Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00134

Research Article

Prediction of compressive strength class of concrete with dominance based rough set approach*

Ahmet TOPAL[1,2,], Nilgün GÜLER BAYAZIT[1], Yasemen UÇAN[1]**

[1]*Department of Mathematical Engineering, Yildiz Technical University, Istanbul, 34349, Türkiye*
[2]*Department of Mathematics, Istanbul Technical University, Istanbul, 34469, Türkiye*

ARTICLE INFO

Article history
Received: 22 November 2021
Revised: 08 February 2022
Accepted: 12 March 2022

Keywords:
Dominance Based Rough Set; Decision Rules; Compressive Strength Of Concrete; Decision-Support Model

ABSTRACT

Dominance based rough set approach is important in studies conducted with datasets containing uncertainty. In this study, a dataset consisting of 1030 samples obtained in the laboratory regarding compressive strength of concrete has been considered. The decision attribute, which has continuous values, has been made discrete for applying dominance relation. In order to measure performance, samples in the dataset have been divided into two groups: the training set and the testing set. This process has been done in a way that corresponds to the distribution of each class within the dataset. On the other hand, since there is a class which has more or less samples than the others, synthetic data generation has been done with Synthetic Minority Oversampling Technique (SMOTE) in order to handle the between-class imbalance problem and equalize the number of samples in the classes. As a result, the training set has been made perfectly balanced. A decision-support model which extracts "if… then…" exact decision rules has been designed to be used in determining the quality or compressive strength of the concrete samples by using dominance based rough set approach. Performance of these rules on the testing set through the confusion matrix has been discussed. The experimental results show that performance of the exact decision rules induced by the dominance rough set approach on the testing set is significant.

Cite this article as: Topal A, Güler Bayazıt N, Uçan Y. Prediction of compressive strength class of concrete with dominance based rough set approach. Sigma J Eng Nat Sci 2023;41(6):1088−1095.

INTRODUCTION

Determining compressive strength of concrete is an important task in the concrete-structure industry. The most considerable reason is to better understand the nature of the concrete and how the mixture will be optimized [1]. In 1918, the water-cement ratio (w/c) regarding compressive strength of concrete presented by Abrams [2] appeared to be quite a useful and important advancement in this field. According to this ratio, compressive strength of concrete decreases as w/c ratio increases, and increases as this ratio decreases. However, this is not totally true; experimental results [3] show that concretes with the same w/c ratio can

*A different version of the article was presented at the 22nd National Mechanical Congress (2021) and published in the proceedings book.

Corresponding author.
**E-mail address: atopal@itu.edu.tr

This paper was recommended for publication in revised form by Regional Editor Abdullahi Yusuf

Published by Yıldız Technical University Press, İstanbul, Turkey

have different compressive strength. This situation also reveals that other ingredients in the concrete affect the compressive strength. Hence, it is essential to consider the different ingredients that concrete contains.

There are many studies in the literature on compressive strength of concrete. However, most of them are estimation studies based on regression [1,4-8]. Yeh [1] showed that the concrete compressive strength model based on the artificial neural network is better than the regression analysis model with respect to R^2 metric. Silva et al. [6] developed the Decision Tree and Random Forest models in order to predict the compressive strength of concrete, and compared the methods with regard to their predicting capabilities. Özturan et al. [7] designed five system models considering various input variables and applied Artificial Neural Network (ANN) models as well as multiple regression and Abrams' Law to each of them. Ahmad et al. [8] trained supervised learning algorithms such as Decision Tree, AdaBoost and Random Forest on 165 experimental data and assessed these methods comparatively on the remaining 42 samples using statistical indices. Unlike previous studies, in our study, the decision-support model has been designed using Dominance Based Rough Set Approach (DRSA) to decide compressive strength. As is well known, the decision-support model is more preferable than the regression-based model, because it gives more concrete and explainable results and is easy for anyone to interpret without technical knowledge.

