

1 Prediction of self-healing in engineered cementitious composite:  
2 Machine learning comparative analysis

3 Guangwei Chen<sup>1,2</sup>, Waiching Tang<sup>1,\*</sup>, Shuo Chen<sup>1</sup>, Shangyong Wang<sup>1</sup>, and Hongzhi Cui<sup>3</sup>

4 <sup>1</sup>The University of Newcastle, Callaghan, 2308, NSW,Australia

5 <sup>2</sup>Qiannan Normal College of Nationalities, Guizhou, China

6 <sup>3</sup>College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China

7 **Abstract**

8 Engineered cementitious composite (ECC) is a unique material which can significantly contribute  
9 to self-healing behaviour based on ongoing hydration. However, it's difficult to model and predict the  
10 self-healing of ECC. Although different machine learning (ML) algorithms have been utilized to predict  
11 several properties of concrete, the application of ML on self-healing prediction is considerably rare. This  
12 paper aims to provide a comparative analysis on the performance of various machine learning models  
13 in predicting self-healing capability of ECC. These models include four individual methods (linear re-  
14 gression (LR), back-propagation neural network (BPNN), classification and regression tree (CART), and  
15 support vector regression(SVR)), and three ensemble methods (bagging, AdaBoost, and stacking) with  
16 the individual models used as the based learner. A series of experimental works on self-healing per-  
17 formance of ECC samples was conducted and the results were used to develop the ML models and compare  
18 the accuracy among the ML models. Among the individual models studied, the BPNN model performed  
19 the best in terms of RMSE and  $R^2$ . In general, all ensemble methods could improve the prediction per-  
20 formance of individual models, however the improvement varies. Among all the ML models studied  
21 including both individual and ensemble methods, the Stack LR model demonstrated the best prediction  
22 results on self-healing of ECC. The results concluded that the individual and ensemble methods can be  
23 used to predict the self-healing of ECC. However, selecting an appropriate base learner and ensemble  
24 method is critical. To improve the performance accuracy, researchers should employ different ensemble  
25 methods to compare their effectiveness with different ML models.

26 **Keywords** ECC, self-healing, machine learning, ensemble method

27 **1 Introduction**

28 According to a research project commissioned by Materials for Life (M4L), the issues associated with crack-  
29 ing in concrete experienced by clients, design team members and contractors were more than any other prob-  
30 lems [1]. Moreover, cracks are primarily responsible for the reduction of strength and stiffness of concrete.  
31 In European countries, the annual cost spent on maintenance, refurbishment, and repair of concrete cracks  
32 in prolonging the service life of infrastructure is estimated around 50% of their annual construction budget  
33 [2]. It has been suggested by M4L that self-healing cementitious materials is of great potential to address the  
34 problems associated with concrete cracking and reduce the maintenance costs over a structure's lifetime [1].

---

\*Correspondence: Weiching Tang

Eamil: patrick.tang@newcastle.edu.au;

Tel.: (+61) 249217246

35 The inspiration of self-healing comes from the biomimicry concept and the healing process in living  
36 nature [3]. For example, the skin of humans or animals can biologically repair itself from simple injuries.  
37 In cement-based materials, the process of crack self-healing can be categorised into two major mechanisms,  
38 autogenous healing and autonomous healing [4]. The former indicates the self-healing ability resulted from  
39 the physical and/or chemical composition of the cementitious matrix, whereas the self-healing mechanism  
40 of the latter is triggered by some biological agents, such as bacteria which are deliberately introduced into  
41 the cementitious matrix.

42 Generally, the autogenous self-healing of concrete is mainly controlled by two mechanisms including  
43 (1) further hydration of cement particles and/or swelling of calcium silicate hydrate; (2) calcium hydroxide  
44 carbonation [5, 6]. It has been reported that the crack widths of 10 $\mu\text{m}$  [7], 100  $\mu\text{m}$  [8], 200  $\mu\text{m}$  [9], 205  $\mu\text{m}$   
45 [5] and 300  $\mu\text{m}$  [10] of ECC can be self-healed completely [11].

46 Engineered cementitious composite (ECC) is a high performance fiber-reinforced cementitious com-  
47 posite and its matrix design is strongly associated with the autogenous self-healing mechanism [12]. ECC  
48 features high tensile ductility with a typical fiber-volume fraction of 2% [13, 14] to promote the self-healing  
49 ability [4]. However, the intrinsic self-healing ability of ECC is complex and difficult to predict because of  
50 different mineral admixture types, interactivity between different composites in the cementitious matrix and  
51 its interaction with the exposed environment [15], and unpredictable crack location, orientation and width  
52 [16]. Previous studies have explored the influence of several factors such as limestone powders (LP) [17, 18],  
53 fly ash (FA) [19, 20], hydrated lime [21], water/binder ratio [22], water permeation [23] and different cur-  
54 ing conditions (air, carbon dioxide, wet/dry and water) [24] on self-healing behaviour of ECC. However,  
55 the relationship between multiple factors is unclear and non-linear, so it's difficult to predict self-healing of  
56 ECC mathematically based on the available data. Moreover, mathematical models based on empirical data  
57 are generally in regression forms, which cannot be used when the problem (e.g. prediction of self-healing  
58 potential of ECC) contains too many independent variables and more assumptions are required [25].

59 To compensate for the drawbacks of mathematical models with multiple interaction variables, machine  
60 learning (ML) techniques have been used for solving many civil engineering problems with multiple vari-  
61 ables. They are model-free approaches that do not rely on predefined models [26]. Many research works have  
62 been conducted using ML algorithms for the prediction of various properties of concrete. Gilan et al.[27] de-  
63 veloped a hybrid Support Vector Regression (SVR) - Particle Swarm Optimization (PSO) algorithm model to  
64 predict the compressive strength and Rapid Chloride Penetration Test (RCPT) results of concretes containing  
65 metakaolin. Yan et al. [28] predicted bond strength of glass fiber-reinforced polymer bar in concrete by Arti-  
66 ficial Neural Network (ANN) with Genetic Algorithm (GA). Yaseen et al.[29] proposed a ML method called  
67 Extreme Learning Machine (ELM) to predict the compressive strength of lightweight foamed concrete.

68 In the literature, the performance of various ML algorithms in predicting concrete properties have been  
69 evaluated and compared. Yan and Shi [30] reported that SVR is better than other individual methods in  
70 predicting elastic modulus of normal and high strength concrete. Chou [31] compared the performance of  
71 individual and ensemble methods for predicting the mechanical properties of high performance concrete, the  
72 results reveal that ensemble learning strategies outperform individual learning techniques in predicting the  
73 compressive strength of high performance concrete. Reuter et al. [26] employed three individual approaches  
74 for modeling concrete failure surfaces. They found the three approaches are able to fit the experimental data  
75 with low error. Sobhani et al. [32] suggested that their proposed fuzzy inference system and ANN are more  
76 reliable than traditional regression models on predicting no-slump concrete. Omran et al. [33] compared the  
77 compressive strengths of an environmentally friendly concrete predicted by using three individual methods,  
78 two ensemble methods, and four regression tree models. Their results showed that the individual gaussian  
79 process regression model and its related ensemble models outperformed other models.

80 Although different ML algorithms have been utilized to predict several properties of concrete, the appli-  
81 cation of ML on self-healing prediction is considerably rare. Recently, Mauludin and Oucif [34] reviewed  
82 the common methods used for modeling autogenous self-healing of concrete, and stated that the methods  
83 can be classified into two categories: (1) numerical simulation and (2) ML. However, the only ML model

84 reviewed in their study was the GA-ANN method proposed by Ramadan et al. [3]. They predicted the self-  
85 healing ability of cement-based materials using a dataset collected from literature. The results showed that  
86 the GA-ANN model was capable of capturing the complex effects of various self-healing agents (e.g., bio-  
87 chemical material, silica-based additive, expansive and crystalline components) on self-healing performance  
88 of cement-based materials.

89 Chaitanya et al. [36] used an ANN model to predict the self-healing property of concrete containing  
90 ground granulated blast furnace slag in terms of compressive strength recovery based on 51 samples collected  
91 from their experimental studies. Generally, the predicted results obtained from the ANN model were in good  
92 agreement with the experimental values. Zhuang and Zhou [37] conducted a comparative study on six ML  
93 algorithms including SVR, Decision Tree Regression (DTR), Gradient Boosting Regression (GBR), ANN,  
94 Bayesian Ridge Regression (BRR) and Kernel Ridge Regression (KRR) for crack-repairing capacity of the  
95 bacteria-based self-healing concrete. The results showed that GBR performed much better than other models  
96 with  $R^2$  values of 0.93 and 0.74 for the training set and testing set, respectively. However, the  $R^2$  values  
97 of most models were less than 0.7 on both training and testing sets. Although extensive experiments with  
98 different combinations of influencing variables were utilized to generate the empirical dataset, their study  
99 only selected three variables including the number of bacteria, the healing time and the initial crack width to  
100 predict the crack closure percentage as the output.

101 To the best of our knowledge, there has been no study to date to predict the self-healing of ECC using  
102 ML approach. The information about prediction performance of individual and ensemble ML models on  
103 self-healing of ECC would be very useful to the design of ECC with self-healing capacity. Thus, this study  
104 aims to provide a comparative analysis on the performance of various ML models in predicting self-healing  
105 capability of ECC. The ML model with the best performance can be used as a baseline prediction model for  
106 developing advanced models in the future.

107 In this paper, four ML individual methods including linear regression (LR), SVR, back-propagation neural  
108 network (BPNN), and classification and regression tree (CART) were proposed to predict the self-healing  
109 capability of ECC. To improve prediction accuracy, three ensemble methods namely bagging, AdaBoost and  
110 stacking were used to construct ensemble models using the individual models as the base learners. A series of  
111 experimental works on self-healing performance of ECC samples was conducted and the results were used to  
112 develop the ML models and compare the accuracy among the ML models. Experimental data collected from  
113 the experiments were first preprocessed and then divided into a 10-fold cross-validation algorithm (details  
114 refer to Section 4.1) to avoid overfitting. Figure 1 summarizes the steps that were performed when predicting  
115 the self-healing capability of ECC.

116 This paper is organized as follows. Section 2 presents the experimental program detailing the materials  
117 used for ECC specimen preparation and the test set-up for crack data measurement. The concepts and  
118 formulations of individual and ensemble models used for predicting the self-healing capability of ECC are  
119 presented in Section 3, whereas the validation and evaluation methods are described in Section 4. In Section  
120 5, the computational results are presented and compared, and the model with best prediction performance is  
121 identified. Finally, Section 6 draws the major conclusions from this work and suggests some directions for  
122 future research.

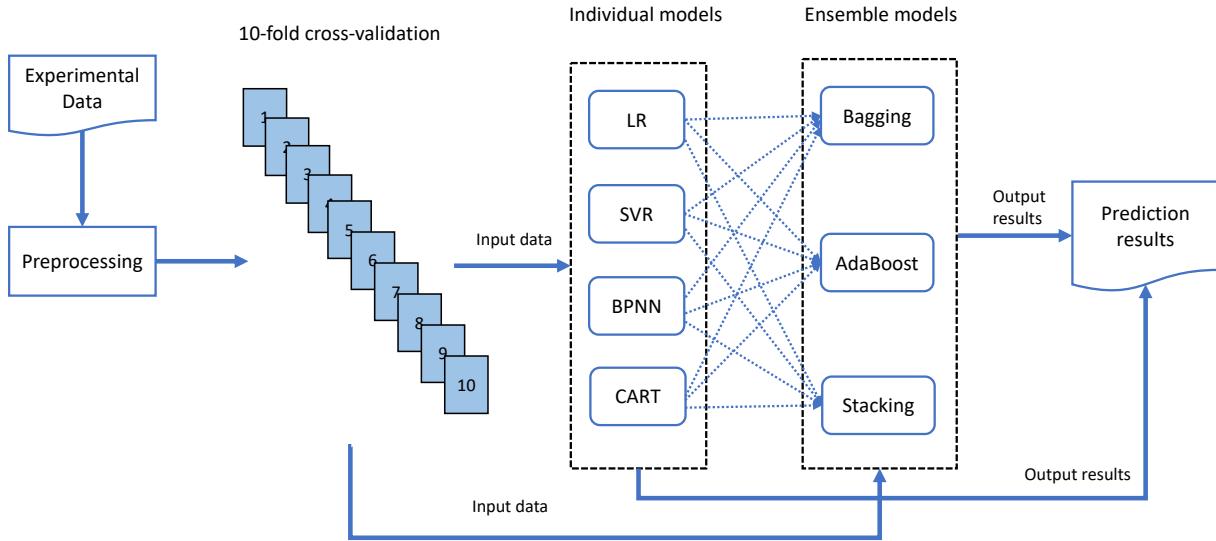


Figure 1: Flow chart of implementing prediction models for self-healing capability of ECC

## 2 Experimental Program

### 2.1 Materials and Mixture Proportion

In the experimental part, samples of ECC with different mineral admixtures were prepared. The materials including general purpose cement (GPC), fly ash (FA), silica fume (SF), hydrated lime powder(LP), fine sand, polyvinyl alcohol (PVA) fibers, as well as water and high range water reducing admixture (HRWR) were used. GPC and FA were supplied by Boral in accordance with Australian Standard AS 3972-2010 [38], while LP was the Adelaide Brighton Hydrated Lime with a specific gravity of 2.2-2.3, and a typical fineness of 0.1% retained on a  $75\text{ }\mu\text{m}$  sieve and less than 0.05% on a  $250\text{ }\mu\text{m}$  sieve. The physical and chemical properties of cementitious materials are shown in Table 1. Fine sand with an average grain size of  $150\text{ }\mu\text{m}$  and a fineness modulus of 2.01 was used. The PVA fibers were supplied by Domocrete and their mechanical and geometrical properties are described in Table 2.

