Data Mining Techniques for Prediction of Concrete Compressive Strength (CCS)

تقنيات التنقيب في البيانات للتنبؤ بالقوة الانضغاطية الخرسانية

Prof. Yousef Saleh Abu Zir
Professor/ Al-Quds Open University/ Palestine
yabuzir@qou.edu

Eng. Saleh Yousef Abu Zir
Graduate Student/ University of Brescia / Italy
s.abuzir@studenti.unibs.it

أ.د. يوسف صالح يوسف ابو زر
أستاذ دكتور/ جامعة القدس المفتوحة/ فلسطين

م. صالح يوسف أبو زر
طالب دراسات عليا/ جامعة بريشا/ إيطاليا

Received: 23/ 8/ 2019, **Accepted:** 25/ 1/ 2020
DOI: 10.33977/2106-000-003-006
http://journals.qou.edu/index.php/PJTAS

تاريخ الاستلام: 23/ 8/ 2019م، **تاريخ القبول:** 25/ 1/ 2020م.
E - ISSN: 2307 - 4655
P - ISSN: 2307 - 4647

Abstract

The main aim of this research is to use data mining techniques to explore the main factors affecting the strength of concrete mix. In this research, we are interested in finding some of the factors that influence the high performance of concrete to increase the Concrete Compressive Strength (CCS) mix. We used Waikato's Knowledge Analysis Environment (WEKA) tool and algorithms such as K-Means, Kohonen's Self Organizing Map (KSOM) and EM to identify the most influential factors that increase the strength of the concrete mix. The results of this research showed that EM is highly capable of determining the main components that affect the compressive strength of high performance concrete mix. The other two algorithms, K-Means and KSOM, were noted to be an advanced predictive model for predicting the strength of the concrete mix.

Keywords: Data Mining, Concrete Compressive Strength (CCS), K-means, EM Algorithm, Kohonen's Self-Organizing Map (KSOM) , Clustering.

ملخص:

هدف البحث الرئيس، هو استخدام تقنيات استخراج البيانات لاكتشاف العوامل الرئيسية التي تؤثر في قوة مزيج الخرسانة. إن جل اهتمامنا في هذا البحث، هو إيجاد بعض العوامل التي تؤثر في الأداء العالي للخرسانة لزيادة مزيج قوة ضاغطة الخرسانة. لتحقيق هذا الهدف، استخدمنا أداة Waikato's Environment Analysis Environment (WEKA) وخوارزميات مثل K-Means وخريطة كوهن ذاتية التنظيم (KSOM) و EM لتحديد العوامل الأكثر تأثير والّتي تزيد من قوة مزيج الخرسانة. أظهرت نتائج هذا البحث أن EM يظهر أهمية كبيرة لتحديد المكونات الرئيسية التي تؤثر في قوة الضغط للمزيج الخرساني عالي الأداء. بينما تعد الخوارزميات K-Means و KSOM نموذجًا تنبؤيًا متقدمًا لقوة الخلطة الخرسانية.

كلمات مفتاحية: تعدين البيانات، قوة الضغط الخرسانية (CCS)،K-means، EM خوارزمية ، خريطة كوهن ذاتية التنظيم (KSOM).

INTRODUCTION

Technical engineers and laboratories are required to obtain and test the strength and the accuracy of concrete. Testing modeling at the laboratory is both, time and cost consuming (Agrawal V. and Sharma A., 2010) as it includes most of the ingredients or components that are required for designing concrete. The traditional approaches focus on understanding and modeling the effects of the components on the strength of concrete (Chen L. and Wang T. S., 2010). Nowadays, the situation has changed with the rapid spread of information technology. In the recent years, different techniques of Artificial Intelligence (AI) and Data Mining were used to predict the main factors that affect the concrete strength. Recently, there has been many applications and approaches that are based on Artificial Intelligence and Data Mining in Civil Engineering (Chen L. and Wang T. S., 2010; Jain et al., 1994; Flood I., and Kartam N., 1994).

Concrete is the major building material that is used around the world. Concrete mainly consists of three basic components that are mixed in measured proportions. These components are, water, portland cement and aggregate (gravel, sand and rock). They all form a solid material called concrete. Concrete is well known for its high compressive strength, impermeability, fire resistance, durability and abrasion resistance.

There are several factors that affect the strength of High Performance Concrete (HPC). Ration of water to cement may be considered the main factor, but it is also induced by the components of the concrete like cement, blast furnaces slag, fly ash, water, super plasticizer, coarse aggregate, fine aggregate and age.

Using Data Mining will provide advice, assistance and indication of signs to enhance Concrete Compressive Strength (CCS) by finding the main factors that influence the compressive strength of concrete and its high performance.

This study focuses on identifying the list of components that affect Concrete Compressive Strength (CCS) by using data mining Algorithms to assist in predicting and identifying the main necessary components to identify high performance compressive strength of concrete.

This study is based on publicly available resources of UCI Machine Learning Repository with eight parameters and one output. We used three different data mining algorithms. These algorithms are K-Means, Kohonen's Self Organizing Map (KSOM) and EM and applied them on the dataset. According to the analysis of the data and the results, the most accurate result is achieved by EM for predicting the key components that affect the compressive strength of concrete. On the other hand, K-Means and KSOM can be used as an effective tool for predicting concrete compressive strength.

In this paper, we start with the literature review of the research papers (section 2). An overview of data mining techniques in civil engineering is then presented in section3. In Section 4, we explain in details the methodological approaches used throughout this study, followed by a discussion regarding the findings of this research (Section 5). Finally, a summary and conclusion are presented in Section 6.

Literature Review

In the recent years, Artificial Intelligence played essential roles in solving problems that are difficult to address through the traditional programming or human experts. Researcher used data mining and ANN for solving many problems in many fields such as, tourism, finance, banking, aerospace, airplane navigation, life insurance, automotive, terrorism, defense, fault detection in electric and electronics, telecommunications, entertainment, control systems in industry, automotive of manufacturing, transportation (Arciszewski, et al, 1994), agriculture (Abuzir Y., 2018), smart cities, civil engineering, medicine, image processing, robotics, speech recognition and information securities.

Data Mining Technique (DMT) applications have become more numerous and more important in many areas. By using DMT, we are able to see a transformation and obtain new knowledge or skills in many fields, as well as allow or plan for a certain possible new future applications (Shu et al, 2011).

In the literature Review, there are different approaches and studies that focus on finding the appropriate properties for designing concrete and predicting the Concrete Compressive Strength (CCS) using Artificial Intelligence techniques.

Ozcan et al., in their research proposed Artificial Neural Networks to predict long-term compressive strength of silica fume concrete (Ozcan et al, 2009). Another researcher used neural network for predicting Concrete Compressive Strength (CCS) with different water/cement ratios. In the input layer of the neural network model, they used the following five input parameters: water/binder ratio, binder/sand ratio, metakaolin percentage, superplasticizer percentage, and age. The proposed neural network model predicts the compressive strength of mortars only (Saridemir M., 2009).