Concrete compressive strength (*CCS*) attribute values have been converted from continuous form to discrete form in order to apply DRSA on the dataset. This discretization process has been carried out according to [9]. As a result of the discretization period, four different concrete classes- low, normal, medium and high- have been generated. In order to measure the performance of the decision support model on classifying the concrete samples, the testing set consists of 100 samples considering the distribution of classes in the dataset. However, in the dataset, the between-class imbalance problem arises due to the fact that one class has significantly fewer samples than the others. Because of this situation, the learning process becomes more difficult. Synthetic data generation, which equalizes the number of samples in each class, has been made based on the SMOTE algorithm to avoid the between-class imbalance problem. Exact decision rules have been extracted with DRSA on the perfectly balanced training set. Performance analysis of these decision rules has been discussed on the testing set.

This paper's key contribution is that it develops a decision-support system model based on DRSA to better understand the effect of ingredients in the concrete mix. For this purpose, the data [10] about compressive strength of concrete created by Chung Hua University Civil Engineering faculty member Prof. I-Cheng Yeh have been used. The rest of the paper is organized as follows: DRSA is explained in the second section. Section 3 summarizes the method about classifying samples in the testing set. Section 4 presents the experimental design and results. The last part of the paper presents the concluding remarks.

MATERIALS AND METHODS

The rough set theory was first introduced by Pawlak [11] as a new mathematical tool to deal with vagueness and uncertainty. Rough set methodologies can be applied as a component of hybrid solutions in data mining and machine learning. It is preferred especially in the following cases: discovering hidden information, extracting decision rules, feature selection, sample selection and performance on datasets with missing values. In addition, it can be said that it has extended application including areas such as machine learning, decision analysis, expert systems, decision-support systems and pattern recognition. Also, the rough set approach, based on data analysis methodologies, seems quite useful in economics, finance, medicine, signal and image processing, robotics and engineering [12-15].

Rough set of Pawlak [11] is based on the indiscernibility relation. However, this approach cannot deal with inconsistency originating in the dominance principle. Therefore, the rough set theory based on dominance relation instead of indiscernibility relation was developed by Greco et al. [16] (Dominance Based Rough Set Approach). DRSA is an improvement of rough set theory for multi-criteria decision analysis (MCDA) [17,18].

Dominance Based Rough Set Approach

In this section, the mathematical background of our study is summarized [19-22]. A decision table is defined as $KT = \{U, A = C \cup D, V, f\}$ in DRSA, where U is a finite set of objects, C and D are the condition and decision attribute sets, respectively, V is the set of values taken by all attributes (i.e. Let V_a be the set of values taken by attribute a, then $V = \bigcup_{a \in A} V_a$), and $f: U \times A \to V$ is called the information function. For the sake of simplicity, we suppose that $D = \{d\}$.

Let O_a be the outranking relation on U with respect to attribute a. xO_ay refers "object x is at least as good as object y" on attribute a. Furthermore, let $\delta = \{1, 2, \cdots, r\}$ be the set of different values taken by decision attribute d, then decision attribute d makes a partition of the universal set U into r distinct classes such that $Cl = \{Cl_t, t \in \delta\}$, where $Cl_t = \{x_i \in U: f(x_i, d) = t\}$. Each $x_i \in U$ belongs to one and only one $Cl_t \in Cl$. The downward and upward unions of these classes Cl_t are defined as $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$ $(t = 1, \cdots, r-1)$ and $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$ $(t = 2, \cdots, r)$, respectively.

In DRSA, if xO_ay for all $a \in P \subseteq C$, then "object x is dominating object y" and is denoted by xD_py. The relation D_p satisfies reflexive and transitive properties, i.e. D_p is a preorder relation. Given an element $x_i \in U$ and a set $P \subseteq C$, P-dominated and P-dominating sets of object x_i are defined as $D_P^-(x_i) = \{x_j \in U: x_i D_P x_j\}$ and $D_P^+(x_i) = \{x_j \in U: x_j D_P x_i\}$, respectively. In other words, the P-dominated set of x_i contains the objects which are at most as good as itself with respect to all attributes of the set P. Similarly, the P-dominating set of x_i contains the objects which are at least as good as itself with respect to all attributes of the set P.

