Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2025, 19(10), 123–135
https://doi.org/10.12913/22998624/207915
ISSN 2299-8624, License CC-BY 4.0

Received: 2025.06.05
Accepted: 2025.08.15
Published: 2025.09.01

Comparison of machine learning models for predicting the compressive strength of cement mixtures with zeolite

Justyna Michaluk[1*], Monika Kulisz[1], Justyna Kujawska[2], Edyta Wojtaś[2], Aliya Aldungarova[3]

[1] Management Faculty, Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland
[2] Faculty of Environmental Engineering and Energy, Lublin University of Technology, Nadbystrzycka 40B, 20-618 Lublin, Poland
[3] International Educational Corporation LLP, Ryskulbekov Street 28, 050043 Almaty, Republic of Kazakhstan
* Corresponding author's e-mail: j.michaluk@pollub.pl

ABSTRACT

This study investigates the applicability of machine learning algorithms for predicting the compressive strength of cement mixtures with zeolite. The research compares the performance of four predictive models – Elastic Net regression, support vector machines (SVM), multilayer perceptron (MLP) neural networks, and Decision Trees – trained on experimentally obtained data describing mix composition and curing conditions. The input features included zeolite percentage, water-to-cementitious-material ratio, curing time, cement mass, and zeolite content. The output variable was compressive strength. Among the evaluated models, the SVM algorithm exhibited the optimal generalization capability, attaining the minimal prediction error on the validation set while sustaining elevated correlation between actual and predicted values. The MLP neural network demonstrated the optimal fit to the training data, however, this was achieved at the expense of heightened sensitivity to overfitting. Decision trees demonstrated robust training efficacy but exhibited diminished generalization capabilities, while the linear elastic net model encountered challenges in replicating the nonlinear characteristics of the material system. The study corroborates the viability of nonlinear machine learning models in facilitating the design and optimization of zeolite-enhanced cementitious mixtures. These findings signify a significant stride towards data-driven modeling in the field of construction materials engineering, thereby facilitating enhanced prediction of mechanical performance with minimized experimental effort. The study also underscores avenues for future exploration, encompassing model hybridization, multi-output prediction frameworks, and integration with optimization algorithms for automated mix design.

Keywords: machine learning, cement mixtures with zeolite, compressive strength, neural network, SVM, predictive modeling, mix design, sustainable materials.

INTRODUCTION

With increasing demands for durability, strength and sustainability in the construction industry, alternative materials are increasingly being sought that can partially replace traditional Portland cement. One such material is natural zeolites, which, due to their crystalline structure, high specific surface area and pozzolanic activity, have found use as additives to concrete mixtures. Their properties make it possible to improve the microstructure of the cement slurry, increase the strength of concrete and reduce its shrinkage and permeability, resulting in increased durability of the material [1–3]. Ensuring the availability and reliability of machinery is a key factor in achieving high efficiency and timely completion of construction projects. Modern approaches to maintenance increasingly rely on data-driven techniques and advanced analytics, with machine learning methods enabling not only better prediction of equipment failures but also supporting the optimization of material compositions and technological parameters. The integration of machine learning models into the design and maintenance offers the potential to reduce experimental costs, improve resource efficiency, and enhance the durability of construction materials [4].

Zeolites, especially their clinoptilolite variety, show the ability to react with calcium hydroxide during the hydration process of cement, leading to the secondary formation of C-S-H gels, resulting in improved mechanical and durability properties of concrete [5, 6].

Studies have also shown that a properly selected proportion of zeolite in a cement mix can lead to an increase in compressive strength, while reducing autogenous shrinkage and improving resistance to environmental factors [7].

One of the most promising materials used as additives to concrete mixtures are natural zeolites, which, due to their ordered crystalline structure, high specific surface area and content of reactive silicon and aluminum oxides, exhibit significant pozzolanic activity. Their structure allows them to absorb and gradually release water, which promotes hydration processes and acts as an internal conditioning agent. As a result of the reaction of zeolites with calcium hydroxide ($Ca(OH)_2$), formed during the hydration of cement, secondary C-S-H gels are formed, which seal the microstructure of the cement slurry, increase the density of the matrix and improve the mechanical properties of concrete, especially over a longer maturation period [8].

Compared to traditional mineral additives such as fly ash or blast furnace slag, zeolites have several significant advantages. First, they are natural materials and readily available in many regions of the world, making them a more predictable and sustainable raw material, independent of industrial changes such as the shift away from energy [9]. Zeolites also help improve concrete resistance to aggressive environments such as chlorides and sulfates and reduce autogenous shrinkage and improve resistance to freeze-thaw cycles [10].

However, the use of zeolites comes with some challenges. Their porous structure results in an increased water content of the mix, which can lead to reduced workability and requires the use of chemical admixtures to maintain adequate rheological properties [11]. In addition, the pozzolanic activity of zeolites is relatively slower than that of some synthetic additives, which can result in lower early strength and require a longer maturation period. Differences in mineralogical composition due to local geological conditions also make it difficult to standardize their use in concrete mixtures [12].

Given the complexity of the hydration mechanisms of cement with zeolite and the variability in material properties depending on local geological sources, classical empirical approaches are increasingly being supplemented – or even replaced – by predictive modeling techniques. These methods allow for a reliable evaluation of concrete performance without the need for time-consuming experimental procedures, especially in the case of novel additives such as natural zeolites.

One such algorithm is Elastic Net, a regularized regression algorithm that combines the features of Lasso and Ridge regression, effectively addressing the problem of variable collinearity. In the context of concrete strength prediction, it allows for simultaneous feature selection and model stabilization, offering good prediction accuracy even with limited data. This model is highly resistant to overfitting and works well for modeling nonlinear relationships in engineering data [13]. In studies on concrete with rubber additives, Elastic Net has been shown to offer balanced predictive performance while maintaining model stability even in the case of strongly correlated parameters such as rubber, aggregate, and cement content. Analyses have shown that Elastic Net effectively reduces the problem of overfitting, achieving lower prediction errors (RMSE and MAE) than classical multiple regression. Thanks to its flexibility and generalization ability, Elastic Net is an attractive tool for modeling the mechanical properties of concrete mixtures containing rubber waste [14].