All ECC mixtures were prepared with a constant water to cementitious materials (W/CM) ratio of 0.29 and a constant sand to CM (PC + FA + LP+SF) ratio of 0.36. All fine aggregates were in saturated surface dried condition prior to mixing. The abbreviations for labelling specimens were adopted in such a way that the letters FA, SF and LP stand for samples with fly ash, silica fume and limestone as binder materials, respectively. The number after the letters shows the percentage of materials into the binder system. For example, the FA70 mixture is related to an ECC sample with binder containing 70% FA by weight, whereas FA60-SF10 was the mixture with 60% FA and 10% SF. A total of nine ECC mixtures were prepared and the details of mix proportion are shown in Table 3.

### 2.2 Sample preparation and crack measurement

A planetary-type mixer of 50 L capacity was used to produce ECC specimens. During the mixing process, the solid ingredients including cement, mineral admixtures and sand were initially placed into the mixer and dry mixed for 30 seconds. Then, the water with HRWR was added and the mixture was mixed for 2 minutes. After that, the PVA fibers were slowly added and mixing was continued until uniform distribution of fibers in the mix. After mixing, ECC pastes were cast into standard moulds with dimension of  $\varnothing 100\text{mm} \times 200\text{mm}$ .

Table 1: Physical and chemical properties of cementitious materials

<i>Chemical composition (%)</i>	GPC	FA	LP	SF
Silica ( $\text{SiO}_2$ )	19.8	65.90	1.8	95.10
Alumina ( $\text{Al}_2\text{O}_3$ )	5.3	24.0	0.5	0.21
Iron oxide ( $\text{Fe}_2\text{O}_3$ )	3.0	2.87	0.6	0.29
Calcium oxide ( $\text{CaO}$ )	64.2	1.59	72.0	-
Magnesia ( $\text{MgO}$ )	1.3	0.42	1.0	-
$\text{R}_2\text{O}$	0.6	1.93	-	-
Sulfur trioxide ( $\text{SO}_3$ )	2.7	-	-	-
Titanium oxide ( $\text{TiO}_2$ )	0.28	0.91	-	-
Manganic oxide ( $\text{Mn}_2\text{O}_3$ )	0.22	-	-	-
Zirconia ( $\text{ZrO}_2$ ) + Hafnium ( $\text{HfO}_2$ )	-	-	-	3.46
Loss on ignition (%)	2.8	1.53	24.0	1.4
Density ( $\text{g/cm}^3$ )	3.08	2.43	2.25	2.26
Specific surface area ( $\text{m}^2/\text{kg}$ )	-	655	460	$1.5 \times 10^4$

Table 2: Properties of PVA

Length (mm)	Length/ diameter ratio	Young's modulus (MPa)	Elongation (%)	Tensile strength (MPa)	Density ( $\text{g/cm}^3$ )
8	200	42000	7	1600	1.3

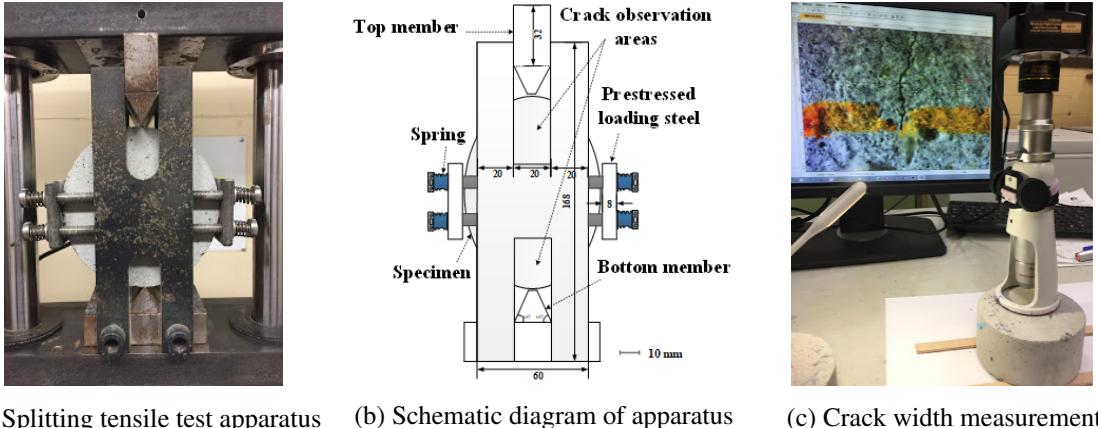
Table 3: Mix proportion of all ECC mixtures

Mix	Water/CM	Sand	Water	fibre (V)	GPC	Fly ash	SF	LP	HRWR
FA70	0.29	419.67	338.07	26	349.73	816.03	0.00	-	5.13
FA65-SF5	0.29	419.67	338.07	26	349.73	757.74	58.29	-	5.13
FA60-SF10	0.29	419.67	338.07	26	349.73	699.45	116.58	-	5.13
FA55-SF15	0.29	419.67	338.07	26	349.73	641.16	174.86	-	5.13
FA65-LP5	0.29	419.67	338.07	26	349.73	757.74	-	58.29	5.13
FA60-LP10	0.29	419.67	338.07	26	349.73	699.45	-	116.58	5.13
FA55-LP15	0.29	419.67	338.07	26	349.73	641.16	-	174.86	5.13
FA55-SF5-LP10	0.29	419.67	338.07	26	349.73	641.16	58.29	116.58	5.13
FA55-SF10-LP5	0.29	419.67	338.07	26	349.73	641.16	116.58	58.29	5.13

148 The specimens were demolded 24 hours after casting and stored in a curing room with a temperature of  
 149  $23 \pm 2^\circ\text{C}$  and the relative humidity (RH) of  $90 \pm 5\%$  for 28 days for 28 days. To prepare splitting tensile  
 150 test samples, the cylinder specimens were cut into specimens with a diameter of 100 mm and a thickness of  
 151 50 mm using a diamond blade saw.

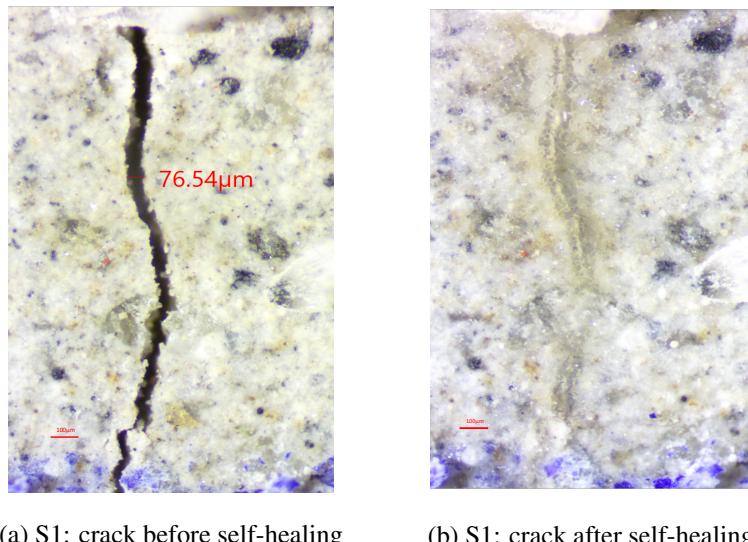
152 A newly developed splitting tensile test apparatus was used to generate micro-cracks as shown in Figure  
 153 2 (a). It consisted of a steel frame, top member, bottom member, prestressed loading steel plates (5 mm  
 154 thick) on both sides with loading nuts and wire springs, as shown in Figure 2 (b). Both steel plates were  
 155 connected to the steel frame by nuts and wire springs. The specimen was placed inside the steel frame and  
 156 then pre-stressed by the steel plates from both sides limiting the propagation and size of crack and preventing  
 157 excessive crack growth.

158 Micro-cracks less than 150  $\mu\text{m}$  were produced by pre-loading the ECC samples up to 70% of their  
 159 maximum splitting strength. A digital microscope was used to measure the crack width on the surface of  
 160 specimens as shown in Figure 2 (c). After the pre-loading, the cracked specimens were subjected to wet-dry  
 161 (W/D) cycles to promote self-healing. Each W/D cycle consisted of submersion in water for 24 hours and  
 162 drying in laboratory conditions at  $23 \pm 2^\circ\text{C}$  and a RH of  $50 \pm 5\%$  for 24 hours. After 10 W/D cycles, the  
 163 cracks were measured again by the digital microscope to examine the extent of crack recovery. Figure 3  
 164 illustrated the self-healing of cracks of an ECC specimen before and after the 10 W/D cycles.



(a) Splitting tensile test apparatus    (b) Schematic diagram of apparatus    (c) Crack width measurement

Figure 2: Splitting tensile test apparatus and microscope used in experiment for creating and measuring ECC cracks



(a) S1: crack before self-healing    (b) S1: crack after self-healing

Figure 3: Comparison of crack width changes in two ECC specimens, S1 and S2, before and after self-healing

### 165 2.3 Data Collection

166 Experimental data for prediction were gathered with four features, including crack width before self-healing  
 167 (representing the influencing factor of self-healing), and the mineral contents of FA, SF, and LP. It is note-  
 168 worthy that the factors such as GPC, sand, W/CM, and healing time were kept constant and hence, they were

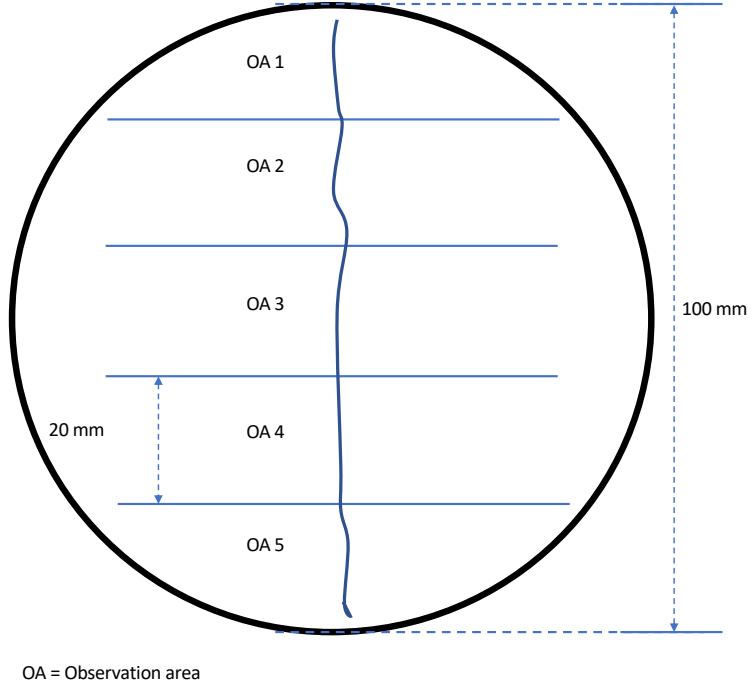


Figure 4: Schematic diagram of measuring observation areas on the surface of ECC mixture specimen

169 excluded in the prediction modeling. For each ECC mixture, there were 6 identical test specimens. After  
 170 pre-loading, the crack widths of the specimens were measured using the digital microscope before and after  
 171 the self-healing. Four horizontal lines were drawn on the surface of each specimen along the direction of  
 172 vertical force, which divided the specimen into five observation areas. The schematic diagram of the mea-  
 173 surement is shown in Figure 4. In each observation area, only one crack data was recorded if the crack width  
 174 showed little or no change along the vertical force, otherwise, multiple crack data would be collected. To-  
 175 tally, 617 crack data samples were collected from nine mixtures to construct the ML training-testing dataset  
 176 [39]. Table 4 shows the number of collected samples and range of crack widths before and after self-healing  
 177 in each mixture.

