APPLICATION OF MACHINE LEARNING FOR PREDICTION OF MECHANICAL PROPERTIES OF MORTARS AND CONCRETES

Matěj Prudil*,[1]

*Matej.Prudil@vut.cz
[1]Brno University of Technology, Faculty of Civil Engineering, Veveří 331/95, 602 00 Brno

Abstract
This paper deals with the application of machine learning (ML) in the field of concrete technology. Two databases of test mortars and concretes were created from selected academic theses, which include mechanical properties in relation to their composition. These databases were used to develop two ML models that predict the mechanical properties of mortars and concretes depending on their composition. The mortar test database contains a total of 242 mechanical property records and the concrete test database contains 111 records. The materials in the database are CEM I, CEM II and CEM III cements combined with additives such as ground granulated blast furnace slag, high temperature fly ash and micro-ground limestone.

Keywords
Concrete technology, machine learning, mechanical properties, compressive strength, flexural strength

1 INTRODUCTION

In recent years, the use of advanced computing technologies has become widespread, especially in solving complex optimization and engineering problems. This growth has been supported by the integration of machine learning (ML) techniques into various industries. For example, ML plays a key role in finance (investment forecasting and analysis), healthcare (diagnosis and treatment recommendation), and transportation (traffic monitoring and management). One such challenging problem is concrete mix design, a process that traditionally relies on the expertise of technologists, adherence to engineering standards, and the use of empirical relationships that are not commonly employed. The optimization problem lies in the selection of the appropriate material composition to achieve the desired properties of the concrete in both its fresh and hardened states [1].

Creating a machine learning model involves several key steps. The first step is data acquisition, where it is crucial to collect and prepare the data, including any necessary transformation. The data is then split into training and test sets, allowing the model to be validated on independent data. The actual training of the model is done using the training data to teach the model the patterns and relationships in the data. This is followed by validating and fine-tuning the model on the validation set to minimize overfitting and optimize parameters. Then, the model is tested on the test set, and its performance is evaluated using appropriate metrics such as RMSE or R^2.

If the model performs successfully, it can be used to make predictions on new data. An important step is to continuously monitor the performance of the model and maintain or update it as necessary to capture changes in the data or environment.

ML models can be divided into supervised and unsupervised learning.

In supervised learning, the learning system is provided with a set of examples where both input and output values or functions are given. The goal is to generate a hypothesis that can accurately predict the relationship between the input and output.

In contrast, unsupervised learning involves a set of examples where only the inputs are known and no information about the correct output is given.

The tasks performed by ML systems can be summarized as follows:

- Classification (supervised ML algorithm): the goal of this step is to determine the category (or class) to which the input data belongs [2].
- Regression (supervised ML algorithm): the goal is to model the relationships between the inputs and the numerical outputs [2].
- Clustering (unsupervised ML algorithm): the goal is to identify hidden data patterns in two or more datasets [2].

DOI 10.13164/juniorstav.2024.24085

This paper will use a regression task, i.e. supervised model learning.

A large amount of data is already available in the field of concrete technology that could be used for innovations and improvements in the concrete industry. Despite this fact, there is a relatively low number of freely available databases in public repositories in this field.

Prof. I-Cheng Yeh has compiled two concrete mix design databases licensed for unrestricted use, which can be used to create ML models to predict compressive strength and consistency [3], [4]. The most common form of these databases describes the composition of concrete in terms of cement content, water content, admixtures, additives, aggregates, and age of concrete which serve as input variables, followed by fresh and hardened properties, which serve as output variables.

In the field of concrete technology, ML has already been used by many experts to predict various concrete properties, most commonly compressive strength, tensile strength, shear strength and modulus of elasticity. In terms of concrete type, special types of concrete such as high-strength concrete, self-compacting concrete, lightweight concrete are predominant [5], [6].

Völker C. and colleagues presented an open-source application using machine learning algorithms to inverse design and predict the properties of newly designed concrete formulations. This application aims to assess the carbon footprint and improve the quality of materials while considering socioeconomic factors in materials design [7].

