JOURNAL OF NANO- AND ELECTRONIC PHYSICS
Vol. 18 No 1, 01028(10pp) (2026)

ЖУРНАЛ НАНО- ТА ЕЛЕКТРОННОЇ ФІЗИКИ
Том 18 № 1, 01028(10cc) (2026)

OPEN ACCESS

REGULAR ARTICLE

Application of Ensemble Machine Learning Algorithms for Modeling the Thermomechanical Properties of Nano-Filled Epoxy Composites

O. Pastukh[1], Yu. Petrov[1], A.V. Buketov[2], K. Dyadyura[3,*], O. Lyashuk[1], O. Totosko[1], V. Sotsenko[2]

[1] *Ternopil Ivan Puluj National Technical University, 46001 Ternopil, Ukraine*
[2] *Kherson State Maritime Academy, 73000 Kherson, Ukraine*
[3] *Odesa State Agrarian University, 65012 Odesa, Ukraine*

(Received 02 December 2025; revised manuscript received 21 February 2026; published online 25 February 2026)

The article is devoted to a comprehensive approach to predicting the heat resistance of plasticized epoxy nanocomposites. The aim of the study was to apply ensemble machine learning algorithms to identify key mechanical factors in predicting the heat resistance of epoxy nanocomposites and to develop a reliable method for their selection for further modeling. To achieve this goal, a two-stage hybrid methodology was applied. At the first stage, ensemble machine learning algorithms, in particular the $f_{\text{regression}}$ method, were used to calculate feature importance, assess their information value, and determine the consistency score between different models. In the second stage, classical statistical analysis, including Pearson's correlation and one-dimensional F-test, was performed to evaluate linear relationships. The results showed high consistency between the methods. It was found that Elastic modulus and Adhesive strength have consistently high significance (Information Value = 0.15 and 0.12, respectively; Consistency Score 82% and 81%, respectively) and are key determinants of heat resistance. Conversely, Destruction stresses and Residual Stresses are completely uninformative (Information Value = 0) and are recommended for exclusion from the models. Filler concentration showed minimal influence in the studied range. Thus, the effectiveness of a hybrid approach combining classical statistics and machine learning for objective and reasonable feature selection in predicting the properties of composite materials has been proven. The proposed algorithm allows the creation of simplified, interpretable, and accurate prediction models for engineering design.

Keywords: Epoxy nanocomposites, Heat resistance, Mechanical properties, Feature selection, Correlation analysis, Machine learning, Ensemble machine learning algorithms, Information value.

DOI: 10.21272/jnep.18(1).01028 PACS numbers: 81.05.t, 81.05.Zx

1. INTRODUCTION

1.1 The Concept of Creating Epoxy Nanocomposites with Predictable Thermomechanical Properties for the Water Transport Industry

Epoxy composites are widely used in the aviation, automotive, electronics, and construction industries. In water transport, they are used for the manufacture of hull structures, ship repairs, and anti-corrosion coatings due to their increased resistance to aggressive environments. The operation of products under dynamic and thermal loads requires these materials to be highly reliable and durable. Therefore, the combination of high mechanical properties and heat resistance is critical for the effective functioning of water transport. Thermomechanical characteristics determine the ability of a material to maintain structural integrity when exposed to thermal fields and mechanical loads. Epoxy composites are characterized by low impact strength and brittleness, which limits their areas of application. Increasing heat resistance allows for an expansion of the temperature range in which composites can be used.

One promising way to improve the performance characteristics of epoxy composite materials (CM) is to introduce nanoscale fillers into the binder [1-12]. Nanofillers, such as carbon nanotubes, fullerenes, and nano-soot, are characterized by high specific surface area and strength. Their introduction into the oligomeric epoxy binder improves mechanical properties due to the redistribution of stresses during the structure formation of materials. At the same time, they affect heat and thermal stability by forming barriers to heat flow [9]. It should be noted that the effectiveness of such nanomodification significantly depends on the filler content and dispersion conditions. Particle agglomeration at high content can lead to a deterioration in the thermomechanical properties of epoxy CMs. Therefore, studying the effect of nanoparticle content, including carbon nanotubes, on the thermomechanical complex of

* Correspondence e-mail: dyadyura.k.o@op.edu.ua

Cite this article as: O. Pastukh et al., *J. Nano- Electron. Phys.* 18 No 1, 01028 (2026) https://doi.org/10.21272/jnep.18(1).01028

properties of epoxy composites is a relevant task today. Optimizing their content allows achieving a synergistic effect in increasing strength, stiffness, and heat resistance. Particular attention in research is paid to the mechanisms of interphase interaction at the "matrix-nanofiller" interface [1, 8]. This interaction determines the degree of gel formation in CM and the density of the polymer structural network after cross-linking of materials. Experimental data indicate the presence of a critical concentration of filler (nanotubes) in the polymer matrix [1]. The introduction of nanoparticles into the polymer beyond the critical content leads to a significant deterioration in the properties of CM [8]. Thus, rational selection of the composition allows the creation of materials with predictable characteristics. The paper presents statistical data on the influence of carbon nanotubes on the thermomechanical properties of epoxy composites in order to determine the optimal modification conditions for use in water transport.

1.2 Integration of Ensemble Algorithms into the Process of Designing Nano-Filled Epoxy Composites and Predicting their Properties

Modern material design methods increasingly integrate artificial intelligence tools to accelerate research. Predicting the thermomechanical properties of nanocomposites is a challenging task due to the nonlinearity of component effects. Traditional empirical approaches require large amounts of experimental data. Ensemble machine learning algorithms offer an effective tool for analyzing complex relationships between material properties. Methods such as Random Forest, Decision Trees, and Gradient Boosting combine the predictions of multiple simple models [13, 14]. This allows for increased robustness and accuracy of prediction on relatively small data sets. For epoxy composites, important predicted parameters are strength, modulus of elasticity, and heat resistance. The physical nature, dispersion, and filler content are commonly used as input variables for models. The use of machine learning algorithms allows us to identify hidden patterns between the composition and properties of CM. In particular, it is possible to quantitatively assess the influence of nanoparticle content on specific CM properties. A significant advantage is the ability of models to interpolate within the training sample. This opens up avenues for virtual screening of composite compositions without costly experiments. For epoxy systems modified with nanotubes, it is important to predict a specific property while minimizing experiments. Ensemble methods can determine the critical concentration at which the strengthening mechanisms change. Modeling also allows us to evaluate the relative importance of properties, the relationship between them, and their impact on the predicted property. Integrating such models into the material design cycle significantly reduces development time. In the context of water transport, this allows for the rapid creation of composites for specific operating conditions. The introduction of machine learning marks a new stage in the creation of materials with predictable properties. Thus, ensemble algorithms open up new opportunities not only for the targeted creation of highly efficient nanocomposites, but also for predicting their properties.