## 178 2.4 Preprocessing of Data

179 Since the input and output data of different features vary in range and units, the features with bigger number  
 180 would steer the model performance. As shown in Table 3, the range of FA varies from 641.16 to 816.03 kg,  
 181 but the range of SF varies from 0 to 174.86 kg. From Table 4, the range of crack width varies from 0 to  
 182 135.47  $\mu\text{m}$ . To eliminate this potential bias, the experimental data was preprocessed through the min-max  
 183 normalization to scale the range of all features into [0,1] with the following equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

184 Where  $x'$  was the scaled value of the variable  $x$ ,  $x_{max}$ ,  $x_{min}$  were the maximum and minimum values of  
 185 variable  $x$  respectively.

Table 4: Number of crack samples and range of crack width before and after self-healing collected from the ECC mixes

Mix	Number of crack samples	Crack width before self-healing		Crack width after self-healing	
		Min ( $\mu m$ )	Max ( $\mu m$ )	Min ( $\mu m$ )	Max ( $\mu m$ )
FA70	87	3.28	134.69	0	121.37
FA65-SF5	77	4.37	135.47	0	124.01
FA60-SF10	88	5.18	121.78	0	113.11
FA55-SF15	88	3.45	115.8	0	109.53
FA65-LP5	112	7.65	119.45	0	105.65
FA60-LP10	37	5.62	126.82	0	110.97
FA55-LP15	61	6.42	132.65	0	115.95
FA55-SF5-LP10	34	8.74	123.09	0	110.78
FA55-SF10-LP5	33	4.64	131.57	0	119.79

### 186 3 Proposed Machine Learning Models

187 To predict the self-healing capability of ECC, four individual ML models including LR, SVR, BPNN and  
 188 CART, and three ensemble methods including bagging, AdaBoost and stacking were proposed. Ensemble  
 189 models were constructed using individual models as the base estimators. To establish a baseline for compar-  
 190 ison, the modeling parameters were set to be the same in both individual models and ensemble models. The  
 191 reason for choosing these techniques was due to their popularity and some of them were even recognized  
 192 as the top data mining algorithms in related fields of concrete [31]. The proposed individual and ensemble  
 193 techniques are described in the following subsections.

#### 194 3.1 Linear Regression

195 LR attempts to determine the relationship between a dependent variable (response variable) and one or more  
 196 independent variables (explanatory variables) by fitting a linear regression equation [40]. Given our dataset  
 197  $T = \{(x_i, y_i), i = 1, 2, \dots, n\}$ , where  $n = 617$  was the size of sample dataset.  $x_i \in R^n$  was independent  
 198 variables representing a sample of selected features from FA, SF, LP and crack width before self-healing,  $R^n$   
 199 was  $n$ -dimensional space,  $y_i \in R^1$  was the target output (crack width after self-healing) that corresponded  
 200 to  $x_i$ . Let  $d = 4$  denote the number of an independent variable of a random vector  $x = \{x_1; x_2; \dots; x_d\}$ , and  
 201  $y$  was the corresponding output ( dependent variable). The general formula of LR for predicting self-healing  
 202 capability of ECC can be expressed as follows:

$$y = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b \quad (2)$$

203 where  $w_i, (i = 1, 2, \dots, d)$  was denoted as the regression coefficient,  $b$  was an error term. The prediction  
 204 performance of LR was used as a benchmark to compare the performance of other individual and ensemble  
 205 models in this study.

#### 206 3.2 Support Vector Regression

207 The support vector machine (SVM) is a supervised machine learning method first introduced by Vapnik [41,  
 208 42] based on statistical learning theory [43]. Since then, it has gained popularity due to attractive features,  
 209 and promising empirical performance. SVM includes two main categories: support vector classification

210 (SVC) and SVR. For classification purposes, SVMs often used a *kernel* function to map the input data as  
 211 vectors to a high-dimensional feature space so that an optimal separating hyperplane can be constructed [44].

212 For regression purposes, the basic idea is to provide a nonlinear function by mapping input data into a  
 213 high-dimensional feature space, where a special type of hyperplane is constructed. After that, a regression  
 214 model is established in the hyperplane [45].

215 Given our dataset  $T = \{(x_i, y_i), i = 1, 2, \dots, n\}$ , where  $n = 617$  was the size of sample dataset,  $x_i \in R^n$   
 216 was the input vector representing selected features of a sample, including FA, SF, LP and crack width before  
 217 self-healing,  $R^n$  was the  $n$ -dimensional vector space,  $y_i \in R^1$  was the target output indicating crack width  
 218 after self-healing that corresponded to  $x_i$ . The SVR aimed to seek an optimum regression function  $f(x)$  with  
 219 minimal empirical risk, which can be expressed as follow:

$$f(x) = \langle w, x \rangle + b \quad \text{with } w \in T, b \in R \quad (3)$$

220 where  $\langle \cdot, \cdot \rangle$  was denoted as the dot product in  $T$ ,  $w$  and  $b$  were the weight vector and bias value which are  
 221 estimated by minimizing the empirical risk, that was, the distance between the predicted crack width and the  
 222 target crack width after self-healing.

223 SVR adopts an  $\epsilon$ -insensitive loss function penalizing predictions that has a distance between the predicted  
 224 crack width and the target crack width when the self-healing is greater than  $\epsilon$ . Therefore, the problem of  
 225 finding  $w$  and  $b$  to reduce the empirical risk with respect to an  $\epsilon$ -insensitive loss function is equivalent to the  
 226 convex optimization problem that minimizes the margin ( $w$ ) with the full prediction error within the range  
 227 of  $\epsilon$ . Then this problem can be expressed as:

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|w\|^2 \\ & \text{subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon \\ \langle w, x_i \rangle + b - y_i \leq \epsilon \end{cases} \end{aligned} \quad (4)$$

228 By introducing slack variables  $\xi, \xi_i^*$  to allow some errors to cope with infeasible solution of the optimization  
 229 problem, the formulation can be generated as [42]:

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (5)$$

230 The constant  $C$  was the penalty value imposed on predictions that lied outside the  $\epsilon$  margin. Lagrange  
 231 multipliers are included to solve this problem. By constructing the objective function and all constraints, a  
 232 dual set of variables are introduced as follows: [46]:

$$\begin{aligned} L_P = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^n \alpha_i (\epsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^n \alpha_i^* (\epsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \\ s.t. \quad & \alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0 \end{aligned} \quad (6)$$

233 Where  $L_P$  was the Lagrangian and  $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$  were Lagrange multipliers.

234 The optimality can be achieved by the partial derivatives of  $L_P$  with respect to the primal variables  
235 following the saddle point condition. Then the function of SVR is obtained as:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (7)$$

236 As for the nonlinear regression, the input data have to be mapped into a high-dimensional feature space,  
237 in which the dot product can be replaced by a kernel function  $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ , and the function (7)  
238 can be written as:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (8)$$

239 Different SVM algorithms use differing kinds of kernel functions such as linear, polynomial, radial basis  
240 function and sigmoid kernel. In this work, the Gaussian radial basis function (RBF) was chosen, which was  
241 defined as [47]:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

### 242 3.3 Artificial Neural Network

243 Artificial neural network (ANN), also called neural network, is originated from simulating biological neural  
244 networks. Generally, it consists of many neurons in layers including one input layer, one or several hidden  
245 layers and an output layer [48]. The neurons are fully interconnected between the neighboring layers by  
246 weight, and typically no inter-connections between neurons within the same layer [49].

247 There are many possible network structures available, BPNN was utilized in this study because of back-  
248 propagation (BP) algorithms is the most widely used and effective learning algorithm for training an ANN.  
249 A preliminary architecture of the BPNN was determined to be 4 - n - 1, where 4 input neurons represented  
250 the input features standing for FA, LP, SF and crack width before self-healing, n = 5 indicated the num-  
251 ber of neurons in the hidden layer, and 1 target neuron in the output layer for the predicted crack width  
252 after self-healing. This is a three-layer network with one hidden layer capable to approximate most contin-  
253 uous functions, of which the complex nonlinear relationship could be approximated in accuracy [28]. The  
254 architecture of the BPNN model for predicting self-healing is demonstrated in Figure 5.

255 Given a set of inputs  $\{x_1, x_2, x_3, \dots, x_n\}$ , while information was passed through the input layer to the  
256 hidden layer, each neuron in the input layer was multiplied by respective weights added by a bias and are  
257 summed together. After that, an activation function  $f$  was applied to form the output  $z$ . This can be expressed  
258 in the following equation [25]:

$$z = f\left(\sum_{i=1}^n w_{ij} x_i + b_j\right) \quad (10)$$

259 where  $w_{ij}$  was the connection weights between the  $i$ th neuron of input and the  $j$ th neuron in the hidden  
260 layer, and  $b_j$  was the bias of the  $j$ th neuron. The sigmoid function was applied as the activation function  
261 between the input, hidden, and output neurons to form the output.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

262 The goal of training a neural network is to determine the values of the connection weights and the biases  
263 of the neurons. The back propagation indicates an iterated method to adjust the weights from output layer

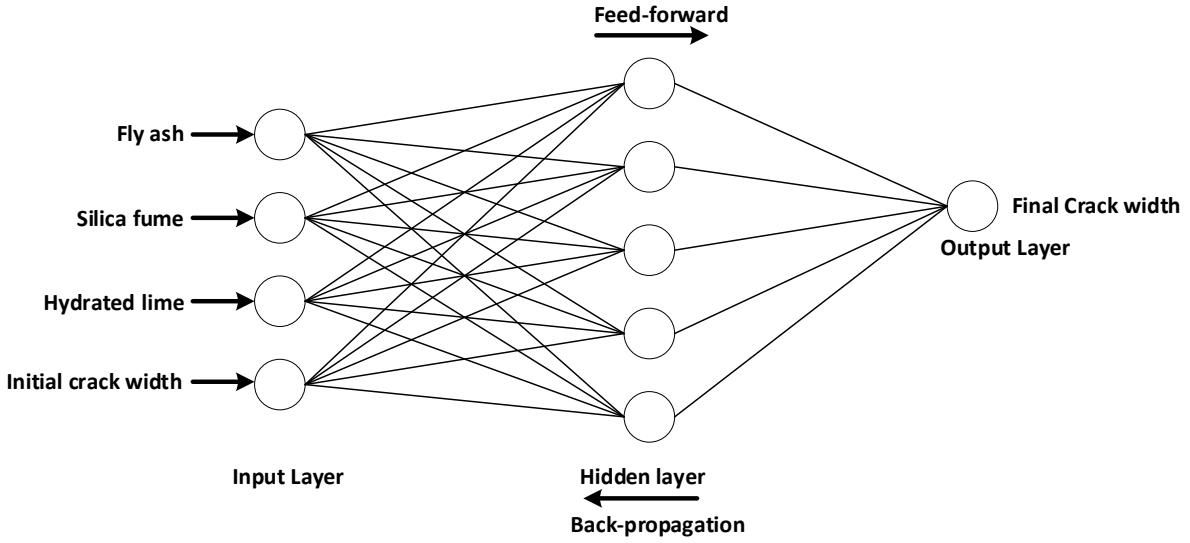


Figure 5: Schematic diagram of BPNN model for predicting self-healing capability of ECC

264 to input layer. At first, the outputs were calculated feed-forward from the input layer via the hidden layer  
 265 to the output layer. Then an error was generated by comparing the output with the target output. After that,  
 266 the error was back propagated to the hidden layer and input layer. By adjusting the connection weights and  
 267 biases, the error was further reduced. The process was repeated until the error was minimised or reaching  
 268 the termination to avoid over-fitting.

### 269 3.4 Classification and Regression Tree

270 The CART [50] is a tree decision algorithm that splits data into mutually exclusive subgroups based on  
 271 recursive binary partitioning procedure. It develops the relationship between the target variables (the crack  
 272 width after self-healing of ECC) and the independent variables (the input features of FA, SF, LP and crack  
 273 width before self-healing of ECC) to create decision rules to form subgroups as branches and leaves as shown  
 274 in Figure 6. The process of CART starts from the root node which contains the entire data set to construct  
 275 two sub-nodes representing two categories. Then this recursion process is applied to each sub-node until all  
 276 divided sub-nodes are leaf nodes. The CART tree can be either a classification tree [51] or regression tree  
 277 [52] depending on the type of target and independent variables which may be categorical or numerical.