Another group of popular methods are support vector machines (SVM), which allow for the mapping of complex nonlinear relationships in data. A special variant of these is support vector regression (SVR), which has been used in concrete analysis using non-destructive testing. Li and Zhang (2024) presented an alternative approach to predicting concrete compressive strength, focusing on SVR and non-destructive testing (NDT) techniques instead of artificial neural networks (ANN). The authors used SVR as an analytical tool to model the relationship between non-destructive test results and actual concrete strength. Two well-known NDT methods were used: the Schmidt hammer test, which assesses mechanical rebound, and ultrasonic pulse velocity

(UPV) measurement, which assesses the homogeneity and quality of the material by analyzing the time it takes for the wave to travel through the concrete. The results suggest that integrating SVR with NDT test data allows for effective and non-destructive prediction of concrete strength, making this method particularly useful in assessing the technical condition of existing structures [15]. The goal of SVR is to find a function that best fits the data, allowing for a small deviation (ε) from the actual values. The algorithm maximizes the distance (the so-called margin) between the data points and the predicted line, ignoring minor errors. To obtain the best results, the authors tuned the model parameters (gamma = 0.6 and C = 33), which enabled effective prediction of concrete strength, even for non-linear data, precisely through the use of kernel functions. [16]

In the field of civil engineering, methods such as SVM can support the classification of the technical condition of building components (e.g., concrete cracks, installation defects) based on limited sensory data, enabling faster and more efficient diagnosis with a minimum of field measurements [17].

It is worth noting that the use of models such as linear regression, SVM, decision trees, and random forests allows for the capture of nonlinear relationships between material variables and concrete strength [18]. The SVM formula reflects the principle of structural risk minimization (SRM), which has proven to be superior to the more traditional principle of empirical risk minimization (ERM) used in many other modeling techniques. SRM imposes an upper bound on the expected risk, unlike ERM, which minimizes error only in the training data. It is this difference that gives SVM greater generalization ability compared to traditional neural network methods [19].

Among the machine learning techniques used to predict the properties of building materials, the decision tree (DT) algorithm plays an important role. It is an intuitive and effective classification method that generates decision rules by analyzing input data and systematically dividing the feature space based on attribute values. This process involves building a tree based on training data and pruning it to prevent overfitting the model to the data. Due to their simplicity, high accuracy, and resistance to noise in the data, decision trees are very popular in concrete analysis, allowing for a quick and transparent determination of the influence of individual components of the mixture on its final compressive strength [20]. Decision trees can effectively support the classification of high-performance concrete mix design methods. They allow design techniques to be distinguished based on key parameters such as water-cement ratio or type of additives. Thanks to their transparent structure, they are easy to interpret and useful in engineering practice. This approach promotes automation and increases the efficiency of the mix selection process [21]. In agricultural machinery, decision trees outperformed ANN and SVM in predicting PTO shaft power demand, achieving nearly 99% accuracy [22]. Machine learning is increasingly being applied to the analysis and forecasting of complex engineering processes, including areas beyond construction materials. For instance, in the prediction of wastewater inflow to the treatment plant in Rzeszów, over 1,000 models were evaluated, including neural networks, k-nearest neighbors, and advanced statistical approaches such as ARIMA and SARIMAX. The results of these studies confirm that the careful selection of machine learning models and parameters enables high prediction accuracy, even in complex environmental processes influenced by multiple variables, such as weather conditions [23]. These examples underline the universality of machine learning methods in solving complex diagnostic and predictive tasks, further supporting their application in modeling cementitious materials.

It is in the face of these difficulties that predictive methods, especially artificial neural networks (ANNs), which make it possible to create advanced predictive models from experimental data, are becoming particularly important. Neural networks, due to their ability to model complex and nonlinear relationships, are particularly useful in data analysis where the connections between variables are intricate; however, a significant limitation of this approach is the risk of overfitting, especially when working with a small number of observations, which necessitates careful selection of parameters and model architecture. [24] Networks such as CNN (convolutional neural network) or LSTM (deep recurrent network with short-term memory) can contribute tothe detection of errors in construction [25]. The main purpose of their use is to create reliable predictive models that allow the prediction of parameters such as compressive strength, tensile strength, flexural strength, thermal conductivity, shrinkage or durability, without the need for expensive and time-consuming experimental tests.

Artificial neural networks are widely used in the construction industry, especially in the areas of energy management, cost prediction and user comfort assessment. They make it possible to process complex environmental and technological data to support decision-making on projects and operations. The article highlights the growing role of ANNs in the automation of construction processes and the need for further research into their integration with other smart technologies [26].

The application of ANNs is based on the ability of these models to map complex, nonlinear relationships between multiple input variables (e.g., concrete mix composition, curing time, mineral additive content, care conditions) and output variables (e.g., strength or thermal conductivity). This makes it possible not only to quickly predict the properties of cementitious composites, but also to optimize their composition and technological parameters. Models of this type learn from experimental data and, when properly trained, are able to generate accurate predictions even for data that were not used in the learning process. Compared to classical testing methods, ANNs offer a number of advantages. First of all, they allow a significant reduction in the time needed to evaluate material properties, which is particularly important for analyses that require a long period of sample maturation (e.g., 28-day strength). Eliminating the need for numerous laboratory tests also reduces costs and material consumption. In addition, predictive models enable what-if simulations to support the process of designing concrete mixtures with specific target properties, as well as to quickly assess the impact of changing a single parameter (such as the amount of zeolite) on the final characteristics of the material.

In recent years, ANNs have received particular attention, showing high performance in modeling nonlinear relationships between input variables and output properties of construction materials [27]. Siddique et al. (2011) developed ANN models for predicting the compressive strength of self-compacting concrete containing bottom ash as a partial replacement for sand, using both literature and experimental data, which made it possible to evaluate the effectiveness of the models in predicting the compressive strength of concrete at different stages of maturation All the obtained models had correlation coefficient values above 0.9 [28].

Lin and Wu (2021) used seven input parameters: weight of water, cement, fine aggregate, coarse aggregate, blast furnace slag, fly ash, and superplasticizer, normalizing the data and using an ANN network with one hidden layer, the resulting R^2 determination coefficient exceeded 0.98 [29].