2 METHODOLOGY

Materials

In this article, a database of test mortars according to [8] (hereinafter referred to as the mortar database) and a database of test concretes prepared in a laboratory environment (hereinafter referred to as the concrete database) using raw materials from the Czech and Slovak Republics will be used.

Tab. 1 lists the materials used, indicating their origin, designation and occurrence in the mortar and concrete databases.

Tab. 1 List of materials used in the mortar and concrete database.

Name of material used	Origin	ID	Concrete database	Mortar database
CEM I 42.5 R	Mokrá, CZ	CEM I Mokrá	X	X
CEM II/A-LL 42.5 R	Hranice, CZ	CEM II Hranice		X
CEM II/B-S 42.5 N	Horné Srnie, SK	CEM II Horné Srnie	X	X
CEM III/A 42.5 N	Hranice, CZ	CEM III Hranice	X	X
High-temperature coal fly ash	Dětmarovice, CZ	FA Dětmarovice		X
High-temperature lignite fly ash	Chvaletice, CZ	FA Chvaletice		X
High-temperature lignite fly ash	Opatovice, CZ	FA Opatovice	X	X
Ground granulated blast furnace slag	Štramberk, CZ	GGBS Štramberk	X	X
Ground granulated blast furnace slag	Dětmarovice, CZ	GGBS Dětmarovice		X
Finely ground limestone	Štramberk, CZ	FGL Štramberk	X	X
Fine aggregate 0/4	Žabčice, CZ	0/4	X	
Coarse aggregate 4/8	Žabčice, CZ	4/8	X	
Coarse aggregate 8/16	Olbramovice, CZ	8/16	X	
Naphthalene Formaldehyde Superplasticizer	-	Naphthalene	X	
Polycarboxylate Superplasticizer	-	Polycarboxylate	X	

Data collection

In this paper, a database of concretes and mortars will be used, which was compiled from the theses of the Faculty of Civil Engineering of BUT [9], [10].

DOI 10.13164/juniorstav.2024.24085

The mortar database was created by laboratory preparation of test mortars with a total dosage of 450 g of cement and admixtures with a constant water dosage of 225 ml. The cement replacement ratio with admixture ranged from 10% to 70%. Three prismatic test specimens of 40 × 40 × 160 mm were prepared each time (total mixing volume was therefore 256 ml) and stored in water in the laboratory environment. After reaching the required age of the specimens (usually 2, 7, 28, 60 and 90 days), the specimens were subjected to tensile flexural strength and compressive strength tests.

Unlike the mortar data collection, the concrete database was compiled by laboratory preparation of concrete mix design with different proportions of binder, aggregate, and water. The replacement of cement by admixture ranged from 10 to 70%, similar to the mortars. Three 150 × 150 × 150 mm (in some cases 100 × 100 × 100 mm) cubic test specimens were prepared and placed in water in a laboratory environment. Compressive strength testing was usually carried out at 7, 28, 60 and 90 days.

Mortar database overview

The mortar database contains a total of 242 records. Tab. 2 gives an overview of this database.

Tab. 2 Statistical values of mortar database.

ID	Unit	Count	Mean	Standard deviation	Min	Max
CEM I Mokrá	g/256 ml	162	290.6	82.6	135	450
CEM II Hranice	g/256 ml	50	301.2	66.7	225	450
CEM II Horné Srnie	g/256 ml	15	303.6	116.9	135	450
CEM III Hranice	g/256 ml	15	296.9	109.6	135	450
FA Dětmarovice	g/256 ml	30	168.2	73.9	45	315
FA Chvaletice	g/256 ml	30	165.3	46.8	113	225
FA Opatovice	g/256 ml	46	170.4	71.1	45	315
GGBS Štramberk	g/256 ml	61	170.6	76.8	45	315
GGBS Dětmarovice	g/256 ml	15	165.3	47.6	113	315
FGL Štramberk	g/256 ml	31	172.9	80.7	45	315
Specimen age	day	242	37.6	30.2	2	90
Compressive strength	MPa	242	35.0	16.3	4.3	71.5
[1]**Flexural strength**	MPa	190	6.4	2.2	1.6	10.5

Concrete database overview

The concrete database contains a total of 111 records. Tab. 3 gives an overview of this database.