2. MATERIALS AND METHODS

2.1 Research Materials

During the formation of polymer compositions, DER 331 epoxy resin (ISO 9001) was selected as the base. In order to improve the rheological properties of the epoxy binder, aliphatic resin DEG-1 (TU 2225-390-04872688-98) was added to the DER 331 resin as a plasticizer. MWCNT nanotubes (Guangzhou Hongwu Material Technology Co., Ltd., China) were used as a filler. The work used multilayer nanotubes with the following parameters: diameter – 10 … 30 nm, length – 5 … 20 μm.

The following properties of CM were studied: destructive stress and modulus of elasticity in bending, impact toughness, adhesive strength, residual stress, and heat resistance.

The destructive stress and modulus of elasticity in bending were determined in accordance with ASTM D 790–03. The impact strength of KM was tested using a pendulum impact tester (ISO 179-1). The adhesive strength at tear was tested in accordance with ISO 4624:2016. The heat resistance of composites was analyzed according to ISO 75-2.

In addition, residual stresses in coatings after their polymerization were analyzed. The method consists in measuring the stresses that arise in the polymer coating during its compaction (bonding) on a metal substrate. The essence of the method: the metal substrate was fixed as a console. A thin layer of adhesive was applied to the substrate. During polymerization, the adhesive shrunk. This resulted in stresses that bent the metal substrate. After determining the amount of deflection of the free end of the substrate and taking into account the elastic modulus of the base, the geometric parameters of the substrate and the adhesive, the residual stresses in the polymer were calculated using the method [5, 6]. Thus, using a simple method and an uncomplicated mechanical model, the residual stresses were quantitatively evaluated without complex equipment.

3. DISCUSSION OF RESEARCH RESULTS

3.1 Experimental Results

The results of the experiments yielded the properties of epoxy composite materials (CM) shown in Table 1. CM based on DER-331 resin (100%), DEG-1 plasticizer (12%), PEP hardener (12%), and nanotube filler with different powder contents (0.05 and 0.1%) was used.

Based on experimental data (Table 1), the paper considers the problem of predicting the values of the variable (property indicators) Heat resistance depending on the values of the following CM properties: Filler, Destruction stresses, Elastic modulus (Elasticity), Impact strength (Viscosity), Adhesive strength (Solidity), Residual stresses. Note that in Table 1, only data obtained for CM without powder (Filler – 0%) and all CM property indicators with filler, except for heat resistance, are used for machine learning. The variable Heat resistance is a prediction task using ensemble machine learning (ML) algorithms.

Table 1 – Properties of CM based on DER-331 resin (100%), DEG-1 plasticizer (12%), PEA hardener (12%), and nanotube filler (powder content, %)

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, K
0	51.6	2.2	9.6	29.9	1.42	356
	60.9	3.2	6.9	36.1	1.66	348
	50.4	1.8	6.7	30.5	0.95	356
	85,2	2.5	13.8	38.6	1.86	344
	61.5	3.4	14.7	29.8	0.48	351
0.025	78.2	2.2	12.4	39.1	1.15	358
	64.7	3.4	15.9	36.3	0.88	348
	65.6	2.8	19.0	45.8	1.34	356
	67.3	2.4	14.8	44.0	1.18	359
	76.4	3.3	13.1	36.5	1.11	353
0.050	97.2	2.6	16.4	61.2	1.22	353
	68.0	2.8	14.2	54.6	0.99	358
	67.6	3.4	10.7	43.9	0.81	353
	75.8	2.9	12.8	43.5	1.22	360
	78.0	3.2	13.9	47.1	0.48	354
0.075	81.7	3.0	10.9	45.4	0.66	355
	73.3	3.1	8.3	44.6	1.24	356
	94.3	3.3	10.6	49.4	0.88	362
	89.2	3.2	14.4	32.7	0.46	347
	82.3	3.7	18.1	49.4	0.82	352
0.100	84.0	3.2	8.8	39.4	0.48	356
	73.3	2.7	12.5	44.8	0.83	355
	66.4	3.5	10.1	38.7	1.03	352
	78.1	3.4	12.4	48.1	0.81	351
	60.3	3.6	12.2	42.5	0.68	350
0.125	53.2	3.1	11.6	43.4	0.66	349
	70.3	3.4	12.3	35.5	0.97	349
	83.5	3.3	9.4	31.6	0.48	356
	70.9	3.1	8.6	44.5	0.66	356
	62.1	3.6	12.1	46.5	0.84	354

To solve the forecasting problem, ensemble machine learning algorithms were used to establish the dependence of the Heat resistance variable on the values of the following variables: Filler, Destruction stresses, Elastic modulus, Impact strength, Adhesive strength, Residual stresses. Ensemble machine learning algorithms form models of the above dependence of CM properties.

When predicting the values of the Heat resistance variable, the dependence was formed in the form of the following models:

$$\mathrm{T} = m_i\,(n_0,\ n_1,\ n_2,\ n_3,\ n_4,\ n_5),\ i\ =\ 0 \ldots 5, \qquad (1)$$

where T – variable Heat resistance; $m_0 \ldots m_6$ – ensemble machine learning algorithms accordingly: Ada Boost (ada_boost), Bagging (bagging), Decision Trees (decision_trees), Gradient Boost (gradient_boost), Random Forest (random_forest), Stacking (stacking), Voting (voting); $n_0 \ldots n_5$ – mechanical properties of CM in the form of variables (in models): n_0 – variable Filler; n_1 – variable Destruction stresses; n_2 – variable Elastic modulus (Elasticity); n_3 – variable Impact strength (Viscosity); n_4 – variable Adhesive strength (Solidity); n_5 – variable Residual stresses.