Amar et al. (2022) used the ANN model to predict the strength of concretes with waste additives. They used 18 input variables to create models, including the type of cement, water content, water/soil ratio, degree of cement replacement by mineral additives (such as metakaolin, fly ash, slag, marble dust, ceramic waste), amount of superplasticizer and others, obtaining very high agreement between predicted and experimental values (R^2 = 0.9888, MAPE = 2.87%) [30].

The potential of ANN was also confirmed by Başyiğit et al. who used four inputs in predicting the strength of heavy concrete with barite addition: the amount of cement, the amount of water, the curing time (7, 28, 90 days) and the percentage of barite, achieving high prediction agreement using both ANN and fuzzy logic (FL). This confirms the effectiveness of creating models even for specialized types of cement mixtures [31].

In the context of trends related to the digitization of construction and the development of the concept of so-called "smart concrete," tools such as ANN are becoming not just an alternative, but often a necessity in the modern design of construction materials. This becomes especially important in the case of materials with new-generation additives, such as zeolite, for which full, standardized empirical data is still lacking. ANN makes it possible to analyze them quickly and implement them in engineering practice.

Recent research also confirms the effectiveness of deep and recurrent neural networks in other engineering applications, such as defect detection in aluminum casting using process data [32] or moisture imaging in historical brick walls using LSTM networks and electrical impedance tomography [33].

In view of the growing impact of the cement industry on global CO_2 emissions, there is a need to develop methods to reduce cement consumption while maintaining structural requirements. One innovative approach is to predict the cement content needed to achieve the nominal compressive strength of concrete after 90 days, instead of the traditional 28 days. A longer curing period may better reflect the actual load conditions in multistory structures and also allows for the optimization of concrete composition in terms of emission reduction. To this end, a number of machine learning algorithms were used, such as decision tree regression, extra trees, random forest, and elastic net. Although the authors also mention the use of methods such as ANN and SVM in the literature, particularly when modeling the properties of concrete with additives such as fly ash, slag, or recycled aggregates – due to the moderate complexity of the data used in the study and the lack of alternative additives, the focus was on simpler regression models. SVM, as an algorithm capable of modeling nonlinear relationships, can be an effective tool in predicting concrete strength, especially when data characterized by high variability and complex structure are available. However, in the context of a limited data set and a focus on optimizing a single component – cement – the authors of the study opted for linear regression with regularization (Elastic Net), achieving high prediction accuracy ($R^2 \approx 0.94$) without the risk of overfitting the model [34].

The novelty of this research lies in its targeted application of machine learning algorithms to cement mixtures with zeolite, which represent a new generation of eco-efficient construction materials. Unlike prior studies that often investigate multiple secondary additives or rich composite formulations, this work isolates zeolite as the sole additive, enabling a focused assessment of its impact on compressive strength prediction. Considering the above, the objective of this study is to conduct a comparative analysis of the effectiveness of selected machine learning algorithms in predicting the compressive strength of cement mixtures with zeolite. The study focuses on evaluating the accuracy and applicability of four models: Elastic Net, SVM, neural networks, and Decision Trees, using experimental data related to the composition of cement mixtures and the curing conditions of the samples.

MATERIALS AND METHODS

Cement-zeolite mixtures

Portland cement (class 42.5 N) and zeolite, clinopliolite were used to prepare the cement mixtures. The chemical properties of the cement and zeolite are shown in Table 1, the physicochemical properties are shown in Table 2. The water used to make the mixtures met the quality standards of PN-EN 1008:2004.

The mixtures created were labeled according to the MIX-X-Y scheme, where X corresponds to the percentage of zeolite additive to the weight of the cement (in the range of 0% to 20%), while Y denotes the value of the water/cement ratio (W/CM) ranging from 0.45 to 0.70.

Testing of cementitious mixtures for compressive strength

The specimens for compressive strength testing were in the form of cubes with dimensions of 70.7 × 70.7 × 70.7 mm. They were made from a cement slurry containing different proportions of cement and zeolite. Nine specimens were prepared for each mix, to be tested after 7, 28 and 70 days of maturation, three specimens for each term.

The sample preparation process involved several steps. First, fresh slurry was poured into cubic molds, which were then placed on a vibrating table to compact the material and remove air bubbles. After the top surface of the molds were leveled, they were covered with plastic film to reduce water evaporation. The samples remained in the molds for the first 24 hours at 24 ± 2 °C. This was followed by unmolding and transferring the samples to a chamber with controlled ripening conditions (20 ± 2 °C, relative humidity above 95%), where they were stored until tested.

The strength was measured in accordance with EN 12390-3:2011, by compressing the specimens in a laboratory press. The specimens were loaded until failure, and the maximum failure force obtained was divided by the cross-sectional area of the specimen to calculate the compressive strength in MPa. The specimens were subjected to compressive strength testing after 7, 28 and 70 days of maturation. The result for each mix and maturation period was the average value of the three samples.

Machine learning methods methodology

In this study, four predictive models were developed to estimate the compressive strength of cement mixtures with zeolite. The selection of these methods was deliberate, with each representing a distinct class of machine learning algorithms and offering different computational and statistical perspectives for analyzing the problem. The models that were implemented included Elastic Net, SVM, MLP, neural networks, and Decision Trees.

The Elastic Net regression was selected as a regularized linear model that combines the strengths of Ridge (L2) and Lasso (L1) regression, thereby allowing for simultaneous variable selection and complexity control. It functions as a robust baseline model with high interpretability, which is especially useful when multicollinearity among predictors is anticipated. SVMs were selected for their effectiveness in modeling nonlinear relationships, particularly in high-dimensional spaces and with small to medium-sized datasets. The utilization of kernel functions enables SVMs to effectively capture intricate boundaries and facilitate robust generalization. MLP neural networks were incorporated due to their capacity to learn nonlinear mappings between inputs and outputs through hidden layers of interconnected neurons. In the domain of composite materials, where the interactions between mix parameters may be intricate and multidimensional, MLPs are particularly well-suited for implementation. Decision trees were implemented to generate interpretable models capable of unveiling hierarchical relationships between predictors and the target variable. The tree-based structure of these models enables the extraction of clear rules, thereby facilitating insight into the influence of individual input features. This set of models enables not only the comparison of predictive accuracy but also an exploration of trade-offs between complexity, interpretability, training time, and generalization ability in the context of data-driven modeling of cementitious materials enhanced with natural zeolite.

