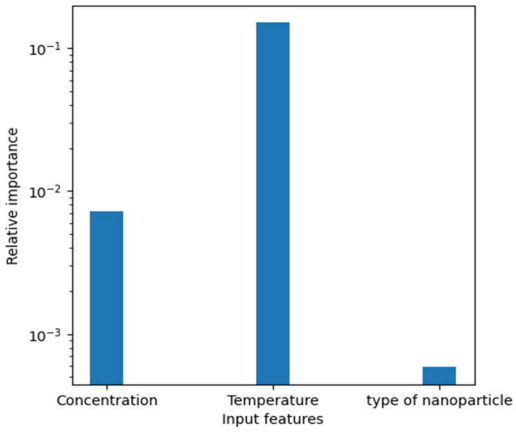

Figure 1. Relevance of input parameters in viscosity prediction Case 1 (3-input model).

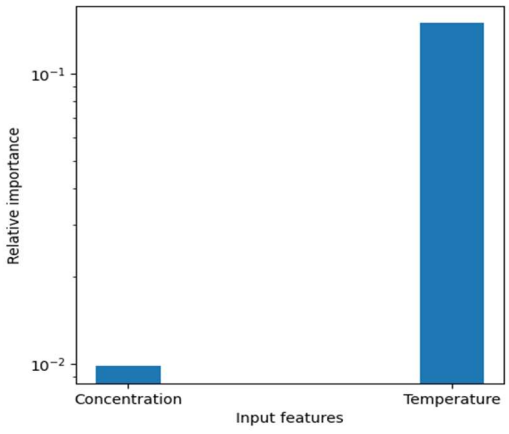

Figure 2. Relevance of input parameters in viscosity prediction Case 2 (2-input model).

Table 5. Evaluated metrics of viscosity for two cases.

Models	Viscosity (3 input model)			Viscosity (2 input model)		
	R^2	MSE	MAE	R^2	MSE	MAE
Random Forest	0.9719	0.0019	0.0391	0.9607	0.0028	0.0453
Extra Tree	0.9771	0.0014	0.0255	0.9665	0.0024	0.0318
Decision Tree	0.9744	0.0015	0.0271	0.9678	0.0023	0.0313
K-Nearest Neighbour	0.9692	0.0020	0.0286	0.9361	0.0051	0.0363
Non-linear regression	0.9471	0.0042	0.0588	0.9050	0.0077	0.0738

When predicting viscosity, it can be seen that for the 3-input model, extra tree gave the best results. And in case of 2-input model, decision tree gave the best results. Random forest, KNN and polynomial regression is also observed to give good results. In theory, a random forest outdoes decision trees because it does not over fit when we add additional trees. Nevertheless, for this data set decision trees also gives good performance, which may be due to the less noisy data used for training and fewer input features. To obtain an explicit mathematical expression between input and output features, multivariate polynomial regression (MPR) can be utilized. In modelling using MPR, all the skewed features were transformed and the input parameters were scaled. Hence the predicted output is linear combinations of function compositions. The mathematical expression of the general non-linear regression model in the case of 2-input for the prediction of viscosity is given in equation 7 and the unknown coefficients are provided in Table 6.

$$\frac{\mu_{nf}}{\mu_{bf}} = a_1 * C' + a_2 * T' + a_3 * C'^2 + a_4 * C'T' + a_5 * T'^2 + a_6 * C'^3 + a_7 * C'^2T' + a_8 * C'T'^2 + a_9 * T'^3 + a_{10} * C'^4 + a_{11} * C'^3T' + a_{12} * C'^2T'^2 + a_{13} * C'T'^3 + a_{14} * T'^4 + a_{15} * C'^5 a_{16} * C'^4T' + a_{17} * C'^3T'^2 + a_{18} * C'^2T'^3 + a_{19} * C'T'^4 + a_{20} * T'^5 + b \quad (7)$$

where, $T' \approx -(T^{-2.5995} + 1.488E - 05)/4.58E - 08$, T is the temperature, $C' \approx (C^{0.214} - 0.38514)/0.13468$, and C is the concentration.

Table 6. Coefficients of general regression equation for viscosity.

a_1 = -0.2646	a_2 = 0.1287	a_3 = 0.0197	a_4 = -0.0036
a_5 = -0.0136	a_6 = 0.0129	a_7 = -0.0180	a_8 = 0.0000
a_9 = -0.0539	a_{10} = 0.0025	a_{11} = -0.0008	a_{12} = 0.0006
a_{13} = -0.0017	a_{14} = 0.0000	a_{15} = -0.0025	a_{16} = 0.0009
a_{17} = -0.0016	a_{18} = 0.0034	a_{19} = 0.0003	a_{20} = 0.0064
b = 0.8611			

The scatter plot of the output from all five models for both cases compared to the experimental data is shown in Figures 3-22. Figure 3-7 represents the actual viscosity values versus the predicted viscosity values of 3-input models for test data obtained from different algorithms. Figure 8-12 represents the actual viscosity values versus the predicted viscosity values of 3-input models for train data obtained from different algorithms. Figure 13-17 represents the actual viscosity values versus the predicted viscosity values of 2-input models for test data obtained from different algorithms. Figure 18-22 represents the actual viscosity values versus the predicted viscosity values of 2-input models for train data obtained from different algorithms. Majority of the data points are located near the line $y = x$, which visually illustrates the model's accuracy. Specifically, it indicates how well the predicted value and experimental values match.

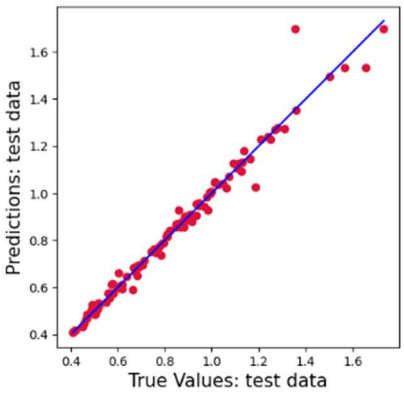

Figure 3. True values vs predicted values using RF.

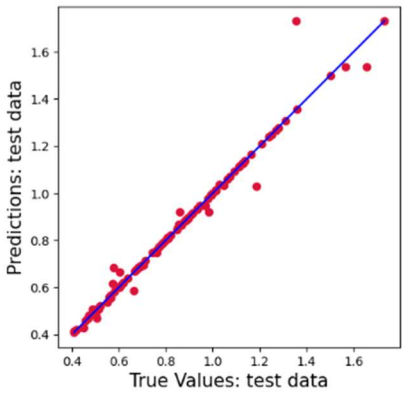

Figure 4. True values vs predicted values using DT.

