THE RESULTS OF THE SPECIALISTS' AND SCIENTISTS' RESEARCHES

Original article
https://doi.org/10.15828/2075-8545-2023-15-2-171-186

CC BY 4.0

Compressive strength prediction and composition design of structural lightweight concretes using machine learning methods

Artemy S. Balykov*, **Elena A. Kaledina**, **Sergey V. Volodin**
National Research Mordovia State University, Saransk, Russia
* Corresponding author: e-mail: artbalrun@yandex.ru

ABSTRACT: Introduction. Reducing the density, increasing the strength and other physical-technical characteristics of lightweight concretes are urgent tasks of modern building materials science. To solve them, it is necessary to consider new approaches to the development of compositions of cement systems using effective porous aggregates, binders, chemical and mineral additives, including different nanomodifiers (carbon nanotubes, fullerenes, nanoparticles of SiO_2, Al_2O_3, Fe_2O_3, etc.). The complexity of designing modified cement concretes is largely due to their multicomponent nature and a large number of parameters affecting the key characteristics of material. The qualitative solution of such multicriteria problems is possible with the complex implementation of rational physical and computational experiments using mathematical modeling and computer technology. New opportunities for modeling of structure formation processes and predicting properties of multicomponent building materials are emerging with the development of machine learning methods. The purpose of this study is to develop machine learning algorithms that can efficiently establish quantitative dependences for the compressive strength of modified lightweight concretes on their composition, as well as to identify the optimal variation ranges of prescription parameters based on the obtained multifactor models to achieve the required level of controlled mechanical characteristic. **Methods and materials.** The processing and analysis of experimental research results were carried out using modern methods of machine learning with a teacher used in the problems of regression recovery, knowledge extraction and forecasting. To implement the developed machine learning algorithms, libraries in the Python programming language, in particular NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, were used. **Results and discussion.** It is established that the gradient boosting model is the most accurate type among the obtained machine learning models. It is characterized by the following quality metrics: R^2 = 0.9557; MAE = 2.4847; MSE = 12.7704; RMSE=3.5736; MAPE = 11.1813%. According to the analysis of this multifactor model, the optimal dosages of pozzolanic and expanding modifiers amounted to 4.5–6.0% and 6.0–7.5% of the binder weight (Portland cement + modifier), respectively, which ensured achievement of the required level of compressive strength (40–70 MPa) of lightweight concretes at the age of 28 days at material density reduced by 3–10% (the range under consideration is 1200–1900 kg/m^3). **Conclusions.** Thus, the study results show the prospects of using machine learning methods for design compositions and predicting properties of multicomponent cement systems.

KEYWORDS: lightweight concrete, nanomodifier, complex additive, nanoparticle, hollow microsphere, compressive strength, design, prediction, optimization, machine learning, algorithm, model, quality metric.

ACKNOWLEDGMENTS: The research was supported by grant from the Russian Science Foundation (project No. 21-73-00228), https://rscf.ru/en/project/21-73-00228/.

FOR CITATION: Balykov A.S., Kaledina E.A., Volodin S.V. Compressive strength prediction and composition design of structural lightweight concretes using machine learning methods. *Nanotechnologies in Construction.* 2023; 15(2): 171–186. https://doi.org/10.15828/2075-8545-2023-15-2-171-186. – EDN: CCPGIF.

INTRODUCTION

The development of physico-chemical and technological bases for the obtaining of modified cement systems characterized by complex of high performance characteristics is actual direction of modern building materials science [1–7]. It is known that the display properties of such concretes during operation depends on their composition, structure and state [8].

The work [9] summarizes the main generalized principles of forming the structure of high performance cement concretes: providing increased density at each of the

© Balykov A.S., Kaledina E.A., Volodin S.V., 2023

scale levels of material structure (macro-, meso-, micro-, nano-) along with a maximally increased area of intercomponent interface; increasing the strength of structural bonds; achieving optimal uniformity – non-uniformity of structure; optimizing content of structural components to ensure the required level of material performance properties (strength, elasticity, impact toughness, permeability, corrosion resistance, etc.) under mechanical loading and exposure to the environment.

High-strength lightweight concretes, which have increased strength characteristics at reduced material density, are one of the promising types of modified cement systems. Global experience shows that it is possible to achieve high specific strength of lightweight concretes by using specially selected porous aggregates, by optimization of the binder and frame-forming components, and by using nanomodifiers [10–14]. At the same time, among porous aggregates, expanded clay, glass and aluminosilicate hollow microspheres, porous rocks of volcanic origin, processing products of multi-tonnage technogenic wastes from metallurgy and thermal power engineering, etc. have shown their efficiency in the formulation of structural lightweight concretes [14–16].

Separately, it is worth dwelling on the importance of using various kinds of modifiers in the compositions of high perfomance cement systems, namely chemical and mineral additives used individually or combined into complexes according to the principles of additivity and synergy. In particular, the following modifiers are most effective for the formulation of high-strength concretes:

- plasticizing additives to significantly improve the technological characteristics of concrete mixtures and the physical-mechanical properties of concretes due to the ability to exert a water-reducing and plasticizing effects on cement systems [17–19];
- finely dispersed pozzolanic additives of natural and technogenic origin, characterized mainly by silicate and aluminosilicate composition with an increased content of amorphous silica and alumina, as well as other reactive phases: opal-cristobalite rocks [20–22], metakaolin and calcined polymineral clays [23–26], silica fume [27, 28], fly ash [29, 30], blast furnace slag [30], etc.;
- expanding additives of the sulfoaluminate type [31–33], which have the ability to control deformations of cement stone by stimulating and intensifying the formation of crystalline hydrates with increased volume (ettringite, etc.);
- promising nanomodifiers (carbon nanotubes, fullerenes, nanoparticles of SiO_2, Al_2O_3, Fe_2O_3, etc.) that improve the elastic-strength characteristics of cement stone, reduce shrinkage deformations of concretes and their permeability to aggressive environments due to the possibility of controlling the structural and energy state of interphase boundaries, directional change in the quality of solid phase and pore space geometry [4, 34–36].

Currently, despite the unremitting interest in high performance concretes, relatively simple and universal methods for designing compositions and predicting their properties have not yet been found. This is largely due to the multicomponent nature of modified cement systems, as well as a large number of parameters affecting the key characteristics of material.

At the moment, one of the most popular methods used to optimize the prescription-technological parameters for obtaining multicomponent building materials is experimental and statistical modeling [36–39]. However, this method not only requires the correct formulation of a relatively complex multicriteria problem, but is often characterized by relatively low accuracy of ES-models obtained as a result of processing small amounts of data in the presence of numerous assumptions about them.

With the development of artificial intelligence technologies, new opportunities appear in the design of multicomponent building materials, including modified cement systems. The research results obtained by some foreign authors concerning the application of machine learning methods for predicting the properties of high performance concretes [40–43] testify to this. It is worth noting that the increasing number of publications on the subjects under consideration confirms the intensive development of this scientific direction.

The purpose of this study was to develop machine learning algorithms that can efficiently establish quantitative dependences for the compressive strength of modified lightweight concretes on their composition, as well as to identify the optimal variation ranges of prescription parameters based on the obtained multifactor models to achieve the required level of controlled mechanical characteristic.

The following tasks were solved during the research:

1) the required amount of experimental data was obtained;

2) the machine learning task by precedents was formulated;

3) the experimental data preprocessing and extraction of signs from them were performed;

4) the choice of types of multiparameter models and development of machine learning algorithms for their obtaining were carried out;

5) the data preparation for training and evaluation (training and test samples) was performed; training, as well as solving the problems of optimization and retraining of models were carried out;

6) the quality of machine learning models was evaluated;

7) based on the analysis of the most effective model the optimization of the content of complex mineral ad-

ditives (pozzolanic and expanding modifiers) to achieve the required level of compressive strength of lightweight concretes at the age of 28 days was carried out.