Neural Network model are based on four input parameters prediction models used for predicting compressive strength of concrete. The input layer employed the following four parameters: Water-to-binder ratio, cement content, curing conditions, and age (Yaprak et al, 2011).

(Tinoco et al., 2010) used Data Mining technique as a prediction model for uniaxial compressive strength (UCS) of JG materials. They showed their model are able to identify with high accuracy the complex relationship between the UCS of JG material and its contributing factors.

Another approach is based on combining conventional method with the artificial intelligence method to design a predictive model for a concrete compressive strength. The results showed that their model is accurate and suitable for predicting the compressive strength development (Liu G. and Zheng J., 2019).

DATA MINING IN CIVIL ENGINEERING

An Overview of Data Mining and Weka

Data mining is a process or a technique of applying different algorithms on a large dataset for extracting beneficial information or knowledge.

Intelligent tools are required to apply data mining techniques to manipulate datasets.

Data mining is often used as a combination of intelligent and unconventional sciences like business analytics, mathematics, logic, statistics, artificial intelligence, machine learning and artificial neural networks (Mohammed, 2016), (Abuzir Y. and Baraka A.M, 2019).

The analytic techniques used in data mining often share or use the following Data Mining algorithms (Brown, 2012), (Patel et al., 2014):

- Classification
- Clustering
- Association
- Prediction
- Sequential patterns
- Decision trees

Data mining involves five steps: Data selection, data cleaning, data transformation, pattern evaluation and knowledge presentation and finally decisions / use of discovered knowledge as shown in the Figure 1

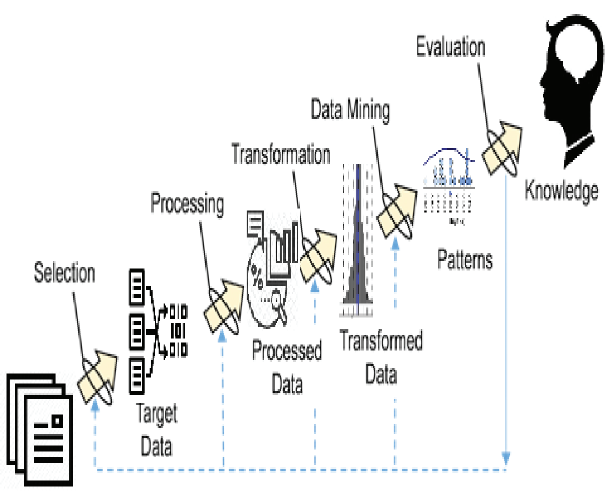


Figure 1

The main steps in Data mining

WEKA is abbreviation for Waikato's Knowledge Analysis Environment. It is an open source tool developed by the University of Waikato in New Zealand. WEKA is a Java based tool that involves many open source data mining and machine learning algorithms. WEKA has the following features (Alka, et al. 2017):

- Data processing tools.
- Classification, clustering algorithms and association.
- User interface and graphical interface.
- WEKA data mining and machine learning tools

The Use of Data Mining in Civil Engineering

Nowadays, a lot of data and information related to civil engineering field are available on online repositories of the research centers. Researchers can use this information and apply different data analysis to obtain important information to support their research papers. They can use data mining techniques in many areas of Civil Engineering.

In the field of civil engineering, many research papers apply different approaches of Data Mining and Artificial Neural Network (ANN) technologies (Deepa et al. 2010; Guneyisi, et al., 2009; Topcu I.B, Sarıdemir M., 2007). Different studies applied data mining techniques and ANN the following areas of civil engineering (Kaplinski, et al., 2016; Topcu, et al., 2009):

- Predicting properties of conventional concrete (Guneyisi, et al., 2009).
- Predicting high performance compressive strength of concretes (Ozcan et al, 2009; (Nikoo, et al., 2015), (Han, et al., 2019; Young et al., 2018).
- Concrete mix proportions (Topcu I.B, Sarıdemir M. , 2007) [15] (Young et al., 2018).
- Predict the concrete durability (Yaprak, et al., 2009) [5] (Pann et al, 2003).
- Modeling of material behavior (Bock et al, 2019),
- Detection of structural damage (Fanga, et al., 2005),
- Structural system identification (Chou, et al., 2014),
- Structural optimization (Tanyildizi, H. 2009),
- Structural control, ground water monitoring (El-Kholy, A. M. 2019),
- Prediction of settlement of shallow foundation (Pann, et al. 2003)

MATERIALS AND METHODS

Most of the previous studies attempted to investigate, study and model the effects of the components on the strength of the concrete. In the recent years, new approaches utilized Artificial Intelligence (AI) and Data Mining techniques to predict the main factors that affect the concrete strength.

The main contributions of our approach is twofold. First, it focuses on using all the different components that compose the concrete, to study the main factors that influence the high performance of the concrete, to increase the Concrete Compressive Strength (CCS) mix. Second, we try to find a better and more accurate prediction model for CCS. We can summarize our contributions in the following points:

- The study uses three different algorithms K-Means, Kohonen’s Self Organizing Map (KSOM) and EM.
- The study determines which is the best algorithm that can be used to identify the main factors that influence the strength of concrete.

The study identifies the best algorithm that can be used as an advanced prediction model for the strength of concrete mix.

This research utilized data mining techniques to predict the key components that affect the strength of concrete. WEKA tool provides us with different tools to analyze the dataset and apply different algorithms such as EM, Kohonen’s Self Organizing Map (KSOM) and K-Means. The following paragraphs and subsections discuss the characteristics of the datasets and algorithms used in this study. It discusses in details the methodological approach used to develop the prediction model of the main key factors that affect the compressive strength of concrete.

Data Sets

Compressive strength concrete dataset from UCI Machine learning Repository (Yeh I. C., 1998) is used as the experimental data sets of 1030 cases. In the data set, there are eight input parameters and one output value Concrete Compressive Strength (CCS). These parameters are cement, blast furnaces slag, fly ash, water, super plasticizer, coarse aggregate, fine aggregate and age. For the first seven parameters, we use kg/m3 and for the eighth parameter age, we use number of days for the laboratory test of the concrete sample.

We obtained the statistical analysis using Weka to create Table 1 and represent it using graphs in Figure 2. Table 1 lists a general statistical information on the eight factors. These statistics are computed by WEKA. The table shows the maximum, minimum , the average, the mean and the Standard deviation for each factors. Weak supports users though two methods to split data.

The first method is training and supplied test set. The second method is a percentage split and these groups are not included with each other during the training phase. To conduct the statistical analysis of the datasets, we divided the dataset into two groups: A training set (721 samples) amounting to70%, and a testing set (309 samples) amounting to 30% of the group. After splitting the data into training and testing sets, the statistical analysis and data mining algorithms were applied to present the results.

