Construction and Building Materials

A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis

--Manuscript Draft--

Manuscript Number:	CONBUILDMAT-D-22-09224R1
Article Type:	Research Paper
Keywords:	Multifractal analysis; load-level assessment; beams and slabs; Machine Learning; score analysis
Corresponding Author:	Jack Banahene Osei, PhD Kwame Nkrumah University of Science and Technology Kumasi, GHANA
First Author:	Jack Banahene Osei, PhD
Order of Authors:	Jack Banahene Osei, PhD
	Mark Adom-Asamoah, PhD
	Jones Owusu Twumasi, PhD
	Peter Andras, PhD
	Hexin Zhang
Abstract:	This paper presents a machine learning model for load-level estimation for shear-critical reinforced concrete (RC) beams and slabs using multifractal features of their characteristic crack patterns to automate and provide well-informed decisions for RC damage assessment. Multifractal analysis was conducted on a database of 508 images, of which critical features were extracted from the singularity and generalized dimension spectra. These features are used as predictors for the load-level estimation model. The extreme gradient boosting algorithm yielded the best performance among the four machine learning models considered. The mean of the predicted-to-true ratio for the developed model was 1.04 with a coefficient of variation of 0.27. Upon applying Shapley additive explanations, the fractal dimension, information dimension, correlation dimension and the area under the left branch of the singularity spectrum were the critical features influencing load-level estimation. The proposed model can be useful to RC building inspectors.
Suggested Reviewers:	Salvatore Salamone, PhD Professor, The University of Texas at Austin salamone@utexas.edu He is an expert in employing fractal analysis for damage assessment of reinforced concrete members
	Gregory Miller, PhD Professor, University of Washington gmiller@u.washington.edu His research revolves around the application of fractal analysis for damage assessment of reinforced concrete members
	Shirley Dyke, PhD Professor, Purdue University sdyke@purdue.edu His research works revolve around the application of computer-vision techniques for bridge inspection
	Christian Koch, PhD Professor, University of Nottingham christian.koch@nottingham.ac.uk He has considerable knowledge on the application of computer vision techniques for concrete and civil infrastructure
	Koichi Maekawa, PhD Professor, The University of Tokyo

	maekawa@concrete.t.u-tokyo.ac.jp He has used several machine learning algorithms for estimating in-service fatigue life assessment of road bridge decks

Department of Civil Engineering,
Kwame Nkrumah University of Science and Technology,
Kumasi, Ghana

Jack Banahene Osei,
Tel: +2330550766883
Email: jobanahene.coe@knust.edu.gh

Dear Editors
Editors-in-Chief
Construction and Building Materials

Manuscript: A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis.

Authors: Jack Banahene Osei, Mark Adom-Asamoah, Jones Owusu-Twumasi, Peter Andras and Hexin Zhang

I have pleasure in submitting to you our paper entitled "A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis" for review and possible publication in the Journal of Construction and Building Materials. This paper presents an automated approach for quantifying the extent of damage a shear-critical reinforced concrete beam or slab would exhibit under loading, by using critical multifractal features of their crack patterns. Further insight on how these critical features influence the failure ratio estimation model is discussed. Model predictions are very much in agreement with observed responses from an image database of the crack patterns of the class of RC beams and slabs considered.

Thank you very much.

Yours faithfully

Jack Banahene Osei

A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis

Jack Banahene Osei[a], Mark Adom-Asamoah[a], Jones Owusu Twumasi[a], Peter Andras[b], Hexin Zhang[b]

[a]Department of Civil Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[b]School of Computing, Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, EH10 5DT, Scotland, UK

Coressponding Author: Jack Banahene osei
Email Address: jobanahene.coe@knust.edu.gh

Abstract

This paper presents a machine learning model for load-level estimation for shear-critical reinforced concrete (RC) beams and slabs using multifractal features of their characteristic crack patterns to automate and provide well-informed decisions for RC damage assessment. Multifractal analysis was conducted on a database of 508 images, of which critical features were extracted from the singularity and generalized dimension spectra. These features are used as predictors for the load-level estimation model. The extreme gradient boosting algorithm yielded the best performance among the four machine learning models considered. The mean of the predicted-to-true ratio for the developed model was 1.04 with a coefficient of variation of 0.27. Upon applying Shapley additive explanations, the fractal dimension, information dimension, correlation dimension and the area under the left branch of the singularity spectrum were the critical features influencing load-level estimation. The proposed model can be useful to RC building inspectors.

Keywords: Multifractal analysis; load-level assessment; beams and slabs; machine learning; score analysis

1. Introduction

The performance characteristics of many civil engineering infrastructure systems play a dominant role in structural safety evaluation [1], as well as public safety [2]. In practice, evaluating the service performance of such systems is typically facilitated by non-destructive techniques. Visual inspection techniques remain one of the most widely used approaches for the non-destructive evaluation of such systems [3]. They are used in many contexts, including but not limited to structural condition monitoring and damage assessment. The results from such techniques usually give a firsthand insight into whether the infrastructure should be repaired or replaced, or an estimate of the remaining life of the system at both local and global levels. For reinforced concrete (RC) structures, the available visual inspection techniques heavily rely on patterns in concrete cracking and propagation (width, length and orientation), spanning a significant period of deterioration. This methodology has been fairly justified since characteristic crack patterns of RC structures can be used as a proxy to ascertain the stress and strain levels induced in the system during deterioration. In other words, they become a valuable piece of information during load-level assessment of reinforced concrete structures. Typically, the damage assessment and structural condition monitoring phase of RC structures as done by visual inspectors is conducted in three stages; (1) using crack detection equipment such as lidars to determine their locations; (2) documenting the damage by capturing images of the cracked regions [4,5] and (3) and determining the internal load levels and damage states of inspected elements. One notable and conventional way to assess the damage of RC structures has been to augment data from crack pattern characterization and analysis, with condition rating systems[6–8]. Nevertheless, with regards to either estimating the residual strength of an RC member/structure (load-level assessment) or categorizing the extent of structural damage (damage assessment), condition rating systems oftentimes result in a qualitative assessment and hence do not necessarily provide building inspectors with the necessary information [7,9]. In particular, guidelines on condition rating of civil infrastructure systems allows for engineering judgment to be used in damage evaluation, hence subjective and highly reliant on the experience of the inspector[1]. With regards to documentation during the assessment phase, visual inspectors do take a considerable amount of time to complete such tasks, and therefore can causes delays. A case in point is the bridge collapse at the Florida International University [10], where although damage documentation was conducted, results were not accessible in a timely manner to aid in collapse prevention and mitigation. Hence, a major drawback of the application of this visual inspection approach has been it's time-consuming nature (damage documentation) amidst subjectivity in making well-informed decisions. To this end, the relevance of developing automated infrastructure inspection methods for load-level and damage assessment of RC structures has presented itself an interesting area of research.

Structural design and industrial guidelines such as ACI [11] , IAEA [6] and AASHTO [12] make available procedures for load-level and damage evaluation of RC components via crack analysis. A real-world application of how crack patterns can be used to predict the strength and stiffness characteristics of RC shear walls that were damaged during an earthquake was conducted by Madani and Dolatshahi [13]. A significant number of research efforts [5,14–22] have been conducted on crack detection and measurement, which is one of the key stages in crack analysis. Nonetheless, the task of using information (width, length, orientation and number of cracks) obtained from crack analysis to correlate the level of damage still remains a challenge with research efforts still at an early stage. In recent times, artificial intelligence-based data-driven techniques keep transforming the field of structural engineering. To this end, automated computer-aided visual inspection approaches have been developed for the identification and characterization of structural damage of RC structures through crack assessment [4,5,20,21,23–27]. These approaches are heavily reliant on two fields: machine learning and computer vision. The fundamental problem of image segmentation (automatically retrieving cracks from images), coming from the computer vision perspective, for RC members has been studied extensively in recent times [28–30]. This has made it possible to extend machine learning algorithms to quantitatively predict the level of damage of many RC structural components. For instance, Ebrahimkhanlou [25] developed a probabilistic graphical model (Bayesian Belief Network) that could visually evaluate the extent of damage of an RC shear wall and also prognosticate the most likely mode of failure for such members. Fatigue life evaluation of bridge deck was presented Fathalla [31] by using an artificial neural network. Davoudi et al. [2,32] employed computer-vision-based inspection methodologies for quantitative damage and load estimation of RC beams and slabs.

The theory of fractals has been extensively applied in the field of structural engineering for performance evaluation and damage assessment of RC components. Farhidzadeh et al.[33] reported that the extent of structural damage of an RC shear wall under reverse-cyclic loading can be quantified from the fractal characteristic of their crack patterns. Experimental validation of how fractal characteristics of surface cracks of RC members can be utilized in damage classification was investigated by Carrillo et al. [34]. Athanasiou et al. [1] and Liu et al [35] have recently developed data-driven machine learning models for damage classification of RC shells using multifractal and fractal analysis respectively.

The present work seeks to extend the application of multifractal analysis of crack patterns in damage evaluation of shear-critical monotonically-loaded simply-supported RC beams and one-way slabs. In order to facilitate this, a database of segmented images of shear-critical RC beams and slabs as compiled by Davoudi et al. [2] is utilized. In particular, this study builds on the work done by Athanasiou et al. [1] that explored the utilization of multifractal features for damage evaluation of RC shells. The singularity spectrum (a parabolic curve, concave in nature) remains the most dominant output of any multifractal analysis. As shown in Athanasiou et al., [1] geometric features of the singularity spectrum can be extracted and utilized as inputs in a machine learning-based damage classification model of RC shells, with significant accuracy. Although four candidate multifractal features (peak, width, and the area under the left and right branch of the singularity spectrum) were used in their approach, which was seemly motivated by trying to reduce the dimensionality of the model, the authors could not exhaust all potential features that can be obtained from the multifractal analysis, which could equally impact the damage evaluation process positively. The primary distinction in the present study is on the identification of the critical multifractal features relating to both geometry and dimensionality of the basic output of multifractal analysis. The secondary distinction is the proposition of a machine learning regression model that utilizes multifractal features for damage evaluation (structural load estimation) of shear-critical simply-supported RC beams and slabs with a monotonic loading protocol, as opposed to the load estimation models developed by Davoudi et al. [2,32] using machine vision. The overall goal motivating this study is to provide an automated model that takes in captured images of RC beams and slabs and can provide a fairly quick estimation of the extent of damage before sophisticated and computationally expensive assessment techniques can be utilized for rigorous cracking assessment of RC structural components.

2. Overview of Fractal Analysis

Fractal theory [36] since its inception in the 1970s has been successful applied in many fields including astrophysics [37], financial engineering[38], structural engineering[33,34,39,40], medicine [41,42] and manufacturing [43]. The theory seeks to characterize the geometry of irregular and complex objects occurring in nature that the classical Euclidean geometry may seem non-applicable. As noted by Mandelbrot [36], ‘clouds are not spheres, mountains are not cones, nor does lightning travel in a straight line. There are two main facets of fractal theory;

dimensionality and self-similarity. The dimensionality concept hinges on the fractal geometry of the object. As popularly known in literature, Euclidean geometry reveals that the topological dimension of a point, straight line and plane is 0, 1 and 2 respectively, without any intermediate values. However, fractal geometry permits the use of fractional or fractal dimensions. To illustrate this, consider the crack pattern of an RC beam in Fig. 1 which has an estimated fractal dimension of 1.4.

Fig. 1 Typical crack pattern of an RC beam with fractal characteristics

The self-similarity property of many fractal objects is related to an observation about how the method of construction of such objects at both local and global scales appear to be identical. Crack patterns of many reinforced concrete structures under both cyclic and monotonic loading have been shown to exhibit this self-similar behavior. An illustrative example is the crack surface of a prestressed RC girder as shown in Fig. 2 [40]. It possesses fractal behavior since the crack patterns contain replicas of itself at microscopical and macroscopical scales. In order words, if one zooms in or out the crack surfaces (Fig. 2), the geometrical shape has similar appearance. If there exist more than one replica of this self-similarity charateristics, the considered crack pattern is categorized as a multifractal crack pattern. Other technical background for categorizing a digital image as either having monofractal or multifractal characterisitics is discussed below. Nevertheless, for this particular example, since there exists some form of self-similarity at more than one location, there is reason to believe that the crack patterns have multifractal characteristics.

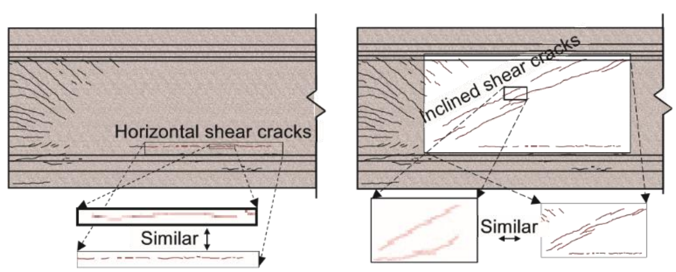


Fig. 2 Self-similarity of RC cracks

2.1 Monofractal Analysis

Several implementation procedures exist for conducting monofractal analysis of images, for fractal dimension determination [44–46]. The box-counting algorithm being the most popular is used in this study. In its abstract form, the fractal analysis seeks to establish the relation between two quantities; the scaling factor, ε, and the number of coverings, $N(\varepsilon)$ of the fractal set, for instance, a digital image profile. Eq. 1 provides the power law relationship that exists between these two quantities.

$$N(\varepsilon) \propto \varepsilon^{-D} \tag{1}$$

where D denotes the fractal dimension. However, for the box-counting algorithm, the scaling factor, ε, is approximated with the size of the box (a) used in discretizing the image pattern. The number of boxes that contains at least an active pixel ($N(a)$) is also used as a proxy for the number of coverings ($N(\varepsilon)$). Linearizing the power law from Eq. 1, the fractal dimension, D, as per the box-counting algorithm, can be computed as:

$$D = \lim_{a \to 0} \frac{\log(N(a))}{\log(1/a)} \tag{2}$$

Alternatively, D can be estimated from the gradient between the number of boxes that contains at least an active pixel, $N(a)$, and the inverse of the box size, a, in the logarithmic space. Fractal dimension, D, depicts the global behavior of fractal sets or digital images through the scaling law presented in Eq. 1, and is the primary output of any monofractal analysis. Monofractal analysis typically do not provide the necessary information for quantifying local fractal characterization. There is a possibility that different images with varying levels of complexities, irregularities and roughness, will yield the same fractal dimension, D, when a monofractal anlysis is conducted [43,47]. In such situation, the utilization of a generalized fractal analysis, known as multifractal analysis could be employed to gain much more insight.

2.2 Multfractal Analysis

Multifractal analysis seeks to provide a detailed local description of the fractal characteristics of a digital image profile. The local pixel density of a particular box, $P_i(a)$, in the digital image is first computed as given in Eq. 3.

$$P_i(a) = \frac{N_i(a)}{\sum_i^{N(a)} N_i(a)} \tag{3}$$

where $N_i(a)$ is the number of pixels in the *i*th box. In the special case where the image in question is a crack pattern of an RC element, $P_i(a)$ denotes the crack density. As an illustrative example, consider the crack pattern of a beam shown in Fig. 1.

Using four candidate boxes, the spatial distribution of the pixel intensities (crack density $P_i(a)$) for the above RC beam is presented in Fig. 3. Evidently, the spatial crack density distribution seems to converge to the original crack pattern of the beam when the size of the box decreases.

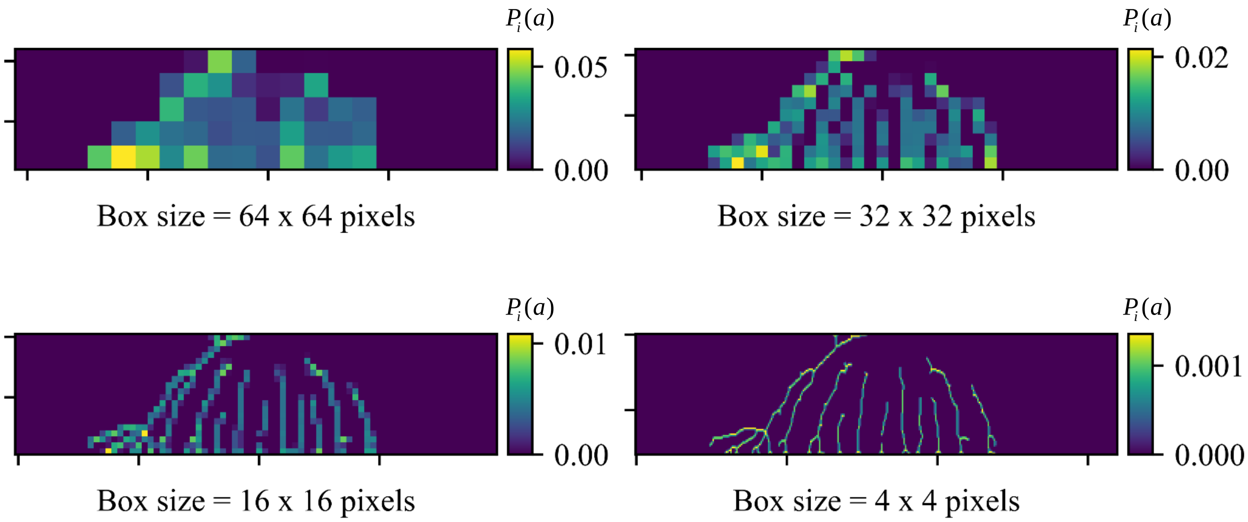


Fig. 3 Spatial distribution of pixel intensities for the crack pattern of an RC beam (color printed)

It turns out that, a similar power law exists between how the pixel density $P_i(a)$ scales, and size of the box a (see Eq. 4).

$$P_i(a) \propto a^{\alpha_i} \tag{4}$$

where α_i is the singularity exponent, depicting the local scaling/fractal behaviour for the ith box. In other words, each box characterized by $P_i(a)$ will have its own singularity exponent α_i. For an infinitesimally small difference $\Delta\alpha$, the number of boxes $N(\alpha)$ for which their singularity exponents fall within the closed interval $[\alpha, \alpha + \Delta\alpha]$ is obtained, and follows a power law with the box size (a), similar to that of Eq. 1.

$$N(\alpha) \propto a^{-f(\alpha)} \tag{5}$$

where $f(\alpha)$ is the fractal dimension of the boxes with the same local scaling α. An $\alpha - f(\alpha)$ plot is commonly called the singularity spectrum is typically used to summarize the output of any multifractal analysis study. The $f(\alpha)$ can be computed from Eq. 6 as:

$$f(\alpha) = \lim_{a \to 0} \frac{\log(N(\alpha))}{\log(1/a)} \tag{6}$$

Traditionally, Legendre Transformation as suggested by Hasley et al. [48] is used to estimate $f(\alpha)$. Nevertheless, a direct numerical approach developed by Chhabra and Jensen [49] is used in this study. It begins with obtaining distorted versions of the spatial distribution of the pixels using the following exponential mapping:

$$P_i(a) \to P_i^q(a) \tag{7}$$

where q it is typically known as the distortion parameter or the order of the probability moment [50]. For a range of values of q ([-5,+5] as recommended by Ebrahimkhanlou et al. [51] for shear-critical RC elements), a normalized form of $P_i^q(a)$ is computed.

$$\mu_i(q,a) = \frac{P_i^q(a)}{\sum_{i=1}^{N(a)} P_i^q(a)} \tag{8}$$

For a given value of q, the singularity exponent $\alpha(q)$ and its corresponding fractal dimension $f(\alpha(q))$ can then be estimated as:

$$\alpha(q) = \lim_{a \to 0} \frac{\sum_{i=1}^{N(a)} \mu_i(q,a)\log(P_i^q(a))}{\log(a)} \tag{9}$$

$$f(\alpha(q)) = \lim_{a \to 0} \frac{\sum_{i=1}^{N(a)} \mu_i(q,a)\log(\mu_i(q,a))}{\log(a)} \tag{10}$$

As already mentioned, a plot of the set of values of α against $f(\alpha)$ for the range of q values, produces the so-called singularity spectrum. Similarly, a $q-\alpha$ plot yields the generalized dimension spectrum. These spectra upon application of multifractal analysis on the crack pattern of the above beam, is shown in Fig. 4 below.