METHODS AND MATERIALS

Materials

The following main components were used to obtain the high-strength lightweight concretes:

- Portland cement CEM I 42.5R (PC) produced by Mordovcement PJSC;
- medium natural quartz sand (QS) of the Khromtsovsky deposit (Ivanovo region) with fineness modulus (module size) $M_{fm} = 2.1$;
- hollow glass microspheres (HGM) grade ForeSphere 3000 with predominant particle size of 30–160 microns, hydrostatic compressive strength of at least 20 atm (3000 psi), true and bulk density of 0.53 g/cm^3 and 0.3 g/cm^3, respectively, produced by Russian company Fores LLC;
- polycarboxylate superplasticizer (PS) grade Melflux 1641 F produced by BASF Construction Additives;
- two types of complex mineral additives (MA) with particle size distribution in the micrometer and upper nanometer ranges:

1) silicon pozzolanic modifier (SPM) is two-component powder material with specific surface area of S_{ss} = 1.85 m^2/g obtained by grinding the mixture of opal-cristobalite rocks from deposits of the Republic of Mordovia (diatomite + opoka);

2) sulfoaluminosilicate expanding modifier (SEM) is two-component powder material with specific surface area of S_{ss} = 0.6 m^2/g obtained by grinding and calcination the mixture of polymineral clay rock from deposit of the Republic of Mordovia and semi-aquatic molding gypsum grade G-6 B III (the Russian State Standard GOST 125-2018) produced by Magma LLC.

Methods

The processing and analysis of experimental research results were carried out using modern methods of machine learning with a teacher used in the problems of regression recovery, knowledge extraction and forecasting. To implement the developed machine learning algorithms, libraries in the Python programming language, in particular NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, were used.

The process of solving machine learning tasks included the following main stages: problem statement; preprocessing experimental data and extracting features from them; selection of types of multiparameter models; preparation of data for training and evaluation (training and test samples); training, solving problems of optimization and retraining, evaluating the quality of models; selection of the most effective model; the final presentation of results.

Data preprocessing

We used the correlation matrix and the *variance inflation factor (VIF)* to test the data for multicollinearity. The formula for calculating VIF looks like:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad (1)$$

where R^2_j is the determination coefficient of the j-th attribute.

When the value of $VIF_j > 10$ it is considered that the j-th factor has multicollinearity.

The principal component analysis (PCA) was used to eliminate the multicollinearity of attributes.

Since the input parameters were characterized by different scales and ranges of variation, their normalization was performed at the final stage of data preprocessing. The following data normalization formula was used in the paper to prevent an imbalance between the influence of the input variables, and hence to avoid obtaining incorrect dependencies:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (2)$$

where x_{norm} and x are the normalized and current values of each input variable;

x_{min} and x_{max} are the minimum and maximum values of each input variable.

Machine learning models used

The study used following machine learning models:

1. Linear regression is model of type:

$$a(x) = \langle w, x \rangle. \quad (3)$$

The model parameters are the weights coefficients w, which are found through gradient learning methods.

2. Automatic relevance determination (ARD regression) is type of Bayesian regression in which posteriori variance estimate is derived for each coefficient. In the next step, the coefficients characterized by low value of variance are zeroed.

3. Decision tree with hyperparameter search is nonparametric supervised machine learning algorithm. This model is characterized by its ability to predict the value of the target variable based on the study of simple decision rules derived from the characteristics of the data. The decision tree can be viewed as piecewise constant approximation. The functional of quality $Q(R_m, j, s)$ in this case has the following form:

$$Q(R_m, j, s) = H(R_m) - \frac{|R_l|}{|R_m|}H(R_l) - \frac{|R_r|}{|R_m|}H(R_r), \quad (4)$$

where R_l and R_r are objects falling, respectively, into the left and right subtree at given predicate;

R_m is set of objects that fell into the vertex being split at the given step;

j is the input parameter number, according to which the partition is carried out in this predicate;

s is the classifier threshold dividing the set R_m into the right and left subtree;

$H(R_l)$, $H(R_r)$ and $H(R_m)$ are the informativity criterions evaluating the quality of the distribution of the target variable among the objects of the sets R_l, R_r и R_m, respectively.

For regression problems, the informativeness criterion $H(R_m)$ looks as follows:

$$H(R_m) = \frac{1}{|R_m|}\sum_{(x_i, y_i)\in R_m}(y_i - \frac{1}{|R_m|}\sum_{(x_j, y_j)\in R_m} y_j)^2, \quad (5)$$

where y_i and x_i are the values of the output and input parameters for the i-th sample object belonging to the set R_m, respectively;

y_j and x_j are the values of the output and input parameters for the j-th sample object belonging to the set R_m, respectively.

To improve the quality of this machine learning model, its hyperparameters were configured, i.e., the parameters defined before the start of the learning process. The search for the optimal set of hyperparameters was performed using the *GridSearch* method.

4. Bagging regressor is ensemble algorithm that selects baseline regressors for each of the random subsets of the original data set and then combines and averages their individual predictions to obtain the final predicted result.

5. Random forest is ensemble of independent decision trees, in the training of which for each partition the attributes are selected from some random subset of features. The final classifier $a(x)$ for the regression problem is the average of trees $b_i(x)$:

$$a(x) = \frac{1}{K}\sum_{i=1}^{K} b_i(x), \quad (6)$$

where K is the total number of decision trees $b_i(x)$ in the ensemble, it is the matched parameter.

6. Gradient boosting is machine learning method that creates decisive prediction model $a_K(x)$ as ensemble of K basic algorithms $b_1, \ldots, b_K$:

$$a_K(x) = \sum_{i=1}^{K} \gamma_i b_i(x), \quad (7)$$

where K is the total number of basic prediction algorithms $b_i(x)$. Decision trees of fixed length were used as the basic algorithm in this model.

γ_i are numerical coefficients for the basic algorithms $b_i(x)$, $i = \overline{1, K}$.

The use of this method allows building the model step by step with the possibility of optimizing an arbitrary differentiable loss function. At the stage when the composition from the $K-1$ algorithm is obtained, the next basic algorithm $b_K(x)$ is formed based on minimizing the construction error and approximating the gradient of loss function s_{iK} on the training sample:

$$\sum_{i=1}^{N} L(y_i, a_{K-1}(x_i) + \gamma b_K(x_i)) \to \min_{b_K, \gamma_K}, \quad (8)$$

$$b_K(x) = \arg\min_b \sum_{i=1}^{N} (b(x_i) - s_{iK})^2 \quad (9)$$

where L is the selected loss function, most often it is quadratic;

N is the training sample size;

y_i and x_i are the current values of the output and input parameters for the i-th sample object, respectively;

$a_{K-1}(x)$ is the regression model at step $K-1$.

After finding the base algorithm $b_K(x)$, the coefficient for it γ_K is determined:

$$\gamma_K = \arg\min_{\gamma\in R} \sum_{i=1}^{N} L(y_i, a_{K-1}(x_i) + \gamma b_K(x_i)). \quad (10)$$

7. Gradient boosting with the search for optimal hyperparameters by using the GridSearch method.

Model quality metrics

We used the *coefficient of determination* R^2 to assess the quality of machine learning models. The formula for calculating this metric looks like:

$$R^2(y, \hat{y}) = 1 - \left(\sum_{i=1}^{l}(y_i - \hat{y}_i)^2 \Big/ \sum_{i=1}^{l}(y_i - \bar{y})^2\right), \quad (11)$$

where y_i and $\hat{y}_i$ are the actual and predicted values of the target variable for the i-th object of the test sample, respectively;

$\bar{y}$ is average of the actual values;

l is the total number of test sample objects.

The following regression error metrics were used to estimate the deviation of model predictions from true values:

1. The mean absolute error (MAE), which is calculated by the formula:

$$MAE(y,\hat{y}) = \frac{1}{l}\sum_{i=1}^{l}|y_i - \hat{y}_i|. \quad (12)$$

When using this metric, model error is calculated as the average of absolute differences between targets and predictions. MAE is linear estimate in which the distinctions for each object are weighted equally on average.

2. The mean squared error (MSE). For each point, the square of the difference between the predicted and target values of the model is calculated, and then these values are averaged. The calculation formula for this metric is the following:

$$MSE(y,\hat{y}) = \frac{1}{l}\sum_{i=1}^{l}(y_i - \hat{y}_i)^2. \quad (13)$$

3. The root mean squared error (RMSE), which is the square root of the MSE:

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(y_i - \hat{y}_i)^2}. \quad (14)$$

4. The mean absolute percentage error (MAPE), which is calculated by the formula:

$$MAPE(y,\hat{y}) = \frac{1}{l}\sum_{i=1}^{l}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%, \quad (15)$$

where y_i and $\hat{y}_i$ in l formulas (12), (13), (14), and (15) are the same as in formula (11).

