Automatic categorization of chloride migration into concrete modified with CFBC ash

Maria Marks, Daria Jóźwiak-Niedźwiedzka and Michał A. Glinicki*

Institute of Fundamental Technological Research, Polish Academy of Sciences Pawińskiego 5 B, 02-106 Warszawa, Poland

(Received January 13, 2010, Revised June 14, 2011, Accepted January 5, 2012)

Abstract. The objective of this investigation was to develop rules for automatic categorization of concrete quality using selected artificial intelligence methods based on machine learning. The range of tested materials included concrete containing a new waste material - solid residue from coal combustion in fluidized bed boilers (CFBC fly ash) used as additive. The rapid chloride permeability test - Nordtest Method BUILD 492 method was used for determining chloride ions penetration in concrete. Performed experimental tests on obtained chloride migration provided data for learning and testing of rules discovered by machine learning techniques. It has been found that machine learning is a tool which can be applied to determine concrete durability. The rules generated by computer programs AQ21 and WEKA using J48 algorithm provided means for adequate categorization of plain concrete and concrete modified with CFBC fly ash as materials of good and acceptable resistance to chloride penetration.

Keywords: concrete durability; chloride ions migration; circulated fluidized bed combustion fly ash (cfbc fly ash); machine learning; classification rules; database.

1. Introduction

Increasing the use of fly ash in cement and concrete industry can considerably enhance the environmental friendliness of concrete production. Current practice for using fly ash as type II concrete additive according to EN 206-1 standard, does not cover the use of solid by-products resulting from advanced coal burning technologies, like Circulating Fluidized Bed Combustion (CFBC). This ‘clean coal technology’ for power production is used in several countries, e.g. Czech Republic, Estonia, France, Germany, Japan, Poland, USA, (Nowak 2003), China (Fu *et al.* 2008). The solid residue from coal combustion in fluidized bed boilers contains noncombustible mineral matter, sorbent material and unburned carbon (Giergiczny 2006). Mainly because of high sulfur content, high free lime content, high loss on ignition LOI and the lack of glassy phase CFBC ash does not meet the requirements defined by European standard EN 450-1 or in ASTM C618-03 in order to be used for cement or concrete production. The potential for using CFBC fly ash in concrete was recently investigated and the adequate strength and frost durability was revealed for selected kinds of CFBC fly ash used to replace 20% of cement mass in the binder (Glinicki and Zielinski 2009). Moreover, the efficient methods for selection of adequate CFBC fly ash to provide the long term durability of concrete are still required and possibility of 30-40% replacement is looked for.

* Corresponding author, Professor, E-mail: mglinic@ippt.gov.pl

Modern computation methods that belong to the group of artificial intelligence soft methods could aid in searching for relationships between the composition of concrete modified with CFBC ash, its microstructure and technical properties, including durability in aggressive environments. Artificial intelligence methods are successfully used in many civil engineering problems (Melhem and Cheng 2003, Alterman and Kasperkiewicz 2006, Kasperkiewicz and Alterman 2007). Kasperkiewicz and Alterman concentrate on three basic concept: artificial neural networks, machine learning and genetic algorithms. In all these approaches the user is not obliged to bother about the model of the process or phenomenon, because the system itself gains adequate knowledge from the examined examples. It can generate thereupon answers in the form of unknown values of the attributes, classification of new examples of the same format or formulation of rules (hypotheses, generalisations) concerning the process under consideration. More details are given in relation to the applied solutions of Fuzzy ARTMAP and ML program AQ19.

The objective of current research was to develop rules for automatic categorization of concrete quality using machine learning techniques. The undertaken research was focused on the resistance of concrete with fluidized bed fly ash to chloride ions aggression. Performed experimental tests on chloride migration provided data for learning and testing of rules discovered by machine learning techniques.

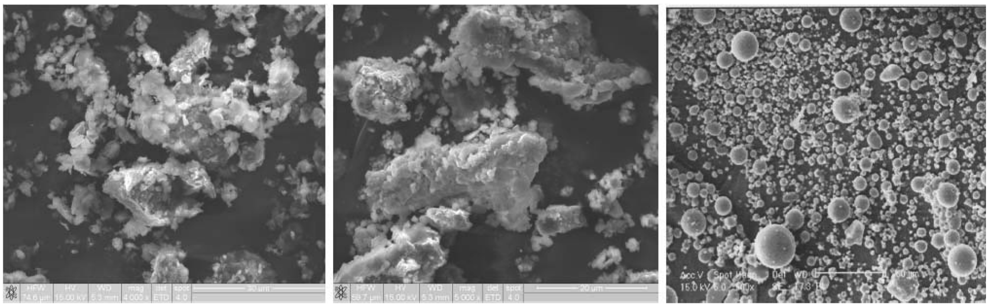
2. Laboratory tests

2.1 Materials and mixture proportions

The chloride migration coefficient in concrete specimens with different content of fluidized bed fly ash was measured (Jóźwiak-Niedźwiedzka 2009). Ordinary Portland cement CEM I 32.5 R from

Table 1 Chemical composition and physical properties of Portland cement CEM I, conventional fly ash and fluidized bed fly ashes from combustion of hard and brown coal (Małolepszy and Kołodziej 2009)

Chemical compounds	PC type I	Conventional fly ash	CFBC fly ash	
			From hard coal *K*	From lignite *T*
SiO_2	21.4	50.8	47.18	36.47
Fe_2O_3	3.5	8.6	6.8	4.4
Al_2O_3	5.7	23.9	25.62	28.4
TiO_2	NA	1.11	1.08	3.84
CaO	64.1	4.0	5.84	15.95
MgO	2.1	2.8	0.15	1.65
SO_3	2.1	0.8	3.62	3.8
Na_2O	0.5	0.8	1.18	1.64
K_2O	0.92	2.9	2.36	0.62
Cl^-	0.029	0.02	0.1	0.03
CaO_{free}	0.9	0.6	0.3	1.4
Specific gravity [g/cm^3]	3.15	2.16	2.68	2.75
Loss on ignition, 1000^{o}C/1h	1.1	2.9	3.4	2.73



(a) (b) (c)

Fig. 1 The shape of ash particles from fluidized bed combustion of lignite (a) - 5000x and hard coal (b) 5000x, and from conventional combustion of hard coal, 500x

Małogoszcz cement plant, gravel fractions 2÷8 mm and 8÷16 mm, and sand fraction 0÷2 mm, were used for composition of concrete specimens. Two kinds of fluidized fly ash were tested: from hard coal combustion in the thermal-electric power station Katowice K and from brown coal - lignite in the power plant Turów T. Chemical and physical properties of Portland cement type I and both CFBC fly ashes are shown in Table 1. Solid residues from coal combustion in circulated fluidized bed boilers are characterized by different mineral and phase compositions than conventional fly ash, by angular shape of grains Fig. 1 and by lack of glassy phase.

Three chemical admixtures: a plasticizer (magnesium lignosulfonates), a high range water reducer

Table 2 Composition of concrete mixes and compressive strength tested after 28 and 90 days

Concrete mix		Cement	Addition		Aggregate	Water	Plasticizer	HRWR	AEA	f_{c28}	f_{c90}
			T	*K*							
		Content [kg/m^3]								[MPa]	
Series B	B0	360	-	-	1859	162	3.2	4.3	-	55.0	70.0
	B15K	306	-	54	1854	162	3.2	3.2	-	56.2	64.3
	B30K	252	-	108	1847	162	3.2	3.2	-	51.6	61.0
	B15T	306	54	-	1850	162	3.2	4.7	-	60.3	70.4
	B30T	252	108	-	1841	162	3.2	5.6	-	58.7	72.0
Series C	C0	380	-	-	1822	171	3.4	2.7	0.4	46.3	49.8
	C15K	323	-	57	1813	171	3.4	2.5	0.6	47.2	48.4
	C30K	266	-	114	1803	171	3.4	3.4	0.6	46.8	56.4
	C15T	323	57	-	1810	171	3.4	3.8	0.6	45.3	50.1
	C30T	266	114	-	1800	171	3.4	4.8	0.6	46.3	47.7
Series D	D0	406	-	-	1586	175	-	0.0	3.2	22.7	26.3
	D20T	290	73	-	1431	151	-	2.0	2.9	21.0	23.3
	D40T	217	145	-	1423	150	-	4.0	5.8	26.1	25.3
	D20K	323	-	81	1593	167	-	2.2	3.2	38.3	41.8
	D40K	244	-	162	1606	157	-	4.5	6.5	43.0	43.4

HRWR- high range water reducer, AEA- air-entraining admixture
0-no addition, *T* - fluidized fly ash from lignite, *K* - fluidized fly ash from hard coal

(polycarboxylane ether) and an air-entraining admixture (synthetic surfactants) were used to achieve approximately the same workability and porosity of fresh mix. Three concrete mixes were designed: series B with water to binder ratio $w/b = 0.45$, air-entrained series C with $w/b = 0.45$ and series D with $w/b = 0.42$. In Table 2 the mixture proportions of tested concretes and the compressive strength of hardened concrete are shown.

The composition of concrete mixes was based on the experimental method with replacement of cement mass by fluidized fly ash: 15% and 30% in series B and C, 20% and 40% in series D. The specimens were cast in cubical moulds 150 mm and in cylinder moulds ø100 mm × 200 mm. Fresh mixes were consolidated by vibration. After 48 hours the specimens were demoulded and cured in high humidity conditions RH > 90%, at temperature 18÷20°C until the age of 28 days.

2.2 Testing procedure

The chloride penetration test for this study was based on the standard of Nordtest Build 492 - Non-Steady State Migration Test (NT Build 492 1999). The principle of the test is to subject the concrete to external electrical potential applied across a specimen and to force chloride ions to migrate into it (Antoni *et al.* 2005). After the specified period of time, depending of the initial current intensity, the specimen is split open and sprayed with silver nitrate solution, which reacts to give white insoluble silver chloride on contact with chloride ions. This provides a simple physical measurement of the depth Fig. 2 to which the sample has been penetrated.

The conformity criteria for concretes according to Non-Steady State Migration Test (NT Build 492 1999) are based on the voltage magnitude, temperature of anolite measured on the beginning and end of test and the depth of chloride ions penetration, are shown in Table 3 (Tang 1996). The non-steady-state migration coefficient, D_{nssm}, is calculated from equation derived from the second Fick's law:

$$D_{nssm} = \frac{0.0239(273+T)L}{(U-2)t}\left(x - 0.0238\sqrt{\frac{(273+T)Lx}{U-2}}\right) \quad (1)$$

here:
D_{nssm} – non-steady-state migration coefficient, $\times 10^{-12}$ [m²/s],
U – absolute value of the applied voltage [V],
T – average value of the initial and final temperature in the anolyte solution [°C],
L – thickness of the specimen [mm],
x – average value of the penetration depths [mm],
t – test duration [h].

Table 3 Estimation of the chloride resistance to chloride ions penetration

Non-steady-state migration coefficient	Resistance to chloride penetration
$< 2 \times 10^{-12}$ m²/s	Very good
$2 - 8 \times 10^{-12}$ m²/s	Good
$8 - 16 \times 10^{-12}$ m²/s	Acceptable
$> 16 \times 10^{-12}$ m²/s	Unacceptable

2.3 Test results of chloride migration coefficient

Tables 4 and 5 present the values of chloride migration coefficient determined after 28 and 90 days of maturity period for concretes series B, C and D.

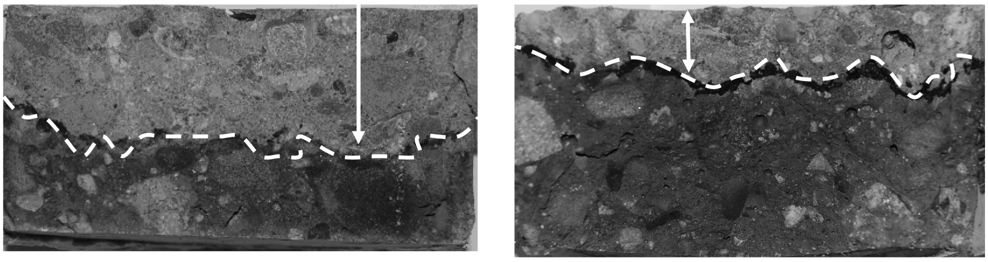
The results show the same general trend in almost all concrete mixtures that values of D_{nssm}

Table 4 Results of tests of chloride ions penetration after 28 days, series B, C and D (mean values from 3 specimens)

Series	Depth of chloride penetration [mm]	D_{nssm} [$\times 10^{-12}$ m^2/s]	Resistance to chloride penetration
B0	27.2	15.25	Acceptable
B15K	20.3	8.68	Acceptable
B30K	15.2	4.98	Good
B15T	17.9	6.40	Good
B30T	12.2	3.02	Good
C0	26.3	13.83	Acceptable
C15K	19.0	7.53	Good
C30K	18.7	6.57	Good
C15T	23.1	9.35	Acceptable
C30T	28.2	10.08	Acceptable
D0	23.3	10.60	Acceptable
D20T	22.5	7.83	Good
D40T	21.7	5.69	Good
D20K	19.4	6.19	Good
D40K	14.1	1.58	Very good

Table 5 Results of tests of chloride ions penetration after 90 days, series B, C and D (mean values from 3 specimens)

Series	Depth of chloride penetration [mm]	D_{nssm} [$\times 10^{-12}$ m^2/s]	Resistance to chloride penetration
B0	20.5	9.29	Acceptable
B15K	18.0	6.29	Good
B30K	12.1	2.93	Good
B15T	14.0	4.81	Good
B30T	11.7	2.66	Good
C0	22.1	10.31	Acceptable
C15K	15.2	4.75	Good
C30K	15.1	4.19	Good
C15T	12.9	4.36	Good
C30T	18.7	4.67	Good
D0	26.6	10.3	Acceptable
D20T	22.7	5.68	Good
D40T	20.6	2.33	Good
D20K	18.9	4.58	Good
D40K	17.9	0.99	Very good



50 x 100 mm
B0-28 days B30T - 28 days

Fig. 2 Example of the depth of chloride ions penetration in concrete series B, without FBCFA and with 30% of FBCFA from lignite

decreased with increased FBCFA content because of the changes in concrete microstructure. The concretes without FBCFA were the ones that showed the highest values of D_{nssm} only acceptable resistance to chloride penetration according to criteria shown in Table 3. In all series of concrete specimens the chloride migration coefficient tested after 90 days showed relative stabilization.

The example of depth of chloride ions penetration in series B (B0 and B30T) tested after 28 days is showed in Fig. 2.