The software implementation of all models mi, $i = 0 \ldots 6$ (ensemble machine learning algorithms) was performed using the scikit-learn library version 1.7.2 [15].

A dataset N was formed for training and testing models. This dataset was obtained by combining the data rows in Table 1. Training $N_{training}$ and test $N_{testing}$ data samples were formed from the obtained dataset N. The training sample $N_{training}$ included rows of the dataset N that were in Table 1 only with the values of the variable Filler = 0%, Filler = 0.025%, Filler = 0.075%, Filler = 0.1%, Filler = 0.125%. The test sample $N_{testing}$ included rows of the dataset N, which were in Table 1 only with the values of the variable Filler = 0.05% and Filler = 0.1%.

Thus, initially, all models mi, i = 0 … 6 were trained on the $N_{training}$ data sample. The next step was to test them on the $N_{testing}$ data sample. Note that by testing the models on the $N_{testing}$ sample, we solved the problem of predicting the values of the Heat resistance variable depending on the values of the variables in the form of mechanical properties of CM using models m_i, $i\ =\ 0 \ldots 6$.

Tables 2 – 8 show the results of predicting the values of the variable Heat resistance depending on the values of variables $n_0 \ldots n_5$ using models mi, $i\ =\ 0 \ldots 6$ on the test sample $N_{testing}$.

Table 2 – Results of forecasting the values of the Heat resistance variable using the m_0 (Ada Boost) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, K	Heat resistance predicted, K
0.05	97.2	2.6	16.4	61.2	1.22	353	358.0
0.05	68.0	2.8	14.2	54.6	0.99	358	356.0
0.05	67.6	3.4	10.7	43.9	0.81	353	352.8
0.05	75.8	2.9	12.8	43.5	1.22	360	357.2
0.05	78.0	3.2	13.9	47.1	0.48	354	353.66
0.1	84.0	3.2	8.8	39.4	0.48	356	352.8
0.1	73.3	2.7	12.5	44.8	0.83	355	356.0
0.1	66.4	3.5	10.1	38.7	1.03	352	351.0
0.1	78.1	3.4	12.4	48.1	0.81	351	352.8
0.1	60.3	3.6	12.2	42.5	0.68	350	349.0

Table 3 – Results of forecasting the values of the Heat resistance variable using the m_1 (Bagging) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	361.2
0.05	68.0	2.8	14.2	54.6	0.99	358	360.2
0.05	67.6	3.4	10.7	43.9	0.81	353	350.4
0.05	75.8	2.9	12.8	43.5	1.22	360	356.8
0.05	78.0	3.2	13.9	47.1	0.48	354	355.4
0.1	84.0	3.2	8.8	39.4	0.48	356	350.4
0.1	73.3	2.7	12.5	44.8	0.83	355	354.8
0.1	66.4	3.5	10.1	38.7	1.03	352	350.4
0.1	78.1	3.4	12.4	48.1	0.81	351	353.8
0.1	60.3	3.6	12.2	42.5	0.68	350	350.4

Table 4 – Results of forecasting the values of the Heat resistance variable using the m_2 (Decision Trees) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	356
0.05	68.0	2.8	14.2	54.6	0.99	358	362
0.05	67.6	3.4	10.7	43.9	0.81	353	349
0.05	75.8	2.9	12.8	43.5	1.22	360	359
0.05	78.0	3.2	13.9	47.1	0.48	354	355
0.1	84.0	3.2	8.8	39.4	0.48	356	349
0.1	73.3	2.7	12.5	44.8	0.83	355	356
0.1	66.4	3.5	10.1	38.7	1.03	352	353
0.1	78.1	3.4	12.4	48.1	0.81	351	354
0.1	60.3	3.6	12.2	42.5	0.68	350	349

Table 5 – Results of forecasting the values of the Heat resistance variable using the m_3 (Gradient Boost) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	355.59
0.05	68.0	2.8	14.2	54.6	0.99	358	354.31
0.05	67.6	3.4	10.7	43.9	0.81	353	354.06
0.05	75.8	2.9	12.8	43.5	1.22	360	353.39
0.05	78.0	3.2	13.9	47.1	0.48	354	353.53
0.1	84.0	3.2	8.8	39.4	0.48	356	352.93
0.1	73.3	2.7	12.5	44.8	0.83	355	353.53
0.1	66.4	3.5	10.1	38.7	1.03	352	352.21
0.1	78.1	3.4	12.4	48.1	0.81	351	353.25
0.1	60.3	3.6	12.2	42.5	0.68	350	352.53

Table 6 – Results of forecasting the values of the Heat resistance variable using the m_4 model (Random Forest)

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	360.2
0.05	68.0	2.8	14.2	54.6	0.99	358	360.0
0.05	67.6	3.4	10.7	43.9	0.81	353	350.0
0.05	75.8	2.9	12.8	43.5	1.22	360	356.6
0.05	78.0	3.2	13.9	47.1	0.48	354	355.2
0.1	84.0	3.2	8.8	39.4	0.48	356	350.4
0.1	73.3	2.7	12.5	44.8	0.83	355	354.8
0.1	66.4	3.5	10.1	38.7	1.03	352	350.4
0.1	78.1	3.4	12.4	48.1	0.81	351	353.8
0.1	60.3	3.6	12.2	42.5	0.68	350	350.4

Table 7 – Results of forecasting the values of the Heat resistance variable using the m_5 (Stacking) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	352.12
0.05	68.0	2.8	14.2	54.6	0.99	358	353.12
0.05	67.6	3.4	10.7	43.9	0.81	353	352.45
0.05	75.8	2.9	12.8	43.5	1.22	360	353.84
0.05	78.0	3.2	13.9	47.1	0.48	354	352.38
0.1	84.0	3.2	8.8	39.4	0.48	356	351.77
0.1	73.3	2.7	12.5	44.8	0.83	355	353.90
0.1	66.4	3.5	10.1	38.7	1.03	352	352.36
0.1	78.1	3.4	12.4	48.1	0.81	351	353.34
0.1	60.3	3.6	12.2	42.5	0.68	350	352.15