The MAPE indicator is relative error calculated for the analyzed object from the training data set by dividing the absolute error by the target value. Thus, MAPE can also be considered as weighted version of MAE. This metric can be used to compare the efficiency of models on different training samples.

In conditions of small data set, *cross-validation (CV)* was used to accurately calculate the metrics of the analyzed models. In the *k-fold CV* approach, the training sample is split into k parts, and k iterations is performed. At each iteration, the model is trained on the $k-1$ group and tested on the remainder of the data not used in training. The performance measure $CV\,(a, X^L)$ reported by k-fold cross-validation is the average of the parameters computed in the loop:

$$CV\,(a,X^L) = \frac{1}{k}\sum_{i=1}^{k} Q(a(X^L \backslash X^{k_i}), X^{k_i}). \quad (16)$$

where $a(x)$ is the machine learning model;
Q is metric;
X^L is the entire training sample;
X^{k_i} is part of the training sample with the number k_i.

Additional validation tests

In addition to the quality metric R^2 and the above-mentioned error metrics, the following tests to evaluate their quality used for additional comparison of the most effective models:

1. The REC curve analysis of the model. To compare the regression models with each other, as well as with the baseline forecast, we built REC-curves, the graph of which shows the accuracy of the model depending on the acceptable error size. Next, we calculated the Area Over Curve (AOC) for the model under study (AOC_{model}) and the baseline forecast ($AOC_{baseline}$), followed by an assessment of its quality based on the ratio $AOC_{model}/AOC_{baseline}$.

The area over the REC-curve is asymptotically equal to the error mathematical expectation, which allows using this metric for model comparison.

2. Spearman correlation coefficient, which is estimate of the measure of the linear relationship between random variables and calculated by the formula:

$$\rho(y,\hat{y}) = \frac{cov(rg_y, rg_{\hat{y}})}{\sigma_{rg_y}\sigma_{rg_{\hat{y}}}}, \quad (17)$$

where rg_y, $rg_{\hat{y}}$ are the ranks of actual and predicted values;

σ_{rg_y}, $\sigma_{rg_{\hat{y}}}$ are the variance of actual and predicted values, respectively.

Calculation of this coefficient allows for assessing the degree of consistency between the predicted and actual values of the studied parameter.

3. Building the learning curve, which is graph of changes in the learning rate of the model. This test makes allows to establish the degree of initial learning difficulty and determine the level of accuracy of the model fitting, its retraining, as well as the data representativeness.

RESULTS AND DISCUSSION

Description of the analyzed dataset

Experimental data set of 407 records (points) was used to build machine learning models. The inputs were 8 varying parameters: the age of lightweight concrete (days) and the consumption of the main prescription components (kg/m^3), in particular, Portland cement (PC); silicon pozzolanic (SPM) and sulfoaluminosilicate expanding (SEM) modifiers; polycarboxylate superplasticizer (PS), natural quartz sand (QS), hollow glass microspheres

Table 1
Statistical characteristics of the dataset

Statistical indicators	Target parameter	Input parameters							
	Compressive strength, MPa	Age of concrete, days	Consumption of prescription components, kg/m³						
			PC	SPM	SEM	PS	QS	HGM	W
count	407.00								
mean	45.41	21.64	675.85	26.56	31.17	7.33	420.58	163.12	285.08
std	16.40	10.71	44.91	38.27	41.57	0.19	261.78	45.62	19.05
min	3.70	1.00	595.00	0.00	0.00	7.00	56.00	97.60	255.20
25%	37.85	7.00	641.20	0.00	0.00	7.20	194.45	124.40	266.60
50%	47.80	28.00	675.90	0.00	0.00	7.30	410.80	165.00	285.90
75%	57.50	28.00	711.80	54.80	57.20	7.50	638.60	203.20	302.60
max	72.10	28.00	764.10	112.60	114.00	7.60	794.70	227.80	313.20

(HGM), and water (W). The target variable under study was compressive strength (MPa).

Table 1 shows the statistical characteristics of the experimental data under study. Analysis of the values of the target and input parameters showed that none of the variables had outliers.

Figure 1 shows plots of pairwise dependences of the compressive strength of lightweight concrete on the consumption of prescription components and the age of the material. According to the study results, it was found that there was no linear correlation between the target indicator and each input parameter, which indicated the presence of complex nonlinear relationships in the binary system "Compressive strength – flow rate of the prescription component / age of concrete".

The next step was to check the input parameters for collinearity by using the results of data correlation analysis and calculating the variance inflation factor (VIF). Figure 2 shows the correlation matrix in the form of a so-called "heat map".

The correlation matrix shows linear relationship between the parameters "Superplasticizer", "Sand", "Microspheres" and "Water". In turn, the results of the VIF calculation indicate the presence of multicollinearity in these input parameters, which can reduce efficiency of the developed machine learning algorithms. The study used the principal component analysis method to overcome multicollinearity in data preprocessing.

Model training

The training dataset was divided into training and test samples in a ratio of 65/35 (265 objects (data points) were used for training models, and 142 were used for testing).

Figures 3, 4, and 5 show changes in the actual values of the investigated output parameter over the entire set of experimental data compared to the values predicted by each developed machine learning algorithm. It was found that the models "Linear Regression", "ARD", "Bagging Regressor" and "Random Forest" were prone to some overestimation of the values of compressive strength of lightweight concrete.

Table 2 shows the quality metrics of the developed machine learning models (values of the determination coefficient R^2 and regression errors MAE, MSE, RMSE, MAPE).

The following machine learning models were found to have the highest values of the determination coefficient: gradient boosting ($R^2 = 0.9557$), gradient boosting with the search for optimal hyperparameters ($R^2 = 0.9465$), bagging regressor ($R^2 = 0.9411$), and random forest ($R^2 = 0.9401$). The other algorithms, despite sufficiently high values of the analyzed parameter of model quality, showed reduced efficiency, especially on test data. In particular, it is worth noting the significant deviation of the test and training data from the model straight line for linear regression and ARD (Fig. 6, a, b, c, d).

In addition, based on the results of the analysis of the training graphs, we can conclude the overfitting of the obtained decision tree model. This is confirmed by the crowding of the training data around the model line with minimum number of outliers (Fig. 6, e), as well as a noticeable decrease in the determination coefficient as a result of algorithm processing of the test data (Fig. 6, f).