Tab. 3 Statistical values of concrete database.

ID	Unit	Count	Mean	Standard deviation	Min	Max
CEM I Mokrá	kg/m^3	101	210.7	79.8	99.5	359.5
CEM II Horné Srnie	kg/m^3	5	378.6	35.4	339.8	404.4
CEM III Hranice	kg/m^3	5	391.7	37.1	351.0	418.8
FA Opatovice	kg/m^3	28	132.5	60.2	34.2	232.1
GGBS Štramberk	kg/m^3	32	143.1	66.9	34.3	239.1
FGL Štramberk	kg/m^3	24	137.2	66.3	34.3	237.4
0/4	kg/m^3	111	858.7	23.4	826.0	935.9
4/8	kg/m^3	111	242.3	6.6	233.0	264.0

[1] Flexural strength was not used as an input variable in the development of the ML model.

DOI 10.13164/juniorstav.2024.24085

8/16	kg/m^3	111	651.9	17.8	627.1	710.5
Naphthalene	kg/m^3	34	3.1	0.0	2.9	3.2
Polycarboxylate	kg/m^3	9	3.3	0.1	3.2	3.4
Water	kg/m^3	111	202.4	22.5	135.1	226.9
Specimen age	day	111	43.5	31.1	7.0	90
Compressive strength	MPa	111	34.7	15.7	4.0	76.1

Performance evaluation methods

Two measures are used to assess the accuracy of model predictions.

The Root Mean Square Error (RMSE) represents the standard deviation between the predicted and actual measured values:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{1}$$

R-squared (R^2) represents the proportion of variability in measured and predicted values:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

where x_i represents the i-th measured value, y_i represents the i-th predicted value and n is the number of records.

RMSE values have the same unit as the model outputs and reach positive values in the interval $\langle 0; \infty)$. Lower RMSE values represent higher model accuracy and optimal data fitting.

R^2 measures how well the model fits the data by assessing the proportion of variance in the dependent variable explained by the independent variables. The value of R^2 reaches the interval $\langle 0; 1\rangle$. A value of 1 indicates a perfect fit of the model to the data, while a value of 0 indicates that the model does not explain any part of the variability in the dependent variable.

The K-fold cross-validation (CV) method was used in training the models. This method consists of splitting the database into folders of a certain amount, in this case, k = 5 was chosen. The model is then learned from k-1 folders, where one folder is kept for validation and measures of accuracy are determined for this model (i.e. RMSE and R^2 values are calculated). The model is then trained on another dataset of k-1 folders, with a different folder selected for validation than in the previous step. This process is repeated until the combination of validation folders is exhausted. The performance of k-fold cross-validation is then obtained by averaging the measurements from all learning steps. While this method can be computationally intensive, it exploits the higher potential even with a relatively small database size, unlike the scenario of selecting a random validation set.

Machine learning models

There are many different tools for automatic training of ML models. In this paper, the Regression Learner application in MATLAB software was used.

The Regression Learner application trains regression models to predict data. Using this application, it is possible to explore data, select features, enter validation schemes, train models, and evaluate results. It is possible to use automatic training and search for the best regression type from a list of defined models: linear regression models, regression trees, Gaussian process regression models, support vector machines, efficiently trained linear regression models, kernel approximation models, ensembles of regression trees, and neural network regression models.

The proposed algorithms were used to train models for predicting the compressive strength of concrete and predicting the compressive strength of mortar. These most accurate models were further optimized by adjusting the basic hyperparameters, resulting in models with even more accurate predictions. In the case of mortar compressive strength prediction, the Gaussian Process Regression (GPR) algorithm was identified as the best model, while the Boosted Trees algorithm was selected for concrete compressive strength prediction.

DOI 10.13164/juniorstav.2024.24085

3 RESULTS

This chapter will present the resulting accuracies of the mortar compressive strength prediction model (hereinafter referred to as the mortar model) and the concrete compressive strength prediction model (hereinafter referred to as the concrete model). Tab. 4 shows the resulting evaluation measures for the mortar model and the concrete model, including hyperparameters, model identification and training time – the models were trained using the CV method. Fig. 1 and Fig. 2 graphically present the deviation of predicted and measured (actual) values. Tab. 5 and Tab. 6 show the predicted compressive strengths of unknown mortar and concrete mix designs that were not included in the original database.