278 The key idea of constructing a CART tree is achieved by selecting a variable at each node that best splits  
 279 the empirical data. To locate splits, *Gini* index was used to measure the impurity of the two child nodes  
 280 containing subsets of data that were as homogeneous as possible with respect to the target variable.

281 Given a dataset had  $K$  classes and the probability of a record in the dataset which belongs to class  $i$  is  
 282  $p_i, i \in \{1, 2, 3, \dots, K\}$ , the *Gini* impurity can be expressed as:

$$G(p) = \sum_{i=1}^K p_i(1 - p_i) = 1 - \sum_{i=1}^K p_i^2 \quad (12)$$

### 283 3.5 Ensemble Methods

284 In contrast to many ML approaches such as SVM and CART (which develop a single learner from training  
 285 data), ensemble methods train multiple base learners and combine them [31] to improve generalizability

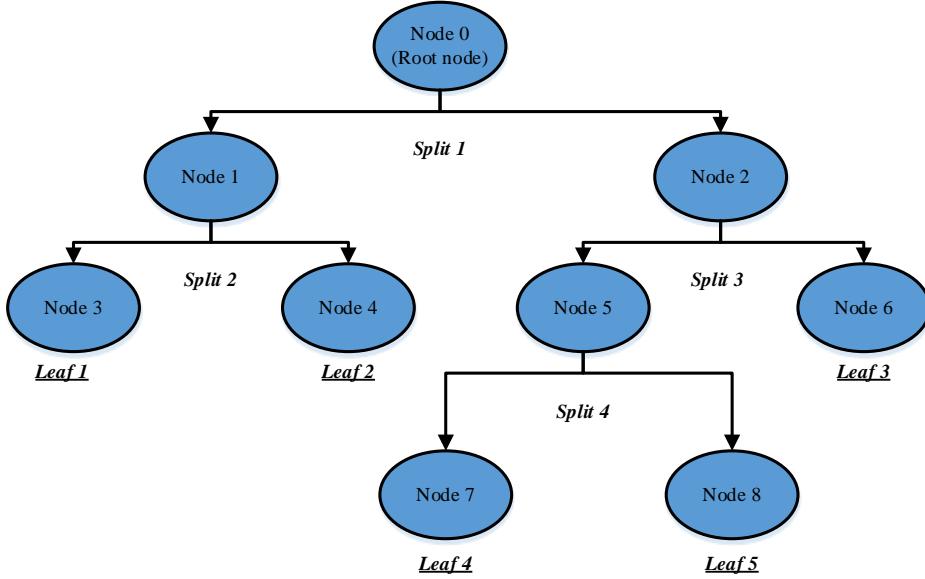


Figure 6: Structure of a classification and regression tree [52]

286 over a single estimator. Therefore, weak learners (base learners) can be boosted to become strong learner  
 287 [53] in an ensemble method. The base learners in an ensemble were developed from an individual learning  
 288 algorithm such as decision tree, SVM, or other kinds of learning algorithms. Breiman [54] showed that  
 289 ensemble methods are usually more accurate than individual learning methods.

290 The input features of FA, SF, LP, and crack width before self-healing of ECC were considered as the  
 291  $d$ -dimensional predictor variable  $X$ , whereas, the crack widths after self-healing of ECC were the one dimen-  
 292 sional output  $Y$ . Each estimator used an individual algorithm to provide one estimated function  $g(\cdot)$ .  
 293 The output presented by ensemble-based function  $g_{\text{en}}(\cdot)$  was obtained by a linear combination of individual  
 294 functions. This ensemble approach can be expressed mathematically as:

$$g_{\text{en}}(\cdot) = \sum_{j=1}^N c_j * g(\cdot) \quad (13)$$

295 Where  $c_j$  expressed as the combination coefficients, dependent on the used ensemble models.

### 296 3.5.1 Bagging

297 Bagging method (bootstrap aggregating) can generate multiple versions of a predictor to obtain an aggre-  
 298 gated predictor [55]. It generates multiple models independently on different versions of dataset via random  
 299 bootstrapping of the original training set. In other words, several training examples could repeatedly ap-  
 300 pear in different bootstrap replicates. Then the individual predictions are aggregated through a combination  
 301 method (either voting or averaging) to form the final prediction. Bagging method can be used to reduce the  
 302 variance of a base estimator (e.g. a regression tree), by introducing randomization into its construction pro-  
 303 cedure and making an ensemble out of it. This study used four individual models to build bagging ensemble  
 304 models including a LR bagging ensemble model (abbreviated as Bag\_LR), a SVR bagging ensemble model  
 305 (abbreviated as Bag\_SVR), a BPNN bagging ensemble model (abbreviated as Bag\_BPNN), and a CART  
 306 bagging ensemble model (abbreviated as Bag\_CART).

307 **3.5.2 AdaBoost**

308 Similar to bagging, AdaBoost method [56] manipulates the training examples to generate multiple predictions  
309 to form the final prediction. The main difference with bagging is that AdaBoost applies a weight to each  
310 of the training examples. In each iteration, the weights are individually updated to minimize the weighted  
311 error on the training set. For example, weights on those training examples incorrectly predicted in previous  
312 iteration increase, whereas the weights of the correctly predicted training examples decrease. Therefore,  
313 AdaBoost tends to construct progressively more difficult learning problems in subsequent iterations. Once  
314 the training process has finished, the predictions are combined through a weighted majority vote (or sum) to  
315 produce the final prediction. So, the final classifier usually can achieve a high degree of accuracy in the test  
316 set.

317 By combining four individual models as base estimators in AdaBoost, this study obtained four AdaBoost  
318 ensemble models. They are a LR AdaBoost ensemble model (abbreviated as Ada\_LR), a SVR AdaBoost en-  
319 semble model (abbreviated as Ada\_SVR), a BPNN AdaBoost ensemble model (abbreviated as Ada\_BPNN),  
320 and a CART AdaBoost ensemble model (abbreviated as Ada\_CART).

321 **3.5.3 Stacking**

322 Stacking regression combines multiple regression models via a meta-regressor, using out-of-fold prediction  
323 concept [57]. The stacking method used in this work splits the data set into k folds, in which the k-1 folds are  
324 used to train the first level regressors in k successive rounds. In each round, the first level regressors are used  
325 to predict based on the remaining 1 subset. After that, the prediction results are used and stacked as input  
326 data to the second level regressors to form a final set of predictions [58]. The schematic diagram of stacking  
327 model is shown in Figure 7. In this study, one stacking based ensemble model (abbreviated as Stack\_LR)  
328 was proposed based on two levels scheme. SVR, BPNN and CART were used as regression models in the  
329 first level to get the prediction results, and LR was used as meta-regressor in the second level to combine and  
330 generate the final prediction results.

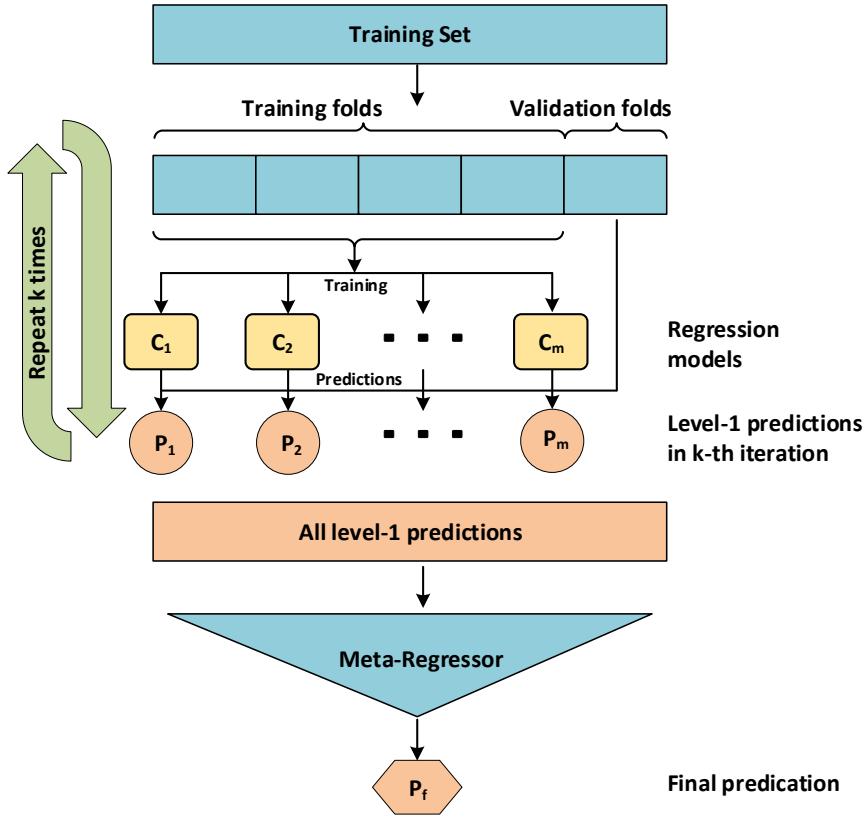


Figure 7: Schematic diagram of Stacking model [58]

## 331 4 Validation and Evaluation

### 332 4.1 Cross-validation Method

333 Generally, dataset is split to generate a training subset and a validation subset keeping the properties of the  
 334 original dataset as much as possible to avoid misleading estimates. To minimize bias of random data splitting,  
 335 the K-fold cross-validation is commonly used as it can yield optimal computational time and reliable variance  
 336 [31, 59]. In this study, a ten-fold cross-validation approach was applied to assess model performance as  
 337 shown in Figure 8. The dataset was split randomly into 10 equal-size subsets with a similar distribution. In  
 338 each validation process, nine of the subsets were used for training and the rest for testing. The process was  
 339 repeated 10 times [60]. The average accuracy after 10 times validation was reported as the model accuracy.

### 340 4.2 Performance Evaluation

341 To show and validate the accuracy of the proposed ML models, three statistical indices namely mean ab-  
 342 solute Error (MAE), root mean square error (RMSE), and the coefficient of determination  $R^2$  were used  
 343 and expressed in equations (14), (15), and (16), respectively. The average deviation of the performance of  
 344 an individual model or an ensemble model from a benchmark model in terms of three statistical measures  
 345 (MAE, RMSE and  $R^2$ ) was calculated using equation (17).

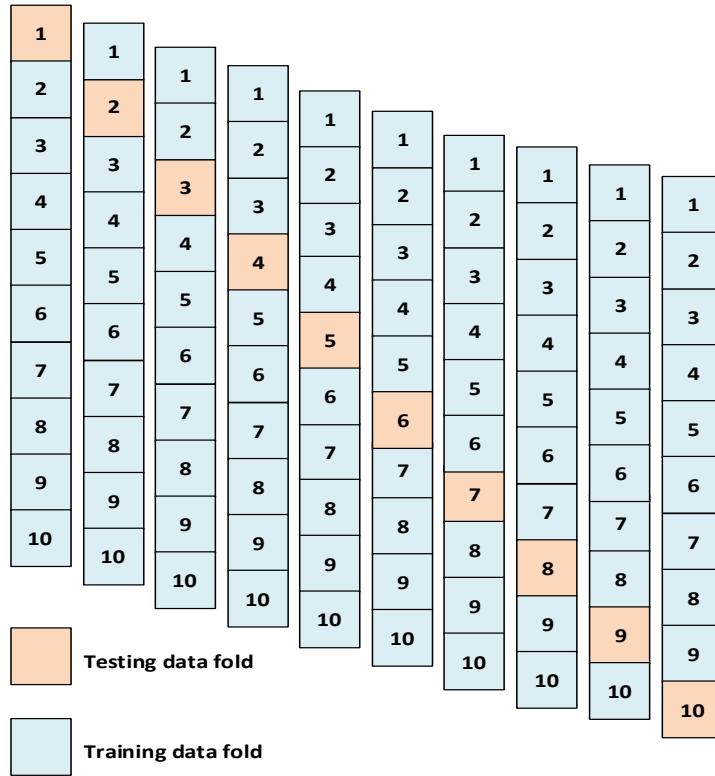


Figure 8: Ten-fold cross-validation approach

346

- Mean absolute error (MAE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (14)$$

347

- Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (15)$$

348

- Coefficient of determination ( $R^2$ )

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

349

- Deviation ( $Dev$ )

$$Dev(\%) = \frac{P_i - P_j}{P_j} * 100 \quad (17)$$

350

Where  $y_i$  was the target output,  $y'_i$  was the predicted output,  $n$  was the number of samples,  $\bar{y}$  was the mean of the target output.  $Dev$  indicated the statistical performance improvement compared with a benchmark model,  $P_i$  was the statistical performance (MAE, RMSE or  $R^2$ ) of an individual or ensemble method, and  $P_j$

353 was the corresponding performance of a benchmark model, LR or an individual method used in the ensemble  
 354 method as the base learner.

