Predicting Specific Heat Capacity of Nanofluids Using Machine Learning Methods: Applications in Thermophysics

Y.E. Barablina, V.A. Raudsep

ITMO University

This study presents the development of a predictive model for the specific heat capacity of nanofluids based on the Random Forest Regressor method, optimized using GridSearchCV. The final model parameters demonstrated high prediction accuracy (MSE = 4.16, R^2 = 0.99999), as confirmed by residual analysis and comparison between actual and predicted values. The model was successfully tested on data for isopropyl alcohol with nanoparticles, showing minimal deviations from the experimental values. Despite limitations associated with the clustered structure of the data, the model exhibits potential for application to other base fluids and nanoparticles, making it a valuable tool for studying the thermophysical properties of nanofluids and developing new materials.

Keywords: nanofluids, heat capacity, RandomForestRegressor, machine learning, thermophysical properties.

1. Introduction

Nanofluids, consisting of nanoparticles suspended in base fluids, have garnered significant attention due to their enhanced thermophysical properties [1]. These fluids exhibit improved thermal conductivity and altered viscosity, making them promising for heat transfer applications [2]. Nanoparticle characteristics, such as concentration, size, and shape, significantly impact nanofluid properties and performance [3]. Metallic oxide nanoparticles like MgO, TiO_2, and ZnO can increase heat transfer efficiency by approximately 30% compared to base fluids [3]. Stability is crucial for nanofluid usability, with various methods employed to enhance it, including surfactant addition and ultrasonic mixing [3]. While increasing nanoparticle concentration can improve heat transfer efficiency, exceeding optimal levels may reduce Brownian motions due to higher viscosity and density [3]. Density, thermal conductivity, viscosity, and heat capacity are key parameters influencing nanofluid performance in heat transfer applications [4].

Recent studies highlight the growing importance of machine learning (ML) techniques in predicting thermophysical properties of nanofluids. These advanced computational methods have proven superior to traditional approaches in handling complex, non-linear relationships between nanofluid parameters [5, 6]. Various ML algorithms, including Artificial Neural Networks (ANNs), Support Vector Regression (SVR), and genetic algorithms, have been employed to accurately predict properties such as thermal conductivity, viscosity, and specific heat capacity [6–8]. The input parameters typically include temperature, concentration, nanoparticle size, and base fluid type [7, 6]. Researchers have found that extra trees and decision trees often provide the best results for estimating thermal conductivity and viscosity, respectively [7]. These ML methods offer efficient, rapid, and practical alternatives to experimental approaches, making them valuable tools for researchers and engineers in the field of nanofluids [8–11].

Machine learning techniques have emerged as powerful tools for predicting and optimizing nanofluid thermophysical properties, offering superior accuracy compared to traditional methods [12, 13]. Various approaches, including Artificial Neural Networks (ANNs), Support Vector Regression (SVR), and genetic algorithms, have demonstrated effectiveness in capturing the complex dynamics of nanofluids [12, 14]. These AI-based models excel at processing large datasets and identifying intricate patterns, making them particularly suitable for predicting properties such as thermal conductivity, viscosity, and specific heat capacity [14, 15]. The integration of machine learning with empirical data has shown promise in enhancing the efficiency and accuracy of predictions, outperforming classical approaches [12, 14]. This advancement in predictive capabilities facilitates the optimization of nanofluid

compositions for specific technical applications, marking a significant step forward in thermal engineering and management [12, 15].

Recent research highlights the potential of machine learning in nanofluid design and analysis. A novel machine learning framework coupled with molecular dynamics methods has been proposed to model multi-component mixing nanofluidic systems, significantly reducing computational costs and improving prediction accuracy for surfactant adsorption properties and heat transfer performance [16]. Machine learning techniques have been shown to efficiently predict thermophysical properties of nanofluids, offering advantages over experimental methods [13]. Various machine learning algorithms, including artificial neural networks, genetic algorithms, and ensemble techniques like Boosted regression and XGBoost, have been applied to nanofluid-based heat transfer studies in renewable energy systems [17]. In a specific study on ethylene glycol- and glycerol-based SiO2 nanofluids, decision tree-based models performed well in predicting thermal performance, with random forest and extreme gradient boosting models also showing high accuracy [18]. These advancements demonstrate the growing importance of machine learning in nanofluid research and applications.

