Journal of Building Engineering 76 (2023) 107248

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Machine learning methods in assessing the effect of mixture composition on the physical and mechanical characteristics of road concrete

I.G. Endzhievskaya [a], A.S. Endzhievskiy [a], M.A. Galkin [a], M.S. Molokeev [a,b,c,*]

[a] Siberian Federal University, Svobodny Ave., 79, Krasnoyarsk, 660041, Russia
[b] Laboratory of Theory and Optimization of Chemical and Technological Processes, University of Tyumen, Tyumen, 625003, Russia
[c] Laboratory of Crystal Physics, Kirensky Institute of Physics SB RAS, Krasnoyarsk, 660036, Russia

ARTICLE INFO

Keywords:
Concrete optimization
Random forest
Decision tree
Machine learning
Cement concrete roads

ABSTRACT

Current manuscript presents a study on the use of 48 experimental data points containing parameters of concrete production technological process and its properties, such as strength, density, and bending strength. It was revealed that temporal characteristics, specifically - compressive strength at the age of 3, 7, 28 days, R3, R7, and R28, are significantly correlated with each other, indicating that only one characteristic, such as R28 or R_{fl} 28, is sufficient for prediction. The absence of multiple correlations between parameters and properties suggests that linear regression analysis may not be accurate. Therefore, the use of Machine Learning is optimal; specifically Random Forest method is preferable due to ease of use and minimum hyperparameters for tuning. Low prediction errors (~1–11%) for 30% of the test data, as determined by the cross-validation method, confirm a relationship between the experimental parameters and the concrete properties. The most important parameters for achieving high values of compressive and bending strengths, R28 and R_{fl} 28, were identified, namely: air-entraining additives, granite crushed stone consisting of a mixture of fractions 5–20 mm, crushed stone derived from gravel of high strength grains of large fractions 10–20 mm. To obtain explanatory model, another Machine Learning method, that was used, called Decision Tree. The model showed that a high amount of crushed stone 10–20 mm from gravel, more than 212 (kg per 1 m^3 of concrete mix), leads to a higher number of strong grains with smooth, rounded surface, thereby, reducing the bending strength of concrete. However, a large concentration of crushed stone mix fractions of 5–20 mm from granite, more than 537 (kg per 1 m^3 of concrete mix), leads to the maximum roughness, which makes a significant contribution to the increased strength of concrete due to the adhesion of the matrix and aggregates to each other.

1. Introduction

The primary goal of road construction is to improve traffic safety by enhancing the quality and service life of road surfaces. However, construction workers face multiple issues such as climatic conditions: low temperatures in winter, large depth of soil freezing, significant amount of precipitation, long roads, and remoteness of the road works sites from the objects of mining or production of building materials. In addition to these limitations, the expenses, reliability, durability are key factors in the road and

* Corresponding author. Siberian Federal University, Svobodny Ave., 79, Krasnoyarsk, 660041, Russia.
E-mail address: msmolokeev@mail.ru (M.S. Molokeev).

https://doi.org/10.1016/j.jobe.2023.107248
Received 1 May 2023; Received in revised form 20 June 2023; Accepted 1 July 2023
Available online 1 July 2023

highway construction industry [1–6].

The construction of cement concrete roads (Fig. 1) is particularly relevant in connection with the development of vacuum residue processing technologies in the oil refining industry. The quantity of building bitumen is declining because the tar used for its production is being converted into motor fuel as much as possible. One of the main problems in the Nordic countries may be the lack of high-quality local materials in various regions for the production of aggregates. In most of the northern remote areas, the rocks are metamorphic and have low frost resistance.

The resistance to frost destruction of crushed stone grains in concrete depends on both its own resistance and the strength of the "shielding" barrier of the cement-sand matrix. Strength should not be lower than B30-B40.

The features of the capillary-porous structure of the matrix are very important. Reduced capillary porosity is achieved by using a complex of additives - super water-reducing and air-entraining. The use of microfillers also leads to the formation of dense cement shells around the aggregate grains. These methods contribute to a prolonged effect of the saturation reduction of aggregate grains. That is, under frost exposure, destructive processes will not manifest themselves so intensively.

The utilization of Machine Learning techniques to address various issues is experiencing a continuous. For example, the study [7] has examined the use of established statistical and Machine Learning methods to predict the compressive strength of concrete as a function of its mixture proportions using a large dataset with over 10,000 measured compressive strengths. However, despite the massive dataset, the model could still predict compressive strength with an average relative error of less than 10%.Precise prediction model was obtained Neural Net Machine Learning method [8], using the database with 741 records with 17 input variables. However, it is widely understood that Neural Nets have relatively low explain ability compared to other models [9,10]. Comparison of several Machine Learning methods, namely Artificial Neural Net (ANN), Support Vector Machine (SVM), Decision Trees (DT) and Evolutionary Algorithms (EA) to predict the compressive strength, shear strength, tensile strength, and elastic modulus of concrete has been reviewed in manuscript [11]. It was emphasized that the 'black-box' method (ANN) in complex models can lead to overfitting, which means that the complexity of concrete properties can be overestimated. Regarding SVM models, they have shown powerful nonlinear mapping and generalization abilities, but it also considered as "black-box" and cannot be used to rules extraction. In summary, the Decision Tree method appears to be the most effective way to forecast the strength of concrete and determine associated rules. Recently the DT was used for prediction of the mechanical properties of roller-compacted concrete pavement utilizing 290 data records [12]. It was shown that ensemble of DT, named Random Forest, significantly better than the other classification-based regression methods.

The purpose of this work was to study the influence of the composition components of the cement concrete mixture, chemical and mineral additives on the physical and mechanical characteristics of cement road concrete on a large number of samples using the Random Forest and Decision Tree methods. To achieve this aim, 46 concrete compositions were selected with compressive strength class not lower than B30 and tensile strength in bending - Btb 4.8. Cubic samples were made on mixtures with mobility (cone draft 2–8 cm) using various inert materials, plasticizing and air-entraining additives. The prediction of concrete strength through Machine Learning rarely utilizes a comprehensive mix of 29 feature parameters, making this study unique in its approach. Furthermore, the current focus of the work is to extract rules from the model, a process that is rarely observed.

2. Data and method

2.1. Dataset description

The dataset of 46 experiments with 29 feature parameter sand 8 property parameters (Table 1, Appendix A, B) was chosen for analysis. The descriptive statistics and the plot of the considered dataset are summarized in Appendix C, D, E, respectively. The materials used and properties of mix and concrete are listed in Table 1.

Preliminary data analysis revealed correlation between some feature parameters, mainly between different size of crushed stones, densities and airs feature parameters (Table 2, Fig. 2). Also there are correlations between property values, for example R3, R7 and R28, which means, for example, that strongest values R3 lead to strongest R28 (Appendix E). Therefore we took only "R 28 days", "p 28 days" and "R_{fl} 28 days" property parameters as representative. It should be noted that feature "Crushed stone 5–20 mm. Granite" has correlation with property "R 28 days" under consideration (Table 2, Appendix E), therefore this feature has the biggest effect on R 28 days. However, the correlation matrix (Fig. 2) proves absence on linear relationships between selected features and "R_{fl} 28 days", "ρ 28 days" properties, which justifies the need to use Machine Learning methods.

2.2. Random forest method

The prediction tool used in this study is random forest (RF), an ensemble method based on regression trees [13]. The regression

Fig. 1. Cement concrete road construction.

Table 1
The materials used and properties of mix and concrete.