355 MAE statistics is a measure of errors between the predicted values (the estimated value of crack width of  
 356 ECC after self-healing) with the target values (the observed value of crack width of ECC after self-healing in  
 357 empirical data). RMSE statistics computes the square root of the average residual error between the predicted  
 358 values and the target values. A lower value of MAE or RMSE indicates a better prediction performance of  
 359 the model.  $R^2$  measures the strength of association between the predicted values and the target values, based  
 360 on the proportion of total variation of outcomes. A greater value close to 1 represents a better prediction  
 361 performance that commendably replicates the observed crack width of ECC after self-healing. Deviation  
 362 statistics indicates the improvement of the prediction performance of an individual or an ensemble model  
 363 from a benchmark model that can be the LR model or the individual model used as base learners in the  
 364 corresponding ensemble model.

## 365 5 Results and Discussion

366 In this section, the prediction performance of individual and ensemble methods are examined by MAE,  
 367 RMSE and  $R^2$  according to the ten-fold cross-validation. The abbreviation for labelling models were adopted  
 368 in a such a way that the letters Bag, Ada and Stack stand for the ensemble methods of Bagging, AdaBoost  
 369 and Stacking, respectively. The letters LR, SVR, BPNN and CART stand for the base estimators. However,  
 370 Stack\_LR refers to the model combining the base methods including SVR, BPNN, and CART in the first  
 371 level and using LR as a meta-regressor in the second level.

### 372 5.1 Prediction performance of the proposed models

373 Table 5 shows the average performance of individual and ensemble models. The ten-fold cross-validation  
 374 results (MAE, RMSE, and  $R^2$ ) for both individual and ensemble models and their deviation with respect to  
 the results of LR model.

Table 5: Average performances of machine learning models for self-healing prediction of ECC

	Models	MAE	Dev(%)	RMSE	Dev(%)	$R^2$	Dev(%)
Individual models	LR	5.012	-	7.680	-	0.860	-
	BPNN	4.329	-13.6	6.515	-15.2	0.899	4.5
	CART	4.305	-14.1	6.811	-11.3	0.887	3.1
	SVR	4.296	-14.3	6.826	-11.1	0.883	2.7
Ensemble models	Ada_LR	4.784	-4.6	7.400	-3.6	0.867	0.8
	Ada_BPNN	4.226	-15.7	6.435	-16.2	0.900	4.7
	Ada_CART	4.207	-16.1	6.455	-15.9	0.898	4.4
	Ada_SVR	4.145	-17.3	6.577	-14.4	0.893	3.8
	Bag_LR	5.014	0.0	7.689	0.1	0.860	0.0
	Bag_BPNN	4.143	-17.3	6.341	-17.4	0.901	4.8
	Bag_CART	4.093	-18.3	6.358	-17.2	0.901	4.8
	Bag_SVR	4.302	-14.2	6.820	-11.2	0.883	2.7
	Stack_LR	3.934	-21.5	6.118	-20.3	0.904	5.1

375 Generally, most of the proposed models were able to learn and predict empirical data with an acceptable  
 376 degree of precision. Based on the results, the Stack\_LR model showed the best prediction performance as  
 377 it has the highest  $R^2$  value and the lowest MAE and RMSE values. Among the individual models, SVR  
 378 performed the best in terms of MAE (4.296), whereas BPNN has the lowest RMSE value (6.515) and the  
 379 highest  $R^2$  of 0.899. Among the individual models boosted by either AdaBoost or bagging, Bag\_CART gave  
 380

381 the best performance in terms of MAE (4.093), while Bag\_BPNN performed better on RMSE value (6.341).  
 382 In terms of  $R^2$ , Bag\_CART and Bag\_BPNN models showed the same performance (0.901) and better than  
 383 other ensemble methods except Stack\_LR. The performances of all ML models described in Table 5 are  
 384 depicted in Figure 9 (a), (b), and (c) in terms of MAE, RMSE and  $R^2$ , respectively.

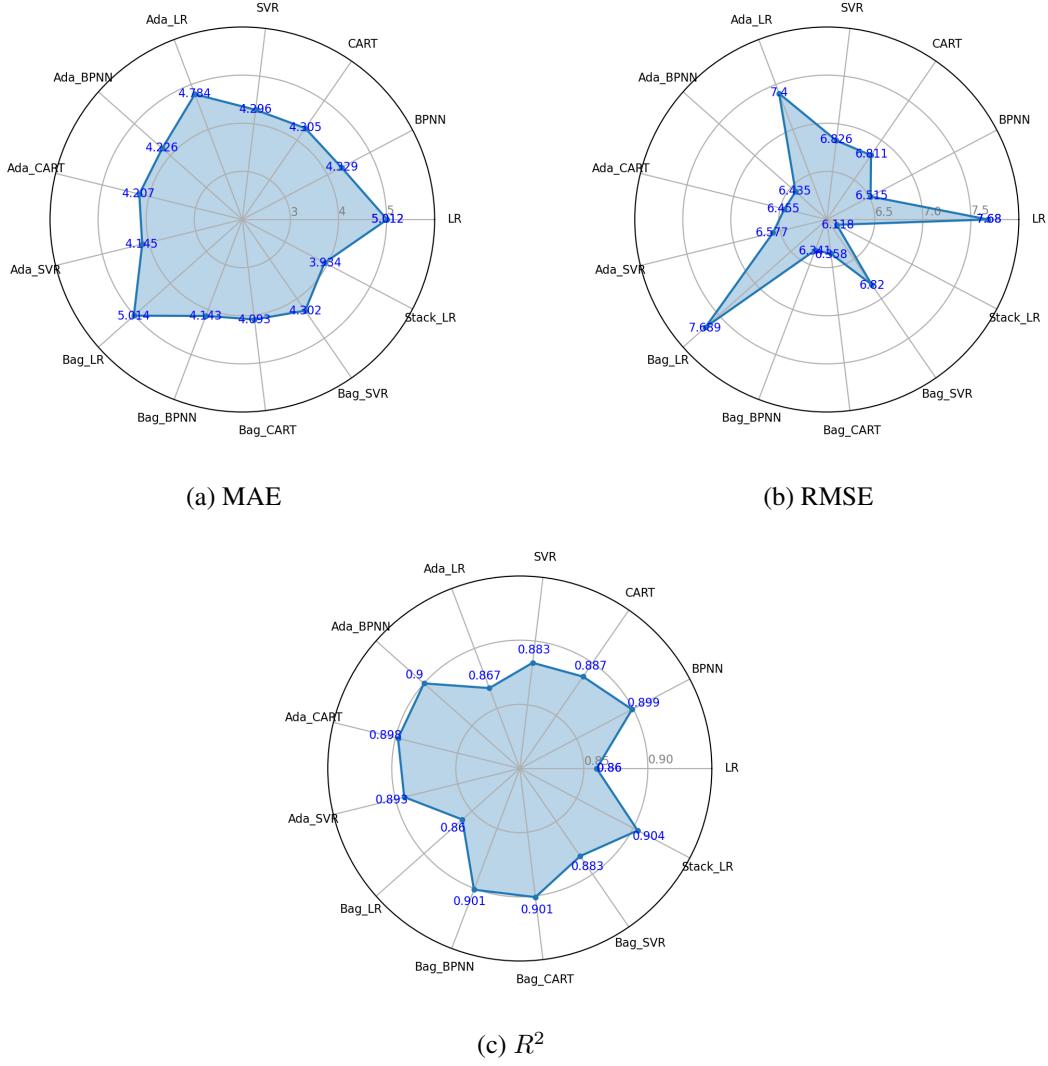


Figure 9: Average prediction performance of 10-fold cross-validation on all machine learning models for predicting self-healing ability of ECC

385 Overall, all models can reduce noticeably the error values and increase the prediction accuracy compared  
 386 with LR, except Bag\_LR. Among the models boosted by AdaBoost, Ada\_SVR performed the best with the  
 387 lowest MAE value, whereas Ada\_BPNN performed the best on RMSE value showing the highest  $R^2$  value.  
 388 In case of bagging, both Bag\_CART and Bag\_BPNN performed better in terms MAE, RMSE and  $R^2$  than  
 389 those of the corresponding models boosted by AdaBoost. However, Bag\_LR showed a poor performance  
 390 compared to LR on the MAE and RMSE values. For a better comparison among the ensemble methods used,  
 391 the performance results between the ensemble models and their corresponding individual (or benchmark)  
 392 models are indicated in Table 6. The results indicate that most ensemble methods improved the performance  
 393 of individual models. For example, the MAE and RMSE values of BPNN after bagging reduced by 4.3%  
 394 and 2.7%, respectively, and its  $R^2$  was much higher than that of the individual BPNN model. Among all the

395 ensemble methods studied, stacking showed the best improvement on all performance measures.

Table 6: Performance deviation of ensemble models from benchmark models on self-healing of ECC

Benchmark	Model	MAE	RMSE	$R^2$	Benchmark	Model	MAE	RMSE	$R^2$
		<i>Dev(%)</i>					<i>Dev(%)</i>		
LR	Ada_LR	-4.6	-3.6	0.8	LR	Bag_LR	0.0	0.1	0.0
BPNN	Ada_BPNN	-2.4	-1.2	0.1	BPNN	Bag_BPNN	-4.3	-2.7	0.2
CART	Ada_CART	-2.3	-5.2	1.2	CART	Bag_CART	-4.9	-6.6	1.6
SVR	Ada_SVR	-3.5	-3.6	1.1	SVR	Bag_SVR	0.1	-0.1	0.0
Ada_LR	Stack_LR	-17.8	-17.3	4.3	Bag_LR	Stack_LR	-21.5	-20.4	5.1

396 However, the results showed that the effectiveness of ensemble methods on individual models varied.  
 397 For instance, bagging method enhanced the performance of BPNN and CART substantially, but not for both  
 398 LR and SVR models. On the other hand, the AdaBoost method brought a considerable improvement for  
 399 LR and SVR models. To improve the performance accuracy, researchers should employ different ensemble  
 400 methods to compare their effectiveness on different ML models.

## 401 5.2 Prediction performance comparison

402 To reveal the accuracy of the proposed ML models in self-healing prediction, the comparison of observed  
 403 crack widths of ECC after self healing with predicted crack widths are shown in Figures 10 to 13. Figure  
 404 10 a shows the observed crack widths compared with the crack widths predicted by different individual  
 405 ML models. Figures 10 b-e show the variations between the observed and the crack widths predicted by  
 406 each individual ML model corresponding to their initial crack widths before self-healing. In other words,  
 407 the prediction performance of models in a particular range of crack widths can be revealed. It should be  
 408 noted that the horizontal line located at the vertical coordinate of zero ( $y = 0$ ) is considered as the target  
 409 line [25, 28]. Generally, the smaller the variation (i.e. closer to the target line), the better the self-healing  
 410 prediction, which means the smaller or even no variation between the observed and the predicted crack  
 411 widths after self-healing.

412 As shown in Figure 10, the SVR model generally exhibited better prediction results than other individual  
 413 models, while the LR model is the worst showing substantial deviation from the target line (denoting relative  
 414 large differences between the observed and the predicted crack widths). For the initial crack widths less than  
 415 20  $\mu m$  and over 100  $\mu m$  before self-healing, the variations between the observed and the ones predicted by  
 416 the SVR model are smaller than other individual ML models. The corresponding MAE values are 1.357 and  
 417 2.724. However, for the crack widths between 20 and 60  $\mu m$ , the CART model performed the best with the  
 418 lowest MAE of 5.045, while the BPNN model has the lowest MAE of 9.565 for the crack widths between  
 419 60 and 100  $\mu m$ . It seems that the choice of ML models may depend on the initial crack widths. However, in  
 420 terms of overall accuracy among the individual models, SVR performed the best, followed by CART, BPNN  
 421 and LR. This is consistent with the results shown in Table 5.

422 The performance of ensemble methods using Bagging and Adaboost are shown in Figures 11 and 12. In  
 423 general, the ensemble models bag\_CART and Ada\_CART exhibited lower variations in self healing results  
 424 compared to other ensemble models. In particular, the MAE values of bag\_CART and Ada\_CART for crack  
 425 widths between 20 and 60  $\mu m$  are 5.000 and 5.037, respectively. These values are smaller than that of CART  
 426 (5.045) as shown in Figure 10 e. However, the variations among BPNN, Ada\_BPNN and bag\_BPNN are not  
 427 significant. Similar variations can be found when comparing SVR with Ada\_SVR and bag\_SVR.