The aim of this study is to develop and optimize a predictive model for estimating the specific heat capacity of nanofluids using the Random Forest Regressor method. The relevance of this work is driven by the widespread application of nanofluids in various fields, including heat exchange systems, energy technologies, and cooling systems, where accurate knowledge of their thermophysical properties plays a critical role. However, the experimental determination of specific heat capacity is often associated with significant time and resource costs, making predictive models a promising tool for addressing this challenge.

The research addresses the following objectives: analyzing data on the specific heat capacity of various nanofluids, developing a predictive model based on the Random Forest Regressor method, optimizing its hyperparameters using the GridSearchCV method, and evaluating the model's accuracy using standard metrics (MSE, R^2). Additionally, the model is tested on new data, including experimental values for isopropyl alcohol with Al_2O_3 nanoparticles, to assess its applicability and generalizability. Particular attention is given to analyzing the model's limitations related to data clustering and its potential adaptation for other base fluids and types of nanoparticles.

The findings of this study may contribute to accelerating the development of new nanofluids with desired thermophysical properties and enhancing the understanding of heat transfer mechanisms in such systems.

2. Methods

In the study, data on nanofluids were used, which are available in the dataset on the Kaggle platform [19]. This dataset contains information on various properties of nanofluids, including parameters such as nanoparticle size, nanoparticle composition, concentration, temperature, type of base fluid, and specific heat capacity. The objective of this research is to predict the specific heat capacity of nanofluids based on these properties.

At the first stage, data preprocessing was performed. The first two columns, containing categorical data unrelated to numerical values, were removed from the dataset, and all missing commas were replaced with periods. This resulted in a cleaned sample suitable for further analysis. Outliers were also removed. To analyze the distribution of the data, plots were created to display the distribution of various characteristics of the nanofluids (Fig. 1).

At the first stage, data preprocessing was performed. The first two columns, containing categorical data unrelated to numerical values, were removed from the dataset, and all missing commas were replaced with periods. This resulted in a cleaned sample suitable for further analysis. Outliers were also removed. To analyze the distribution of the data, plots were created to display the distribution of various characteristics of the nanofluids (Fig. 1).