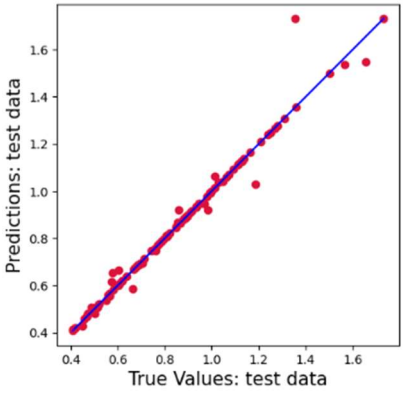

Figure 5. True values vs predicted values using ET.

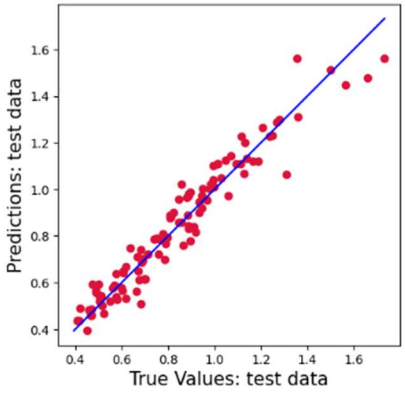

Figure 6. True values vs predicted values using MPR.

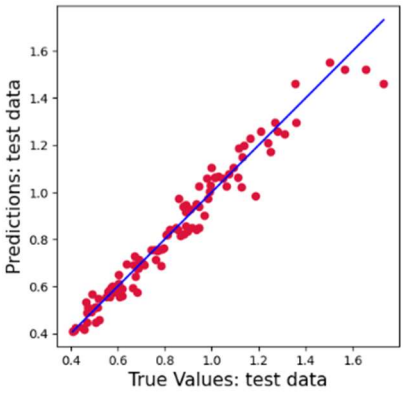

Figure 7. True values vs predicted values using KNN.

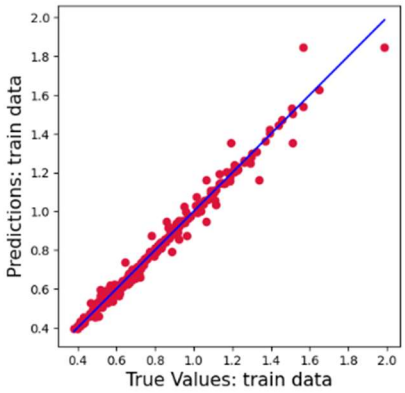

Figure 8. True values vs predicted values using RF.

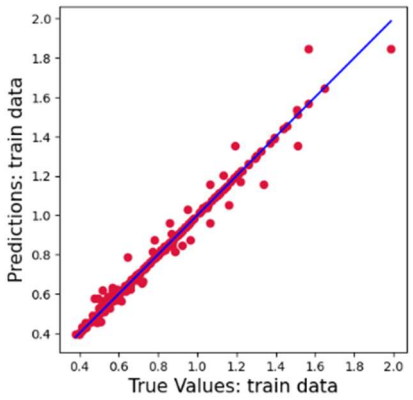

Figure 9. True values vs predicted values using DT.

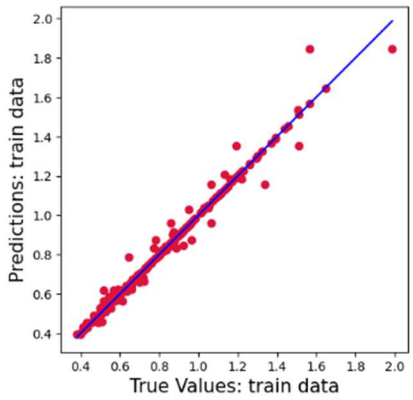

Figure 10. True values vs predicted values using ET.

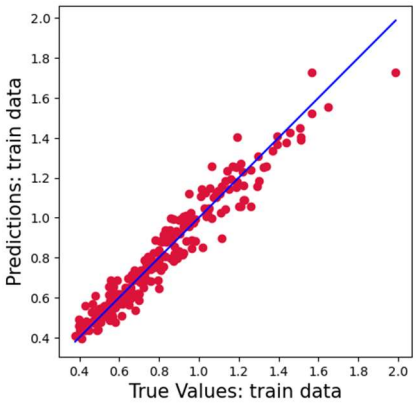

Figure 11. True values vs predicted values using MPR.

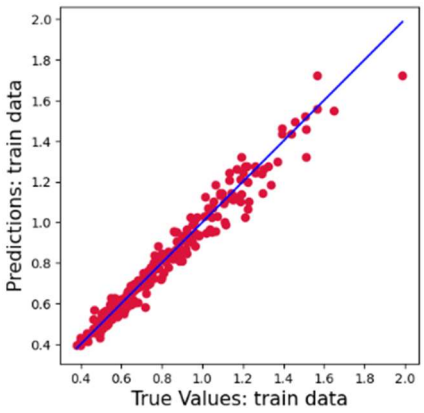

Figure 12. True values vs predicted values using KNN.

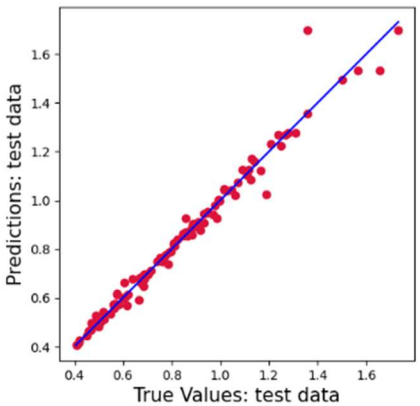

Figure 13. True values vs predicted values using RF.

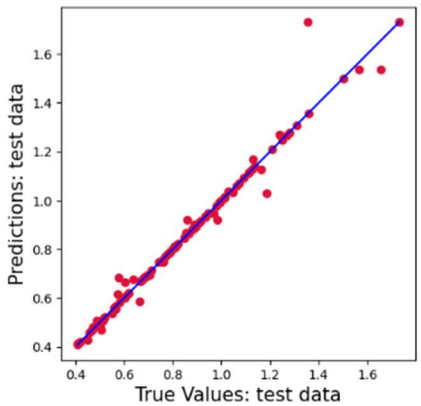

Figure 14. True values vs predicted values using DT.