Lower and upper approximations of the upward union of the classes Cl_t are defined as follows by using set D_P^+:

$$\underline{P}(Cl_t^{\geq}) = \{x_i \in U: D_P^+(x_i) \subseteq Cl_t^{\geq}\},$$
$$\overline{P}(Cl_t^{\geq}) = \cup_{x_i \in Cl_t^{\geq}} D_P^+(x_i),\ (t = 2, \cdots, r). \quad (1)$$

In a similar way, lower and upper approximations of the downward union of the classes Cl_t are defined as follows by using set D_P^-:

$$\underline{P}(Cl_t^{\leq}) = \{x_i \in U: D_P^-(x_i) \subseteq Cl_t^{\leq}\},$$
$$\overline{P}(Cl_t^{\leq}) = \cup_{x_i \in Cl_t^{\leq}} D_P^-(x_i),\ (t = 1, \cdots, r-1). \quad (2)$$

P-boundaries for $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as follows:

$$BND_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}),\ (t = 2, \cdots, r), \quad (3)$$

$$BND_P(Cl_t^{\leq}) = P(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}),\ (t = 1, \cdots, r-1). \quad (4)$$

In DRSA, a decision table can be seen as a set of decision rules in the form of IF {elementary condition(s)}, THEN {decision} [22]. Here, the condition part refers to the value taken by one or more condition attributes and the decision part refers to the assignment to one or more decision classes. Decision rules can be categorized into three groups: certain decision rules extracted from lower approximations, possible decision rules extracted from upper approximations, and approximate decision rules extracted from boundary regions. In our study, exact decision rules extracted from lower approximations have been considered, and the following are their syntaxes:

- Exact $Dec_{\geq}$-rules: **IF** $f(x, a_1) \geq v_{a_1}$ and $f(x, a_2) \geq v_{a_2}$ and … $f(x, a_P) \geq v_{a_P}$, **THEN** $x \in Cl_t^{\geq}$.
- Exact $Dec_{\leq}$-rules: **IF** $f(x, a_1) \leq v_{a_1}$ and $f(x, a_2) \leq v_{a_2}$ and … $f(x, a_P) \leq v_{a_P}$, **THEN** $x \in Cl_t^{\leq}$.

where $P = \{a_1, a_2, \cdots, a_p\} \subseteq C$, $(v_{a1}, v_{a2}, \cdots, v_{ap}) \in Va_1 \times Va_2 \times \cdots \times V_{ap}$ and $t \in \delta$.

Classification Method for Samplesin Testing Set

In this section, the classification method that determines the decision class of any test samples has been summarized in the context of exact decision rules obtained from the transformed decision tables [23,24]. For this, let x be any test sample and σ_x be the set of rules which covers the object x. In order to decide the class of the object x, score calculations should be done considering the following three situations:

1. The object x is not covered by any rules, i.e. $|\sigma_x| = 0$.
2. The object x is covered by only one rule, i.e. $|\sigma_x| = 1$.
3. The object x is covered by more than one rule, i.e. $|\sigma_x| > 1$.

In Situation 1, assignment of x to any decision class has the same probability. In other words, let $\rho(x, Cl_t)$ be the score indicating that object x can belong to class Cl_t. In this case $\rho(x, Cl_t) = \frac{1}{r}$.

For the second situation, score calculation is done by looking at the decision part of the rule. In this context, the expression "$THEN\ Dec.Attr. \leq 0$" will be used to denote assignment to class $\neg Cl_t$, whereas "$THEN\ Dec.Attr. \geq 1$" will be used to denote assignment to Cl_t.

i. If object x is covered by rule r having decision part "$THEN\ Dec.Attr. \geq 1$", then the following formula is used to calculate the score:

$$\rho_{r_1}(x, Cl_t) = \frac{|\|\varphi_{r_1}\| \cap Cl_t|^2}{|\|\varphi_{r_1}\|| |Cl_t|}. \quad (5)$$

ii. If object x is covered by rule r having decision part "$THEN\ Dec.Attr. \leq 0$", then the score is calculated using the formula below:

$$\rho_{r_0}(x, \neg Cl_t) = \frac{|\|\varphi_{r_0}\| \cap \neg Cl_t|^2}{|\|\varphi_{r_0}\|| |\neg Cl_t|}. \quad (6)$$

where $\|\varphi_{r_1}\|$ and $\|\varphi_{r_0}\|$ represent the set of objects covered by rules having a decision part "$THEN\ Dec.Attr. \geq 1$" and "$THEN\ Dec.Attr. \leq 0$", respectively, $|\cdot|$ denotes the cardinality of the sets, and the set $\neg Cl_t$ is the complement of Cl_t, i.e. $\neg Cl_t = U - Cl_t$.