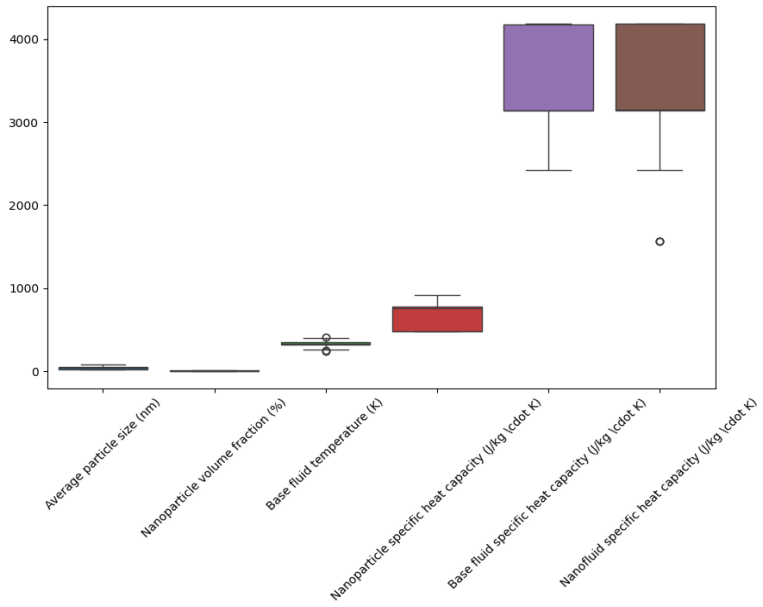

Fig. 1. Distribution of variables in the processed dataset

To gain a deeper understanding of the relationships between the different characteristics of nanofluids, a correlation analysis between the features was conducted. To assess the degree of association between the variables, a Pearson correlation analysis was performed, which quantitatively evaluates the linear relationship between numerical features. As a result, correlation matrices were obtained, clearly illustrating the degree of interconnection between all the variables. For a more visual representation of the correlations, a heatmap was created, displaying the correlation coefficients between all features in the dataset. The most highly correlated features are the heat capacity of the nanofluid and the heat capacity of the base fluid (Fig. 2).

To train the model and assess its performance, the data was split into two parts: a training set and a test set. The standard train_test_split function from the sklearn library was used to randomly divide the data, with 80% of the data allocated for training the model and the remaining 20% for testing. The target variable selected was the nanofluid specific heat capacity, which is predicted based on other features. To ensure that the variables were evenly distributed across the samples, distribution plots were created (Fig. 3).