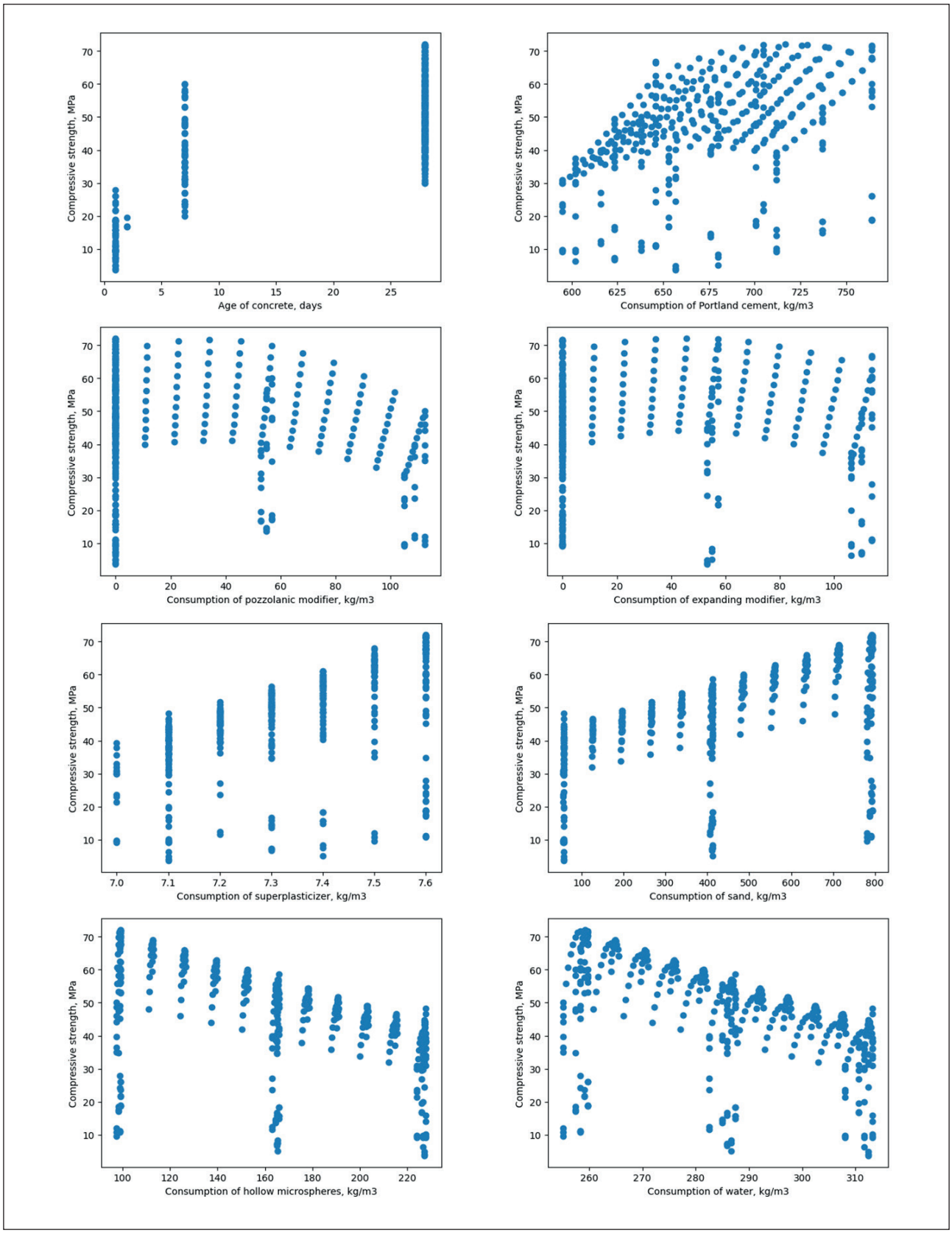

Fig. 1. **Graphs of pairwise dependencies of the target indicator "Compressive strength of lightweight concrete" on separate input parameters (consumption of prescription components and age of the material)**

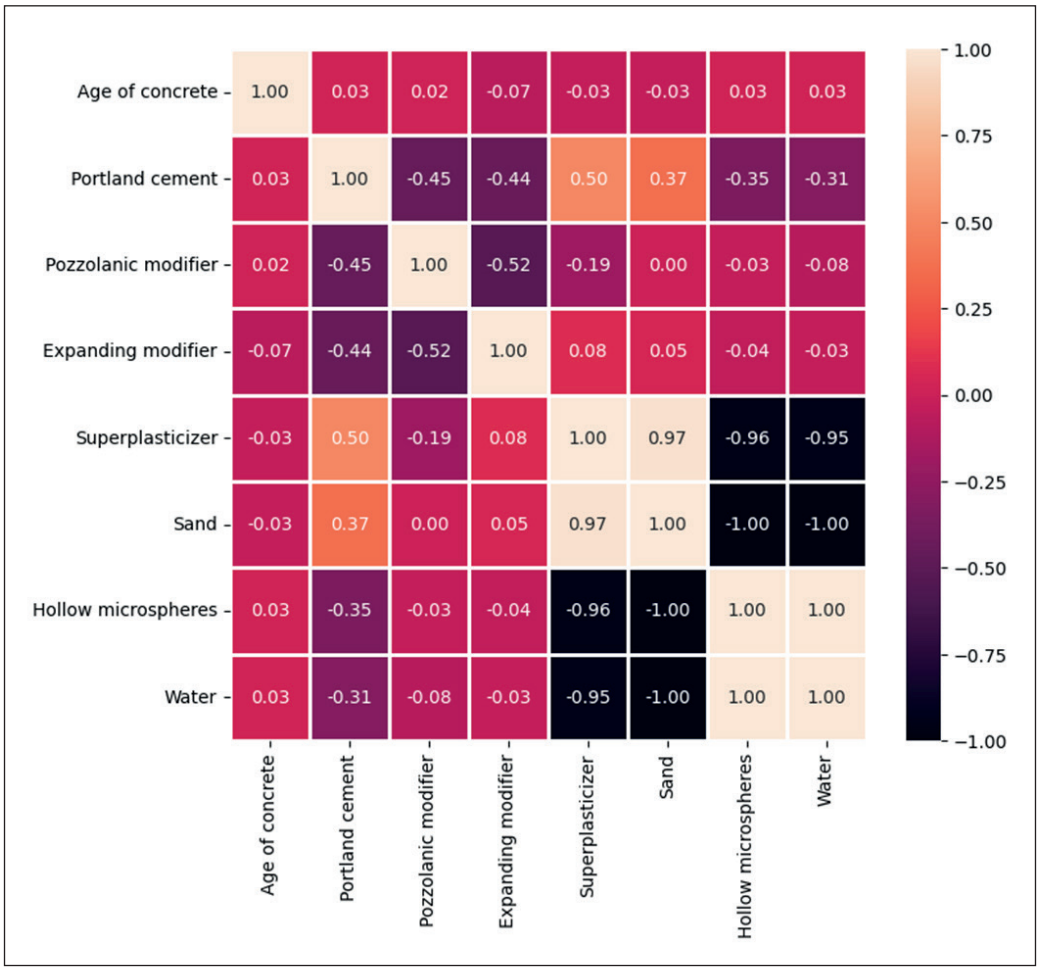

Fig. 2. **Correlation matrix of input parameters**

It is worth noting that, according to Table 2, the lowest regression error values have the "Gradient Boosting" and "Gradient Boosting with the Search for Optimal Hyperparameters" models: MAE = 2.4847–2.4917; MSE = 12.7704–15.4094; RMSE = 3.5736–3.9255; MAPE = 11.1813–11.6773%. In this case, among the developed machine learning models, the most accurate is the gradient boosting model characterized by the following parameters :Huber loss function; maximum tree depth is 10; minimum number of samples required for a finite node is 6; minimum number of samples required to divide an internal node is 10; the number of trees is 1100. Figure 7 shows the results of evaluating the efficiency of the gradient boosting algorithm on the training and test parts of the experimental data.

Optimization of the content of complex mineral modifiers in the compositions of high-strength lightweight concretes

At the final stage, based on the analysis of the developed multifactorial model of gradient boosting, optimization of the content of complex mineral additives (silicon pozzolanic (SPM) and sulfoaluminosilicate expanding (SEM) modifiers) was conducted to achieve the required level of the studied strength characteristics of concretes at the age of 28 days. To effectively solve the task, graphs were plotted in the form of isolines of changes in the compressive strength of lightweight concretes depending on the content of hollow glass microspheres, as well as dosages of pozzolanic and expanding additives (Figure 8).

According to the results of the analysis of obtained graphic dependencies, it was found that the optimum dosages of silicon pozzolanic and sulfoaluminosilicate expanding modifiers amounted to 4.5–6.0 and 6.0–7.5% of the binder weight (Portland cement (PC) + mineral additive (MA)), respectively. The use of additives SPM and SEM in the indicated concentrations ensured the achievement of the required level of strength indicator (40–70 MPa) at reduced by 3–10% (the range under consideration is 1200–1900 kg/m^3) due to the possibility of increasing the content of microspheres (by 2–4% of the binder weight) without compromising the mechanical characteristic. In this case, according to Figure 8, the