The comparable tests results based on eight concrete mixtures was obtained. The ordinary Portland cement replacement by ground fly ash varied from 0% to 70% in steps of 10%. For high volume fly ash concrete better chloride resistance than in ordinary concrete has been achieved (Sengul *et al.* 2005).

3. Machine learning methods

Data mining can be defined as the process of discovering patterns in a dataset. By a dataset we mean a *database* i.e., collection of logically related records. Each record can be called an *example* or *instance* and each one is characterized by the values of a set of predetermined *attributes*. A few different styles of learning appear in data mining applications but the most common is a *classification*. The aim of the classification process is to learn a way of classifying unseen examples based on the knowledge extracted from the provided set of classified examples. In order to extract the knowledge from the provided dataset the attribute set characterizing the example has to be divided into two groups: the *class* attribute or attributes and the *non-class* attributes. It is obvious that for an unseen examples only *non-class* attributes are known, therefore the aim of data mining algorithms is to build such a knowledge model that allows predicting the example class membership based only on non-class attributes. The knowledge model is dependent on the way how the classifier is constructed and it can be represented by decision trees (e.g. algorithm C4.5) or classification rules (the AQ algorithms family). Regardless of the representation both types of algorithms create hypotheses.

In order to evaluate the classifier i.e., to judge the hypotheses generated from the provided *training set* we have to verify the classifier performance on the independent dataset which is called *testing set*. Of course both sets of training data and test data should be representative samples of the considered problem. The classifier predicts the class of each instance from the test set; if it is correct, that is counted as a success; if not it is an error. In order to measure the overall performance of the classifier some quantitative analysis should be done.

The example of such a quantitative measure are a success rate usually called a *classification accuracy.* This is the number of correct classifications of the instances from the test set divided by the total number of these instances, its measure is expressed as a percentage.

In order to get a deeper understanding which types of errors are the most frequent the result obtained from a test set is often displayed as a two-dimensional *confusion matrix* with a row and a column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. Good results correspond to large numbers down the main diagonal and small, ideally zero, off-diagonal elements. The sum of the numbers down the main diagonal divided by the the total number of test examples determine classification accuracy.

Lets consider what can be done when the number of data for training and testing is limited. The simplest way is to reserve a certain number for testing and to use the remainder for training. Of course, the selection should be done randomly. In practical terms, it is common to hold out one-third of the data for testing and use the remaining two-thirds for training (Witten and Frank 2005). The main disadvantage of this simple method is that this random selection may be not representative. A more general way to mitigate any bias caused by the particular sample chosen for holdout is to repeat the whole process, training and testing, several times with different random samples. This process is called the *k-fold cross-validation.* In this technique a fixed number of folds – k is arbitrary described. Then the data set U is split into k approximately equal portions ($U = E_1 \cup \ldots \cup E_k$) (Krawiec and Stefanowski 2003). In each iteration i the set E_i is used for testing and the remainder $U \setminus E_i$ is used for training.

Overall classification accuracy is calculated as an average from the classification accuracy for each iteration $\eta(E_i)$, i.e., is defined as

$$\overline{\eta} = \frac{1}{k}\sum_{i=1}^{k} \eta(E_i) \tag{2}$$

In order to generate rules describing the concrete resistance to chloride penetration several numerical experiments were performed using program AQ21 and algorithm J48 from the WEKA workbench. Algorithm AQ21, invented in the Machine Learning and Inference Laboratory of George Mason University (Wojtusiak 2004) is based on covering approach alike most of the rule-based data mining algorithms. Therefore, the AQ21 algorithm generates subsequent rules until all the examples (sometimes not all) are covered. The idea of adding a new rule or a new term to existing rule is to include as many instances of the desired class (*positive examples*) as possible and to exclude as many instances of other classes (*negative examples*) as possible.

The second considered algorithm, J48, is available as a part of WEKA workbench, which was developed at the University of Waikato in New Zealand (Witten and Frank 2005). Algorithm J48 is an implementation of the last publicly available version of an algorithm C4.5 devised by J. Ross Quinlan. Construction of decision trees is based on a simple divide and conquer approach, which is well known in computer science. The main problem is connected with a selection of tests (splits of attributes) which should be placed in the nodes. The test is good if it allows to shorten the way from the root to the leaves representing classes. Decision trees can be converted to classification rules with ease.

4. Seeking for the rules describing chloride ions penetration

4.1 Chloride ions penetration after 28 days

4.1.1 Results obtained from AQ21

As the results of the experiments carried on the specimens with different contents of fluidized fly ash, as shown in tables 2 and 4, the following database consisted of 15 records was introduced. This database was used to determine the rules describing the concrete resistance to chloride penetration after 28 days. The database with one nominal and 6 numerical attributes is presented in Table 6 (Marks *et al.* 2009).

where:
*C*1 – cement content, [kg/m^3],
pfT – fluidized fly ash from brown coal content (power plant Turów), [kg/m^3],
pfK – fluidized fly ash from hard coal content (power station Katowice), [kg/m^3],
W – water content, [kg/m^3],
A_fr – air content in fresh mix, [%],
*fc*28 – compressive strength tested after 28 days, [MPa],
Resistance – concrete resistance to chloride ions penetration (Acceptable, Good).

The last attribute – resistance - is a nominal one which takes on two possible values: Acceptable, Good. In the considered database to the class [Resistance=Acceptable] belongs 6 examples and to the class [Resistance=Good] belongs 9 examples.

The aim of an experiment is to generate the rules, which allow us to determine concrete resistance to chloride ions penetration. As an training set all the instances from the database were considered. The rules generated by an AQ21 algorithm are presented below

Table 6 The database

C1	*pfT*	*pfK*	*W*	*A_fr*	*fc*28	Resistance
360	0	0	162	2.1	55.0	Acceptable
306	0	54	162	1.8	56.2	Acceptable
252	0	108	162	1.3	51.6	Good
306	54	0	162	1.6	60.3	Good
252	108	0	162	1.6	58.7	Good
380	0	0	171	6.2	46.3	Acceptable
323	0	57	171	6.8	47.2	Good
266	0	114	171	5.8	46.8	Good
323	57	0	171	6.6	45.3	Acceptable
266	114	0	171	6.2	46.3	Acceptable
406	0	0	175	4.9	22.7	Acceptable
290	73	0	151	6.9	21.0	Good
217	145	0	150	7.8	26.1	Good
323	0	81	167	4.6	38.3	Good
244	0	162	157	4.6	43.0	Good

[Resistance=Good]
  # Rule 1
<-- [*pfK*>=55] : *p*=5, *n*=0, *q*=0.556

  # Rule 2
<-- [*C*1<=258] : *p*=4, *n*=0, *q*=0.444

  # Rule 3
<-- [*pfT*>=27] [*W*<=166] : *p*=4, *n*=0, *q*=0.444 (3)

[Resistance=Acceptable]
  # Rule 1
<-- [*pfK*<=55] [*A_fr*=1.7..6.75] : *p*=6, *n*=0, *q*=1

  # Rule 2
<-- [*pfK*<=55] [*fc*28=44.15..57.45] : *p*=5, *n*=0, *q*=0.833

where *p* denotes the number of positive examples covered by the rule, *n* denotes the number of negative examples covered by the rule (i.e., the number of records from the other classes satisfying the rule) and *q* denotes the quality of the rule.

The rules showed in Eq. (3) can be interpreted as follows but it should be underlined that the presented rules concern concretes with the overall mass of cement and additions equal 360, 380 or 406 [kg/m^3] (Table 2).

[Resistance is **Good**]
IF
[*pfK* >= 55]
OR
[*C*1 <= 258]
OR
[*pfT* >= 27] and [*W* <=166]

[Resistance is **Acceptable**]
IF
[*pfK* <= 55] and [*A_fr* = 1.7..6.75]
OR
[*pfK* <= 55] and [*fc*28 = 44.15..57.45]

In order to evaluate the classifier, i.e., to judge the hypotheses (classification rules, decision trees) generated from the provided training set, we have to verify the classifier performance on the independent testing set. When we have only one database consisting of a very small number of records, the estimation of classification accuracy (measure of the overall performance of the classifier) can be done using the *n*-fold cross validation, where *n* is the number of examples in the database (Witten and Frank 2005). In this method each example in turn is left out, and the learning method is trained on all the remaining examples. It is judged by its correctness on the remaining example – one or zero for success or failure, respectively. The results from *n* judgments, one for each member of the database, are averaged, and that average represents the classification accuracy (Witten and Frank 2005). This method, named *leave-one-out* cross validation, is useful to the database of a very small number of records. It seems to offer a chance of squeezing the maximum out of a small dataset and obtaining as accurate an estimate as possible.

The results from *n* judgments may be displayed as a two-dimensional confusion matrix with a row and a column for each class. Each confusion matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. The numbers of examples down main diagonal are predicted correctly. The classification accuracy is the sum of numbers down the main diagonal divided by the total number of data set examples.

Applying the *n*-fold cross validation for n = 15 (number of examples in Table 6) we obtain a confusion matrix in the following form:

	Acceptable	Good	Other
Acceptable	4	2	0
Good	4	4	1

The value of classification accuracy is equal to 53.3%.

4.1.2 Results obtained from J48

In order to generate the rules, which allow us to determine the concrete resistance against chloride ion penetration the J48 algorithm was also used. As the training set all the instances from the database (Table 6) were considered. The decision tree generated by the J48 algorithm is presented in Fig. 3.

where the first number in brackets denotes the number of examples from the training set covered by a selected leaf, and the second number – just after the sign "/" – indicates the number of incorrectly classified instances (negative examples). When there is only one number in brackets, then it indicates the number of examples correctly classified (positive examples).

The obtained decision tree (Fig. 3) can be easily transformed into the following rules:

[Resistance=Good]
Rule1 [$C1$ <= 323] and [pfK <= 54] and [W <= 162]
Rule2 [$C1$ <= 323] and [pfK > 54] (4)

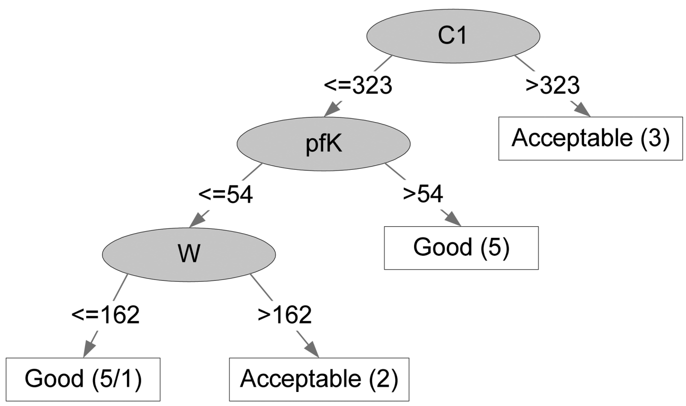


Fig. 3 The decision tree generated by the J48 algorithm for chloride penetration after 28 days

[Resistance=Acceptable]
Rule1 [*C*1 <= 323] and [*pfK* <= 54] and [*W* > 162]
Rule2 [*C*1 > 323]

Using the *n*-fold cross validation for J48 algorithm we obtain the confusion matrix in the following form:

	Acceptable	Good
Acceptable	2	4
Good	2	7

and the classification accuracy equal 60%.

4.2 Chloride ions penetration after 90 days

4.2.1 Results obtained from AQ21

In order to generate rules describing the concrete resistance to chloride penetration after 90 days a database was used, that was very similar to the database shown in Table 6. The first five numerical attributes are identical as in Table 6. The last numerical attribute *fc*90 determines compressive strength tested after 90 days, [MPa]. In the considered database three examples belong to the class [Resistance=Acceptable] and 12 examples belong to the class [Resistance=Good] (Table 7).

As a training set all the instances from the database were considered. The rules generated by an AQ21 algorithm are presented below:

[Resistance=Good]
 # Rule 1
<-- [*C*1<=341] : *p*=12, *n*=0, *q*=1

Table 7 The database

*C*1	*pfT*	*pfK*	*W*	*A_fr*	*fc*90	Resistance
360	0	0	162	2.1	70.0	Acceptable
306	0	54	162	1.8	64.3	Good
252	0	108	162	1.3	61.0	Good
306	54	0	162	1.6	70.4	Good
252	108	0	162	1.6	66.3	Good
380	0	0	171	6.2	49.8	Acceptable
323	0	57	171	6.8	48.4	Good
266	0	114	171	5.8	56.4	Good
323	57	0	171	6.6	50.1	Good
266	114	0	171	6.2	47.7	Good
406	0	0	175	4.9	26.3	Acceptable
290	73	0	151	6.9	23.3	Good
217	145	0	150	7.8	25.3	Good
323	0	81	167	4.6	41.8	Good
244	0	162	157	4.6	43.4	Good

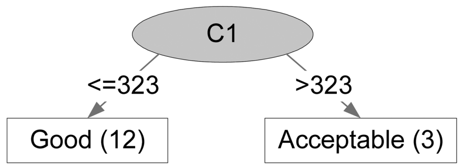


Fig. 4. The decision tree generated by the J48 algorithm for chloride penetration after 90 days

Rule 2
<-- [C1<=351] [fc90<=68.15] : p=11, n=0, q=0.987

[Resistance=Acceptable]
# Rule 1
<-- [C1>=342] : p=3, n=0, q=1

In order to estimate the classification accuracy the n-fold cross validation was used for n =15. The results of this method are described by the following confusion matrix:

	Acceptable	Good
Acceptable	3	0
Good	1	11

Here one example from Acceptable class is classified incorrectly to Good class, the remaining examples are classified correctly and the classification accuracy is equal 93.3%.

4.2.2 Results obtained from J48

In order to generate the rules, which allow us to determine concrete resistance against chloride ions penetration the J48 algorithm was used also. As the training set all the instances from the database (Table 7) were considered. The decision tree generated by an J48 algorithm is presented in Fig. 4.

When n-fold cross validation was used we obtain the following confusion matrix:

	Acceptable	Good
Acceptable	3	0
Good	0	12

and the classification accuracy was equal 100%.

5. Conclusions

The rules generated by computer programs AQ21 and WEKA using J48 algorithm have provided means for automatic categorization of plain concretes and concretes modified with CFBC fly ash as materials of good or acceptable resistance to chloride penetration. Due to a small number of tested specimens the rules are applicable only to concrete mix compositions with similar binder content and similar values of water to cement ratio.