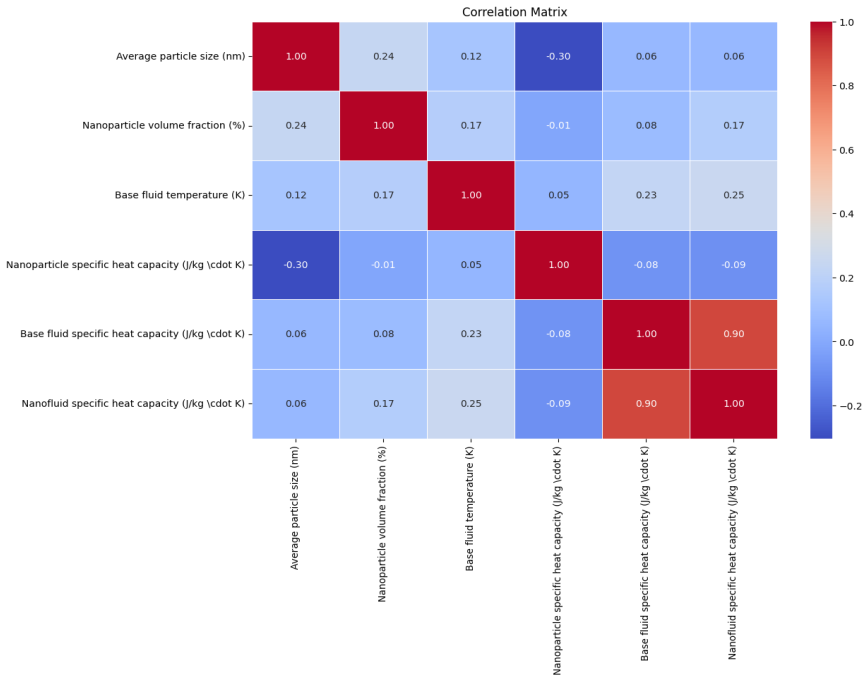

Fig. 2. Correlation matrix of variables in the studied dataset

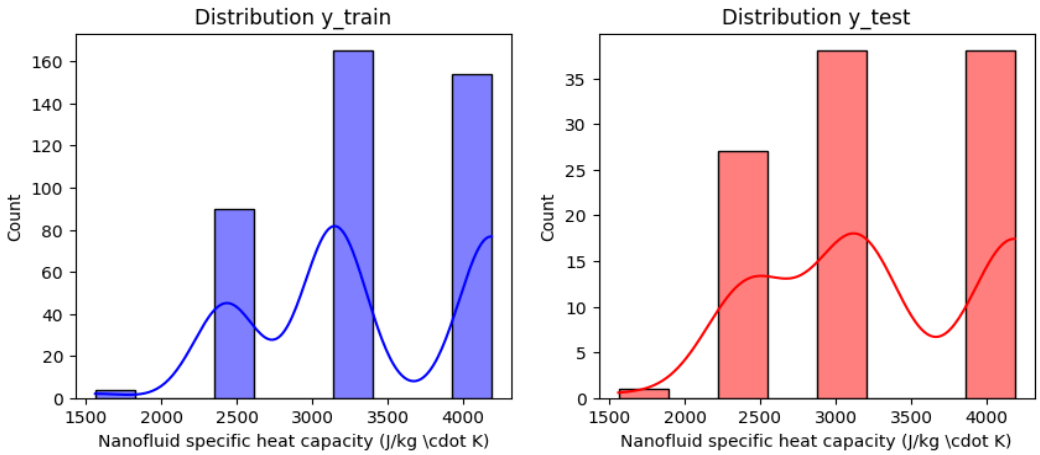

Fig. 3. Distribution plots for the test and training datasets

For the correct functioning of the model, all numerical features were standardized using the StandardScaler method. This step was necessary because different features may have varying scales (for example, temperature and density), which could impact the model's performance.

To predict the specific heat capacity, the RandomForestRegressor method was chosen. This algorithm is well-suited for regression tasks as it can model complex relationships between features and the target variable without making assumptions about the data's distribution. The Random Forest model constructs an ensemble of decision trees and uses the average of their predictions to produce a more accurate result.

To improve model performance, hyperparameter tuning was performed using GridSearchCV. This technique helps find the optimal parameters for the model, such as the number of trees in the forest (n_estimators), the maximum depth of the trees (max_depth), the minimum number of samples required to split a node (min_samples_split), and others. This enhances the model's accuracy and prevents overfitting.

During the training phase, cross-validation with 3 folds was employed for hyperparameter optimization, running the model 3600 times for 1200 different hyperparameter combinations. The best hyperparameters obtained after optimization were: max_depth=None, max_features=None, min_samples_leaf=1, min_samples_split=2, and n_estimators=42.

After optimizing the hyperparameters, the Random Forest model showed excellent results on the test set, with a mean squared error (MSE) of 4.16 and a coefficient of determination (R^2) of 0.99999, indicating extremely high prediction accuracy and the closeness of the predicted values to the actual data.

Mean Squared Error (MSE):

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

where y_i — are the actual values, $\hat{y}_i$ — are the predicted values, and n is the number of observations.

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \underline{y}\right)^2},$$

where $\underline{y}$ — is the mean of the actual values.

These results indicate that the selected model with optimized parameters is capable of predicting the specific heat capacity of nanofluids with high accuracy based on the input features.

To further improve the results and assess the possibility of applying other methods, a Gradient Boosting model based on the XGBoost algorithm was used. XGBoost is one of the most efficient methods for regression and classification tasks. Gradient Boosting helps reduce overfitting and improve prediction accuracy by sequentially training weak models, where each model corrects the errors of the previous one. For model tuning, the following parameter set was chosen: number of trees (n_estimators=1000), learning rate (learning_rate=0.01), and a fixed random state (random_state=80).

3. Results

The final parameters of the model's learning quality are as follows: the mean squared error (MSE) was 0.0138, the mean absolute error (MAE) was 0.0855, and the coefficient of determination (R^2) reached a value of 0.9999999734. These results confirm the high accuracy of the heat capacity prediction for nanofluids, indicating the model's ability to precisely predict the system's behavior, even with minimal errors.

To evaluate the prediction quality, several graphs were constructed.

1. Comparison of actual and predicted values (Fig. 4): The first graph presents the real and predicted values of heat capacity. The line of perfect agreement, which runs diagonally, represents the theoretical accuracy of predictions. The graph demonstrates that the model provides predictions that are close to the actual values, which is further supported by the model's high accuracy.
2. Residual distribution in Fig. 5: The second graph illustrates the distribution of residuals (the difference between actual and predicted values). The distribution is close to normal, indicating good model quality and the absence of systematic errors. The small magnitude of the standard deviation of the residuals further confirms this.