Depending upon the decision part of the rule, the final score for each Cl_t is either $\rho^+(x, Cl_t) = \rho_{r_1}(x, Cl_t)$ or $\rho^-(x, Cl_t) = \rho_{r_0}(x, \neg Cl_t)$. As a result of the calculations, object x is assigned to the decision class that has the highest score for $\rho^+(x, Cl_t)$, while it is assigned to the decision class that has the lowest score for $\rho^-(x, Cl_t)$.

For more than one rule, Situation 3 is an extension of Situation 2. By using the decision part of the rules, the set of rules covering the object x is separated into two subsets: those that propose assignment of x to Cl_t and those that propose assignment of x to $\neg Cl_t$. We will denote the assignment of x to Cl_t as "$THEN\ Dec.Attr. \geq 1$" and assignment of x to $\neg Cl_t$ as "$THEN\ Dec.Attr. \leq 0$" in the decision part of the rules. Similar to (i) and (ii), the scores $\rho_{r_1}(x, Cl_t)$ and $\rho_{r_0}(x, \neg Cl_t)$ are calculated like this:

$$\rho_{r_1}(x, Cl_t) = \frac{|(\|\varphi_{r_1}^1\| \cap Cl_t) \cup \cdots \cup (\|\varphi_{r_1}^k\| \cap Cl_t)|^2}{|\|\varphi_{r_1}^1\| \cup \cdots \cup \|\varphi_{r_1}^k\|| |Cl_t|}, \quad (7)$$

$$\rho_{r_0}(x, \neg Cl_t) = \frac{|(\|\varphi_{r_0}^1\| \cap \neg Cl_t) \cup \cdots \cup (\|\varphi_{r_0}^l\| \cap \neg Cl_t)|^2}{|\|\varphi_{r_0}^1\| \cup \cdots \cup \|\varphi_{r_0}^l\|| |\neg Cl_t|}. \quad (8)$$

where $\|\varphi_{r_1}^i\|$ indicates the set of elements covered by the i^{th} of rules having the decision part "$THEN\ Dec.Attr. \geq 1$", and $\|\varphi_{r_0}^{j}\|$ indicates the set of elements covered by the j^{th} of rules having the decision part "$THEN\ Dec.Attr. \leq 0$".

In Situation 3, the following formula is used to determine the final score for each decision class:

$$\rho(x, Cl_t) = \rho^{+}(x, Cl_t) - \rho^{-}(x, Cl_t) = \rho_{r_1}(x, Cl_t) - \rho_{r_0}(x, \neg Cl_t). \quad (9)$$

So, object x is assigned to the decision class with the highest value of $\rho(x, Cl_t)$.

RESULTS AND DISCUSSION

Data on compressive strength of concrete created by Yeh [10], with consideration of various components, have been used in this study in which we examine the strength of the concrete samples through the rules extracted on the basis of the dominance based rough set approach. This dataset consists of 1030 concrete samples and 9 continuous attributes that describe each concrete sample in terms of various characteristics. Eight of these nine attributes listed in Table 1 correspond the materials that comprise the concrete mix, and the remaining one is the compressive strength.

Table 1. Decision and condition attributes in the dataset

Attribute Name	Attribute Type
Cement (*kg/m*^{3})	Condition attr.
Blast Furnace Slag (*kg/m*^{3})	Condition attr.
Fly Ash (*kg/m*^{3})	Condition attr.
Water (*kg/m*^{3})	Condition attr.
Super Plasticizer (*kg/m*^{3})	Condition attr.
Coarse Aggregate (*kg/m*^{3})	Condition attr.
Fine Aggregate (*kg/m*^{3})	Condition attr.
Age (Day)	Condition attr.
CCS (MPa)	*Decision attr.*

The first eight attributes in Table 1 specify the condition attributes, whereas the final attribute, *CCS*, specifies the decision attribute. It is necessary that all values of the decision attribute should be discrete with the aim of applying dominance relation. For this reason, the continuous *CCS* attribute is discretized as in Table 2. As a result of the discretization process, it is observed that there are 181 (17.57 %) low, 436 (42.33 %) normal, 374 (36.31 %) medium, and 39 (3.78 %) high types of concrete within the dataset. In order to analyze the performance of the exact decision rules extracted in DRSA, the dataset has been divided into two parts: the testing set and the training set. In the testing set, there are 17 low, 42 normal, 36 medium and 5 high concrete type samples in parallel with the number of samples in each class. The training set is made up of the remaining 930 samples.