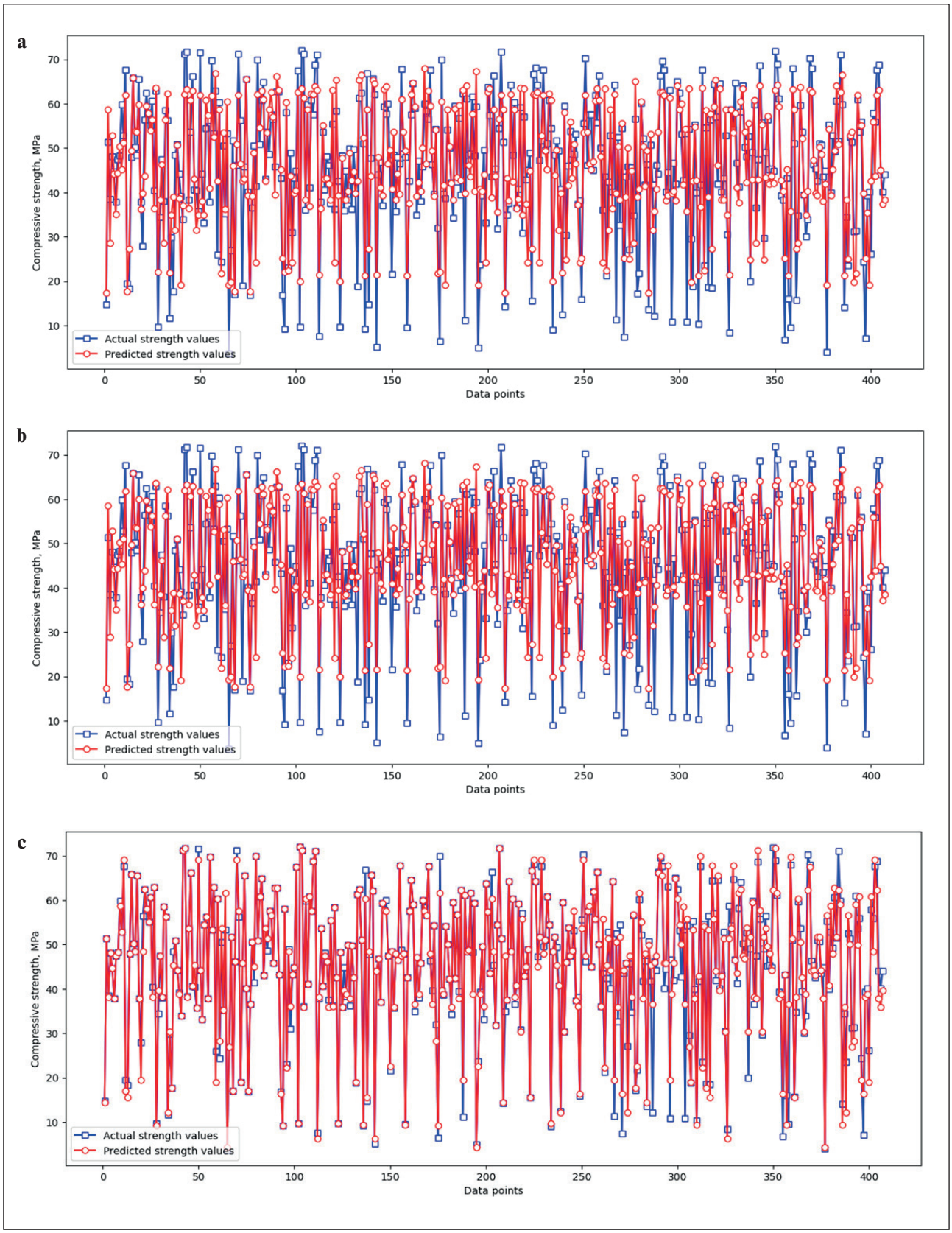

Fig. 3. **Comparison of actual and predicted values of compressive strength of lightweight concrete on the entire dataset using the following machine learning models:** (a) linear regression; (b) ARD; (c) decision tree

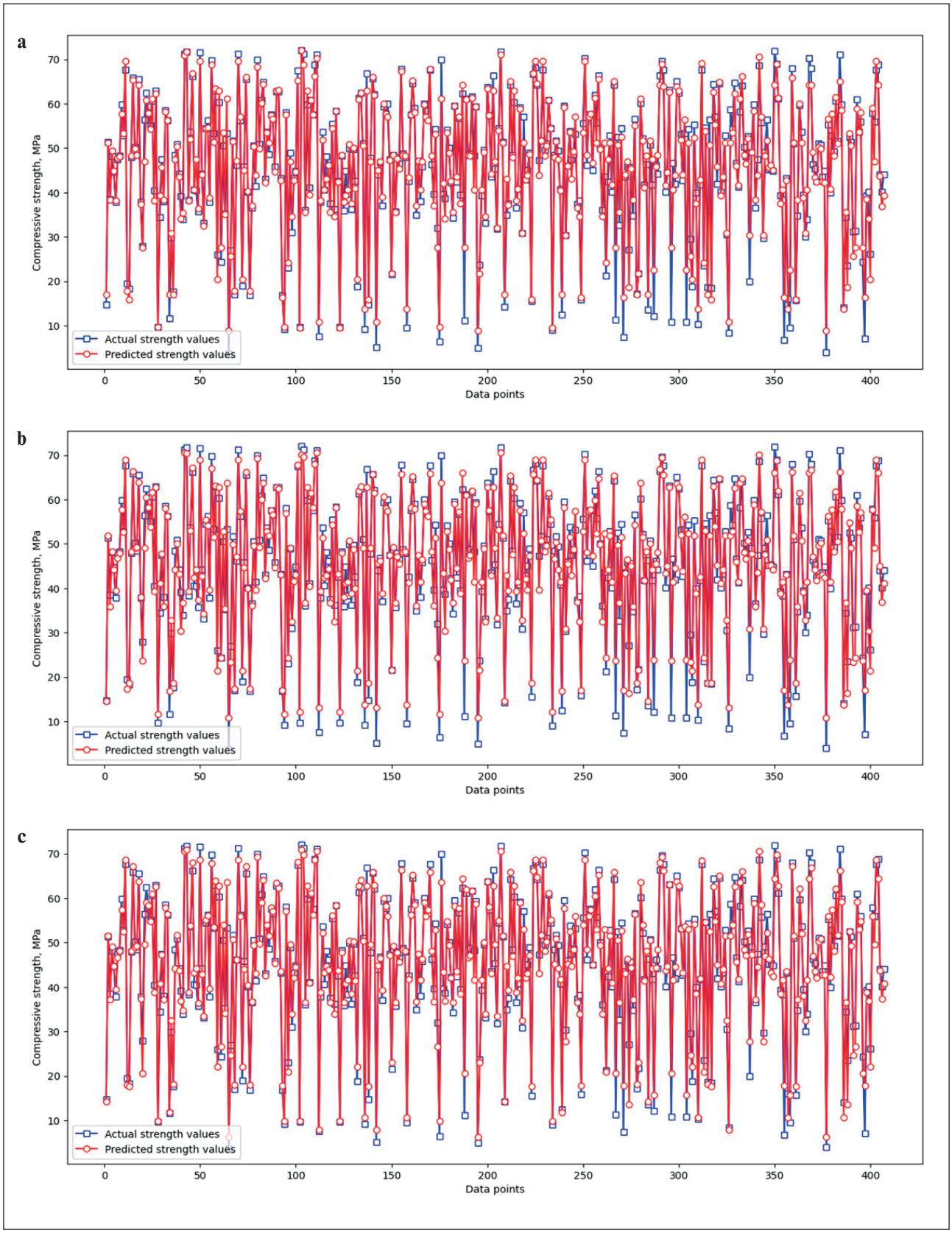

Fig. 4. **Comparison of actual and predicted values of compressive strength of lightweight concrete on the entire dataset using the following machine learning models:** (a) bagging regressor; (b) random forest; (c) gradient boosting