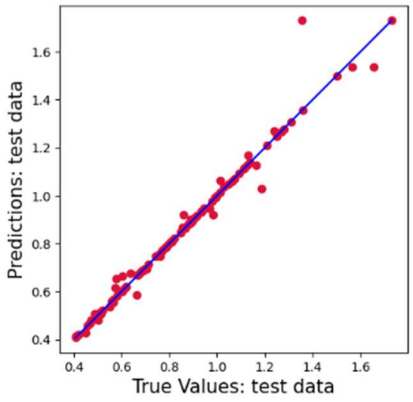

Figure 15. True values vs predicted values using ET.

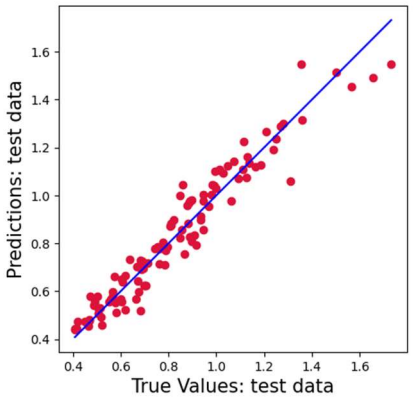

Figure 16. True values vs predicted values using MPR.

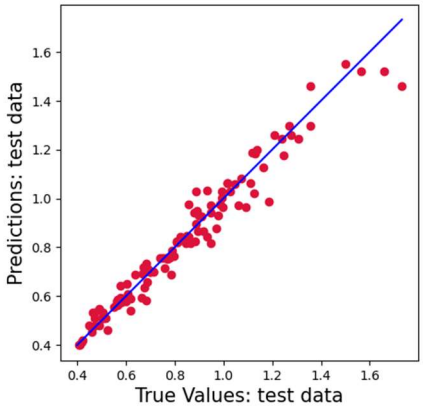

Figure 17. True values vs predicted values using KNN.

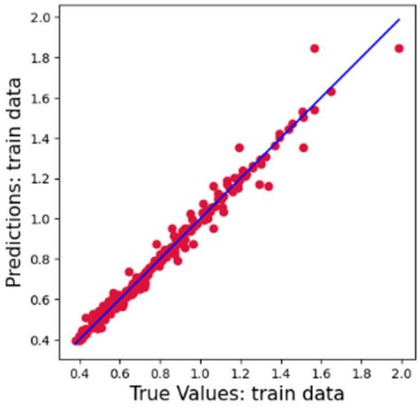

Figure 18. True values vs predicted values using RF.

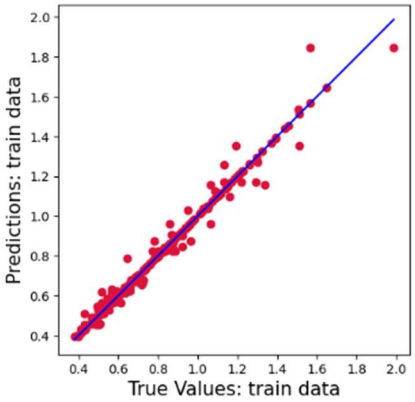

Figure 19. True values vs predicted values using DT.

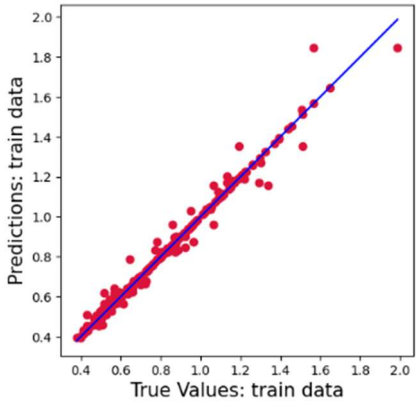

Figure 20. True values vs predicted values using ET.

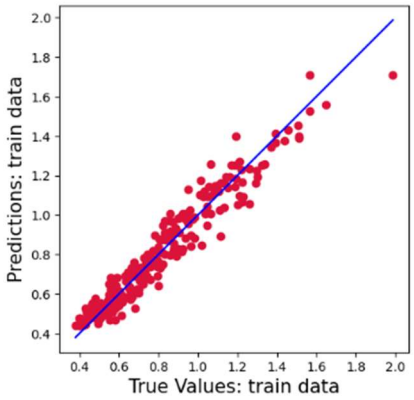

Figure 21. True values vs predicted values using MPR.

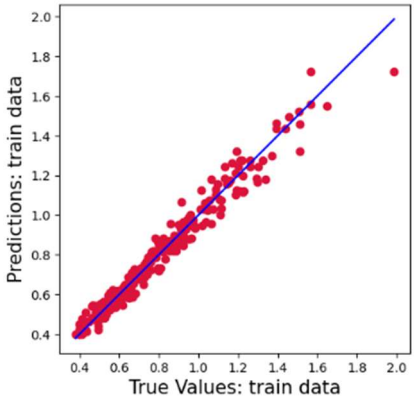

Figure 22. True values vs predicted values using KNN.

4.2 Thermal Conductivity Model

Among the parameters affecting the thermal conductivity of nanofluids, temperature, concentration, and thermal conductivity of nanoparticles were considered as the input parameters. The type of nanoparticle given in the data-set was replaced with the thermal conductivity of the corresponding nanoparticle. Considered nanofluids include: Copper Oxide/water, Aluminium oxide/water, Titanium dioxide/water, Zinc oxide/water, and Silver/water. And the measurements of thermal conductivity of the nanoparticles considered were 18W/mK, 32W/mK, 8.3W/mk, 50W/mK and 406W/mK respectively. The normalized thermal conductivity value is considered as the output parameter for the model prediction. The most governing parameters in predicting thermal conductivity are found in the order T, the thermal conductivity of nanoparticles, and ϕ (Figure 23). For the purpose of predicting thermal conductivity across a larger range of concentrations and temperatures, five different machine learning algorithms were implemented. For the same, different evaluation criteria including coefficient of determination (R^2), Mean squared error (MSE) and mean absolute error (MAE) are used to assess the performance of each model, and the details are summarized in Table 7.