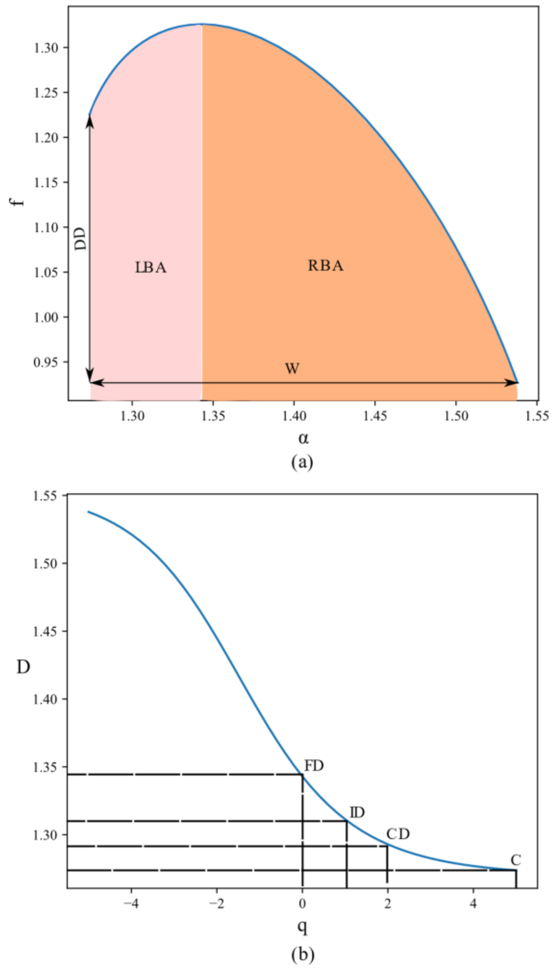


Fig. 4 Key features of (a) the singularity spectrum and (b) generalized dimension spectrum; W: width, FD: fractal dimension, LBA: area under the left branch, RBA: area under the right branch , ID: information dimension, CD: correlation dimension, C: capacity and DD:dimensional difference. (color printed)

3. Extracted features from singularity and generalized dimension spectra

Past research efforts have revealed that specific features that can be extracted from the singularity spectrum of RC shells, can be utilized in structural damage level assessment. In particular, the width (W), area of the left branch (LBA), area of the right branch (RBA) and the peak (FD) of the singularity spectrum have been suggested as critical parameters for damage level identification of RC shells [1] (see Fig. 4a). The geometric width (W) of the singularity spectrum has been deemed to be influential at characterizing RC crack inclination. Generally, the width of the singularity spectrum quantifies the image's heterogeneity. Larger values of the width would usually imply a more severe uneven spatial crack density distribution. In addition, due to the typical asymmetry shape of the singularity spectrum (see Fig. 4a) the area under the left (LBA) and right branch (RBA) of the singularity spectrum has been proven to be key features that influence cracking properties. Also, as noted by Athanasiou et al. [1], the peak of the singularity spectrum (FD) is highly correlated with crack inclination [1].

The fractal dimension (FD) can also be obtained from the generalized dimension spectrum (Fig. 4b) when q = 0. Nevertheless, some well know generalized dimensions (D_q ,i.e, α for a particular q) can be candidate features that can significantly characterize the damage performance of RC elements. Information dimension (ID) is the ordinate of the generalized dimension spectrum when $q = 1$. It characterizes the rate at which information contained in the image profile changes with box size. To this end, the information dimension (ID) is explored in this study. The generalized dimensions D_q, corresponding to q > 1 accentuates the more singular regions (regions with significant cracking behaviour), whereas for q < 1, reflects the regular regions of the RC crack pattern. The correlation dimension (CD) is also used in this study for RC damage assessment. It quantifies correlation for the heterogeneity of a pair of boxes. The generalized dimension corresponding to the maximum q value is usually referred to as the capacity (C) (see Fig. 4b). The capacity reflects segments of the RC crack patterns with low densities ($P_i(a)$). The capacity, C, can also be used as a proxy for heterogeneity since, larger values signify a higher degree of homogeneity within the singular regions. To this end the capacity (C) is also used in this study. Finally, the dimensional difference (DD) defined as the difference between the fractal dimension of the most singular event $f(\alpha_{\min})$ and the most regular event $f(\alpha_{\max})$ is utilized (see Fig. 4). It reflects the frequency ratio or the proportion of the number of regular regions to singular regions. In summary, eight geometric and generalized dimension multifractal features are extracted from crack patterns of selected shear-critical RC beams and slabs for damage assessment; width (W), peak (FD), area of left (LBA) and right (RBA) branch of the singularity spectrum, information dimension (ID), correlation dimension (CD), capacity (C) and dimensional difference (DD).

4. Image database of RC beams and Slabs

In order to develop a reliable model for structural load estimation, the load-level of RC beams and slabs of an existing database was compiled by Davoudi et al.[2] is utilized in this study. It comprises a variety of experimental programs ranging from uniform to monotonic loading of RC beams and one-way slabs without transverse reinforcement. Table 1 presents a summary of the various independent sources of experimental programs that have been aggregated to form the database used in this study.

To this end, a complied database of the multifractal features considered in this study was presented for 508 RC beams and slabs. The eight multifractal features (see section 3.0) served as input features for the estimation model, whereas the load level (LL) served as the output. LL is defined as:

$$LL = V \,/\, V_{failure} \qquad (11)$$

where V and $V_{failure}$ represents the nominal applied shear during loading and at failure, respectively. Pragmatic use of the load level (LL) would be to anticipate the degree to which an RC member has been subjected to a load that would cause failure (an LL of 0.7 would imply that the RC member has been given a load of 70% of what it can sustain (capacity)). Some descriptive statistics of both input and output features is presented in Table 2, whereas Fig. 5 and 6 displays statistical distributions, in particular pairwise relationship between some selected features. Each row and column of the matrix of subplots in Fig. 5 and 6 signifies a single feature. The diagonal plots reveal the univariate marginal distribution of a particular feature, whereas the annotations inserted in the upper half of the off-diagonal plots are used to quantify the correlation between two features. All variables were positively correlated with each other, except the LBA and DD which was negatively correlated. The various forms of generalized dimensions are very highly correlated (see Fig. 5), whereas the other features are fairly correlated (see Fig. 6). In order to obtain more insight into how these features could be used to provide a meaningful estimate of the load-level of shear-critical RC beams and slabs, sophisticated machine learning model implementation were explored as opposed to the basic statistical measures presented in Fig. 5 and 6.

Table 1. Summary of experimental testing programs from which database is compiled.

Reference	#S	#I	Test / Specimen Type	*a/d*	*ρ (%)*	*fc'*
Sneed[51]	8	52	3-point load, beam	2.3	0.55-0.85	18.6-32.4
Murray[52]	8	88	3-point load, beam	2.97-3	1.2-1.3	64.8-74.8
McCain[53]	10	82	3-point load, beam	2.3-2.9	0.63-0.98	22.8-33.8
Sherwood[54]	30	197	3-point load, beam & slab	2.79-3.4	0.3-1.33	29.1-77.3
Quach[55]	1	10	3-point load, deep beam	3.1	0.70	40.0
Yoshida[56]	1	4	3-point load, deep beam	2.9	0.70	31.8
Cao[57]	2	12	3-point load, deep beam	2.8-2.9	0.4-1.5	26.2-28.3
Perkins[58]	6	35	Uniform loading	1.62-3.24	0.98	39-64
Nghiep[59]	3	28	3-point load, haunched beam	3-5.0	1.57-3.1	35.4-59.1
Overall	**69**	**508**	-	**1.1-5.0**	**0.3-3.1**	**18.6-77.3**

Note: #S = number of specimens; #I = number of images; *a/d* = shear span-to-depth ratio; *ρ* = tensile reinforcement ratio; *fc'* = compressive strength.

Table 2. Multifractal features of database of RC beams and slabs

Reference	Statistic	FD	ID	CD	C	LBA	RDA	DD	W
	Minimum	0.79	0.79	0.78	0.75	0.03	0.14	0.44	0.25
Sneed[51]	Mean	1.22	1.21	1.20	1.18	0.04	0.24	0.48	0.26
	Maximum	1.45	1.44	1.44	1.41	0.05	0.30	0.55	0.27
	Minimum	0.38	0.36	0.34	0.19	0.03	0.03	0.02	0.19
Murray[52]	Mean	1.03	1.01	0.99	0.88	0.09	0.15	0.19	0.26
	Maximum	1.38	1.36	1.34	1.28	0.13	0.25	0.29	0.27
	Minimum	0.34	0.33	0.33	0.21	0.01	0.02	0.01	0.11
McCain[53]	Mean	1.05	1.04	1.02	0.94	0.07	0.17	0.29	0.26
	Maximum	1.33	1.31	1.29	1.23	0.09	0.23	0.41	0.27
Sherwood[54]	Minimum	0.36	0.34	0.32	0.19	0.02	0.04	0.03	0.20

	Mean	1.12	1.10	1.08	1.00	0.08	0.18	0.26	0.26
	Maximum	1.46	1.44	1.42	1.32	0.13	0.25	0.48	0.27
Quach[55]	Minimum	0.91	0.90	0.89	0.85	0.04	0.15	0.23	0.25
	Mean	1.37	1.36	1.34	1.27	0.10	0.24	0.26	0.27
	Maximum	1.57	1.55	1.53	1.46	0.12	0.28	0.41	0.27
Yoshida[56]	Minimum	0.47	0.45	0.43	0.34	0.03	0.06	0.25	0.25
	Mean	0.93	0.92	0.89	0.81	0.07	0.15	0.26	0.26
	Maximum	1.29	1.27	1.25	1.17	0.10	0.22	0.28	0.26
Cao[57]	Minimum	0.32	0.32	0.31	0.30	0.24	0.01	0.13	0.24
	Mean	1.01	1.00	0.97	0.87	0.26	0.09	0.20	0.26
	Maximum	1.33	1.31	1.29	1.19	0.28	0.12	0.52	0.28
Perkins[58]	Minimum	0.74	0.73	0.70	0.60	0.06	0.11	0.09	0.23
	Mean	1.19	1.17	1.15	1.05	0.10	0.19	0.20	0.26
	Maximum	1.38	1.37	1.34	1.27	0.12	0.24	0.27	0.27
Nghiep[59]	Minimum	0.61	0.59	0.58	0.50	0.04	0.08	0.26	0.23
	Mean	1.16	1.14	1.12	1.05	0.08	0.19	0.30	0.25
	Maximum	1.41	1.40	1.38	1.32	0.10	0.25	0.34	0.26

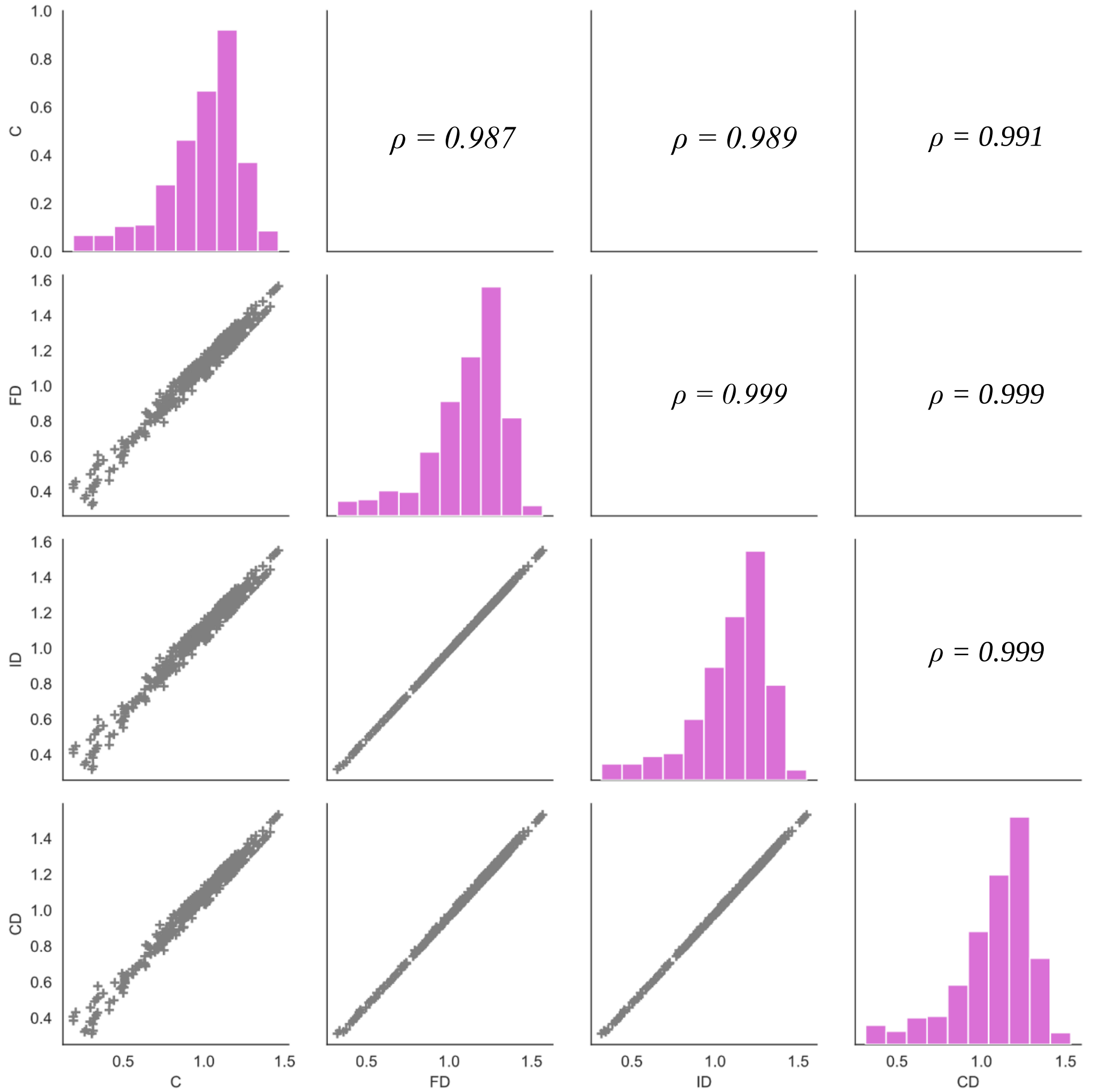


Fig. 5. Pair-plot of input features (C, FD, ID, CD) showing statistical distribution and correlation. (color printed)

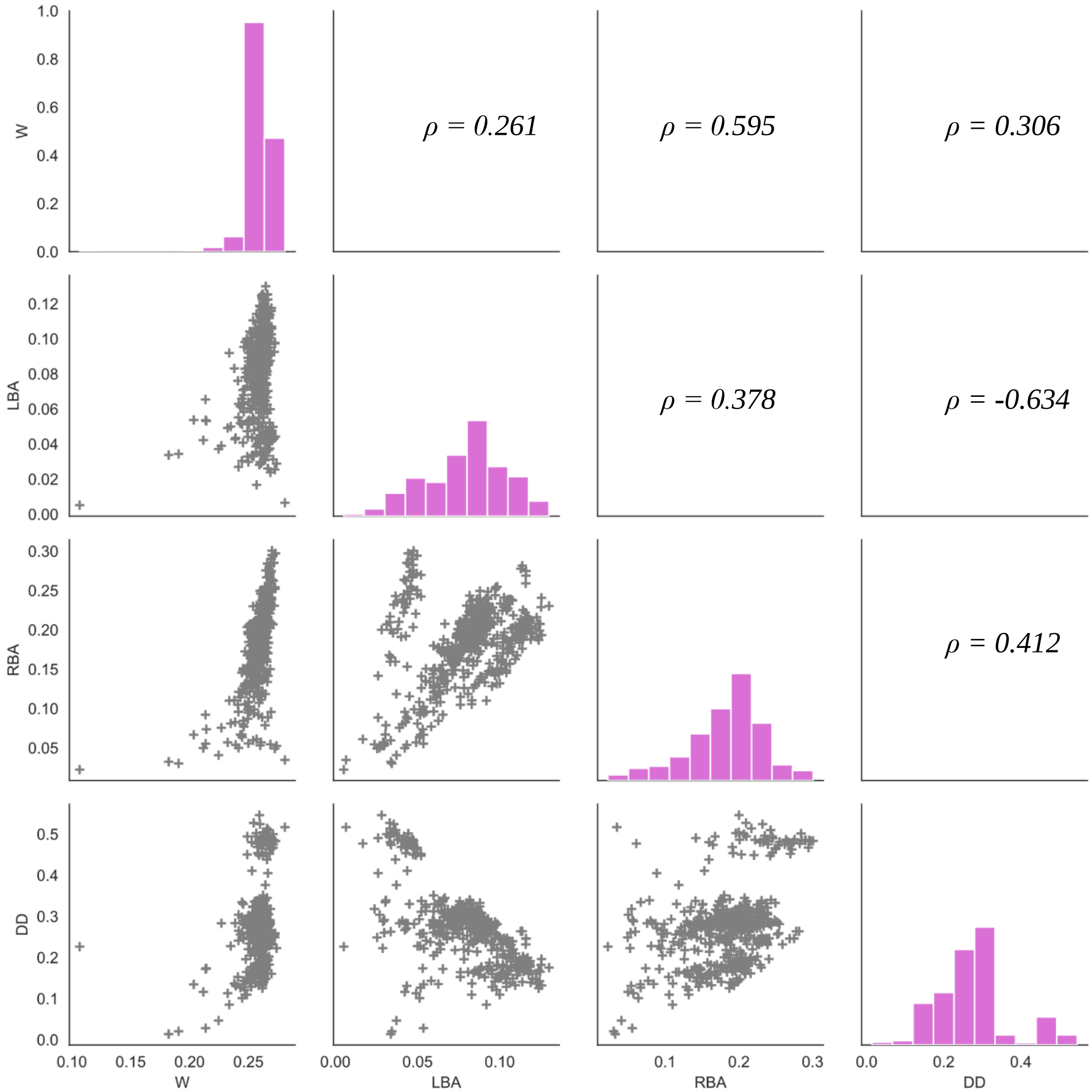


Fig. 6. Pair-plot of input features (W, LBA, RBA, DD) showing statistical distribution and correlation. (color printed)

5. Machine Learning Model Implementation

5.1 Training-Testing Data Splitting

Fig. 7 shows a schematic representation of the proposed machine learning model implementation procedure. Firstly, the image database of RC beams and one-way slabs is split into training and testing data. In this study, random samples of 70% of the entire database was assigned to the training data, whereas the remaining 30% was assigned as testing data. Four regression-like machine learning techniques were implemented using the training data (see Fig. 7). A brief background on these four regression techniques is presented as follows:

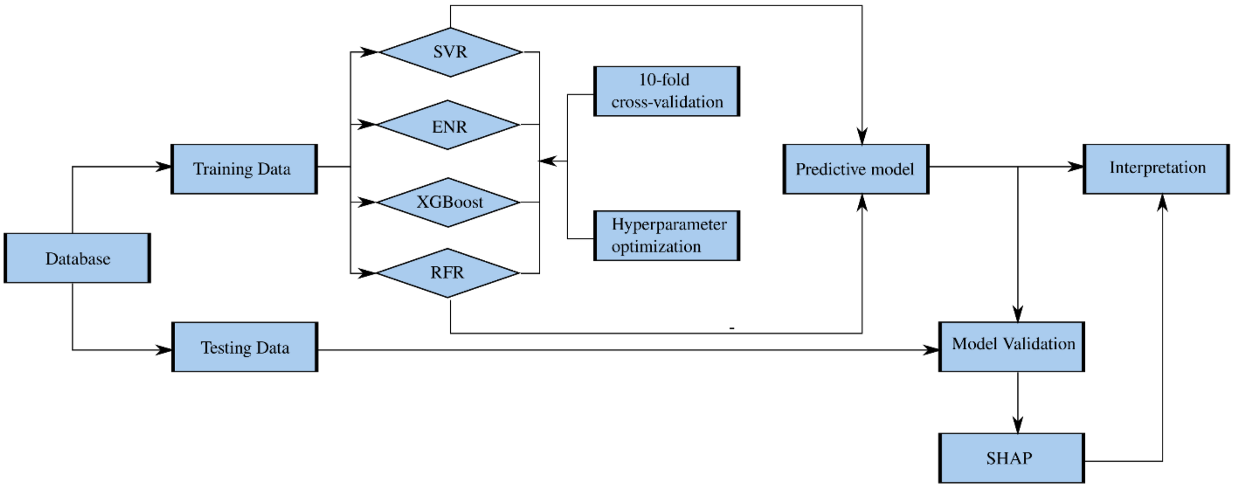


Fig. 7. Machine learning model implementation. (color printed)

5.2 Machine learning Algorithms

In the present context, a predictive model that could map the set of multifractal features into a load level (*FR*) estimate for the database of RC beams and one-way slabs is sought after. The Support Vector Regression (SVR), Random Forest Regression (RFR), linear Elastic-Net Regression (ENR) and the Extreme Gradient Boosting (XGboost) algorithm were adopted in this study. All these machine learning techniques have been successfully employed in solving similar structural engineering-related problems [60–62] which usually comprises a relatively limited number of data points in a dataset.

5.2.1 Elastic-Net regression (ENR)

The basic linear regression model seeks to provide a solution to finding the best fit between a set of input points and an output. In the present context, given an input vector of multifractal features, $X_i = (x_{i1}, x_{i2}, x_{i3}, ...x_{ip})$ and an output load level, LL, of an RC beam or one-way slab, the linear regression model has the following functional form [52]:

$$LL_i = \beta_0 + \sum_{i=j}^{p} \beta_j x_{ij} \tag{12}$$

where β_j are the unknown parameters and p is the number of input features. Given a training dataset $((X_1, LL_1), (X_2, LL_2), (X_3, LL_3), ...(X_N, LL_N))$, β_j are estimated by using the most popular loss function; the sum of squared error (SSE) as given in Eq. 13.

$$SSE(\beta) = \sum_{i=1}^{N} \left[LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right]^2 \tag{13}$$

It turns out that the estimates obtained from minimizing the *SSE*, have the smallest variance for all available linear unbiased estimators [53]. Nevertheless, biased estimators tend to have a fairly relatively low variance compared to their unbiased counterpart. The emphasis of most regression-like machine learning models is to determine model parameters that will reduce the generalization or test error, hence the variance. To this end, the regularized variable selection regression model, Elastic-Net Regression (ENR) is able to mitigate this drawback of the original regression model. It consists of minimizing the aggregate sum of a loss and penalty function. The unknown parameters $\beta_{elastic}$ are estimated from Eq. 14.

$$\beta_{elastic} = \underset{\beta}{\operatorname{argmin}} \left(\sum_{j=1}^{N} \left[LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right]^2 + \lambda \sum_{j=1}^{p} \left[\alpha \beta_j^2 + (1-\alpha) \left| \beta_j \right| \right] \right) \tag{14}$$

The penalty term as seen in Eq. 14, requires the specification of two hyperparameters; λ and α. A comprehensive description of ENR can be found in Hastie et al. [54].