Number of parameter	Material characteristic	Unit	Designation
1	Microsilica MKU 85 (active)	kg per 1 m^3 of concrete mix	Microsilica MKU 85
2	Stone flour from gravel (inert)	kg per 1 m^3 of concrete mix	Stone flour (from grave)
3	Natural sand with fineness modulus 2.0	kg per 1 m^3 of concrete mix	Natural Sand $M_f = 2,0$
4	Sand enriched with crushing screenings, with fineness modulus 2.3	kg per 1 m^3 of concrete mix	Artificial sand, $M_f = 2,3$
5	Granite crushed stone of a mixture of fractions 5–20 of the optimal granulometric composition	kg per 1 m^3 of concrete mix	Crushed stone 5–20 mm. Granite
6	Crushed stone of medium quality with inclusions of grains of weak rocks of fractions 5–10 in various ratios	kg per 1 m^3 of concrete mix	Crushed stone with carbonate inclusion 5–10 mm.
7	Crushed stone of medium quality with inclusions of grains of weak rocks of fractions 10–20 in various ratios	kg per 1 m^3 of concrete mix	Crushed stone with carbonate inclusion 10–20 mm.
8	Crushed stone from gravel of high strength grains of fractions 5–10 in various ratios	kg per 1 m^3 of concrete mix	Crushed gravel 5–10 mm.
9	Crushed stone from gravel of high strength grains of fractions 10–20 in various ratios	kg per 1 m^3 of concrete mix	Crushed gravel 10–20 mm.
10	Water	kg per 1 m^3 of concrete mix	Water
11	Water cement ratio	–	W/C
12	$r = \frac{\text{Sand}}{\text{Sand} + \text{Crushed stone}}$	–	r
13	Additive based on polycarboxylate ethers (RSEmix)	% by weight of cement	MC-TECHNIFLOW 170
14	Additive based on polycarboxylate ethers (RSEmix)	% by weight of cement	MC-TECHNIFLOW 173
15	Additive based on polycarboxylate ethers (RSEmix)	% by weight of cement	MC-POWERFLOW 7951
16	Additive based on naphthalene formaldehyde sulfonic acid (SA)	% by weight of cement	MURAPLAST FK 49
17	Additions based on SA + inhibitor	% by weight of cement	MC-POWERFLOW 7951 + CENTRAMENT RETARD 390
18	MURAPLAST	% by weight of cement	MURAPLAST FK 88
19	Plasticizing and air-entraining complex	% by weight of cement	CENTRAMENT N11
20	Air-entraining additives	% by weight of cement	CENTRAMENT AIR 202
21	Mixture curing retardant	% by weight of cement	CENTRAMENT RETARD 390
Properties of the concrete mixturex			
22	Estimated density	kg/m^3	Estimated density
23	Density of mix after 10 min.	kg/m^3	Density fact 10 min
24	Density of mix after 60 min.	kg/m^3	Density fact 60 min
25	Density of mix after 90 min.	kg/m^3	Density fact 90 min
26	The air content in the concrete mixture after 15 min.	%	air 15 min.
27	The air content in the concrete mixture after 60 min.	%	air 60 min.
28	The air content in the concrete mixture after 90 min.	%	air 90 min.
29	Cone draft	cm	Cd
Concrete properties			
30	Compressive strength of concrete at the age of 3 days	MPa	R 3 days
31	Concrete density at 3 days	kg/m^3	ρ 3 days
32	Compressive strength of concrete at the age of 7 days	MPa	R 7 days
33	Concrete density at 7days	kg/m^3	ρ 7 days
34	Compressive strength of concrete at the age of 28 days	MPa	R 28 days
35	Concrete density at 28 days	kg/m^3	ρ 28 days
36	Flexural strength of concrete at the age of 7 days	MPa	R_{fl} 7 days
37	Flexural strength of concrete at the age of 28 days	MPa	R_{fl} 28 days

trees are built by recursive binary partitioning of the multidimensional predictor space into regions by constructing a multitude of "decision trees" at training time and outputting the class that is the "mode" of the classes (classification) or mean/average prediction (regression) of the individual trees [13].

To begin with, the dataset is partitioned into two distinct groups: the training dataset, which typically contains between 70 and 90% of the total dataset, and the test dataset, which contains the remaining 10–30%. The training dataset is utilized to randomly select relevant features and samples for constructing the Decision Tree model (Fig. 3). This procedure is repeatedly performed until an

Table 2
Selected parameters with strongest correlation coefficients.

Number of parameter 1	Designation of parameter 1	Number of parameter 2	Designation of parameter 2	Correlation coefficient
3	Natural Sand $M_f = 2,0$	4	Artificial sand, $M_f = 2,3$	−0.94
3	Natural Sand $M_f = 2,0$	5	Crushed stone 5–20 mm. Granite	0.71
5	Crushed stone 5–20 mm. Granite	13	MC-TECHNIFLOW 170	0.84
5	Crushed stone 5–20 mm. Granite	34	R 28 days	0.71
6	Crushed stone with carbonate inclusion 5–10 mm.	8	Crushed gravel 5–10 mm.	0.77
6	Crushed stone with carbonate inclusion 5–10 mm.	7	Crushed stone with carbonate inclusion 10–20 mm.	0.91
6	Crushed stone with carbonate inclusion 5–10 mm.	18	MURAPLAST FK 88	0.87
6	Crushed stone with carbonate inclusion 5–10 mm.	20	CENTRAMENT AIR 202	0.85
7	Crushed stone with carbonate inclusion 10–20 mm.	9	Crushed gravel 10–20 mm.	−0.78
7	Crushed stone with carbonate inclusion 10–20 mm.	18	MURAPLAST FK 88	0.93
7	Crushed stone with carbonate inclusion 10–20 mm.	20	CENTRAMENT AIR 202	0.89
10	Water	11	W/C	0.73
15	MC-POWERFLOW 7951	20	CENTRAMENT AIR 202	0.75
18	MURAPLAST FK 88	20	CENTRAMENT AIR 202	0.93
23	Density fact 10 min	24	Density fact 60 min	0.95
23	Density fact 10 min	25	Density fact 90 min.	0.90
24	Density fact 60 min	25	Density fact 90 min.	0.95
26	air 15 min % device	27	air 60 min %	0.99
26	air 15 min % device	28	air 90 min %	0.94
27	air 60 min %	28	air 90 min %	0.97
30	R 3days	34	R 28 days	0.83
30	R 3days	32	R 7 days	0.90
31	ρ 3 days	35	ρ 28 days	0.73
31	ρ 3 days	33	ρ 7 days	0.74
32	R 7 days	34	R 28 days	0.91
33	ρ 7 days	35	ρ 28 days	0.87
36	R_{fl} 7 days	37	R_{fl} 28 days	0.76

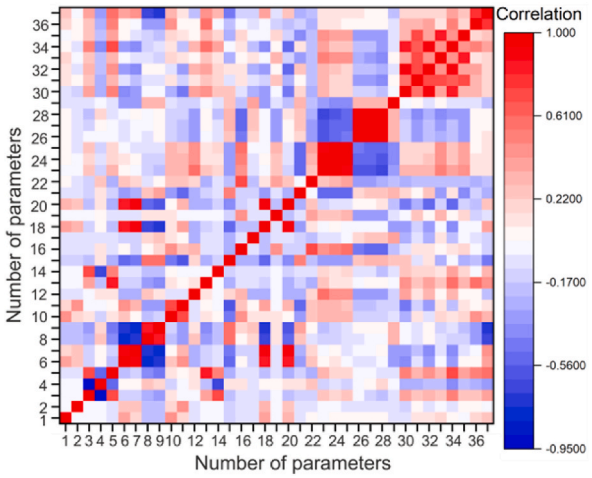


Fig. 2. The correlation matrix of feature and property parameters. There is correlation between one feature "Crushed stone 5–20 mm. Granite" (parameter 5) and property "R28 days" (parameter 34) meaning existing relationships.

optimal ensemble of N trees is formed, where N represents the only essential hyperparameter that requires tuning in the Random Forest method. Following this, the ensemble of trees is evaluated with the test dataset. If it attains a small Mean Absolute Error (MAE) or high accuracy, the model is considered dependable and well-fit for making predictions. Predictions are done by passing new data parameters from the root through the internal nodes until a terminal node is reached.