Table 1

Concrete Strength Data Sets Components Ranges (WEKA)

Name of Component	Maximum (kg/m3 mixture)	Minimum (kg/m3 mixture)	Average Value (kg/m3 mixture)	Mean	SDV
Cement	540	102	321	281.16	104.50
Blast Furnace	359.40	0	179.7	73.896	86.279
Fly Ash	200.10	0	100.05	54.188	63.997
Water	247	121.75	184.375	181.56	21.354
Superplasticizer	32.20	0	16.1	6.205	5.974
Coarse Aggregate	1145	801	973	972.91	77.754
Fine Aggregate	992.60	594	793.3	773.58	80.176
Age of testing	365 days	1 day	183 days	45.662	63.17

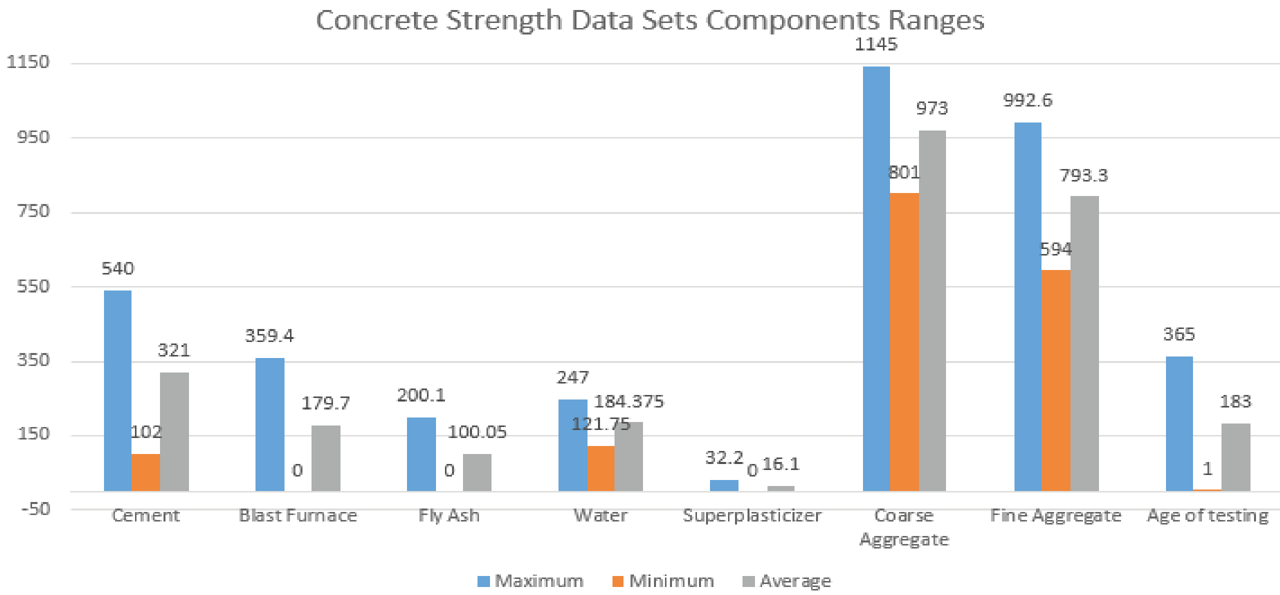


Figure 2.

Concrete Strength Data Sets Components Ranges

DATA MINING ALGORITHMS

This section presents the different machine learning algorithms used in finding the main factors that affect the compressive strength of the concrete. EM is one of the clustering algorithms used in data mining. It uses two iterative steps called E-step and M-step:

- E-step, where each object assigned to the most likely cluster(centroids).
- M-step, where the model (centroids) are recomputed (Least Squares Optimization).

Another algorithm is Kohonen Self-Organizing Map (KSOM). It is one of the most adopted neural network in unsupervised learning. (Fernando, 2015).

K-means algorithm is a clustering algorithm, given the data <x1, x2,…,xn> and K, assign each xi to one K clusters, C1…Ck, minimizing equation 1 (Khedr et. al, 2014), equation (1) used to find Sum of Squared Error (SSE)

$$SSE = \sum_{j=1}^{K} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2 . \qquad (1)$$

Where

K is the number of desired clusters

is mean over all points in cluster Cj.

The following Algorithm is used to apply K-Means:

1. Set randomly

2. Repeat until convergence:

- Assign each point xi to the cluster with closest mean
- Calculate the new mean for each cluster (equation 2)

$$\mu j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Figure 3 presents a schematic illustration of prediction mechanisms using the three machine-learning algorithms of simple K-Means, KSOM and EM.

We utilized the EM, KSOM and K-means algorithms for finding the main components in concrete mix that affect the compressive strength of concrete. The study applied these algorithms with different configurations of both, the algorithms and the dataset. Then the study analyzed the results along with an evaluation of the different configurations of the results. The simplest approach is to find the parameter that minimizes scores of the different parameters like Standard Deviation (STD) and Root Mean Squared Error (RMSE).

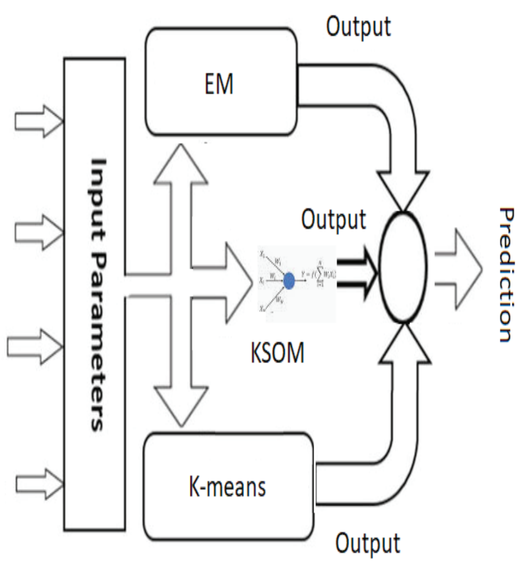


Figure 3

Schematic illustration of prediction using EM, KSOM and K-Means Algorithms.

RESULTS AND DISCUSSION

In this research, the dataset is first selected, then data mining techniques are utilized in finding the parameters. In general, eight parameters (cement, blast furnaces slag, fly ash, water, super plasticizer, coarse aggregate, fine aggregate and age) were examined against concrete compressive strength using three data mining algorithms. This section discusses, compares and evaluates these algorithms using concrete dataset to investigate the main factors that affect concrete mix strength.

Table 2 represents the primary results of EM algorithm. To get the result, different datasets were used with different numbers of clusters (K=3,5,7,and 9) as shown in Table 2.