5.2.2 Support Vector Regression (SVR)

The general support vector machine which was originally described to solve classification problems, can be adapted for regression analysis [52]. Similar to the elastic-net model presented above, the algorithm minimizes the following objective function:

$$\beta_{svr} = \underset{\beta}{\operatorname{argmin}} \left(\sum_{i=1}^{N} V_{\varepsilon} \left(LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right) + \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2 \right) \tag{15}$$

$$\text{where } V_{\varepsilon}(r) = \begin{cases} 0 & \text{if } |r| < \varepsilon \\ |r| - \varepsilon, & \text{otherwise} \end{cases} \tag{16}$$

This support-vector formalism is usually referred to as the ϵ-insensitive or error-insensitive SVR model. It basically requires the determination of two hyperparameters, epsilon (ε) and lambda (λ). However, the general minimization problem is solved numerically by making use of kernels after approximating the regression function given in Eq. 12 with a set of basis functions [55]. Some of the widely used kernels are the polynomial, sigmoid, and the gaussian radial basis kernel function. The selection of the most appropriate kernel as well as other hyperparameters is oftentimes determined via cross-validation.

5.2.3 Random Forest Regression (RFR)

Random forest leverages the superiority of considering an ensemble of regression trees for decision making, in this case, predicting a quantitative response value (see Fig. 8). The algorithm begins with bootstrapping a sample from the training data, from which a regression tree that utilizes a random selection of a subset of features can be developed [52]. This procedure is repeated for different bootstrap samples and features. The prediction of unseen or test data can then be computed by taking the mean of the predictions obtained from the various regression trees already developed. Fig. 8 provides a schematic presentation of the Random Forest Regression (RFR) implementation procedure. A couple of hyperparameters influence the performance of an RFR scheme; the number of trees or estimators, maximum depth of tree, and the number of features to select at each split, and the minimum number of samples in each split.

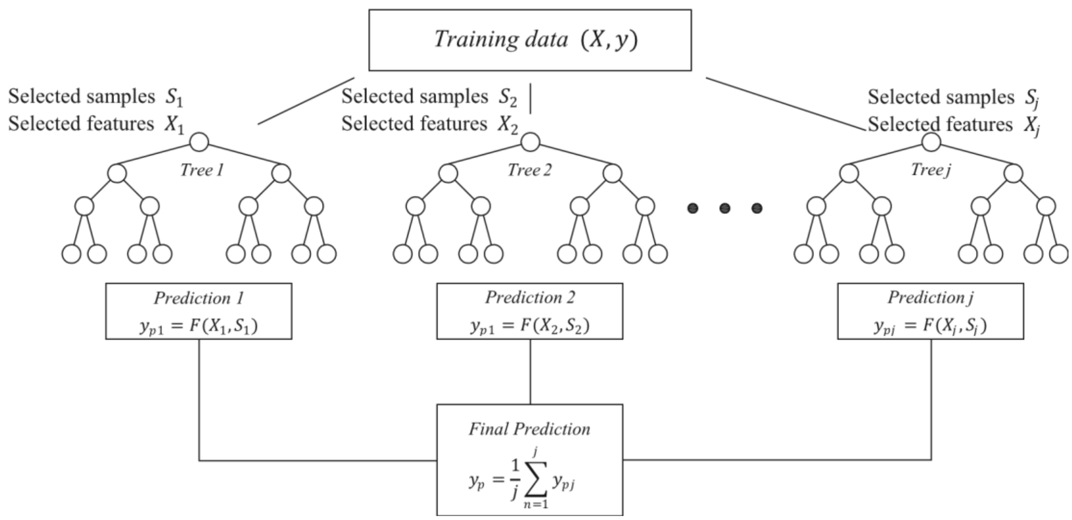


Fig. 8 A random forest regression implementation scheme.

5.2.4 Extreme Gradient Boosting (XGBoost)

This fairly recent developed machine learning technique is an extension of the popular ensemble learning method, gradient descent decision tree [56,57]. The XGBoost aggregate a collection weak learner that is usually obtained from a decision tree model. Whereas random forest regression outputs the mean of different trees, XGBoost incrementally improves the prediction through a weighted aggregation of weak learners to form a strong learner. In this study, decision trees are used as weak learners. The XGBoost regressor seeks to provide a mapping between the input set of features and the output of a training dataset using the following Equation.

$$LL_i = \sum_{k=1}^{K} \sigma_k f_k(X_i) \tag{17}$$

where, K is the number of weak learners or estimators, σ_k is the learning rate, and $f_k(X_i)$ is the weak leaner obtained from a decision tree. In determining the most appropriate learner at a particular stage, and other hyperparameters, the loss and penalty functions that need to be minimized is given in Equation 18 below.

$$f_t = \underset{f \in F}{\operatorname{argmin}} \left[\sum_{i=1}^{N} \left(LL_i - \sum_{k=1}^{t} \sigma_k f_k(X_i) \right)^2 + \sum_{k=1}^{t} \left(\gamma T + \frac{1}{2} \lambda \, \| w_k \| \right) \right] \tag{18}$$

where f_t is the weak learner to be determined at the *t*-th step, γ and λ are the hyperparameters of the penalty term, and T and w_k are the number of leaf nodes and weights, respectively. It is worth noting that, the sequential nature of the XGBoost algorithm only permits the determination of the optimal weak leaner and penalty coefficients at the *t*-th step (f_t, γ and λ), since all other parameters and learners before the *t*-th step would have been determined. The output of the regression model is sequentially updated to a point where *t* equals to *K*, the number of weak learners to be considered. Further details on how the weak learners with its accompanying hyperparameters are determined can be found elsewhere in Chen and Guestrin [57].

5.3 Hyperparameter Optimization

In the implementation process, a 10-fold cross-validation scheme was utilized in hyperparameter optimization via a random search, in order to determine the best set of parameter combinations for each model training. The performance measure used in determining the optimal hyperparamter was the mean squared error. This analysis is performed for 1000 runs, and the modal values of the hyperparameters that were optimal for each machine learning model is presented in Table 3. As observed, the optimal number of estimators for the random forest and extreme gradient boosting machine were different (see Table 3), after hyperparameter optimization. The number of estimators refers to the number of decision trees that constitutes the meta model. Informed comparisons between these two models can be made since their learning algorithms are different. For instance, whereas random forest assigns equal weight to each decision tree during the aggregation process to make a final prediction, the weighting scheme for the extreme gradient boosting machine model is adjustable or adaptive and depends on the loss function to be minimized. With this inherent difference in the two algorithms, the number of estimators does not have to be necessarily equal to make well-informed comparison during model evaluation.

Table 3. Tuned hyperparameters for various machine learning models

Model	Hyperparameter	Modal Value
SVR	Kernel	Radial Basis
	Epsilon (ε)	0.1
	Lambda (λ)	1000
ENR	Alpha (α)	0.9
	Lambda (λ)	0.001
RFR	Number of Estimators	800
	Maximum depth of tree	6
	Minimum samples for split	3
	Maximum number of features	3
XGboost	Number of Estimators	500
	Learning rate	0.01
	Maximum depth of tree	6

Minimum samples for split	3
Lambda (λ)	0.1
Gamma (γ)	0.1

5.4 Performance Measures

One of the four machine learning models obtained from the training data after hyperparameter optimization was then selected as the final proposed predictive model. In order to make valuable comparison of the various machine learning models, suitable performance or error measures are needed to be selected, for the acquisition of illustrative estimation accuracy of the output variable. To that end, the four-regression performance metrics were used in this study, with a brief description of them given below.

5.4.1 Root-Mean-Squared Error (RMSE)

This performance measure assesses the difference between the true and predicted output of an entire dataset as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (LL_i - LL_i)^2}{N}} \tag{19}$$

where LL_i is the true value of the load-level for a particular datapoint i, LL_i is the predicted value, and N represents the total number of samples in the dataset.

5.4.2 Correlation Coefficient (R)

The strength and direction of the linear relation between the predicted and true values of the output can be measured using the correlation coefficient, *R*. Values of *R* are usually bounded between -1 and 1, and it depicts the strength of the correlation, with positive values presenting positive correlation and vice-versa. The correlation coefficient, *R*, can be computed as:

$$R = 1 - \frac{\sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right) \left(LL_i - \overline{LL_i} \right)}{\sqrt{\sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right)^2 \sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right)^2}} \tag{20}$$

where $\overline{LL_i}$ and $\overline{LL_i}$ are the averages of the true and predicted load-levels, respectively.

5.4.3 Explained Variance Score (EV)

The explained variance score measures the extent to which the variance in the output of the dataset is captured by the predictive model. Values of EV closer to 1.0 signifies a higher correlation between predicted and true values of the output. Mathematical, Explained Variance Score, EV, is computed as:

$$EV = 1 - \frac{\sum_{i=1}^{N}\left(LL_i - LL_i - \overline{LL_i} + \overline{LL_i}\right)^2}{\sum_{i=1}^{N}\left(LL_i - LL_i\right)^2} \tag{21}$$

5.4.4 Index of Agreement (IA)

It establishes a level of agreement between the predicted and their corresponding true values. It is a dimensionless measure of model accuracy and has been argued by some researchers as a remarkable improvement to the more popular coefficient of determination. Values of Index of Agreement (IA) closer to 1.0 signifies better agreement. Although similar to the correlation coefficient, R, IA is less sensitive to outliers or extreme values and is computed as follows:

$$IA = 1 - \frac{\sum_{i=1}^{N}\left(LL_i - LL_i\right)^2}{\sum_{i=1}^{N}\left(\left|LL_i - \overline{LL_i}\right| + \left|LL_i - \overline{LL_i}\right|\right)^2} \tag{22}$$

The best performing machine model is selected by assessing the aforementioned performance metrics on the testing data. The model is then validated by considering the full dataset and predicting the load-level of the RC beams and one-way slabs.

5.5 Model Interpretation

The various forms of machine learning techniques differ in their level of complexity, and hence influence how they can be interpreted. Generally, linear models are more likely to be interpreted with ease, and thus can give a fair understanding of the underline process being modelled. Also, they tend to give valuable insight and information needed for model improvement. Conversely, linear models are not sophisticated enough to yield very accurate results compared to non-linear machine linear models. For instance, the XGBoost regression model usually tends to produce more accurate results than linear regression models on many datasets. On the other end, interpretating a model developed from the XGboost algorithm or any flexible machine learning model, is quite challenging. To this end, the recently developed SHapley Additive exPlanation (SHAP) tool can be used for model interpretability of very complex machine learning models. SHAP results in the provision of a so-called explanation model useful for (1) demonstrating the importance of any feature in the dataset; (2) quantifying how each feature affects the model prediction on both local and global scales; (3) ascertaining how the prediction model output changes with variations in the input values of the feature. A brief description of Shapley Additive Explanation (SHAP) for model interpretation is presented below.

Once again, consider an example input vector of features $X_i = (x_{i1}, x_{i2}, x_{i3}, ...x_{ip})$ for which a machine model $f(X_i)$ is developed to predict a quantitative response LL_i. The SHapley Additive ExPlanation (SHAP) for machine learning model interpretation begins with mapping the original input vector of features X_i into a binary simplified input vector $X_i' \in \{0,1\}^p$, which serves as input for the explanation model $g(X_i')$. The X_i' which contains either 0 or 1, depicts whether a feature is present ($x_{ij}' = 1$) or absent ($x_{ij}' = 0$) in the explanation model yet to be determined. The explanation model is usually obtained by a weighted summation of the simplified input vector of features X_i' and a constant term as represented in Eq. 23.

$$g(X_i') = \theta_0 + \sum_{j=1}^{p} \theta_j x_{ij}' \tag{23}$$

where $X_i' \in \{0,1\}^p$ is a vector of binary simplified inputs features, x_{ij}', which are mapped to the original input features x_{ij}, and θ_j is the attribution value for feature j. To this end, SHAP is usually referred as a class of feature attribution methods, amongst others such as LIME [58], deepLIFT [59] etc.

The advantage of using SHAP as opposed to other feature attribution methods is how it presents three key desirable properties that any feature attribution method should have. The first property deals with local accuracy, where the output of the explanation is expected to match that of the model prediction for any data point in the dataset (see Eq. 24).

$$f(X_i) = g(X_i') \tag{24}$$

Secondly, if a feature does not contribute to the predictive model's output, then the feature attribution value should be zero in the explanation model (see Eq. 25).

$$x_{ij}' = 0 \Rightarrow \theta_j = 0 \tag{25}$$

To conclude, the third property states that if the predictive model changes and causes a particular simplified input contribution to increase or stay the same regardless of other simplified inputs, then the attribution from that input should not decrease. In explaining the third property, known as consistency, consider two predictive models $f_1(X_i)$ and $f_2(X_i)$. Mathematically, the consistency property can be presented as:

$$f_1(X_i) - f_1(X_i \setminus j) \geq f_2(X_i) - f_2(X_i \setminus j) \Rightarrow \theta_j(f_1) \geq \theta_j(f_2) \tag{26}$$

where $f_1(X_i \setminus j)$ and $f_2(X_i \setminus j)$ denote prediction values of models $f_1(X_i)$ and $f_2(X_i)$ with feature j absent, respectively. Similarly, $\theta_j(f_1)$ and $\theta_j(f_2)$ are the feature attribution values for $f_1(X_i)$ and $f_2(X_i)$ respectively.

It turns out the only solution for the feature attribution values θ_j that satisfies these three properties, are the Shapley values of the conditional expectation function of the original model[60]. These Shapley values can be computed from Eq. 27 as:

$$\theta_j(f, X_i) = \sum_{Z'_i \subseteq X'_i} \frac{|Z'_i|!\left(P - |Z'_i| - 1\right)!}{P!}\left[f\left(Z'_i\right) - f\left(Z'_i \setminus j\right)\right] \tag{27}$$

where $\theta_j(f, X_i)$ is the Shapley regression value or feature attribution value for the feature j in the model $f(X_i)$, Z' is a vector of binary values representing one of the subsets of X', P is the number of input features, $|Z'|$ represents the number of non-zero elements in Z', $f\left(Z'_i\right)$ denotes the model prediction for Z' and $f\left(Z'_i \setminus j\right)$ represents the prediction for Z' without feature j. These Shapley values $\theta_j(f, X_i)$, once obtained, can be used to explain the model output. The magnitude and sign of $\theta_j(f, X_i)$ will determine whether a particular feature impacts the model output negatively or positively. The θ_0 from Eq. 23 represents the average value of the model prediction assuming the model has no input feature and usually represents a base value for the model output before the various Shapley values obtained from Eq. 27 are aggregated to obtain the output $f(X_i)$. Further details on techniques available to compute the Shapley values can be found elsewhere in [60].

6. Results and Discussions

6.1 Model Predictions and Evaluation

6.1.2 Global Level

The performance of the four selected machine learning models for load-level estimation of the class of structural elements under consideration is presented. Following the training-testing splitting rule of 70/30 as previously mentioned, the accuracy of these models was drawn for each group of data (training and testing data). Typically, the performance of the model on the testing data is used to determine its generalization capacity. Table 4 shows a summary of the four performance measures for each dataset, across the machine learning models developed. It presents the mean and standard deviation of the performance measures for 1000 runs of the developed models having different randomly sampled training and testing data. Multiple runs of the developed models were necessary to help ascertain how statistically significant the model predictions might differ. It is worth mentioning that high values of the correlation coefficient (R), explained variance (EV) and index of agreement (IA) for a particular model signifies greater performance. Similarly, models with lower root-mean squared error (RMSE) also presents a case for better predictability.

Among the four machine learning models, the RFR and XGBoost models yielded the best performance on the training and testing data respectively (see Table 4). They produced relatively high values of the correlation coefficient (R), explained variance (EV) and index of agreement (IA) when compared to the ENR and SVR models. Similarly, lower average values were recorded for the root-mean squared error (RMSE) of these models, when compared to the ENR and SVR models, during the training and testing phase. However, the diferrence between the mean estimate for these models (RFR and XGBoost) were comparatively similarly, as well as their deviations. To assess the statistical significance of the differences of the mean values of these two high performing models we calculated the t-statistic, compared this to the critical t-value, and calculated the corresponding p-values as well. Details on how the t-statistic is computed when comparing means of different populations can be found elsewhere [61,62] .

Table 4. Performance measures of various machine learning models

Data	Algorithm	Statistic	Performance Metrics			
			RMSE	R	EV	IA
Training	SVR	Mean	0.150	0.811	0.628	0.851
		SD	0.004	0.012	0.020	0.011
	ENR	Mean	0.151	0.785	0.616	0.867
		SD	0.004	0.012	0.020	0.009
	RFR	Mean	0.0897	0.934	0.867	0.961
		SD	0.003	0.005	0.009	0.003
	XGBoost	Mean	0.104	0.913	0.819	0.941
		SD	0.003	0.005	0.011	0.004
Testing	SVR	Mean	0.151	0.810	0.625	0.849
		SD	0.009	0.028	0.035	0.016
	ENR	Mean	0.152	0.784	0.611	0.864
		SD	0.008	0.028	0.043	0.017
	RFR	Mean	0.138	0.827	0.681	0.900
		SD	0.009	0.026	0.044	0.014
	XGBoost	Mean	0.136	0.831	0.687	0.895
		SD	0.008	0.026	0.042	0.014

SD: standard deviation; SVR: Support vector regression; ENR: elastic-net regression; RFR: random forest regression; XGBoost: extreme gradient boosting.

Table 5 and 6 presents the calculated t-values and p-values for the comparisons of the performance mean values for the RFR and XGBoost models. The t-values were compared to a critical t-value of 1.96, obtained from the student's-t distribution at a 5% significance level with 1998 degrees of freedom. All t-values computed for these two models, and across various performance measures were higher than this critical value (see Table 5 and 6). The calculated p-values show that the actual levels of statistical significance are all below 1%.

The data shown in Tables 4, 5 and 6 mean that the differences between the mean values of RFR and XGBoost for the various performance measures are statistically significant. From Table 4, the XGBoost model outperformed the RFR model when the RMSE, R and EV are considered, whiles a higher IA values was observed for the RFR model, during the testing phase. To this end, we recommed the XGBoost model as the optimal model for load level estimation of shear-critical RC beams and slabs. Since the generalization capability of a model is usually assessed by considering how it performs during the testing phase, further comparisons between these two models are drawn.

Table 5. T-values of various model comparisons

Performance Measure	Dataset	Model	t-value			
			SVR	ENR	RFR	XGB
RMSE	Training	SVR		-5.59017	381.3707	290.9295
		ENR	5.59017		387.6952	297.2541
		RFR	-381.371	-387.695		-106.586
		XGB	-290.93	-297.254	106.5859	
	Testing	SVR		-2.62613	32.29876	39.39193
		ENR	2.626129		36.7658	44.72136
		RFR	-32.2988	-36.7658		5.252257
		XGB	-39.3919	-44.7214	-5.25226	
R	Training	SVR		48.44814	-299.2	-248.117
		ENR	-48.4481		-362.446	-311.363
		RFR	299.2001	362.4457		93.91486
		XGB	248.1172	311.3627	-93.9149	
	Testing	SVR		20.76349	-14.0693	-17.3797
		ENR	-20.7635		-35.5871	-38.8975
		RFR	14.0693	35.58705		-3.4401
		XGB	17.37972	38.89748	3.440105	
EV	Training	SVR		13.41641	-344.608	-264.615
		ENR	-13.4164		-361.91	-281.24
		RFR	344.608	361.9105		106.7986
		XGB	264.615	281.2401	-106.799	
	Testing	SVR		7.985022	-31.4975	-35.8615
		ENR	-7.98502		-35.9803	-39.9834
		RFR	31.49748	35.98033		-3.11925
		XGB	35.86152	39.98339	3.119251	
IA	Training	SVR		-35.5995	-305.085	-243.154
		ENR	35.59953		-313.333	-237.6
		RFR	305.0851	313.3333		126.4911
		XGB	243.1545	237.5997	-126.491	
	Testing	SVR		-20.3186	-75.8579	-68.4209
		ENR	20.31856		-51.693	-44.5134
		RFR	75.85792	51.69299		7.985957
		XGB	68.42087	44.51341	-7.98596	

Table 6. P-values of various model comparisons

Performance Measure	Dataset	Model	p-value			
			SVR	ENR	RFR	XGB
RMSE	Training	SVR		2.58E-08	0	0
		ENR	2.58E-08		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		0.008702	1.6E-184	1.2E-251
		ENR	0.008702		1.8E-226	2.9E-303
		RFR	1.6E-184	1.8E-226		1.66E-07
		XGB	1.2E-251	2.9E-303	1.66E-07	
R	Training	SVR		0	0	0
		ENR	0		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		7.19E-87	6.12E-43	4.07E-63
		ENR	7.19E-87		2.9E-215	6.9E-247
		RFR	6.12E-43	2.9E-215		0.000593
		XGB	4.07E-63	6.9E-247	0.000593	
EV	Training	SVR		2.32E-39	0	0
		ENR	2.32E-39		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		2.35E-15	3.8E-177	7.2E-218
		ENR	2.35E-15		5.4E-219	2.4E-257
		RFR	3.8E-177	5.4E-219		0.001839
		XGB	7.2E-218	2.4E-257	0.001839	
IA	Training	SVR		2.2E-215	0	0
		ENR	2.2E-215		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		1.39E-83	0	0
		ENR	1.39E-83		0	3E-301
		RFR	0	0		2.33E-15
		XGB	0	3E-301	2.33E-15	

In corroborating this finding, a score analysis is conducted. Score analysis basically entails assigning a score to the various values of the performance measures across different models. In this study, with the number of machine learning models being 4, a model that yields the greatest performance is assigned a score of 4, whereas the least performing model is assigned a value of 1. Considering that 1000 runs of the developed models for different training and testing data was conducted, the average score for a particular model is used. Subsequently, a summation of the average scores of the various performance measures for each machine learning model is computed to obtain the total score (see Table 7).