Tab. 4 Overview of ML models, algorithms and hyperparameters used with resulting evaluation measures.

Model identification	**Algorithm**	**Hyperparameters**	**RMSE** [MPa]	**R^2** [-]	**Training time** [s]
Mortar model	*GPR*	*sigma SD = 11.65*	*2.77*	*0.97*	*80.0*
Concrete model	*Boosted Trees*	*min leaf size = 10, n. of learners = 35*	*4.23*	*0.93*	*2.4*

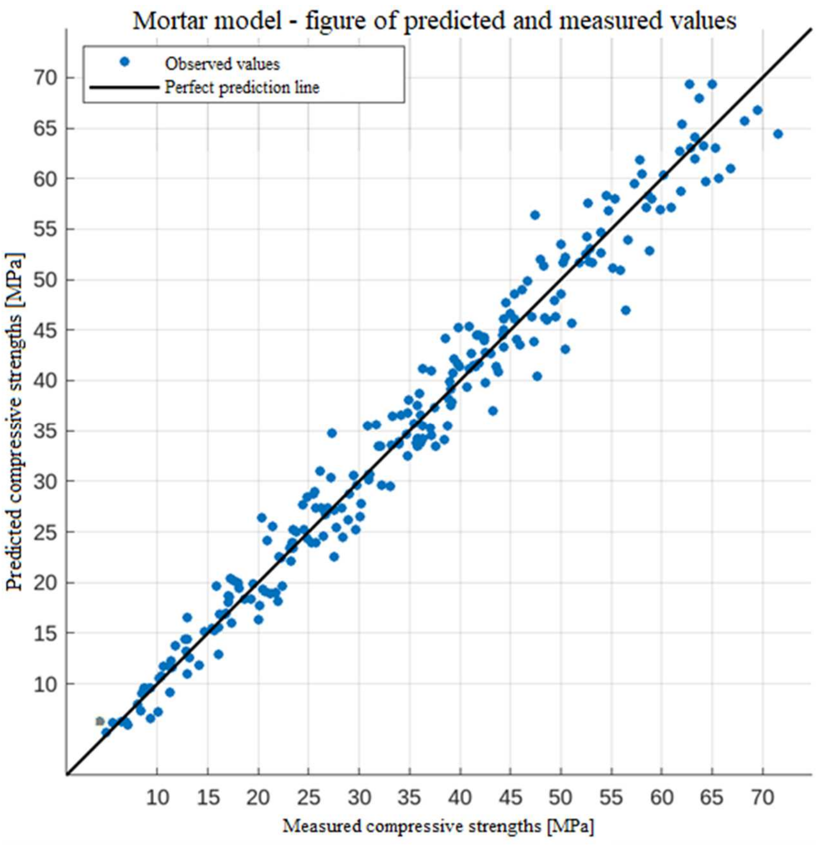

Fig. 1 Predicted and measured values of the compressive strengths of the mortar model (GPR).