Table 2. Grouping of concrete based on compressive strength [9]

Concrete Type	Compressive Strength (MPa)
Low Strength	0-19
Normal Strength	20-39
Medium Strength	40-69
High Strength	70-119
Ultra-High Strength	120-1000

Since the attributes in DRSA have preference ordered domains, it is required that the preference order for each attribute in the dataset should be decided. An attribute is called a *gain* type attribute if higher values of its domain are better, and a *cost* type attribute if lower values of its domain are better. Deciding the preference order for an attribute can be done by a domain expert, but without a domain expert it should be handled with the approach presented in [23]. In this study, this approach has been undertaken in deciding the preference order of attributes using the following steps for transforming the original decision table:

1. The decision table is copied the same number of times as the different values that the decision attribute takes. That is, a decision table for each class (low, normal, medium and high) is replicated from the original one.
2. Decision attribute values in each replicated decision table are modified as 0 or 1 depending on whether the value is from the class which the decision table is generated or not. For example, low values in the decision field are modified by 1, and all others are modified by 0 when the decision table generated from the low class is taken into account.
3. In the last stage, decision tables are set up so that they have two of each condition attribute, one is *gain* and the other is *cost*

In Table 3, 5-sample piece of the dataset is provided in the form of a decision table and Table 4 shows an example transformation for the class medium using the three steps listed above. The *gain* type and *cost* type attributes are represented by the symbols (↑) and (↓) in Table 4, respectively.

The issue of between-class imbalance occurs when one class has numerically more samples than the others. Such a situation leads to the dataset becoming unbalanced without an equal number of samples in each class. A class which has more samples is called majority class, and a class which has less samples is called minority class. In this paper, we chose *normal strength* class as the majority and *high strength* class as the minority by analyzing the superiority of classes from the standpoint of sample numbers. The number of samples in the majority class divided by the number of samples in the minority class simply states the unbalance ratio [25]. The unbalance ratio is calculated as $\frac{436}{39} = 11.17$.

Table 3. View of 5-sample piece of the decision table

	Condition Attributes			Decision Attribute
Sample No	**Cement (kg/m³)**	**Blast Furnace Slag (kg/m³)**	**… …**	***CCS* (MPa)**
1	540	0	… …	High
2	332.5	142.5	… …	Medium
3	266	114	… …	Medium
4	380	95	… …	Normal
5	139.6	209.4	… …	Low

Table 4. An example transformation of the original table piece shown in Table 3 for class medium

	Condition Attributes					Decision Attribute
Sample No	***Cement-1* (*kg/m³*) (↑)**	***Cement-2* (*kg/m³*) (↓)**	***Blast Furnace Slag-1* (*kg/m³*) (↑)**	***Blast Furnace Slag-2* (*kg/m³*) (↓)**	**…**	***CCS* (MPa) (↑)**
1	540	540	0	0	…	0
2	332.5	332.5	142.5	142.5	…	1
3	266	266	114	114	…	1
4	380	380	95	95	…	0
5	139.6	139.6	209.4	209.4	…	0

Datasets having an unbalance ratio greater than 4 are seen as extremely unbalanced and then it is difficult to learn the patterns in the minority [26]. Therefore, the SMOTE algorithm (see Figure 1) proposed by Chawla et al. [27] was used to generate new data points and avoid unequal distribution between the classes in the extremely unbalanced training dataset.

After the preprocessing steps, exact decision rules were extracted from each transformed decision table (see Table 4) with the help of DomLem [19] algorithm on JAMM [28] software tool and the performance analysis of the decision support system on the testing set was assessed with the Python program that we developed. Sample exact decision rules derived from each transformed decision table are illustrated in Figure 2, and Table 5 shows the total number of exact decision rules.