Table 7. Score analysis results of various machine learning models

Data	Algorithm	Score				Total Score
		RMSE	R	EV	IA	
Training	SVR	1.63	1.00	2.00	1.92	6.55
	ENR	1.37	2.00	1.00	1.08	5.45
	RFR	4.00	4.00	4.00	4.00	16.00
	XGBoost	3.00	3.00	3.00	3.00	12.00
Testing	SVR	1.63	1.00	2.25	1.85	6.73
	ENR	1.4.0	2.00	1.02	1.20	5.62
	RFR	3.32	3.88	3.10	3.31	13.61
	XGBoost	3.65	3.12	3.63	3.64	14.04

The model producing the highest total score is deemed to be the best performing model. As seen in Table 7, the RFR and XGBoost models dominated the score analysis by being the best models during the training and testing phase respectively. Nevertheless, since the generalization capability of model can be evaluated by considering its performance on the testing data, the XGBoost model is deemed the optimal model for load-level estimation of shear-critical RC beams and slabs. The attained total score were 13.61 and 14.04 for the RFR and XGboost models respectively, during the testing phase. However, the RFR model tends to outperforms the XGBoost model during the training phase (see Table 7). This observation might imply that there is an inherent overfitting problem with the RFR model. The least performing model was the Elastic-Net Regression (ENR), which yielded total scores of 5.45 and 5.62 during the training and testing phase respectively. This observation also suggests that the linear statistical method of analysis may not be optimal for predicting the load-level of shear-critical beams and slabs using multifractal analysis.

Graphical presentation of the score analysis is given in a form of a radar chart as shown in Fig. 9, to facilitate interpretation. It is observed that the Random Forest Regression (RFR) model tends to perform well on the training data (see Fig. 9) than any other model across the various performance measures. Similarly, the radar charts indicates the the XGboost model performance better than the RFR model during the testing phase. This suggests that non-linear models, in particular tree-based models such as random forest and the extreme gradient boosting machine, tends to produce better estimates of the load-level of shear-critical concrete beams and slabs using the proposed framework.

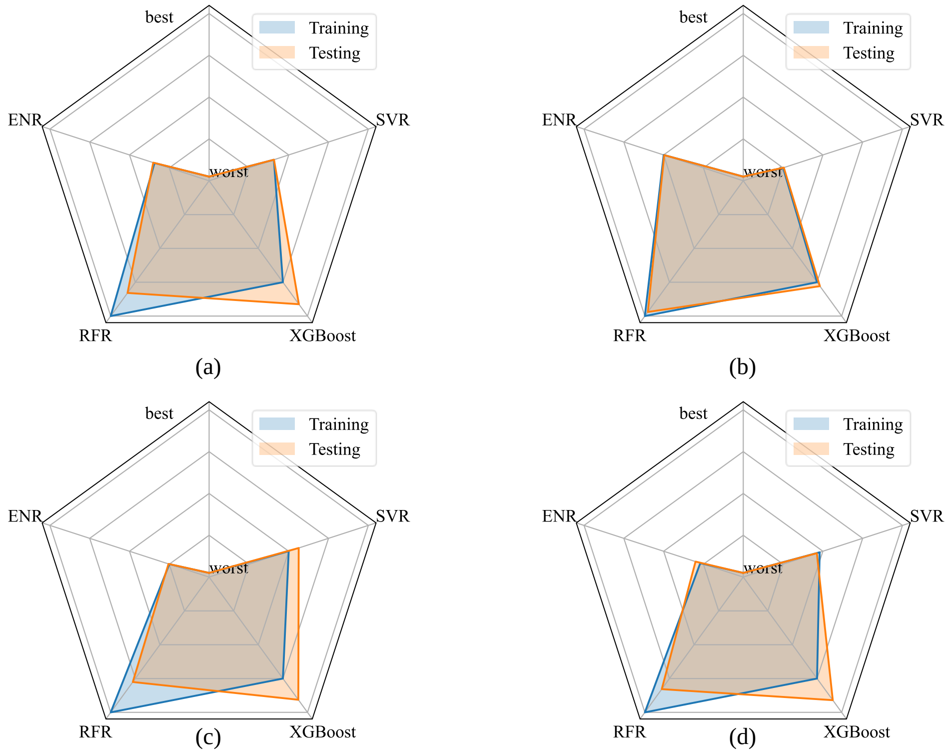


Fig. 9 Radar charts for various performance measures: (a) RMSE; (b) R; (c) EV; (d) IA. (color printed)

6.1.3 Local Level

In order to gain insight into the predictive performance of these models at the local level, Fig. 10 shows a typical scatter plot, to help establish the correlation between predicted and true values of the load-level for each data point in the training and testing dataset. This visualization will also assist in determining which regions across the load-level range, tends to produce better estimates. Evidently, the XGboost produces the lower scatter or deviation with a narrow prediction interval compared to the other models investigated in this study (see Fig. 10). The mean of the predicted-to-tested ratio for this model was 1.04 with a coefficient of variation of 27%. Nonetheless, there seems to be significant error or outliers for some data points, particularly in the testing data. The majority of these data points yielded a prediction of load-level higher than their true values, and hence conservative for damage assessment or design. Although there exist works on estimating the load-level of beams and slabs using fractal analysis and other data-driven machine learning algorithms[2,35,63], fair comparison cannot be generally drawn for most of them due to the disparity in specimens that make up the database as well as its size. Nevertheless, a closely related work that used about 95% of the database in this study is that of Davoudi et al. [2] who provided another alternative to damage assessment of shear-critical concrete beams and slabs using machine vision In their assessment, scatter plot and performance metric values similarly those presented in Fig. 10 and Table 4 were plotted. By comparison, the developed model produced comparable performance measures as against those reported by Davoudi et al. [2]

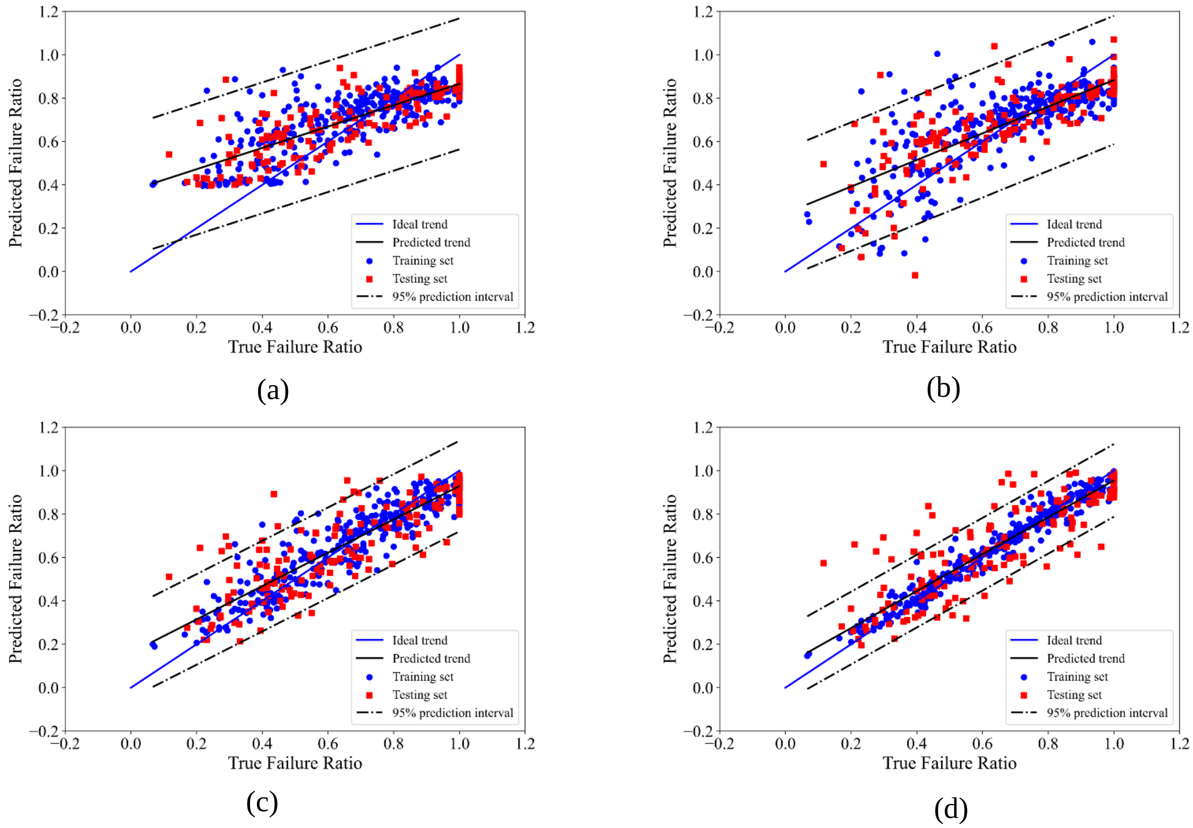


Fig. 10 Scatter plot of load-level results predicted by different machine learning model: (a) SVR; (b) ENR; (c) RFR; (d) XGBoost. (color printed)

A typical regression error characteristic (REC) curve as constructed in Fig. 11 for the various models is used to facilitate model predictability at the local scale. The REC curve is a cumulative distribution function which tends to establish a relationship between the absolute error or deviation (x-axis) as against the proportion of datapoints (y-axis) with absolute error lesser than or equal to the current level. It is analogous to the receiver operating characteristic (ROC) curve in classification problems for model assessment. Whereas the ROC curve uses the area under the curve (AUC) to evaluate performance, it has been widely established that the area over curve (ROC) be used to provide a valid measure for regression problems. The ROC can be simply computed by subtracting the AUC from 1. A regression model is known to perform well if the AOC value of an REC curve is low.

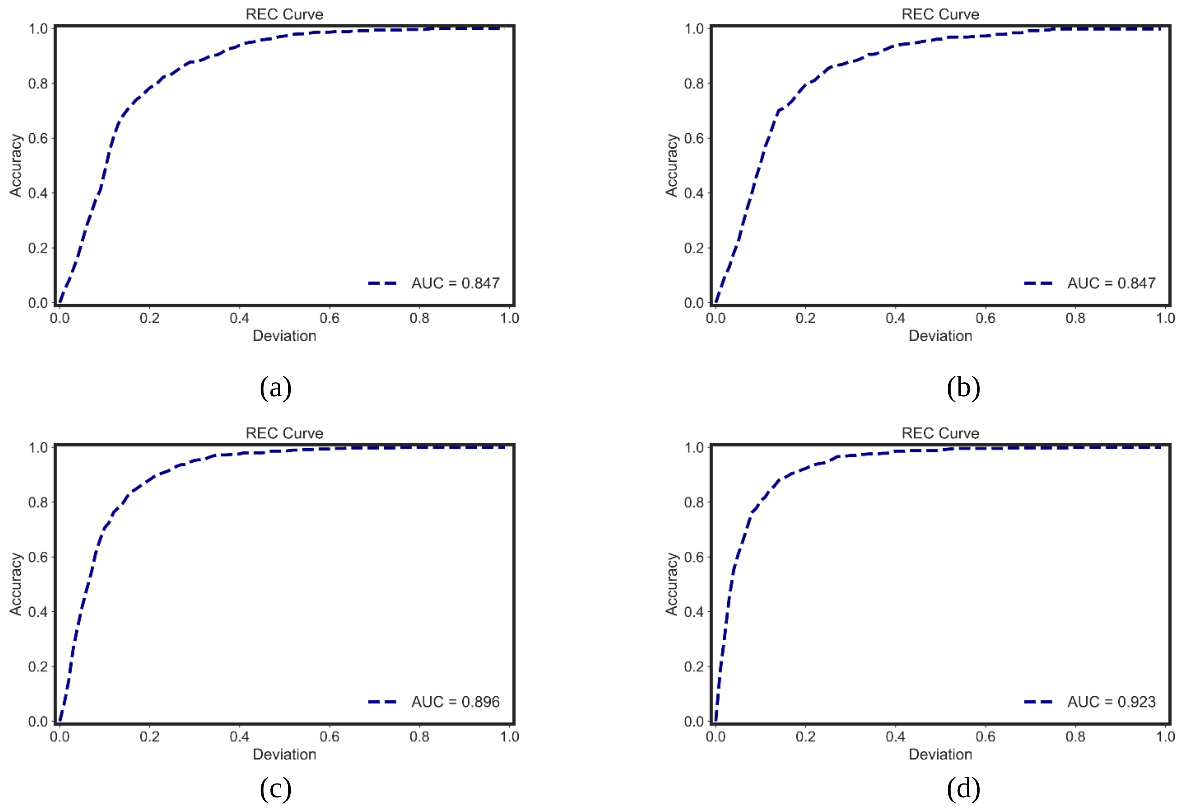


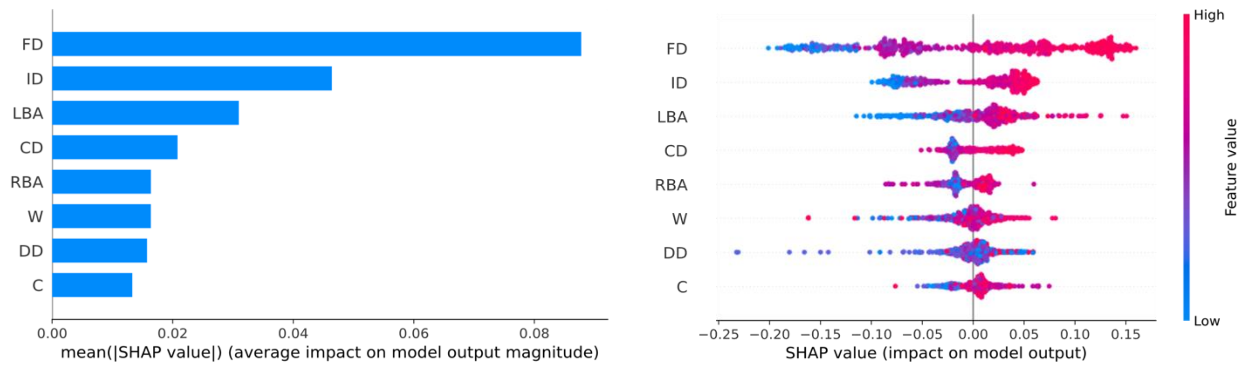
Fig. 11 Regression error characteristic curves for various machine learning model: (a) SVR; (b) ENR; (c) RFR; and (d) XGBoost

From Fig. 11, which shows the REC curve using the full dataset, the XGBoost model produced the lowest ROC of 0.077, hence corroborating findings attained at the global level of assessment. The ROC for both SVR and ENR models were the same, hence suggesting equal performance. 80% of the datapoints produced absolute errors of load-level lesser than 0.1 for the XGBoost model (see Fig. 11d). The RFR, ENR and SVR models yielded predictions of which 80% had absolute errors within 0.17, 0.21 and 0.21 respectively. In this study, the XGBoost model developed remains the optimal model at both local and global levels for estimating the load-level of shear-critical RC beams and slabs.

6.2 Model Interpretation

6.2.1 Global Level

A simplified explanation model was developed for the optimal predictive model, i.e, XGBoost, for interpretation using SHapley Additive ExPlanation (SHAP). On the global scale (entire dataset), the relative importance of each feature is given in Fig. 12. It provides the mean of the absolute SHAP values computed for each feature in the full dataset. These mean values are then used to ascertain the impact of each feature on the predictions made.



(b) (c)

Fig. 12 Global interpretations of XGBoost model: (a) SHAP feature importance; and (b) SHAP summary plot. (color printed)

Generally, it was observed that the so-called generalized dimensions (FD, ID and CD), which were obtained from the multifractal analysis of the crack patterns considered, has significant impact on the estimation of the load-level, as opposed to the other geometric features acquired from the singularity spectrum. For the generalized dimensions, the box-counting fractal dimension (FD) was arguably the most critical parameter (see Fig. 12a). Many of previous works on the application of multifractal analysis for crack damage assessment of RC elements have always considered FD as the most influential feature, with the findings from this study affirming it. The area under the left branch of the singularity spectrum (LBA) tends to contribute the most to the model predictions for the geometric features considered, providing about 35% of that produced by FD. The least contributing feature as seen in Fig. 12a is the capacity (C), whose mean absolute SHAP value was about 17% as important as the most critical feature.

In order to determine how the original values of the features within the dataset affects the model prediction or load-level, Fig. 12b demonstrate a summary plot for such analysis. Each point in the plot shows the SHAP value (x-axis) of a particular feature (y-axis). For each feature, the distribution of SHAP values are shown along the x-axis, which are colour-coded to differentiate between high (red dots) and low (blue dots) values of the original feature. For instance, for high values of the fractal dimension (FD) as seen in the upper right corner of Fig. 12b, there is an expected increase in the load-level of about 16%. Nevertheless, there are instances for which higher values of FD cause a reduction in the load-level (red dots on the left-hand side of the summary plot for FD). To this end, the average value of the distribution of SHAP values is used to ascertain whether a feature impacts the load-level positively or negatively. In general, for the critical features, an increase in the fractal dimension FD, information dimension ID, and correlation dimension CD causes an increase in the load-level. Conversely, the load-level tends to decrease when the area under the left branch (LBA), is low.

6.2.2 Local Level

SHAP also provides interpretation for each individual prediction. In assessing the impact of the various feature at the local level, four RC beams were sampled from the database considered. These samples had load-levels spanning various damage states (low, moderate, near failure and ultimate failure). For the sample exhibiting a lower degree of damage, a simplified explanation model which comprises the aggregation of SHAP values for each feature and a base value to yield a final prediction is given in the second column of Table 8. This sample had a true load-level of 17.1% and a predicted value of 22%. It is worth noting that the base value depicts the default prediction when the attribution from each feature is excluded.

Table 8. Relative SHAP values of features for four selected samples

Feature	Shapley values of selected sample scenarios (%)			
	Low	Moderate	Near failure	Failure
C	-3.6	1.7	0.2	0.6
FD	-18.1	-6.8	1.9	12.2
W	-1.8	0.2	0.7	2.5
LBA	-8.7	3.1	5.6	3.1
RBA	-1.8	-1.5	-3.3	0.9
ID	-8.7	-5.3	1.7	6.3
CD	-2.2	-1.8	-0.2	3.6
DD	-0.8	0.4	5.9	1.3
Base Prediction	67.7	67.7	67.7	67.7
Prediction	**22.0**	**57.7**	**80.2**	**98.2**
True Value	**17.1**	**59.0**	**81.5**	**100**

It is observed that, FD, LBA, ID and C are the most critical features that influence the predictions of RC beams with a low load-level (see Fig. 13a). These features negatively impact the final prediction by reducing the base value. For this particular sample, FD, LBA ID and C caused a reduction in the base value of about 18.1%, 8.7%, 8.7% and 3.6%, respectively.

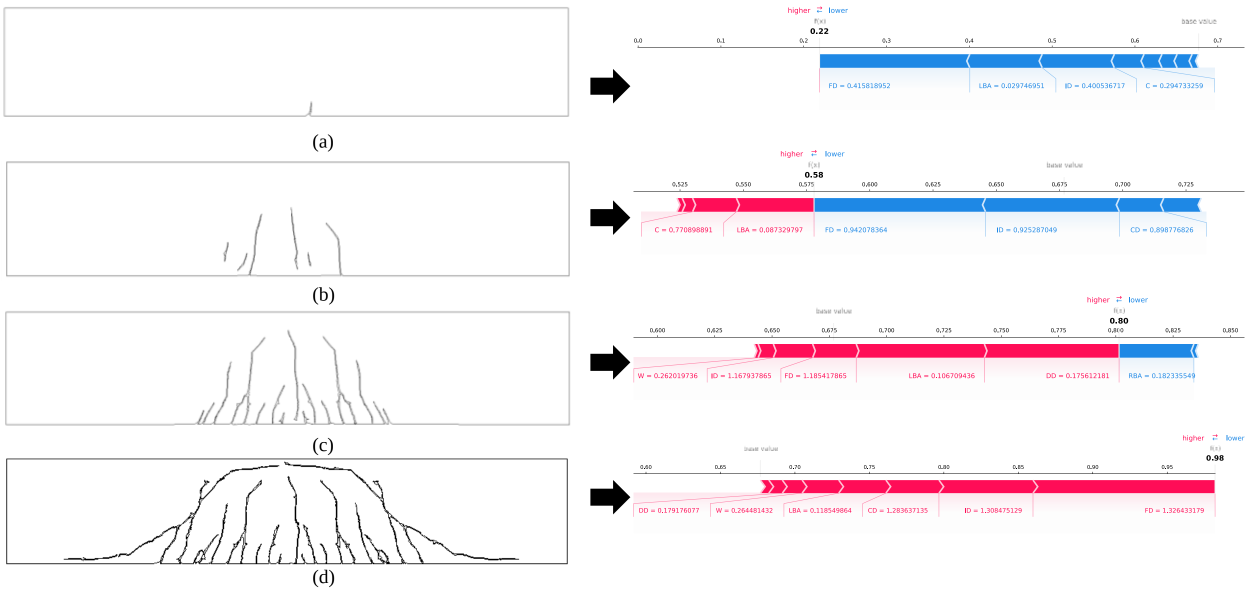


Fig. 13 Local interpretations of selected RC beams with different damage levels: (a) low-Sherwood [54] ; (b) moderate Cao[57] - ; (c) near failure - Cao[57] ; and (d) ultimate failure - Cao[57]. (color printed)

The second sample was selected to depict an instance where the RC beam is moderately damaged. The true and predicted load-level for this sample is 59% and 57.7%, respectively. The SHAP values of each feature for this sample are given in Table 8. Fig. 13b illustrates the critical features that influence the prediction made for this sample. The red bars represent contributions from features that increase the load-level, with the blue bars outlining features that affect the load-level prediction negatively. It is observed that whereas LBA and C reduce the load-level for the slightly damaged beams (Fig. 13a), they rather tend to increase the load-level for moderately damaged RC beams (Fig. 13b). The original values of LBA and C are relatively higher for the moderately damaged beams when compared to the slightly damaged beams, and hence could be a contributing factor to explain this observation (see annotations in Fig. 13a and 13b). As the level of damage of the RC beam increases and approaches failure, the SHAP values for the features assume positive values (Table 8). This is evident in the two other samples which were used to represent near failure and ultimate failure cases (see Table 8 and Fig. 13). The fractal characteristics of these beams produced relatively high values of the original features and hence can partly give a physical reason why the predictions are increased from the base value to the final output. In all cases, FD and ID appears to dominate the most critical features for the four samples considered and either affect the load-level prediction positively or negatively, depending on the level of damage the RC beam in question has sustained.