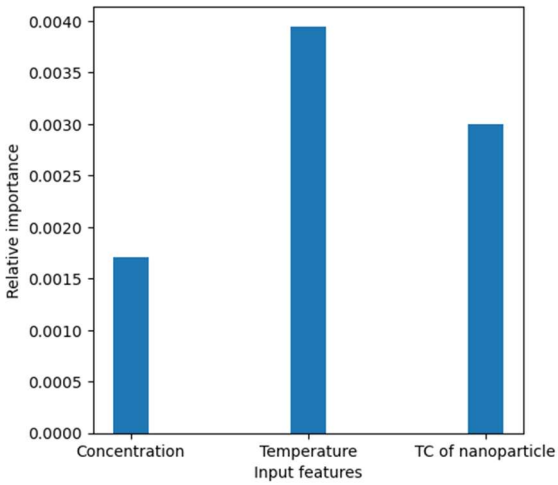

Figure 23. Relevance of input parameters in TC prediction.

Table 7. Evaluated metrics of thermal conductivity.

Models	R^2	MSE	MAE
Extra Tree	0.9403	0.0002	0.0090
Random Forest	0.9234	0.0003	0.0143
K-Nearest Neighbour	0.9206	0.0002	0.0088
Decision Tree	0.8835	0.0003	0.0105
Non-linear regression	0.8218	0.0007	0.0256

It can be observed that the extra trees give the highest R^2 values in the prediction of thermal conductivity, followed by random forest, KNN, and decision trees. From the obtained results for this data-set, extra tree is suggested as a better algorithm. Though relatively, the accuracy of multivariate polynomial regression (MPR) obtained in this study is less, to obtain an explicit mathematical expression between input and output features, this modelling can be utilized. In modelling using MPR, all the skewed features were transformed and the input parameters were scaled. Hence the predicted output is linear combinations of function compositions. So, the mathematical expression of the general non-linear regression model with 3-input for the prediction of thermal conductivity is given in equation 8 and the unknown coefficients are provided in Table 8.

$$\frac{\kappa_{nf}}{\kappa_{bf}} = a_1 * C' + a_2 * T' + a_3 * n' + a_4 * C'^2 + a_5 * C'T' + a_6 * C'n' + a_7 * T'^2 + a_8 * T'n' + a_9 * n'^2 + a_{10} * C'^3 + a_{11} * C'^2T' + a_{12} * C'^2n' + a_{13} * C'T'^2 + a_{14} * C'T'n' + a_{15} * C'n'^2 + a_{16} * T'^3 + a_{17} * T'^2n' + a_{18} * T'n'^2 + a_{19} * n'^3 + a_{20} * C'^4 + a_{21} * C'^3T' + a_{22} * C'^3n' + a_{23} * C'^2T'^2 + a_{24} * C'^2T'n' + a_{25} * C'^2n'^2 + a_{26} * C'T'^3 + a_{27} * C'T'^2n' + a_{28} * C'T'n'^2 + a_{29} * C'n'^3 + a_{30} * T'^4 + a_{31} * T'^3n' + a_{32} * T'^2n'^2 + a_{33} * T'n'^3 + a_{34} * n'^4 + a_{35} * C'^5 + a_{36} * C'^4T' + a_{37} * C'^4n' + a_{38} * C'^3T'^2 + a_{39} * C'^3T'n' + a_{40} * C'^3n'^2 + a_{41} * C'^2T'^3 + a_{42} * C'^2T'^2n' + a_{43} * C'^2T'n'^2 + a_{44} * C'^2n'^3 + a_{45} * C'T'^4 + a_{46} * C'T'^3n' + a_{47} * C'T'^2n'^2 + a_{48} * C'T'n'^3 + a_{49} * C'n'^4 + a_{50} * T'^5 + a_{51} * T'^4n' + a_{52} * T'^3n'^2 + a_{53} * T'^2n'^3 + a_{54} * T'n'^4 + a_{55} * n'^5 + b \quad (8)$$

where, $T' \approx -(T^{-2.8123} - 2.3473E-06)/1.66E-08$, T is the temperature,
$C' \approx (C^{0.3311} - 0.2161)/0.077917$, C is the concentration,
$n' \approx -(n^{-0.4764} - 0.18296)/0.06953$, and n is the thermal conductivity of nanoparticle.

Table 8. Coefficients of general regression equation for thermal conductivity.

a_1 = 0.0601	a_2 = -0.0148	a_3 = -0.0817	a_4 = -0.0109
a_5 = 0.0058	a_6 = -0.0807	a_7 = -0.0131	a_8 = 0.1300
a_9 = -0.0740	a_{10} = -0.0018	a_{11} = 0.0024	a_{12} = -0.0123
a_{13} = -0.0125	a_{14} = -0.0017	a_{15} = -0.0657	a_{16} = 0.0306
a_{17} = 0.0276	a_{18} = 0.1265	a_{19} = 0.0728	a_{20} = 0.0071
a_{21} = 0.0001	a_{22} = 0.0216	a_{23} = 0.0091	a_{24} = 0.0040
a_{25} = -0.0060	a_{26} = -0.0013	a_{27} = 0.0107	a_{28} = -0.0001
a_{29} = 0.0254	a_{30} = -0.0022	a_{31} = -0.0102	a_{32} = -0.0040
a_{33} = -0.0474	a_{34} = 0.0067	a_{35} = 0.0031	a_{36} = -0.0015
a_{37} = -0.0039	a_{38} = -0.0014	a_{39} = -0.0046	a_{40} = -0.0202
a_{41} = -0.0020	a_{42} = -0.0079	a_{43} = 0.0042	a_{44} = 0.0153
a_{45} = 0.0047	a_{46} = 0.0055	a_{47} = 0.0011	a_{48} = -0.0011
a_{49} = 0.0148	a_{50} = -0.0052	a_{51} = -0.0111	a_{52} = -0.0152
a_{53} = -0.0027	a_{54} = -0.0276	a_{55} = -0.0106	b = 1.1405

The scatter plot of the output from all five models compared to the experimental data is shown in Figures 24-33. Figures 24-28 represents the actual TC values versus the predicted TC values of 3-input models for test data obtained from different algorithms. Figure 29-33 represents the actual TC values versus the

predicted TC values of 3-input models for train data obtained from different algorithms. Most of the data points in the cases of RF, DT, ET, and KNN lie close to the line $y = x$, which visually illustrates the model's accuracy. However, comparing to other models, in the case of non-linear regression, certain data points are found to be lying away from the diagonal. For the prediction of thermal conductivity, comparison of the non-linear regression model and extra tree model, used in this study is carried out with two existing models namely maxwell and Xue model. The overall MSE and MAE of the considered models with the exact experimental values are given in Table 9.