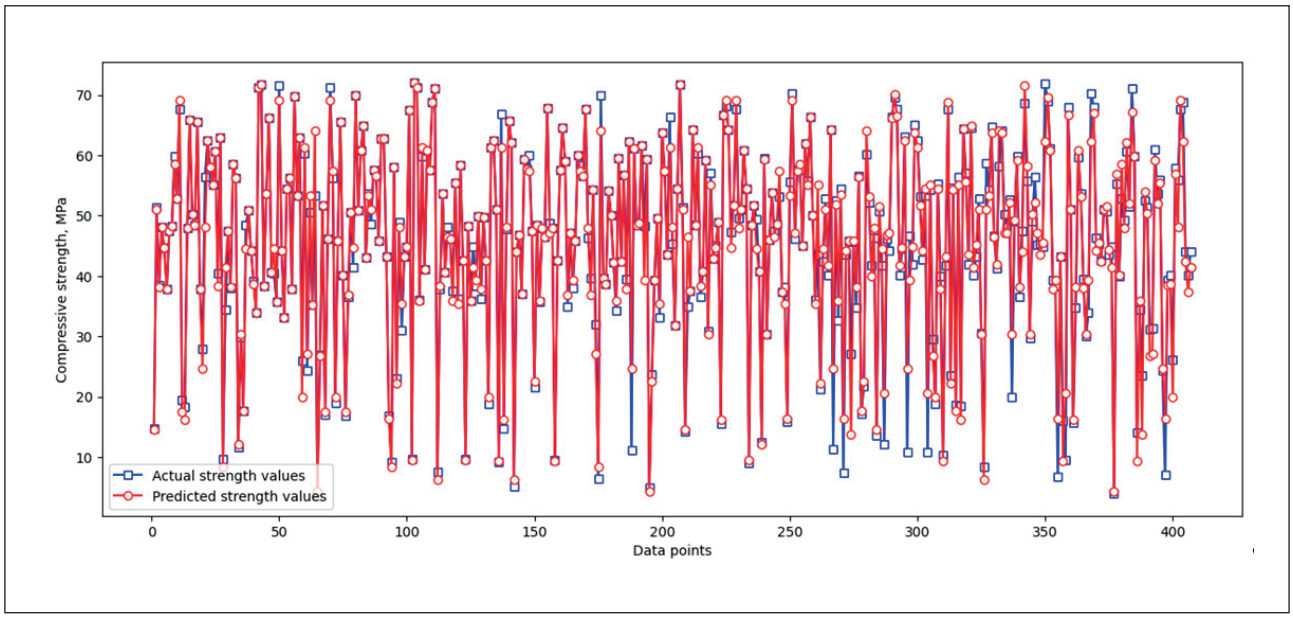

Fig. 5. **Comparison of actual and predicted values of compressive strength of lightweight concrete on the entire dataset using the machine learning model "Gradient boosting with the search for optimal hyperparameters"**

Table 2
Quality metrics of the developed machine learning models (values of determination coefficient and regression errors)

Machine learning models	Quality metrics of models				
	R^2	*MAE*	*MSE*	*RMSE*	*MAPE* (%)
Linear regression	0.7374	6.0752	75.6671	8.6987	31.5038
ARD	0.7361	6.0861	76.0399	8.7201	31.6589
Decision tree	0.8953	3.4091	30.1583	5.4917	15.5677
Bagging regressor	0.9411	2.7292	16.9749	4.1201	13.3829
Random forest	0.9401	2.7790	17.2584	4.1543	14.0700
Gradient boosting	0.9557	2.4847	12.7704	3.5736	11.1813
Gradient boosting with the search for optimal hyperparameters	0.9465	2.4917	15.4094	3.9255	11.6773

increase in compressive strength of lightweight concretes up to the level of 70–73 MPa can be achieved at the dosages of hollow microspheres, pozzolanic and expanding modifiers 13%; 2–7% and 2–10% of binder weight (PC + MA), respectively.

CONCLUSIONS

The conducted scientific studies allowed to develop machine learning algorithms that can effectively establish quantitative dependences for the compressive strength of modified lightweight concretes on their composition. Based on the results of the analysis of the obtained multifactor models, the optimal variation ranges of dosages of the pozzolanic and expanding modifiers were identified, which ensured the achievement of the required level of controlled mechanical characteristic at reduced material density.

The study results showed the prospects of using machine learning methods for design compositions and predicting properties of multicomponent lightweight concretes.

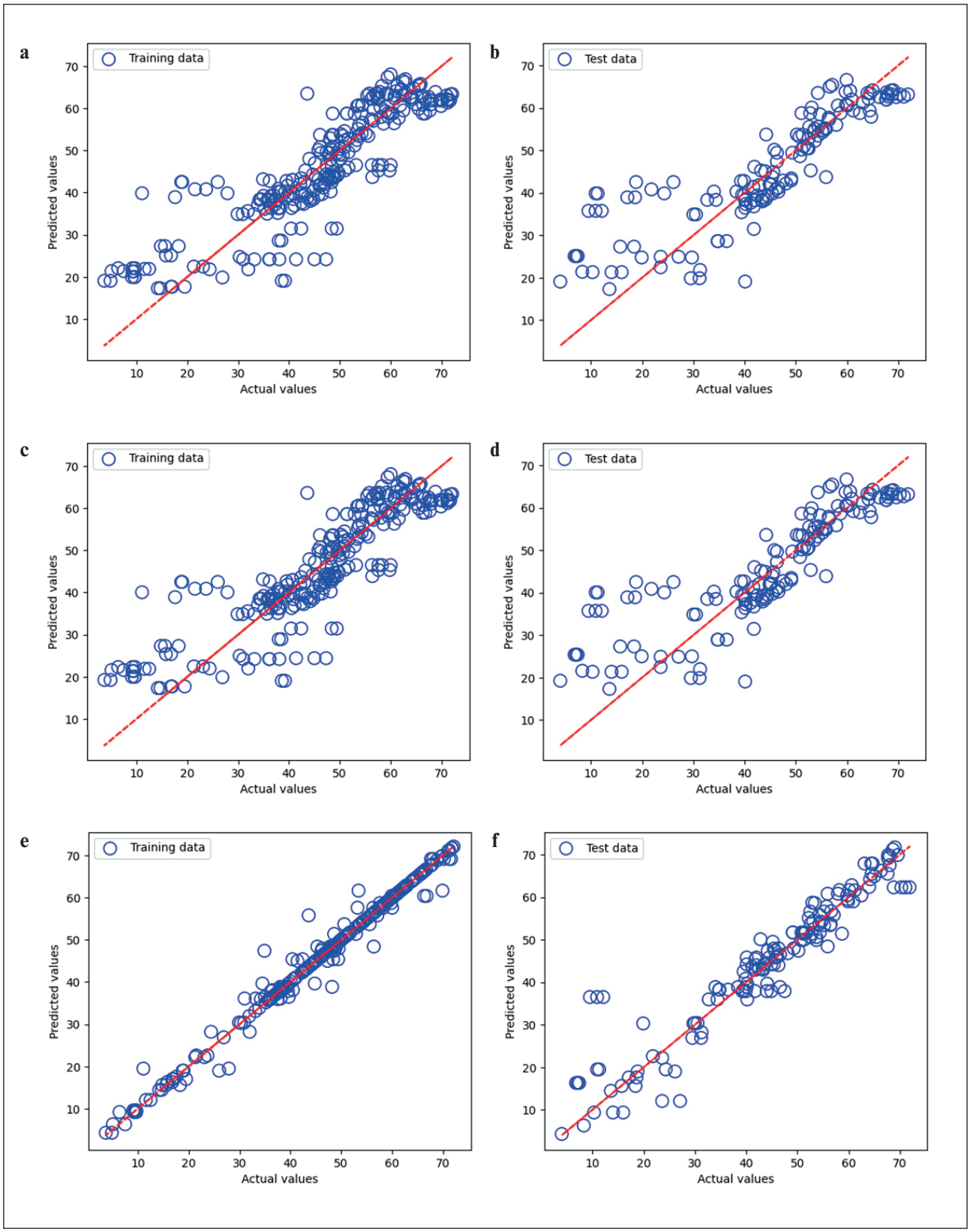

Fig. 6. **Efficiency of the linear regression (a, b), ARD (c, d), and decision tree (e, f) algorithms on the training (a, c, e) and test (b, d, f) parts of the experimental data** (the red line indicates the model line)

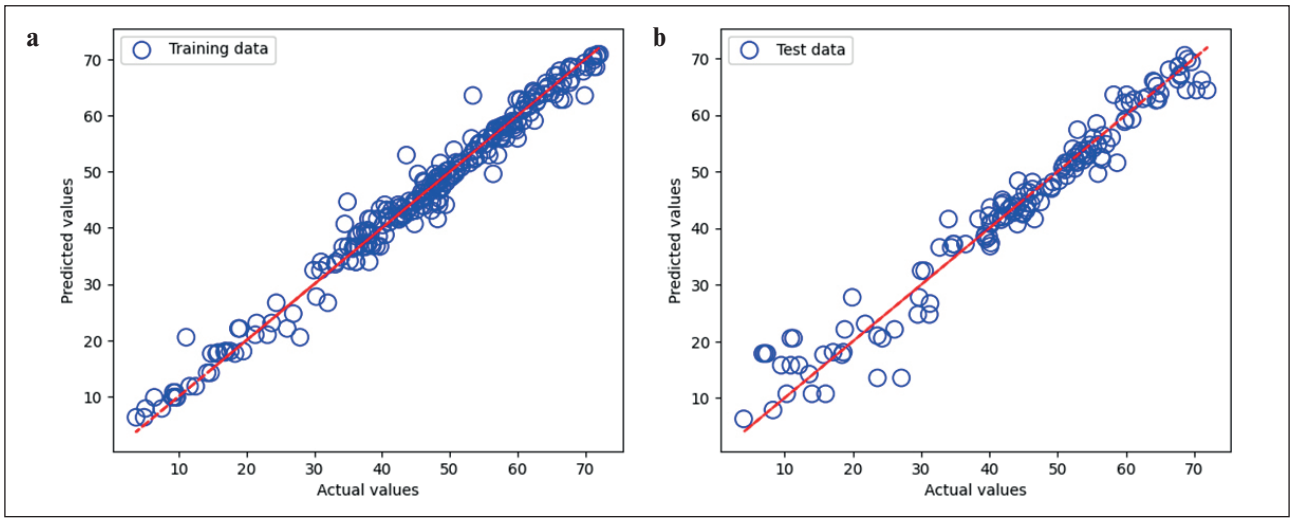

Fig. 7. **Efficiency of the gradient boosting algorithm on the training (a) and test (b) parts of the experimental data** (the red line indicates the model line)