The greatest advantage provided by a rule-based decision algorithm is that it can be simply understood and interpreted by someone who lacks technical knowledge and skills. For example, rule number 7 shows that concrete with *Cement* quantity less than or equal to 273 kg/m^3 does not have high strength. In mathematical notation, we can simply write this rule as $Cement \leq 273 \Rightarrow CCS \in \{Low, Normal, Medium\}$. Another example is rule number 2, which states that concrete with *Fly Ash* content of 133.6 to 134 kg/m^3 has low strength. Analogously, this rule can be written as $Fly\ Ash \in [133.6, 134] \Rightarrow CCS \in \{Low\}$ in mathematical form. Another advantage of rule-based decision algorithms is the ability to recognize what the main factors are in an estimating process and to evaluate whether the mechanism for a certain sample condition performs correctly or not.

1. K-minority class nearest neighbor of each sample in the minority class is identified.
2. Number of samples to be generated is shared equally among the minority class samples. If the number of samples to be generated is less than the number of minority class samples, a random selection is made to determine which minority samples will be used in the data generation phase.
3. Depending upon the amount of oversampling required in Step 2, a random selection is made between each minority sample's K-minority class nearest neighbor.
4. Synthetic data are obtained along the line segments between the minority class sample and its selected K-minority class nearest neighbor samples. In other words, it is a convex combination of two points on n-dimensional attribute(feature) space.

Figure 1. Steps in the SMOTE algorithm.

In Figure 2, the decision part of rules will be interpreted that $CCS \leq 0$ denotes assignment to outside the class, whereas $CCS \geq 1$ denotes assignment to the class itself.

A confusion matrix was constructed to observe the performance of the decision support model on the testing set. Since the state of the sample assigned to more than one class would be ambiguous, two samples in the testing set have been excluded from the confusion matrix. Due to the extremely unbalanced dataset, a macro-f1 score, which is more sensitive to the changes in the minority class, was utilized as a performance metric instead of overall accuracy.

In a confusion matrix, rows indicate the true classes of the samples in the testing set, while columns indicate the predicted classes. For instance, when examined on the basis of row, there are 34 samples in class medium, with 25 of them correctly labeled as medium and the remaining

Decision table transformed for class Low:

1. IF Age-2 ≥ 7 & Water-1 ≤ 157.9 THEN CCS ≤ 0

2. IF Fly Ash-1 ≥ 133.6 & Fly Ash-2 ≤ 134 THEN CCS ≥ 1

Decision table transformed for class Normal:

3. IF Age-2 ≥ 91 & Blast Furnace Slag-2 ≥ 15 & Blast Furnace Slag-1 ≤ 236 THEN CCS ≤ 0

4. IF Water-1 ≥ 192.7 & Water-2 ≤ 192.94 & Age-2 ≤ 90 & Age-1 ≥ 7 THEN CCS ≥ 1

Decision table transformed for class Medium:

5. IF Fine Aggregate-1 ≤ 594 & Blast Furnace Slag-2 ≥142.8 THEN CCS ≤ 0

6. IF Coarse Aggregate-1 ≥ 446 & Coarse Aggregate-2 ≤ 450.1 & Super Plasticizer-2 ≤ 11.61 & Age-1 ≥ 7 THEN CCS ≥ 1

Decision table transformed for class High:

7. IF Cement-1 ≤ 273 THEN CCS ≤ 0

8. IF Water-2 ≤ 151 & Coarse Aggregate-1 ≥ 1120 THEN CCS ≥ 1

Figure 2. Examples of exact decision rule assigned to the class itself and outside the class for each transformed table.

Table 5. Number of exact decision rules extracted for decision tables

Decision Tables	Number of rules assigned to outside the class	Number of rules assigned to the class itself	Total number of rules
Low Strength	57	55	112
Normal Strength	120	117	237
Medium Strength	91	100	191
High Strength	31	24	55

9 incorrectly labeled as normal. Given that there are 36 medium class samples in the testing set, it is understood that 2 samples have been assigned to more than one class.