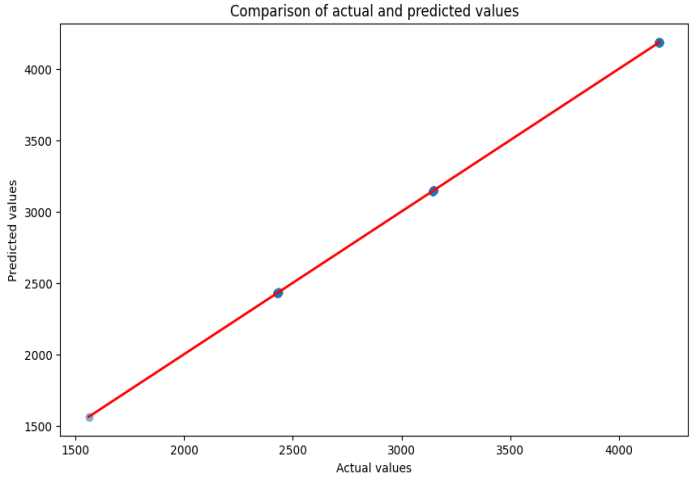

Fig. 4. Visualization of Actual vs Predicted Values

3. Residuals vs Predicted Values (Fig. 6): The third graph displays the distribution of residuals relative to the predicted values. The line positioned at zero helps to visually demonstrate that the residuals barely deviate from zero, further confirming the high accuracy of the model.

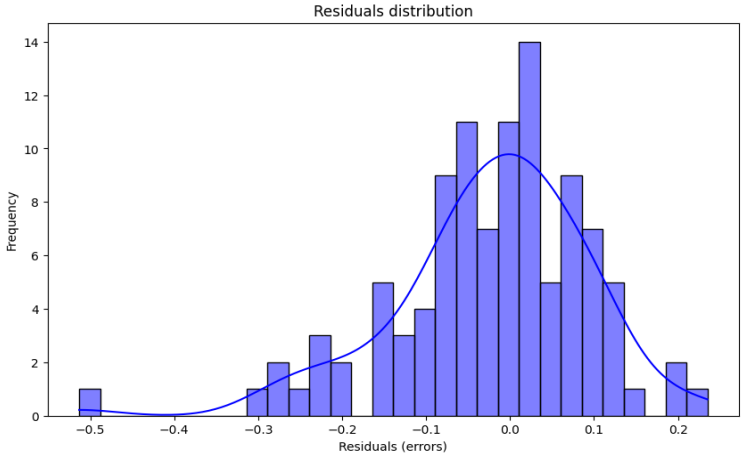

Fig. 5. Visualization of Residuals

4. Residuals vs Predicted Values (Fig. 6): The third graph displays the distribution of residuals relative to the predicted values. The line positioned at zero helps to visually demonstrate that the residuals barely deviate from zero, further confirming the high accuracy of the model.

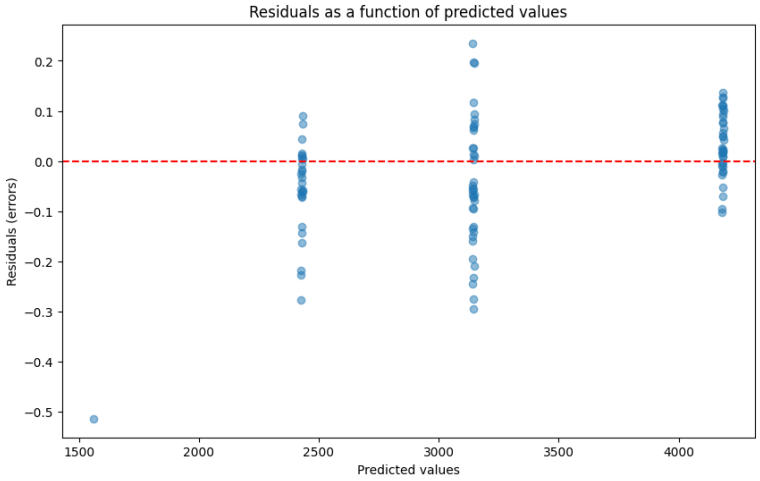

Fig. 6. Residuals vs Predicted Values Plot

Mean and Standard Deviation of Residuals: The mean value of the residuals was close to zero, and the standard deviation was very low, further confirming the high accuracy of the model's performance.