Maxwell model and Xue model (Zhang et al., 2021) respectively are described as below:

$$\frac{\kappa_{nf}}{\kappa_{bf}} = \frac{\kappa_p + 2\kappa_{bf} - 2\varphi(\kappa_{bf} - \kappa_p)}{\kappa_p + 2\kappa_{bf} + \varphi(\kappa_{bf} - \kappa_p)} \tag{9}$$

$$\frac{\kappa_{nf}}{\kappa_{bf}} = \frac{1 - \varphi + 2\varphi\left(\frac{\kappa_p}{\kappa_p - \kappa_{bf}}\right) ln\left(\frac{\kappa_p + \kappa_{bf}}{2\kappa_{bf}}\right)}{1 - \varphi + 2\varphi\left(\frac{\kappa_{bf}}{\kappa_p - \kappa_{bf}}\right) ln\left(\frac{\kappa_p + \kappa_{bf}}{2\kappa_{bf}}\right)} \tag{10}$$

where, κ_p is the particle thermal conductivity, κ_{bf} is the thermal conductivity of the base fluid, φ is the volume fraction and κ_{nf} is the thermal conductivity of the nanofluid. And we can observe that the model that performs the best (ET) and the least (MPR) relatively in this study, gives less error as compared to the existing developed models. Hence the models used in this study are reliable in the prediction of thermal conductivity of nanofluids.

Table 9. Overall MSE and MAE of the considered models with the exact experimental values.

Model	MAE	MSE
ET	5.47E-03	1.82E-04
MPR	2.12E-02	7.28E-04
MAXWELL	6.77E-02	7.49E-03
XUE	7.55E-02	1.01E-02

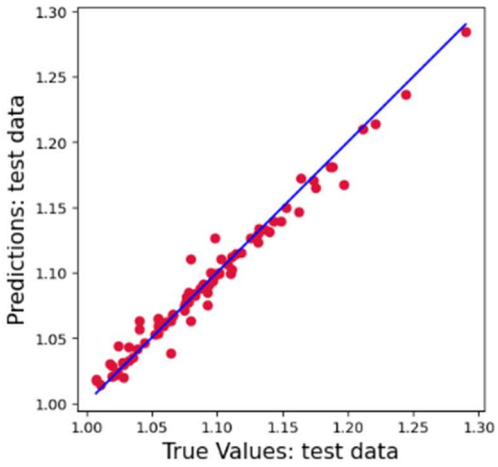

Figure 24. True values vs predicted values using RF.

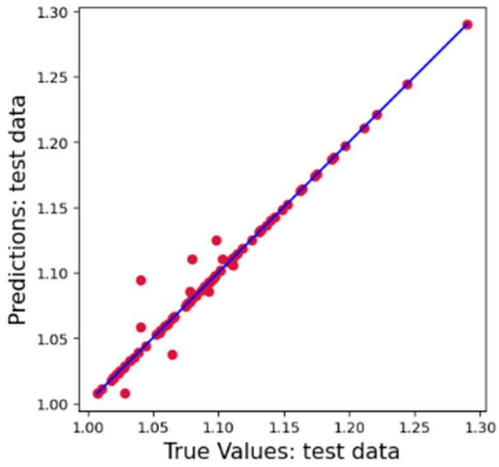

Figure 25. True values vs predicted values using DT.

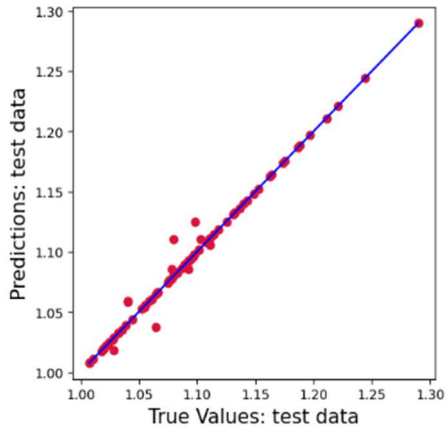

Figure 26. True values vs predicted values using ET.

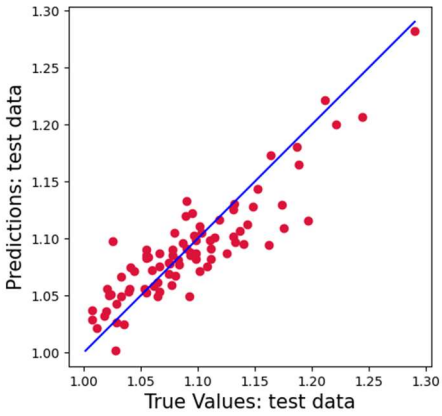

Figure 27. True values vs predicted values using MPR.

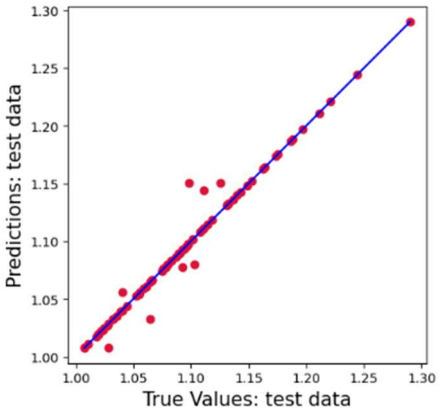

Figure 28. True values vs predicted values using KNN.

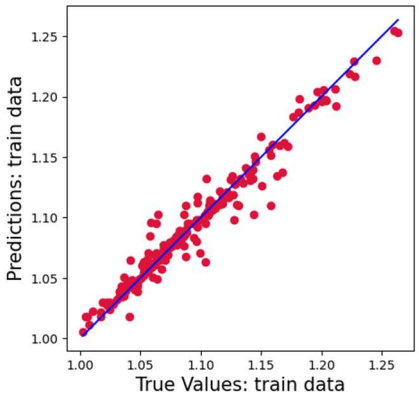

Figure 29. True values vs predicted values using RF.

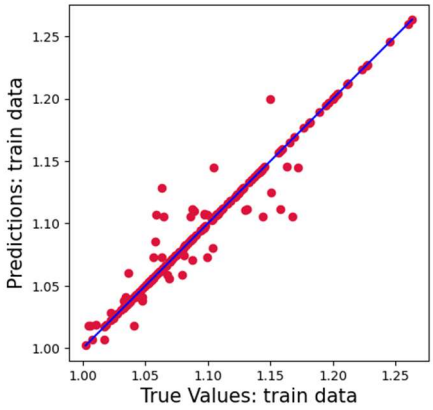

Figure 30. True values vs predicted values using DT.