6.3 Feature Dependency plot

The correlation between SHAP values and features values can give a detailed insight into which scenarios can either cause a decrease or increase in the load-level. Fig.14 shows feature dependency plots to facilitate such analysis. For brevity, the variation of SHAP values for six selected features is presented.

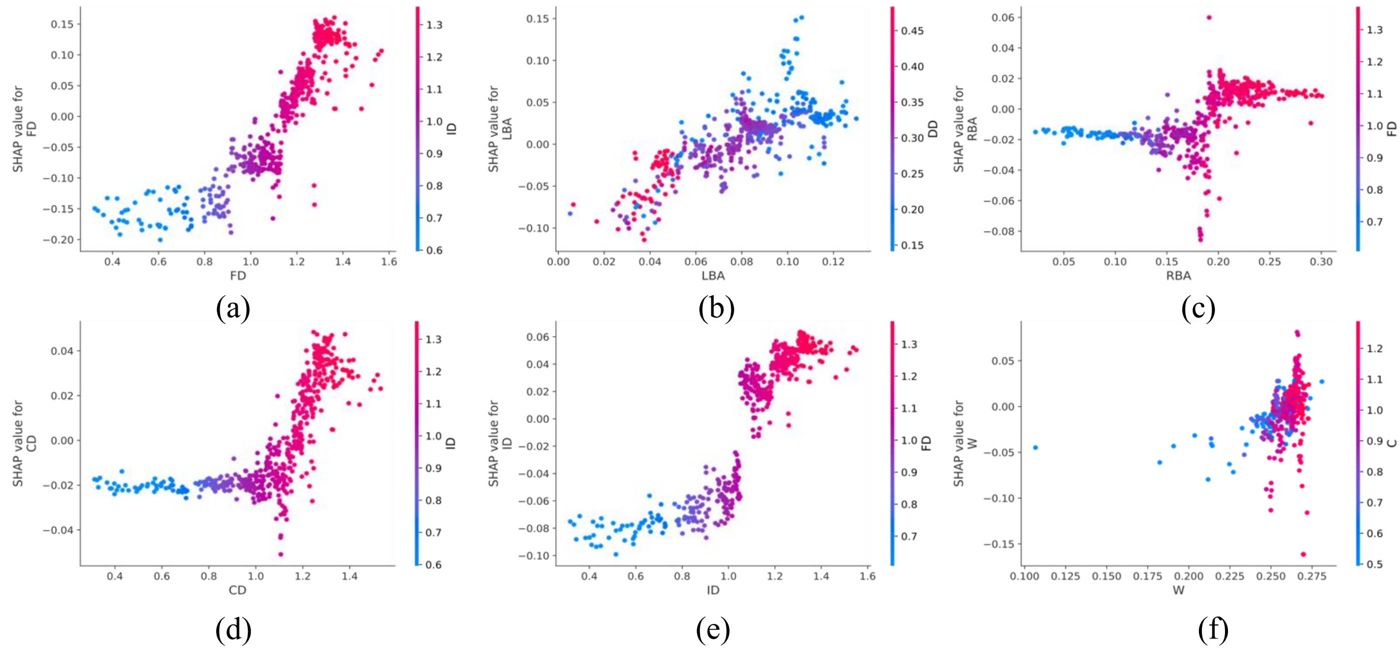


Fig. 14 Plots of feature dependency: (a) FD; (b) LBA; (c) RBA; (d) CD; (e) ID; and (f) W. (color printed)

SHAP values increase with increasing values of FD, LBA, CD and ID. This indicates that FD, LBA, CD and ID are positively correlated with load-level estimation of shear-critical RC beams. From Fig. 14a, RC beams with FD greater than 1.05 tend to cause an increase in load-level. Note that many of these RC beams tend to have values of ID greater than 1.1 (Fig. 14e). Nevertheless, whereas the maximum increase in load-level considering ID is about 6%, FD can contribute an increase of about 16% in load-level (Fig. 14a and 14e). Beams with LBA values greater than 0.05 and DD less than 0.3, do cause an increase in load-level (Fig. 14b). Even though RC beams with CD greater than 1.1 tend to cause an increase in load-level, its contribution is not so significant with a maximum increase of about 4.5%. For W and RBA, the pattern is inconclusive and hence insignificantly affect load-level estimates. Findings from this analysis can be used to develop closed form solutions to load-level estimation for damage assessment of shear-critical RC beams and slabs.

7. Conclusions

This paper explored the application of multifractal analysis to shear-critical RC beams and slabs for load-level estimation. A database of 508 RC beams and slabs were used for model training (70%) and testing (30%). Multifractal analysis was first conducted on images of crack patterns of these beams, with critical features extracted from the singularity and generalized dimension spectra to form the design input matrix in the model development phase, whereas the load-level for each specimen served as the output. The efficiency of four regression-like machine learning models (elastic-net regression (ENR), support vector regression (SVR), random forest regression (RFR) and extreme gradient boosting (XGBoost)) were explored on the dataset. Hyperparameter optimization was conducted for these models using a random search algorithm. For performance measures (root-mean squared error (RMSE), correlation (R), explained variance (EV) and index of agreement (IA)) were used to facilitate model evaluation and selection. Shapley additive explanations (SHAP) was later used for model interpretation. The primary findings from this study are listed below:

- The XGBoost model was the most effective model for estimating the load-level of shear-critical RC beams and slabs. The mean of the predicted-to-tested ratio was 1.04 with coefficient of variation of 27%.
- Upon comparing the XGBoost model with the other models, it was found out that tree-based methods perform significantly better than linear and non-linear methods of regression.
- For model interpretation at the global level, it was revealed by SHAP that the so-called generalized dimensions (fractal dimension (FD), information dimension (ID) and correlation dimension (CD)) which was obtained from the multifractal analysis of the crack patterns considered, had significant impact on the estimation of the load-level, as opposed to the other geometric features acquired from the singularity spectrum. The fractal dimension (FD) was arguably the most critical feature whereas the capacity (C) was the least influential.
- Shear-critical RC beams with FD greater than 1.05 tend to cause an increase in load-level, which can be as high as high as 16%. Even though RC beams with CD greater than 1.1 tend to cause an increase in load-level, its contribution is not so significant with a maximum increase of about 4.5%.
- It was observed that depending on how high or low the original values of the multifractal features are, which is heavily related to the level of damage, the obtained SHAP values will either increase or decrease the load-level estimates. For instance, whereas the area under the left branch (LBA) and C reduce the load-level for the slightly damaged beams (Fig. 14a), they rather tend to increase the load-level for moderately damaged RC beams (Fig. 14b).

To facilitate the practical application of the developed model as well as reproducibility, the source code and database will be made available to the public on a GitHub account. Users may use the proposed model to either get a firsthand insight on the level of damage sustained by such structural elements in service, before another sophisticated framework can be applied.

8. Limitations

Despite the successful development of the structural load estimation model based on multifractal features, some limitations have been identified. The present study only considers RC beams and slabs that have been designed to exhibit shear dominant failure. In order words, the developed model is not generally applicable, as it cannot be utilized for other structural failure phenomena. Future studies should continuously explore the combined application of machine-learning and multifractal analysis to other modes of structural failure, type of RC element and loading conditions. This could assist in the development of a unified model for structural load level estimation for a wide variety of RC structural elements. Secondly, valuable damage parameters on crack patterns such concrete spalling and crack width were not considered in the present study. The generalization error of the develop model can be improved if information relating these parameters are provided and well documented. Therefore, future experimental testing programs should grant the research community access to raw data if possible. Despite these limitations, findings from this research have revealed the need for continuous research in the application of machine-learning based multifractal analysis of reinforced concrete structures for structural load-level assessment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

9. References

[1] Athanasiou A, Ebrahimkhanlou A, Zaborac J, Hrynyk T, Salamone S. A machine learning approach based on multifractal features for crack assessment of reinforced concrete shells. Comput-Aided Civ Infrastruct Eng 2020;35:565–78. https://doi.org/10.1111/mice.12509.

[2] Davoudi R, Miller GR, Kutz JN. Structural Load Estimation Using Machine Vision and Surface Crack Patterns for Shear-Critical RC Beams and Slabs. J Comput Civ Eng 2018;32:04018024. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000766.

[3] Rens KL, Wipf TJ, Klaiber FW. Review of Nondestructive Evaluation Techniques of Civil Infrastructure. J Perform Constr Facil 1997;11:152–60. https://doi.org/10.1061/(ASCE)0887-3828(1997)11:4(152).

[4] Cha Y-J, Choi W, Büyüköztürk O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Comput-Aided Civ Infrastruct Eng 2017;32:361–78. https://doi.org/10.1111/mice.12263.

[5] Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, et al. Automated Crack Detection on Concrete Bridges. IEEE Trans Autom Sci Eng 2016;13:591–9. https://doi.org/10.1109/TASE.2014.2354314.

[6] Guidebook on Non-destructive Testing of Concrete Structures 2019. https://www.iaea.org/publications/6347/guidebook-on-non-destructive-testing-of-concrete-structures (accessed March 16, 2022).

[7] Maffei J, Comartin CD, Kehoe B, Kingsley GR, Lizundia B. Evaluation of Earthquake-Damaged Concrete and Masonry Wall Buildings. Earthq Spectra 2000;16:263–83. https://doi.org/10.1193/1.1586111.

[8] Hartle RA, Ryan TW, Mann E, Danovich LJ, Sosko WB, Bouscher JW. Bridge Inspector's Reference Manual: Volume 1 and Volume 2. 2002.

[9] Yumpu.com. ATC-20-1 Field Manual: Postearthquake Safety Evaluation of ... YumpuCom n.d. https://www.yumpu.com/en/document/view/47765334/atc-20-1-field-manual-postearthquake-safety-evaluation-of- (accessed March 16, 2022).

[10] Ayub M. Investigation of March 15, 2018 Pedestrian Bridge Collapse at Florida International University, Miami, FL. Washington, D.C.: US Department of Labour; 2019.

[11] ACI PRC-201.1-08 Guide for Conducting a Visual Inspection of Concrete in Service n.d. https://www.concrete.org/store/productdetail.aspx?ItemID=201108&Language=English&Units=US_AND_METRIC (accessed March 23, 2022).

[12] Manual for Bridge Element Inspection (2nd Edition) - Vanderbilt University n.d. https://catalog.library.vanderbilt.edu/discovery/fulldisplay/alma991043715690603276/01VAN_INST:vanui (accessed March 23, 2022).
[13] Madani HM, Dolatshahi KM. Strength and stiffness estimation of damaged reinforced concrete shear walls using crack patterns. Struct Control Health Monit 2020;27:e2494. https://doi.org/10.1002/stc.2494.
[14] Tan X, Bao Y. Measuring crack width using a distributed fiber optic sensor based on optical frequency domain reflectometry. Measurement 2021;172:108945. https://doi.org/10.1016/j.measurement.2020.108945.
[15] Felice MV, Velichko A, Wilcox PD. Accurate depth measurement of small surface-breaking cracks using an ultrasonic array post-processing technique. NDT E Int 2014;68:105–12. https://doi.org/10.1016/j.ndteint.2014.08.004.
[16] Hamia R, Cordier C, Dolabdjian C. Eddy-current non-destructive testing system for the determination of crack orientation. NDT E Int 2014;61:24–8. https://doi.org/10.1016/j.ndteint.2013.09.005.
[17] Aggelis DG, Kordatos EZ, Soulioti DV, Matikas TE. Combined use of thermography and ultrasound for the characterization of subsurface cracks in concrete. Constr Build Mater 2010;24:1888–97. https://doi.org/10.1016/j.conbuildmat.2010.04.014.
[18] Huang J, Liu W, Sun X. A Pavement Crack Detection Method Combining 2D with 3D Information Based on Dempster-Shafer Theory. Comput-Aided Civ Infrastruct Eng 2014;29.
[19] Park SE, Eem S-H, Jeon H. Concrete crack detection and quantification using deep learning and structured light. Constr Build Mater 2020;252:119096. https://doi.org/10.1016/j.conbuildmat.2020.119096.
[20] Yamaguchi T, Hashimoto S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. Mach Vis Appl 2010;21:797–809. https://doi.org/10.1007/s00138-009-0189-8.
[21] Yeum CM, Dyke SJ. Vision-Based Automated Crack Detection for Bridge Inspection. Comput-Aided Civ Infrastruct Eng 2015;30:759–70. https://doi.org/10.1111/mice.12141.
[22] Kong X, Li J. Vision-Based Fatigue Crack Detection of Steel Structures Using Video Feature Tracking. Comput-Aided Civ Infrastruct Eng 2018;33:783–99. https://doi.org/10.1111/mice.12353.
[23] Wu R-T, Singla A, Jahanshahi MR, Bertino E, Ko BJ, Verma D. Pruning deep convolutional neural networks for efficient edge computing in condition assessment of infrastructures. Comput-Aided Civ Infrastruct Eng 2019;34:774–89. https://doi.org/10.1111/mice.12449.
[24] Destrebecq J-F, Toussaint E, Ferrier E. Analysis of Cracks and Deformations in a Full Scale Reinforced Concrete Beam Using a Digital Image Correlation Technique. Exp Mech 2011;51:879–90. https://doi.org/10.1007/s11340-010-9384-9.
[25] Ebrahimkhanlou A, Salamone S. A probabilistic model for visual inspection of concrete shear walls. Sens. Smart Struct. Technol. Civ. Mech. Aerosp. Syst. 2017, vol. 10168, SPIE; 2017, p. 247–53. https://doi.org/10.1117/12.2258614.
[26] Sohn H-G, Lim Y-M, Yun K-H, Kim G-H. Monitoring Crack Changes in Concrete Structures. Comput-Aided Civ Infrastruct Eng 2005;20:52–61. https://doi.org/10.1111/j.1467-8667.2005.00376.x.
[27] Lin Y, Nie Z, Ma H. Structural Damage Detection with Automatic Feature-Extraction through Deep Learning. Comput-Aided Civ Infrastruct Eng 2017;32:1025–46. https://doi.org/10.1111/mice.12313.
[28] Jahanshahi MR, Kelly JS, Masri SF, Sukhatme GS. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. Struct Infrastruct Eng 2009;5:455–86. https://doi.org/10.1080/15732470801945930.
[29] Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv Eng Inform 2015;29:196–210. https://doi.org/10.1016/j.aei.2015.01.008.
[30] Ito A, Aoki Y, Hashimoto S. Accurate extraction and measurement of fine cracks from concrete block surface image. IEEE 2002 28th Annu. Conf. Ind. Electron. Soc. IECON 02, vol. 3, 2002, p. 2202–7 vol.3. https://doi.org/10.1109/IECON.2002.1185314.

[31] Fathalla E, Tanaka Y, Maekawa K. Remaining fatigue life assessment of in-service road bridge decks based upon artificial neural networks. Eng Struct 2018;171:602–16. https://doi.org/10.1016/j.engstruct.2018.05.122.
[32] Davoudi R, Miller GR, Kutz JN. Data-driven vision-based inspection for reinforced concrete beams and slabs: Quantitative damage and load estimation. Autom Constr 2018;96:292–309. https://doi.org/10.1016/j.autcon.2018.09.024.
[33] Farhidzadeh A, Dehghan-Niri E, Moustafa A, Salamone S, Whittaker A. Damage Assessment of Reinforced Concrete Structures Using Fractal Analysis of Residual Crack Patterns. Exp Mech 2013;53:1607–19. https://doi.org/10.1007/s11340-013-9769-7.
[34] Carrillo J, Dominguez D, Garcia NP. Damage Index Based on Fractal Dimension of Cracking on Thin Reinforced Concrete Walls. Struct J 2017;114. https://doi.org/10.14359/51700919.
[35] Liu Y, Dai K, Li D, Luo M, Liu Y, Shi Y, et al. Structural performance assessment of concrete components based on fractal information of cracks. J Build Eng 2021;43:103177. https://doi.org/10.1016/j.jobe.2021.103177.
[36] Mandelbrot B. Les objets fractals. 4th ed. Paris: ChampsFlamarion,: 1995.
[37] Teles S, Lopes AR, Ribeiro MB. Fractal Analysis of the UltraVISTA Galaxy Survey. Phys Lett B 2021;813:136034. https://doi.org/10.1016/j.physletb.2020.136034.
[38] Evertsz CJ g. Fractal geometry of financial time series. Fractals 1995;03:609–16. https://doi.org/10.1142/S0218348X95000539.
[39] Multifractal analysis of crack patterns in reinforced concrete shear walls - Arvin Ebrahimkhanlou, Alireza Farhidzadeh, Salvatore Salamone, 2016 n.d. https://journals.sagepub.com/doi/abs/10.1177/1475921715624502 (accessed March 16, 2022).
[40] Ebrahimkhanlou A, Athanasiou A, Hrynyk TD, Bayrak O, Salamone S. Fractal and Multifractal Analysis of Crack Patterns in Prestressed Concrete Girders. J Bridge Eng 2019;24:04019059. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001427.
[41] Raghavendra BS, Narayana Dutt D. A note on fractal dimensions of biomedical waveforms. Comput Biol Med 2009;39:1006–12. https://doi.org/10.1016/j.compbiomed.2009.08.001.
[42] Heymans O, Fissette J, Vico P, Blacher S, Masset D, Brouers F. Is fractal geometry useful in medicine and biomedical sciences? Med Hypotheses 2000;54:360–6. https://doi.org/10.1054/mehy.1999.0848.
[43] Yao B, Imani F, Sakpal AS, Reutzel EW, Yang H. Multifractal Analysis of Image Profiles for the Characterization and Detection of Defects in Additive Manufacturing. J Manuf Sci Eng 2018;140. https://doi.org/10.1115/1.4037891.
[44] Badii R, Politi A. Hausdorff Dimension and Uniformity Factor of Strange Attractors. Phys Rev Lett 1984;52:1661–4. https://doi.org/10.1103/PhysRevLett.52.1661.
[45] Mandelbrot B. Les objets fractals: Forme, hasard et dimension. Paris, France:Flammarion: 1975.
[46] Clarke KC. Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. Comput Geosci 1986;12:713–22. https://doi.org/10.1016/0098-3004(86)90047-6.
[47] Thomas TR, Rosén B-G. Implementation of Whitehouse's method for calculating properties of self-affine fractal profiles. Proc Inst Mech Eng Part C J Mech Eng Sci 2008;222:1547–50. https://doi.org/10.1243/09544062JMES938.
[48] Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI. Fractal measures and their singularities: The characterization of strange sets. Nucl Phys B - Proc Suppl 1987;2:501–11. https://doi.org/10.1016/0920-5632(87)90036-3.
[49] Chhabra null, Jensen null. Direct determination of the f(alpha) singularity spectrum. Phys Rev Lett 1989;62:1327–30. https://doi.org/10.1103/PhysRevLett.62.1327.
[50] Lopes R, Betrouni N. Fractal and multifractal analysis: A review. Med Image Anal 2009;13:634–49. https://doi.org/10.1016/j.media.2009.05.003.
[51] Ebrahimkhanlou A, Farhidzadeh A, Salamone S. Multifractal analysis of two-dimensional images for damage assessment of reinforced concrete structures. In: Lynch JP, editor., San Diego, California, United States: 2015, p. 94351A. https://doi.org/10.1117/12.2084052.

[52] Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition. n.d. https://hastie.su.domains/ElemStatLearn/ (accessed April 18, 2022).
[53] Mangalathu S, Jeon J-S, DesRoches R. Critical uncertainty parameters influencing seismic performance of bridges using Lasso regression. Earthq Eng Struct Dyn 2018;47:784–801. https://doi.org/10.1002/eqe.2991.
[54] Hastie T, Tibshirani R, Friedman J. Linear Methods for Regression. In: Hastie T, Tibshirani R, Friedman J, editors. Elem. Stat. Learn. Data Min. Inference Predict., New York, NY: Springer; 2009, p. 43–99. https://doi.org/10.1007/978-0-387-84858-7_3.
[55] Hastie T, Tibshirani R, Friedman J. Support Vector Machines and Flexible Discriminants. In: Hastie T, Tibshirani R, Friedman J, editors. Elem. Stat. Learn. Data Min. Inference Predict., New York, NY: Springer; 2009, p. 417–58. https://doi.org/10.1007/978-0-387-84858-7_12.
[56] Feng D-C, Wang W-J, Mangalathu S, Taciroglu E. Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. J Struct Eng 2021;147:04021173. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115.
[57] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., New York, NY, USA: Association for Computing Machinery; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.
[58] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier 2016. https://doi.org/10.48550/arXiv.1602.04938.
[59] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 2014;41:647–65. https://doi.org/10.1007/s10115-013-0679-x.
[60] Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. ArXiv170507874 Cs Stat 2017.
[61] Rice JA. Mathematical Statistics and Data Analysis. Cengage Learning; 2006.
[62] Adom-Asamoah M, Banahene JO. Nonlinear seismic analysis of a super 13-element reinforced concrete beam-column joint model. Earthq Struct 2016;11:905–24. https://doi.org/10.12989/EAS.2016.11.5.905.
[63] Dai K, Li D, Luo M, Shi Y, Zhang S, Huang Z. Machine vision-based concrete beam crack pattern identification using fractal theory. Nondestruct. Charact. Monit. Adv. Mater. Aerosp. Civ. Infrastruct. Transp. XIII, vol. 10971, SPIE; 2019, p. 232–7. https://doi.org/10.1117/12.2514124.