For each number of clusters, we computed different statistical values. In our case, we used standard deviation as a statistical measure to select the main factors that affect the CCS. Table 2 summarizes our calculations and shows only the most influential factors on the CCS. For example, when K=5, we find that the following three factors blast furnaces slag, fly ash, and super plasticizer imapct the CCS.

The values show the different results of predicting the main factors that affect Concrete Compressive Strength (CCS) using EM algorithm based on eight components of concrete mix. These results are computed and visualized using WEKA Tool.

Figure 4 shows the relationship between the main components that affect the concrete mix and the parameter Concrete Compressive Strength (CCS) using EM Algorithm. As shown in these figures, the values of Concrete Compressive Strength (CCS) as a function computed based on Superplasticizer, Fly Ash and Blast Furnace Slag serve obtained high similarity values.

Table 2.

Screen dumps of the results of EM Algorithms Using WEKA

Number of Clusters	Results of EM
EM (with K= 3)	Superplasticizer mean std. dev.
EM (with K= 5)	Blast Furnace Slag mean 71.726 std. dev. 69.762 Fly Ash mean 8.6447 std. dev. 30.1856 Superplasticizer mean 0.34 std. dev. 1.18

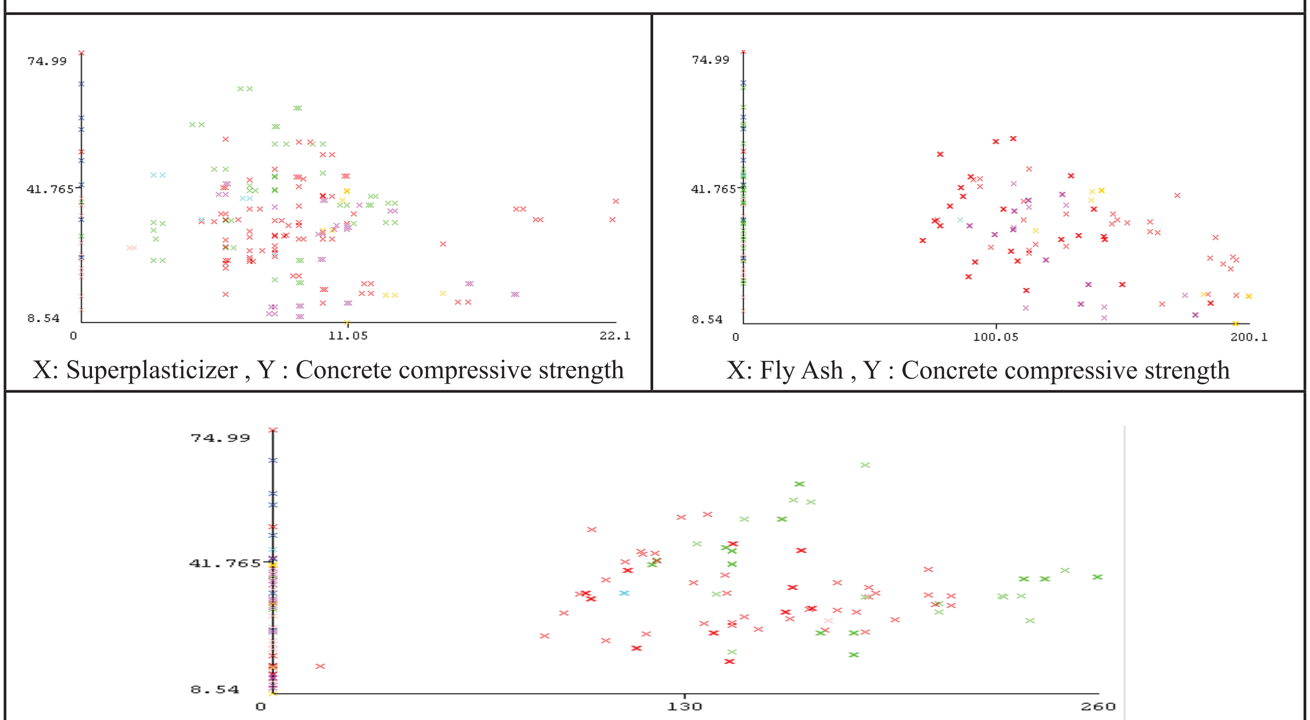
Number of Clusters	Results of EM										
EM (with K= 7)	Blast Furnace Slag										
	mean	23.1165	140.6642	82.718	0	26.5337	192.7328	22.9621			
	std. dev.	23.3759	65.9588	73.0192	0.0002	51.3771	61.611	40.8663			
	Fly Ash										
	mean	110.9131	0.0035	0.0107	0	1.5761	0.4274	120.6712			
	std. dev.	26.8864	0.2928	1.0604	61.4471	8.8978	3.2074	32.6375			
	Superplasticizer										
	mean	10.2752	14.6176	0.0006	0	3.0277	0.5002	7.9891			
	std. dev.	3.8984	6.4238	0.0616	0.0004	5.3601	1.9652	2.9272			
EM (with K= 9)	Blast Furnace Slag										
	mean	4.0644	90.9507	82.6599	8.8915	175.5803	0.4714	21.0026	33.7256	182.1985	
	std. dev.	13.4621	48.8117	72.6721	29.1616	78.743	2.3478	5.7829	44.2412	48.6432	
	Fly Ash										
	mean	110.9131	0.0035	0.0107	0	1.5761	0.4274	120.6712			
	std. dev.	26.8864	0.2928	1.0604	61.4471	8.8978	3.2074	32.6375			
	Superplasticizer										
	mean	0.9103	17.5907	0.0002	0.5885	0.0253	7.479	10.4851	8.4821	11.8749	
	std. dev.	2.3356	7.5873	0.036	1.8861	0.3856	3.3652	3.8487	2.4399	3.313	

After designing EM model for predicting the main factors that affect concrete compressive strength and analyzing the results obtained by the EM algorithm, it is clear that the EM algorithm achieves the optimal mix of the concrete components.

After running the EM algorithm on the dataset for a number of times with varied values for number of clusters, the best parameters were selected based on their Standard deviation values. Table 3 shows the list of the main factors that affect Concrete Compressive Strength (CCS) with their standard deviations.

Figure. 4 illustrates the values of concrete compressive strength predicted by the EM algorithm versus the other components such as Superplasticizer, Fly Ash and Blast Furnace Slag, for both training and testing datasets. As shown in figure 4, there is a consistent indication among the different combinations of the three components and the concrete compressive strength. It is clear that the distribution of points in the three planes shows the same picture.

EM with 9 Clusters



X: Superplasticizer , Y : Concrete compressive strength

X: Fly Ash , Y : Concrete compressive strength

X: Blast Furnace Slag , Y : Concrete compressive strength

Figure 4.