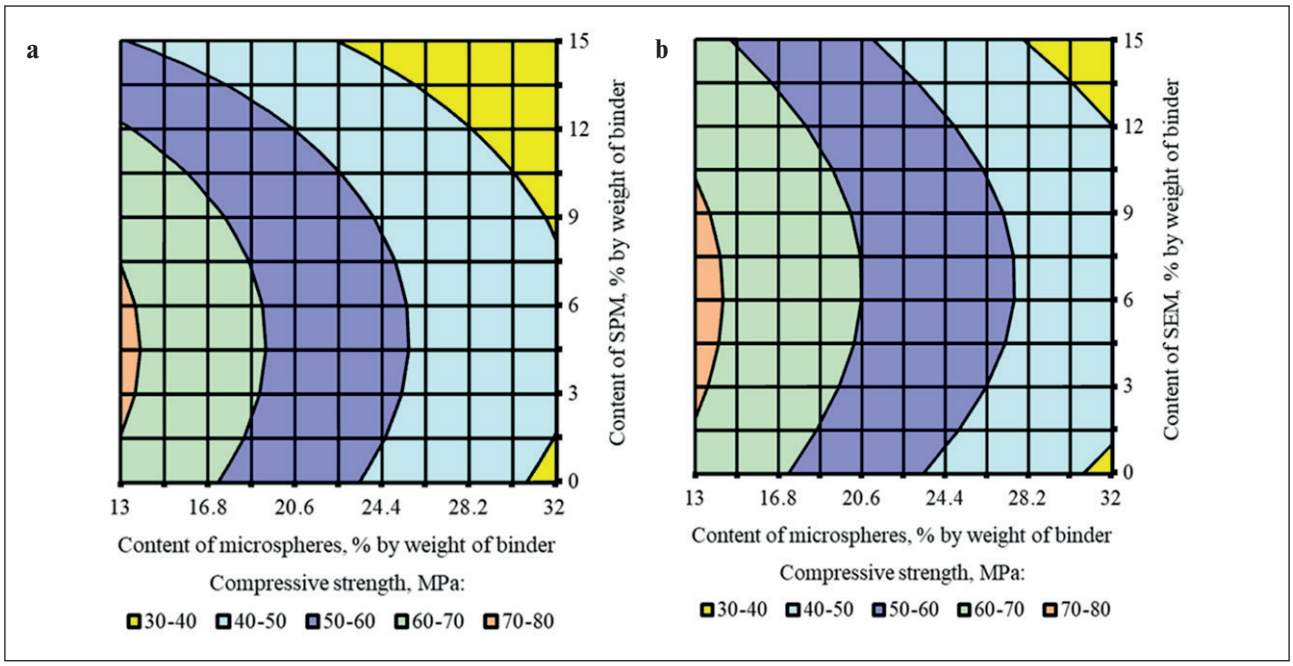

Fig. 8. **Isolines of changes in the compressive strength of lightweight concretes at the age of 28 days depending on the content of hollow glass microspheres, dosages of pozzolanic (a) and expanding (b) additives**

REFERENCES

1. Tayeh B.A., Akeed M.H., Qaidi S., Bakar B.H.A. Ultra-high-performance concrete: Impacts of steel fibre shape and content on flowability, compressive strength and modulus of rupture. *Case Studies in Construction Materials.* 2022; 17: e01615. https://doi.org/10.1016/j.cscm.2022.e01615

2. O'Hegarty R., Kinnane O., Newell J., West R. High performance, low carbon concrete for building cladding applications. *Journal of Building Engineering.* 2021; 43: 102566. https://doi.org/10.1016/j.jobe.2021.102566

3. Shin H.O., Yoo D.Y., Lee J.H., Lee S.H., Yoon Y.S. Optimized mix design for 180 MPa ultra-high-strength concrete. *Journal of Materials Research and Technology*. 2019; 8: 4182–4197. https://doi.org/10.1016/j.jmrt.2019.07.027

4. Nizina T.A., Balykov A.S., Korovkin D.I., Volodin V.V. Physical and mechanical properties of modified fine-grained fibre-reinforced concretes containing carbon nanostructures. *International Journal of Nanotechnology*. 2019; 16: 496–509. https://doi.org/10.1504/IJNT.2019.106621

5. Tolstoy A.D., Lesovik V.S., Zagorodnyuk L.Kh., Kovaleva I.A. Powder concretes with technogenic materials. *Proceedings of Moscow State University of Civil Engineering*. 2015; 11: 101–109.

6. Kaprielov S.S., Sheinfeld A.V., Kardumyan G.S., Chilin I.A. About selection of compositions of high-quality concretes with organic-mineral modifiers. *Construction Materials*. 2017; 12: 58–63.

7. Kalashnikov V.I. Evolution of Development of Concretes Compositions and Change in Concrete Strength. Concretes of Present and Future. Part 1. Change in Compositions and Strength of Concretes. *Construction Materials*. 2016; 1-2: 96–103.

8. Chernyshov E.M., Makeev A.I. On the Problem of Control of Prescription-Technological Factors of Concrete Production in the Course of Design and Synthesis of its Optimal Structure. *Academia. Architecture and Construction*. 2018; 3: 135–143. https://doi.org/10.22337/2077-9038-2018-3-135-143

9. Bazhenov Yu.M., Chernyshov E.M., Korotkikh D.N. Designing of modern concrete structures: determining principles and technological platforms. *Construction materials*. 2014; 3: 6–14.

10. Sifan M., Nagaratnam B., Thamboo J., Poologanathan K., Corradi M. Development and prospectives of lightweight high strength concrete using lightweight aggregates. *Construction and Building Materials*. 2023; 362: 129628. https://doi.org/10.1016/j.conbuildmat.2022.129628

11. Zhou Y., Gong G., Huang Y., Chen C., Huang D., Chen Z., Guo M. Feasibility of incorporating recycled fine aggregate in high performance green lightweight engineered cementitious composites. *Journal of Cleaner Production*. 2021; 280: 124445. https://doi.org/10.1016/j.jclepro.2020.124445

12. Wei H., Liu Y., Wu T., Liu X. Effect of Aggregate Size on Strength Characteristics of High Strength Lightweight Concrete. *Materials*. 2020; 13: 1314. https://doi.org/10.3390/ma13061314

13. Ye Y., Liu J., Zhang Z., Wang Z., Peng Q. Experimental study of high-strength steel fiber lightweight aggregate concrete on mechanical properties and toughness index. *Advances in Materials Science and Engineering*. 2020; 2020: 1–10. https://doi.org/10.1155/2020/5915034

14. Inozemtcev A.S., Korolev E.V., Smirnov V.A. Nanoscale modifier as an adhesive for hollow microspheres to increase the strength of high-strength lightweight concrete. *Structural Concrete*. 2017; 18: 67–74. https://doi.org/10.1002/suco.201500048

15. Yarmakovsky V.N. Multifunctional lightweight concretes for energy-saving industrial housing construction. *Construction materials*. 2012; 4: 4–12.

16. Inozemtsev A.S., Korolev E.V. Comparative analysis of the effect of nanomodification and microdisperse reinforcement on the process and parameters of destruction of high-strength lightweight concretes. *Construction materials*. 2017; 7: 11–15.