DOI 10.13164/juniorstav.2024.24085

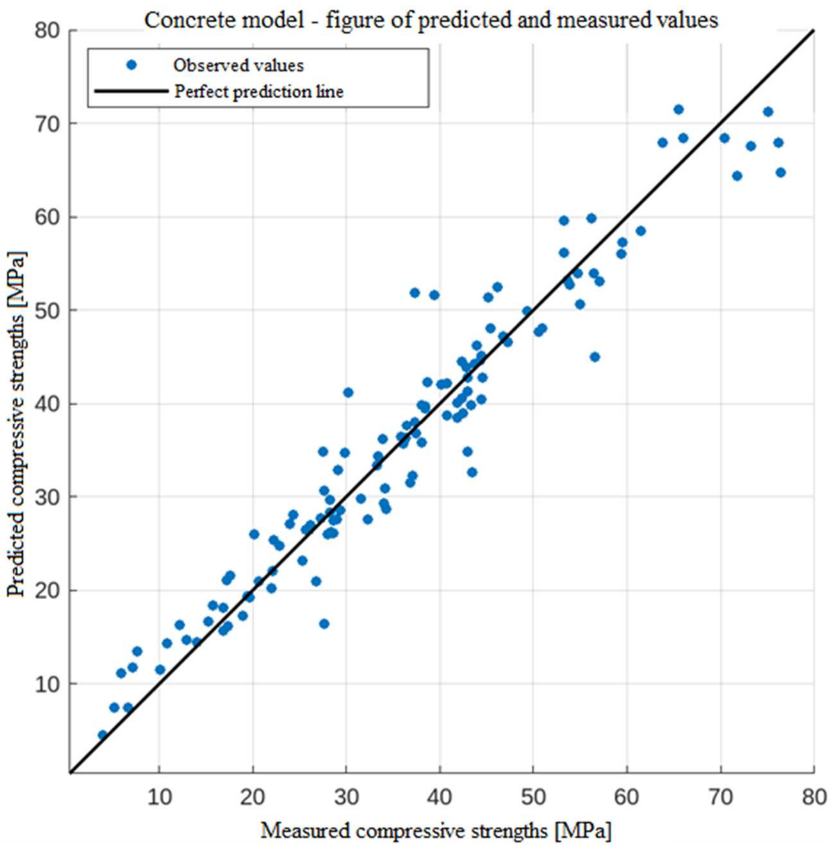

Fig. 2 Predicted and measured values of the compressive strengths of the concrete model (Boosted Trees).

Tab. 5 Predicted compressive strengths of mortars not included in the training data set.

CEM I Mokrá [g/256 ml]	**CEM II Hranice** [g/256 ml]	**GGBS Štramberk** [g/256 ml]	**FA Chvaletice** [g/256 ml]	**FA Opatovice** [g/256 ml]	**Specimen age** [day]	**Predicted comp. strength** [MPa]
350		*100*			*7*	*36.5*
350				*100*	*7*	*33.0*
	350		*100*		*7*	*25.7*
350		*100*			*28*	*54.2*
350				*100*	*28*	*44.7*
	350		*100*		*28*	*36.4*

Tab. 6 Predicted compressive strengths of concretes not included in the training data set.

CEM I Mokrá [kg/m^3]	**FA Opat.** [kg/m^3]	**GGBS Štram.** [kg/m^3]	**FGL Štram.** [kg/m^3]	**Water** [kg/m^3]	**0/4** [kg/m^3]	**4/8** [kg/m^3]	**8/16** [kg/m^3]	**Naphthal.** [kg/m^3]	**Polycar.** [kg/m^3]	**Spec. age** [day]	**Predicted comp. strength** [MPa]
273	*61*			*187*	*883*	*249*	*670*			*7*	*33.7*
286		*102*		*184*	*869*	*245*	*659*	*3.5*		*7*	*34.8*
306			*50*	*199*	*865*	*244*	*657*		*1.5*	*7*	*31.7*
273	*61*			*187*	*883*	*249*	*670*			*28*	*44.3*
286		*102*		*184*	*869*	*245*	*659*	*3.5*		*28*	*50.2*
306			*50*	*199*	*865*	*244*	*657*		*1.5*	*28*	*50.0*

DOI 10.13164/juniorstav.2024.24085

4 DISCUSSION

The mortar database contains 242 entries, and in terms of material, there are 4 types of cements, 3 types of fly ash, 2 types of slag, and 1 finely ground limestone. Ideally, the database should contain sufficient records to present a clear correlation with the output variation. For some cements (CEM II Horné Srnie and CEM III Hranice) and the additive GGBS Dětmarovice there are only 15 records in the database. Therefore, the model may be inaccurate in predicting the strengths of mortars with cements and additives that had a low number of records in the training database. However, CEM I Mokrá is represented in the database by 162 records, and the additive GGBS Štramberk by 61 records, which may lead to relatively high accuracy in predicting the strengths of mortars with these materials. Compared to the concrete database, the input variables of water and aggregate are not represented in the mortar database because these materials were always constant during mortar preparation.