Additionally, the model was tested on random experimental data obtained from the study "Effect of Al_2O_3 Nanoparticle Impurities on the Heat Capacity of Isopropyl Alcohol" [20]. The experimental data included various concentrations of nanoparticles in isopropyl alcohol, allowing for further validation of the model's applicability in real-world chemical systems. The following sample characteristics were used during the validation:

- average particle size: 53 nm;
- volume fraction of nanoparticles: 2%;
- base fluid temperature: 293 K;
- specific heat capacity of nanoparticles: 773 J/kg·K;
- specific heat capacity of base fluid: 2400 J/kg·K.

Using these data, the model predicted the specific heat capacity of the nanofluid to be 2428.32 J/kg·K. This value was found to be close to the actual experimental data, confirming the high accuracy of the model's predictions.

The results demonstrated that the model effectively predicts the heat capacity of isopropyl alcohol with Al_2O_3 nanoparticles, confirming its universality and precision.

4. Conclusion

The proposed solution for predicting the specific heat capacity of nanofluids, despite its high accuracy, still demonstrates some deviation from ideal results. This can be attributed to the fact that the initial dataset exhibits a clustered structure, which may affect the model's ability to generalize to new, previously unseen data. Nevertheless, the model achieves excellent performance on both test and experimental datasets, confirming its effectiveness in predicting the thermophysical properties of nanofluids.

It is worth noting that the developed model for predicting the specific heat capacity of nanofluids, based on the current dataset, can be successfully applied to other fluids and nanoparticles not included in the original dataset. This opens up opportunities for its broader application in various fields, including the study of new materials and systems where accurate predictions of thermophysical properties are required.

Future work will focus on expanding the current approach by increasing the size of the dataset and incorporating a wider range of input parameters. This is expected to enhance the model's quality and improve its generalization capabilities. Additionally, further research will involve comparing different

predictive models, including neural networks, which may potentially yield better results, particularly when working with larger and more diverse datasets.

Thus, further improvement of the model and its adaptation to new types of data present promising prospects for developing even more accurate and versatile tools for predicting the thermophysical properties of nanofluids.