RESPONSE TO REVIEWER'S COMMENT

We value the time spent in reviewing our manuscript. We believe that the technical comments raised will help improve on our current submission and enlighten us on areas we couldn't capture well. Below are the responses to the comments raised.

Reviewer One

Comment Number 1: It seems that the predictors utilized are extracted from images, while the response/output is the ratio of nominal applied shear during loading to the one at failure (i.e., V/Vfailure). I wonder if the definition "failure ratio" for output is appropriate under this circumstance? In addition, how do authors determine the value of V/Vfailure from image? More importantly, it seems that the input predictors extracted from images do not have physical meaning. If so, how do authors apply their proposed model for future prediction only based on images?

Response: The paper primarily sought to develop a model for estimating the internal load levels in shear critical RC concrete beams and slabs using images of their crack patterns upon loading. To that end, all RC specimens used in the analysis were designed to have shear dominant failure. For each experimental program, a captured sequence of images that are linked to their recorded load levels were obtained for each specimen. Because these images were captured at multiple load levels, we are trying to explore patterns between the images and the loading, as well as correlations between them. Specifically, we intend to predict how close a specimen is to failure based on the captured image.

We have used ***V/Vfailure*** to quantify what we mean by "how close a specimen is to failure" because all specimens under consideration were shear-critical. For clarity we have edited the definition of V***/Vfailure*** as "load level" instead of "failure ratio".

With regards to your question on how we determined the value of V/Vfailure from images, it worth noting that they were load levels that were recorded when capturing such images.

With regards to the question on the application of this model for future prediction, the present research findings as communicated in this manuscript is based solely on utilizing multifractal features for damage level evaluation. The inputs features can be considered as additional piece of information that can assist in the quantifying the load level sustained by a shear-critical RC beams or slabs. The model can be used as it is or combined with other sophisticated approaches that utilizes some physical and measurable design parameters.

Comment Number 2: The validation and comparison of ML models should be rearranged to comprehensively consider the sensitivities of ML algorithms to training and testing sets. Different training and testing sets applied to ML algorithms may generate totally different results. It seems that the authors only split training and testing sets once and use the results for the final comparison and decision-making. Please implement your method (i.e., Figure 7) at least ten times with different random splits to consider the sensitivities of ML algorithms to the split of training and testing sets.

Response: The validation and comparison of the ML models were conducted in our original manuscript as submitted. We have considered different training and testing sets as spelt in section 6.1.2. To be precise, 1000 runs on different training and testing sets were conducted. Table 5 and 6 presents results on their comparison and subsequent decision-making process.

Comment number 3: Please add a section to discuss the limitations of your proposed method in a VERY detailed way.

Response: A section on assumptions and limitations has been inserted in the revised manuscript. See section 8 of revised manuscript.

Reviewer Two

Comment number 1: P.1, ll.28-29: What is "reinforced concrete civil engineering infrastructure system"? The meaning of this term is not clear. Is it "reinforced concrete structure"?

Response: We were generally referring to "civil engineering infrastructure systems". We have made the necessary corrections in the revised manuscript.

Comment number 2: P.2, ll.54-55: Please provide suitable reference for a case of the bridge collapse.

Response: A suitable reference has been cited and also inserted in the list of bibliography. See reference number 10.

Comment number 3: PP.1-2, introduction part: At pages 1-2, authors mentioned the importance about how to assess the damage of existing structures which has cracks. But the main focus of this study is estimating load level of RC members. The damage evaluation and the load level estimation are different, so the introduction part doesn't match with the main content of the paper. It will be better to modify the introduction part.

Response: The introduction part has been modified to address this important difference. See highlight portion of the introductory part.

Comment number 4: Table 1: The loading conditions or structural characteristics should be different for each reference. This point should be summarized, because they are strongly affect the cracking behavior during loading.

Response: Table 1 has been edited to capture the suggestions given above.

Comment number 5: Table 2: It is recommended that the multifractal features of each experiment from references are shown separately, because different experimental condition (such as 3-point loading or 4-point loading, deep beam or not deep beam) can have largely different cracking patterns and it can result in different behavior of multifractal features.

Response: Table 2 has been edited to capture the suggestions given above.

Comment number 6: Chapter 4: Is the information of shear cracks extracted effectively from the crack patterns? Because most of the cracks in the crack pattern shown in Fig.3 are bending cracks which is less related to shear capacity of RC members, we are not sure whether it is effective to use the whole crack patterns to estimate the load level. Especially, at low load level, all cracks should be bending cracks. Even if all cracks are bending crack, is the information used to estimate load level for shear failure?

Response: The paper primarily sought to develop a model for estimating the internal load levels in shear critical RC concrete beams and slabs using images of their crack patterns upon loading. As we rightly acknowledge, at low levels, all cracks can be considered to be tensile cracks. Nevertheless, all RC specimens used in the analysis were designed to have shear dominant failure. The intention is to estimate how close these specimens are to failure, and as such when the develop model outputs a low-level prediction, for which bending cracks dominate, its relevance may not be that significant in this context.

Comment number 7: P.29, section 6.2.2: From which reference were four RC beams selected? And, if possible, please provide crack patterns and the value of multifractal features for four cases. It will be helpful for better understanding of behaviors.

Response: The crack patterns and values of multifractal features are presented in Fig. 13. The appropriate references have been inserted in this figure's caption.

Comment number 8: References: Please provide necessary information for references. For example, the information for[55]-[58] is poor.

Response: Thank you for drawing our attention to this. We have added extra information to such references.

Highlights

- A novel structural load-level estimation model for RC beams using multifractal features is developed.
- Multifractal features from both singularity and generalized dimension spectra are extracted for model development.
- SHapley Additive explanation (SHAP) is used for model interpretation.
- Dataset spans reinforced concrete beams and slabs without transverse reinforcements.

AUTHORSHIP STATEMENT

A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis

Authorship contributions

Jack Banahene Osei

Conceptualization, Data curation, Formal analysis, Methodology, Writing - original draft

Mark Adom-Asamoah

Validation, Investigation, Writing - review & editing

Jones Owusu Twumasi

Data curation, Validation, Writing - original draft, Writing - review & editing

Peter Andras

Formal analysis, Validation, Investigation, Writing - review & editing

Hexin Zhang

Formal analysis, Validation, Writing - original draft, Writing - review & editing

This statement is signed by the corresponding authors

Author's name	Author's signature	Date
Jack Banahene Osei		21/9/2022

A Machine Learning-based Structural Load Estimation Model for Shear-Critical RC Beams and Slabs using Multifractal Analysis

Jack Banahene Osei[a], Mark Adom-Asamoah[a], Jones Owusu Twumasi[a], Peter Andras[b], Hexin Zhang[b]

[a]Department of Civil Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[b]School of Computing, Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, EH10 5DT, Scotland, UK

Coressponding Author: Jack Banahene osei
Email Address: jobanahene.coe@knust.edu.gh

Abstract

This paper presents a machine learning model for load-level estimation for shear-critical reinforced concrete (RC) beams and slabs using multifractal features of their characteristic crack patterns to automate and provide well-informed decisions for RC damage assessment. Multifractal analysis was conducted on a database of 508 images, of which critical features were extracted from the singularity and generalized dimension spectra. These features are used as predictors for the load-level estimation model. The extreme gradient boosting algorithm yielded the best performance among the four machine learning models considered. The mean of the predicted-to-true ratio for the developed model was 1.04 with a coefficient of variation of 0.27. Upon applying Shapley additive explanations, the fractal dimension, information dimension, correlation dimension and the area under the left branch of the singularity spectrum were the critical features influencing load-level estimation. The proposed model can be useful to RC building inspectors.

Keywords: Multifractal analysis; load-level assessment; beams and slabs; machine learning; score analysis

1. Introduction

The performance characteristics of many civil engineering infrastructure systems play a dominant role in structural safety evaluation [1], as well as public safety [2]. In practice, evaluating the service performance of such systems is typically facilitated by non-destructive techniques. Visual inspection techniques remain one of the most widely used approaches for the non-destructive evaluation of such systems [3]. They are used in many contexts, including but not limited to structural condition monitoring and damage assessment. The results from such techniques usually give a firsthand insight into whether the infrastructure should be repaired or replaced, or an estimate of the remaining life of the system at both local and global levels. For reinforced concrete (RC) structures, the available visual inspection techniques heavily rely on patterns in concrete cracking and propagation (width, length and orientation), spanning a significant period of deterioration. This methodology has been fairly justified since characteristic crack patterns of RC structures can be used as a proxy to ascertain the stress and strain levels induced in the system during deterioration. In other words, they become a valuable piece of information during load-level assessment of reinforced concrete structures. Typically, the damage assessment and structural condition monitoring phase of RC structures as done by visual inspectors is conducted in three stages; (1) using crack detection equipment such as lidars to determine their locations; (2) documenting the damage by capturing images of the cracked regions [4,5] and (3) and determining the internal load levels and damage states of inspected elements. One notable and conventional way to assess the damage of RC structures has been to augment data from crack pattern characterization and analysis, with condition rating systems[6–8]. Nevertheless, with regards to either estimating the residual strength of an RC member/structure (load-level assessment) or categorizing the extent of structural damage (damage assessment), condition rating systems oftentimes result in a qualitative assessment and hence do not necessarily provide building inspectors with the necessary information [7,9]. In particular, guidelines on condition rating of civil infrastructure systems allows for engineering judgment to be used in damage evaluation, hence subjective and highly reliant on the experience of the inspector[1]. With regards to documentation during the assessment phase, visual inspectors do take a considerable amount of time to complete such tasks, and therefore can causes delays. A case in point is the bridge collapse at the Florida International University [10], where although damage documentation was conducted, results were not accessible in a timely manner to aid in collapse prevention and mitigation. Hence, a major drawback of the application of this visual inspection approach has been it's time-consuming nature (damage documentation) amidst subjectivity in making well-informed decisions. To this end, the relevance of developing automated infrastructure inspection methods for load-level and damage assessment of RC structures has presented itself an interesting area of research.

Structural design and industrial guidelines such as ACI [11] , IAEA [6] and AASHTO [12] make available procedures for load-level and damage evaluation of RC components via crack analysis. A real-world application of how crack patterns can be used to predict the strength and stiffness characteristics of RC shear walls that were damaged during an earthquake was conducted by Madani and Dolatshahi [13]. A significant number of research efforts [5,14–22] have been conducted on crack detection and measurement, which is one of the key stages in crack analysis. Nonetheless, the task of using information (width, length, orientation and number of cracks) obtained from crack analysis to correlate the level of damage still remains a challenge with research efforts still at an early stage. In recent times, artificial intelligence-based data-driven techniques keep transforming the field of structural engineering. To this end, automated computer-aided visual inspection approaches have been developed for the identification and characterization of structural damage of RC structures through crack assessment [4,5,20,21,23–27]. These approaches are heavily reliant on two fields: machine learning and computer vision. The fundamental problem of image segmentation (automatically retrieving cracks from images), coming from the computer vision perspective, for RC members has been studied extensively in recent times [28–30]. This has made it possible to extend machine learning algorithms to quantitatively predict the level of damage of many RC structural components. For instance, Ebrahimkhanlou [25] developed a probabilistic graphical model (Bayesian Belief Network) that could visually evaluate the extent of damage of an RC shear wall and also prognosticate the most likely mode of failure for such members. Fatigue life evaluation of bridge deck was presented Fathalla [31] by using an artificial neural network. Davoudi et al. [2,32] employed computer-vision-based inspection methodologies for quantitative damage and load estimation of RC beams and slabs.

The theory of fractals has been extensively applied in the field of structural engineering for performance evaluation and damage assessment of RC components. Farhidzadeh et al.[33] reported that the extent of structural damage of an RC shear wall under reverse-cyclic loading can be quantified from the fractal characteristic of their crack patterns. Experimental validation of how fractal characteristics of surface cracks of RC members can be utilized in damage classification was investigated by Carrillo et al. [34]. Athanasiou et al. [1] and Liu et al [35] have recently developed data-driven machine learning models for damage classification of RC shells using multifractal and fractal analysis respectively.

The present work seeks to extend the application of multifractal analysis of crack patterns in damage evaluation of shear-critical monotonically-loaded simply-supported RC beams and one-way slabs. In order to facilitate this, a database of segmented images of shear-critical RC beams and slabs as compiled by Davoudi et al. [2] is utilized. In particular, this study builds on the work done by Athanasiou et al. [1] that explored the utilization of multifractal features for damage evaluation of RC shells. The singularity spectrum (a parabolic curve, concave in nature) remains the most dominant output of any multifractal analysis. As shown in Athanasiou et al., [1] geometric features of the singularity spectrum can be extracted and utilized as inputs in a machine learning-based damage classification model of RC shells, with significant accuracy. Although four candidate multifractal features (peak, width, and the area under the left and right branch of the singularity spectrum) were used in their approach, which was seemly motivated by trying to reduce the dimensionality of the model, the authors could not exhaust all potential features that can be obtained from the multifractal analysis, which could equally impact the damage evaluation process positively. The primary distinction in the present study is on the identification of the critical multifractal features relating to both geometry and dimensionality of the basic output of multifractal analysis. The secondary distinction is the proposition of a machine learning regression model that utilizes multifractal features for damage evaluation (structural load estimation) of shear-critical simply-supported RC beams and slabs with a monotonic loading protocol, as opposed to the load estimation models developed by Davoudi et al. [2,32] using machine vision. The overall goal motivating this study is to provide an automated model that takes in captured images of RC beams and slabs and can provide a fairly quick estimation of the extent of damage before sophisticated and computationally expensive assessment techniques can be utilized for rigorous cracking assessment of RC structural components.

2. Overview of Fractal Analysis

Fractal theory [36] since its inception in the 1970s has been successful applied in many fields including astrophysics [37], financial engineering[38], structural engineering[33,34,39,40], medicine [41,42] and manufacturing [43]. The theory seeks to characterize the geometry of irregular and complex objects occurring in nature that the classical Euclidean geometry may seem non-applicable. As noted by Mandelbrot [36], ‘clouds are not spheres, mountains are not cones, nor does lightning travel in a straight line. There are two main facets of fractal theory;

dimensionality and self-similarity. The dimensionality concept hinges on the fractal geometry of the object. As popularly known in literature, Euclidean geometry reveals that the topological dimension of a point, straight line and plane is 0, 1 and 2 respectively, without any intermediate values. However, fractal geometry permits the use of fractional or fractal dimensions. To illustrate this, consider the crack pattern of an RC beam in Fig. 1 which has an estimated fractal dimension of 1.4.

Fig. 1 Typical crack pattern of an RC beam with fractal characteristics

The self-similarity property of many fractal objects is related to an observation about how the method of construction of such objects at both local and global scales appear to be identical. Crack patterns of many reinforced concrete structures under both cyclic and monotonic loading have been shown to exhibit this self-similar behavior. An illustrative example is the crack surface of a prestressed RC girder as shown in Fig. 2 [40]. It possesses fractal behavior since the crack patterns contain replicas of itself at microscopical and macroscopical scales. In order words, if one zooms in or out the crack surfaces (Fig. 2), the geometrical shape has similar appearance. If there exist more than one replica of this self-similarity charateristics, the considered crack pattern is categorized as a multifractal crack pattern. Other technical background for categorizing a digital image as either having monofractal or multifractal characterisitics is discussed below. Nevertheless, for this particular example, since there exists some form of self-similarity at more than one location, there is reason to believe that the crack patterns have multifractal characteristics.

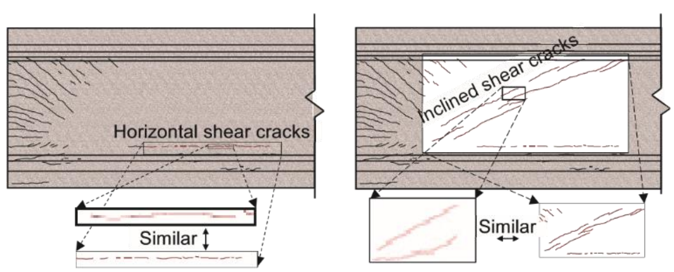


Fig. 2 Self-similarity of RC cracks

2.1 Monofractal Analysis

Several implementation procedures exist for conducting monofractal analysis of images, for fractal dimension determination [44–46]. The box-counting algorithm being the most popular is used in this study. In its abstract form, the fractal analysis seeks to establish the relation between two quantities; the scaling factor, ε, and the number of coverings, $N(\varepsilon)$ of the fractal set, for instance, a digital image profile. Eq. 1 provides the power law relationship that exists between these two quantities.

$$N(\varepsilon) \propto \varepsilon^{-D} \tag{1}$$

where D denotes the fractal dimension. However, for the box-counting algorithm, the scaling factor, ε, is approximated with the size of the box (a) used in discretizing the image pattern. The number of boxes that contains at least an active pixel ($N(a)$) is also used as a proxy for the number of coverings ($N(\varepsilon)$). Linearizing the power law from Eq. 1, the fractal dimension, D, as per the box-counting algorithm, can be computed as:

$$D = \lim_{a \to 0} \frac{\log(N(a))}{\log(1/a)} \tag{2}$$

Alternatively, D can be estimated from the gradient between the number of boxes that contains at least an active pixel, $N(a)$, and the inverse of the box size, a, in the logarithmic space. Fractal dimension, D, depicts the global behavior of fractal sets or digital images through the scaling law presented in Eq. 1, and is the primary output of any monofractal analysis. Monofractal analysis typically do not provide the necessary information for quantifying local fractal characterization. There is a possibility that different images with varying levels of complexities, irregularities and roughness, will yield the same fractal dimension, D, when a monofractal anlysis is conducted [43,47]. In such situation, the utilization of a generalized fractal analysis, known as multifractal analysis could be employed to gain much more insight.

2.2 Multfractal Analysis

Multifractal analysis seeks to provide a detailed local description of the fractal characteristics of a digital image profile. The local pixel density of a particular box, $P_i(a)$, in the digital image is first computed as given in Eq. 3.

$$P_i(a) = \frac{N_i(a)}{\sum_i^{N(a)} N_i(a)} \tag{3}$$

where $N_i(a)$ is the number of pixels in the *i*th box. In the special case where the image in question is a crack pattern of an RC element, $P_i(a)$ denotes the crack density. As an illustrative example, consider the crack pattern of a beam shown in Fig. 1.

Using four candidate boxes, the spatial distribution of the pixel intensities (crack density $P_i(a)$) for the above RC beam is presented in Fig. 3. Evidently, the spatial crack density distribution seems to converge to the original crack pattern of the beam when the size of the box decreases.