Table 8 – Results of forecasting the values of the Heat resistance variable using the m_6 (Voting) model

Filler, %	Destruction stresses, MPa	Elastic modulus, GPa	Impact strength, kJ/m²	Adhesive strength, MPa	Residual stresses, MPa	Heat resistance, К	Heat resistance predicted, К
0.05	97.2	2.6	16.4	61.2	1.22	353	355.94
0.05	68.0	2.8	14.2	54.6	0.99	358	355.34
0.05	67.6	3.4	10.7	43.9	0.81	353	353.71
0.05	75.8	2.9	12.8	43.5	1.22	360	353.79
0.05	78.0	3.2	13.9	47.1	0.48	354	354.19
0.1	84.0	3.2	8.8	39.4	0.48	356	352.90
0.1	73.3	2.7	12.5	44.8	0.83	355	354.97
0.1	66.4	3.5	10.1	38.7	1.03	352	350.92
0.1	78.1	3.4	12.4	48.1	0.81	351	354.03
0.1	60.3	3.6	12.2	42.5	0.68	350	351.88

Next, we evaluated the relative error of predicting the values of the Heat resistance variable using models m_i, $i = 0 \dots 6$. On the test sample $N_{testing}$, we calculated the Mean Absolute Percentage Error (MAPE) using the formula:

$$\delta = \frac{1}{P}\sum_{n=0}^{P-1}\left(\frac{T_n - T_{p_n}}{T_n}\right)\cdot 100\%, \qquad (2)$$

where P – test sample size $N_{testing}$; T_n – actual heat resistance values; T_{pn} – predicted heat resistance indicators using ensemble machine learning algorithms.

Fig. 1 shows the average relative error values for each ML algorithm.

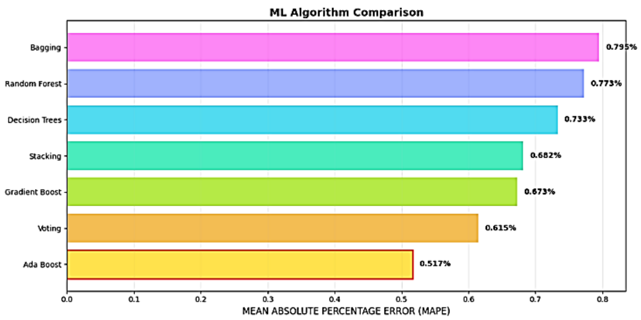

Fig. 1 – Error in predicting heat resistance (MAPE) using ensemble machine learning algorithms

3.2 Assessment of the impact of independent variables (mechanical properties) on the predicted variable (heat resistance) based on models

To determine the key factors influencing the variable Heat resistance, feature importance scores of trained models were calculated. Note that feature importance is an indirect measure of the influence of one property on another (such as correlation or regression coefficient). It is a measure of how much a particular feature improves (or worsens) the accuracy of predicting the target variable–Heat Resistance–within the selected model.

Seven machine learning algorithms $m_0 \dots m_6$ were used to assess the impact of input parameters on heat resistance. Each algorithm provided its own assessment of feature importance (feature_importances score), which determines the contribution of each variable (Filler, Destruction stresses, Elastic modulus, Impact strength, Adhesive strength, Residual stresses, Material type) to the accuracy of the model prediction. The summary results are shown in Table 9.

Numerical estimates of the importance of variables based on Ada Boost, Decision Trees, Gradient Boost, and Random Forest were determined as the normalized sum of the decrease in the splitting criterion [16], and based on Bagging, Stacking, Voting, they were determined using the model-independent Permutation Importance method [17] built into the scikit-learn v1.7.2 library.

Table 9 – Feature importance score

ML algorithms	Filler	Destruction stresses	Elastic modulus (Elasticity)	Impact strength (Viscosity)	Adhesive strength (Solidity)	Residual Stresses
bagging	0.011	0.2694	0.2249	0.069	0.2954	0.1304
decision_trees	0.0074	0.2732	0.2269	0.0595	0.296	0.1371
random_forest	0.0194	0.2643	0.2285	0.0643	0.2915	0.1321
gradient_boost	0.002	0.0473	0.3093	0.0848	0.1787	0.3777
stacking	0.0109	0.1728	0.183	0.1698	0.3354	0.1281
ada_boost	0.0099	0.1193	0.1947	0.103	0.3295	0.2436
voting	0.0082	0.0599	0.2242	0.1825	0.2768	0.2485

To interpret the data in Table 9 and for the purpose of comparative analysis, histograms were constructed showing the importance of the influence of variables on heat resistance indicators. The results are shown in Fig. 2.

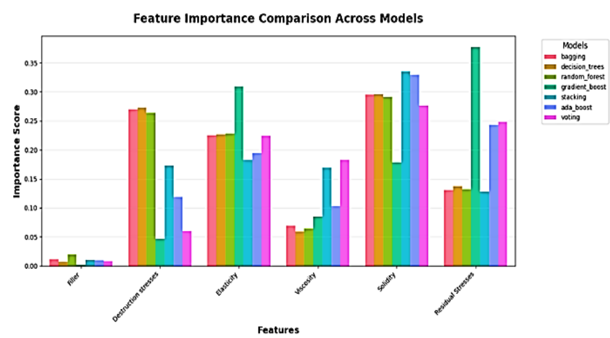

Fig. 2 – Importance of variables (features) for selected ML algorithms (models)

Table 9 shows that the importance values of the variables Filler, Destruction stresses, Elastic modulus (Elasticity), Impact strength (Viscosity), Adhesive strength (Solidity), and Residual stresses vary within different ML algorithms. Therefore, their average values and standard deviations were calculated and are shown in Fig. 3 and Fig. 4, respectively.