The rules describing the concrete resistance to chloride penetration after 90 days, which were determined by AQ21 algorithm as well by J48 algorithm, are similar. According to generated rules, resistance was qualified as acceptable for tested concrete without fluidized fly ash, whereas resistance was good for the same concrete with replacement of cement mass from 15% to 40% by fluidized fly ash from hard coal or brown coal. Therefore, application of CFBC fly ash improved the resistance of concrete in respect to chloride penetration.

Application of AQ21 and WEKA programs provided similar estimation of the concrete resistance to chloride ion penetration. Further tests are needed in order to enlarge the experimental data basis and to cover larger variety of concrete compositions.

References

Alterman, D. and Kasperkiewicz, J. (2006), "Evaluating concrete materials by application of automatic reasoning", *Bull. Pol. Acad. Sci. Tech.,* **54**(4), 353-361.

Antoni, A., Horiguchi, T. and Saeki, N. (2005), "Chloride penetration into fiber reinforced concrete under static and cyclic compressive loading", *Proceedings of 10 DBMC International Conference on Durability of Building Materials and Components*, Lyon, April.

Fu, X., Li, Q., Zhai, J., Sheng, G. and Li, F. (2008), "The physical-chemical characterization of mechanically-treated CFBC fly ash", *Cement Concrete Comp.,* **30**(3), 220-226.

Giergiczny, Z. and Puźak, T. (2008), "Properties of concrete with fluidal fly ash addition", *Proceedings of 3rd International Symposium Non-Traditional Cement Concrete*, 263-271, Brno, June.

Glinicki, M.A. and Zielinski, M. (2009), "Frost salt scaling resistance of concrete containing CFBC fly ash", *Mater. Struct.*, **42**(7), 993-1002.

Jóźwiak-Niedźwiedzka, D. (2009), "Effect of fluidized bed combustion fly ash on the chloride resistance and scaling resistance of concrete", *Proceedings of RILEM Conference - Concrete in Aggressive Aqueous Environments, Performance, Testing and Modeling*, 556-563, Toulouse, June.

Kasperkiewicz, J. and Alterman, D. (2007), "Holistic approach to diagnostics of engineering materials", *Comput. Assist. Mech. Eng. Sci.*, **14**(2), 197-207.

Krawiec, K. and Stefanowski, J. (2003), *Machine learning and neural networks (in Polish)*, Poznan University of Technology, Poland.

Melhem, H.G. and Cheng, Y. (2003), "Prediction of remaining service life of bridge decks using machine learning", *J. Comput. Civil Eng.*, **17**(1), 1-9.

Małolepszy, J. and Kołodziej, Ł. (2009), "Resistance of cements with high amount of ashes from fluidized bed furnace to chloride diffusion", *Proceedings of 5th International Conference Concrete and Concrete Structures*, Zilina, October.

Marks, M., Jóźwiak-Niedźwiedzka, D. and Glinicki, M.A. (2009), "Application of machine learning for prediction of concrete resistance to migration of chlorides", *Proceedings of 9th International Symposium on Brittle Matrix Composites BMC9*, 227-236, Warsaw, October.

NT BUILD 492 (1999), *Concrete, mortar and cement-based repair materials: chloride migration coefficient from non-steady-state migration experiments,* Nordtest Method 492, Finland.

Nowak, W. (2003), "Clean coal fluidized-bed technology in Poland", *Appl. Energ*, **74**(3-4), 405-413.

Tang, L. (1996), *Chloride transport in concrete - measurement and prediction*, PhD thesis, Department of Building Materials, Chalmers University of Technology, Göteborg, Sweden.

Sengul, O., Tasdemir, C. and Tasdemir, M.A. (2005), "Mechanical properties and rapid chloride permeability of concretes with ground fly ash", *ACI Mater. J.*, **102**(6), 414-421.

Witten, I.H. and Frank, E. (2005), *Data mining. Practical machine learning tools and techniques*, Elsevier, San Francisco, CA.

Wojtusiak, J. (2004), *AQ21 user's guide,* Reports of the Machine Learning and Inference Laboratory, MLI 04-3, George Mason University.

CC

Assessment of Scaling Durability of Concrete with CFBC Ash by Automatic Classification Rules

Maria Marks[1]; Daria Jóźwiak-Niedźwiedzka[2]; Michał A. Glinicki[3]; Jan Olek[4]; and Michał Marks[5]

Abstract: The objective of this investigation was to develop rules for automatic assessment of concrete quality by using selected artificial intelligence methods based on machine learning. The range of tested materials included concrete containing nonstandard waste material—the solid residue from coal combustion in circulating fluidized bed combustion boilers (CFBC ash) used as an additive. Performed experimental tests on the surface scaling resistance provided data for learning and verification of rules discovered by machine learning techniques. It has been found that machine learning is a tool that can be applied to classify concrete durability. The rules generated by computer programs AQ21 and WEKA by using the J48 algorithm provided a means for adequate categorization of plain concrete and concrete modified with CFBC fly ash as materials resistant or not resistant to the surface scaling. **DOI:** 10.1061/(ASCE)MT.1943-5533.0000464. *© 2012 American Society of Civil Engineers.*

CE Database subject headings: Classification; Databases; Concrete; Durability; Fly ash.

Author keywords: Machine learning; Automatic classification rules; Database; Concrete durability; Scaling resistance; Circulated fluidized bed combustion fly ash (CFBC fly ash).

Introduction

In Europe, approximately 50% of electricity is generated through combustion of solid fuels, primarily hard coal and lignite and, to a lesser extent, also through the combustion of oil shale. The consequence of the above occurrence is the production of more than 160 Mt of combustion by-products of which 100 Mt can be attributed to the 27 states of the European Union (Szczygielski and Hycnar 2009). The forecast for the next years predict the production of 17 Mt of coal combustion products in Poland, fluidized bed combustion ash—approximately 5 Mt. Fly ash is known to be a valuable additive to concrete mixes, and it has also become a significant factor in concrete technology to reduce the negative environmental effect of concrete production. The most desired advantage of using fly ash in concrete technology concerns an improvement of the mix workability, the long-term strength of concrete and its resistance to aggressive environments (Aïtcin 1998; Gencel Brostow et al. 2011; Helmuth 1987; Khurana and Saccone 2001). Commonly used pozzolan material is derived from pulverized coal combustion systems. Other systems including Fluidized Bed Combustion (FBC), which has the advantage of, for example, lower thermal NO_x emissions and greater fuel flexibility, also supplies fly ash. However, the fly ash collected from FBC boilers have characteristics that place it outside of the standard requirements in the European Union. According to the ASTM C618 standard, the American FBC fly ashes cannot be used in concrete technology because of too high a sulfate content (Stevens et al. 2009). Current European practice for using fly ash as a type II concrete additive (inorganic addition, pozzolanic or latent hydraulic addition) according to the EN 206-1 standard (European Committee for Standardization 2000), does not allow the use of solid by-products resulting from advanced coal burning technologies, including circulating fluidized bed combustion (CFBC). Also, the European standard EN 450-1 (European Committee for Standardization 2005) does not anticipate to use the CFBC fly ash either for cement or for concrete production. CFBC fly ash differs in physical and chemical properties from the traditionally used fly ashes (Table 1).

The solid residue from coal combustion in fluidized bed boilers contains noncombustible mineral matter, sorbent material, and unburned carbon, because of a high sulfur content, high free lime content, high loss on ignition and the lack of glassy phase (Nowak 2003; Giergiczny and Pużak 2008; Goodarzi 2006; Fu et al. 2008; Glinicki and Zielinski 2009). Moreover, according to the definition indicated by the European Standard, fly ash is a fine powder of mainly spherical, glassy particles derived from burning the pulverized coal, with or without cocombustion materials. The positive effect of the spherical shape and smooth surface of fly ash particles is primarily related to a reduction of the water demand for a given mix workability (Helmuth 1987). In the case of CFBC fly ash, grains are nonspherical and the glassy phase is not present, so that this fly ash is beyond its scope (Glinicki and Zieliński 2008). The potential of CFBC fly ash as an additive to concrete mix is not well established. Limited experimental tests on concrete containing CFBC fly ash as an additive revealed satisfactory strength and durability of such concretes (Jóźwiak-Niedźwiedzka 2009). The

[1]Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland. E-mail: mmarks@ippt.gov.pl

[2]Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland. E-mail: djozwiak@ippt.gov.pl

[3]Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland (corresponding author). E-mail: mglinic@ippt.gov.pl

[4]Purdue Univ., School of Civil Engineering, West Lafayette, IN. E-mail: olek@purdue.edu

[5]Institute of Control and Computation Engineering, Warsaw Univ. of Technology, Nowowiejska 15/19, 00-665 Warsaw, E-mail: mmarks@ia.pw.edu.pl

Note. This manuscript was submitted on June 27, 2011; approved on December 20, 2011; published online on June 15, 2012. Discussion period open until December 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Materials in Civil Engineering***, Vol. 24, No. 7, July 1, 2012. ©ASCE, ISSN 0899-1561/2012/7-860–867/$25.00.

Table 1. Chemical Requirements for Fly Ash According to ASTM C618 and EN 450-1 and the Actual Characteristics of CFBC Fly Ash

	ASTM C618		EN 450	CFBC fly ash	
Constituent	*F*	*C*	Type II concrete additive	From hard coal	From lignite
ΣSiO_2, Al_2O_3, Fe_2O_3, min, %	70	50	70	79.6	69.27
SO_3, max, %	5	5	3	3.62	3.8
Na_2O	max% 1.5	max% 1.5	max% 5 (Na_2O_{eq})	1.18	1.64
CaO free, max, %	—	—	2.5	0.3	1.4
Moisture content, max, %	3	3	—	—	—
Cl^- content, max, %	—	—	0.1	0.1	0.03
Loss on ignition 1,000°C/1 h, max, %	6–12	6	5	3.4	2.73

J. Mater. Civ. Eng. 2012.24:860-867.

test results of Glinicki and Zielinski (2009) revealed both long-term strength and proper scaling resistance of concrete containing 20% by mass of selected CFBC ash provided that the air-void characteristics were adequate. Another result (Małolepszy and Kołodziej 2009) showed that substitution of cement by CFBC fly ash decreases the chloride ion diffusion coefficients. However, efficient methods for selecting the appropriate type of CFBC ash to provide the long-term durability of concrete are still not available.

Modern computation methods that belong to the group of artificial intelligence methods could aid in searching for relationships between the composition of concrete modified with CFBC ash, its microstructure and technical properties, including durability in aggressive environments. Artificial intelligence methods, (AIM), including artificial neural networks (ANN) and machine learning (ML) are successfully used in many civil engineering problems. Neural networks have been applied to the prediction of the mechanical properties of cement, including materials such as the compressive strength of concrete (Ni and Wang 2000), strength of high-performance concrete (Yeh 1998), compressive and splitting tensile strengths of recycled aggregate concretes containing silica fume (Topçu and Saridemir 2008), or the abrasive wear of concrete (Gencel and Kocabas et al. 2011). Machine learning has been used in classification and prediction problems such as estimating the remaining service life of bridge decks (Melhem and Cheng 2003), in classification of plain concrete, and concrete modified with CFBC fly ash as materials of good and acceptable resistance to chloride penetration (Marks et al. 2009). In the paper by Alterman and Kasperkiewicz (2006), the authors proposed to combine artificial neural networks and machine learning methods in one system to estimate and predict various properties of concrete materials.

The aim of the present paper is to generate the rules describing the scaling resistance of concrete modified with CFBC ash by using selected machine learning algorithms. The rules generated by selected algorithms provided a means for required classification of modified concrete as materials resistant or not resistant to the surface scaling.

Laboratory Tests

Materials and Mixture Proportions

Concrete specimens with a different CFBC ash content were manufactured by using ordinary Portland cement CEM I 32.5 R and two kinds of CFBC ash: from hard coal combustion in the thermal-electric power station Katowice (*K*) and from brown coal combustion in the power plant Turów (*T*).

The chemical and physical properties of Portland cement used and both CFBC fly ashes are shown in Table 2. Solid residues from coal combustion in circulated fluidized bed boilers are

Table 2. Chemical Composition and Physical Properties of Portland Cement CEM I, Pulverized Fuel Fly Ash, and CFBC Ashes from Combustion of Hard and Brown Coal

			CFBC ash	
Chemical compounds	PC type I	Pulverized fuel fly ash	From hard coal *K*	From lignite *T*
SiO_2	19.6	50.8	47.18	36.47
Fe_2O_3	3.1	8.6	7.84	4.37
Al_2O_3	5.7	23.9	26.62	28.35
TiO_2	—	1.11	1.08	3.84
CaO	62.1	4	5.84	15.95
MgO	2.1	2.8	0.18	1.65
$\mathbf{SO_3}$	3.11	0.8	3.62	3.8
Na_2O	0.5	0.8	1.18	1.64
K_2O	0.92	2.9	2.36	0.62
Cl^-	0.029	0.02	0.1	0.03
$\mathbf{CaO_{free}}$	0.9	0.6	0.34	1.4
Specific gravity [g/cm^3]	3.15	2.16	2.68	2.75
Loss on ignition, 1,000°C/1 h	1.1	2.9	3.4	2.73

characterized by a different chemical composition than conventional fly ash and by the angular shape of grains (Fig. 1) in contrast to round-shape particles of conventional fly ash. The increased content of SO_3 and CaO in CFBC is evident.

Coarse aggregate fractions 2–8 mm and 8–16 mm and sand fraction of 0–2 mm were used. Regular chemical admixtures were used in sets as recommended by the manufacturers: high-range water reducers, air-entraining admixtures and a plasticizer. The amount of admixtures was changeable to achieve approximately the same slump and the designed air-void content in the mix. Four series of air-entrained concrete mixes were designed: series C with water to binder ratio $w/b = 0.45$, series D, F with $w/b = 0.42$; and series G with $w/b = 0.44$. In Table 3, the proportions of binder materials are shown. Natural gravel aggregates were used in series C and D; crushed basalt aggregates were used in series F and G.