428 After stacking, the error variations shown in Figure 13 are much reduced when compared those shown  
 429 in Figures 10 to 12. More specifically, the MAE of stack\_LR for cracks widths less than 20, between 20 to  
 430 60, between 60 to 100 and over 100  $\mu m$  are 1.361, 4.932, 9.789 and 3.177, respectively. These MAE values

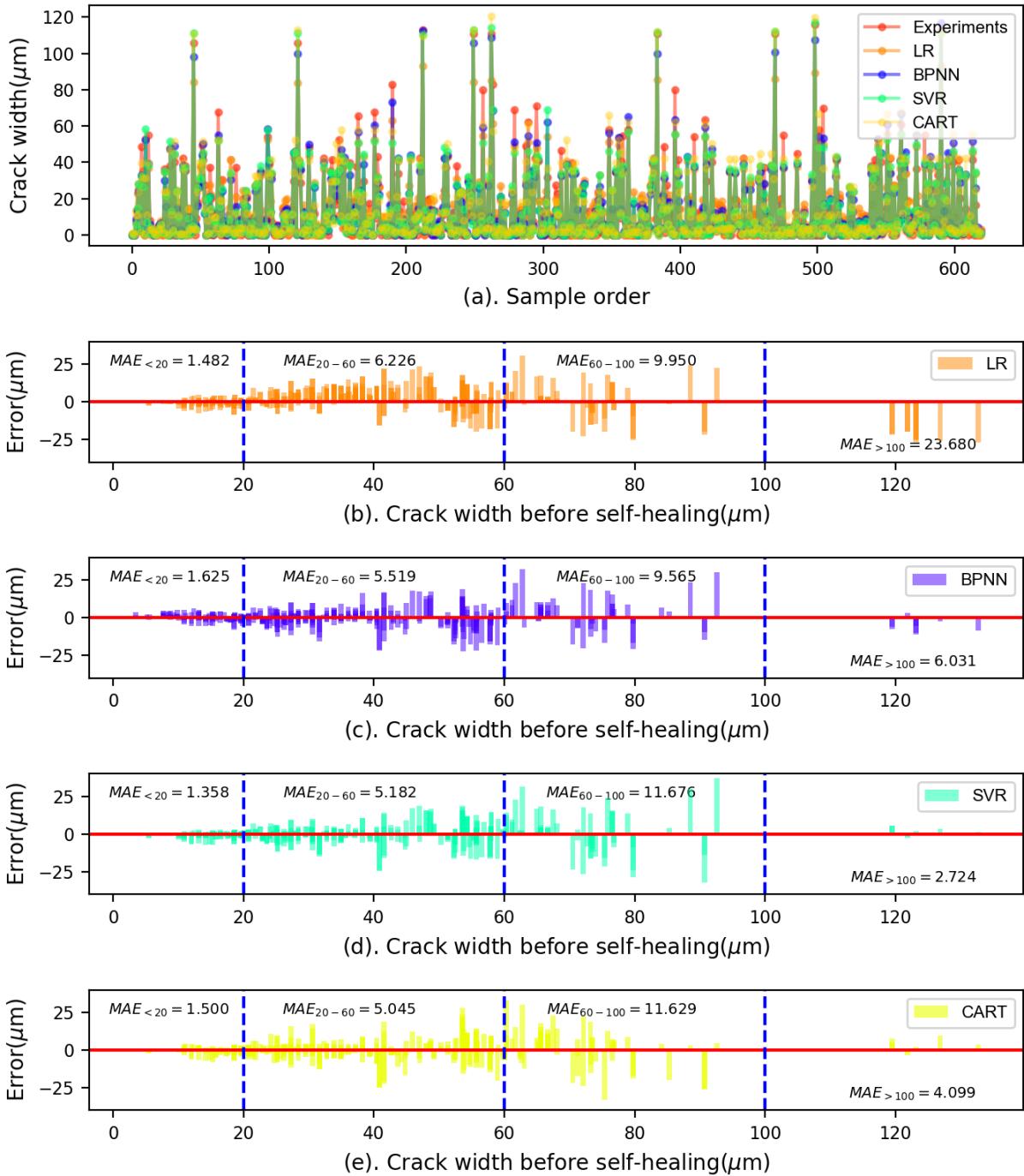


Figure 10: Comparison of observed crack widths of ECC after self healing with crack widths predicted by individual models

431 are the lowest among all the ML models studied. Based on the results, it can be concluded that the stack\_LR  
 432 model performed the best.

433 It is known that smaller crack width is favourable for autogenous healing in concrete [64, 65] as small  
 434 cracks consume less repair products to complete self-healing [66]. However, larger crack width will not heal  
 435 completely or just heal partially. As shown in Figures b in 10 to 13, the variations between the observed and

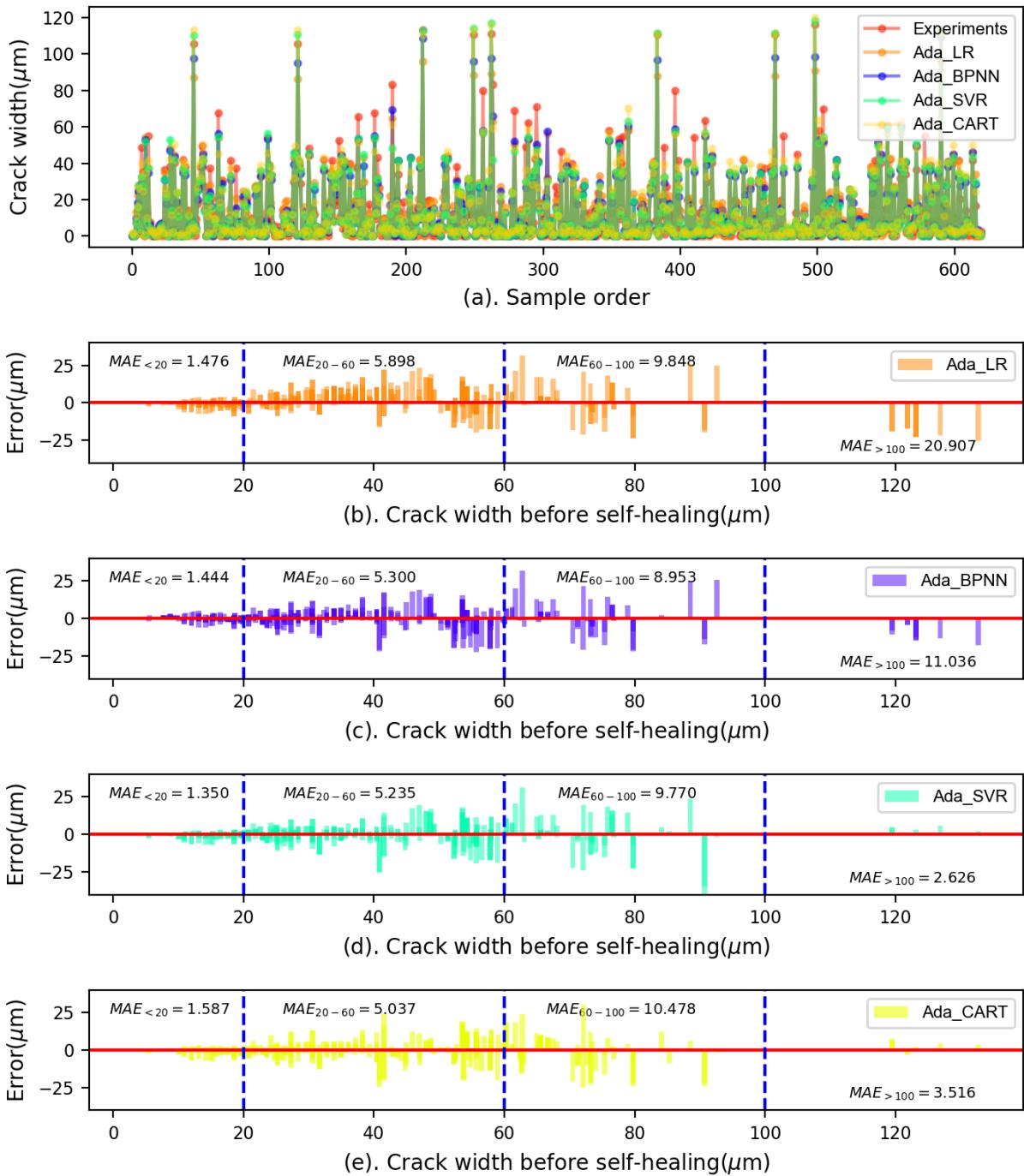


Figure 11: Comparison of observed crack widths of ECC after self healing with crack widths predicted by bagging ensemble models

predicted results for the LR, bag\_LR and Ada\_LR models increased with the increase of crack width. For the crack width below 20  $\mu\text{m}$ , the MAE values were less than 1.5, which is much lower than those for crack widths between 20 and 60  $\mu\text{m}$  (i.e. 6.23) and between 60 and 100  $\mu\text{m}$  (around 10). Similar trends were observed in other models but with smaller variations. Specifically, for crack width over 100  $\mu\text{m}$ , the LR, bag\_LR and Ada\_LR models showed much higher variations. Their MAE values are over 20 and higher than

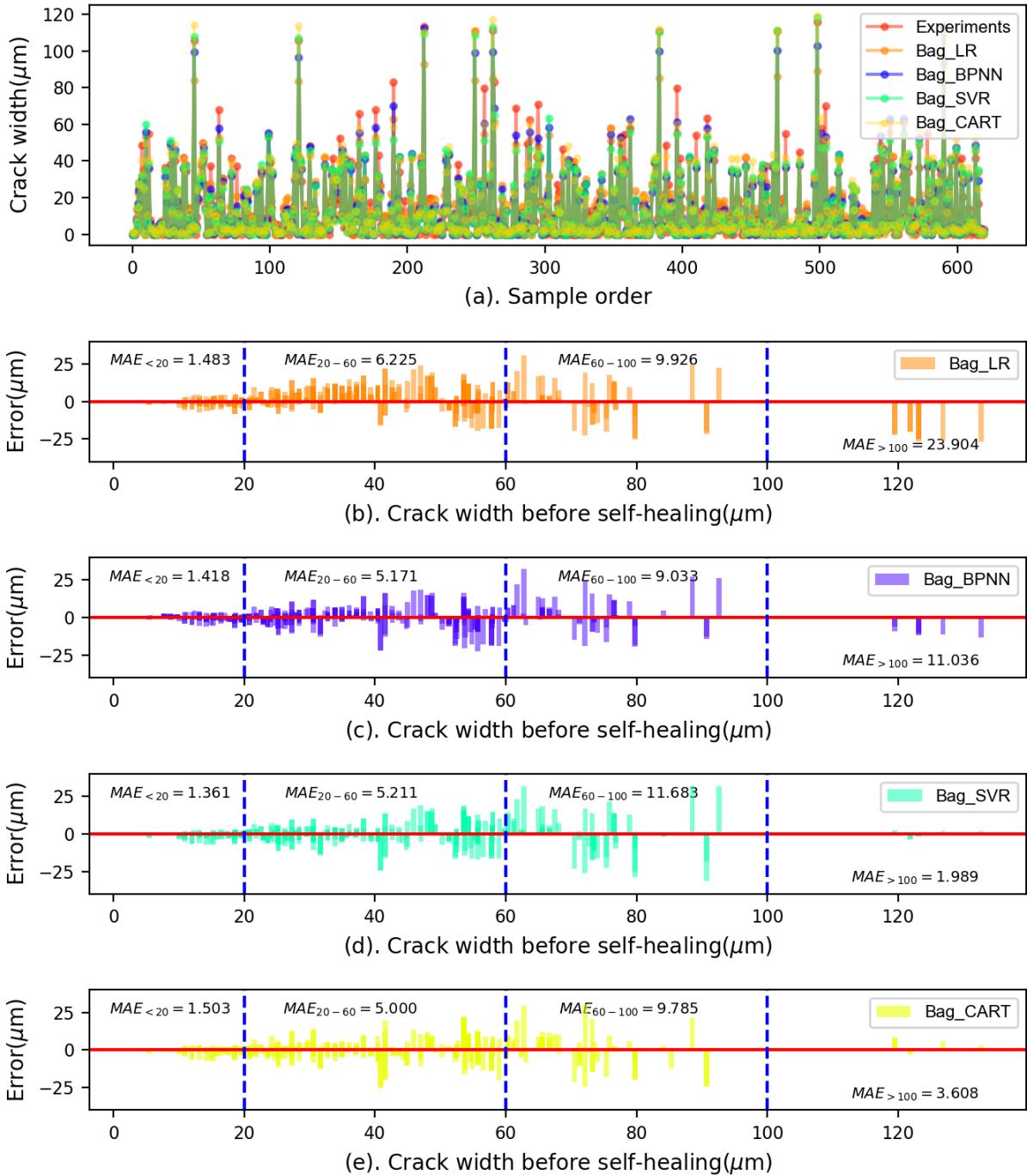


Figure 12: Comparison of observed crack widths of ECC after self healing with crack widths predicted by adaBoost ensemble models

441 other ML models with the MAE less than 10.