The implementation and training of all models was conducted within the MATLAB R2024b environment, leveraging the built-in functions of the Statistics and Machine Learning Toolbox and the Deep Learning Toolbox. The modeling process was based on an experimental dataset, where the input variables (predictors) included the percentage of zeolite (%), the water-to-cementitious-material ratio (W/CM), curing period (days), cement mass (g), and the amount of superfine zeolite (SFZ, g). The output variable (response) was the compressive strength of cement mixtures with zeolite. (MPa), which was labeled in the dataset as "Strength (MPa)." A total of 90 different experimental observations were used, covering diverse combinations of input variable values.

All four models were trained using a systematic hyperparameter tuning procedure based on grid search combined with 5-fold cross-validation. For each algorithm, a predefined range of key hyperparameters was explored in order to identify the combination that minimized the mean squared error (MSE) during validation. Various configurations were assessed to ensure both robustness and generalization of the models. The list of parameters and ranges considered for each model is summarized in Table 3.

In order to assess generalization performance and ensure fairness in comparison, all models were evaluated using consistent 5-fold cross-validation. For each fold, models were trained on the training set and tested on the validation fold to produce out-of-fold predictions. These were then compared with actual values to estimate unbiased performance.

The following performance metrics were calculated:

- Mean squared error (MSE) – overall prediction error magnitude,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i{}^{*} - y_i{}')^2 \tag{1}$$

- Mean absolute error (MAE) – average prediction deviation,

Table 1. Chemical properties of the additives used in the study

Parameter	SiO_2 [%]	Al_2O_3 [%]	Fe_2O_3 [%]	CaO [%]	K_2O [%]
Cement	20.65	6.54	3.46	64.6	0.99
Clinopliolit	72.40	12.09	0.85	2.2	2.56

Table 2. Physicochemical properties

Properties	Cement	Properties	Clinopliolite
Specific surface (cm^2/g)	3650	Surface density, kg/m^3	1735
Bulk density, kg/m^3	1145	Porosity, %	27
Particle density, kg/m^3	3150	Relative weight, kg/m^3	1632

Table 3. Hyperparameters and parameter ranges analyzed for each model

Parameter	Hyperparameters considered	Parameter range explored
Elastic Net	Alpha (L1/L2 mixing) Lambda (regularization)	α ∈ {0.1, 0.5, 0.9} λ ∈ {0.01, 0.1, 1}
SVM	Kernel type BoxConstraint	Kernels: linear, RBF, polynomial C ∈ {0.1, 1, 10}
Neural network	Hidden layer size Training algorithm	Neurons 10÷20 Algorithms: Levenberg–Marquardt (trainlm), Scaled Conjugate Gradient (trainscg)
Decision tree	MaxNumSplits MinLeafSize	MaxNumSplits ∈ {5, 10, 20} MinLeafSize ∈ {1, 5, 10}

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i{}^* - y_i{}'| \quad (2)$$

- Root mean squared error (RMSE) – error in physical units,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i{}^* - y_i{}')^2} \quad (3)$$

Pearson correlation coefficient (R) – strength of correlation between predicted and actual values.

$$R(y', y^*) = \frac{cov(y', y^*)}{\sigma_{y'}\sigma_{y^*}} \; R\epsilon < 0.1 > \quad (4)$$

where: n – number of observations in the dataset, yi^* – actual (measured) value of the response variable for the ii-th observation, yi' – predicted value of the response variable for the ii-th observation, $cov(y', y^*)$ – covariance between predicted and actual values, $\sigma y'$, σy^*– standard deviations of the predicted and actual values, respectively.

The metrics were computed separately for the entire dataset, the training folds, and the validation folds. In order to support visual analysis, regression plots were generated to illustrate model fit across these sets.

To mitigate the issue of overfitting, all models underwent a process of hyperparameter tuning. This process involved the implementation of grid search combined with cross-validation. The implementation of cross-validation ensured that optimal configurations were selected based on their performance on unseen data rather than on the training set. Furthermore, regularization techniques were employed in applicable cases, including Elastic Net regression and SVM box constraints. Additionally, model complexity was explicitly controlled through architectural constraints, such as the number of hidden neurons in MLP models or the maximum tree depth and leaf size in Decision Trees. The implementation of these strategies resulted in a collective enhancement of the models' capacity to generalize to novel, previously unobserved input combinations.

RESULT AND DISCUSSION

Figure 1 shows the compressive strength values of cement-zeolite mixtures. Mixtures containing 5% and 10% zeolite, at low W/CM ratios of 0.45 and 0.5, achieved higher strength values than mixtures without the mineral additive. For higher zeolite additions, above 10%, lower compressive strength values were observed for both 7-day and 70-day strength tests. Mixtures with 15% and 20% zeolite had significantly lower strength than those with 5% and 10% additive.

The water/cement ratio shows a strong influence on the strength of mixtures. For all analyzed mixes, a relationship is evident: the lower the W/CM ratio, the higher the compressive strength. Mixtures with W/CM = 0.45 showed the highest strength, regardless of the content of added zeolite, while mixtures with W/CM = 0.7 showed the lowest strength parameters.

The increase in strength between days 28 and 70 is more pronounced in mixtures containing zeolite than in those without it, confirming the activation of zeolite's pozzolanic properties during the later maturation period. The best mechanical properties were shown by Mix-5-0.45 and Mix-10-0.45, whose strength after 70 days exceeded 70 MPa. In contrast, the lowest values were obtained for Mix-0-0.7 and Mix-20-0.7, confirming the adverse effects of both high water/cement