The concrete mix design was based on the experimental method with replacement of cement mass by CFBC ash: 15 and 30% in series C, 20 and 40% in series D, 20, 30, 40% in series F and 30% in series G. The specimens were cast in 150-mm cubical molds and in cylinder molds $\phi 100$ mm $\times$ 200 mm. Fresh mixes were consolidated by vibration. After 48 h, the specimens (because of elongated setting time) were demoulded and cured in high humid conditions $RH > 90\%$, at a temperature of 18–20°C until the age of 28 days. Properties of fresh mix and the compressive strength of concrete at 28 days are given in Table 4. The compressive strength

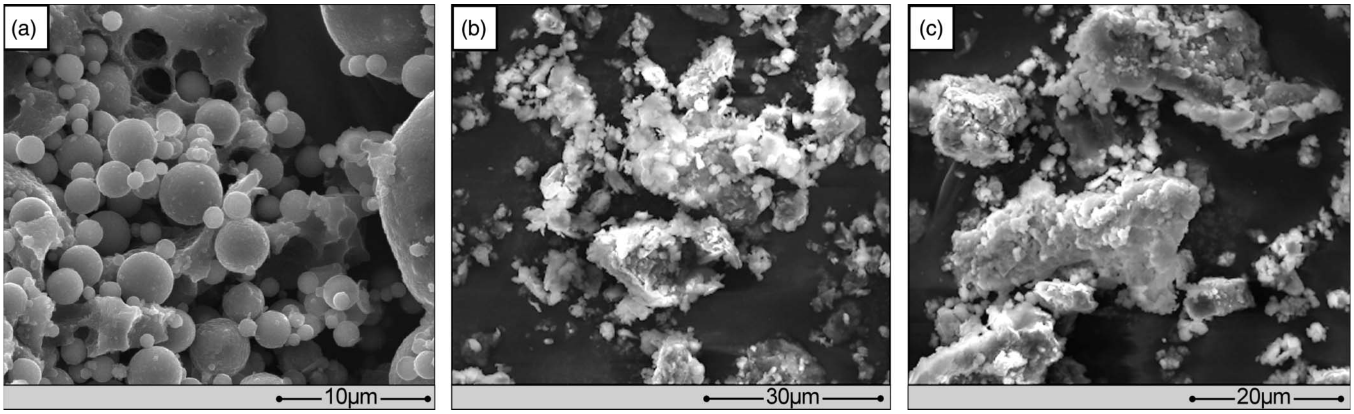


Fig. 1. The shape of ash particles: (a) from pulverized coal combustion, magnification 10,000 (reprinted with permission from Giergiczny and Pużak 2008); (b) from fluidized bed combustion of hard coal—Katowice, magnification 4,000 (image by authors); (c) from fluidized bed combustion of brown coal—Turów, magnification 5,000 (image by authors)

Table 3. Proportions of Binder Materials in Concrete Mixes

Concrete mix		Cement	CFBC ash *T*	CFBC ash *K*	Water	Plasticizer	HRWR	AEA
		Content [kg/m^3]						
Series C	C0	380	—	—	171	3.4	2.7	0.4
	C15K	323	—	57	171	3.4	2.5	0.6
	C30K	266	—	114	171	3.4	3.4	0.6
	C15T	323	57	—	171	3.4	3.8	0.6
	C30T	266	114	—	171	3.4	4.8	0.6
Series D	D0	406	—	—	175	—	0.0	3.2
	D20T	290	73	—	151	—	2.0	2.9
	D40T	217	145	—	150	—	4.0	5.8
	D20K	323	—	81	167	—	2.2	3.2
	D40K	244	—	162	157	—	4.5	6.5
Series F	F0	360	—	—	150	—	1.8	0.36
	F20K	288	—	72	150	—	2.16	1.44
	F30K	252	—	108	150	—	2.88	2.16
	F40K	216	—	144	150	—	3.6	2.88
Series G	G0	354	—	—	156	—	3.2	0.05
	G30K	246	—	105	155	—	7.4	0.4
	G30T	246	105	—	155	—	8.8	0.11

Note: HRWR—high-range water reducer, AEA—air-entraining admixture, *T*—from lignite combustion, *K*—from hard coal combustion.

Table 4. Properties of Fresh Mix and the Compressive Strength of Concrete

Concrete mix		Slump [mm]	Air content [%]	Cube compressive strength at 28 days [MPa]
Series C $w/b = 0.45$	C0	115	6.2	46.3
	C15K	95	6.8	47.2
	C30K	120	5.8	46.8
	C15T	135	6.6	45.3
	C30T	115	6.2	46.3
Series D $w/b = 0.42$	D0	130	4.9	22.7[a]
	D20T	150	6.9	21.0[a]
	D40T	150	7.8	26.1[a]
	D20K	150	4.6	38.3[a]
	D40K	240	4.6	43.0[a]
Series F $w/b = 0.42$	F0	30	7.6	49.7
	F20K	30	7.0	56.4
	F30K	20	6.6	55.8
	F40K	40	6.6	54.3
Series G $w/b = 0.44$	G0	90	7.1	62.5
	G30K	40	6.1	65.3
	G30T	80	4.7	66.8

[a]Reduced compressive strength at 28 days of curing was because of high content of air-voids in the hardened concrete. The results of the total air-void content A_hr are further given in the database table.

of concrete containing CFBC ash was similar to the strength of the reference concrete. The observed differences in the compressive strength of concrete within each series can be primarily attributed to the air content variability.

Testing Procedure

Frost salt scaling tests were performed according to a European Standard procedure (European Committee for Standardization 2006) that was established following the Swedish standards 137244 (SS 1995). The upper horizontal surface of the specimens (the cut surface) was exposed to freezing and thawing; whereas, the remaining surfaces were isolated against humidity and heat transfer. After 28 days of curing, the top exposed surface was covered with 3% NaCl solution. Standard cooling and thawing cycles were applied. The temperature in the saline solution layer on the top of specimens was recorded every hour with the digital thermometer LB-711 system working with 6 surface probes. The scaled material was collected and weighed after a given number of freeze-thaw cycles, and the results expressed as mass per unit area have been recorded. The mean mass of scaled material after 28 (m_{28}) and 56 (m_{56}) cycles is used for evaluating the scaling resistance, according to the criteria presented in Table 5 and according to Swedish standard (SS 1995). For concrete paving flags evaluated by using European Standard procedure EN 1339 (European Committee for Standardization 2003), the allowable limit of mass of scaled material is 1 kg/m^2 after 28 cycles.

Air-void parameters in hardened concrete were determined by using a computer-driven image analysis system described in Glinicki and Zieliński (2008). The automatic measurement procedure was designed to comply with the requirements imposed by ASTM C457 (ASTM 1991) and the European Standard EN

Table 5. Criteria of the Surface Scaling Resistance Evaluation

Scaling resistance	Requirements
Very good	$m_{56} < 0.10\ kg/m^2$
Good	$m_{56} < 0.20\ kg/m^2$
	or $m_{56} < 0.50\ kg/m^2$ and $m_{56}/m_{28} < 2$
Admissible	$m_{56} < 1.00\ kg/m^2$ and $m_{56}/m_{28} < 2$
Inadmissible	$m_{56} > 1.00\ kg/m^2$ or $m_{56}/m_{28} > 2$

480-11 (European Committee for Standardization 2008). Results of measurements were available as a set of standard parameters for air-void structure characterization: spacing factor $\bar{L}$ (mm), specific surface $\alpha(1/mm)$, air content A (%), denoted as A_hr in section 4, content of air voids with a diameter less than 0.3 mm A300 (%).

Test Results of Scaling Resistance of Concrete

Values of the mass of scaled material determined after 28 and 56 cycles of freezing and thawing in presence of 3% NaCl solution for concretes series C, D, F, G and the results of evaluation of the surface scaling resistance according to the Borås method and to the European standard procedure EN 1339 are given in Table 6.

A general tendency to decrease the frost salt scaling resistance with AN increasing content of CFBC ash is observed. A more detailed analysis of material composition effect on the frost salt scaling of concrete was undertaken by using machine learning methods.

Machine Learning Methods

Data mining can be defined as the process of discovering patterns in a dataset. By a dataset a *database* is meant i.e., collection of logically related records. Each record can be called an *example* or *instance* and each one is characterized by the values of predetermined *attributes*. Four basically different types of learning appear in data mining applications: *classification, association, clustering*, and *numeric prediction* (Witten and Frank 2005). The most common of them is *classification*. The aim of the classification process is to learn a way of classifying unseen examples on the basis of the knowledge extracted from the provided set of classified examples. To extract the knowledge from the provided dataset, the attribute set characterizing the example has to be divided into two groups: the *class* attribute and the *nonclass* attributes. For unseen examples, only nonclass attributes are known, therefore, the aim of data mining algorithms is to build such a knowledge model that allows predicting the example class membership only on the basis of nonclass attributes. The knowledge model is dependent on the way of how the classifier is constructed, and it can be represented by decision trees (e.g., algorithm C4.5, Quinlan 1993) or classification rules (the AQ algorithms family, Michalski 1983). Regardless of the representation, both types of algorithms create hypotheses.

To evaluate the classifier i.e., to judge the hypotheses generated from the provided *training set* it is necessary to verify the classifier performance on the independent dataset, which is called the *testing set*. It is also important to ensure that both the training data and the test data are representative for the considered problem. The classifier predicts the class of each instance from the test set; if it is correct, that is counted as a success; if not it is an error. To measure the overall performance of the classifier, some quantitative analysis needs to be performed.

The example of such a measure resulting from quantitative analysis is a success rate, which is usually called a *classification accuracy*. This is the number of correct classifications of the instances from the test set divided by the total number of these instances expressed as a percentage.

To get a deeper understanding of what types of errors are the most frequent, the result obtained from a test set is often displayed as a two-dimensional *confusion matrix* with a row and a column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. Good results correspond to large numbers down the main diagonal and small, ideally zero, for the elements off the diagonal.

Table 6. Results of Tests of Scaling Resistance of Concrete after 28 and 56 Cycles, Series C, D, F, G and the Results of Evaluation of Scaling Resistance according to Assumed Criteria

	Mass of scaled material		Evaluation of frost salt scaling resistance according to	
Series	After 28 cycles [kg/m^2]	After 56 cycles [kg/m^2]	SS 137244:1995	EN 1339:2003
C0	0.04	0.17	Good	Resistant
C15K	0.19	0.58	Inadmissible	Resistant
C30K	1.22	1.50	Inadmissible	Nonresistant
C15T	0.23	0.56	Inadmissible	Resistant
C30T	0.90	1.30	Inadmissible	Resistant
D0	0.08	0.11	Good	Resistant
D20T	0.10	0.15	Good	Resistant
D40T	0.50	0.67	Admissible	Resistant
D20K	0.20	0.36	Good	Resistant
D40K	1.20	1.41	Inadmissible	Nonresistant
F0	0.04	0.09	Very good	Resistant
F20K	0.25	0.52	Inadmissible	Resistant
F30K	0.31	0.85	Inadmissible	Resistant
F40K	0.88	1.52	Inadmissible	Resistant
G0	0.81	0.83	Admissible	Resistant
G30K	1.94	2.03	Inadmissible	Nonresistant
G30T	1.32	1.36	Inadmissible	Nonresistant

The sum of the numbers down the main diagonal divided by the total number of test examples determine classification accuracy.

Let's consider what can be done when the number of data for training and testing is limited. The simplest way to handle this situation is to reserve a certain number of examples for testing and to use the remainder for training. Of course, the selection should be done randomly. In practical terms, it is common to hold out one-third of the data for testing and use the remaining two-thirds for training (Witten and Frank 2005). The primary disadvantage of this simple method is that this random selection may be not representative. A more general way to mitigate any bias caused by the particular sample chosen for holdout is to repeat the whole process, training and testing, several times with different random samples. The random selection repeated many times can be treated as the basis of an statistical technique called *cross-validation*. In the *k-fold cross-validation*, the data set U is split into k approximately equal portions $(U = E_1 \cup \cdots \cup E_k)$ (Krawiec and Stefanowski 2003; Witten and Frank 2005). At each iteration i, the set E_i is used for testing, and the remainder $UmathordE_i$ is used for training. Overall, classification accuracy is calculated as an average from the classification accuracy for each iteration.

To generate rules describing the scaling resistance of concrete, several numerical experiments were performed by using program AQ21 (Wojtusiak 2004) and algorithm J48 from the WEKA workbench (Witten and Frank 2005). Algorithm AQ21, invented in the Machine Learning and Inference Laboratory of George Mason University, is based on the *covering approach* alike most of the rule-based data mining algorithms. Therefore, the AQ21 algorithm generates subsequent rules until all the examples (sometimes not all) are covered. The idea of adding a new rule or a new term to existing rule is to include as many instances of the desired class (*positive examples*) as possible and to exclude as many instances of other classes (*negative examples*) as possible.

The second considered algorithm, J48, is available as a part of the WEKA workbench, which was developed at the University of Waikato in New Zealand. The algorithm J48 is an implementation of the last publicly available version of an algorithm C4.5 devised by J. Ross Quinlan (Quinlan 1993). Construction of decision trees is based on a simple divide and conquer approach, which is well known in computer science. The primary problem is connected to a selection of tests (splits of attributes), which should be placed in the nodes. The test is good if it allows shortening of the way from the root to the leaves representing classes.

Application of program AQ21 and algorithm J48 from workbench WEKA provided the means for automatic classification of plain concretes and concretes modified with CFBC fly ash as materials of good or acceptable resistance to chloride penetration (Marks et al. 2009).

Search for the Rules by Describing Surface Scaling Resistance of Concrete

Generation of Rules by Using Criteria Defined in EN 1339

As a result of the experiments carried out on the specimens with different contents of fluidized fly ash, as shown in Tables 3 and 6, the following database consisting of 17 records was created. This database was used to determine the rules describing the scaling resistance of concrete after 28 cycles. The database with one nominal and 6 numerical attributes is presented in Table 7.

The attributes A_hr and $\bar{L}$ are standard parameters describing the air-void microstructure of concrete according to EN 480-11

Table 7. Database Containing Attributes of Concrete with CFBC Ash Additions

Record number	C1	pfT	pfK	w/b	A_hr	$\bar{L}$	Resistance
1	380	0	0	0.45	4.46	0.38	YES
2	323	0	57	0.45	4.83	0.28	YES
3	266	0	114	0.45	4.33	0.35	NO
4	323	57	0	0.45	4.7	0.34	YES
5	266	114	0	0.45	6.88	0.24	YES
6	406	0	0	0.42	10.08	0.07	YES
7	323	0	81	0.42	5.94	0.12	YES
8	290	73	0	0.42	18.41	0.05	YES
9	244	0	162	0.42	6.07	0.14	NO
10	217	145	0	0.42	16.56	0.08	YES
11	360	0	0	0.42	6.25	0.13	YES
12	288	0	72	0.42	6.25	0.13	YES
13	252	0	108	0.42	7.08	0.14	YES
14	216	0	144	0.42	6.03	0.18	YES
15	354	0	0	0.44	3.7	0.17	YES
16	246	0	105	0.44	2	0.27	NO
17	246	105	0	0.44	3.8	0.56	NO

Note: where: C1—cement content, [kg/m^3], pfT—the content of CFBC ash from brown coal content (power plant Turów), [kg/m^3], pfK—the content of CFBC ash from hard coal content (power station Katowice), [kg/m^3], w/b—water-to-binder ratio, A_hr—air content in the hardened concrete measured on polished sections by using microscopic image analysis, [%], $\bar{L}$—air voids spacing factor, which characterizes the spatial distribution of bubbles generated by the air entrainment [mm], Resistance—result of evaluation of the scaling resistance of concrete according to EN 1339 (YES, NO).