We used self-written python script named RandomForest.py using the Python 3.6 programming language [14] in order to build the

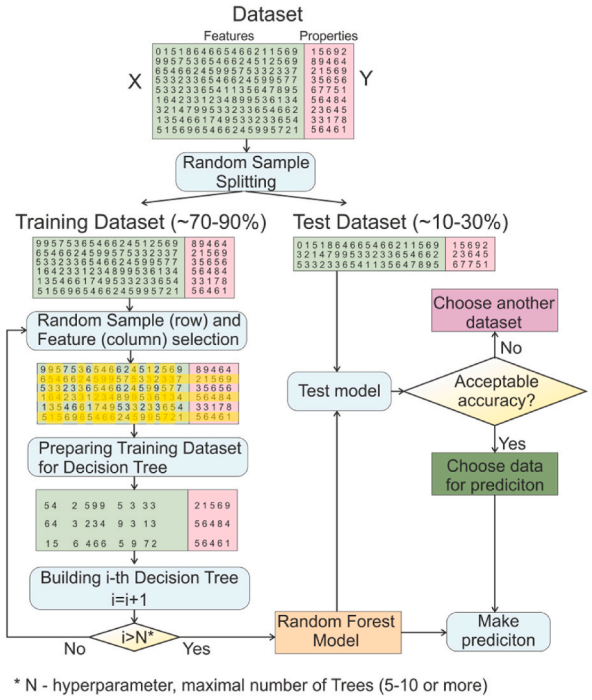


Fig. 3. The primary process flow involves dataset processing, building, validating, and generating predictions using the Random Forest model.

described RF model. The number of standard libraries were used in the program: numpy, pandas, sklearn, matplotlib and mpl_toolkits. Since this machine-learning algorithm is stochastic, we performed ten repeats of cross-validation and averaged the performance across them. Each time, the data were split into the two random datasets: a set for training procedure (80% of total data), and another set for test (20% of total data). The MAE of training set and test dataset for "R 28 days", "ρ 28 days" and "R_{fl} 28 days" are presented in Table 3. Additionally, we have made 5-fold cross-validation test on whole dataset, which also summarized in Table 3. The error is low, and we can conclude that the main correlation between experimental features and the main properties have been revealed (Fig. 4).

It should be noted that RF allows measuring the importance of the feature after training. The selected value is permuted among the training data and the error is computed on this perturbed data set. The importance score for the selected feature is computed by averaging the difference in error before and after the permutation over all trees [15]. The score is normalized by the standard deviation of these differences. Features which produce large values for this score are ranked as more important than features which produce small values (Fig. 4). Now we can highlight the most important parameters which influence on "R 28 days", "ρ 28 days", "R_{fl} 28 days": "CENTRAMENT AIR 202", "Crushed stone mix 5–20 mm Granite","Cd", "CENTRAMENT RETARD 390" and "Crushed gravel 10–20 mm", respectively.

2.3. Decision tree method

In the previous chapter we obtained an RF prediction model and identified importance feature parameters, that exert the greatest influence on concrete properties. This chapter is focused on presenting the rules that will lead to best properties, i.e. interpretation of prediction. It is well known that Decision Tree (DT) is one of the best methods to get interpretation [16,17].

The process of building a Decision Tree begins with sorting the dataset by a specific feature parameter, denoted as i, and dividing the properties into two subsets, Y1 and Y2, based on a value of m_i (Fig. 5). The goal is to minimize the normalized sum of dispersions (D), or entropies (E), of the subsets (determined by D = N1*D1 + N2*D2, or E = N1*E1 + N2*E2), as shown in Fig. 5. This step is repeated for all feature parameters, and the parameter with the lowest D or E value is selected. The rule $i < m_i$ is used to construct the upper portion of the Decision Tree. Y1 and Y2 may be further subdivided using the same procedure, until the full Decision Tree is formed. The final tree can be used to predict outcomes by passing new data parameters from the root to the internal nodes, and ultimately to the terminal node.

Table 3
Mean Absolute Errors of training and test datasets after RF model building.

Property	MAE(Training Dataset)	MAE(Test Dataset)	MAE(Cross-Validation)
R 28 days	1.83	5.17	5.13
ρ 28 days	14.71	24.27	33.37
R_{fl} 28 days	0.14	0.24	0.36

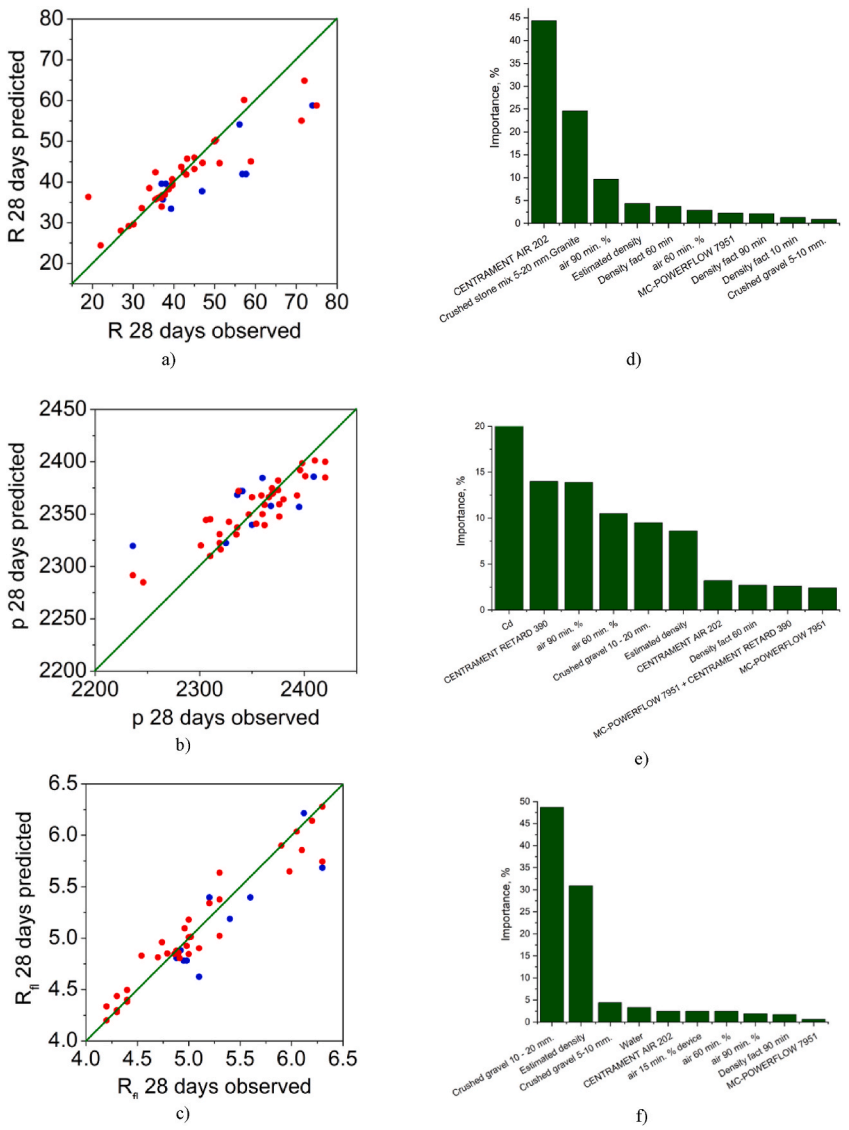


Fig. 4. Comparative plot of observed property values per calculated values obtained from RF model for: (a) "R 28 days"; (b) "ρ 28 days"; (c) "R_{fl} 28 days". Red dots represent training dataset and blue dots – test dataset. Linear fit proves correctness of model. Importance of main feature parameters on properties: d) "R 28 days"; e) "ρ 28 days"; f) "R_{fl} 28 days". (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

We used self-written python script named DecisionTree.py using the Python 3.6 programming language [14] in order to build the DT model. The standard libraries were used in this program: numpy, pandas, sklearn, matplotlib and mpl_toolkits. The DT with the depth equal to 2–3 was used to simplify the task. The DT models which lead to best data segregation are presented in Fig. 6.

It was presented that the main rule to get high values of "R 28 days" property is "CENTRAMENT AIR 202" <0.221 and "Crushed stone mix 5–20 mm. Granite" > 537.5. There are four experiments which fulfill this condition with average "R 28 days" = 73.1+/-1.5. The "air 15 min" feature parameter also important and should not be bigger than 8.5, otherwise it leads to low "R 28 days" values (Fig. 6a).

The same rule "Crushed stone mix 5–20 mm. Granite" > 537.5 is important to get high values "ρ 28 days" (Fig. 6b).Again four experiments, which fulfill this rule have the highest property values "ρ 28 days" = 2401+/-5. So we can conclude that this rule is general for two properties of concrete.

The "R_{fl} 28 days" property demands another rule, which is depicted in Fig. 6c. This main rule is "Crushed gravel 10–20 mm"<212. There are 14 experiments which fulfill this inequality and have high values of "R_{fl} 28 days" = 5.8+/-0.5. Additional rule is "Estimated density"<2428 which segregates 6 experiments from these 14 experiments with the highest "R_{fl} 28 days" = 6.20+/-0.09.