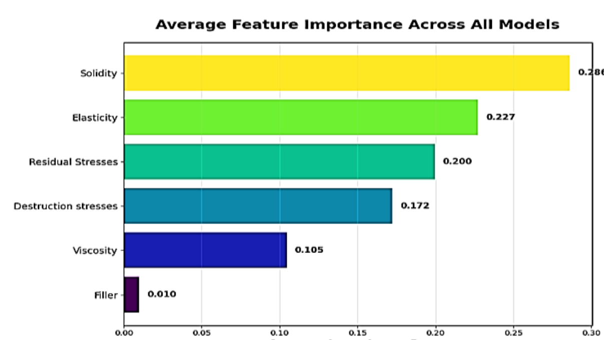

Fig. 3 – Average values of the importance of variables (features) within ML algorithms (models)

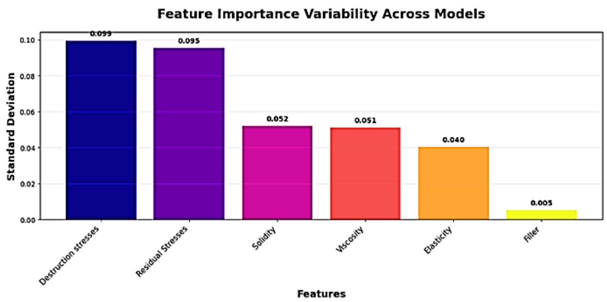

Fig. 4 – Mean square deviations of the importance values of variables (features) within ML algorithms (models)

Taking into account the average values (Fig. 3) and standard deviations (Fig. 4) of the importance of the influence of variables within ML algorithms (models), the consistency score of the results between ML algorithms (models) was calculated using the following formula:

$$cs(n_i) = 100 - \left[\frac{\sigma_{n_i}}{m_{n_i}} \cdot 100\right] \tag{3}$$

where cs(n_i) – this is a consistency score of ML algorithms (models) regarding the importance of variable n_i; n_i – variables in the formula (1); σ_{n_i}– mean square deviation of estimates of the importance of variable ni within ML algorithms (models); m_{n_i}– average value of estimates of the importance of variable ni within ML algorithms (models). The numerical values of the consistency assessment are given in Table 10.

Table 10 – Consistency Score

Feature	Score (%)
Elastic modulus (Elasticity)	82
Adhesive strength (Solidity)	81
Residual Stresses	52
Impact strength (Viscosity)	51
Filler	46
Destruction stresses	42

The models demonstrate moderate overall consistency. The average Consistency Score is ≈ 59%, indicating the presence of both consensus and controversial features among the model family. High consistency was found for the properties: Elastic modulus (Elasticity) (82%) and Adhesive strength (Solidity) (81%). For these two mechanical characteristics, there is almost complete consensus among the entire family of models. They are unanimously recognized as consistently important for the forecast. These are the most reliable factors in terms of model consistency. Their influence is confirmed regardless of the chosen ensemble machine learning algorithm, which makes conclusions about them statistically robust.

Moderate consistency was found for the properties: Residual Stresses (52%) and Impact strength (Viscosity) (51%). Consistency regarding these features is unstable. The models were divided roughly in half in their assessment of their importance. Their influence is ambiguous and depends on the architecture of the specific model. Low agreement was found for the properties: Filler (46%) and Destruction stresses (42%). These features cause the greatest disagreement within the model family. Most models did not agree on their role. The particularly low agreement for the Destruction stresses property (42%) may indicate that this feature interacts with other factors in a nonlinear manner.

Thus, it can be stated that the analysis of the consistency score between models (Table 10) showed that the highest consensus was reached regarding the importance of the Elastic modulus (82%) and Adhesive strength (81%) features. Conversely, the lowest consistency was found for the Filler (46%) and Destruction stresses (42%) features, indicating significant differences in their interpretation by different algorithms. The features Residual Stresses (52%) and Impact strength (51%) occupy an intermediate position. These results indicate that the influence of elastic modulus and adhesive strength is the most stable and independent of the choice of model, while the influence of filler concentration, destructive stresses, and, partially, impact strength requires further study using more complex models capable of accounting for their nonlinear nature or interaction with other factors.

3.3 Statistical Tests to Determine the Significance of the Influence of Characteristics

Additionally, the importance of the influence of variable values in the form of mechanical properties on the value of the Heat resistance variable was assessed based on statistical tests. In particular, Pearson's correlation statistical test and Fisher's one-factor F-test were used.

Pearson's correlation statistical test (Pearson's correlation coefficient, r) is a parametric method that estimates the magnitude and direction of the linear statistical relationship between two continuous variables. It is calculated as the ratio of the covariance of two variables to the product of their standard deviations. It takes values in the range from -1 to +1. A value close to +1 indicates a strong direct linear relationship (an increase in one variable is accompanied by an increase in the other); a value close to -1 indicates a strong inverse relationship; a value close to 0 indicates no linear relationship between the input and output variables.

The p-value obtained as a result of the test indicates the probability of obtaining the coefficient r, assuming that the null hypothesis is true; a low p-value (usually <0.05) allows the null hypothesis to be rejected.

Pearson's correlation coefficient for the sample (r) was calculated using a formula that takes into account covariance and standard deviations:

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

where: r_{xy} – Pearson's correlation coefficient between variables x and y; n – number of observation pairs; x_i, y_i

– individual values of the i-th observation pair; $\bar{x}, \bar{y}$ – the arithmetic mean values of variables x and y, respectively; $cov\,(x, y)$ – covariance between x and y; σ_x, σ_y – standard deviations (root mean square deviations) for x and y.

In our case (based on model (1)), Pearson's correlation formula for T and n_j is as follows:

$$r_{T,n_j} = \frac{\sum_{i=0}^{6}\left(n_j^{(i)} - \bar{n}_j\right)\cdot\left(T^{(i)} - \bar{T}\right)}{\sqrt{\sum_{i=0}^{6}\left(n_j^{(i)} - \bar{n}_j\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(T^{i} - \bar{T}\right)^2}} \quad (5)$$

where j – argument index (0, 1, 2, 3, 4, 5); i – observation index (0, 1, ..., 6); $n_j^{(i)}$– value of argument nj in the i-th observation; $T^{(i)} = f_i(n_0^{(i)}, n_1^{(i)}, \ldots, n_5^{(i)})$ – value of function T in the i-th observation; $\bar{n}_j = \frac{1}{7}\sum_{i=0}^{6} n_j^{(i)}$ – the arithmetic mean of the values of argument nj across all 7 observations; $\bar{T} = \frac{1}{7}\sum_{i=0}^{6} T^{(i)}$ – arithmetic mean of function values T.