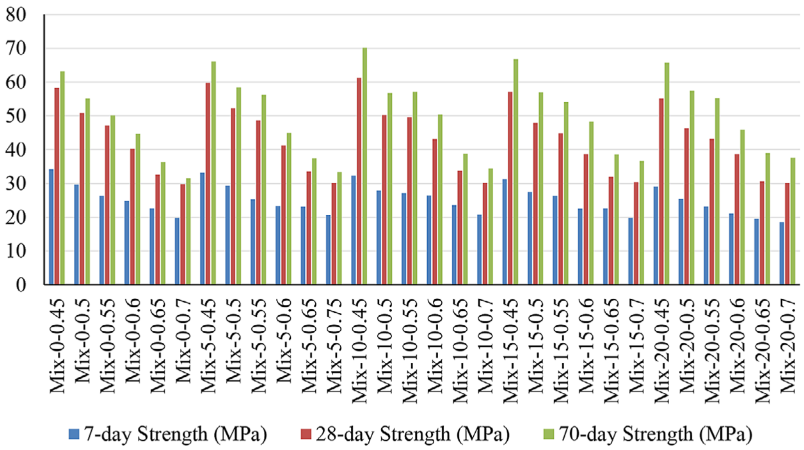

Figure 1. Compressive strength values of cement-zeolite mixtures

ratio and excessive zeolite. These results suggest that optimal concrete strength parameters can be obtained with low W/CM and moderate zeolite addition, preferably in the range of 5–10%, Vaiciukyniene et al. (2012) noted that zeolite has a beneficial effect on the compressive strength of hardened cement paste, which is due to its pozzolanic activity and the formation of hydroaluminate phases in the structure of the hardened material [35].

In terms of predictive performance, the Elastic Net model yielded the most favorable results with alpha set to 0.9 and lambda to 0.1.The high alpha value (approximately 1) guided the regularization strategy toward Lasso regression, promoting sparsity by selecting only the most pertinent input variables. The moderate lambda value applied a mild penalty on the coefficients, achieving a balance between feature selection and model flexibility. This configuration reflects a trade-off between generalization and model simplicity.

The optimal SVM configuration was achieved using a polynomial kernel and a BoxConstraint of 10. The polynomial kernel facilitated the model's ability to capture nonlinear relationships in the data, while the relatively stringent regularization imposed by a high box constraint prevented overfitting of the model. This combination has proven to be highly effective for structured datasets with complex variable interactions.

The neural network that demonstrated the highest level of performance incorporated 10 neurons within the hidden layer, in conjunction with the Levenberg–Marquardt training algorithm (trainlm). The modest network size guaranteed effective learning without over-parameterization. Concurrently, the training algorithm facilitated expeditious convergence and superior precision, rendering it particularly well-suited for medium-sized regression tasks. This configuration enabled the model to capture nonlinear patterns while preserving numerical stability.

The construction of the optimal regression tree was achieved through the implementation of MaxNumSplits = 20 and MinLeafSize = 1. This configuration enabled the tree to develop in a profound and highly granular manner, ensuring a close fit with the training data by facilitating numerous branch divisions and minimal pruning. While this increased the risk of overfitting, it also enabled precise modeling of local variations within the dataset.

The subsequent Table 4 provides a succinct synopsis of the primary evaluation metrics for each model, meticulously delineated for the training and validation sets.

Comparing the four predictive models—Elastic Net, SVM, MLP neural networks, and Decision Trees – based on MSE, MAE, RMSE, and correlation coefficient R, clear differences emerge in their effectiveness for predicting the compressive strength of cement mixtures with zeolite. The Elastic Net model demonstrated the least effective performance, exhibiting high MSE values of 33.46 in the training set and 38.47 in the validation set, as well as the lowest correlation coefficient R value of 0.91 in the training set and 0.89 in the validation set. These findings suggest that a linear model is incapable of effectively capturing the nonlinear relationships present in the data. The SVM model consistently yielded commendable outcomes, evidenced by its MSE of 2.59 on the training set and 4.24 on the validation set. This is further substantiated by the high R values

Table 4. Comparative summary of model performance

Quality indicators/model		Elastic Net	SVM	Neural network	Decision tree
MSE	Train	33.4591	2.5865	1.6864	2.6021
	Validation	38.4747	4.2411	6.3288	10.1458
	All dataset	34.5213	2.6752	2.7835	4.7321
MAE	Train	4.9142	1.3703	0.9497	1.2827
	Validation	5.2226	1.6576	2.0163	2.6153
	All dataset	5.0251	1.4231	1.4357	1.7432
RMSE	Train	5.7843	1.6082	1.2986	1.6131
	Validation	6.2027	2.0593	2.5157	3.1852
	All dataset	5.8935	1.7873	1.7835	1.9982
R	Train	0.9055	0.9932	0.9955	0.9929
	Validation	0.8903	0.9888	0.9839	0.9723
	All dataset	0.8987	0.9899	0.9899	0.9899

of 0.99 for both the training and validation sets, which underscores the model's capacity for effective balancing of accuracy and generalization. The MLP neural network exhibited excellent performance during training, with an MSE of 1.69 and a correlation coefficient of 0.996, demonstrating its strong modeling capacity. However, slightly reduced performance on the validation set (MSE of 6.33, R of 0.98) indicates slightly diminished generalization ability, although the model still maintains a high level of predictive accuracy. The Decision Tree model performed well on the training set (MSE = 2.60, R = 0.99) and demonstrated acceptable performance on the validation set (MSE = 10.15, R = 0.97). While the difference between training and validation metrics is more pronounced than in the case of other models, this outcome is consistent with the known characteristics of decision tree algorithms, which tend to closely fit training data. Despite the application of pruning and hyperparameter tuning to control model complexity, the tree-based structure may exhibit limited flexibility in capturing broader patterns across the entire input space. Nevertheless, the validation results confirm that the model retains a good level of predictive capability.

To further illustrate the predictive performance of each model, regression plots were generated for both the training and validation datasets. These correlation plots display the relationship between actual and predicted compressive strength values, providing a visual assessment of model fit and generalization capacity across the entire dataset (Figure 2).

The comparative analysis presented in this study clearly demonstrates the potential of machine learning models in predicting the compressive strength of cement mixtures with zeolite. Among the four tested algorithms, the SVM emerged as the most balanced model, combining high prediction accuracy with strong generalization. The MLP neural network also performed very well, showing excellent fit to the training data and maintaining high accuracy on validation. Decision Trees, while interpretable and well-aligned with training data, showed slightly reduced performance on validation. The Elastic Net model, though transparent and computationally efficient, was not well-suited to capturing the nonlinear relationships present in the dataset.