References

1. Kalsi S., Kumar S., Kumar A. et al. Thermophysical properties of nanofluids and their potential applications in heat transfer enhancement: A review // Arabian Journal of Chemistry. 2023. Vol. 16, no. 11. P. 105272. DOI: 10.1016/j.arabjc.2023.105272.
2. Ali A.R.I., Salam B. A review on nanofluid: preparation, stability, thermophysical properties, heat transfer characteristics and application // SN Applied Sciences. 2020. Vol. 2. P. 1636. DOI: 10.1007/s42452-020-03427-1.
3. Rahman M.A., Hasnain S.M.M., Pandey S. et al. Review on Nanofluids: Preparation, Properties, Stability, and Thermal Performance Augmentation in Heat Transfer Applications // ACS Omega. 2024. Vol. 9, no. 30. P. 328–339. DOI: 10.1021/acsomega.4c03279.
4. Arslan R., Özdemir V.A., Akyol E. et al. Thermophysical Properties of Nanofluids // Current Nanoscience. 2021. Vol. 17, no. 5. DOI: 10.2174/1573413716999201228130519.
5. Basu A., Saha A., Banerjee S. et al. A Review of Artificial Intelligence Methods in Predicting Thermophysical Properties of Nanofluids for Heat Transfer Applications // Energies. 2024. P. 56–89.
6. Ganga S., Uddin Z., Asthana R. et al. Modelling of Viscosity and Thermal Conductivity of Water-Based Nanofluids using Machine-Learning Techniques // International Journal of Mathematical, Engineering and Management Sciences. 2023. P. 24–27.
7. Wang H.B., Chen X. A Comprehensive Review of Predicting the Thermophysical Properties of Nanofluids Using Machine Learning Methods // Industrial Engineering & Chemistry Research. 2022. Vol. 61, no. 40. P. 14711–14730. DOI: 10.1021/acs.iecr.2c02059.
8. Apmann K., Fulmer R., Elias Soto A.A., Vafaei S. Thermal Conductivity and Viscosity: Review and Optimization of Effects of Nanoparticles // Materials. 2021. Vol. 14, no. 5. P. 78. DOI: 10.3390/ma14051291.
9. Urmi W.T., Rahman M.M., Kadirgama K. et al. A Comprehensive Review on Thermal Conductivity and Viscosity of Nanofluids // Journal of Advanced Research in Fluid Mechanics and Thermal Sciences. 2022. P. 5.
10. Arslan R., Özdemir V.A., Akyol E. et al. Thermophysical Properties of Nanofluids // Current Nanoscience. 2020. P. 16–18.
11. Ali A.R.I., Salam B. A review on nanofluid: preparation, stability, thermophysical properties, heat transfer characteristics and application // SN Applied Sciences. 2020. Vol. 2. P. 78–91. DOI: 10.1007/s42452-020-03427-1.
12. Saha A., Basu A., Banerjee S. Enhancing thermal management systems: a machine learning and metaheuristic approach for predicting thermophysical properties of nanofluids // Engineering Research Express. 2024. P. 89–97.
13. Wang H.B., Chen X.Y. A Comprehensive Review of Predicting the Thermophysical Properties of Nanofluids Using Machine Learning Methods // Industrial Engineering & Chemistry Research. 2022. Vol. 61, no. 40. P. 101–107. DOI: 10.1021/acs.iecr.2c02059.
14. Maleki A., Haghighi A., Mahariq I. Machine learning-based approaches for modeling thermophysical properties of hybrid nanofluids: A comprehensive review // Journal of Molecular Liquids. 2021. Vol. 322. P. 114843. DOI: 10.1016/j.molliq.2020.114843.

15. Saha A., Basu A., Banerjee S. Enhancing thermal management systems: a machine learning and metaheuristic approach for predicting thermophysical properties of nanofluids // Engineering Research Express. 2024. Vol. 6, no. 4. P. 7–23. DOI: 10.1088/2631-8695/ad8536.
16. Li G., Sun H., Han D. et al. Applying machine learning to reveal the microscopic heat transfer mechanism of nanofluids as coolants // Thermochimica Acta. 2024. P. 1–9. DOI: 10.1016/j.tca.2024.179788.
17. Sharma P., Said Z., Kumar A. et al. Recent Advances in Machine Learning Research for Nanofluid-Based Heat Transfer in Renewable Energy System // Energy & Fuels. 2022. Vol. 36, no. 13. P. 6626–6658. DOI: 10.1021/acs.energyfuels.2c01006.
18. Akilu S., Sharma K.V., Baheta A.T. et al. Machine learning analysis of thermophysical and thermohydraulic properties in ethylene glycol- and glycerol-based SiO2 nanofluids // Scientific Reports. 2024. Vol. 14. Article 14829. P. 21–27. DOI: 10.1038/s41598-024-65411-8.
19. NanoFluid Dynamics. Revolutionizing Thermal Management: Advanced Insights from Nanofluid Experiment. URL: https://www.kaggle.com/datasets/kumaraugustya/nanoparticle-basefluid (дата обращения: 05.01.2025).
20. Лозовский Т., Шимчук Н., Мотовой И., Железный В. Влияние примесей наночастиц Al2O3 на теплоемкость изопропилового спирта // Холодильна техніка та технологія. 2016. № 52. С. 52–79.