The statistical test for Pearson's correlation was calculated using the '$r_{\text{regression}}$' function from the scikit-learn v1.7.2 library [18]. The results are shown in Table 11.

Table 11 – Results of the statistical test for Pearson's correlation

Feature	Correlation coefficient	p-value (α = 0.05)
Filler	0.021	0.90
Destruction stresses	0.097	0.60
Elastic modulus (Elasticity)	– 0.319	0.08
Impact strength (Viscosity)	– 0.155	0.41
Adhesive strength (Solidity)	0.284	0.12
Residual Stresses	– 0.069	0.71

Based on the results of Pearson's correlation, the following conclusion can be made: none of the independent variables studied has a statistically significant linear relationship with the target variable (Heat resistance) at the standard significance level of $\alpha = 0.05$, since all p-values > 0.05.

Filler feature. The correlation coefficient $r = 0.021$ is practically zero ($p = 0.90$). This indicates the absence of any linear relationship. A change in the concentration of the additive in the studied range correlates nonlinearly with a change in the target property.

Destruction stresses feature. The weak positive correlation ($r = 0.097$) is statistically insignificant ($p = 0.60$). There is no reason to claim a linear relationship between tensile strength and the target variable.

Elastic modulus feature. The highest coefficient among all ($r = -0.319$), indicating a moderate inverse trend. However, p-value $= 0.08\,(> 0.05)$, which indicates the statistical insignificance of this relationship. It can be considered marginally significant or indicate a potential trend that may become significant as the sample size increases.

Impact strength. The weak inverse correlation ($r = -0.155$) is not statistically significant ($p = 0.41$). The linear relationship between impact strength and the target property is not confirmed.

Adhesive strength feature. Moderate direct trend ($r = 0.284, p = 0.12$). Similar to the elastic modulus feature, the relationship is not statistically significant at the 0.05 level, but it shows the highest strength among direct correlations, which may be important for formulating hypotheses.

Residual stresses feature. A very weak inverse relationship ($r = -0.069$) is statistically insignificant ($p = 0.71$). Residual stresses are practically not linearly related to the target variable.

The calculated correlation coefficients indicate that the target property Heat resistance within the scope of this experiment does not have a pronounced linear dependence on any of the individual mechanical characteristics or filler concentration. This may be due to several factors:

1. Nonlinearity of the relationship: the actual dependence may be complex and nonlinear (e.g., exponential, etc.), which is not revealed by Pearson's linear correlation.

2. Combined influence of variables: a property may depend not on a single factor, but on their synergistic interaction, which requires the use of multiple regression or machine learning methods. In this context, a possible solution to the problem is to use multiple correlation analysis or regression models to assess the cumulative effect of all variables.

Thus, the results of Pearson's correlation analysis do not confirm simple linear dependencies, which indicates the need to use more complex analysis models to predict the target property.

We also used Fisher's one-way F-test (one-way analysis of variance, one-way ANOVA) to determine the importance of the influence of characteristics on the target variable. Its null hypothesis (H_0) states that all group mean values of the features are collectively equal, i.e., the factor under study has no statistically significant effect on the result. The alternative hypothesis (H_1) is that at least one pair of groups mean values is statistically different.

Fisher's one-way F-test (One-Way ANOVA) was calculated using the formula derived from (4):

$$F_i = \frac{r_i^2}{1 - r_i^2} \cdot (K - 2) \quad (5)$$

where K is number of observations; r_i are Pearson's correlation coefficient between features n_i and the target variable T.

Fisher's one-factor F-test was calculated using the '$f_{\text{regression}}$' function from the scikit-learn v1.7.2 library [19]. The results are shown in Table 12.

No statistically significant linear relationship was found between any of the factors studied and the target variable Heat resistance at the standard significance level $\alpha = 0.05$, since all p-values $>$ 0.05. This means that with this sample size and data dispersion, the null hypothesis (about the absence of a linear relationship) cannot be rejected.

Table 12 – Results of Fisher's one-way F-test

Feature	F score	p-value ($\alpha = 0.05$)
Filler	0.01	0.90
Destruction stresses	0.26	0.60
Elastic modulus (Elasticity)	3.18	0.08
Impact strength (Viscosity)	0.69	0.41
Adhesive strength (Solidity)	0.13	0.12
Residual Stresses	0.02	0.71

Filler feature: $F = 0.01, p = 0.90$. The F-statistic is minimal. There is practically no linear relationship between concentration and the target property. The change in concentration in the studied range is not described by a linear model.

Destruction stresses feature: $F = 0.26, p = 0.60$. A rather weak correlation, statistically insignificant. Destructive stresses during bending are not a linear predictor for the target variable.

Elastic modulus feature: $F = 3.18, p = 0.08$. This is the strongest linear trend found among all factors. Although the p-value (0.08) is formally > 0.05, the indicator is on the verge of significance. This may indicate:

– a potential tendency toward an inverse linear relationship (since the F-test for correlation does not show the direction, but based on the preliminary correlation analysis, we assume that $r = -0.319$);

– an effect that may become statistically significant as the sample size increases (type II error).

Impact strength: $F = 0.69, p = 0.41$. Weak and insignificant correlation. No linear relationship between impact strength and target property has been confirmed.

Adhesive strength: $F = 0.13, p = 0.12$. A weak trend, statistically insignificant at the 0.05 level, but with a p-value lower than most other factors (except for the modulus of elasticity). This may again indicate a potentially weak linear relationship. However, to increase statistical power, the sample size needs to be increased.

Residual Stresses feature: $F = 0.02, p = 0.71$. No linear relationship. Residual stresses are not a linear predictor.

A one-factor F-test performed using the $f_{\text{regression}}$ method to assess the significance of linear relationships showed that none of the independent variables studied (filler concentration, destructive stresses, modulus of elasticity, impact strength, adhesive strength, residual stress) had a statistically significant ($p < 0.05$) linear effect on the target variable. The strongest, but still statistically insignificant (at the 0.05 level), trend was found for the modulus of elasticity ($F = 3.18, p = 0.08$). This indicates the absence of simple linear dependencies and points to the need to use more complex nonlinear or multiple models to analyze the impact of the factors under study.