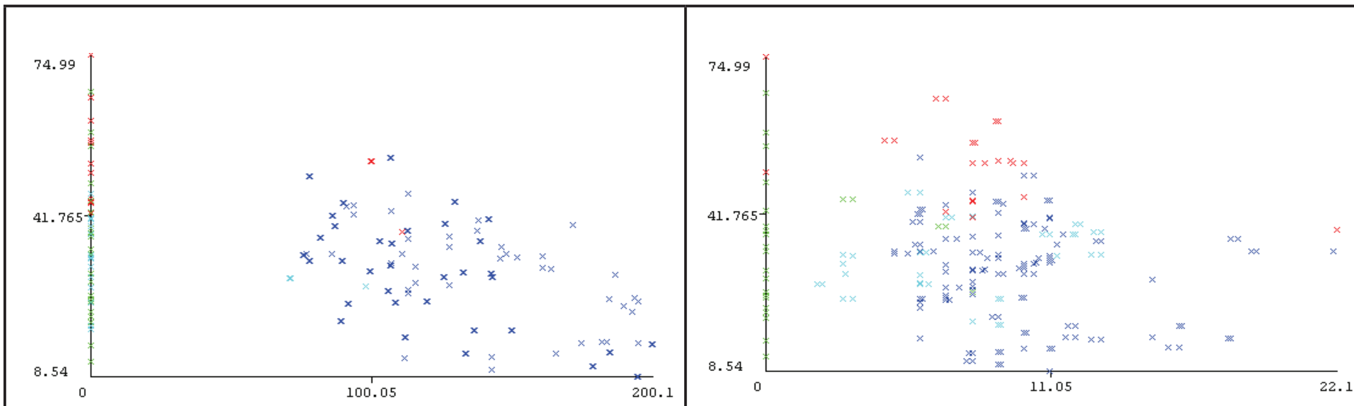
Plotting of the main components the affect the concrete using EM Algorithm

Table 3.

List of the main components that affect compressive strength of concrete (EM)

Number of Clusters	Standard. Deviation	Predict Components
3	0.0001	Superplasticizer
	0.015	Blast Furnace Slag
5	3.265	Fly Ash
	0.2488	Superplasticizer
	0.0004	Superplasticizer
7	0.2928	Fly Ash
	0.0002	Blast Furnace Slag
	0.0001	Superplasticizer
9	0.0148	Fly Ash
	2.3478	Blast Furnace Slag

The second model use the KSOM algorithm. This algorithm is employed to illustrate the components that affect concrete compressive strength. In the KSOM algorithm, the main components that affect the Concrete Compressive Strength (CCS) are Fly Ash and Superplasticizer. Figure 5 shows the results.



X: Fly Ash , Y : Concrete Compressive Strength

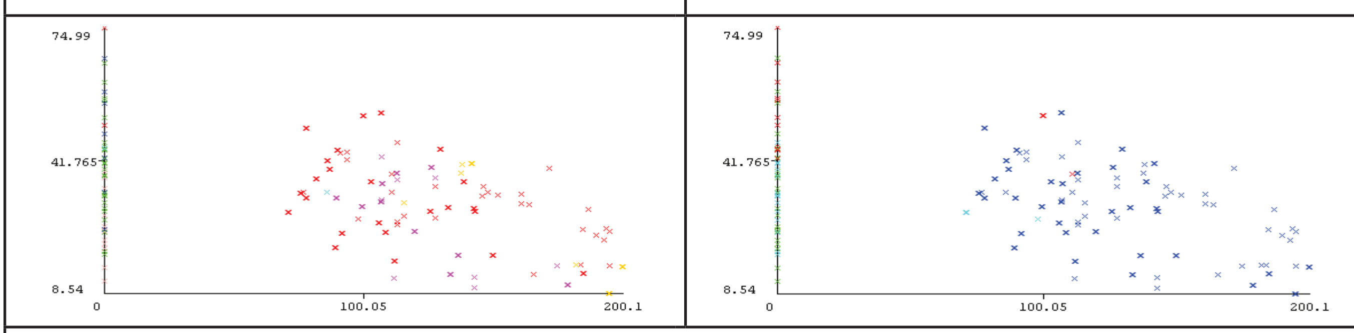
X: Superplasticizer, Y : Concrete Compressive Strength

Figure 5.

Fly Ash and Superplasticizer versus Concrete Compressive Strength (KSOM)

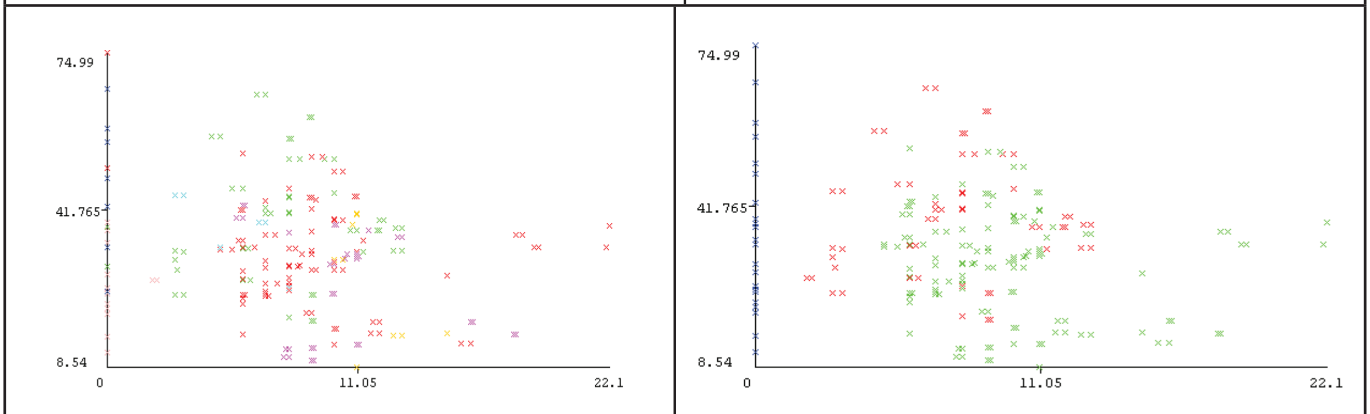
Figure 6 illustrates a comparison between the EM and KSOM algorithms. As the figure shows, the predicted models for the two components are highly similar. The performance of fly ash on concrete compressive strength has the same significant effect. The analysis of the two graphs shows that the two algorithms have the same effect among the potentially used two input parameters, fly ash and Superplasticizer.

EM | **KSOM**



X: Fly Ash , Y : Concrete Compressive Strength

EM	KSOM



X: Superplasticizer, Y : Concrete Compressive Strength

Figure 6.

Comparing EM and KSOM Algorithms

The K-means algorithm is applied to the datasets, using different value for k = 3,5,7 and 9. Table 4 shows the results of clustering with the different value for K= 3,5, 7 and 9.