(EN 480-11 2008). The last attribute—resistance—is a nominal one that takes on two possible values: yes, no. In the considered database, 13 examples belonged to the [Resistance = YES] class and 4 examples belonged to the [Resistance = NO] class.

The aim of an experiment is to generate the set of rules which allow the determination of the surface scaling resistance of concrete. As a training set, all the instances from the database were considered. The rules generated by an AQ21 algorithm are presented next:

[Resistance = YES]

Rule 1

$$[\text{pfK} \leq 110] \text{ and } [A\_\text{hr} \geq 4.13]: p = 11, \quad n = 0, \quad q = 0.846 \tag{1a}$$

Rule 2

$$[\text{pfK} \leq 152] \text{ and } [\bar{L} \leq 0.255]: p = 10, \quad n = 0, \quad q = 0.769 \tag{1b}$$

[Resistance = NO]

Rule 1

$$[C1 \leq 294] \text{ and } [A\_\text{hr} \leq 4.58]: p = 3, \quad n = 0, \quad q = 0.75 \tag{1c}$$

Rule 2

$$[C1 = 230..248]: p = 3, \quad n = 0, \quad q = 0.75 \tag{1d}$$

where p = of positive examples in the training set covered by the rule; n = number of negative examples covered by the rule (i.e., the number of records from the other classes satisfying the rule); and q = quality of the rule, defined as:

$$q = \text{cov}^{1/2} \times \text{consig}^{1/2}$$

where

$$\text{cov} = \frac{p}{P} \quad \text{and} \quad \text{consig} = \left(\frac{p}{p+n} - \frac{P}{P+N}\right) \times \frac{P+N}{N}$$

are measures of the rule: coverage and consistency gain, respectively. In this equation, P and N are the numbers of positive and negative examples in the training data for that class, respectively (for the [Resistance = YES] class $P = 13$, $N = 4$ and for the [Resistance = NO] class $P = 4$, $N = 13$).

The rules shown in Eqs. (1*a*)–(1*d*) can be interpreted as described next.

The concrete is expected to be scaling resistant if

$$\text{pfK} \leq 110 \quad \text{and} \quad A\_\text{hr} \geq 4.13 \qquad \text{or}$$

$$\text{pfK} \leq 152 \quad \text{and} \quad \bar{L} \leq 0.255$$

The concrete will be nonscaling resistant if

$$C1 \leq 294 \quad \text{and} \quad A\_\text{hr} \leq 4.58 \qquad \text{or}$$

$$230 \leq C1 \leq 248$$

However, it should be emphasized that the presented rules are only applicable to concrete with the defined overall mass of cement and coal ash additions as given in Table 3.

To estimate the classification accuracy, the k-fold cross-validation was used, where k is the number of examples in the database. In this method, each example in turn is left out, and the learning method is trained on all the remaining examples. The results of k judgments, one for each member of the database, are averaged, and that average represents the classification accuracy (Witten and Frank 2005). This method, named *leave-one-out* cross-validation, is useful to the database of a small number of records. It seems to offer a chance of squeezing the maximum out of a small dataset and obtaining as accurate an estimate as possible.

The results of k judgments may be displayed as a two-dimensional confusion matrix with a row and a column for each class. Each confusion matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. The numbers of examples down the main diagonal are predicted correctly. The classification accuracy is the sum of numbers down the main diagonal divided by the total number of data set examples.

Applying the k-fold cross-validation for $k = 17$ (number of examples in Table 6), the confusion matrix is obtained as shown in Table 8.

The value of classification accuracy is equal to 76.5%. It is inferred from that value that the use of a classifier will provide fairly reliable results.

Table 8. Confusion Matrix

	Predicted class	
Actual class	Yes	No
Yes	11	2
No	2	2

Generation of Rules by Using Criteria Defined in SS 137244

The scaling resistance criteria defined in Table 5 depend on the mass of scaled material at the surface of concrete specimens after 56 cycles and the ratio of scaled material after 56 cycles to scaled material after 28 cycles. The classes of the scaling resistance of concrete determined by using this method are given in Table 6. Because of a small number of data (only 17 records), the classification was simplified by introducing only two classes of the surface scaling resistance (YES and NO). The class "YES" was assigned to cover the following categories of resistance: very good, good, and admissible. The class "NO" was equivalent to inadmissible frost salt scaling resistance. Table 9 provides the database generated by using those criteria along with other parameters, (C1, pfT, pfK, w/b, A_hr and $\bar{L}$) previously discussed.

The last attribute—resistance—is a nominal one and can only take two possible values: yes or no. In the considered database, 7 records belonged to the [Resistance = YES] class, and 10 records belonged to the [Resistance = NO] class.

To generate the rules, which allow us to determine the scaling resistance of concrete, the J48 algorithm was also used. As a training set, all the instances from the database were considered. The decision tree generated by an J48 algorithm is presented in Fig. 2 in which the number in brackets denotes the number of examples correctly classified (positive examples).

The obtained decision tree (Fig. 2) can be easily transformed into the following rules:

[Resistance = YES]
Rule1 [$C1 \leq 323$] and [[$\bar{L} \leq 0.12$]: $p = 3$, $n = 0$
Rule2 [$C1 > 323$]: $p = 4$, $n = 0$
[Resistance = NO]
Rule1 [$C1 \leq 323$] and [$\bar{L} > 0.12$]: $p = 10$, $n = 0$

Applying the k-fold cross-validation for $k = 17$ (value of parameter k is the number of examples), the result was obtained on the test set displayed as a two-dimensional confusion matrix for both classes as shown in Table 10.

Table 9. Database Containing Attributes of Concrete with CFBC Ash Additions

Record number	C1	pfT	pfK	w/b	A_hr	$\bar{L}$	Resistance
1	380	0	0	0.45	4.46	0.38	Yes
2	323	0	57	0.45	4.83	0.28	No
3	266	0	114	0.45	4.33	0.35	No
4	323	57	0	0.45	4.7	0.34	No
5	266	114	0	0.45	6.88	0.24	No
6	406	0	0	0.42	10.08	0.07	Yes
7	323	0	81	0.42	5.94	0.12	Yes
8	290	73	0	0.42	18.41	0.05	Yes
9	244	0	162	0.42	6.07	0.14	No
10	217	145	0	0.42	16.56	0.08	Yes
11	360	0	0	0.42	6.25	0.13	Yes
12	288	0	72	0.42	6.25	0.13	No
13	252	0	108	0.42	7.08	0.14	No
14	216	0	144	0.42	6.03	0.18	No
15	354	0	0	0.44	3.7	0.17	Yes
16	246	0	105	0.44	2	0.27	No
17	246	105	0	0.44	3.8	0.56	No

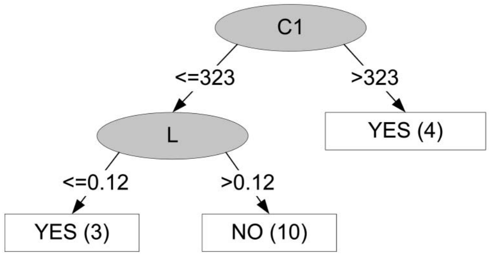


Fig. 2. The decision tree generated by an J48 algorithm describing the surface scaling resistance of concrete according to simplification of the SS 137244 criteria

Table 10. Confusion Matrix

	Predicted class	
Actual class	Yes	No
Yes	5	2
No	3	7

The value of classification accuracy is equal to 70.6%, so it is predicted correctly. Obtained rules show the significance of cement content and air-void characteristics.

Conclusions

The classifiers generated by computer programs AQ21 and WEKA by using the J48 algorithm have provided a simple automatic classification of the scaling resistance of the plain concretes and concretes modified with CFBC ash according to the introduced criteria. It has been found that both air-void microstructure parameters and the content of cement and CFBC ash play a significant role in providing the required concrete scaling resistance. The classifiers were evaluated by using the *k*-fold cross-validation, where *k* was the number of instances in the data set. The obtained values of the classification accuracy on both data sets, determined according to SS 137244 and EN 1339 standard, was respectively 70.6 and 76.5%. These values seem to be sufficient to acknowledge the correctness of the classifiers. Moreover, both classifiers have an additional and very useful property. They are able not only to predict a proper value of the class attribute for unseen examples, but their knowledge is represented in a clear way, which is understandable for an expert. The relationship between the attributes and decision class, which is expressed in an explicit way, can be used for the design of new concrete mixes. Because of a small number of tested specimens, the rules are applicable only to concrete mix compositions of similar binder content and similar values of water-to-binder ratio. Further tests are needed to enlarge the experimental database and to cover a broader range of concrete compositions.

References

Aïtcin, P.-C. (1998). *High performance concrete*, E&FN SPON, London.

Alterman, D., and Kasperkiewicz, J. (2006). "Evaluating concrete materials by application of automatic reasoning." *Bull. Pol. Ac.: Tech.*, 54(4), 353–361.

ASTM. (1991). "Standard test method for microscopical determination of parameters of the air-void system in hardened concrete." *Annual Book of ASTM Standards*, *C457*, West Conshohocken, PA.

European Committee for Standardization. (2000). "*Concrete-Specification, performance, production and conformity.*" *EN 206-1*, Brussels.

European Committee for Standardization. (2003). "*Concrete paving flags—Requirements and test methods*." *EN 1339*, Brussels.

European Committee for Standardization. (2005). "*Fly ash for concrete—Part 1. Definition, specifications and conformity criteria*." *EN 450-1*, Brussels.

European Committee for Standardization. (2006). "*Testing hardened concrete—Part 9: Freeze—thaw resistance—scaling*." *CEN/TS 12390-9*, Brussels.

European Committee for Standardization. (2008). "*Admixtures for concrete, mortar and grout. Test methods. Determination of air void characteristics in hardened concrete*." *EN 480-11*, Brussels.

Fu, X., Li, Q., Zhai, J., Sheng, G., and Li, F. (2008). "The physical-chemical characterization of mechanically-treated CFBC fly ash." *Cem. Concr. Compos.*, 30(3), 220–226.

Gencel, O., Brostow, W., Tea, D., and Thedford, M. (2011). "Workability and mechanical performance of steel-fiber-reinforced self-compacting concrete with fly ash." *Composite Interfaces*, 18(2), 169–184.

Gencel, O., Kocabas, F., Gok, M. S., and Koksal, F. (2011). "Comparison of artificial neural networks and general linear model approaches for the analysis of abrasive wear of concrete." *Constr. Build. Mater.*, 25(8), 3486–3494.

Giergiczny, Z., and Pużak, T. (2008). "Properties of concrete with fluidal fly ash addition." *Proc., 3rd Int. Symp. Non-Traditional Cement & Concrete*, Brno University of Technology, Brno, 263–271.

Glinicki, M. A., and Zielinski, M. (2008). "Air void system in concrete containing circulating fluidized bed combustion fly ash." *Mater. Struct.*, 41(4), 681–687.

Glinicki, M. A., and Zielinski, M. (2009). "Frost salt scaling resistance of concrete containing CFBC fly ash." *Mater. Struct.*, 42(7), 993–1002.

Goodarzi, F. (2006). "Charakteristics and composition of fly ash from canadian coal-fired power plants." *Fuel*, 85(10–11), 1418–1427.

Helmuth, R. (1987). *Fly ash in cement and concrete*, Portland Cement Association, Skokie, IL.

Jóźwiak-Niedźwiedzka, D. (2009). "Effect of fluidized bed combustion fly ash on the chloride resistance and scaling resistance of concrete." *Proc., RILEM Conf.—Concrete in Aggressive Aqueous Environ., Performance, Testing and Modeling*, RILEM, Bagneux, 556–563.

Khurana, R., and Saccone, R. (2001). "Fly ash in self-compacting concrete." *Proc., 7th CANMET/ACI Int. Conf. Fly Ash, Silica Fume, Slag and Natural Pozzolans in Concrete*. American Concrete Institute SP-199, Farmington Hills, MI, 259–274.

Krawiec, K., and Stefanowski, J. (2003). *Machine learning and neural networks (in Polish)*, Poznan Univ. of Technology, Poznań, Poland.

Małolepszy, J., and Kołodziej, Ł. (2009). "Resistance of cements with high amount of ashes from fluidized bed furnace to chloride diffusion." *Proc., 5th Int. Conf. Concrete and Concrete Structures*, University of Žilina, Žilina.

Marks, M., Jóźwiak-Niedźwiedzka, D., and Glinicki, M. A. (2009). "Application of machine learning for prediction of concrete resistance to migration of chlorides." *Proc., 9th Int. Symp. on Brittle Matrix Composites BMC9*, Woodhead, Cambridge and IFTR, Warsaw, 227–236.

Melhem, H. G., and Cheng, Y. (2003). "Prediction of remaining service life of bridge decks using machine learning." *J. Comput. Civ. Eng.*, 17(1), 1–9.

Michalski, R. S. (1983). "A theory and methodology of inductive learning." *Machine learning. An artificial intelligence approach*, R. S. Michalski, T. J. Carbonell, and T. M. Mitchell, eds., TIOGA Publishing, Palo Alto, CA, 83–134.

Ni, H. G., and Wang, J. Z. (2000). "Prediction of compressive strength of concrete by neural networks." *Cem. Concr. Res.*, 30(8), 1245–1250.

Nowak, W. (2003). "Clean coal fluidized-bed technology in Poland." *Appl. Energy*, 74(3–4), 405–413.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*, Morgan Kaufmann Publishers, San Mateo, CA.

Stevens, W., Robl, T., and Mahboub, K. (2009). "The cementitious and pozzolanic properties of fluidized bed combustion fly ash." *Proc., World of Coal Ash (WOCA) Conf.*, University of Kentucky, Lexington, KY.

Swedish Standards (SS). (1995). “Concrete testing—hardened concrete—scaling at freezing.” *137244*, SIS, Stockholm, Sweden.

Szczygielski, T., and Hycnar, J. (2009). “HCFA and CO_2 savings in Europe—Resources, markets, technologies.” *Proc., World of Coal Ash (WOCA) Conf.*, University of Kentucky, Lexington, KY.