The concrete database contains 111 records, and in terms of materials, it contains 3 types of cements, 1 type of fly ash, 1 type of slag, 1 type of finely ground limestone, 2 types of aggregates, 2 types of plasticizing admixture, various doses of water and various doses of aggregates. Most of the records in the concrete database contain CEM I Mokrá (101 records), which could lead to more accurate predictions with this cement. On the other hand, the cements CEM II Horné Srnie and CEM III Hranice are represented in the database with only 5 records each. This low number of records in the database of cements, coupled with the large variety of input variables, may lead to a reduction in model accuracy. In addition, compared to the mortar database, the concrete database contains different doses of water and different doses of aggregate (divided into 3 fractions). Moreover, it is evident from the standard deviations of aggregate and water in the database, which are up to 25 kg/m^3, that these variables are represented in the database with low variety. While this may lead to more optimal model fitting, the prediction of strengths outside the range of input variables from training may be highly inaccurate.

The resulting concrete and mortar model accuracies are based on the quality of the available database. The GPR was chosen as the most accurate algorithm for predicting the compressive strength of mortars, achieving RMSE = 2.77 and R^2 = 0.97. This model, according to the resulting evaluation gauges, achieves relatively high accuracy compared to the concrete model.

Boosted Trees was chosen as the most accurate algorithm for predicting the compressive strength of concretes, achieving RMSE = 4.23 MPa and R^2 = 0.93. This model did not achieve very high accuracy according to the resulting evaluation gauges.

In this study [11], the accuracy of the model with GPR algorithm for predicting the strength of high-performance concrete (HPC) was analysed. The results showed that the models developed in this study performed well in predicting the compressive strength of HPC, but GPR-32 (R^2 = 0.893; RMSE = 5.46; MAE = 3.86) was the most effective model compared to the other algorithms.

In this study [12] different prediction models for predicting the compressive strength of lightweight concrete (LWC) were compared. One of them was the GPR algorithm, which achieved the highest training accuracy: R^2 = 0.99; RMSE = 1.83; MAE = 1.44.

In this study [13] extensive data collection of compressive strengths of mortars with 10%–40% fly ash addition was carried out, with specimens aged 1-365 days with a total of 8 different fly ashes. The study further developed four prediction models for estimating compressive strength, of which the genetic algorithm was found to be the most accurate, showing R^2 = 0.95 and RMSE = 5.11.

Using the trained mortar model and concrete model, predictions were made for mixtures that were not included in the training database. These mixes were designed with materials that have sufficient representation of records in the training database to obtain plausible results.

5 CONCLUSION

Two databases were created from the final academic theses – the mortar database and the concrete database. The mortar database contains a total of 242 records, while the concrete database contains a total of 111 records. In the concrete database, unlike the mortar database, there are variable dosages of aggregates, water, and plasticizer. In terms of materials, the database contains several types of cements, fly ashes, slags and one type of finely ground limestone. Both of these databases provide a good basis for the creation of a more extensive file with supplemented records of the materials used.

Despite the lack of some records with the materials used, the GPR algorithm was identified as the most accurate for predicting the compressive strength of mortar, achieving a remarkable RMSE of 2.77 and R^2 of 0.97. In contrast, the concrete model using Boosted Trees shows lower accuracy with an RMSE of 4.23 and R^2 of 0.93. Other studies mentioned in the discussion further confirm the effectiveness of concrete algorithms such as GPR in predicting compressive strength for high-strength and lightweight concrete.

DOI 10.13164/juniorstav.2024.24085

Further research will focus on collecting additional data to increase the variety of some of the input variables in the database, and selecting other more accurate models capable of predicting the mechanical properties of mortars and concretes.