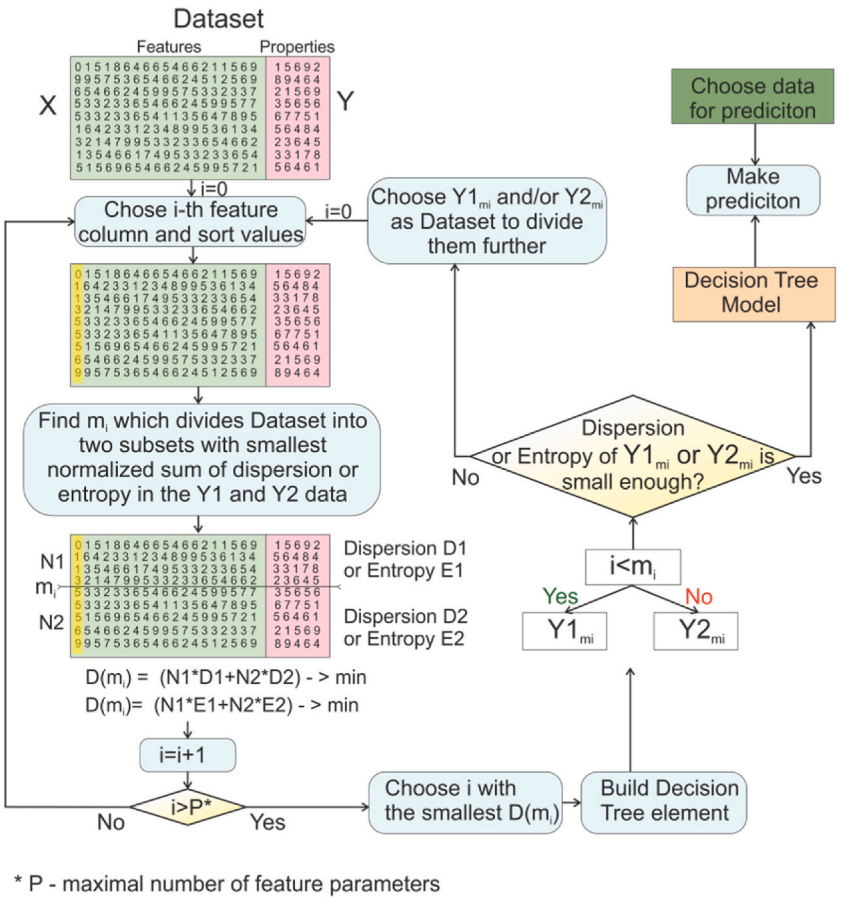


Fig. 5. The process depicted in the flowchart entails main components - dataset processing, construction of a Decision Tree model, and the generation of predictions based on this model.

3. Results and discussion

3.1. Analysis of concrete structural strength by RF and DT methods

The issue of increasing the resistance of road concrete cannot be independently addressed. When designing concrete, an integrated approach is always used to obtain both strength and performance properties such as frost resistance, water permeability, corrosion resistance, etc. Any measures aimed at enhancing the strength properties of concrete invariably lead to concrete with the densest structure. Such a structure has a minimum number of defects in the form of pores and microcracks.

Random Forest method have shown that the most important components for structural strength are “CENTRAMENT AIR 202”, “Crushed stone mix 5–20 mm Granite”, and their sum influence on structural strength is ~70% over all components (Fig. 4). Decision Tree proves their high influence, since these conditions most often appears at the root and nodes of decision trees (Fig. 6). The dataset depicted in the space of these most important parameters really shows nice segregation of S28 with high and low values (Fig. 7). The Decision Tree revealed that an elevated concentration of coarse aggregate notably influences the strength of S28 (as shown in Fig. 4a). The observation is supported by the disparity "Crushed stone mix 5–20 mm. Granite >537.5" which results in the separation of the four most resilient concretes among all specimens (Fig. 7). The rough surface likely augments adhesion to the cement stone, which significantly contributes to the concrete’s strength by enabling better cohesion between the matrix and filler. The level of adhesion is determined by the roughness of the filler’s surface, the shape and size of its particles, and the value of its modulus of elasticity.

The use of air-entraining chemical additives makes it possible to obtain small closed pores evenly distributed throughout the volume of the material. The inequality “CENTRAMENT AIR 202 < 0.21” (Fig. 6a) quite well separates the group of concretes with high compressive strength S28 from the ordinary ones. Since the equation is at the root of the Decision Tree, it has the greatest weight and should be fulfilled in order to obtain desired characteristics. It should be noted that within the node of the Decision Tree, the upper limit of 8.5%for air entrainment in concrete was identified. Beyond this threshold, the strength of R28 concrete decreases significantly (Fig. 4a). This is due to the increase in total and open porosity, as well as pore sizes.

To determine the effect of air entrainment in slow-moving mixtures of road pavement concrete deformation the compressive strength of concrete samples was tested by using in the presence or absence of air entrainment (Fig. 8). Deformations were developed less in concrete specimens with 5.5% air entrainment. The cement matrix with air entrainment becomes less rigid, but more "ductile".

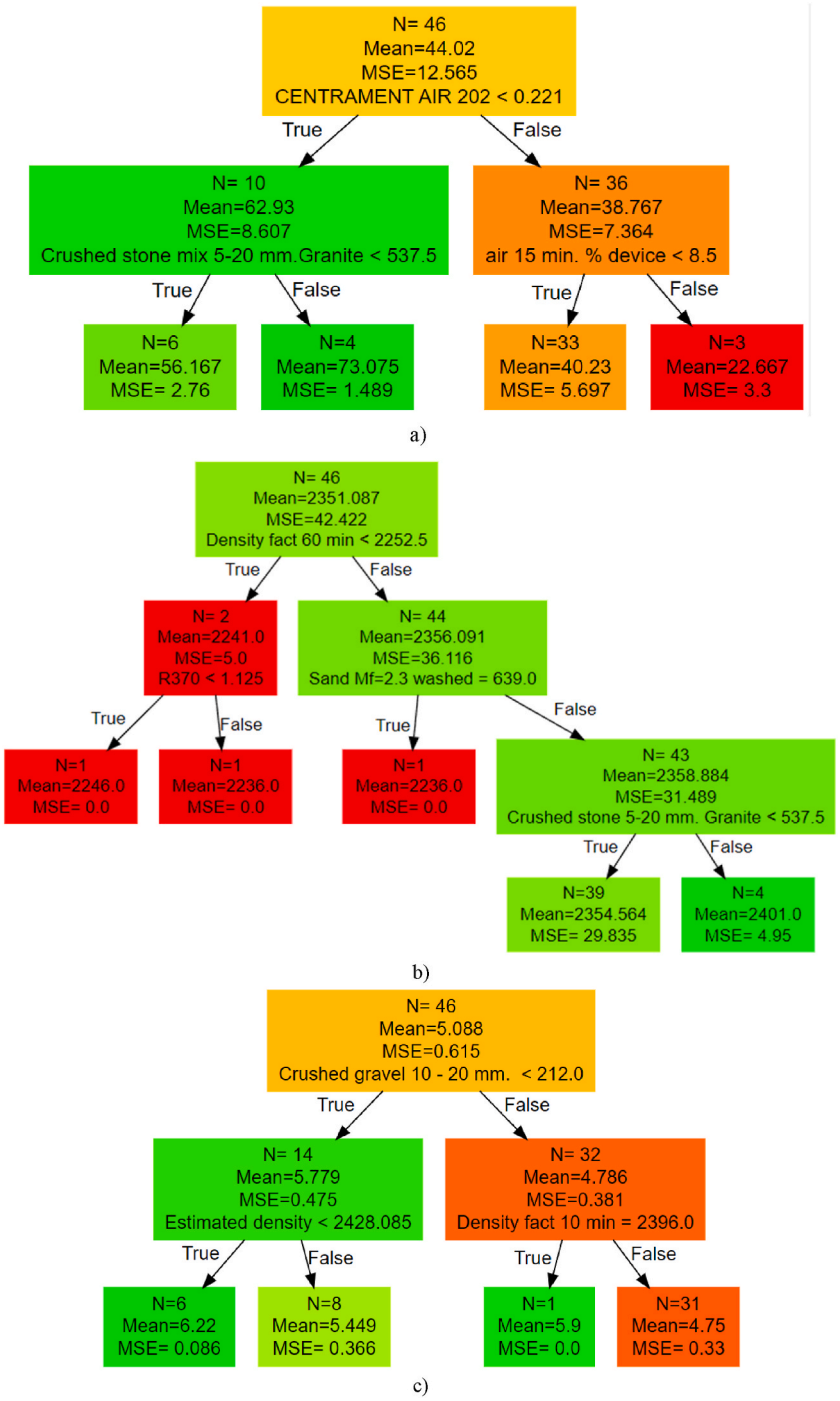


Fig. 6. The Decision Tree models obtained from whole dataset of 46 experiments: (a) "R 28 days" property; (b) "ρ 28 days"; (c) "R_{fl} 28 days". The bars which depict root, nodes or leafs have colors: green (the highest property value), red (the smallest property value) and yellow (average values). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Therefore, the cement matrix is more resistant to cyclic impacts.