Macro-f1 score has been calculated as 85%. Therefore, it can be said that performance of exact decision rules extracted with DRSA on the testing set has been significant. While decision rules can almost correctly label samples in class low and class high, labelling a meaningful number of samples related to class medium as normal is a noteworthy finding.

Optimum amounts in a mixture can be determined for each concrete class, taking into account the score calculations described in Section 3. We will consider Situation 3, the most generalized score calculation, to determine the range of each component in the mixture. Due to the importance of high performance concrete in terms of industry, the interval of each component for high-strength concrete type has been obtained and given in Table 7. The probability of concrete belonging to the class high increases when the value $\rho^{+}(x, Cl_{High}) = \rho_{r_1}(x, Cl_{High})$ is as high as possible and the value $\rho^{-}(x, Cl_{High}) = \rho_{r_0}(x, \neg Cl_{High})$ is as low as possible. As a result, the sample desired to have high durability should include rules with a maximum number of *CCS* ≥ 1 decision parts and rules with a minimum number of

Table 6. Confusion matrix

		Predicted Classes			
		Low	Normal	Medium	High
Actual Classes	Low	14	3	0	0
	Normal	3	35	4	0
	Medium	0	9	25	0
	High	0	0	0	5

Table 7. Amount interval of each component in the mixture to obtain high performance concrete

Attribute name (unit)(abbr.)	Amount interval
Cement (kg/m^3) (a_1)	$a_1 \geq 531.419$
Blast Furnace Slag (kg/m^3) (a_2)	$189.249 \leq a_2 \leq 189.772$
Fly Ash (kg/m^3) (a_3)	a_3 does not any upper or lower bound
Water (kg/m^3) (a_4)	$149.348 \leq a_4 \leq 151.149$
Super Plasticizer (kg/m^3) (a_5)	$11.3064 \leq a_5 < 11.61$
Coarse Aggregate (kg/m^3) (a_6)	$908.923 < a_6 \leq 913.656$
Fine Aggregate (kg/m^3) (a_7)	$744.917 < a_7 \leq 753.647$
Age (*Day*) (a_8)	$a_8 > 29$

$CCS \leq 0$ decision parts among the decision rules extracted from the transformed decision table for high strength.

It is clearly seen that high performance concrete is not affected by the presence of *Fly Ash* component. Moreover, there is no upper limit for the components *Cement* and *Age*. This shows that the increase in *Cement* and *Age* after a certain value cannot impair the high durability of concrete.

CONCLUSION

Concrete, which is a composite material, has a complex structure due to its content. Understanding the complex structure of concrete is critical in order to model its compressive strength. Consequently, the aim of this study was to create a decision-support system that could estimate the compressive strength of concrete while taking into account the various components. Rule-based models, rather than regression models, indicate the impacts of different components on compressive strength in a clearer manner. For instance, there are no upper or lower limits for *Fly Ash* component. Obviously, this highlights the fact that high-performance concrete is independent of *Fly Ash*. Furthermore, some rules extracted from the decision table transformed for high strength can have one elementary condition including only the components *Water* or *Cement*. This indicates that a direct influence of *Water* and *Cement* on compressive strength is feasible, however, the effect of an ingredient other than these two is possible with other ingredients in the mix. Another finding is that high strength concrete is unlimited in the upper direction for the components *Cement* and *Age*. These are worthy results for high performance concrete. On the other hand, since there are few samples in the dataset, the capacity of the decision rules to generalize declines, resulting in ambiguous assignments due to a lack of learning. So, the vague situation associated with assigning a sample to more than one class is reduced thanks to synthetic data generation.

ACKNOWLEDGEMENTS

This research has been supported by Yildiz Technical University, Scientific Research Coordinatorship under project code FYL-2020-4001.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Yeh IC. Modeling of strength of high performance concrete using artificial neural networks. Cement Concrete Res. 1998;28:1797−1808. [CrossRef]

[2] Abrams DA. Design of concrete mixtures. Bull Struct Mater Res Lab Lewis Inst Chic. 1919;1:25.