442 A box plot as shown in Figure 14 is created to show the distribution of RMSE results of each ML model  
 443 based on the ten-fold cross validation. The RMSE values were calculated based on the differences between  
 444 the predicted and observed crack widths. Box plot is a statistical tool that is used to depicting numerical  
 445 data through their quartiles including maximum, minimum, median values of a dataset [67, 68].The medium

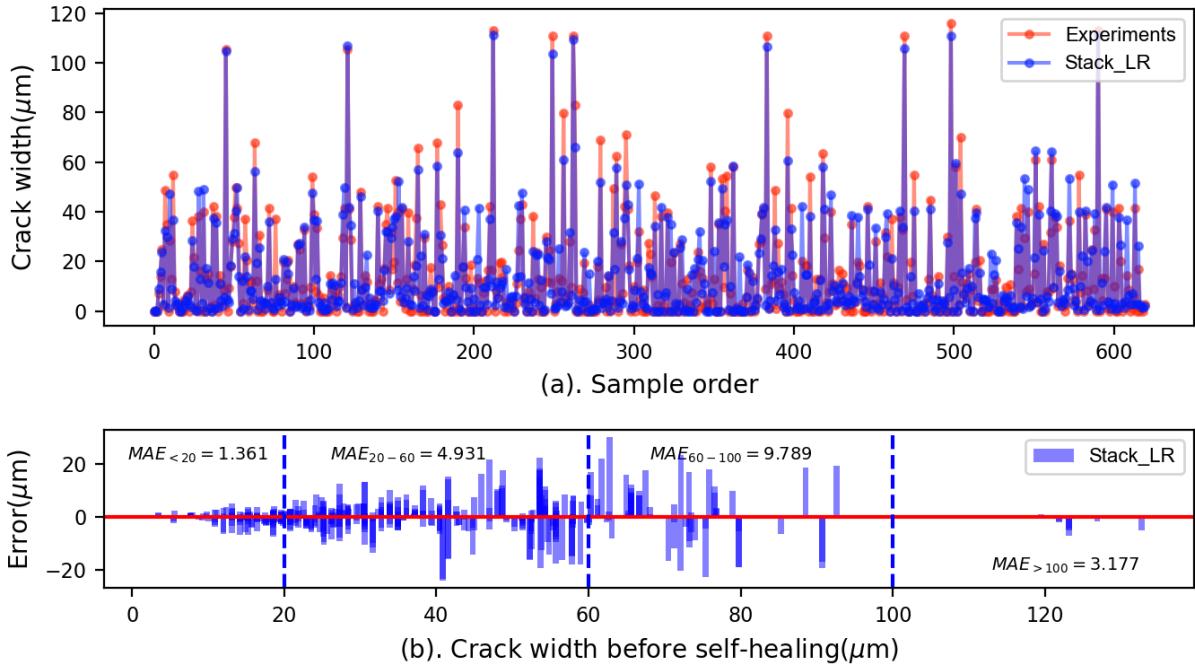


Figure 13: Comparison of observed crack widths of ECC after self healing with crack widths predicted by stacking ensemble models

446 value is shown as the red line within the box. The interquartile range (IQR) in each box covers the 50% (the  
 447 lower 25% to the upper 75% quartiles) of the RMSE data point, while the whiskers drawn up and down to  
 448 the maximum and minimum values represent 1.5 times the IQR from the RMSE data. All other points out of  
 449 the whiskers range are outliers and shown as red dots. A mean value of RMSE equals to zero would indicate  
 450 that the predictions perfectly fit the observed data. However, this is almost never achieved in practice [69].  
 451 In general, the lower RMSE value the better the prediction performance of a model.

452 Assessment of the box plot revealed that the stack\_LR model outperforms all other models because of  
 453 its shortest IQR length and smallest RMSE values as shown in Figure 5. In contrast, the LR and bag\_LR  
 454 models have the longest IQR length and largest RMSE values, thereby suggesting that the LR model and its  
 455 ensemble methods are of low accuracy. Among the individual models, BPNN has the lowest RMSE, while  
 456 SVR has the shortest IQR length but with three outliers (out of ten data points). In general, BPNN gave the  
 457 most stable performance showing reasonable low RMSE values with short IQR length.

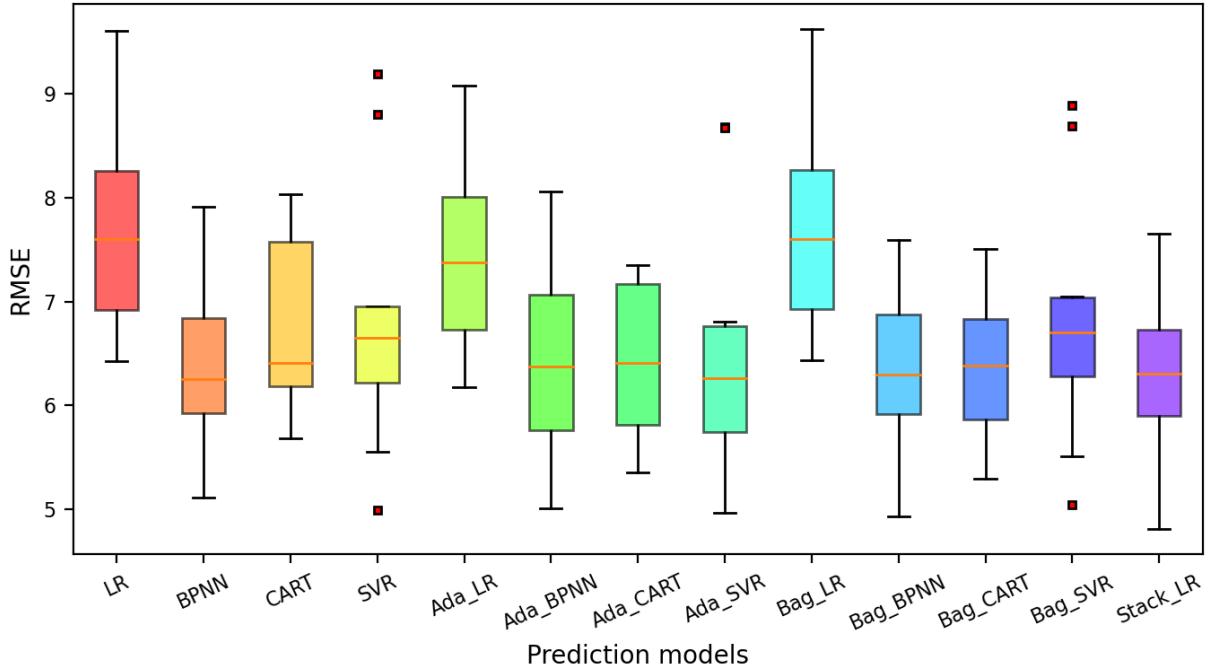


Figure 14: Ten-fold cross validation of RMSE by proposed ML models in prediction of self-healing ability of ECC

## 458 6 Conclusions

459 In this study, several individual and ensemble ML models were proposed to predict the self-healing ability of  
 460 ECC. All the models were trained and validated based on the experimental results from nine ECC mixtures.  
 461 Based on the results, the following conclusions can be drawn.

- 462 1. Among of the individual ML model studies, the BPNN model performed the best in terms of RMSE  
 463 and  $R^2$ .
- 464 2. All ensemble methods can generally improve the prediction accuracy of individual methods, however  
 465 the improvement varies. It is found that Bagging method mainly enhanced the performance of BPNN  
 466 and CART whereas AdaBoost method brought a considerable improvement for LR and SVR models.
- 467 3. Among all the ML models studied, the Stack\_LR model demonstrated great prediction on self-healing  
 468 of ECC and performed the best on MAE, RMSE and  $R^2$  results. The assessment of the box plot also  
 469 revealed that the stackLR model outperforms all other models because of its shortest IQR length and  
 470 smallest RMSE values.
- 471 4. For the initial crack widths less than  $60 \mu m$ , the variations shown in SVR model are smaller than those  
 472 presented in other models. However, the CART model showed smaller variations for the crack widths  
 473 between  $60$  and  $100 \mu m$  compared to the SVR and BPNN models. For crack widths larger than  $100$   
 $\mu m$ , the SVR model performed the best showing the smallest variations.
- 475 5. The computational results indicate that the individual and ensemble methods can be used to predict  
 476 the self-healing ability of ECC. However, how to choose an appropriate base learner and ensemble  
 477 method is critical. To improve the performance accuracy, researchers should employ different ensem-

478       ble methods to compare their effectiveness with different ML models. The proposed individual and  
479       ensemble ML models can be used to predict the self-healing properties of ECC.

480       6. Future investigation and experimentation should be carried to extend the training dataset to include  
481       various crack width distributions and diverse influencing factors such as components, W/MC rate etc..  
482       In addition, more research should be undertaken to optimise parameters in ML models and develop a  
483       hybrid model to improve the prediction accuracy.

## 484       References

- 485       [1] Diane Gardner, Robert Lark, Tony Jefferson, and Robert Davies. A survey on problems encountered  
486       in current concrete construction and the potential benefits of self-healing cementitious materials. *Case  
487       studies in construction materials*, 8:238–247, 2018.
- 488       [2] E Cailleux and V Pollet. Investigations on the development of self-healing properties in protective  
489       coatings for concrete and repair mortars. In *Proceedings of the 2nd International Conference on Self-  
490       Healing Materials, Chicago, IL, USA*, volume 28, 2009.
- 491       [3] Ahmed Ramadan Suleiman and Moncef L Nehdi. Modeling self-healing of concrete using hybrid  
492       genetic algorithm–artificial neural network. *Materials*, 10(2):135, 2017.
- 493       [4] Waiching Tang, Omid Kardani, and Hongzhi Cui. Robust evaluation of self-healing efficiency in ce-  
494       mentitious materials—a review. *Construction and Building Materials*, 81:233–247, 2015.
- 495       [5] Carola Edvardsen. Water permeability and autogenous healing of cracks in concrete. *Materials Journal*,  
496       96(4):448–454, 1999.
- 497       [6] Waiching Tang, Guangwei Chen, and Shanyong Wang. Self-healing capability of ecc incorporating  
498       with different mineral additives - a review. In Changwen Miao, editor, *Proceedings of the 3rd Interna-  
499       tional RILEM Conference on Microstructure Related Durability of Cementitious Composites*, page 668,  
500       2016.
- 501       [7] Stefan Jacobsen, Jacques Marchand, and Hugues Hornain. Sem observations of the microstructure of  
502       frost deteriorated and self-healed concretes. *Cement and Concrete Research*, 25(8):1781–1790, 1995.
- 503       [8] Hans-Wolf Reinhardt and Martin Jooss. Permeability and self-healing of cracked concrete as a function  
504       of temperature and crack width. *Cement and concrete research*, 33(7):981–985, 2003.
- 505       [9] Mustafa Şahmaran and İ Özgür Yaman. Influence of transverse crack width on reinforcement corrosion  
506       initiation and propagation in mortar beams. *Canadian Journal of Civil Engineering*, 35(3):236–245,  
507       2008.
- 508       [10] CA Clear. The effects of autogenous healing upon the leakage of water through cracks in concrete.  
509       Technical report, 1985.
- 510       [11] Mustafa Sahmaran, Gurkan Yildirim, and Tahir K Erdem. Self-healing capability of cementitious  
511       composites incorporating different supplementary cementitious materials. *Cement and Concrete Com-  
512       posites*, 35(1):89–101, 2013.
- 513       [12] Toshiro Kamada and Victor C Li. The effects of surface preparation on the fracture behavior of  
514       ecc/concrete repair system. *Cement and Concrete Composites*, 22(6):423–431, 2000.