3.2. Analysis of concrete density by RF and DT methods

According to importance parameters obtained from Random Forest method (Fig. 4e) there are several important parameters. The first three of them: Cone draft (Cd); mixture curing retardant (CENTRAMENT RETARD 390); the air content in the concrete mixture after 90 min. They account less than 50% of influence on density over all components. It is a small value indicating that density ρ 28

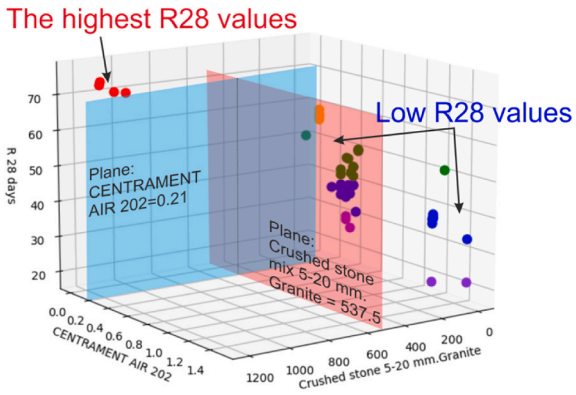


Fig. 7. The experimental dataset showed in 3D space spanned on two most important feature parameters: 1) CENTRAMENT AIR 202; 2) Crushed stone mix 5–20 mm. Granite, and 3) structural strength property R28 days. The depicted and signed colour planes segregates R28 with the highest values (red circles) from R28 with low values (violet, blue, green, orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

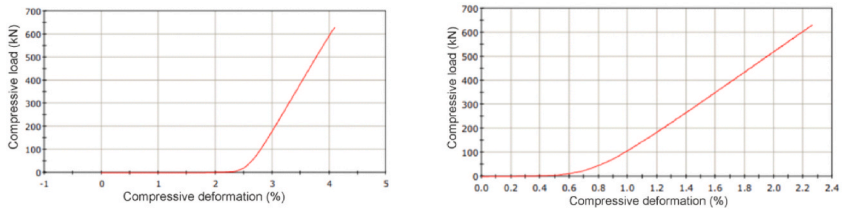


Fig. 8. Diagrams of deformations during the compressive strength test of concrete specimens aged 2 months (a) without air entrainment (b) with air entrainment 5.5%.

days appeared to be really complex value, having nonlinear dependence per many parameters, and this dependence is hard to obtain in a simple form. Indeed, Decision Tree proves it, showing slow segregation rate under deep tree increasing, therefore usual 3D plot cannot show nice segregation. Nevertheless, Random Forest, which combines many Decision Trees with the highest deep produces a nice model with correct prediction and small error (~ 1%) and it can be applied for ρ 28 days estimation using 29 parameters.

3.3. Analysis of bend strength by RF and DT methods

The Random Forest approach has revealed that "Crushed gravel 10–20 mm" and "Estimated density" are the critical factors that determines the bend strength with a combined impact of approximately 75% over all the components (as depicted in Fig. 4e). The Decision Tree analysis further corroborates their significance, as these conditions appeared at the root and the closest nodes of the Decision Tree (as depicted in Fig. 6). The dataset depicted in the space of these most important parameters, indeed demonstrates a clear segregation of R_{fl} 28 with high and low values (Fig. 9).

The detailed analysis of Decision Tree for "R_{fl} 28" (Fig. 4c) revealed the inequality "Crushed gravel 10–20 mm < 212" is at the root which separates the 14 strong concretes from the rest. This is the main condition, which means that crushed gravel, having strong grains, but a smooth and rounded surface, reduces the bending strength of concrete (Fig. 9).

Remarkably, the most critical factor for both structural and bend strength is a large aggregate mixture of fractions of 5–10 and 10–20 mm, i.e. coarse filler (Fig. 4 a,e). It improves the properties of concrete due to a significant decrease in its shrinkage properties, and the appearance of the reinforcing effect by the inclusions having higher elastic moduli. With the optimal ratio of the quantity and quality of aggregates in the matrix, it is possible to obtain concretes of high density and performance properties.

4. Conclusions

The structural and bending strength of road concrete is influenced by many factors, including the aggregate mixture, crushed gravel/granite stone shape and size, and the presence of voids. The scarcity of good quality local materials in Nordic countries poses a significant challenge, and therefore, it is crucial to carefully select the aggregate mix and to pay close attention to the production process to ensure a dense and strong structure.

In this study, Machine Learning techniques were employed to develop predictive models for key properties such as structural strength (S28), density (ρ28), and bending strength (R_{fl} 28) with small errors of 5.13, 33.37, and 0.36, respectively. These errors represent only 1–11% of the average, indicating the high level of accuracy achieved with this approach. The effectiveness of Machine Learning for such tasks was also proved by determination of the most important factors affecting the structural and bending strength of

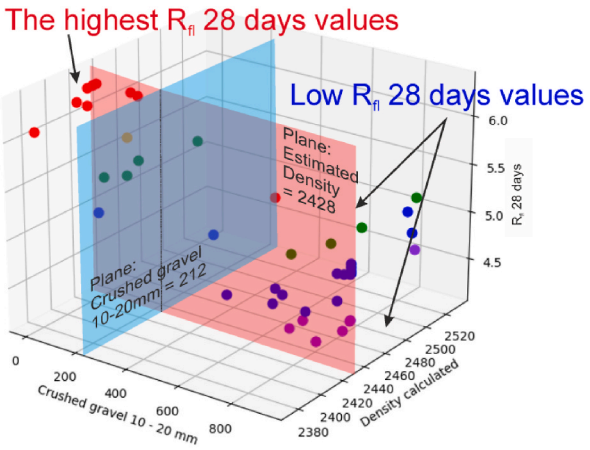


Fig. 9. All experimental dataset showed in 3D space spanned on two most crucial feature parameters: 1) Crushed gravel 10–20 mm; 2) calculated Density, and 3) bend strength property R_{fl} 28 days. The depicted and signed colour planes segregates R_{fl} 28 with the highest values (red circles) from R_{fl} 28 with low values (violet, blue, green, orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

road concrete. The Decision Tree method proved to be particularly effective in extracting valuable rules, including the optimal concentrations of the most important components, to achieve desirable results. The summary of these rules are listed below:

- The most important factor for structural strength of concrete is coarse aggregate 5–10 or 10–20 mm;
- The critical concentration of Crushed gravel 10–20 mm is 212. The values lower this number lead to concretes with the highest R_{fl} 28. It appeared that strong grains and smooth rounded surface of gravel reduces the bending strength of concrete;
- The critical concentration of Crushed stone mix 5–20 mm, granite is 537.5. The values higher this number lead to concretes with a high S28. Apparently, this form of aggregate has the maximum roughness, which makes a significant contribution to increasing the strength of concrete due to the adhesion of the matrix and aggregate to each other.
- Additionally, concretes with high S28 have the unique condition CENTRAMENTAIR $202 < 0.21$, which should probably always be fulfilled in order to get high S28;
- A less important, yet still worthy of attention parameter was: the upper limit of air entrainment in concrete should be about 8.5%, when it is exceeded, the strength of s28 concrete decreases sharply.

Obtained critical values in particular justify the costs and necessary of used components to build road surfaces with high quality and service life.

Contrary to popular belief, recent research suggests that Machine Learning is not as challenging as once thought and doesn't require extensive data. In fact, a modest dataset of only around 50 examples is sufficient. Furthermore, it's unnecessary to adjust hyperparameters such as those of a Decision Tree. This means that even with just a small amount of experience, one can successfully implement these models and achieve dependable results with clear explanations.

In general, the results of this study provide valuable insights for engineers and architects to design road concrete with increased resistance and longevity.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Maxim Molokeev reports financial support was provided by Tyumen Oblast Government.

Data availability

Data will be made available on request.