[3] Popovics S. Analysis of concrete strength versus water-cement ratio relationship. ACI Mater J. 1990;87:517−29. [CrossRef]

[4] Yeh IC. Modeling Concrete Strength with Augment-Neuron Networks. J Mater Civ Eng. 1998;10:263−268. [CrossRef]

[5] Yeh IC. Analysis of strength of concrete using design of experiments and neural networks. J Mater Civ Eng. 2006;18:597−604. [CrossRef]

[6] Silva PF, Moita GF, Arruda VF. A Computational Method to Predict the Concrete Compression Strength Using Decision Trees and Random Forest. In: XLI Ibero-Latin-American Congress on Computational Methods in Engineering, Brazil; November 2020.

[7] Ozturan M, Kutlu B, Ozturan T. Comparison of concrete strength prediction techniques with artificial neural network approach. Build Res J. 2008:56;23−36.

[8] Ahmad M, Hu JL, Ahmad F, Tang XW, Amjad M, Iqbal MJ, et al. Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature. Materials. 2021;14:1983. [CrossRef]

[9] Sojobi AO, Aladegboye OJ, Awolusi TF. Green interlocking paving units. Constr Build Mater. 2018;173:600−614. [CrossRef]

[10] UCI Machine Learning Repository. Concrete Compressive Strength. Available at: http://archive.ics.uci.edu/ml/dataset/concrete+compressive+strength. Accessed March 30, 2021.

[11] Pawlak Z. Rough Sets. Int J Comput Inf Sci. 1982;11:341−356. [CrossRef]

[12] Pawlak Z. Rough Sets and Flow Graphs. Lect Notes Comput Sci. 2005;3641:1−11. [CrossRef]

[13] Cui N, Cui Q. A Rough-Set Based Approach to Predict Consumers' Brand Preference. In: Proceedings of the 2009 International Workshop on Intelligent Systems and Applications; 2009. Wuhan, China. p. 1−4. [CrossRef]

[14] Grzymala-Busse JW. Rule Induction. In: Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook. Boston: Springer; 2009. p. 249–265. [CrossRef]

[15] Liou JJH, Tzeng G. A Dominance-based Rough Set Approach to Customer Behavior in the Airline Market. Inf Sci. 2010;180:2230–2238. [CrossRef]

[16] Greco S, Matarazzo B, Slowinski R. Rough approximation by dominance relations. Int J Intell Syst. 2002;17:153–171. [CrossRef]

[17] Greco S, Matarazzo B, Slowinski R. Multicriteria classification by dominance-based rough set approach. Università degli Studi. 2000:1–14.

[18] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multicriteria decision analysis. Eur J Oper Res. 2001;129:1–47. [CrossRef]

[19] Greco S, Matarazzo B, Slowinski R, Stefanowski J. An algorithm for induction of decision rules consistent with the dominance principle. In: Rough Sets and Current Trends in Computing, Second International Conference, RSCTC 2000, Banff, Canada, October 16-19, 2000, Revised Papers, 304–313. [CrossRef]

[20] Slowinski R, Greco S, Matarazzo B. Rough set and rule based multicriteria decision aiding. Pesqui Oper. 2012;32:213–270. [CrossRef]

[21] Chakhar S, Ishizaka A, Labib A, Saad I. Dominance-based rough set approach for group decisions. Eur J Oper Res. 2016;251:206–224. [CrossRef]

[22] Hu Q, Chakhar S, Siraj S, Labib A. Spare parts classification in industrial manufacturing using the dominance-based rough set approach. Eur J Oper Res. 2017;262:1136–1163. [CrossRef]

[23] BŁaszczyński J, Greco S, SŁowiński R. Inductive discovery of laws using monotonic rules. Eng Appl Artif Intell. 2012;25:284–294. [CrossRef]

[24] BŁaszczyński J, Greco S, SŁowiński R. Multi-criteria classification-A new scheme for application of dominance-based decision rules. Eur J Oper Res. 2007;181:1030–1044. [CrossRef]

[25] Abdellatif S, Hassine MAB, Yahia SB, Bouzeghoub A. ARCID: a new approach to deal with imbalanced datasets classification. In: International Conference on Current Trends in Theory and Practice of Informatics; 2018. p. 569–580. [CrossRef]

[26] Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell. 2016;5:221–232. [CrossRef]

[27] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–357. [CrossRef]

[28] JAMM Software Tool. Available at: http://idss.cs.put.poznan.pl/site/75.html. Accessed April 12, 2021.