Regression plots for both training and validation sets confirmed the superior predictive performance of the SVM and MLP models. These findings highlight the effectiveness of nonlinear approaches in modeling the behavior of cement mixtures with zeolite, which are influenced by complex physical and chemical interactions.

The encouraging outcomes observed in this investigation align with those documented by Covatariu et al. (2024), who employed an ANN to forecast the thermal conductivity of cement-based mortars incorporating natural zeolites. The model was configured with a single hidden layer containing 12 neurons and trained using the Levenberg–Marquardt algorithm. This configuration achieved a correlation coefficient of approximately 0.95 between experimental and predicted values. This substantial degree of agreement confirmed the capacity of MLP networks to model the complex physical properties of cementitious materials, particularly those influenced by porosity, moisture, and compositional heterogeneity.

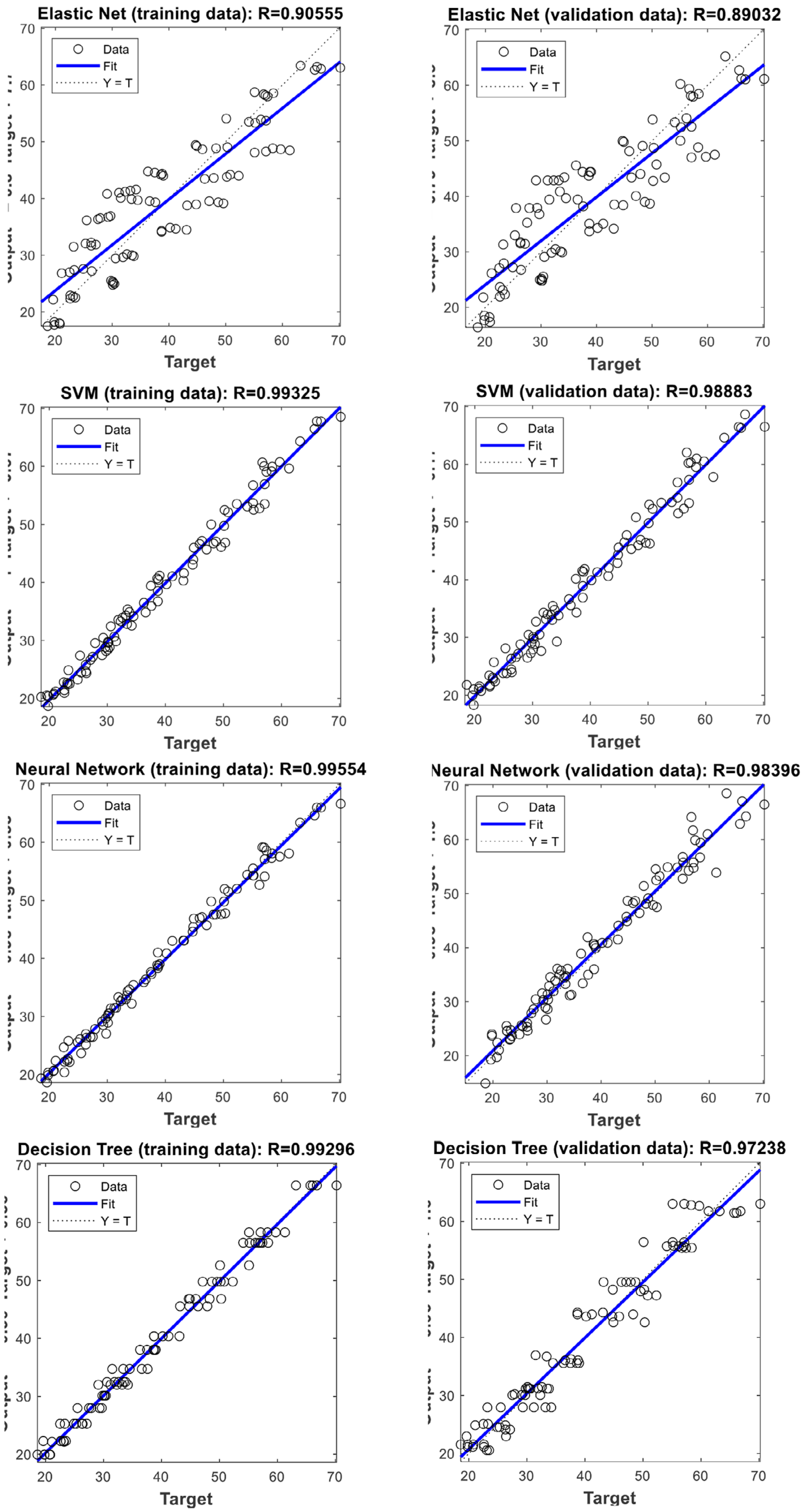

Figure 2. Regression plots showing the correlation between predicted and actual compressive strength of cement mixtures with zeolite values for all four models (Elastic Net, SVM, Neural Network, Decision Tree) for the training and validation datasets

These characteristics are also present in cement mixtures with zeolite. Covatariu et al. also emphasized that the most influential input in their model was bulk density, while moisture content and temperature had comparatively lower predictive importance. This finding highlights the value of feature selection and domain-specific knowledge in neural network modeling [36].

A parallel trend is observed in the study by Onyari and Ikotun (2018), who developed an artificial neural network to forecast both compressive and flexural strengths of mortars containing a modified zeolite additive (MZA). The input parameters encompassed conventional mix design variables, including cement, water, silica sand, and MZA content, as well as curing duration and loading level. The resulting ANN exhibited a three-layer structure and was trained for multiple load scenarios. The model demonstrated remarkable precision, attaining R^2 values more than 0.99 for flexural strength and ranging between 0.99 and 0.67 for compressive strength, contingent upon the applied loading level. These results demonstrate the predictive power of ANN models and their adaptability to different mechanical test conditions [37].

Considering these studies, the present work further reinforces the usefulness of artificial neural networks for property prediction in zeolite-based cementitious systems. However, the current results also highlight those alternative nonlinear methods, such as SVMs, may offer comparable – if not superior – performance when properly tuned. This suggests that hybrid or ensemble approaches could be considered in future research to leverage the complementary strengths of different algorithms.