Based on the analysis of the result of K-Means we find that the factors that mostly affect the compressive strength on concrete mix are Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate (Table 5). According to the results, Table 6 presents a summary of the key attributes that affect the concrete compressive strength using the three different algorithms.

Referring to the results in table 6, K-Means algorithm shows that Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate are the most common components that affect the Concrete Compressive Strength (CCS) mix. In EM and KSOM algorithms two common component are considered, Fly Ash and Superplasticizer. At the same time, EM algorithms includes a distinguished component which is the Blast Furnace Slag. It is clear that, all the three algorithms show intersection and provide different information. In general, the analysis concludes that Fly Ash and Superplasticizer are common components and they are the two main factors that affect concrete compressive strength.

Table 4.

Screen dumps of the Results For K-Means (with K= 3,5, 7 and 9)

Number of Clusters	Results For K-Means (with K= 3,5, 7 and 9)
K-Means (with K= 3)	(screen dump below)

```
Final cluster centroids:
                                            Cluster#
Attribute                      Full Data         0          1          2
                                 (806.0)   (360.0)    (295.0)    (151.0)
=========================================================================
Cement                          292.8646  284.8378   246.1525   403.2603
Blast Furnace Slag               67.3143   83.2689    18.9129   123.8358
Fly Ash                          47.4553    0.9153   124.3339     8.2185
Water                           179.8442  197.2456   169.0492    159.447
Superplasticizer                  5.6511    0.2769     8.3688    13.1543
Coarse Aggregate                 985.786  992.7369  1001.0068   939.4781
Fine Aggregate                  778.3337  755.3672   805.9288   779.1775
Age                              49.5546   64.5111         37    38.4238
Concrete compressive strength    36.5954   28.7541    35.4996    57.4306
```

Number of Clusters	Results For K-Means (with K= 3,5, 7 and 9)
K-Means (with K= 5)	see below

K-Means (with K= 5)

Attribute	Full Data (806.0)	Cluster# 0 (149.0)	1 (297.0)	2 (88.0)	3 (57.0)	4 (215.0)
Cement	292.8646	200.2141	247.4983	373.85	428.8509	350.5428
Blast Furnace Slag	67.3143	187.5711	18.9471	148.1057	95.4404	10.2628
Fly Ash	47.4553	1.3154	124.0286	13.142	0	0.2791
Water	179.8442	196.4242	169.0017	164.6159	150.293	197.3991
Superplasticizer	5.6511	0.594	8.3906	11.0352	17.4947	0.0279
Coarse Aggregate	985.786	974.1993	1000.7778	978.5557	868.3368	1007.2033
Fine Aggregate	778.3337	751.1007	805.2963	729.2545	875.9649	754.1656
Age	49.5546	45.3356	36.798	38.1705	35.0877	78.5953
Concrete compressive strength	36.5954	26.3142	35.5209	59.998	53.5979	31.1184

K-Means (with K= 7)

Final cluster centroids:

Attribute	Full Data (806.0)	Cluster# 0 (128.0)	1 (244.0)	2 (79.0)	3 (60.0)	4 (66.0)	5 (164.0)	6 (65.0)
Cement	292.8646	194.9336	215.3877	368.1253	431.1583	357.6136	329.2226	399.9446
Blast Furnace Slag	67.3143	192.768	18.7746	157.7367	96.6083	66.4136	4.4921	24.9585
Fly Ash	47.4553	0	125.3332	3.2278	0	0	1.0457	111.4031
Water	179.8442	195.3016	168.8914	165.0646	151.8333	217.4318	190.7787	168.5846
Superplasticizer	5.6511	0.2008	7.982	10.6076	17.065	0	0.3262	10.2477
Coarse Aggregate	985.786	970.3328	1020.9295	995.5886	867.525	958.9697	1022.5116	916.1108
Fine Aggregate	778.3337	758.6336	811.9779	722.6987	871.2417	660.6455	789.5463	763.8985
Age	49.5546	30.1406	39.5164	38.2278	36.25	213.5	33.1768	26.3692
Concrete compressive strength	36.5954	24.2441	33.5362	59.637	54.4978	45.3738	25.4502	47.0786

K-Means (with K= 9)

Final cluster centroids:

Attribute	Full Data (806.0)	Cluster# 0 (128.0)	1 (60.0)	2 (67.0)	3 (60.0)	4 (42.0)	5 (153.0)	6 (116.0)	7 (48.0)	8 (132.0)
Cement	292.8646	194.7477	405.5333	349.6239	431.1583	317.5524	312.0131	237.1379	473.0354	198.5303
Blast Furnace Slag	67.3143	191.7852	26.5383	188.6418	96.6083	83.2333	6.4098	2.7414	12.3229	32.5227
Fly Ash	47.4553	0	111.8	1.8284	0	0	1.6013	107.2853	2.4792	140.9803
Water	179.8442	195.8703	167.7	165.3836	151.8333	220.5714	190.4824	176.65	194	162.2659
Superplasticizer	5.6511	0.0461	10.6217	11.7821	17.065	0	0.4725	7.3328	1.0333	8.5288
Coarse Aggregate	985.786	970.4109	915.3833	979.5582	867.525	943.9619	1018.7843	985.4974	1031.3458	1048.3583
Fine Aggregate	778.3337	757.1438	761.7167	727.4851	871.2417	675.1048	802.7582	846.4509	647.5875	782.2326
Age	49.5546	31.0156	25.3333	37.4776	36.25	278.0952	36.7908	37.9397	55.1667	40.9621
Concrete compressive strength	36.5954	24.5193	47.63	58.2196	54.4978	44.5945	24.5222	30.9368	49.6467	35.8519

Table 5.

K-Means - factors that mostly affect the concrete compressive strength.

Number of Cluster	Components
3	Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate
5	Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate
7	Fly Ash, Coarse Aggregate and Fine Aggregate,
9	Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate

Table 6.

Summary of the main components that affect concrete mix using the three algorithms.

K-Means	EM	KSOM
Fly Ash, Superplasticizer, Coarse Aggregate and Fine Aggregate	Fly Ash, Superplasticizer and Blast Furnace Slag	Fly Ash and Superplasticizer

Table 7.

The relation between no. of Cluster, Sum of Squared Errors and Concrete Compressive Strength (CCS) using K-Means.

No. of Clusters	Sum Of Squared Errors (SSE)	Number Of Iterations	Concrete Compressive Strength (CCS) (Average Actual Data is 35.818)
3	286.5	18	56.2506
4	244.2	11	56.9138
5	219.1	18	56.448
6	205.8	10	56.8346

No. of Clusters	Sum Of Squared Errors (SSE)	Number Of Iterations	Concrete Compressive Strength (CCS) (Average Actual Data is 35.818)
7	182.6	30	57.042
8	176.1	12	57.2971
9	159.198	17	56.9463
12	138.4	22	54.764
15	122.486	26	53.4718
20	104.0	24	55.7447
25	94.18	16	56.8314
30	83.77	17	63.3709
50	62.3	16	67.23

Table 8 and figure 5 show the prediction of the Concrete Compressive Strength (CCS) by applying both K-Means and KSOM using WEKA. We find the actual average of Concrete Compressive Strength (CCS) is equal to 35.818. By comparing the results of the Concrete Compressive Strength CCS of both algorithms, we find a slight intersection or similarity between K-Means and KSOM algorithm.