3.4 Determining the Importance of Factors Based on Information Measure

To assess the importance of the influence of variable values in the form of mechanical properties on the value of the target variable Heat resistance, an information measure based on Shannon's entropy, namely Mutual Information, was also used. Numerical values for the Mutual_Information information measure were calculated using the formula:

$$I(n_i, T) = H(n_i) + H(T) - H(n_i, T), \; i = 0\ldots5, \qquad (6)$$

where $I(n_i, T)$ – mutual information between n_i and T; $H(n_i)$ – Shannon entropy of a variable n_i; $H(T)$ – Shannon entropy of a variable T; $H(n_i, T)$ – Shannon's joint entropy.

Numerical values for Mutual Information were calculated using the 'mutual_info_regression' function from the scikit-learn v1.7.2 library [20]. Higher numerical values of Mutual Information for variable n_i indicate that this variable has a greater influence on T. Mutual Information can reveal nonlinear dependencies between n_i and T. The results are presented in Table 13.

Table 13 – Mutual Information

Feature	Info value	Conclusion on importance
Elastic modulus (Elasticity)	0.15	Key variable (highest informational value)
Adhesive strength (Solidity)	0.12	Important additional variable (significant informational value)
Impact strength (Viscosity)	0.10	Low importance (May be included, but its contribution is limited)
Filler	0.03	Low informativeness (for ensemble methods that can detect weak nonlinear dependencies, it can be left for verification, but no significant contribution should be expected from it)
Destruction stresses	0	Non-informative (absolutely independent of the variable Heat resistance. Recommended to remove)
Residual Stresses	0	Non-informative (absolutely independent of the variable Heat resistance. Recommended to remove)

Based on the data obtained, the following can be stated.

Features with moderate informativeness ($Info\ value \geq 0.1$):

- Elastic modulus ($Info\ value = 0.15$) – the highest indicator among all features. The elastic modulus shows the strongest statistical correlation with heat resistance among all variables studied;

– Adhesive strength ($Info\ value = 0.12$) – the second most significant factor;

– Impact strength ($Info\ value = 0.10$) – is on the borderline of the weak informativeness category, but formally falls into the moderate category.

Feature with weak informativeness ($0.02 \leq$

$Info\ value < 0.1$):

– Filler ($Info\ value = 0.03$) – the concentration of filler has a very weak connection with the target variable.

Non-informative features ($Info\ value = 0$):

– Destruction stresses ($Info\ value = 0$);

– Residual Stresses ($Info\ value = 0$).

These two mechanical characteristics show absolute statistical independence from the Heat resistance variable within this sample. Their distribution does not contain information for predicting heat resistance. The recommendation to remove them from further modeling is entirely justified.

The analysis of the informativeness of features ($Information\ Value$) showed that the variables Elastic modulus and Adhesive strength are moderately significant for predicting heat resistance and are key for modeling. The variable Impact strength shows a weak to moderate correlation. The concentration of filler (Filler) proved to be practically uninformative. The variables Destruction stresses and Residual Stresses have an Info value = 0, which indicates their absolute statistical independence from the target parameter. It is recommended to exclude them from further analysis to simplify the model and improve its interpretability. These results are consistent with the conclusions of the previous stages of feature selection. This analysis provides a clear scientific basis for forming the final set of variables for the prediction model.

4. CONCLUSIONS

1. The work presents results on predicting the heat resistance of epoxy nanocomposites based on their mechanical properties. Ensemble machine learning algorithms were used for this purpose. It has been proven that all the models used for prediction are highly effective, since their average relative error is insignificant: 0.517% for Ada Boost, 0.615% for Voting, 0.673% for Gradient Boost, 0.682% for Stacking, 0.733% for Decision Trees, 0.773% for Random Forest, and 0.795% for Bagging. These results should be integrated into practical applications when developing epoxy nanocomposites for various functional purposes.

2. Key mechanical factors affecting the heat resistance of epoxy nanocomposites have been identified. Based on a comprehensive statistical analysis and assessment of the informativeness of variables, it has been established that the modulus of elasticity and adhesive strength are of the highest significance for predicting heat resistance. These variables are characterized by moderate informativeness (Infor Value is 0.15 and 0.12, respectively) and high consistency between different machine learning models (Consistency Score is 82% and 81%). This indicates their stability and statistical significance in the formation of thermomechanical properties of materials.

3. A slight linear effect of nanofiller concentration on heat resistance in the studied range was found. The filler concentration showed very low informativeness ($Infor\ Value = 0.03$) and insignificant correlation with heat resistance. However, low agreement between models (46%) may indicate possible nonlinear effects or interactions with other parameters.

4. It was found that destructive stresses and residual stresses are statistically uninformative for predicting heat resistance. Both variables have zero Information Value and low agreement between models (42% and 52%). This allows us to recommend excluding them from further prediction models to simplify and improve the efficiency of machine learning algorithms, as they do not carry a useful signal for predicting the target variable.

5. Ensemble machine learning algorithms demonstrate high consensus in identifying important features, confirming the reliability of the results obtained. Consistency Score analysis showed that different models agree on the importance of the modulus of elasticity and adhesive strength, while for other variables the consensus was lower. This confirms the objectivity of the identified patterns and shows that the choice of a specific algorithm is not critical for identifying key factors.

6. The complementary use of classical statistics (Pearson's correlation, F-test), machine learning (ensemble machine learning algorithms), and an information-based approach ensures comprehensive and objective validation of the feature selection results. The high consistency of conclusions obtained by different independent methods (for example, consensus on the importance of the modulus of elasticity and the insignificance of destructive stresses) significantly increases the reliability of scientific results and provides a clear basis for the selection of variables at the final stage of composite development.