Acknowledgements

The work was carried out within the framework of the Strategic Academic Leadership Program "Priority-2030" for the Siberian Federal University. The research in field of Machine Learning application was supported by the Tyumen Oblast Government, as part of the West-Siberian Interregional Science and Education Center's project No. 89-DON (3).

Appendix A. Dataset of 46 experiments with 29 feature parameters

Microsilica MKU 85	Stone flour (from gravel)	Natural Sand $M_f = 2,0$	Artificial sand, M_f 2,3	Crushed stone mix 5–20 mm.Granite	Crushed stone with carbonate inclusion 5–10 mm.	Crushed stone with carbonate inclusion 10–20 mm.	Crushed gravel 5–10 mm.	Crushed gravel 10–20 mm.	Water	W/C	r	MC-TECHNIFLOW 170
0	0	0	540	0	393	917	0	0	150	0.357	0.292	0
0	0	0	639	0	373	922	0	0	150	0.358	0.33	0
0	0	0	560	0	393	917	0	0	150	0.356	0.3	0
0	0	0	560	0	393	917	0	0	150	0.357	0.3	0
0	0	0	560	0	393	917	0	0	160	0.37	0.3	0
0	0	0	560	0	393	917	0	0	160	0.37	0.3	0
0	26	0	534	0	446	818	0	0	170	0.4	0.3	0
0	0	0	534	0	446	818	0	0	170	0.37	0.3	0
27	0	0	533	0	538	716	0	0	183	0.38	0.298	0
0	0	0	600	0	466	0	0	818	175	0.39	0.32	0
0	0	0	600	0	0	0	466	818	167	0.36	0.32	0
0	0	0	534	0	0	0	393	917	160	0.37	0.29	0
0	0	0	560	0	0	0	466	818	160	0.38	0.3	0
0	0	0	645	0	0	0	775	424	160	0.38	0.35	0
0	0	0	583	0	0	0	454	806	156.8	0.348	0.3	0
0	0	0	505	0	0	0	466	818	164	0.36	0.296	0
0	0	0	505	0	0	0	466	818	158	0.35	0.296	0
0	0	0	505	0	0	0	466	818	156	0.35	0.283	0
0	0	0	505	0	0	0	385	899	165	0.37	0.283	0
0	0	0	534	0	0	0	385	899	165	0.37	0.293	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	560	0	0	0	385	895	149	0.33	0.3	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	545	0	0	0	370	880	150	0.33	0.3	0
0	0	0	560	0	0	0	385	895	165	0.37	0.3	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	605	0	0	0	495	735	156	0.35	0.33	0
0	0	0	534	0	446	818	0	0	170	0.37	0.3	0
0	0	0	505	0	0	0	466	818	158	0.35	0.296	0
0	0	0	534	0	0	0	385	899	165	0.37	0.293	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	560	0	0	0	385	895	156	0.35	0.3	0
0	0	0	605	0	0	0	495	735	156	0.35	0.33	0
0	0	0	600	0	0	0	466	818	175	0.37	0.47	0

(*continued on next page*)

(*continued*)

Microsilica MKU 85	Stone flour (from gravel)	Natural Sand $M_f = 2,0$	Artificial sand, M_f 2,3	Crushed stone mix 5–20 mm.Granite	Crushed stone with carbonate inclusion 5–10 mm.	Crushed stone with carbonate inclusion 10–20 mm.	Crushed gravel 5–10 mm.	Crushed gravel 10–20 mm.	Water	W/C	r	MC-TECHNIFLOW 170
0	0	0	600	0	0	0	466	818	167	0.36	0.47	0
0	0	0	534	0	0	0	393	917	160	0.37	0.41	0
0	0	0	560	0	0	0	466	818	160	0.37	0.44	0
0	0	0	560	0	0	0	385	895	156	0.35	0.44	0
0	0	0	545	0	0	0	495	735	150	0.33	0.44	0
0	0	0	560	0	0	0	385	895	156	0.35	0.44	0
0	0	0	605	0	0	0	495	735	145	0.35	0.49	0
0	0	610	0	1195	0	0	0	0	164	0.36	0.34	0.7
0	0	610	0	1195	0	0	0	0	164	0.36	0.34	0
0	0	0	671	1135	0	0	0	0	164	0.37	0.37	0.7
0	0	0	700	1075	0	0	0	0	160	0.36	0.39	0.7

MC-TECHNIFLOW 173	MC-POWERFLOW 7951	MURAPLASTFK 49	MC-POWERFLOW 7951 + CENTRAMENTRETARD 390	CENTRAMENTN10	MURAPLASTFK 88	CENTRAMENTN11	CENTRAMENTAIR 202	CENTRAMENTRETARD 390
0	0	0	0	0	1.8	0	1.2	0
0	0	0	0	0	1.8	0	1.2	0
0	0	0	0	0	1.8	0	1.2	0
0	0	0	0	0	1.8	0	1.2	0
0	0	0	0	0	2	0	1.5	0
0	0	0	0	0	2	0	1.5	0
0	0	0	0	0	2.2	0	1.2	0
0	0	0	0	0	2.2	0	1.2	0
0	0	0	0	0	2.4	0	1.3	0
0	0	0.6	0	0	0	0	0.51	0
0	0	0.5	0	0	0	5.64	0	0
0	0	0	0	0	0.6	0	0.35	0
0	0	0	0	0	0.6	0	0.35	0
0	0	0	0	0	0.6	0	0.35	0
0	0.7	0	0	0	0	0	0.49	0
0	0.5	0	0	0	0	0	0.44	0
0	0.44	0	0	0	0	0	0.4	0
0	0	0	0.54	0	0	0	0.44	0
0	0	0	0.4	0	0	0	0.44	0
0	0	0	0.4	0	0	0	0.4	0
0	0.44	0	0	0	0	0	0.42	0.9
0	0.44	0	0	0	0	0	0.42	1.35
0	0.44	0	0	0	0	0	0.42	1.8
0	0.44	0	0	0	0	0	0.13	1.35
0	0.44	0	0	0	0	0	0.13	1.35
0	0.44	0	0	0	0	0	0.47	1.2
0	0.44	0	0	0	0	0	0.47	1.2

(*continued on next page*)

(*continued*)

MC-TECHNIFLOW 173	MC-POWERFLOW 7951	MURAPLASTFK 49	MC-POWERFLOW 7951 + CENTRAMENTRETARD 390	CENTRAMENTN10	MURAPLASTFK 88	CENTRAMENTN11	CENTRAMENTAIR 202	CENTRAMENTRETARD 390
0	0.44	0	0	0	0	0	0.47	1.2
0	0	0	0	0	0.48	0	0.26	0
0	0.44	0	0	0	0	0	0.4	0
0	0	0	0.4	0	0	0	0.4	0
0	0.44	0	0	0	0	0	0.13	1.35
0	0.44	0	0	0	0	0	0.47	1.2
0	0.44	0	0	0	0	0	0.47	1.2
0	0	0.6	0	0	0	0	0.51	0
0	0	0.5	0	0	0	0	0.51	0
0	0.6	0	0	0	0	0	0.35	0
0	0.6	0	0	0	0	0	0.35	0
0	0.45	0	0	0	0	0	0.13	0
0	0.45	0	0	0	0	0	0.13	0
0	0.45	0	0	0	0	0	0.47	0
0	0.45	0	0	0	0	0	0.47	0
0	0	0	0	0	0	0	0.157	0
0.65	0	0	0	0	0	0	0.166	0
0	0	0	0	0	0	0	0.182	0
0	0	0	0	0	0	0	0.182	0

Estimated density	Density fact 10 min	Density fact 60 min	Density fact 90 min	air 15 min % device	air 60 min %	air 90 min %	Cd
2433.649	2385	2400	2405	3	2.6	2.3	3
2517.688	2432	2488	2497	9	7.3	5.9	2
2453.656	2389	2406	2409	4	3.5	3.2	3
2453.657	2396	2409	2411	4.5	4.1	3.4	3
2464.17	2259	2298	2369	9.5	7.3	6.2	3
2464.17	2376	2398	2432	6	5.1	4.3	3
2428.1	2368	2406	2406	4.9	4.3	3.7	4
2428.07	2371	2400	2420	5.2	4.3	3.9	4
2457.378	2388	2422	2442	6.5	5.2	4.7	3
2530.82	2462	2480	2480	4	3.4	3	3
2527.82	2492	2500	2520	2	1.8	1.6	3
2435.61	2374	2389	2410	5	4.2	3.4	4
2435.63	2376	2398	2417	4.7	4.1	3.5	4
2435.68	2386	2410	2428	3.7	3.3	2.9	2.7
2451.638	2290	2350	2410	8	7	6	7
2404.596	2208	2279	2358	9	7.1	5.9	7
2398.486	2395	2442	2489	5.7	4.8	3.9	3.7
2396.613	2320	2358	2397	7.5	6.3	5.2	15
2405.493	2345	2386	2425	8	6.8	5.2	5
2434.463	2397	2408	2423	5.9	5.2	4.7	5.5
2448.41	2178	2241	2289	7	6.2	5.3	7.5