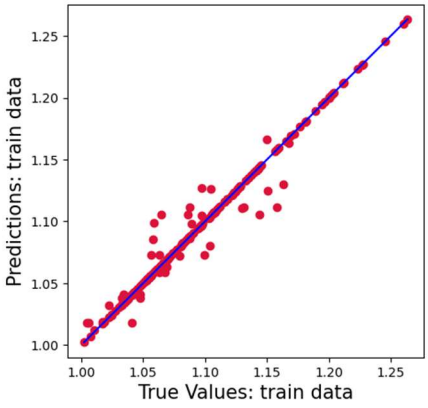

Figure 31. True values vs predicted values using ET.

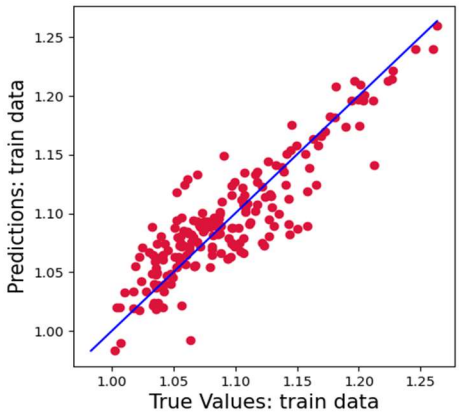

Figure 32. True values vs predicted values using MPR.

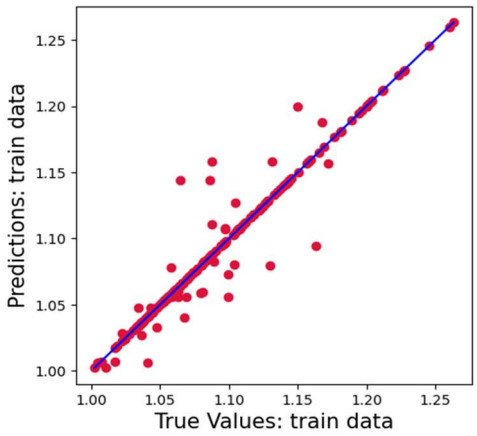

Figure 33. True values vs predicted values using KNN.

5. Conclusion

Five machine learning techniques were used in this study to estimate the thermal conductivity and viscosity of several nanofluids based on water. For the same, in predicting thermal conductivity, the influential properties of the nanoparticles including temperature, concentration, and thermal conductivity of nanoparticles are used as input features and the order of significance of these parameters is determined. And, for the prediction of viscosity, 2-input and 3-input models were considered. While the three-input model also takes into account the type of nanoparticle, both models take into account the temperature and concentration. The order of significance of these parameters is also determined. 523 data points, with concentrations between 0.000221 and 22.9512, and temperatures between 278.15 and 348.15, were considered in predicting viscosity. 302 data points, with concentrations between 0.002 and 5, and temperatures between 278.15 and 363.08 were considered in predicting thermal conductivity. Scatter plots and statistical measures are used to assess the models. From this study, it can be inferred that the prediction of nanofluid's viscosity and thermal conductivity using random forest, decision tree, extra tree, MPR and KNN yields reliable results. When predicting viscosity, it can be seen that for the 3-input model, extra tree gave the best results. And in the case of the 2-input model, decision tree gave the best results and in the prediction of thermal conductivity, extra tree gave the best results. However, to obtain an explicit expression MPR can be utilized. Also, the comparison with the existing thermal conductivity model of Xue and Maxwell is carried out to validate the results obtained in this study and the results are observed to be reliable.

When we have to study heat transfer phenomena of nanofluids, it is important to have the mathematical models for viscosity and thermal conductivity, that satisfy the wide range of experimental data sets. In our future work we will make use of the obtained MPR model and apply it to study the problems arising in heat transfer phenomena.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgements

The authors are thankful for the reviewers for their valuable suggestions.

References

Abdollahi, A., Karimi Darvanjooghi, M.H., Karimipour, A., & Safaei, M.R. (2018). Experimental study to obtain the viscosity of CuO-loaded nanofluid: effects of nanoparticles' mass fraction, temperature and basefluid's types to develop a correlation. *Meccanica*, *53*, 3739-3757.

Afrand, M., Toghraie, D., & Sina, N. (2016). Experimental study on thermal conductivity of water-based Fe3O4 nanofluid: Development of a new correlation and modeled by artificial neural network. *International Communications in Heat and Mass Transfer*, *75*, 262-269.

Agarwal, R., Verma, K., Agrawal, N.K., & Singh, R. (2017). Sensitivity of thermal conductivity for Al2O3 nanofluids. *Experimental Thermal and Fluid Science*, *80*, 19-26.

Ahmadi, M.H., Baghban, A., Ghazvini, M., Hadipoor, M., Ghasempour, R., & Nazemzadegan, M.R. (2020a). An insight into the prediction of TiO2/water nanofluid viscosity through intelligence schemes. *Journal of Thermal Analysis and Calorimetry*, *139*, 2381-2394.

Ahmadi, M.H., Mohseni-Gharyehsafa, B., Farzaneh-Gord, M., Jilte, R.D., Kumar, R., & Chau, K.W. (2019). Applicability of connectionist methods to predict dynamic viscosity of silver/water nanofluid by using ANN-MLP, MARS and MPR algorithms. *Engineering Applications of Computational Fluid Mechanics*, *13*(1), 220-228.

Ahmadi, M.H., Mohseni-Gharyehsafa, B., Ghazvini, M., Goodarzi, M., Jilte, R.D., & Kumar, R. (2020b). Comparing various machine learning approaches in modeling the dynamic viscosity of CuO/water nanofluid. *Journal of Thermal Analysis and Calorimetry*, *139*, 2585-2599.

Alade, I.O., Rahman, M.A.A., Hassan, A., & Saleh, T.A. (2020). Modeling the viscosity of nanofluids using artificial neural network and Bayesian support vector regression. *Journal of Applied Physics*, *128*(8), 085306. https://doi.org/10.1063/5.0008977.

Azmi, W.H., Sharma, K.V., Mamat, R., Alias, A.B.S., & Misnon, I.I. (2012, September). Correlations for thermal conductivity and viscosity of water based nanofluids. In *IOP Conference Series: Materials Science and Engineering* (Vol. 36, No. 1, p. 012029). IOP Publishing. Kuantan, Pahang, Malaysia.