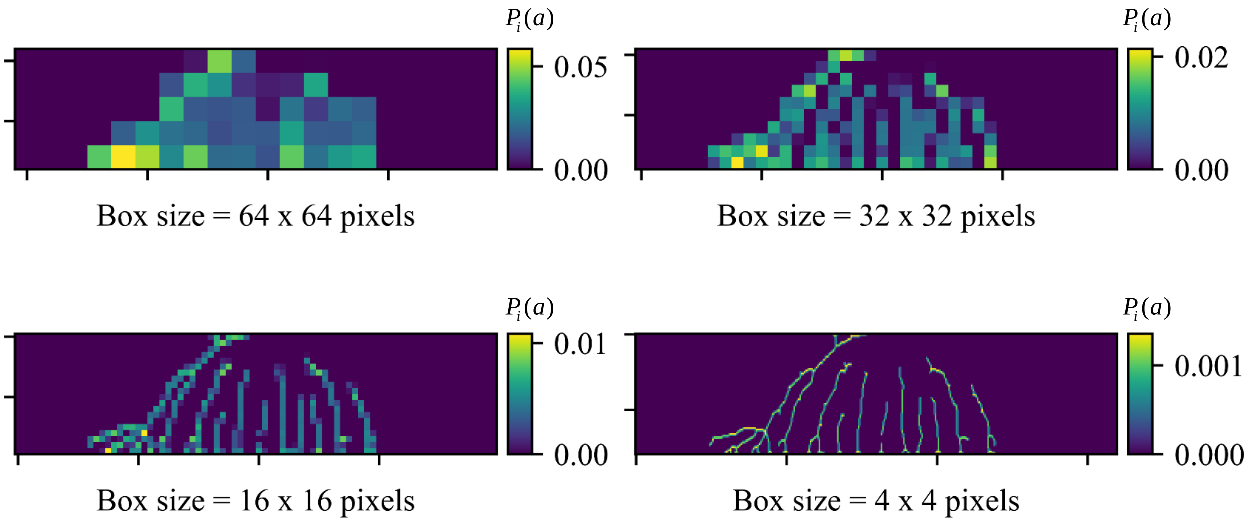


Fig. 3 Spatial distribution of pixel intensities for the crack pattern of an RC beam (color printed)

It turns out that, a similar power law exists between how the pixel density $P_i(a)$ scales, and size of the box a (see Eq. 4).

$$P_i(a) \propto a^{\alpha_i} \tag{4}$$

where α_i is the singularity exponent, depicting the local scaling/fractal behaviour for the ith box. In other words, each box characterized by $P_i(a)$ will have its own singularity exponent α_i. For an infinitesimally small difference $\Delta\alpha$, the number of boxes $N(\alpha)$ for which their singularity exponents fall within the closed interval $[\alpha, \alpha + \Delta\alpha]$ is obtained, and follows a power law with the box size (a), similar to that of Eq. 1.

$$N(\alpha) \propto a^{-f(\alpha)} \tag{5}$$

where $f(\alpha)$ is the fractal dimension of the boxes with the same local scaling α. An $\alpha - f(\alpha)$ plot is commonly called the singularity spectrum is typically used to summarize the output of any multifractal analysis study. The $f(\alpha)$ can be computed from Eq. 6 as:

$$f(\alpha) = \lim_{a \to 0} \frac{\log(N(\alpha))}{\log(1/a)} \tag{6}$$

Traditionally, Legendre Transformation as suggested by Hasley et al. [48] is used to estimate $f(\alpha)$. Nevertheless, a direct numerical approach developed by Chhabra and Jensen [49] is used in this study. It begins with obtaining distorted versions of the spatial distribution of the pixels using the following exponential mapping:

$$P_i(a) \to P_i^q(a) \tag{7}$$

where q it is typically known as the distortion parameter or the order of the probability moment [50]. For a range of values of q ([-5,+5] as recommended by Ebrahimkhanlou et al. [51] for shear-critical RC elements), a normalized form of $P_i^q(a)$ is computed.

$$\mu_i(q, a) = \frac{P_i^q(a)}{\sum_{i=1}^{N(a)} P_i^q(a)} \tag{8}$$

For a given value of q, the singularity exponent $\alpha(q)$ and its corresponding fractal dimension $f(\alpha(q))$ can then be estimated as:

$$\alpha(q) = \lim_{a \to 0} \frac{\sum_{i=1}^{N(a)} \mu_i(q, a) \log(P_i^q(a))}{\log(a)} \tag{9}$$

$$f(\alpha(q)) = \lim_{a \to 0} \frac{\sum_{i=1}^{N(a)} \mu_i(q, a) \log(\mu_i(q, a))}{\log(a)} \tag{10}$$

As already mentioned, a plot of the set of values of α against $f(\alpha)$ for the range of q values, produces the so-called singularity spectrum. Similarly, a $q - \alpha$ plot yields the generalized dimension spectrum. These spectra upon application of multifractal analysis on the crack pattern of the above beam, is shown in Fig. 4 below.

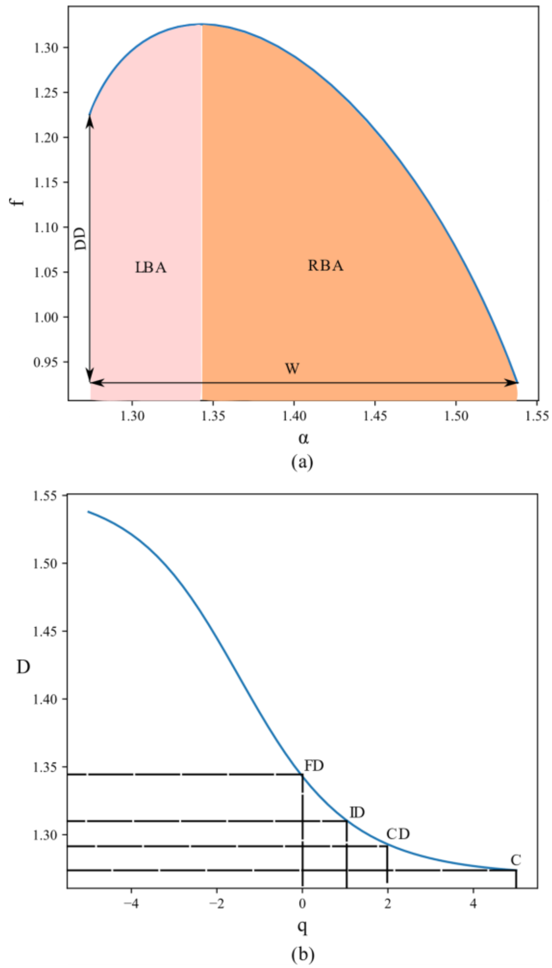


Fig. 4 Key features of (a) the singularity spectrum and (b) generalized dimension spectrum; W: width, FD: fractal dimension, LBA: area under the left branch, RBA: area under the right branch , ID: information dimension, CD: correlation dimension, C: capacity and DD:dimensional difference. (color printed)

3. Extracted features from singularity and generalized dimension spectra

Past research efforts have revealed that specific features that can be extracted from the singularity spectrum of RC shells, can be utilized in structural damage level assessment. In particular, the width (W), area of the left branch (LBA), area of the right branch (RBA) and the peak (FD) of the singularity spectrum have been suggested as critical parameters for damage level identification of RC shells [1] (see Fig. 4a). The geometric width (W) of the singularity spectrum has been deemed to be influential at characterizing RC crack inclination. Generally, the width of the singularity spectrum quantifies the image's heterogeneity. Larger values of the width would usually imply a more severe uneven spatial crack density distribution. In addition, due to the typical asymmetry shape of the singularity spectrum (see Fig. 4a) the area under the left (LBA) and right branch (RBA) of the singularity spectrum has been proven to be key features that influence cracking properties. Also, as noted by Athanasiou et al. [1], the peak of the singularity spectrum (FD) is highly correlated with crack inclination [1].

The fractal dimension (FD) can also be obtained from the generalized dimension spectrum (Fig. 4b) when $q = 0$. Nevertheless, some well know generalized dimensions (D_q ,i.e, α for a particular q) can be candidate features that can significantly characterize the damage performance of RC elements. Information dimension (ID) is the ordinate of the generalized dimension spectrum when $q = 1$. It characterizes the rate at which information contained in the image profile changes with box size. To this end, the information dimension (ID) is explored in this study. The generalized dimensions D_q, corresponding to q > 1 accentuates the more singular regions (regions with significant cracking behaviour), whereas for $q < 1$, reflects the regular regions of the RC crack pattern. The correlation dimension (CD) is also used in this study for RC damage assessment. It quantifies correlation for the heterogeneity of a pair of boxes. The generalized dimension corresponding to the maximum q value is usually referred to as the capacity (C) (see Fig. 4b). The capacity reflects segments of the RC crack patterns with low densities ($P_i(a)$). The capacity, C, can also be used as a proxy for heterogeneity since, larger values signify a higher degree of homogeneity within the singular regions. To this end the capacity (C) is also used in this study. Finally, the dimensional difference (DD) defined as the difference between the fractal dimension of the most singular event $f(\alpha_{\min})$ and the most regular event $f(\alpha_{\max})$ is utilized (see Fig. 4). It reflects the frequency ratio or the proportion of the number of regular regions to singular regions. In summary, eight geometric and generalized dimension multifractal features are extracted from crack patterns of selected shear-critical RC beams and slabs for damage assessment; width (W), peak (FD), area of left (LBA) and right (RBA) branch of the singularity spectrum, information dimension (ID), correlation dimension (CD), capacity (C) and dimensional difference (DD).

4. Image database of RC beams and Slabs

In order to develop a reliable model for structural load estimation, the load-level of RC beams and slabs of an existing database was compiled by Davoudi et al.[2] is utilized in this study. It comprises a variety of experimental programs ranging from uniform to monotonic loading of RC beams and one-way slabs without transverse reinforcement. Table 1 presents a summary of the various independent sources of experimental programs that have been aggregated to form the database used in this study.

To this end, a complied database of the multifractal features considered in this study was presented for 508 RC beams and slabs. The eight multifractal features (see section 3.0) served as input features for the estimation model, whereas the load level (LL) served as the output. LL is defined as:

$$LL = V / V_{failure} \tag{11}$$

where V and $V_{failure}$ represents the nominal applied shear during loading and at failure, respectively. Pragmatic use of the load level (LL) would be to anticipate the degree to which an RC member has been subjected to a load that would cause failure (an LL of 0.7 would imply that the RC member has been given a load of 70% of what it can sustain (capacity)). Some descriptive statistics of both input and output features is presented in Table 2, whereas Fig. 5 and 6 displays statistical distributions, in particular pairwise relationship between some selected features. Each row and column of the matrix of subplots in Fig. 5 and 6 signifies a single feature. The diagonal plots reveal the univariate marginal distribution of a particular feature, whereas the annotations inserted in the upper half of the off-diagonal plots are used to quantify the correlation between two features. All variables were positively correlated with each other, except the LBA and DD which was negatively correlated. The various forms of generalized dimensions are very highly correlated (see Fig. 5), whereas the other features are fairly correlated (see Fig. 6). In order to obtain more insight into how these features could be used to provide a meaningful estimate of the load-level of shear-critical RC beams and slabs, sophisticated machine learning model implementation were explored as opposed to the basic statistical measures presented in Fig. 5 and 6.

Table 1. Summary of experimental testing programs from which database is compiled.

Reference	#S	#I	Test / Specimen Type	*a/d*	*ρ (%)*	*fc'*
Sneed[51]	8	52	3-point load, beam	2.3	0.55-0.85	18.6-32.4
Murray[52]	8	88	3-point load, beam	2.97-3	1.2-1.3	64.8-74.8
McCain[53]	10	82	3-point load, beam	2.3-2.9	0.63-0.98	22.8-33.8
Sherwood[54]	30	197	3-point load, beam & slab	2.79-3.4	0.3-1.33	29.1-77.3
Quach[55]	1	10	3-point load, deep beam	3.1	0.70	40.0
Yoshida[56]	1	4	3-point load, deep beam	2.9	0.70	31.8
Cao[57]	2	12	3-point load, deep beam	2.8-2.9	0.4-1.5	26.2-28.3
Perkins[58]	6	35	Uniform loading	1.62-3.24	0.98	39-64
Nghiep[59]	3	28	3-point load, haunched beam	3-5.0	1.57-3.1	35.4-59.1
Overall	**69**	**508**	-	**1.1-5.0**	**0.3-3.1**	**18.6-77.3**

Note: #S = number of specimens; #I = number of images; *a/d* = shear span-to-depth ratio; *ρ* = tensile reinforcement ratio; *fc'* = compressive strength.

Table 2. Multifractal features of database of RC beams and slabs

Reference	Statistic	FD	ID	CD	C	LBA	RDA	DD	W
	Minimum	0.79	0.79	0.78	0.75	0.03	0.14	0.44	0.25
Sneed[51]	Mean	1.22	1.21	1.20	1.18	0.04	0.24	0.48	0.26
	Maximum	1.45	1.44	1.44	1.41	0.05	0.30	0.55	0.27
	Minimum	0.38	0.36	0.34	0.19	0.03	0.03	0.02	0.19
Murray[52]	Mean	1.03	1.01	0.99	0.88	0.09	0.15	0.19	0.26
	Maximum	1.38	1.36	1.34	1.28	0.13	0.25	0.29	0.27
	Minimum	0.34	0.33	0.33	0.21	0.01	0.02	0.01	0.11
McCain[53]	Mean	1.05	1.04	1.02	0.94	0.07	0.17	0.29	0.26
	Maximum	1.33	1.31	1.29	1.23	0.09	0.23	0.41	0.27
Sherwood[54]	Minimum	0.36	0.34	0.32	0.19	0.02	0.04	0.03	0.20

	Mean	1.12	1.10	1.08	1.00	0.08	0.18	0.26	0.26
	Maximum	1.46	1.44	1.42	1.32	0.13	0.25	0.48	0.27
Quach[55]	Minimum	0.91	0.90	0.89	0.85	0.04	0.15	0.23	0.25
	Mean	1.37	1.36	1.34	1.27	0.10	0.24	0.26	0.27
	Maximum	1.57	1.55	1.53	1.46	0.12	0.28	0.41	0.27
Yoshida[56]	Minimum	0.47	0.45	0.43	0.34	0.03	0.06	0.25	0.25
	Mean	0.93	0.92	0.89	0.81	0.07	0.15	0.26	0.26
	Maximum	1.29	1.27	1.25	1.17	0.10	0.22	0.28	0.26
Cao[57]	Minimum	0.32	0.32	0.31	0.30	0.24	0.01	0.13	0.24
	Mean	1.01	1.00	0.97	0.87	0.26	0.09	0.20	0.26
	Maximum	1.33	1.31	1.29	1.19	0.28	0.12	0.52	0.28
Perkins[58]	Minimum	0.74	0.73	0.70	0.60	0.06	0.11	0.09	0.23
	Mean	1.19	1.17	1.15	1.05	0.10	0.19	0.20	0.26
	Maximum	1.38	1.37	1.34	1.27	0.12	0.24	0.27	0.27
Nghiep[59]	Minimum	0.61	0.59	0.58	0.50	0.04	0.08	0.26	0.23
	Mean	1.16	1.14	1.12	1.05	0.08	0.19	0.30	0.25
	Maximum	1.41	1.40	1.38	1.32	0.10	0.25	0.34	0.26

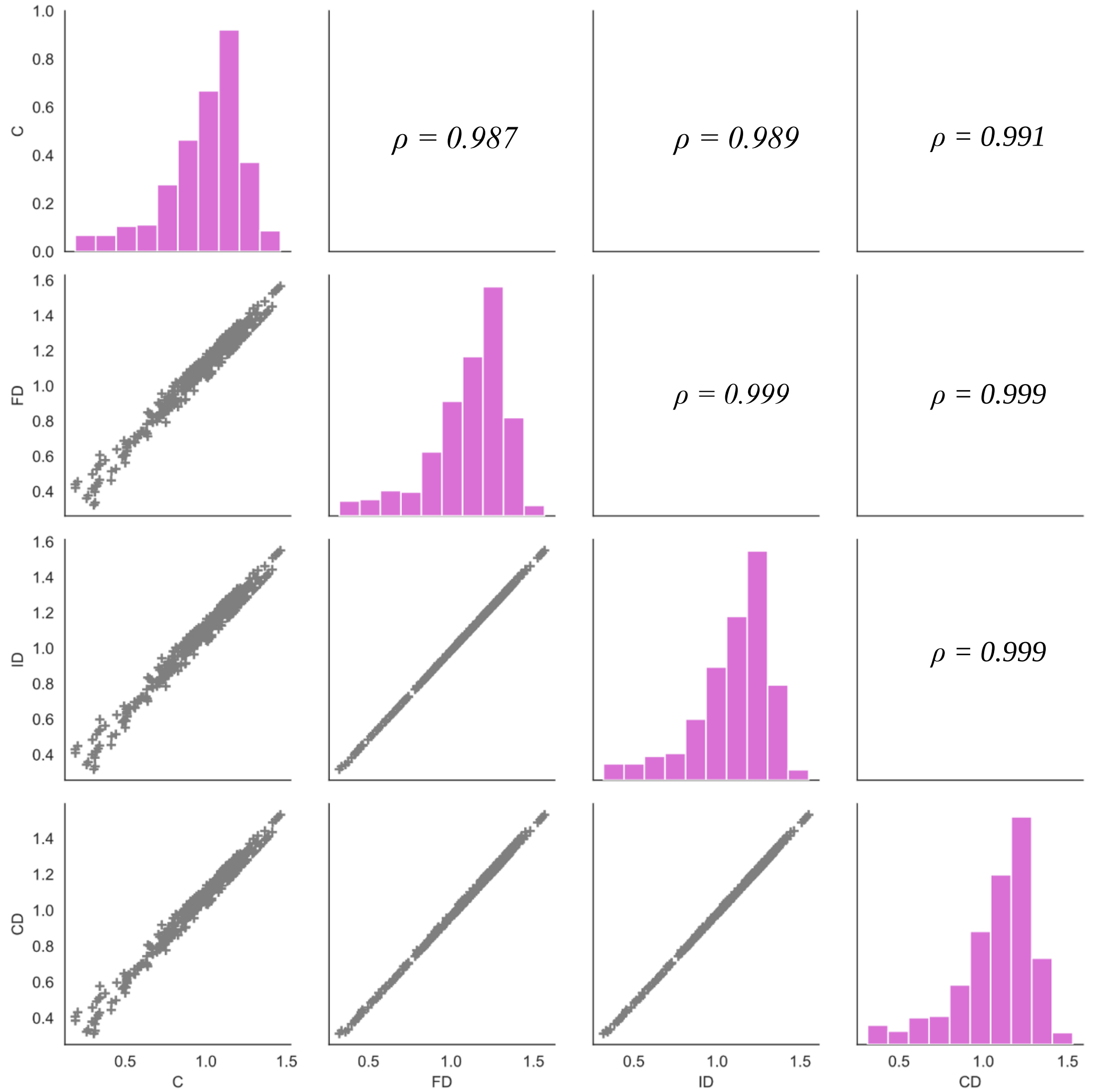


Fig. 5. Pair-plot of input features (C, FD, ID, CD) showing statistical distribution and correlation. (color printed)

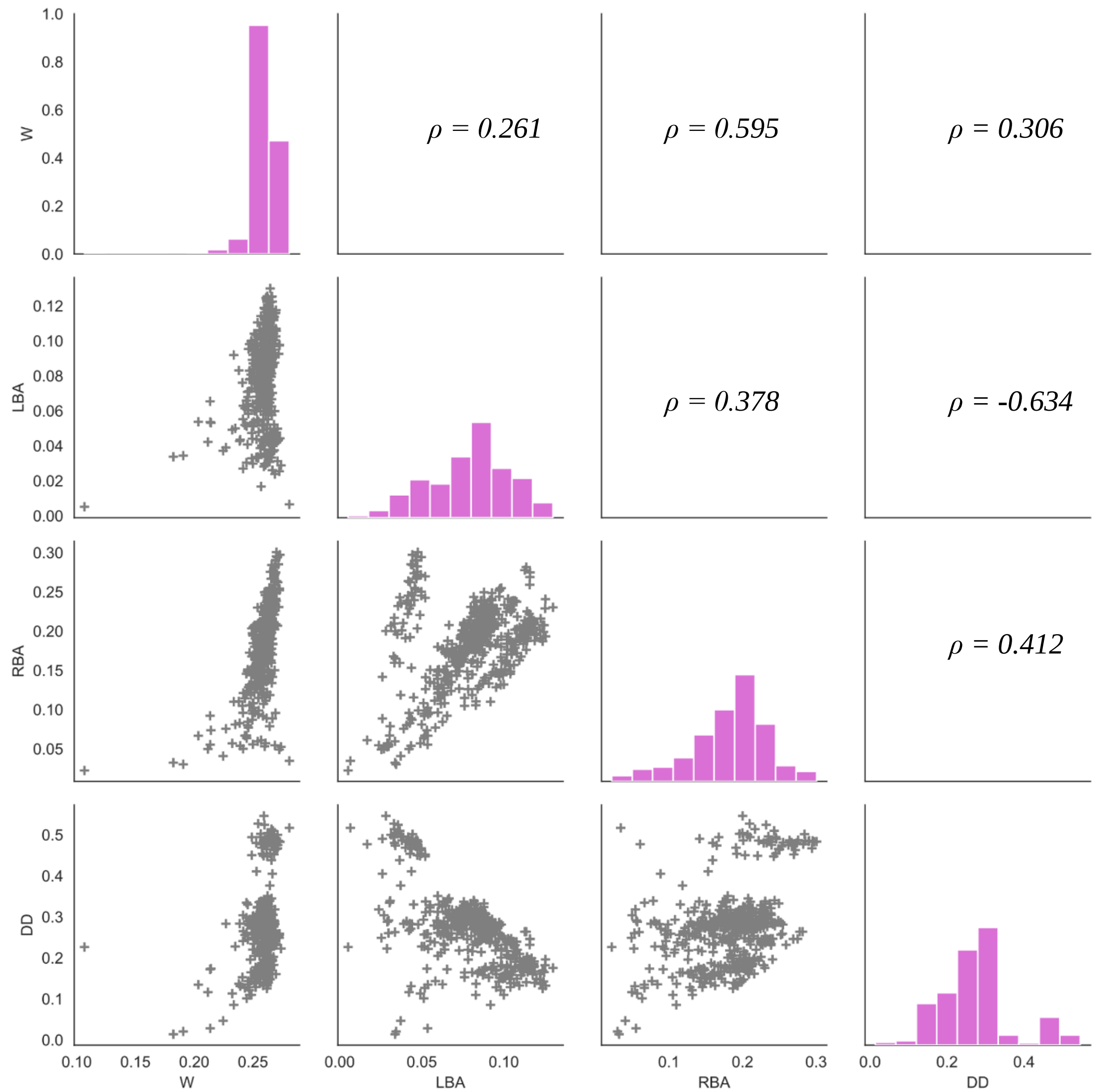


Fig. 6. Pair-plot of input features (W, LBA, RBA, DD) showing statistical distribution and correlation. (color printed)

5. Machine Learning Model Implementation

5.1 Training-Testing Data Splitting

Fig. 7 shows a schematic representation of the proposed machine learning model implementation procedure. Firstly, the image database of RC beams and one-way slabs is split into training and testing data. In this study, random samples of 70% of the entire database was assigned to the training data, whereas the remaining 30% was assigned as testing data. Four regression-like machine learning techniques were implemented using the training data (see Fig. 7). A brief background on these four regression techniques is presented as follows:

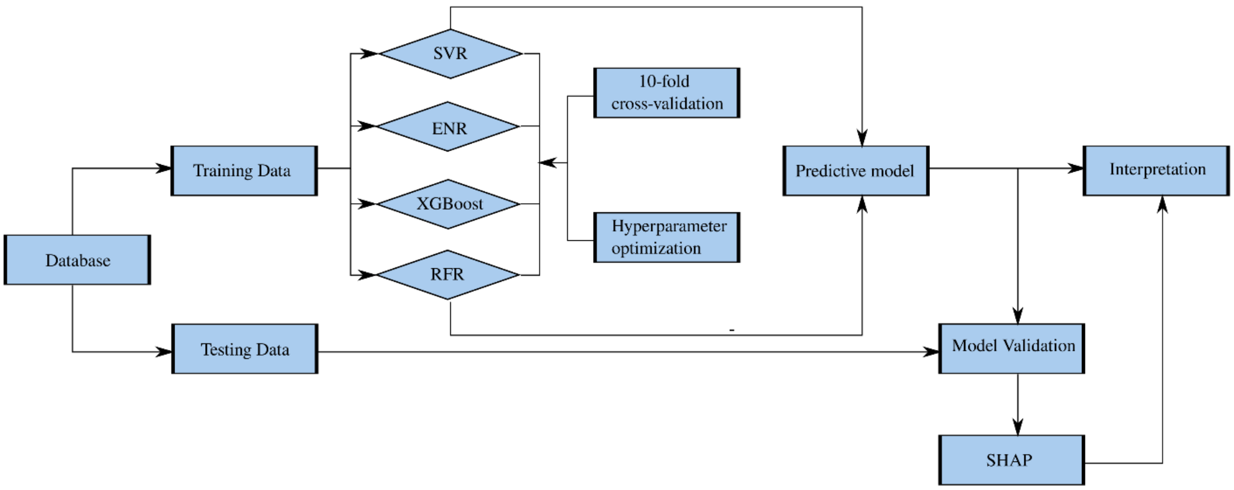


Fig. 7. Machine learning model implementation. (color printed)

5.2 Machine learning Algorithms

In the present context, a predictive model that could map the set of multifractal features into a load level (*FR*) estimate for the database of RC beams and one-way slabs is sought after. The Support Vector Regression (SVR), Random Forest Regression (RFR), linear Elastic-Net Regression (ENR) and the Extreme Gradient Boosting (XGboost) algorithm were adopted in this study. All these machine learning techniques have been successfully employed in solving similar structural engineering-related problems [60–62] which usually comprises a relatively limited number of data points in a dataset.