(*continued on next page*)

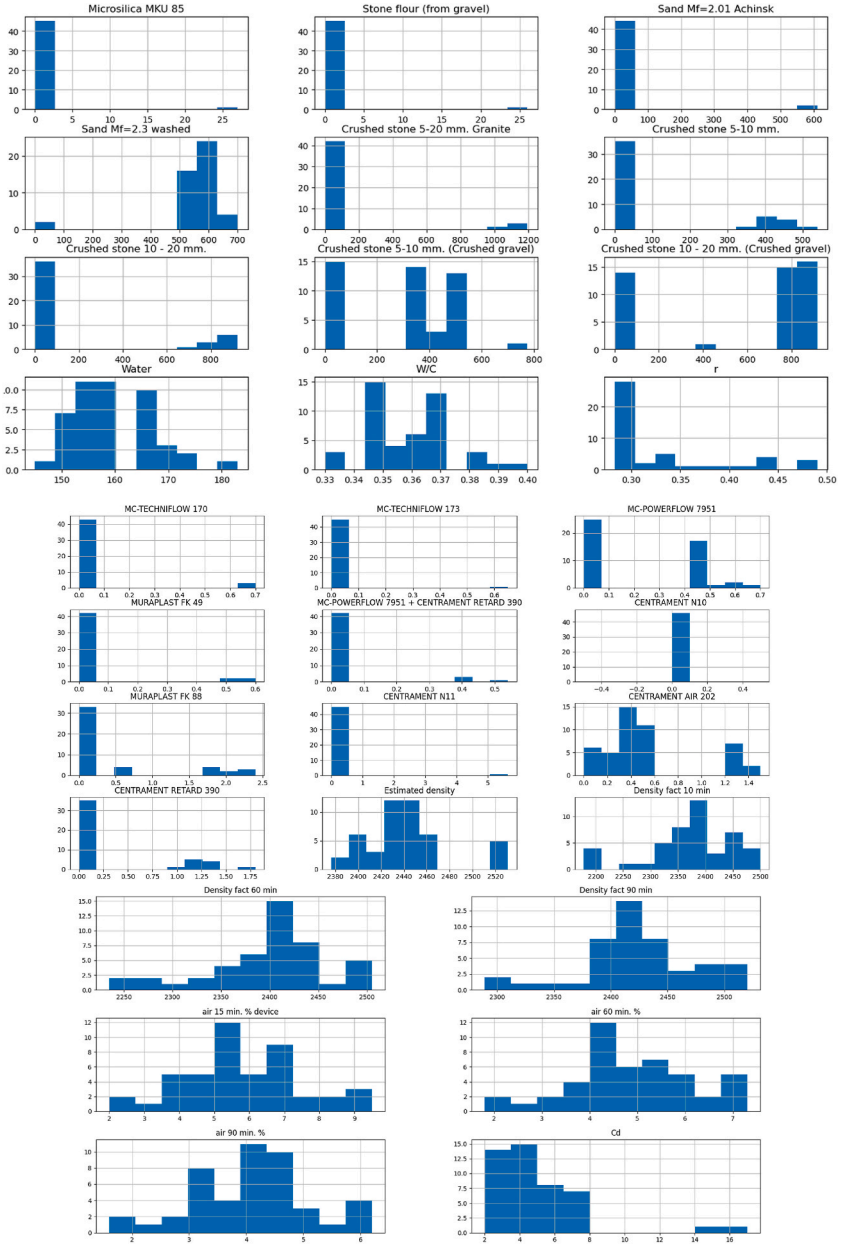
(*continued*)

Estimated density	Density fact 10 min	Density fact 60 min	Density fact 90 min	air 15 min % device	air 60 min %	air 90 min %	Cd
2448.86	2203	2235	2308	6.8	6.1	5.2	5.5
2442.29	2201	2264	2334	5.3	4.7	4.1	6.5
2448.57	2322	2363	2404	7	5.9	4.8	3.9
2397.55	2352	2428	2428	6.5	5.2	3.9	2.5
2457.78	2368	2374	2390	7.5	6	4.8	17
2448.76	2360	2379	2398	5.9	5.4	4.7	6
2443.79	2315	2334	2396	6.2	5.5	4.7	7
2425.41	2368	2382	2396	5.2	4.5	4.1	4
2398.486	2396	2442	2442	5.7	4.8	3.9	3.7
2434.463	2397	2405	2428	5.9	5	4.2	5.5
2448.57	2322	2363	2404	7	5.9	4.8	3.9
2448.76	2360	2379	2420	5.1	4.3	4	6
2443.79	2315	2334	2388	5	4.4	3.9	7
2530.95	2495	2498	2520	4	3.6	3	3
2522.84	2500	2505	2519	2	1.8	1.7	3
2435.73	2450	2446	2460	5	4.3	3.6	4
2435.76	2455	2440	2455	4.7	4.1	3.6	4
2447.37	2450	2440	2420	7	5.9	4.8	3.9
2376.35	2384	2398	2410	6.5	5.2	3.9	2.5
2447.71	2441	2444	2452	4.4	3.9	3.3	6
2431.76	2432	2401	2430	4.2	3.7	3.1	7
2420.557	2468	2452	2478	5.7	4.4	3.1	4.5
2420.516	2467	2440	2475	5.2	4	2.8	4
2411.622	2438	2416	2438	5.3	5	4.7	5.5
2376.632	2420	2400	2415	6.5	5.6	4.8	3.5

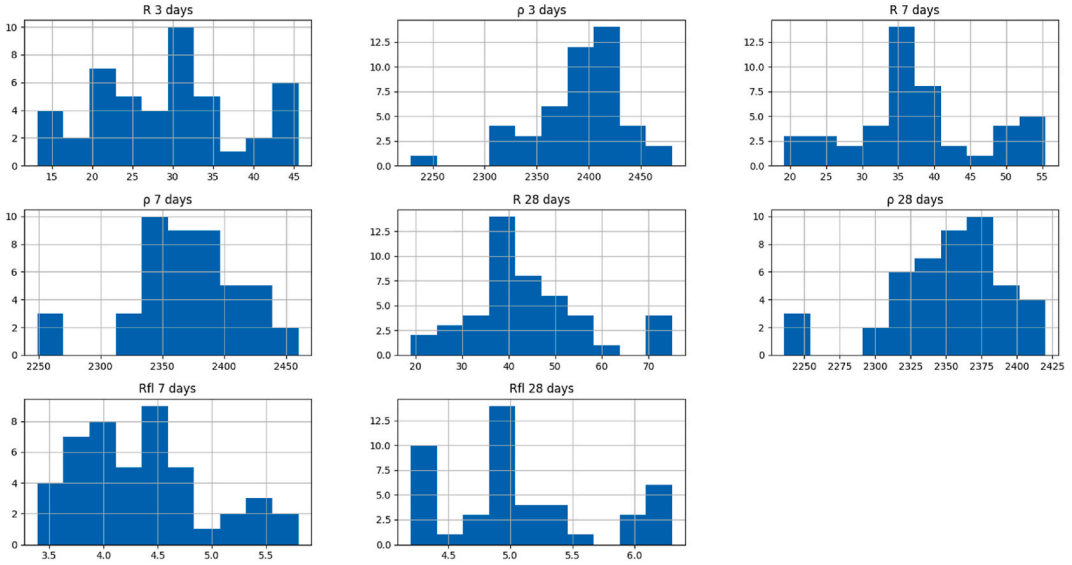
Appendix B. Dataset of 46 experiments with 8 property parameters