Topçu, I. B., and Saridemir, M. (2008). “Prediction of mechanical properties of recycled aggregate concretes containing silica fume using artificial neural networks and fuzzy logic.” *Comput. Mater. Sci.*, 42(1), 74–82.

Witten, I. H., and Frank, E. (2005). *Data mining. Practical machine learning tools and techniques*, Elsevier, San Francisco.

Wojtusiak, J. (2004). “AQ21 user’s guide.” *Reports of the Machine Learning and Inference Laboratory, MLI 04-3*, George Mason Univ., Fairfax, VA.

Yeh, I. C. (1998). “Modeling of strength of high—Performance concrete using artificial neural networks.” *Cem. Concr. Res.*, 28(12), 1797–1808.

materials

MDPI

Article

Prediction of the Chloride Resistance of Concrete Modified with High Calcium Fly Ash Using Machine Learning

Michał Marks [1,*], Michał A. Glinicki [2] and Karolina Gibas [2]

Received: 9 October 2015; Accepted: 30 November 2015; Published: 11 December 2015
Academic Editor: Prabir Sarker

1 Research and Academic Computer Network, Wawozowa 18, Warsaw 02-796, Poland
2 Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, Warsaw 02-106, Poland; mglinic@ippt.gov.pl (M.A.G.); kgibas@ippt.gov.pl (K.G.)
* Correspondence: mmarks@nask.pl; Tel.: +48-22-380-8200; Fax: +48-22-380-8201

Abstract: The aim of the study was to generate rules for the prediction of the chloride resistance of concrete modified with high calcium fly ash using machine learning methods. The rapid chloride permeability test, according to the Nordtest Method Build 492, was used for determining the chloride ions' penetration in concrete containing high calcium fly ash (HCFA) for partial replacement of Portland cement. The results of the performed tests were used as the training set to generate rules describing the relation between material composition and the chloride resistance. Multiple methods for rule generation were applied and compared. The rules generated by algorithm J48 from the Weka workbench provided the means for adequate classification of plain concretes and concretes modified with high calcium fly ash as materials of good, acceptable or unacceptable resistance to chloride penetration.

Keywords: chloride penetration; concrete; durability; high calcium fly ash; machine learning

---

1. Introduction

The increased use of high calcium fly ash (HCFA) for partial replacement of Portland cement in concrete could result in a number of environmental benefits (reduced consumption of cement clinker, reduced CO_2 emissions during cement production, saving natural resources, reduced landfill space and storage costs). The resources of high calcium fly ash are large, it is produced as a by-product of power generation in brown coal burning plants. However, this type of ash is usually characterized by low silica content, a high content of free lime and an increased content of sulfur compounds. It could be used in concrete following the requirements of ASTM (American Society for Testing and Materials) C618 Class C, but in Europe, it does not meet the requirements defined in standard EN 450-1. At present, HCFA is not in common use in European countries in spite of positive examples of its suitability provided by Greek and Turkish researchers. It was shown [1] that in the case of cement replacement with HCFA, the compressive strength of concrete was increased if the content of active silica in the fly ash was higher than that in the cement. Similar results were obtained earlier by Naik, *et al.* [2]: partial replacement of cement by fine-grained HCFA resulted in the same or better compressive strength of concrete; the results for drying shrinkage were also positive. The optimization of fineness coupled with the adjustment of water content were found as the key parameters of the effective utilization of high calcium fly ashes for strength maximization of cement mortars [3]. The application of HCFA as a partial cement replacement in mortar beams stimulated self-healing of cracks and particularly of microcracks [4]. It was also found that concrete specimens incorporating HCFA exposed to long-term chloride ponding experiments exhibited significantly lower total chloride content for all depths from the surface [5]. The key factors for the adequate performance of HCFA in concrete seem to be both the composition and the gradation of fly ash.

The assessment of concrete resistance to chloride ingress is fundamental for the durability of reinforced concrete structures exposed to deicing salt and the marine environment [6]. Numerous papers on chloride penetration resistance of concrete modified with standard siliceous fly ash were recently reviewed in [7]. The addition of fly ash is generally found (and confirmed in [8]) to reduce chloride permeability and also to increase the chloride binding capacity of concrete. Despite lower chloride threshold values, the addition of fly ash was found to provide better corrosion protection to steel reinforcements. There is a need to extend such a study to include high calcium fly ash. For rational use of HCFA in structural concrete, there is also a need to propose tools for the prediction of the chloride penetration resistance of concrete.

The prediction of the engineering properties of composite materials is usually based on experimental test results with a reference to the observed material microstructure. The relevant material characteristics can be extracted from an experimental dataset using various artificial intelligence methods, developed for the last two decades for various engineering applications [9,10]. Artificial neural networks were successfully applied for the prediction of the compressive strength of concrete containing silica fume [11] or coal ash [12]. Moreover, the application of neural networks and optimization technologies created the possibility to search for the optimum mixture of concrete: the mixture with the lowest cost and required performance, such as strength and slump [13]. Machine learning methods were also tested on the classification of concrete modified by fluidized bed fly ash as materials of adequate resistance to chloride penetration [14] and resistance to surface scaling [15]. The application of machine learning for the prediction of the scaling resistance of concrete modified with high calcium fly ash is described in [16]. The authors of [17,18] proposed to combine artificial neural networks and machine learning methods in one system to estimate and predict various properties of concrete materials.

The aim of this study is to generate rules using a machine learning algorithm to evaluate the chloride resistance of concrete modified with high calcium fly ash. The rules are generated using selected attributes from a database created by storing the experimental results of the chloride migration coefficient determined for three concrete series.

2. Composition of Concrete Mixes and Test Results of the Chloride Migration Coefficient

The chloride migration coefficient in concrete specimens with different contents of high calcium fly ash was experimentally measured. Concrete mixes were prepared with high calcium fly ash used for replacement of 15% or 30% of the cement mass. Experimental tests were performed on several mixes. For concrete manufacturing, two types of Portland cement, CEM I 42.5R (with 10% C_3A content) or CEM I 42.5 HSR NA (with 2% C_3A content), siliceous sand fraction $0 \div 2$ mm and amphibolite as a coarse aggregate (two fractions $2 \div 8$ mm and $8 \div 16$ mm) were used. The following admixtures were used: a high range water reducer (based on polycarboxylate ethers) and a plasticizer (lignosufonate). Because of the expected variability of ash properties, three lots of high calcium fly ash were tested from different deliveries from the power plant, namely S1, 16 March 2010, S2, 19 May 2010, and S3, 28 June 2010. The chemical composition of HCFA is given in Table 1. For HCFA beneficiation, a grinding process was applied during 10–28 minutes in a ball mill. The physical properties of ash before and after grinding are given in Table 2 [19]. HCFA was used as an additive to concrete mix in unprocessed form (as collected) and after grinding.

The Nordtest Method Build 492—Non-Steady State Migration Test [20] was used to determine the chloride migration coefficient. The principle of the test is to subject the concrete specimen to external electrical potential applied across it and to force chloride ions to migrate into the concrete. The specimens are then split open and sprayed with silver nitrate solution, which reacts to give white insoluble silver chloride on contact with chloride ions. This provides a possibility to measure the depth to which a sample has been penetrated. The non-steady-state migration coefficient, D_{nssm}, is determined on the basis of Fick's second law. This coefficient is dependent on the voltage magnitude, the temperature of the anolyte measured at the beginning and the end of test and the depth of chloride ions' penetration. The criteria for evaluating the resistance of concrete against chloride penetration proposed by L. Tang [21] are shown in Table 3.

Table 1. The chemical composition of high calcium fly ashes from Bełchatów power plant in Poland, determined using the XRF (X-ray fluorescence) method. Fly ash sampling date and bath designation [19].

Component	Fly Ash Sampling Date and Batch Designation		
	16.03.2010 S1	19.05.2010 S2	28.06.2010 S3
LOI	2.56%	3.43%	1.85%
SiO_2	33.62%	35.41%	40.17%
Al_2O_3	19.27%	21.86%	24.02%
Fe_2O_3	5.39%	6.11%	5.93%
CaO	31.32%	25.58%	22.37%
MgO	1.85%	1.49%	1.27%
SO_3	4.50%	4.22%	3.07%
K_2O	0.11%	0.13%	0.20%
Na_2O	0.31%	0.16%	0.15%
P_2O_5	0.17%	0.16%	0.33%
TiO_2	1.21%	1.22%	1.01%
Mn_2O_3	0.07%	0.06%	0.06%
SrO	0.20%	0.17%	0.16%
ZnO	0.02%	0.02%	0.02%
CaO_{free}	2.87%	1.24%	1.46%

Table 2. Physical properties of high calcium fly ashes before and after processing [19].

Batch	Fly Ash Designation	Density (g/cm^3)	Fineness: The Residue on Sieve 45 μm (%)	Specific Surface by Blaine (cm^2/g)
S1	S1: unprocessed	2.62	38.0	2860
	$S1_{10}$: ground 10 min	2.77	23.0	3500
	$S1_{28}$: ground 28 min	2.75	10.5	3870
S2	S2: unprocessed	2.58	35.4	4400
	$S2_{15}$: ground 15 min	2.70	13.3	6510
S3	S3: unprocessed	2.64	55.6	1900
	$S3_{20}$: ground 20 min	2.71	20.0	4060

Table 3. Criteria for the classification of the concrete resistance to chloride ions' penetration [21].

Chloride Migration Coefficient D_{nssm}	Resistance to Chloride Penetration
$<2 \times 10^{-12}$ m^2/s	Very good
2–8 $\times 10^{-12}$ m^2/s	Good
8–16 $\times 10^{-12}$ m^2/s	Acceptable
$>16 \times 10^{-12}$ m^2/s	Unacceptable

Experimental tests revealed a decrease of the chloride migration coefficient with the increase in the HCFA amount added to the mix. The most significant reduction of D_{nssm} by 36%–75% and 54%–89% after 28 and 90 days of curing, respectively, was obtained when using ground HCFA to substitute 30% of binder mass. With a such reduction of D_{nssm}, the level of chloride resistance changed from acceptable to good or from unacceptable to acceptable, [22]. For a few mixes prepared with a water-to-binder ratio of 0.60, a change of D_{nssm} did not increase the level of chloride penetration resistance. Sieving through a 0.125-mm mesh size sieve was found to improve HCFA performance: it significantly reduced the value of D_{nssm}, which was most evident after 90 days of curing. No clear relationship could be found between D_{nssm} and the water-to-binder ratio or the compressive strength of concrete.

The resistance against chloride ingress of concrete containing low calcium fly ash was previously tested by Baert, *et al.* [23], and at 28 days, the chloride migration coefficient was increased with increasing fly ash content. However at later ages (3, 6 or 12 months), due to the pozzolanic reaction, the D_{nssm} coefficient was lower for all concrete mixes with siliceous fly ash. The effects of blast furnace slag on the chloride migration coefficient summarized by Gjorv [6] were clearly favorable, even at the age of 14 days. After 28 days of water curing, the increasing amounts of slag up to 80% replacement resulted in the reduced apparent chloride diffusion coefficient from 11×10^{-12} down to 2×10^{-12} m/s^2. The comparison with the obtained results on HCFA in concrete reveals almost comparable efficiency as blast furnace slag. This could be attributed to both pozzolanic and hydraulic activity of HCFA. The hydraulic properties of these fly ashes should be related to reactive aluminate phases and their hydration and also to the formation of ettringite in the initial phase of hydration [24]. A high hydraulic and pozzolanic activity index after a prolonged hydration and hardening process is connected with hydraulic phases, mainly belite and gehlenite, as well as with the reactivity of the glassy phase. The complexity of the phenomena involved in chloride ion penetration in concrete containing such a mineral addition of pozzolanic and hydraulic activity justifies an application of machine learning techniques to reveal the possible governing rules.

In Table 4, the database containing data on the composition of the concrete mixes, the specific surface of fly ash obtained by the Blaine method and the chloride migration coefficient determined after 28 days of curing is presented. The estimation of the concrete resistance to chloride penetration, based on the values of the diffusion coefficients according to the criterion presented in Table 3, is placed in the last column of Table 4.

The permeability of concrete is known to be dependent largely on the water-to-cement ratio, (w/c). However the definition of w/c is not unambiguous when using supplementary cementitious materials. Following the EN 206 standard, the effect of active mineral additions on w/c is quantified using the k-efficiency factor: the content of the additive (a) is multiplied with a k-value, and the water to cement ratio (w/c) is replaced by $(w/c)_{eq} = w/(c + k \cdot a)$. The efficiency k factor approach is adequate to address the mix design for compressive strength when using the additives of the established efficiency. Even in such a case, like siliceous fly ash, the efficiency factors are not the same for durability performance and for the compressive strength [25]. The compiled fly ash efficiency data [6,26] revealed a much higher efficiency coefficient k in relation to the compressive strength than the value given in EN 206, even reaching the value of two in relation to the resistance to chloride attack. For nonstandard fly ashes and coal combustion products from so-called clean coal technology, the efficiency factors are not established [27]. Therefore, it is not possible to describe all of the effects of the nonstandard fly ashes, including HCFA, on concrete performance when exposed to various environmental factors with only one efficiency coefficient. In order to avoid an unambiguous (w/c) definition, the content of water in the mix is used as a descriptor in the machine learning database.

Table 4. The database of the composition of concrete mixes and the properties of hardened concretes.