References

[1] ZIOLKOWSKI, Patryk and NIEDOSTATKIEWICZ, Maciej. Machine Learning Techniques in Concrete Mix Design. *Materials* [online]. 2019. [accessed 2024-01-25]. Vol. 12, no. 8. ISSN 1996-1944. Avaible at: https://doi.org/10.3390/ma12081256

[2] SHALEV-SHWARTZ, Shai and BEN-DAVID, Shai. Understanding Machine Learning: From Theory to Algorithms. 1. Cambridge University Press. 2014. ISBN 978-1-10-705713-5

[3] Concrete Compressive Strength Dataset. In: *UC Irvine Machine Learning Repository* [online]. 2007. [Accessed 2023-11-08] Avaible at: https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength

[4] Concrete Slump Test Dataset. online. In: UC Irvine Machine Learning Repository. 2009. Avaible at: https://archive.ics.uci.edu/dataset/182/concrete+slump+test. [Accessed 2023-11-08].

[5] BEN CHAABENE, Wassim; FLAH, Majdi and NEHDI, Moncef. Machine learning prediction of mechanical properties of concrete: Critical review. online. *Construction and Building Materials*. 2020, vol. 260. ISSN 09500618. Avaible at: https://doi.org/10.1016/j.conbuildmat.2020.119889. [Accessed 2023-10-31].

[6] MOHTASHAM MOEIN, Mohammad; SARADAR, Ashkan; RAHMATI, Komeil; GHASEMZADEH MOUSAVINEJAD, Seyed; BRISTOW, James et al. Predictive models for concrete properties using machine learning and deep learning approaches: A review. online. *Journal of Building Engineering*. 2023, vol. 63. ISSN 23527102. Avaible at: https://doi.org/10.1016/j.jobe.2022.105444. [Accessed 2023-10-31].

[7] VÖLKER, Christoph; MORENO TORRES, Benjamí; AHMAD ZIA, Ghezal; FIRDOUS, Rafia; KRUSCHWITZ, Sabine et al. Presenting SLAMD - A Sequential Learning based Software for the Inverse Design of Sustainable Cementitious Materials (First Revision). online. *ResearchGate*. 2023, vol. 2023. Avaible at: https://www.researchgate.net/publication/368510732_Presenting_SLAMD_-_A_Sequential_Learning_based_Software_for_the_Inverse_Design_of_Sustainable_Cementitious_Materials_First_Revision. [Accessed 2023-10-31].

[8] CEN. EN 196-1:2016 Methods of testing cement - Part 1: Determination of strength. 2016.

[9] PRUDIL, Matěj. Design of methodology for determination and practical application of k-value for commonly used admixtures and cements. Diploma Thesis, supervised Rudolf Hela. Brno: Faculty of Civil Engineering BUT, Institute of Technology of Building Materials and Components, 2023. Avaible at: https://www.vut.cz/en/students/final-thesis/detail/143625.

[10] ŠPERLING, Petr. Possibilities of determining the k-value and efficiency index for active admixtures and their assessment and determination for concretes. Diploma Thesis, supervised Ing. Adam Hubáček Ph.D. Brno: Faculty of Civil Engineering BUT, Institute of Technology of Building Materials and Components, 2021. Avaible at: https://www.vut.cz/en/students/final-thesis/detail/141707.

[11] DAO, Dong; ADELI, Hojjat; LY, Hai-Bang; LE, Lu; LE, Vuong et al. A Sensitivity and Robustness Analysis of GPR and ANN for High-Performance Concrete Compressive Strength Prediction Using a Monte Carlo Simulation. online. *Sustainability*. 2020, vol. 12, no. 3. ISSN 2071-1050. Avaible at: https://doi.org/10.3390/su12030830. [Accessed 2023-11-11].

[12] KUMAR, Aman; ARORA, Harish; KAPOOR, Nishant; MOHAMMED, Mazin; KUMAR, Krishna et al. Compressive Strength Prediction of Lightweight Concrete: Machine Learning Models. online. *Sustainability*. 2022, vol. 14, no. 4. ISSN 2071-1050. Avaible at: https://doi.org/10.3390/su14042404. [Accessed 2023-11-11].

[13] SEVIM, Umur, BILGIC, Hasan; CANSIZ, Omer; OZTURK, Murat and ATIS, Cengiz. Compressive strength prediction models for cementitious composites with fly ash using machine learning techniques. online. *Construction and Building Materials*. 2021, vol. 271. ISSN 09500618. Avaible at: https://doi.org/10.1016/j.conbuildmat.2020.121584. [Accessed 2023-11-11].