Barbés, B., Páramo, R., Blanco, E., Pastoriza-Gallego, M.J., Piñeiro, M.M., Legido, J.L., & Casanova, C. (2013). Thermal conductivity and specific heat capacity measurements of Al 2O3 nanofluids. *Journal of Thermal Analysis and Calorimetry*, *111*, 1615-1625.

Bhaumik, B., Chaturvedi, S., Changdar, S., & De, S. (2023). A unique physics-aided deep learning model for predicting viscosity of nanofluids. *International Journal for Computational Methods in Engineering Science and Mechanics*, *24*(3), 167-181.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Buonomo, B., Manca, O., Marinelli, L., & Nardini, S. (2015). Effect of temperature and sonication time on nanofluid thermal conductivity measurements by nano-flash method. *Applied Thermal Engineering*, *91*, 181-190.

Changdar, S., Saha, S., & De, S. (2020). A smart model for prediction of viscosity of nanofluids using deep learning. *Smart Science*, *8*(4), 242-256.

Choi, S.U., & Eastman, J.A. (1995). *Enhancing thermal conductivity of fluids with nanoparticles* (No. ANL/MSD/CP-84938; CONF-951135-29). Argonne National Lab (ANL), Argonne, IL (United States).

Colla, L., Marinelli, L., Fedele, L., Bobbo, S., & Manca, O. (2015). Characterization and simulation of the heat transfer behaviour of water-based ZnO nanofluids. *Journal of Nanoscience and Nanotechnology*, *15*(5), 3599-3609.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21-27.

Dai, X., Andani, H.T., Alizadeh, A.A., Abed, A.M., Smaisim, G.F., Hadrawi, S.K., Karimi, M., Shamsborhan, M., & Toghraie, D. (2023). Using Gaussian process regression (GPR) models with the Matérn covariance function to predict the dynamic viscosity and torque of SiO2/Ethylene glycol nanofluid: A machine learning approach. *Engineering Applications of Artificial Intelligence*, *122*, 106107. https://doi.org/10.1016/j.engappai.2023.106107.

Durgam, S., & Kadam, G. (2021). Investigation on thermal conductivity and viscosity of nanofluids using analytical and machine learning models. *arXiv preprint arXiv:2102.12120*.

Ferrouillat, S., Bontemps, A., Poncelet, O., Soriano, O., & Gruss, J.A. (2013). Influence of nanoparticle shape factor on convective heat transfer and energetic performance of water-based SiO2 and ZnO nanofluids. *Applied Thermal Engineering*, *51*(1-2), 839-851.

Fix, E., & Hodges, J.L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique*, *57*(3), 238-247.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3-42.

Ghodsinezhad, H., Sharifpur, M., & Meyer, J.P. (2016). Experimental investigation on cavity flow natural convection of Al2O3–water nanofluids. *International Communications in Heat and Mass Transfer*, *76*, 316-324.

Gholizadeh, M., Jamei, M., Ahmadianfar, I., & Pourrajab, R. (2020). Prediction of nanofluids viscosity using random forest (RF) approach. *Chemometrics and Intelligent Laboratory Systems*, *201*, 104010. https://doi.org/10.1016/j.chemolab.2020.104010.

Hemmat Esfe, M., Ahangar, M.R.H., Toghraie, D., Hajmohammad, M.H., Rostamian, H., & Tourang, H. (2016). Designing artificial neural network on thermal conductivity of Al2O3–water–EG (60–40 %) nanofluid using experimental data. *Journal of Thermal Analysis and Calorimetry*, *126*(2), 837-843.

Hemmati-Sarapardeh, A., Varamesh, A., Husein, M.M., & Karan, K. (2018). On the evaluation of the viscosity of nanofluid systems: Modeling and data assessment. *Renewable and Sustainable Energy Reviews*, *81*, 313-329.

Hemmati-Sarapardeh, A., Varamesh, A., Nait Amar, M., Husein, M.M., & Dong, M. (2020). On the evaluation of thermal conductivity of nanofluids using advanced intelligent models. *International Communications in Heat and Mass Transfer*, *118*, 104825. https://doi.org/10.1016/j.icheatmasstransfer.2020.104825.

Ho, C.J., Liu, W.K., Chang, Y.S., & Lin, C.C. (2010). Natural convection heat transfer of alumina-water nanofluid in vertical square enclosures: An experimental study. *International Journal of Thermal Sciences*, *49*(8), 1345-1353.

Ho, T.K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278-282). IEEE. Montreal, QC, Canada.

Hojjat, M., Etemad, S.G., Bagheri, R., & Thibault, J. (2011). Thermal conductivity of non-Newtonian nanofluids: Experimental data and modeling using neural network. *International Journal of Heat and Mass Transfer*, *54*(5-6), 1017-1023.

Kim, S.H., Choi, S.R., & Kim, D. (2007). Thermal conductivity of metal-oxide nanofluids: Particle size dependence and effect of laser irradiation. *Journal of Heat Transfer*, *129*(3), 298-307.

Kumar, P.C.M., & Kavitha, R. (2021). Prediction of nanofluid viscosity using multilayer perceptron and Gaussian process regression. *Journal of Thermal Analysis and Calorimetry*, *144*(4), 1151-1160.

Kumar, V., Tiwari, A.K., & Ghosh, S.K. (2016). Effect of variable spacing on performance of plate heat exchanger using nanofluids. *Energy*, *114*, 1107-1119.

Longo, G.A., Zilio, C., Ceseracciu, E., & Reggiani, M. (2012). Application of artificial neural network (ANN) for the prediction of thermal conductivity of oxide–water nanofluids. *Nano Energy*, *1*(2), 290-296.

Madhesh, D., Parameshwaran, R., & Kalaiselvam, S. (2016). Experimental studies on convective heat transfer and pressure drop characteristics of metal and metal oxide nanofluids under turbulent flow regime. *Heat Transfer Engineering*, *37*(5), 422-434.

Mahian, O., Kolsi, L., Amani, M., Estellé, P., Ahmadi, G., Kleinstreuer, C., Marshall, J.S., Siavashi, M., Taylor, R.A., Niazmand, H., Wongwises, S., Hayat, T., Kolanjiyil, A., Kasaeian, A., & Pop, I. (2019). Recent advances in modeling and simulation of nanofluid flows-Part I: Fundamentals and theory. *Physics Reports*, *790*, 1-48.