Concrete Mix	Content (kg/m^3)											Specific Surface of Fly Ash (cm^2/g)	Chloride Migration Coefficient ($\times 10^{-12}$ m^2/s)	Category of Resistance to Chloride Penetration
	Cement CEM I 42.5		High Calcium Fly Ash							Aggregate	Water			
	10% C_3A	2% C_3A												
mix	**C1**	**C2**	**S1**	**S1$_{10}$**	**S1$_{28}$**	**S2**	**S2$_{15}$**	**S3**	**S3$_{20}$**	**K0$_{16}$**	**w**	**surf**	**D$_{nssm}$**	**resistance**
R_38	359	0	0	0	0	0	0	0	0	1945	156	0	10.13	acceptable
R_39	305	0	137	0	0	0	0	0	0	1848	153	2860	7.88	good
R_41	250	0	268	0	0	0	0	0	0	1741	152	2860	3.76	good
R_42	323	0	0	0	0	0	0	0	0	1938	174	0	23.73	unacceptable
R_43	272	0	120	0	0	0	0	0	0	1837	169	2860	12.36	acceptable
R_44	226	0	241	0	0	0	0	0	0	1768	169	2860	8.10	acceptable
R_47	310	0	0	139	0	0	0	0	0	1892	140	3500	5.44	good
R_48	257	0	0	275	0	0	0	0	0	1802	142	3500	3.42	good
R_49	275	0	0	121	0	0	0	0	0	1872	160	3500	17.79	unacceptable
R_50	228	0	0	244	0	0	0	0	0	1800	159	3500	10.37	acceptable
R_51	306	0	0	0	137	0	0	0	0	1852	153	3870	6.37	good
R_52	255	0	0	0	273	0	0	0	0	1780	153	3870	3.85	good
R_53	277	0	0	0	122	0	0	0	0	1871	175	3870	12.22	acceptable
R_54	228	0	0	0	244	0	0	0	0	1784	173	3870	5.52	good
R_75	0	366	0	0	0	0	0	0	0	1997	143	0	11.96	acceptable
R_76	0	312	140	0	0	0	0	0	0	1901	142	2860	6.34	good
R_77	0	251	270	0	0	0	0	0	0	1765	140	2860	4.04	good
R_78	0	328	0	0	0	0	0	0	0	1982	165	0	21.91	unacceptable
R_79	0	278	123	0	0	0	0	0	0	1894	159	2860	10.30	acceptable
R_80	0	226	242	0	0	0	0	0	0	1790	157	2860	7.88	good
R_81	0	304	0	0	0	136	0	0	0	1861	133	4400	5.04	good
R_82	0	277	0	0	0	122	0	0	0	1889	158	4400	7.76	good
R_116	340	0	0	0	0	0	0	0	0	1841	170	0	20.79	unacceptable
R_125	296	0	0	0	0	0	0	75	0	1836	174	1900	8.17	acceptable
R_118	237	0	0	0	0	0	0	145	0	1767	172	1900	10.95	acceptable
R_117	295	0	0	0	0	0	0	0	74	1826	174	4060	12.00	acceptable
R_119	239	0	0	0	0	0	0	0	147	1781	171	4060	5.17	good
R_107	308	0	0	0	0	0	0	0	0	1846	186	0	26.00	unacceptable
R_102	265	0	0	0	0	0	0	67	0	1834	189	1900	22.80	unacceptable
R_103	218	0	0	0	0	0	0	134	0	1814	189	1900	20.86	unacceptable
R_105	265	0	0	0	0	0	0	0	67	1839	189	4060	12.10	acceptable
R_104	219	0	0	0	0	0	0	0	135	1820	190	4060	7.59	good
R_120	0	343	0	0	0	0	0	0	0	1862	172	0	23.09	unacceptable

Table 4. *Cont.*

Concrete Mix	Content (kg/m^3)											Specific Surface of Fly Ash (cm^2/g)	Chloride Migration Coefficient ($\times 10^{-12}$ m^2/s)	Category of Resistance to Chloride Penetration
	Cement CEM I 42.5		High Calcium Fly Ash							Aggregate	Water			
	10% C_3A	2% C_3A												
mix	**C1**	**C2**	**S1**	**$S1_{10}$**	**$S1_{28}$**	**S2**	**$S2_{15}$**	**S3**	**$S3_{20}$**	**$K0_{16}$**	**w**	**surf**	**D_{nssm}**	**resistance**
R_126	0	290	0	0	0	0	0	73	0	1793	170	1900	22.87	unacceptable
R_122	0	239	0	0	0	0	0	146	0	1779	171	1900	21.85	unacceptable
R_121	0	295	0	0	0	0	0	0	74	1824	173	4060	19.61	unacceptable
R_123	0	240	0	0	0	0	0	0	147	1786	171	4060	17.65	unacceptable
R_106	0	312	0	0	0	0	0	0	0	1869	189	0	28.50	unacceptable
R_111	0	265	0	0	0	0	0	67	0	1836	187	1900	31.63	unacceptable
R_112	0	222	0	0	0	0	0	136	0	1840	191	1900	27.44	unacceptable
R_110	0	265	0	0	0	0	0	0	67	1840	187	4060	25.42	unacceptable
R_108	0	223	0	0	0	0	0	0	137	1852	192	4060	23.04	unacceptable
A_0	350	0	0	0	0	0	0	0	0	1890	158	0	14.38	acceptable
A_15	298	0	133	0	0	0	0	0	0	1800	158	2860	7.91	good
B_15	298	0	0	133	0	0	0	0	0	1800	158	3500	6.39	good
C_15	298	0	0	0	133	0	0	0	0	1800	158	3870	5.52	good
A_30	245	0	263	0	0	0	0	0	0	1710	158	2860	5.43	good
B_30	245	0	0	263	0	0	0	0	0	1710	158	3500	1.63	very good
C_30	245	0	0	0	263	0	0	0	0	1710	158	3870	1.52	very good
D_15	298	0	0	0	0	133	0	0	0	1800	158	4400	3.06	good
E_15	298	0	0	0	0	0	133	0	0	1800	158	6510	2.06	good
H_0	0	350	0	0	0	0	0	0	0	1880	175	0	37.04	unacceptable
H_15M	0	298	0	0	0	0	0	0	75	1847	175	4060	34.48	unacceptable
H_15S	0	298	0	0	0	0	0	75	0	1847	175	1900	33.03	unacceptable
H_30M	0	245	0	0	0	0	0	0	150	1813	175	4060	27.41	unacceptable
H_30S	0	245	0	0	0	0	0	150	0	1813	175	1900	27.59	unacceptable

The database presented in Table 4 is a general database, which can be transformed into a “working database” by column selection.

3. Machine Learning Methods Used in the Prediction of the Engineering Properties of Composite Materials

3.1. Introduction to Machine Learning

Determining the relationship between material composition and the chloride resistance of concrete is a difficult and time-consuming process, even in the case of a small dataset, as presented in Table 4. For the considered dataset, it requires simultaneous analysis of 12 attributes (columns) for over 50 examples (rows). This task can be done manually; however, using a computer system to support data exploration is much more efficient. The branch of artificial intelligence concerned with applying algorithms that let computers evolve patterns using empirical data is called machine learning.

The aim of machine learning is to automatically learn to recognize complex patterns and make intelligent decisions based on the dataset. By a dataset, we mean a collection of logically-related records: a database. Each record can be called an instance or example, and each one is characterized by the values of predetermined attributes. The difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too large to be covered by the set of observed examples (training data). Hence, the learner must generalize from the given examples, so as to be able to produce a useful output in new cases.

Patterns recognition associated usually with classification is the most popular example of utilizing machine learning. However machine learning or, more general, statistical algorithms can support the knowledge discovery at different stages from outlier detection and attribute (features) selection to knowledge modeling and model validation.

3.2. Feature Selection

Feature selection, also known as attribute selection or feature reduction, is the technique of selecting a subset of relevant features for building robust learning models. By removing most irrelevant and redundant attributes from the data, feature selection helps improve the performance of learning models by: speeding up the learning process and alleviating the effect of the curse of dimensionality. Moreover, the irrelevant attributes degrade the performance of state-of-the-art decision tree and rule learners [28].

3.3. Classification

As was written earlier in Section 3.1, classification is the most common type of machine learning application. The goal of the classification process is to find a way of classifying unseen examples based on the knowledge extracted from the provided set of classified instances. Extracting the knowledge from the provided dataset requires the attribute set characterizing the example to be divided into two groups: the class attribute and the non-class attributes. For unseen instances, only non-class attributes are known; hence, the aim of data mining algorithms is to create such a knowledge model that allows predicting the example class membership based only on non-class attributes.

The knowledge model depends on the way the classifier is constructed, and it can be represented by classification rules (the algorithm AQ21 [29]), decision trees (e.g., algorithm C4.5, [30]) or many other representations. Regardless of the representation, both classification rules and decision trees algorithms create hypotheses.

In the considered problem, the chloride resistance of concrete (class attribute) depending on the material composition and some predictions of the concrete (non-class attributes) is searched. We concentrated on the most popular representative of decision tree classifiers,, the J48 algorithm, the open-source implementation of the last publicly-available version of a C4.5 method developed by

J. Ross Quinlan [30]. This algorithm was compared to selected algorithms available in Weka [28] in Section 4.2.

3.4. Classifier Evaluation

So as to evaluate the classifier, *i.e.*, to judge the hypotheses generated from the provided training set, we have to verify the classifier performance on the independent dataset, which is called the testing set. The classifier predicts the class of each instance from the test set; if it is correct, it is counted as a success; if not it, is an error. The measure of the overall performance of the classifier is the classification accuracy. This is the number of correct classifications of the instances from the test set divided by the total number of these instances, expressed as a percentage. The greater the classification accuracy, the better is the classifier.

In order to get a deeper understanding of which types of errors are the most frequent, the result obtained from a test set is often displayed as a two-dimensional confusion matrix with a row and a column for each class. Each matrix element shows the number of test examples, for which the actual class is the row and the predicted class is the column. Good results correspond to large numbers down the main diagonal and small, ideally zero, for the elements off the diagonal. The sum of the numbers down the main diagonal divided by the total number of test examples determine the classification accuracy.

Let's consider what can be done when the number of data for training and testing is limited. The simplest way to handle this situation is to reserve a certain number of examples for testing and to use the remainder for training. Of course, the selection should be done randomly. The main disadvantage of this simple method is that this random selection may not be representative. A more general way to mitigate any bias caused by the particular sample chosen for hold out is to repeat the whole process, training and testing, several times with different random samples. The random selection repeated many times can be treated as the basis of a statistical technique called cross-validation. In the k-fold cross-validation, the dataset U is split into k approximately equal portions ($U = E_1 \bigcup ... \bigcup E_k$) [31]. In each iteration i, the set E_i is used for testing, and the remainder $U \setminus E_i$ is used for training. Overall classification accuracy is calculated as an average from the classification accuracy for each iteration.

When we have only one database consisting of a very small number of records, the estimation of classification accuracy (the measure of the overall performance of the classifier) can be done using the n-fold cross-validation, where n is the number of examples in the database. In this method, called leave-one-out cross-validation, each example in turn is left out, and the learning method is trained on all of the remaining examples. It is judged by its correctness on the remaining example, one or zero for success or failure, respectively. The results from n judgments, one for each member of the database, are averaged, and that average represents the classification accuracy [28].

4. Searching for the Rules Describing the Chloride Resistance of Concrete Modified with HCFA

4.1. Feature Selection

In Table 4, the dataset with 12 attributes is presented. It is clear that for database with a few dozens of instances, this number of attributes is too large. Some attributes can be eliminated, but it is important to eliminate the most irrelevant attributes.

Therefore, we decided to evaluate a subset of attributes using the best first and exhaustive approaches to feature selection. The best first method searches the space of attributes by greedy hill climbing augmented with backtracking facility. In both cases, the *CfsSubsetEvaluator*, provided by Weka, was used to assess the predictive ability of each attribute individually and the degree of redundancy among them, preferring sets of attributes that are highly correlated with the class, but have low inter-correlation. Both methods of searching (best first and exhaustive) resulted in selection of C1, $S1_{28}$, w and surf attributes as a percent of tests, as presented in Table 5.

Table 5. Attribute selection cross-validation results.

Attribute	C1	C2	S1	$S1_{10}$	$S1_{28}$	S2	$S2_{15}$	S3	$S3_{20}$	$K0_{16}$	w	surf
Best First	100%	0%	0%	0%	32%	0%	0%	0%	0%	0%	100%	100%
Exhaustive Search	98%	0%	0%	0%	32%	0%	0%	0%	0%	0%	100%	100%

Therefore, in order to generate rules describing the chloride resistance of concrete modified with high calcium fly ash, the subset of attributes (C1, cement content with 10 percent of C_3A content (kg/m^3), $S1_{28}$, high calcium fly ash ground 28 minutes content (kg/m^3), w, water content (kg/m^3), surf, specific surface of fly ash obtained by the Blaine method (cm^2/g), and resistance, concrete resistance to chloride penetration (acceptable, good, unacceptable)) from the database (Table 4) is used. The shrunken database containing 56 records, each one described by four numerical and one nominal attributes, is presented in Table 6. The last attribute, resistance, denotes a class and can take one of three values (good, acceptable or unacceptable). Since the class "very good" representation is not sufficient (only two examples), we decided to assign them to the "good" class, which now covers 22 examples.

Table 6. The database.

Number	C1	$S1_{28}$	w	surf	resistance
1	359	0	156	0	acceptable
2	305	0	153	2860	good
3	250	0	152	2860	good
4	323	0	174	0	unacceptable
5	272	0	169	2860	acceptable
6	226	0	169	2860	acceptable
7	310	0	140	3500	good
8	257	0	142	3500	good
9	275	0	160	3500	unacceptable
10	228	0	159	3500	acceptable
11	306	137	153	3870	good
12	255	273	153	3870	good
13	277	122	175	3870	acceptable
14	228	244	173	3870	good
15	0	0	143	0	acceptable
16	0	0	142	2860	good
17	0	0	140	2860	good
18	0	0	165	0	unacceptable
19	0	0	159	2860	acceptable
20	0	0	157	2860	good
21	0	0	133	4400	good
22	0	0	158	4400	good
23	340	0	170	0	unacceptable
24	296	0	174	1900	acceptable
25	237	0	172	1900	acceptable
26	295	0	174	4060	acceptable
27	239	0	171	4060	good
28	308	0	186	0	unacceptable
29	265	0	189	1900	unacceptable
30	218	0	189	1900	unacceptable
31	265	0	189	4060	acceptable
32	219	0	190	4060	good
33	0	0	172	0	unacceptable
34	0	0	170	1900	unacceptable
35	0	0	171	1900	unacceptable
36	0	0	173	4060	unacceptable
37	0	0	171	4060	unacceptable
38	0	0	189	0	unacceptable
39	0	0	187	1900	unacceptable
40	0	0	191	1900	unacceptable
41	0	0	187	4060	unacceptable
42	0	0	192	4060	unacceptable

Table 6. *Cont.*

Number	C1	$S1_{28}$	w	surf	resistance
43	350	0	158	0	acceptable
44	298	0	158	2860	good
45	298	0	158	3500	good
46	298	133	158	3870	good
47	245	0	158	2860	good
48	245	0	158	3500	good
49	245	263	158	3870	good
50	298	0	158	4400	good
51	298	0	158	6510	good
52	0	0	175	0	unacceptable
53	0	0	175	4060	unacceptable
54	0	0	175	1900	unacceptable
55	0	0	175	4060	unacceptable
56	0	0	175	1900	unacceptable

4.2. Classification

As was mentioned in Section 3.3, the chloride resistance of concrete depending on material composition can be searched using one of many software suites available on the market, and we decided to utilize the Weka workbench. The Weka workbench provides over one hundred algorithms supporting classification. They belong to different types, like: Bayesian classifiers, rule classifiers, tree classifiers or meta classifiers. In our research, we decided to determine the chloride resistance of concrete using the selected 20 algorithms belonging to three different types of algorithms. As a training set, all of the instances from the database (Table 6) were considered. The classification accuracy was evaluated using leave-one-out cross-validation. The obtained results are collected in Table 7.