Table 8.

Concrete Compressive Strength (CCS) Prediction K-Means vs. KSOM

No. of Clusters	Concrete Compressive Strength (CCS) Prediction (Average Actual Data is 35.818)	
	K-Means	KSOM
2	36.9804	34.8796
3	56.2506	55.2417
4	56.9138	56.9722
5	56.448	56.179
6	56.8346	56.88
8	57.2971	58.616
10	58.7342	58.9955
CCS Prediction Average	54.2084	53.9663

The values obtained using K-Means and KSOM in WEKA, indicate that the estimation results of CCS predication for both algorithms are very close. The results show that the K-Means can be successfully used to give a more accurate prediction for increasing the Concrete Compressive Strength (CCS) (54.2084) than the average actual data (35.818) and KSOM.

This study applied three algorithms and compared their results to find the main components that affect the Concrete Compressive Strength (CCS) using the WEKA tool. It was noted that the results of the EM algorithm is one of the most accurate and effective tools for finding the factors affecting the Concrete Compressive Strength. On the other hand, K-Means and KSOM algorithms are the most adequate algorithms for improving Concrete Compressive Strength mix.

Results of this study can be used to predict the main factors that affect the compressive strength of concrete and the mixtures of concrete.

Table 6 shows the main predicted components that affect the concrete compressive strength. These components are Blast Furnace Slag, Fly Ash, Superplasticizer, Coarse Aggregate and Fine

Aggregate. The analysis of the data in Table 8 and Table 9 show a significant correlation between the prediction of improving the CCS and the main factors that affect the CCS. The values for these parameters are similar among the three Data Mining algorithms. These results are very important because they provide us with the threshold values that improve the CCS. These parameters increased the performance of CSS from 35.818 to 58.9955. They were also able to increase the performance model from 36% to 59% of CCS.

Table 9.

Summary of the main components that improve the performance of concrete compressive strength.

Predictive parameters	K-Means	EM	KSOM	Average Value (kg/m3 mixture)	Mean
Blast Furnace Salg	-	0.4714		179.7	73.896
Fly Ash	0.0	0.0	0.0	100.05	54.188
Superplasticizer	0.0	0.0253	0.0355	16.1	6.205
Coarse Aggregate	867.525			973	972.91
Fine Aggregate	727.4851			793.3	773.58

Furthermore, these results reinforce the predication model through improving the CCS and reducing the cost of the concrete mixtures. For example, the cost of fly ash is varying and expressive. In our model, it is important to note that the cost of fly ash is beyond concrete mixture because the three Data Mining algorithms suggest a threshold value of zero for fly ash.

Overall, applying the different algorithms of Data Mining to our datasets proved to be very effective in predicting and improving the concrete compressive strength.

While all input parameters are very important and effective in predicting concrete compressive strength based on the laboratory test, our analysis shows that there are more effective parameters in our input that improve the performance of concrete compressive strength.

Our analysis shows that the performance of each Data Mining algorithm is similar yet with a small difference between them. Moreover, each one of them is appropriate for the prediction for improving CCS.

CONCLUSION

The main aim of this present study is to find the key components that affect Concrete Compressive Strength (CCS). To accomplish this research, the datasets were selected, then three data mining algorithms (EM, KSOM and K-Means) were applied. The actual input parameters consists of eight parameters and one output CCS. The input parameters were examined against CCS using the three data mining algorithms. The results were analyzed and discussed. The study used WEKA as a tool for data mining techniques.

This study focuses on including all the different components of the concrete in our prediction model and in finding the main factors that influence the high performance of concrete to increase the Concrete Compressive Strength (CCS) mix, using three different algorithms.

Results showed that using data mining techniques is highly effective in predicting the main factors that affect CCS. The analysis shows that K-Means and KSOM algorithms are the most accurate algorithm to predict the CCS. At the same time, EM is useful for predicting the main factors that affect the CCS.

In general, data mining techniques are very effective tools in predicting concrete compressive strength as well as the main factors that affect and improve the performance of concrete compressive strength. Our study can be expanded to include additional parameters, such as humidity, moisture, temperature, and methods of mixing etc. These parameters might be able to improve the prediction of CCS.

REFERENCES

1. (Abuzir Y. and Baraka A.M, 2019) Y.Abuzir and A.M. Baraka, Financial Stock Market Forecast Using Data Mining in Palestine, accpeted in Palestinian Journal of Technology and Applied Sciences, pp , No 2 (2019).

2. (Abuzir Y., 2018) Abuzir Y., Predict the Main Factors that Affect the Vegetable Production in Palestine Using WEKA Data Mining Tool, Palestinian Journal of Technology and Applied Sciences, pp 58-71, No 1 (2018).

3. (Agrawal V. and Sharma A., 2010) V. Agrawal and A. Sharma,World Academy of Science, Engineering and Technology International Journal of Civil and Environmental Engineering, Vol:4, No:9, 2010

4. (Alka et al., 2008) Alka A., Malhotra P. K., Sudeep M., Anshu B., and Shashi D., Data Mining Techniques and Tools for Knowledge Discovery in Agricultural Datasets, Indian Agricultural Statistics Research Institute, (ICAR), E-book, online: http://www.iasri.res.in/ebook/win_school_aa/

5. (Arciszewski, Et Al, 1994) Arciszewski, T., Khasnabis, S., Hoda, S. K., &Ziarko, W. (1994). Machine Learning In Transportation Engineering: A Feasibility Study. Applied Artificial Intelligence, 8(1), 109–124. Doi:10.1080/08839519408945434

6. (Bacao, et al., 2015) Fernando Bacao, et al., "Self-organizing Maps as Substitutes for KMeans Clustering," Springer Computational Science – ICCS 2015 Lecture Notes in Computer Science, vol 3516,pp 476- 483,2015.

7. (Bock et al, 2019), Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., &Klusemann, B. (2019). A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics. Frontiers in Materials, 6. doi:10.3389/fmats.2019.00110.