Maleki, A., Haghighi, A., & Mahariq, I. (2021). Machine learning-based approaches for modeling thermophysical properties of hybrid nanofluids: A comprehensive review. *Journal of Molecular Liquids*, *322*, 114843. https://doi.org/10.1016/j.molliq.2020.114843.

McCants, D.A., Ali, M.Y., & Khan, J. (2009). Effective viscosity measurement of CuO and ZnO nanofluids. In *International Conference on Micro/Nanoscale Heat Transfer* (Vol. 43895, pp. 607-615). Shanghai, China.

Mehrabi, M., Sharifpur, M., & Meyer, J.P. (2012). Application of the FCM-based neuro-fuzzy inference system and genetic algorithm-polynomial neural network approaches to modelling the thermal conductivity of alumina–water nanofluids. *International Communications in Heat and Mass Transfer*, *39*(7), 971-977.

Meybodi, M.K., Naseri, S., Shokrollahi, A., & Daryasafar, A. (2015). Prediction of viscosity of water-based Al2O3, TiO2, SiO2, and CuO nanofluids using a reliable approach. *Chemometrics and Intelligent Laboratory Systems*, *149*, 60-69.

Mondejar, M.E., Regidor, M., Krafczyk, J., Ihmels, C., Schmid, B., Kontogeorgis, G.M., & Haglind, F. (2021). An open-access database of the thermophysical properties of nanofluids. *Journal of Molecular Liquids*, *333*, 115140. https://doi.org/10.1016/j.molliq.2020.115140.

Naseri, A., Jamei, M., Ahmadianfar, I., & Behbahani, M. (2020). Nanofluids thermal conductivity prediction applying a novel hybrid data-driven model validated using Monte Carlo-based sensitivity analysis. *Engineering with Computers*, *38*, 815-839.

Pak, B.C., & Cho, Y.I. (2007). Hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. *Experimental Heat Transfer*, *11*(2), 151-170.

Pantzali, M.N., Mouza, A.A., & Paras, S.V. (2009). Investigating the efficacy of nanofluids as coolants in plate heat exchangers (PHE). *Chemical Engineering Science*, *64*(14), 3290-3300.

Pastoriza-Gallego, M.J., Casanova, C., Páramo, R., Barḃs, B., Legido, J.L., & Piñeiro, M.M. (2009). A study on stability and thermophysical properties (density and viscosity) of Al2 O3 in water nanofluid. *Journal of Applied Physics*, *106*(6), 064301. https://doi.org/10.1063/1.3187732.

Patil, V.S., Cera-Manjarres, A., Salavera, D., Rode, C.V., Patil, K.R., & Coronas, A. (2018). Influence of silver nanoparticles morphologies on density, viscosity and thermal conductivity of silver nanofluids and silver Ionanofluids. *Journal of Nanofluids*, *7*(2), 246-257.

Paul, T.C., Morshed, A.K.M.M., McCants, D.A., & Khan, J.A. (2013). Buoyancy driven heat transfer behavior of zinc oxide (zno)–water nanofluids. In *Heat Transfer Summer Conference* (Vol. 55478, p. V001T03A008). American Society of Mechanical Engineers. Minneapolis, Minnesota, USA.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81-106.

Ramalingam, S., Dhairiyasamy, R., Rajendran, S., & Radhakrishnan, M. (2016). Contraction of radiator length in heavy vehicles using cerium oxide nanofluid by enhancing heat transfer performance. *Thermal Science*, *20*, 1037-1044.

Ramezanizadeh, M., Ahmadi, M.A., Ahmadi, M.H., & Alhuyi Nazari, M. (2018). Rigorous smart model for predicting dynamic viscosity of Al2O3/water nanofluid. *Journal of Thermal Analysis and Calorimetry, 137*(1), 307-316.

Sahin, F., Genc, O., Gökcek, M., & Çolak, A.B. (2023). An experimental and new study on thermal conductivity and zeta potential of Fe3O4/water nanofluid: Machine learning modeling and proposing a new correlation. *Powder Technology*, *420*, 118388. https://doi.org/10.1016/j.powtec.2023.118388.

Saxena, R., Gangacharyulu, D., & Bulasara, V.K. (2016). Heat transfer and pressure drop characteristics of dilute alumina–water nanofluids in a pipe at different power inputs. *Heat Transfer Engineering*, *37*(18), 1554-1565.

Sharma, P., Ramesh, K., Parameshwaran, R., & Deshmukh, S.S. (2022). Thermal conductivity prediction of titania-water nanofluid: A case study using different machine learning algorithms. *Case Studies in Thermal Engineering*, *30*, 101658. https://doi.org/10.1016/j.csite.2021.101658.

Shateri, M., Sobhanigavgani, Z., Alinasab, A., Varamesh, A., Hemmati-Sarapardeh, A., & Mosavi, A. (2020). Comparative analysis of machine learning models for nanofluids viscosity assessment. *Nanomaterials*, *10*(9), 1767. https://doi.org/10.3390/nano10091767.

Song, Y.Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130-135.

Suresh, T., Uthayakumar, G., & Srinivasan, R. (2016). Synthesis and characterization of CeO 2/water nanofluids and its thermal properties. In *2016 International Conference on Energy Efficient Technologies for Sustainability* (pp. 202-205). IEEE. Nagercoil, India.

Tso, C.Y., & Chao, C.Y.H. (2015). Study of enthalpy of evaporation, saturated vapor pressure and evaporation rate of aqueous nanofluids. *International Journal of Heat and Mass Transfer*, *84*, 931-941.

Yousefi, F., Karimi, H., & Papari, M.M. (2012). Modeling viscosity of nanofluids using diffusional neural networks. *Journal of Molecular Liquids*, *175*, 85-90.

Zhang, Y., & Xu, X. (2020). Predicting the thermal conductivity enhancement of nanofluids using computational intelligence. *Physics Letters A*, *384*(20), 126500. https://doi.org/10.1016/j.physleta.2020.126500.

Zhang, Y., Shahmir, N., Ramzan, M., Ghazwani, H.A.S., & Malik, M.Y. (2021). Comparative analysis of Maxwell and Xue models for a hybrid nanofluid film flow on an inclined moving substrate. *Case Studies in Thermal Engineering*, *28*, 101598. https://doi.org/10.1016/j.csite.2021.101598.

Zhu, B.J., Zhao, W.L., Li, J.K., Guan, Y.X., & Li, D.D. (2011). Thermophysical properties of Al2O3-water nanofluids. *Materials Science Forum*, *688*, 266-271.