R 3 days	ρ 3 days	R 7 days	ρ 7 days	R 28 days	ρ 28 days	R_{fl} 7 days	R_{fl} 28 days
27.6	2333	31.9	2354	36.7	2310	3.7	5.3
16.6	2229	20	2267	19	2236	3.6	5.2
32.7	2428	36.7	2383	37	2368	3.9	5.6
31.9	2418	35.8	2429	38	2395	3.7	5.2
16.2	2364	19.2	2355	22	2354	3.8	5.3
20.2	2394	27.3	2370	34	2362	4.7	5.98
22.7	2411	33.8	2369	35.5	2370	4.45	4.96
23.3	2390	33.8	2355	37.3	2360	5.36	6.3
23.7	2420	44.1	2428	51.2	2351	5.08	6.05
13.2	2410	36	2384	49.9	2420	4	4.54
28.35	2465	40.3	2460	50.3	2410	3.87	4.74
24.7	2405	34.7	2384	37.8	2376	3.78	4.4
21	2355	33.8	2414	36	2369	3.73	4.2
22	2376	35.6	2373	38.7	2375	4.18	5
23.4	2320	25	2331	32.1	2347	4.2	5.1
18.8	2310	22.5	2318	27	2336	4.3	4.9
29.3	2426	40.1	2428	35.5	2359	4.1	5
32	2480	38	2460	43	2420	3.98	4.86
20.2	2318	31	2376	37	2320	4.38	4.79
13.8	2337	23.5	2344	28.9	2335	4.41	4.98
20	2318	31.2	2249	36.8	2246	4.46	4.91
22.1	2370	33.6	2253	39.3	2236	4.41	4.88
30.1	2421	41.5	2329	46.9	2325	4.36	4.92
41.2	2440	55.4	2347	57.7	2341	4.46	4.98
39.1	2416	50.8	2371	56.1	2350	4.64	5.4
31.4	2404	37	2385	42.4	2319	4.6	5.3
32.3	2427	39.7	2401	45	2380	4.01	4.4
30.6	2401	36	2391	39.6	2375	3.39	4.3
23.3	2390	28.8	2355	37.3	2360	5.51	6.3
29.3	2426	40.1	2428	47	2393	5.1	5.9
13.8	2337	23.5	2344	30.1	2301	4.2	4.7
44.6	2397	55.4	2347	56.8	2339	4.44	4.95
33.3	2369	40.9	2348	45	2310	4.04	4.4
31.5	2355	36.9	2338	39.6	2306	3.39	4.3
34.1	2386	47	2344	50.1	2336	4.5	5.1
33.2	2385	40.3	2353	50.3	2337	4.9	5
29.8	2392	34.7	2347	43.2	2319	4.02	4.4
30.1	2395	33.8	2365	41.8	2328	3.7	4.2
45.6	2423	55.4	2392	57.2	2362	4.69	4.88
42.2	2419	50.8	2396	58.9	2376	4.83	5.02
33.1	2401	40.9	2388	45	2369	4.1	4.4
30.1	2388	36.9	2374	39.6	2366	3.56	4.3
44.8	2449	52	2421	74	2409	5.8	6.12
45.1	2441	51	2417	75	2401	4.35	6.3
44.2	2428	50.1	2415	72	2398	5.6	6.2
44.4	2431	52.8	2413	71.3	2396	5.5	6.1

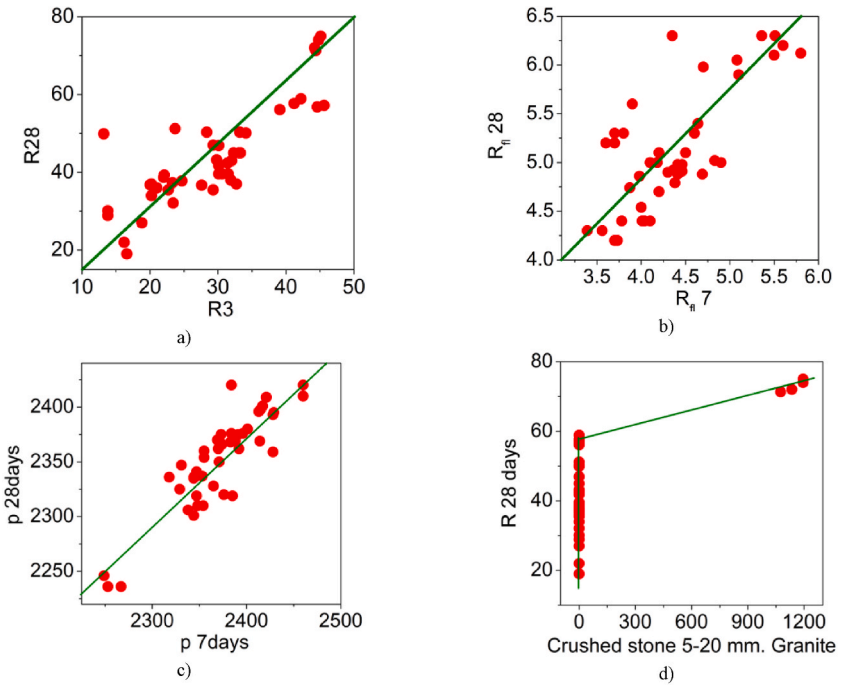
Appendix C. The statistic plots for all features and properties of experiments under consideration



Appendix D. The statistic plots for properties of experiments under consideration



Appendix E. Plots which depicts linear correlations between: (a) "R3 days" and "R28 days"; (b) "R_{fl} 7 days" and "R_{fl} 28 days"; (c) "ρ 3 days" and "ρ 28 days". Additional plot (d) depicts correlation between feature "Crushed stone 5–20 mm. Granite" and "R 28 days" property



References

[1] S. Janaki Raman, Shanmugasundaram: utilizing and optimizing waste resources in paver block, Lecture Notes in Civil Engineering 78 (2021) 407–416.
[2] M. Linek, P. Nita, W. Zebrowski, P. Wolka, Influence of operating media on the parameters of cement concrete intended for airfield pavements, J. Konbin 49 (4) (2020) 103–126.
[3] C. Zhou, G. Lan, P. Cao, C. Tang, Q. Cao, Y. Xu, D. Feng, Impact of freeze-thaw environment on concrete materials in two-lift concrete pavement, Construct. Build. Mater. 262 (2020), 120070.
[4] E. Melese, H. Baaj, S. Tighe, Fatigue behaviour of reclaimed pavement materials treated with cementitious binders, Construct. Build. Mater. 249 (2020), 118565.
[5] J. Skarkova, Concrete pavements in the Czech republic, Cem.Appl. 5 (2015) 102–107.
[6] H. Garrecht, C. Baumert, S. Hampel, W. Lisin, P. Lazik, Challenges of uniform production of road concretes and possibilities of rheology-based mixing process control, Beton- Stahlbetonbau 12 (114) (2019) 888–898.
[7] B.A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: new insights from statistical analysis and machine learning methods, Cement Concr. Res. 115 (2019) 379–388.
[8] P. Ziolkowski, M. Niedostatkiewicz, S.B. Kang, Model-based adaptive machine learning approach in concrete mix design, Materials 14 (7) (2021) 1661.
[9] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, J. Zhu, Explainable AI: a brief survey on history, research areas, approaches and challenges, in: Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8, Springer International Publishing, 2019, pp. 563–574.
[10] I.S. Markham, R.G. Mathieu, B.A. Wray, Kanban setting through artificial intelligence: a comparative study of artificial neural networks and decision trees, Integrated Manuf. Syst. 11 (4) (2000) 239–246.
[11] W.B. Chaabene, M. Flah, M.L. Nehdi, Machine learning prediction of mechanical properties of concrete: critical review, Construct. Build. Mater. 260 (2020), 119889.
[12] A. Ashrafian, M.J. Taheri Amiri, P. Masoumi, M. Asadi-shiadeh, M. Yaghoubi-chenari, A. Mosavi, N. Nabipour, Classification-based regression models for prediction of the mechanical properties of roller-compacted concrete pavement, Appl. Sci. 10 (11) (2020) 3707.
[13] T.K. Ho, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, IEEE, 1995, pp. 278–282.
[14] L.P. Coelho, W. Richert, M. Brucher, Building Machine Learning Systems with Python: Explore Machine Learning and Deep Learning Techniques for Building Intelligent Systems Using Scikit-Learn and TensorFlow, Packt Publishing Ltd., 2018, p. 406.
[15] R.Q. Zhu, D.L. Zeng, M.R. Kosorok, Reinforcement learning trees, J. Am. Stat. Assoc. 110 (512) (2015) 1770–1784.
[16] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[17] L.L. Custode, G. Iacca, Evolutionary Learning of Interpretable Decision Trees, IEEE Access, 2023.