- 515 [13] Victor C Li and Tetsushi Kanda. Innovations forum: engineered cementitious composites for structural  
516 applications. *Journal of Materials in Civil Engineering*, 10(2):66–69, 1998.
- 517 [14] Erdoğan Özbay, Mustafa Şahmaran, Mohamed Lachemi, and Hasan Erhan Yücel. Self-healing of  
518 microcracks in high-volume fly-ash-incorporated engineered cementitious composites. *ACI Materials  
519 Journal*, 110(1), 2013.
- 520 [15] Min Wu, Björn Johannesson, and Mette Geiker. A review: Self-healing in cementitious materials and  
521 engineered cementitious composite as a self-healing material. *Construction and Building Materials*,  
522 28(1):571–583, 2012.
- 523 [16] Haoliang Huang, Guang Ye, and Denis Damidot. Characterization and quantification of self-healing  
524 behaviors of microcracks due to further hydration in cement paste. *Cement and Concrete Research*, 52:  
525 71–81, 2013.
- 526 [17] Ahmed R Suleiman, Andrew J Nelson, and Moncef L Nehdi. Visualization and quantification of crack  
527 self-healing in cement-based materials incorporating different minerals. *Cement and Concrete Com-  
528 posites*, 103:49–58, 2019.
- 529 [18] J Zhou, S Qian, MG Sierra Beltran, G Ye, E Schlangen, and K van Breugel. Developing engineered  
530 cementitious composite with local materials. In *International conference on microstructure related  
531 durability of cementitious composites, Nanjing, China*, 2008.
- 532 [19] Victor C Li and En-Hua Yang. Self healing in concrete materials. In *Self healing materials*, pages  
533 161–193. Springer, 2007.
- 534 [20] Zhigang Zhang, Shunzhi Qian, and Hui Ma. Investigating mechanical properties and self-healing be-  
535 havior of micro-cracked ecc with different volume of fly ash. *Construction and Building Materials*, 52:  
536 17–23, 2014.
- 537 [21] Gürkan Yıldırım, Mustafa Sahmaran, and Hemn Unis Ahmed. Influence of hydrated lime addition on  
538 the self-healing capability of high-volume fly ash incorporated cementitious composites. *Journal of  
539 Materials in Civil Engineering*, 27(6):04014187, 2014.
- 540 [22] Y Yang, M Lepech, and Victor C Li. Self-healing of ecc under cyclic wetting and drying. 2005.
- 541 [23] Mustafa Sahmaran, Mo Li, and Victor C Li. Transport properties of engineered cementitious compos-  
542 ites under chloride exposure. *ACI Materials Journal*, 104(6):604, 2007.
- 543 [24] SZ Qian, J Zhou, and E Schlangen. Influence of curing condition and precracking time on the self-  
544 healing behavior of engineered cementitious composites. *Cement and concrete composites*, 32(9):  
545 686–693, 2010.
- 546 [25] Marai M Alshihri, Ahmed M Azmy, and Mousa S El-Bisy. Neural networks for predicting compressive  
547 strength of structural light weight concrete. *Construction and Building Materials*, 23(6):2214–2219,  
548 2009.
- 549 [26] Uwe Reuter, Ahmad Sultan, and Dirk S Reischl. A comparative study of machine learning approaches  
550 for modeling concrete failure surfaces. *Advances in Engineering Software*, 116:67–79, 2018.
- 551 [27] Siamak Safarzadegan Gilan, Hamed Bahrami Jovein, and Ali Akbar Ramezanianpour. Hybrid sup-  
552 port vector regression–particle swarm optimization for prediction of compressive strength and rcpt of  
553 concretes containing metakaolin. *Construction and Building Materials*, 34:321–329, 2012.

- 554 [28] Fei Yan, Zhibin Lin, Xingyu Wang, Fardad Azarmi, and Konstantin Sobolev. Evaluation and prediction  
555 of bond strength of gfrp-bar reinforced concrete using artificial neural network optimized with genetic  
556 algorithm. *Composite Structures*, 161:441–452, 2017.
- 557 [29] Zaher Mundher Yaseen, Ravinesh C Deo, Ameer Hilal, Abbas M Abd, Laura Cornejo Bueno, Sancho  
558 Salcedo-Sanz, and Moncef L Nehdi. Predicting compressive strength of lightweight foamed concrete  
559 using extreme learning machine model. *Advances in Engineering Software*, 115:112–125, 2018.
- 560 [30] Kezhen Yan and Caijun Shi. Prediction of elastic modulus of normal and high strength concrete by  
561 support vector machine. *Construction and Building Materials*, 24(8):1479–1485, 2010.
- 562 [31] Jui-Sheng Chou, Chih-Fong Tsai, Anh-Duc Pham, and Yu-Hsin Lu. Machine learning in concrete  
563 strength simulations: Multi-nation data analytics. *Construction and Building Materials*, 73:771–780,  
564 2014.
- 565 [32] Jafar Sobhani, Meysam Najimi, Ali Reza Pourkhorshidi, and Tayebeh Parhizkar. Prediction of the  
566 compressive strength of no-slump concrete: A comparative study of regression, neural network and  
567 anfis models. *Construction and Building Materials*, 24(5):709–718, 2010.
- 568 [33] Behzad Abounia Omran, Qian Chen, and Ruoyu Jin. Comparison of data mining techniques for pre-  
569 dicting compressive strength of environmentally friendly concrete. *Journal of Computing in Civil  
570 Engineering*, 30(6):04016029, 2016.
- 571 [34] Luthfi Muhammad Mauludin and Chahmi Oucif. Modeling of self-healing concrete: A review. *Jour-  
572 nal of Applied and Computational Mechanics*, 5(Special Issue: Computational Methods for Material  
573 Failure):526–539, 2019.
- 574 [35] Wenjun Wang, Nicolette G Moreau, Yingfang Yuan, Paul R Race, and Wei Pang. Towards machine  
575 learning approaches for predicting the self-healing efficiency of materials. *Computational Materials  
576 Science*, 168:180–187, 2019.
- 577 [36] M Chaitanya, P Manikandan, V Prem Kumar, S Elavenil, and V Vasugi. Prediction of self-healing char-  
578 acteristics of ggbs admixed concrete using artificial neural network. In *Journal of Physics: Conference  
579 Series*, volume 1716, page 012019. IOP Publishing, 2020.
- 580 [37] Xiaoying Zhuang and Shuai Zhou. The prediction of self-healing capacity of bacteria-based concrete  
581 using machine learning approaches. *Computers, Materials & Continua* 59 (2019), Nr. 1, 2019.
- 582 [38] AS 3972-2010 General purpose and blended cements. URL [https://infostore.saiglobal.  
583 com/en-au/standards/as-3972-2010-122323\\_saig\\_as\\_as\\_268436/](https://infostore.saiglobal.com/en-au/standards/as-3972-2010-122323_saig_as_as_268436/).
- 584 [39] Guangwei Chen. *Repeatability of self-healing in ECC with various mineral admixtures*. PhD thesis,  
585 School of Architecture and Built Environment, University of Newcastle, Australia, 2021.
- 586 [40] John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. *Applied linear  
587 statistical models*, volume 4. Irwin Chicago, 1996.
- 588 [41] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297,  
589 1995.
- 590 [42] Vladimir Naumovich Vapnik. An overview of statistical learning theory. *IEEE transactions on neural  
591 networks*, 10(5):988–999, 1999.
- 592 [43] Xu Juncai, Ren Qingwen, and Shen Zhenzhong. Prediction of the strength of concrete radiation shield-  
593 ing based on ls-svm. *Annals of nuclear energy*, 85:296–300, 2015.

- 594 [44] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural*  
595 *processing letters*, 9(3):293–300, 1999.
- 596 [45] Yankun Li, Xueguang Shao, and Wensheng Cai. A consensus least squares support vector regression  
597 (ls-svr) for analysis of near-infrared spectra of plant samples. *Talanta*, 72(1):217–222, 2007.
- 598 [46] P Yuvaraj, A Ramachandra Murthy, Nagesh R Iyer, SK Sekar, and Pijush Samui. Support vector  
599 regression based models to predict fracture characteristics of high strength and ultra high strength  
600 concrete beams. *Engineering fracture mechanics*, 98:29–43, 2013.
- 601 [47] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and comput-  
602 ing*, 14(3):199–222, 2004.
- 603 [48] Abhijit Mukherjee and Sudip Nag Biswas. Artificial neural networks in prediction of mechanical  
604 behavior of concrete at high temperature. *Nuclear engineering and design*, 178(1):1–11, 1997.
- 605 [49] Hosein Naderpour, Amir Hossein Rafiean, and Pouyan Fakharian. Compressive strength prediction of  
606 environmentally friendly concrete using artificial neural networks. *Journal of Building Engineering*,  
607 16:213–219, 2018.
- 608 [50] Leo Breiman. *Classification and regression trees*. 2017.
- 609 [51] S Dan and P Colla. Cart: Tree-structured non-parametric data analysis. *Salford Systems, San Diego,  
610 CA*, 1995.
- 611 [52] R Put, C Perrin, F Questier, D Coomans, DL Massart, and Y Vander Heyden. Classification and  
612 regression tree analysis for molecular descriptor selection and retention prediction in chromatographic  
613 quantitative structure–retention relationship studies. *Journal of Chromatography A*, 988(2):261–276,  
614 2003.
- 615 [53] Dimitrios Frosyniotis, Andreas Staftlopatis, and Aristidis Likas. A divide-and-conquer method for  
616 multi-net classifiers. *Pattern Analysis & Applications*, 6(1):32–40, 2003.
- 617 [54] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple  
618 classifier systems*, pages 1–15. Springer, 2000.
- 619 [55] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- 620 [56] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *Icml*, volume 96,  
621 pages 148–156. Citeseer, 1996.
- 622 [57] Sebastian Raschka. Mlxtend: Providing machine learning and data science utilities and extensions to  
623 python’s scientific computing stack. *The Journal of Open Source Software*, 3(24), 2018. doi: 10.21105/  
624 joss.00638. URL <http://joss.theoj.org/papers/10.21105/joss.00638>.
- 625 [58] Joseph Sill, Gábor Takács, Lester Mackey, and David Lin. Feature-weighted linear stacking. *arXiv  
626 preprint arXiv:0911.0460*, 2009.
- 627 [59] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selec-  
628 tion. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- 629 [60] Ofer Dor and Yaoqi Zhou. Achieving 80% ten-fold cross-validated accuracy for secondary structure  
630 prediction by large-scale training. *Proteins: Structure, Function, and Bioinformatics*, 66(4):838–845,  
631 2007.

- 632 [61] Thierry Vidal, Arnaud Castel, and Raoul François. Analyzing crack width to predict corrosion in  
633 reinforced concrete. *Cement and concrete research*, 34(1):165–174, 2004.
- 634 [62] Kejin Wang, Daniel C Jansen, Surendra P Shah, and Alan F Karr. Permeability study of cracked  
635 concrete. *Cement and concrete research*, 27(3):381–393, 1997.
- 636 [63] Jishen Qiu, Han Siang Tan, and En-Hua Yang. Coupled effects of crack width, slag content, and  
637 conditioning alkalinity on autogenous healing of engineered cementitious composites. *Cement and*  
638 *Concrete Composites*, 73:203–212, 2016.
- 639 [64] Emily N. Herbert and Victor C. Li. Self-Healing of Microcracks in Engineered Cementitious Compos-  
640 ites (ECC) Under a Natural Environment. *Materials*, 6:2831–2845, 2013. doi: 10.3390/ma6072831.
- 641 [65] Hezhi Liu, Qian Zhang, Chongshi Gu, Huaizhi Su, and Victor Li. Influence of microcrack self-healing  
642 behavior on the permeability of engineered cementitious composites. *Cement and Concrete Compos-*  
643 *ites*, 82:14–22, 2017.
- 644 [66] Nele De Belie, Elke Gruyaert, Abir Al-Tabbaa, Paola Antonaci, Cornelia Baera, Diana Bajare, Aveline  
645 Darquennes, Robert Davies, Liberato Ferrara, Tony Jefferson, et al. A review of self-healing concrete  
646 for damage management of structures. *Advanced Materials Interfaces*, 5(17):1800074, 2018.
- 647 [67] Woubishet Zewdu Taffese, Esko Sistonen, and Jari Puttonen. Caprm: Carbonation prediction model  
648 for reinforced concrete using machine learning methods. *Construction and Building Materials*, 100:  
649 70–82, 2015.
- 650 [68] Oladimeji B Olalusi and Panagiotis Spyridis. Machine learning-based models for the concrete breakout  
651 capacity prediction of single anchors in shear. *Advances in Engineering Software*, 147:102832, 2020.
- 652 [69] Wikipedia contributors. Root-mean-square deviation — Wikipedia, the free encyclopedia, 2021.  
653 URL [https://en.wikipedia.org/wiki/Root-mean-square\\_deviation](https://en.wikipedia.org/wiki/Root-mean-square_deviation). [Online;  
654 accessed 28-August-2021].