17. Kapeluszna E., Chrabąszcz K. Mutual compatibility of superplasticizers (PC, SNF), grinding aids (TEA, glycol) and C_3A in Portland cement systems – Hydration, rheology, physical properties and air void characteristics. *Construction and Building Materials*. 2023; 373: 130877. https://doi.org/10.1016/j.conbuildmat.2023.130877

18. Lin X., Pang H., Wei D., Lu M., Liao B. Effect of superplasticizers with different anchor groups on the properties of cementitious systems. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*. 2021; 630: 127207. https://doi.org/10.1016/j.colsurfa.2021.127207

19. Smirnova O.M. 2016 Compatibility of portland cement and polycarboxylate-based superplasticizers in high-strength concrete for precast constructions *Magazine of Civil Engineering*. 2016; 6: 12–22. https://doi.org/10.5862/MCE.66.2

20. Mota dos Santos A.A., Cordeiro G.C. Investigation of particle characteristics and enhancing the pozzolanic activity of diatomite by grinding. *Materials Chemistry and Physics*. 2021; 270: 124799. https://doi.org/10.1016/j.matchemphys.2021.124799

21. Taoukil D., El meski Y., Lahlaouti M.L., Djedjig R., El bouardi A. Effect of the use of diatomite as partial replacement of sand on thermal and mechanical properties of mortars. *Journal of Building Engineering*. 2021; 42: 103038. https://doi.org/10.1016/j.jobe.2021.103038

22. Balykov A.S., Nizina T.A., Kyashkin V.M., Volodin S.V. Prescription and technological efficiency of sedimentary rocks of various composition and genesis in cement systems. *Nanotechnologies in Construction*. 2022; 14(1): 53–61. https://doi.org/10.15828/2075-8545-2022-14-1-53-61

23. Kocak Y. Effects of metakaolin on the hydration development of Portland–composite cement. *Journal of Building Engineering*. 2020; 31: 101419. https://doi.org/10.1016/j.jobe.2020.101419

24. Gaifullin A.R., Rakhimov R.Z., Rakhimova N.R. The influence of clay additives in Portland cement on the compressive strength of the cement stone. *Magazine of Civil Engineering*. 2015; 7(59): 66–73. https://doi.org/10.5862/MCE.59.7

25. Balykov A.S., Nizina T.A., Volodin S.V. Optimization of technological parameters for obtaining mineral additives based on calcined clays and carbonate rocks for cement systems. *Nanotechnologies in Construction*. 2022; 14(2): 145–155. https://doi.org/10.15828/2075-8545-2022-14-2-145-155

26. Balykov A.S., Nizina T.A., Volodin V.V., Kyashkin V.M. Effects of Calcination Temperature and Time on the Physical-Chemical Efficiency of Thermally Activated Clays in Cement Systems. *Materials Science Forum*. 2021; 1017: 61–70. https://doi.org/10.4028/www.scientific.net/MSF.1017.61

27. Chand G., Happy S.K., Ram S. Assessment of the properties of sustainable concrete produced from quaternary blend of portland cement, glass powder, metakaolin and silica fume. *Cleaner Engineering and Technology*. 2021; 4: 100179. https://doi.org/10.1016/j.clet.2021.100179

28. Rassokhin A.S., Ponomarev A.N., Figovsky O.L. Silica fumes of different types for high-performance fine-grained concrete. *Magazine of Civil Engineering*. 2018; 78: 151–160. https://doi.org/10.18720/MCE.78.12

29. Ribeiro R.P., Jaramillo Nieves L.J., Bernardin A.M. Effect of fiberglass waste and fly ash addition on the mechanical performance of Portland cement paste. *Cleaner Materials*. 2023; 7: 100176. https://doi.org/10.1016/j.clema.2023.100176

30. Nedunuri S.S.S.A., Sertse S.G., Muhammad S. Microstructural study of Portland cement partially replaced with fly ash, ground granulated blast furnace slag and silica fume as determined by pozzolanic activity. *Construction and Building Materials*. 2020; 238: 117561. https://doi.org/10.1016/j.conbuildmat.2019.117561

31. Nontikansak M., Chaiyapoom P., Siriwatwechakul W., Jongvisuttisun P., Snguanyat C. Control the early-stage hydration of expansive additive from calcium sulfoaluminate clinker by polymer encapsulation. *Cement*. 2022; 8: 100021. https://doi.org/10.1016/j.cement.2022.100021

32. Carballosa P., García Calvo J.L., Revuelta D., Sánchez J.J., Gutiérrez J.P. Influence of cement and expansive additive types in the performance of self-stressing and self-compacting concretes for structural elements. *Construction and Building Materials*. 2015; 93: 223–229. https://doi.org/10.1016/j.conbuildmat.2015.05.113

33. Le Saoût G., Lothenbach B., Hori A., Higuchi T., Winnefeld F. Hydration of Portland cement with additions of calcium sulfoaluminates. *Cement and Concrete Research*. 2013; 43: 81–94. https://doi.org/10.1016/j.cemconres.2012.10.011

34. Bazhenov Yu.M., Falikman V.R., Bulgakov B.I. Nanomaterials and Nanotechnologies in the Present-day Concrete Technology. *Proceedings of Moscow State University of Civil Engineering*. 2012; 12: 125–133.

35. Falikman V.R., Sobolev K.G. «There's plenty of room at the bottom», or how nanotechnologies can change the world of concrete. Part 1. *Nanotechnologies in Construction*. 2010; 2(6): 17–31.

36. Nizina T.A., Ponomarev A.N., Balykov A.S., Korovkin D.I. Multicriteria optimization of the formulation of modified fine-grained fibre concretes containing carbon nanostructures. *International Journal of Nanotechnology*. 2018; 15: 333–346. https://doi.org/10.1504/IJNT.2018.094790

37. Lyashenko T.V. On neural networks and experimental and statistical modeling. In: *Modeling and optimization of building composites*. Odessa: Astroprint; 2016. p. 86–90.

38. Radchenko S.G. Analysis of methods for modeling complex systems. *Mathematical machines and systems*. 2015; 4: 123–127.

39. Shinkevich E.S., Chernyshov E.M., Lutskin E.S., Tymnyak A.B. Multicriterian optimization of composition and properties of activated lime-silica composites. *Dry building mixes*. 2013; 2: 33–37.

40. Jiang Y., Li H., Zhou Y. Compressive Strength Prediction of Fly Ash Concrete Using Machine Learning Techniques. *Buildings*. 2022; 12: 690. https://doi.org/10.3390/buildings12050690

41. Feng D.-C., Liu Z.-T., Wang X.-D., Chen Y., Chang J.-Q., Wei D.-F., Jiang Z.-M. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*. 2020; 230: 117000. https://doi.org/10.1016/j.conbuildmat.2019.117000

42. Ziolkowski P., Niedostatkiewicz M. Machine Learning Techniques in Concrete Mix Design. *Materials*. 2019; 12: 1256. https://doi.org/10.3390/ma12081256

43. Naderpour H., Rafiean A.H., Fakharian P. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. Journal of Building Engineering. 2018; 16: 213–219. https://doi.org/10.1016/j.jobe.2018.01.007

INFORMATION ABOUT THE AUTHORS

Artemy S. Balykov – Cand. Sci. (Eng.), Senior Researcher, Research Laboratory of Ecological and Meteorological Monitoring, Building Technologies and Expertises, National Research Mordovia State University, Saransk, Russia, artbalrun@yandex.ru, https://orcid.org/0000-0001-9087-1608

Elena A. Kaledina – Cand. Sci. (Eng.), Associate Professor, Department of Applied Mathematics, Differential Equations and Theoretical Mechanics, National Research Mordovia State University, Saransk, Russia, elena.lizina@gmail.com, https://orcid.org/0000-0002-0992-9540

Sergey V. Volodin – student, National Research Mordovia State University, Saransk, Russia, svolodin12@gmail.com, https://orcid.org/0000-0003-2018-4487

CONTRIBUTION OF THE AUTHORS

Artemy S. Balykov – scientific leadership; development of the concept and development of study methodology; analysis of research results; writing the original text of the article; drawing up final conclusions.

Elena A. Kaledina – processing and analysis of experimental data using machine learning methods; graphical and tabular presentation of research results; writing the original text of the article; drawing up final conclusions.

Sergey V. Volodin – literature review; conducting experimental work; collection and systematization of experimental data.

The authors declare no conflicts of interests.

The article was submitted 03.03.2023; approved after reviewing 31.03.2023; accepted for publication 05.04.2023.