From an application standpoint, the successful use of machine learning techniques in this study opens the way toward data-driven optimization of cement mix design. Especially in cases involving supplementary cementitious materials like natural zeolites, such tools can reduce reliance on time-consuming and costly experimental campaigns by offering reliable, simulation-based estimates of key mechanical properties.

CONCLUSIONS

The main objective of this study was to develop and assess predictive models capable of estimating the compressive strength of cement mixtures with zeolite. The motivation for this approach stems from the recognition that traditional empirical formulas are often insufficient to reflect the nonlinear and multidimensional relationships between mixed composition, curing time, and mechanical performance. Machine learning techniques were therefore applied to model these complex dependencies and support data-driven decision-making in mixed design.

The predictive models were developed using an experimental dataset incorporating a variety of cement mixtures with zeolite. The models were constructed based on five input variables: zeolite content expressed as a percentage of cement replacement, the water-to-cementitious-material ratio (W/CM), the curing time in days, the mass of cement in grams, and the amount of superfine zeolite added to the mix. These input parameters were chosen due to their known influence on hydration kinetics, microstructural development, and the strength evolution of cementitious composites. The output variable in all models was the compressive strength of the cement mixtures with zeolite, expressed in megapascals (MPa), which serves as the primary indicator of structural performance.

This study has demonstrated the applicability and effectiveness of selected machine learning algorithms in predicting the compressive strength of cement mixtures with zeolite. By comparing four distinct models – Elastic Net, SVM, MLP neural networks, and Decision Trees – rained on real experimental data, it was possible to evaluate their predictive accuracy and ability to generalize to unseen input combinations.

The results show that the SVM model offered the most balanced performance, with a training MSE of 2.59, validation MSE of 4.24, and high correlation coefficients (R = 0.99 for both training and validation sets). The MLP neural network achieved the best fit on the training set (MSE = 1.69, R = 0.996), and maintained good validation performance (MSE = 6.33, R = 0.98), confirming its strong predictive capacity. The Decision Tree model also performed well on the training data (MSE = 2.60, R = 0.99), but showed reduced accuracy in validation (MSE = 10.15, R = 0.97), consistent with the typical behavior of tree-based models. The Elastic Net regression, as a linear model, exhibited the weakest performance overall, with training MSE = 33.46, validation MSE = 38.47, and R values of 0.91 and 0.89, respectively, confirming its limited ability to represent the nonlinear relationships present in the dataset.

Nevertheless, the study is subject to certain limitations. The dataset was limited in both size and diversity, focusing on a specific type of natural zeolite and a relatively narrow range of composition and curing variables. Additionally, the modeling targeted compressive strength alone, without incorporating other relevant properties such as flexural strength, durability, or workability. While 5-fold cross-validation was used to address overfitting, external validation on independent datasets would be needed to assess the models' transferability to broader contexts.

REFERENCES

1. Alexa-Stratulat S.-M., Olteanu I., Toma A.-M., Pastia C., Banu O.-M., Corbu O.-C. and Toma I.-O. The use of natural zeolites in cement-based construction materials – a state of the art review. Coatings. 2024; 14: 18. https:// doi.org/10.3390/coatings14010018
2. Kocak Y., Tascı E. and Kaya U. The effect of using natural zeolite on the properties and hydration characteristics of blended cements. Constr. Build. Mater. 2013; 47: 720–727. https://doi.org/10.1016/j.conbuildmat.2013.05.033
3. Nagrockiene D. and Girskas G. Research into the properties of concrete modified with natural zeolite addition. Constr. Build. Mater. 2016; 113: 964–969. https://doi.org/10.1016/j.conbuildmat.2016.03.133
4. Antosz, K., Kaczmarek, M. J., Machado, J., Relich, M. Application of Principle Component Analysis and logistic regression to support Six Sigma implementation in maintenance. Maintenance & Reliability/Eksploatacja i Niezawodność, 2023; 25(4). https://ein.org.pl/pdf-174603-95997?filename=Application%20of%20Principle.pdf
5. Chen J.J., Li L.G., Ng P.L. and Kwan A.K.H. Effects of superfine zeolite on strength, flowability and cohesiveness of cementitious paste. Cem. Concr. Compos. 2017; 83: 101–110. https://doi.org/10.1016/j.cemconcomp.2017.06.010
6. Valipour M., Yekkalar M., Shekarchi M. and Panahi S. Environmental assessment of green concrete containing natural zeolite on the global warming index in marine environments. J. Clean. Prod. 2014; 65: 418–423. https://doi.org/10.1016/j.jclepro.2013.07.055
7. Zhang J., Wang Q. and Zhang J. Shrinkage of internal cured high strength engineered cementitious composite with pre-wetted sand-like zeolite. Constr. Build. Mater. 2017; 134: 664–672. https://doi.org/10.1016/j.conbuildmat.2016.12.182
8. Kocak Y., Tascı E. and Kaya U. The effect of using natural zeolite on the properties and hydration characteristics of blended cements. Constr. Build. Mater. 2013; 47: 720–727.
9. Valipour M., Yekkalar M, Shekarchi M. and Panahi S. Environmental assessment of green concrete containing natural zeolite on the global warming index in marine environments. J. Clean. Prod. 2014; 65: 418–423. https://doi.org/10.1016/j.jclepro.2013.07.055
10. Thang, N.C., Van Tuan, N., Yang, K.-H. and Phung, Q.T. Effect of zeolite on shrinkage and crack resistance of high-performance cement-based concrete. Materials 2020; 13: 3773. https://doi.org/10.3390/ma13173773
11. Chen, J.J., Li, L.G., Ng, P.L. and Kwan, A.K.H. Effects of superfine zeolite on strength, flowability and cohesiveness of cementitious paste. Cem. Concr. Compos. 2017; 83: 101–110. https://doi.org/10.1016/j.cemconcomp.2017.06.010
12. Alexa-Stratulat, S.-M., Olteanu, I., Toma, A.-M., Pastia, C., Banu, O.-M., Corbu, O.-C. and Toma, I.-O. The Use of natural zeolites in cement-based construction materials – a state of the art review. Coatings. 2024; 14: 18. https:// doi.org/10.3390/coatings14010018
13. Salami B. A., Alimi W., Iqbal M., Jamal A., Abdulraheem A., Tafsirojjaman T., Liu Y., Jalal F. E. and Bardhan A. Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approaches. Cem. Concr. Compos. 2023; 145: 105924. https://doi.org/10.1016/j.cemconcomp.2023.105924
14. Yasin B., Maraqa F., Al-Sahawneh E., Al Adwan J. and Alzubi Y. A comparative assessment of regularized regression techniques for modeling the mechanical properties of rubberized concrete. Open Civ. Eng. J. 2022; 16: e187414952208170. https://doi.org/10.2174/18741495-v16-e2208170
15. Wang Y., Li X. and Zhang Q. Prediction of concrete compressive strength using support vector regression and non-destructive testing methods. Case Stud. Constr. Mater. 2024; 25: e02045. https://doi.org/10.1016/j.cscm.2024.e02045
16. Paixão R.C.F., Penido R.E.-K., Cury A.A. and Mendes J.C. Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization. Rev. IBRACON Estrut. Mater. 2022; 15(5): e15503. https://doi.org/10.1590/S1983-41952022000500003
17. Tong S. and Koller D. Support vector machine active learning with applications to text classification. J. Mach. Learn. Res. 2001; 2:45–66. https://doi.org/10.1162/153244302760185243
18. Reddy D. V., Zhang Y. and Li C. Sustainable mix design of recycled aggregate concrete using machine learning approach. J. Clean. Prod. 2024; 446: 140927. https://doi.org/10.1016/j.jclepro.2024.140927