7. The results of the study form a clear algorithm of actions for creating effective models for predicting thermomechanical properties: at the first stage, using Information Value or f_regression, it is necessary to filter out uninformative features (Residual Stresses); at the second stage, focus should be placed on detailed modeling of relationships with key factors (Elastic modulus, Adhesive strength); at the third stage, possible nonlinearity and combined effects of insignificant variables (Impact strength, Filler) should be taken into account using ensemble machine learning algorithms.

REFERENCES

1. M.V. Kindrachuk, Yu.Ya. Dushek, M.V. Luchka, *Powder Metallurgy and Metal Ceramics* **9-10**, 56 (1994).
2. M.V. Kindrachuk, Yu.Ya. Dushek, M.V. Luchka, A.N. Gladchenko, *Powder Metallurgy and Metal Ceramics* **34** No 5-6, 321 (1995).
3. I.G. Dobrotvor, P.D. Stukhlyak, A.G. Mykytyshyn, D.P. Stukhlyak, *Coatings. Strength of Materials*. **53** No 2, 283 (2021).
4. É.A. Lysenkov, *Journal of Engineering Physics and Thermophysics* **88** No 4, 1008 (2015).
5. A. Mostovoy, A. Yakovlev, V. Tseluikin, M. Lopukhova, *Polymers* **12** No 8, 1816 (2020).
6. A.V. Buketov, M.V. Brailo, S.V. Yakushchenko, O.O. Sapronov, S.O. Smetankin, *Journal of Marine Engineering and Technology* **19** No 3, 109 (2020).
7. O.O. Sapronov, A.V. Buketov, P.O. Maruschak, S.V. Panin, M.V. Brailo, S.V. Yakushchenko, et al., *Functional Materials* **26** No 1, 114 (2019).
8. O.V. Totosko, P.D. Stukhlyak, A.H. Mykytyshyn, V.V. Levytskyi, *Functional Materials* **27** No 4, 760 (2020).
9. A.V. Buketov, A.V. Sapronova, O.O. Sapronov, N.M. Buketova, et al., *Composites: Mechanics, Computations, Applications* **11** No 2, 113 (2020).
10. P. Stukhliak, O. Totosko, D. Stukhlyak, O. Vynokurova, I. Lytvynenko, *CEUR Workshop Proceedings* **3896**, 405 (2024).
11. A. Buketov, S. Smetankin, P. Maruschak, O. Sapronov, V. Matvyeyev, A. Menou, *Transport* **35** No 6, 679 (2020).
12. A. Momber, M. Irmer, N. Gluck, *Cold Reg. Sci. Technol.* **127**, 76 (2016).
13. T. Krenický, K. Dyadyura, O. Kokhan, S. Umynskyi, *Management Systems in Production Engineering* **33** No 4, 470 (2025).
14. O.O. Sapronov, A.V. Buketov, D.O. Zinchenko, V.M. Yatsyuk, *Composites: Mechanics, Computations, Applications* **8** No 1, 47 (2017).
15. K.A. Dyadyura, K.V. Berladir, P.V. Rudenko, O.A. Budnik, V.A. Sviderskij, *Proceedings of the 6th International Conference Nanomaterials: Applications and Properties*, NAP 2016, Conference Paper (2016).
16. K.V. Berladir, O.A. Budnik, K.A. Dyadyura, V.A. Svidersky, Ya.O. Kravchenko, *High Temperature Material Processes* **20** No 2, 157 (2016).
17. A. Panda, K. Dyadyura, P. Savchuk, V. Kashytskyi, V. Malets, et al., *MM Science Journal*, 3509 (2019).
18. K. Luo, et al., *Mater. Today Commun.* **47**, 113053 (2025).
19. R. Kumar, S. Kumar, B. Rai, P. Samui, *Structures* **66**, 106850 (2024).
20. *Scikit-learn: Ensemble methods*, (accessed 01.12.2025).

Застосування ансамблевих алгоритмів машинного навчання для моделювання термомеханічних властивостей нанонаповнених епоксидних композитів

О.А. Пастух[1], Ю. Петров[1], А.В. Букетов[2], К. Дядюра[3], О.Л. Ляшук[1], О.В. Тотосько[1], В.В. Соценко[2]

[1] *Тернопільський національний технічний університет імені Івана Пулюя, 46001 Тернопіль, Україна*
[2] *Херсонська державна морська академія, 73000 Херсон, Україна*
[3] *Одеський державний аграрний університет, 65012 Одеса, Україна*

Стаття присвячена комплексному підходу до прогнозування теплостійкості пластифіко-ваних епоксидних нанокомпозитів. Метою дослідження було застосувати ансамблеві алго-ритми машинного навчання для визначення ключових механічних факторів при прогно-зуванні теплостійкості епоксидних нанокомопзитів та створення надійного методу їх відбору для подальшого моделювання. Для досягнення мети застосовано двоетапну гібридну методологію. На першому етапі використано ансамблеві алгоритми машинного навчання, зокрема метод f_regression, для розрахунку важливості ознак (feature importance), оцінки їх інформативності (Information Value) та визначення рівня узгодженості (Consistency Score) між різними моделями. На другому етапі виконано класичний статистичний аналіз, включно з кореляцією Пірсона та одновимірним F-тестом, для оцінки лінійних зв'язків. Результати виявили високу узгодженість між методами. Встановлено, що Elastic modulus та Adhesive strength мають стабільно високу значущість (Information Value = 0.15 та 0.12 відповідно; Consistency Score 82% та 81% відповідно) і є ключовими детермінантами теп-лостійкості. Навпаки, Destruction stresses та Residual Stresses є абсолютно неінформатив-ними (Information Value = 0) та рекомендовані до виключення з моделей. Концентрація Filler показала мінімальний вплив у досліджуваному діапазоні. Таким чином доведено ефективність гібридного підходу, що поєднує класичну статистику та машинне навчання, для об'єктивного та обґрунтованого відбору ознак при прогнозуванні властивостей компо-зитних матеріалів. Запропонований алгоритм дій дозволяє створювати спрощені, інтер-претовані та точні моделі прогнозування для інженерного проєктування.

Ключові слова: Епоксидні нанокомпозити, Теплостійкість, Механічні властивості, Відбір ознак, Кореляційний аналіз, Машинне навчання, Ансамблеві алгоритми машинного навчання, Інформаційна цінність.