8. (Brown M., 2012) Brown M., Data mining techniques, IBM DeveloperWork, December 2012, online https://www.ibm.com/developerworks/library/ba-data-mining-techniques/

9. (Chen L. and Wang T. S., 2010) Li Chen & Tai-Sheng Wang, Modeling slump of concrete using the group method data handling algorithm,Indian Journal of Engineering & Materials Sciences Vol. 17, June 2010, pp. 179-185

10. (Chou, et al., 2014) Chou, J.-S., Tsai, C.-F., Pham, A.-D., & Lu, Y.-H. (2014). Machine learning in concrete strength simulations: Multination data analytics. Construction and Building Materials, 73, 771–780. doi:10.1016/j.conbuildmat.2014.09.054

11. (Deepa et al. 2010) C. Deepa, K. SathiyaKumari and V. PreamSudha, Prediction of the Compressive Strength of High Performance Concrete Mix using Tree Based Modeling,International Journal of Computer Applications (0975 – 8887) Volume 6– No.5, September 2010.

12. (El-Kholy, A. M. 2019) El-Kholy, A. M. (2019). Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects. International Journal of Construction Management, 1–19. doi:10.1080/15623599.2019.1580001.

13. (Fanga, et al., 2005) X.Fanga, H.Luob and J.Tanga, Structural damage detection using neural network with learning rate improvement, Computers & Structures, Volume 83, Issues 25–26, September 2005, Pages 2150-2161.

14. (Flood I., and Kartam N., 1994) I. Flood, and N. Kartam, "Neural Networks in Civil Engineering I:Principles and Understanding," J. Comp. in Civil Eng., vol. 8, no. 2,pp. 149-162, 1994.

15. (Guneyisi, et al., 2009) E. Guneyisi, M. Gesoglu, T. Ozturan, E. Ozbay, Estimation of chloride permeability of concretes by empirical modeling: considering effect of cement type, curing condition and age, Constr. Build. Mater. 23 (2009) 469–481.

16. (Han, et al., 2019) Han, Q., Gui, C., Xu, J., &Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. Construction and Building Materials, 226, 734–742. doi:10.1016/j.conbuildmat.2019.07.315

17. (Hetal P. and Dharmendra P., 2014) Patel Hetal P and Patel Dharmendra, A Brief survey of Data Mining Techniques Applied to Agricultural Data, International Journal of Computer Applications (0975 – 8887), Volume 95– No. 9, June 2014 .

18. (Jain et al., 1994) A. Jain, S. K. Jha,, and S. Misra, "Modeling the compressive strength of concrete using Artificial Neural Networks," Indian Concr. J., pp. 17-22, Oct. 2006.

19. (Kaplinski, et al., 2016) Kaplinski, O., Košeleva, N., &Ropaite, G. (2016). Big Data In Civil Engineering: A State-Of-The-Art Survey. Engineering Structures and Technologies, 8(4), 165–175. doi:10.3846/2029882x.2016.1257373

20. (Khedr et al., 2015) Khedr A., El Seddawy A. , and Idrees A. Performance Tuning of K-Mean Clustering Algorithm a Step towards Efficient DSS, International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Vol. 2, Issue 6, November 2014, pp. 111-118.

21. (Liu G. and Zheng J. , 2019) G. Liu, J. Zheng, Prediction Model of Compressive Strength Development in Concrete Containing Four Kinds of Gelled Materials with the Artificial Intelligence Method, Applied Sciences Journal, Appl. Sci. 2019, 9, 1039; doi:10.3390/app9061039

22. (M. Saridemir, 2009) M. Saridemir, Prediction of the compressive strength of mortars containing metakaolin by artificial neural networks and fuzzy logic, Adv. Eng. Softw. 40 (2009) 920–927.

23. (Nikoo, et al., 2015) Mehdi Nikoo, FarshidTorabianMoghadam, and Lukasz Sadowski, Prediction of Concrete Compressive Strength by Evolutionary Artificial Neural Networks, Advances in Materials Science and Engineering, Volume 2015, Article ID 849126, 8 pages, http://dx.doi.org/10.1155/2015/849126

24. (Ozcan et al., 2009) F. Ozcan, C.D. Atis, O. Karahan, E. Uncuoglu, H. Tanyildizi, Comparison of artificial neural network and fuzzy logic models for prediction of long term compressive strength of silica fume concrete, Adv. Eng. Softw. 40 (2009) 856–863.

25. (Pann et al, 2003). K.S. Pann, T. Yen, C.W. Tang, T.D. Lin, New strength model based on water cement ratio and capillary porosity, ACI Mater. Journal. 100 (2003) 311–317.

26. (Tanyildizi, H. 2009) Tanyildizi, H. (2009). Fuzzy logic model for the prediction of bond strength of high-strength lightweight concrete. Advances in Engineering Software, 40(3), 161–169. doi:10.1016/j.advengsoft.2007.05.013

27. (Tinoco et al., 2010) J. Tinoco , A. G. Correia , P. Cortez, Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time, Construction and Building Materials Journal, (2010), doi:10.1016/j.conbuildmat.2010.09.027 .

28. (Topcu and. Sarıdemir, 2007) I.B. Topcu, M. Sarıdemir, Prediction of properties of waste AAC aggregate concrete using artificial neural network, Comput. Mater. Sci. 41 (2007) 117–125.

29. (Topcu, et al., 2009) I.B. Topcu, A.R. Boga, F.O. Hocaoglu, Modeling corrosion currents of reinforced concrete using ANN, Automat. Const. 18 (2009) 145–152.

30. (Yaprak, et al., 2011) H. Yaprak, A.I. Karacı, I.I. Demir, Prediction of the effect of varying cure conditions and w/c ratio on the compressive strength of concrete using artificial neural networks, Neural Comput. Appl. (2011).

31. (Yeh I. C., 1998) I. Cheng Yeh, Modeling of strength of high performance concrete using artificial neural networks, Cement and Concrete Research 28(12) (1998), 1797–1808. Accessed on 31-08-2016 at http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

32. (Young et al., 2018) Young, B. A., Hall, A., Pilon, L., Gupta, P., &Sant, G. (2018). Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. Cement and Concrete Research. doi:10.1016/j.cemconres.2018.09.006,

33. (Zaki et al., 2016) Mohammed J. Zaki, Wagner Meira, Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, May 2016. ISBN: 9780521766333.

- **Engineer Saleh Abuzir** received the B.S. degree in Civil Engineering from Birziet University, West Bank, in 2018. He is a Master student in Civil and Environmental Engineering at University of Brescia, Brescia, Italy.
- **Prof. Yousef. Abuzir** holds a Ph.D. in Computer Engineering and Applied Sciences from Ghent University in Belgium (2002). He is a researcher, and full professor at Al-Quds Open University (QOU) in Palestine. He served as a coordinator for different Tempus and Erasmus+ projects (2005-current) and ICM Erasmus+ Coordinator at QOU. Abuzir’s research papers have resulted in four books and more than 70 journal articles, conference and book contributions in the fields of Artificial Intelligence, Information Retrieval, Knowledge Discovery, e-Learning, Natural Language Processing, Data Mining and Cloud Computing.