19. Lesmana C. Predicting the shear strength of reinforced concrete beams using support vector machine. J. Technol. Sci. 2018: 2(2); 80. https://doi.org/10.28932/jts.v2i2.1257
20. Parlak İ. E. and Emel E. Deep learning-based detection of aluminum casting defects and their types. Eng. Appl. Artif. Intell. 2023; 118: 105636. https://doi.org/10.1016/j.engappai.2022.105636
21. Alghamdi S. J. Classifying high strength concrete mix design methods using decision trees. Materials. 2022; 15(5): 1950. https://doi.org/10.3390/ma15051950
22. Al-Sammarraie M. A. J., Gierz Ł. A., Özbek O. and Kırılmaz H. Power predicting for power take-off shaft of a disc maize silage harvester using machine learning. Adv. Sci. Technol. Res. J. 2024; 18(5): 1–9. https://doi.org/10.12913/22998624/188666
23. Kowal B., Organiściak P., Kuraś P., Masłoń A., Wadiak B. Machine learning in forecasting the amount of sewage for the wastewater treatment plant of the city of Rzeszów based on weather conditions. Pomiary Automatyka Robotyka, 2025; 29(1): 81–90. https://doi.org/10.14313/PAR_255/81
24. Antosz, K., Kulisz M, Husar J. Evaluation and Comparison of Selected Machine Learning Methods for Improving Maintenance Processes, IFAC-PapersOnLine.- 2024; 58(8): 85–90. https://doi.org/10.1016/j.ifacol.2024.08.055
25. Kłosowski G. and Rymarczyk T. Application of convolutional neural networks in wall moisture identification by EIT method. IAPGOŚ, 2022; 1. https://doi.org/10.35784/iapgos.2883
26. Kaushik A. K., Islam R., Elbahy S. and Arif M. Application of artificial neural networks in construction and built environment: A bibliometric analysis. Buildings. 2024: 14(8): 2423. https://doi.org/10.3390/buildings14082423
27. Arvind Saravan R., Annadurai R. Artificial neural networking prediction of the mechanical properties of high-performance concrete with zeolite replacement with cement. Asian J. Civ. Eng. 2023; 24: 2491–2503. https://doi.org/10.1007/s42107-023-00657-7
28. Siddique R., Aggarwal P. and Aggarwal Y. Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks, Adv. Eng. Softw. 2011; 42: 10. https://doi.org/10.1016/j.advengsoft.2011.05.016
29. Chia-Ju L. and Nan-Jing W. ANN model for predicting the compressive strength of concrete, Appl. Sci. 2021; 11: 3798. https://doi.org/10.3390/app11093798.
30. Amar M., Benzerzour, M., Zentar, R. and Abriak, N.-E. Prediction of the compressive strength of waste-based concretes using artificial neural network. Materials 2022; 15: 7045. https://doi.org/10.3390/ma15207045
31. Başyiğit C., Akkurt I., Kilincarslan S., Beycioglu A. Prediction of compressive strength of heavyweight concrete by ANN and FL model. Neural Comput & Applic. 2010; 19: 507–513. https:// doi.org/10.1007/s00521-009-0292-9
32. Awtoniuk M., Majerek D., Myziak A. and Gajda C. Industrial application of deep neural network for aluminum casting defect detection in case of unbalanced dataset. Adv. Sci. Technol. Res. J. 2022; 16(5): 120–128. https://doi.org/10.12913/22998624/154963
33. Kłosowski G., Hoła A., Rymarczyk T., Skowron Ł., Wołowiec T. and Kowalski M. The concept of using LSTM to detect moisture in brick walls by means of electrical impedance tomography. Energies 2021; 14(22): 7617. https://doi.org/10.3390/en14227617
34. Shahrokhishahraki M., Malekpour M., Mirvalad S. and Faraone G. Machine learning predictions for optimal cement content in sustainable concrete constructions. J. Build. Eng. 2024; 82: 108160. https://doi.org/10.1016/j.jobe.2023.108160
35. Rahman, M. F. A convolutional neural network (CNN) for defect detection of additively manufactured parts (Master's thesis, Missouri University of Science and Technology). Retrieved, 2022. https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=9098&context=masters_theses
36. Covatariu G., Toma O., Alexa-Stratulat S. M., Covatariu D., Bunea G. and Olteanu I. The use of Artificial Neural Network for predicting the thermal conductivity of cement based mortar with natural zeolites, IOP Conf. Ser.: Mater. Sci. Eng. 2024; 1304: 012023. https://doi.org10.1088/1757-899X/1304/1/012023
37. Onyari E.K. and Ikotun B.D. Prediction of compressive and flexural strengths of a modified zeolite additive mortar using artificial neural network, Constr. Build. Mater. 2018; 187: 30. https://doi.org/10.1016/j.conbuildmat.2018.08