Table 7. Results obtained for different classifiers from the Weka workbench.

Number	Classifier	Accuracy
	Bayesian Classifiers	
1	BayesNet	66.07
2	ComplementNaiveBayes	62.50
3	NaiveBayes	73.21
	Tree Classifiers	
4	BFTree	73.21
5	DecisionStump	73.21
6	FT	78.57
7	LADTree	82.14
8	**J48**	**89.29**
9	LMT	82.14
10	NBTree	78.57
11	REPTree	64.29
12	SimpleCart	71.43
	Rule Classifiers	
13	ConjunctiveRule	71.43
14	DecisionTable	71.43
15	DTNB	80.36
16	JRip	62.50
17	NNge	76.79
18	OneR	71.43
19	PART	76.79
20	Ridor	66.07

The best accuracy equaling almost 90% was obtained using the J48 algorithm. The decision tree generated by the J48 algorithm is presented in Figure 1, where the first number in brackets denotes the number of examples from the training set covered by a selected leaf, and the second number, just after the sign "/", indicates the number of incorrectly-classified instances (negative examples).

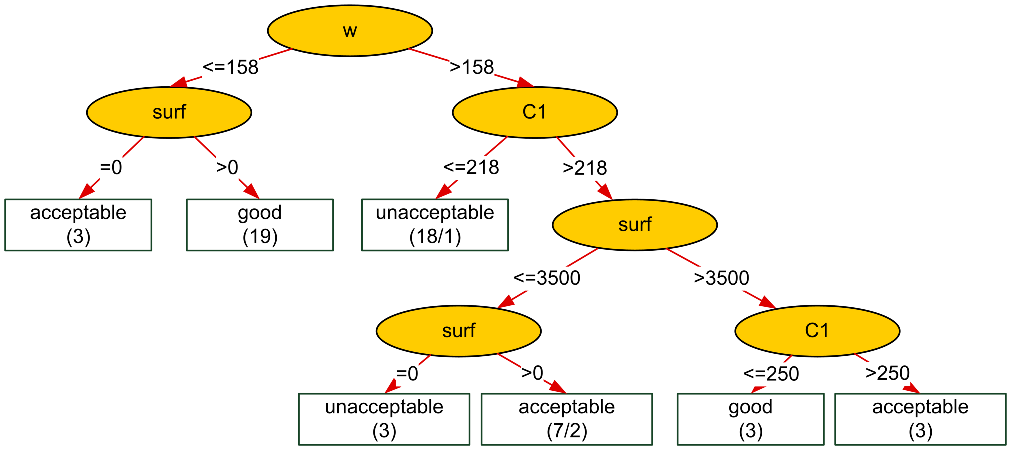


Figure 1. The decision tree for resistance to chloride penetration generated by the J48 algorithm.

The obtained decision tree can be easily transformed into the following rules:

[resistance = good]
Rule 1 [w $\leq$158] and [surf $>$0]: $p = 19, n = 0$,
Rule 2 [w $>$158] and [surf $>$3500] and [218 $<$ C1 $\leq$ 250]: $p = 3, n = 0$.

[resistance = acceptable]
Rule 1 [w $\leq$158] and [surf = 0]: $p = 3, n = 0$,
Rule 2 [w $>$158] and [C1 $>$218] and [0 < surf $\leq$ 3500]: $p = 7, n = 2$,
Rule 3 [w $>$158] and [C1 $>$250] and [surf $>$3500]: $p = 3, n = 0$.

[resistance = unacceptable]
Rule 1 [w $>$158] and [C1 $\leq$218]: $p = 18, n = 1$,
Rule 2 [w $>$158] and [C1 $>$218] and [surf $=$ 0]: $p = 3, n = 0$,

where p denotes the number of positive examples covered by the rule (*i.e.*, the number of records from this class satisfying the rule) and n denotes the number of negative examples covered by the rule (*i.e.*, the number of records from the other classes satisfying the rule).

The obtained decision rules determine the conditions concretes have to fulfill to provide appropriate resistance against chloride penetration.

The good class characterizes:

- concretes with water content below 158 kg/m^3 (w $\leq$ 158) where 15% or 30% of cement mass (C1 or C2) was replaced with high calcium fly ash (surf $>$ 0),
- concretes with water content above 158 kg/m^3 (w $>$ 158) where 30% of cement C1 mass (218 $<$ C1 $\leq$ 250) was replaced by high calcium fly ash S1 ground for 28 minutes or fly ash S3 ground for 20 minutes (surf $>$ 3500).

The acceptable class characterizes:

- concretes without high calcium fly ash (surf = 0) with water content below 158 kg/m^3,
- concretes with water content above 158 kg/m^3 (w $>$ 158) where 15% or 30% of cement C1 mass (C1 $>$ 218) was replaced by unprocessed high calcium fly ash S1, S3 or S1 ground for 10 minutes (surf $\leq$ 3500),

- concretes with water content above 158 kg/m^3 (w $>$ 158) where 15% of cement C1 mass (C1 $>$ 250) was replaced by high calcium fly ash S1 ground for 28 minutes or fly ash S3 ground for 20 minutes (surf $>$ 3500),

The unacceptable class characterizes:

- concretes with water content above 158 kg/m^3 (w $>$ 158) and with a content of cement C1 below 218 kg/m^3 (C1 $\leq$ 218), that is concretes containing cement C2 with or without high calcium fly ash, as well as concretes where 30% of cement C1 mass was replaced by unprocessed high calcium fly ash S3,
- concretes without high calcium fly ash (surf = 0) with water content above 158 kg/m^3 (w $>$ 158).

Using the leave-one-out method (n = 56), we obtained a classification accuracy equal 89.3%. The result obtained from a test set is often displayed as a two-dimensional confusion matrix with a row and a column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column. The sum of the numbers down the main diagonal divided by the total number of test examples determine the classification accuracy. The confusion matrix of the solved problem is determined in the form presented in Table 8.

Table 8. The confusion matrix for leave-one-out validation.

	good	**acceptable**	**unacceptable**
good	22	0	0
acceptable	0	9	3
unacceptable	0	3	19

Such a result can be considered satisfactory with respect to the limited number of records in the database.

5. Conclusions

The rules generated by algorithm J48 from the Weka workbench provided a means for the adequate classification of plain concretes and concretes modified with high calcium fly ash as materials of good, acceptable and unacceptable resistance to chloride penetration.

According to the generated rules, it is found that if the content of water in mixes is small enough (in investigated concretes, w $\leq$ 158 L/m^3), then concretes modified with high calcium fly ash are qualified as materials of good resistance to chloride penetration, whereas concretes without high calcium fly ash are qualified as materials of acceptable resistance. For greater content of water (w $>$ 158 L/m^3), concretes using cement of low C_3A with or without high calcium fly ash are characterized by unacceptable resistance to chloride penetration. However, when using cement of high C_3A, the replacement 15% or 30% of cement mass by high calcium fly ash, particularly by ground fly ash, improves the resistance of concretes to chloride penetration.

It is found that both the specific surface of fly ash and the content of water and cement play a significant role in providing the required concrete resistance. The classifier was evaluated using the leave-one-out method. The obtained classification accuracy was equal to 89.3%. This value seems to be sufficient to acknowledge the correctness of the classifier. Due to a small number of tested specimens, the rules are applicable only to concrete mix compositions of similar binder content. Further tests are needed in order to enlarge the experimental database and to cover a broader range of concrete compositions.

Acknowledgments: The research is a part of the research project "Innovative cement based materials and concrete with high calcium fly ashes" co-financed by the European Union from the European Regional Development Fund.

Author Contributions: Michal A. Glinicki planned and organized the experimental study, made the selection of materials and the mix design. Karolina Gibas conducted the experiments and collected data measurements. Michal Marks performed the machine learning analysis. All authors read and agreed with the final version of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Papadakis, V.G. Effect of fly ash on Portland cement systems—Part II. High-calcium fly ash. *Cem. Concr. Res.* **2000**, *30*, 1647–1654.
2. Naik, T.R.; Singh, S.S.; Hossain, M.M. Properties of high performance concrete systems incorporating large amounts of high-lime fly ash. *Constr. Build. Mater.* **1995**, *9*, 195–204.
3. Felekoglu, B.; Türkel, S.; Kalyoncu, H. Optimization of fineness to maximize the strength activity of high-calcium ground fly ash—Portland cement composites. *Constr. Build. Mater.* **2009**, *23*, 2053–2061.
4. Jóźwiak-Niedźwiedzka, D.; Brandt, A.M.; Ranachowski, Z. Self-healing of cracks in fibre reinforced mortar beams made with high calcium fly ash. *Cem. Lime Concr.* **2012**, *79*, 38–49.
5. Papadakis, V.G. Effect of supplementary cementing materials on concrete resistance against carbonation and chloride ingress. *Cem. Concr. Res.* **2000**, *30*, 291–299.
6. Gjørv, O.E. *Durability Design of Concrete Structures in Severe Environments*; Taylor & Francis: New York, NY, USA, 2009.
7. Shi, X.; Xie, N.; Fortune, K.; Gong, J. Durability of steel reinforced concrete in chloride environments: An overview. *Constr. Build. Mater.* **2012**, *30*, 125–138.
8. Andrade, C.; Buják, R. Effects of some mineral additions to Portland cement on reinforcement corrosion. *Cem. Concr. Res.* **2013**, *53*, 59–67.
9. Kaetzel, L.J.; Clifton, J.R. Expert/knowledge based systems for materials in the construction industry: State-of-the-art report. *Mater. Struct.* **1995**, *28*, 160–174.
10. Mikut, R.; Reischl, M. Data mining tools. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 431–443.
11. Topçu, I.B.; Saridemir, M. Prediction of mechanical properties of recycled aggregate concretes containing silica fume using artificial neural networks and fuzzy logic. *Comput. Mater. Sci.* **2008**, *42*, 74–82.
12. Sebastiá, M.; Olmo, I.F.; Irabien, A. Neural network prediction of unconfined compressive strength of coal fly ash-cement mixtures. *Cem. Concr. Res.* **2003**, *33*, 1137–1146.
13. Yeh, I.C. Computer-aided design for optimum concrete mixtures. *Cem. Concr. Compos.* **2007**, *29*, 193–202.
14. Marks, M.; Jóźwiak-Niedźwiedzka, D.; Glinicki, M.A. Automatic categorization of chloride migration into concrete modified with CFBC ash. *Comput. Concr.* **2012**, *9*, 375–387.
15. Marks, M.; Jóźwiak-Niedźwiedzka, D.; Glinicki, M.A.; Olek, J.; Marks, M. Assessment of scaling durability of concrete with CFBC ash by automatic classification rules. *J. Mater. Civ. Eng.* **2012**, *24*, 860–867.
16. Marks, M.; Marks, M. Prediction of scaling resistance of concrete modified with high-calcium fly ash using classification methods. *Procedia Comput. Sci.* **2015**, *51*, 394–403.
17. Alterman, D.; Kasperkiewicz, J. Evaluating concrete materials by application of automatic reasoning. *Bull. Pol. Acad. Sci. Tech. Sci.* **2006**, *54*, 352–362.
18. Kasperkiewicz, J.; Alterman, D. Holistic approach to diagnostics of engineering materials. *Comput. Assist. Mech. Eng. Sci.* **2007**, *14*, 197–207.
19. Gibas, K.; Glinicki, M.A. Influence of high-calcium fly ashes on the chloride ion penetration into concrete. In Proceedings of the International Conference Non-Traditional Cement & Concrete IV, Brno, Czech Republic, 27–30 June 2011; pp. 419–428.
20. NT Build 492: Concrete, mortar and cement-based repair materials: Chloride migration coefficient from non-steady-state migration experiments. Available online: http://210.42.35.80/G2S/eWebEditor/uploadfile/20110819235419966.pdf (accessed on 4 December 2015).
21. Tang, L. Chloride transport in concrete — Measurement and prediction. Ph.D. Thesis, Department of Building Materials, Chalmers University of Technology, Göteborg, Sweden, 1996.
22. Gibas, K.; Glinicki, M.A.; Nowowiejski, G. Evaluation of impermeability of concrete containing calcareous fly ash in respect to environmental media. *Roads Bridges Drogi i Mosty* **2013**, *12*, 159–171.

23. Baert, G.; Gruyaert, E.; Audenaert, K.; de Belie, N. Chloride ingress in high-volume fly ash concrete. In Proceedings of the First International Conference on Microstructure Related Durability of Cementitious Composites, Nanjing, China, 12–15 October 2008; Sun, W., van Breugel, K., Miao, C., Ye, G., Chen, H., Eds.; RILEM Proceedings PRO 061; Volume 61, pp. 473–482.
24. Giergiczny, Z.; Garbacik, A.; Ostrowski, M. Pozzolanic and hydraulic activity of calcareous fly ash. *Roads Bridges Drogi i Mosty* **2013**, *12*, 71–81.
25. Vollpracht, A.; Brameshuber, W. Performance-Concept, K-Value Approach - Which Concept Offers Which Advantages? In Proceedings of the International RILEM Conference on Material Science, Aachen, Germany, 6–10 September 2010; Brameshuber, W., Ed.; RILEM Proceedings PRO 077; pp. 403–411.
26. Bentur, A.; Mitchell, D. Material performance lessons. *Cem. Concr. Res.* **2008**, *38*, 259–272.
27. Glinicki, M.A.; Nowowiejski, G.; Gibas, K. Strengthening efficiency of nonstandard addition of fluidized bed ash in concrete. In Procceedings of the International RILEM Conference on Advances in Construction Materials Through Science and Engineering, Hong-Kong, China, 5–17 September 2011; Leung, C., WAN, K., Eds.; RILEM Proceedings PRO 079: Hong Kong, China, 2011; pp. 684–691.
28. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*; The Morgan Kaufmann Series in Data Management Systems; Elsevier: Burlington, MA, USA, 2005.
29. Wojtusiak, J. *AQ21 User's Guide. Reports of the Machine Learning and Inference Laboratory*; Technical Report MLI 04-3; George Mason University: Fairfax, VA, USA, 2004.
30. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Series in Machine Learning; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1993.
31. Krawiec, K.; Stefanowski, J. *Machine Learning and Neural Networks*; Poznan University of Technology: Poznań, Poland, 2003. (In Polish)