5.2.1 Elastic-Net regression (ENR)

The basic linear regression model seeks to provide a solution to finding the best fit between a set of input points and an output. In the present context, given an input vector of multifractal features, $X_i = (x_{i1}, x_{i2}, x_{i3}, ...x_{ip})$ and an output load level, LL, of an RC beam or one-way slab, the linear regression model has the following functional form [52]:

$$LL_i = \beta_0 + \sum_{i=j}^{p} \beta_j x_{ij} \tag{12}$$

where β_j are the unknown parameters and p is the number of input features. Given a training dataset $((X_1, LL_1), (X_2, LL_2), (X_3, LL_3), ...(X_N, LL_N))$, β_j are estimated by using the most popular loss function; the sum of squared error (SSE) as given in Eq. 13.

$$SSE(\beta) = \sum_{i=1}^{N} \left[LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right]^2 \tag{13}$$

It turns out that the estimates obtained from minimizing the *SSE*, have the smallest variance for all available linear unbiased estimators [53]. Nevertheless, biased estimators tend to have a fairly relatively low variance compared to their unbiased counterpart. The emphasis of most regression-like machine learning models is to determine model parameters that will reduce the generalization or test error, hence the variance. To this end, the regularized variable selection regression model, Elastic-Net Regression (ENR) is able to mitigate this drawback of the original regression model. It consists of minimizing the aggregate sum of a loss and penalty function. The unknown parameters $\beta_{elastic}$ are estimated from Eq. 14.

$$\beta_{elastic} = \underset{\beta}{\operatorname{argmin}} \left(\sum_{j=1}^{N} \left[LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right]^2 + \lambda \sum_{j=1}^{p} \left[\alpha \beta_j^2 + (1-\alpha) \left| \beta_j \right| \right] \right) \tag{14}$$

The penalty term as seen in Eq. 14, requires the specification of two hyperparameters; λ and α. A comprehensive description of ENR can be found in Hastie et al. [54].

5.2.2 Support Vector Regression (SVR)

The general support vector machine which was originally described to solve classification problems, can be adapted for regression analysis [52]. Similar to the elastic-net model presented above, the algorithm minimizes the following objective function:

$$\beta_{svr} = \underset{\beta}{\operatorname{argmin}} \left(\sum_{i=1}^{N} V_\varepsilon \left(LL_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right) + \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2 \right) \tag{15}$$

$$\text{where } V_\varepsilon(r) = \begin{cases} 0 & \text{if } \left| r \right| < \varepsilon \\ \left| r \right| - \varepsilon, & \text{otherwise} \end{cases} \tag{16}$$

This support-vector formalism is usually referred to as the ϵ-insensitive or error-insensitive SVR model. It basically requires the determination of two hyperparameters, epsilon (ε) and lambda (λ). However, the general minimization problem is solved numerically by making use of kernels after approximating the regression function given in Eq. 12 with a set of basis functions [55]. Some of the widely used kernels are the polynomial, sigmoid, and the gaussian radial basis kernel function. The selection of the most appropriate kernel as well as other hyperparameters is oftentimes determined via cross-validation.

5.2.3 Random Forest Regression (RFR)

Random forest leverages the superiority of considering an ensemble of regression trees for decision making, in this case, predicting a quantitative response value (see Fig. 8). The algorithm begins with bootstrapping a sample from the training data, from which a regression tree that utilizes a random selection of a subset of features can be developed [52]. This procedure is repeated for different bootstrap samples and features. The prediction of unseen or test data can then be computed by taking the mean of the predictions obtained from the various regression trees already developed. Fig. 8 provides a schematic presentation of the Random Forest Regression (RFR) implementation procedure. A couple of hyperparameters influence the performance of an RFR scheme; the number of trees or estimators, maximum depth of tree, and the number of features to select at each split, and the minimum number of samples in each split.

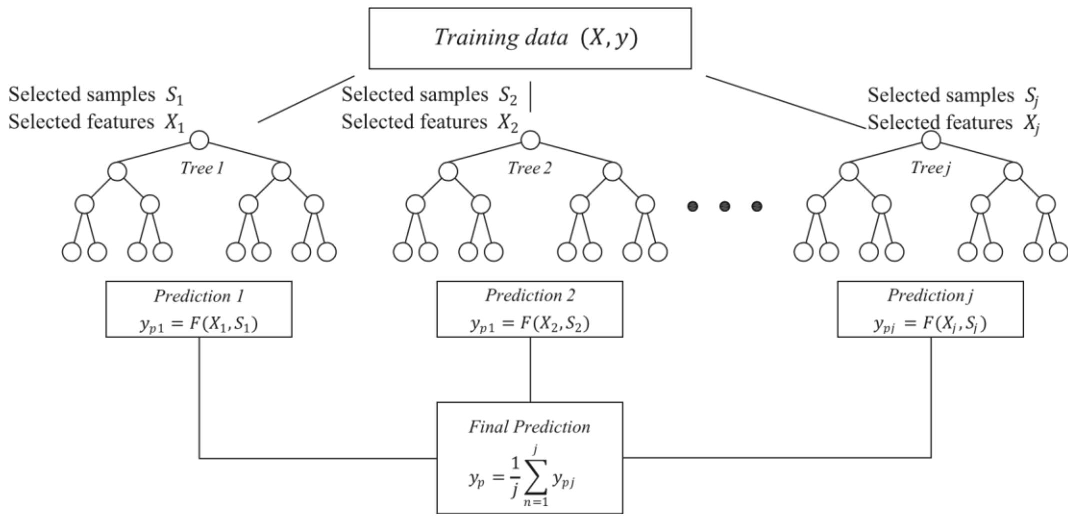


Fig. 8 A random forest regression implementation scheme.

5.2.4 Extreme Gradient Boosting (XGBoost)

This fairly recent developed machine learning technique is an extension of the popular ensemble learning method, gradient descent decision tree [56,57]. The XGBoost aggregate a collection weak learner that is usually obtained from a decision tree model. Whereas random forest regression outputs the mean of different trees, XGBoost incrementally improves the prediction through a weighted aggregation of weak learners to form a strong learner. In this study, decision trees are used as weak learners. The XGBoost regressor seeks to provide a mapping between the input set of features and the output of a training dataset using the following Equation.

$$LL_i = \sum_{k=1}^{K} \sigma_k f_k(X_i) \tag{17}$$

where, K is the number of weak learners or estimators, σ_k is the learning rate, and $f_k(X_i)$ is the weak leaner obtained from a decision tree. In determining the most appropriate learner at a particular stage, and other hyperparameters, the loss and penalty functions that need to be minimized is given in Equation 18 below.

$$f_t = \underset{f \in F}{\operatorname{argmin}} \left[\sum_{i=1}^{N} \left(LL_i - \sum_{k=1}^{t} \sigma_k f_k(X_i) \right)^2 + \sum_{k=1}^{t} \left(\gamma T + \frac{1}{2} \lambda \left\| w_k \right\| \right) \right] \tag{18}$$

where f_t is the weak learner to be determined at the t-th step, γ and λ are the hyperparameters of the penalty term, and T and w_k are the number of leaf nodes and weights, respectively. It is worth noting that, the sequential nature of the XGBoost algorithm only permits the determination of the optimal weak leaner and penalty coefficients at the t-th step (f_t, γ and λ), since all other parameters and learners before the t-th step would have been determined. The output of the regression model is sequentially updated to a point where t equals to K, the number of weak learners to be considered. Further details on how the weak learners with its accompanying hyperparameters are determined can be found elsewhere in Chen and Guestrin [57].

5.3 Hyperparameter Optimization

In the implementation process, a 10-fold cross-validation scheme was utilized in hyperparameter optimization via a random search, in order to determine the best set of parameter combinations for each model training. The performance measure used in determining the optimal hyperparamter was the mean squared error. This analysis is performed for 1000 runs, and the modal values of the hyperparameters that were optimal for each machine learning model is presented in Table 3. As observed, the optimal number of estimators for the random forest and extreme gradient boosting machine were different (see Table 3), after hyperparameter optimization. The number of estimators refers to the number of decision trees that constitutes the meta model. Informed comparisons between these two models can be made since their learning algorithms are different. For instance, whereas random forest assigns equal weight to each decision tree during the aggregation process to make a final prediction, the weighting scheme for the extreme gradient boosting machine model is adjustable or adaptive and depends on the loss function to be minimized. With this inherent difference in the two algorithms, the number of estimators does not have to be necessarily equal to make well-informed comparison during model evaluation.

Table 3. Tuned hyperparameters for various machine learning models

Model	Hyperparameter	Modal Value
SVR	Kernel	Radial Basis
	Epsilon (ε)	0.1
	Lambda (λ)	1000
ENR	Alpha (α)	0.9
	Lambda (λ)	0.001
RFR	Number of Estimators	800
	Maximum depth of tree	6
	Minimum samples for split	3
	Maximum number of features	3
XGboost	Number of Estimators	500
	Learning rate	0.01
	Maximum depth of tree	6

Minimum samples for split	3
Lambda (λ)	0.1
Gamma (γ)	0.1

5.4 Performance Measures

One of the four machine learning models obtained from the training data after hyperparameter optimization was then selected as the final proposed predictive model. In order to make valuable comparison of the various machine learning models, suitable performance or error measures are needed to be selected, for the acquisition of illustrative estimation accuracy of the output variable. To that end, the four-regression performance metrics were used in this study, with a brief description of them given below.

5.4.1 Root-Mean-Squared Error (RMSE)

This performance measure assesses the difference between the true and predicted output of an entire dataset as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (LL_i - LL_i)^2}{N}} \quad (19)$$

where LL_i is the true value of the load-level for a particular datapoint i, LL_i is the predicted value, and N represents the total number of samples in the dataset.

5.4.2 Correlation Coefficient (R)

The strength and direction of the linear relation between the predicted and true values of the output can be measured using the correlation coefficient, *R*. Values of *R* are usually bounded between -1 and 1, and it depicts the strength of the correlation, with positive values presenting positive correlation and vice-versa. The correlation coefficient, *R*, can be computed as:

$$R = 1 - \frac{\sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right) \left(LL_i - \overline{LL_i} \right)}{\sqrt{\sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right)^2 \sum_{i=1}^{N} \left(LL_i - \overline{LL_i} \right)^2}} \quad (20)$$

where $\overline{LL_i}$ and $\overline{LL_i}$ are the averages of the true and predicted load-levels, respectively.

5.4.3 Explained Variance Score (EV)

The explained variance score measures the extent to which the variance in the output of the dataset is captured by the predictive model. Values of EV closer to 1.0 signifies a higher correlation between predicted and true values of the output. Mathematical, Explained Variance Score, EV, is computed as:

$$EV = 1 - \frac{\sum_{i=1}^{N}\left(LL_i - LL_i - \overline{LL}_i + \overline{LL_i} \right)^2}{\sum_{i=1}^{N}\left(LL_i - LL_i \right)^2} \tag{21}$$

5.4.4 Index of Agreement (IA)

It establishes a level of agreement between the predicted and their corresponding true values. It is a dimensionless measure of model accuracy and has been argued by some researchers as a remarkable improvement to the more popular coefficient of determination. Values of Index of Agreement (IA) closer to 1.0 signifies better agreement. Although similar to the correlation coefficient, R, IA is less sensitive to outliers or extreme values and is computed as follows:

$$IA = 1 - \frac{\sum_{i=1}^{N}\left(LL_i - LL_i \right)^2}{\sum_{i=1}^{N}\left(\left| LL_i - \overline{LL_i} \right| + \left| LL_i - \overline{LL_i} \right| \right)^2} \tag{22}$$

The best performing machine model is selected by assessing the aforementioned performance metrics on the testing data. The model is then validated by considering the full dataset and predicting the load-level of the RC beams and one-way slabs.

5.5 Model Interpretation

The various forms of machine learning techniques differ in their level of complexity, and hence influence how they can be interpreted. Generally, linear models are more likely to be interpreted with ease, and thus can give a fair understanding of the underline process being modelled. Also, they tend to give valuable insight and information needed for model improvement. Conversely, linear models are not sophisticated enough to yield very accurate results compared to non-linear machine linear models. For instance, the XGBoost regression model usually tends to produce more accurate results than linear regression models on many datasets. On the other end, interpretating a model developed from the XGboost algorithm or any flexible machine learning model, is quite challenging. To this end, the recently developed SHapley Additive exPlanation (SHAP) tool can be used for model interpretability of very complex machine learning models. SHAP results in the provision of a so-called explanation model useful for (1) demonstrating the importance of any feature in the dataset; (2) quantifying how each feature affects the model prediction on both local and global scales; (3) ascertaining how the prediction model output changes with variations in the input values of the feature. A brief description of Shapley Additive Explanation (SHAP) for model interpretation is presented below.

Once again, consider an example input vector of features $X_i = (x_{i1}, x_{i2}, x_{i3}, ...x_{ip})$ for which a machine model $f(X_i)$ is developed to predict a quantitative response LL_i. The SHapley Additive ExPlanation (SHAP) for machine learning model interpretation begins with mapping the original input vector of features X_i into a binary simplified input vector $X_i' \in \{0,1\}^p$, which serves as input for the explanation model $g(X_i')$. The X_i' which contains either 0 or 1, depicts whether a feature is present (x_{ij}' = 1) or absent (x_{ij}' = 0) in the explanation model yet to be determined. The explanation model is usually obtained by a weighted summation of the simplified input vector of features X_i' and a constant term as represented in Eq. 23.

$$g(X_i') = \theta_0 + \sum_{j=1}^{p} \theta_j x_{ij}' \tag{23}$$

where $X_i' \in \{0,1\}^p$ is a vector of binary simplified inputs features, x_{ij}', which are mapped to the original input features x_{ij}, and θ_j is the attribution value for feature j. To this end, SHAP is usually referred as a class of feature attribution methods, amongst others such as LIME [58], deepLIFT [59] etc.

The advantage of using SHAP as opposed to other feature attribution methods is how it presents three key desirable properties that any feature attribution method should have. The first property deals with local accuracy, where the output of the explanation is expected to match that of the model prediction for any data point in the dataset (see Eq. 24).

$$f(X_i) = g(X_i') \tag{24}$$

Secondly, if a feature does not contribute to the predictive model's output, then the feature attribution value should be zero in the explanation model (see Eq. 25).

$$x_{ij}' = 0 \Rightarrow \theta_j = 0 \tag{25}$$

To conclude, the third property states that if the predictive model changes and causes a particular simplified input contribution to increase or stay the same regardless of other simplified inputs, then the attribution from that input should not decrease. In explaining the third property, known as consistency, consider two predictive models $f_1(X_i)$ and $f_2(X_i)$. Mathematically, the consistency property can be presented as:

$$f_1(X_i) - f_1(X_i \setminus j) \geq f_2(X_i) - f_2(X_i \setminus j) \Rightarrow \theta_j(f_1) \geq \theta_j(f_2) \tag{26}$$

where $f_1(X_i \setminus j)$ and $f_2(X_i \setminus j)$ denote prediction values of models $f_1(X_i)$ and $f_2(X_i)$ with feature j absent, respectively. Similarly, $\theta_j(f_1)$ and $\theta_j(f_2)$ are the feature attribution values for $f_1(X_i)$ and $f_2(X_i)$ respectively.

It turns out the only solution for the feature attribution values θ_j that satisfies these three properties, are the Shapley values of the conditional expectation function of the original model[60]. These Shapley values can be computed from Eq. 27 as:

$$\theta_j(f, X_i) = \sum_{Z'_i \subseteq X'_i} \frac{|Z'_i|!\left(P - |Z'_i| - 1\right)!}{P!}\left[f\left(Z'_i\right) - f\left(Z'_i \setminus j\right)\right] \quad (27)$$

where $\theta_j(f, X_i)$ is the Shapley regression value or feature attribution value for the feature j in the model $f(X_i)$, Z' is a vector of binary values representing one of the subsets of X', P is the number of input features, $|Z'|$ represents the number of non-zero elements in Z', $f\left(Z'_i\right)$ denotes the model prediction for Z' and $f\left(Z'_i \setminus j\right)$ represents the prediction for Z' without feature j. These Shapley values $\theta_j(f, X_i)$, once obtained, can be used to explain the model output. The magnitude and sign of $\theta_j(f, X_i)$ will determine whether a particular feature impacts the model output negatively or positively. The θ_0 from Eq. 23 represents the average value of the model prediction assuming the model has no input feature and usually represents a base value for the model output before the various Shapley values obtained from Eq. 27 are aggregated to obtain the output $f(X_i)$. Further details on techniques available to compute the Shapley values can be found elsewhere in [60].

6. Results and Discussions

6.1 Model Predictions and Evaluation

6.1.2 Global Level

The performance of the four selected machine learning models for load-level estimation of the class of structural elements under consideration is presented. Following the training-testing splitting rule of 70/30 as previously mentioned, the accuracy of these models was drawn for each group of data (training and testing data). Typically, the performance of the model on the testing data is used to determine its generalization capacity. Table 4 shows a summary of the four performance measures for each dataset, across the machine learning models developed. It presents the mean and standard deviation of the performance measures for 1000 runs of the developed models having different randomly sampled training and testing data. Multiple runs of the developed models were necessary to help ascertain how statistically significant the model predictions might differ. It is worth mentioning that high values of the correlation coefficient (R), explained variance (EV) and index of agreement (IA) for a particular model signifies greater performance. Similarly, models with lower root-mean squared error (RMSE) also presents a case for better predictability.

Among the four machine learning models, the RFR and XGBoost models yielded the best performance on the training and testing data respectively (see Table 4). They produced relatively high values of the correlation coefficient (R), explained variance (EV) and index of agreement (IA) when compared to the ENR and SVR models. Similarly, lower average values were recorded for the root-mean squared error (RMSE) of these models, when compared to the ENR and SVR models, during the training and testing phase. However, the diferrence between the mean estimate for these models (RFR and XGBoost) were comparatively similarly, as well as their deviations. To assess the statistical significance of the differences of the mean values of these two high performing models we calculated the t-statistic, compared this to the critical t-value, and calculated the corresponding p-values as well. Details on how the t-statistic is computed when comparing means of different populations can be found elsewhere [61,62] .

Table 4. Performance measures of various machine learning models

Data	Algorithm	Statistic	Performance Metrics			
			RMSE	R	EV	IA
Training	SVR	Mean	0.150	0.811	0.628	0.851
		SD	0.004	0.012	0.020	0.011
	ENR	Mean	0.151	0.785	0.616	0.867
		SD	0.004	0.012	0.020	0.009
	RFR	Mean	0.0897	0.934	0.867	0.961
		SD	0.003	0.005	0.009	0.003
	XGBoost	Mean	0.104	0.913	0.819	0.941
		SD	0.003	0.005	0.011	0.004
Testing	SVR	Mean	0.151	0.810	0.625	0.849
		SD	0.009	0.028	0.035	0.016
	ENR	Mean	0.152	0.784	0.611	0.864
		SD	0.008	0.028	0.043	0.017
	RFR	Mean	0.138	0.827	0.681	0.900
		SD	0.009	0.026	0.044	0.014
	XGBoost	Mean	0.136	0.831	0.687	0.895
		SD	0.008	0.026	0.042	0.014

SD: standard deviation; SVR: Support vector regression; ENR: elastic-net regression; RFR: random forest regression; XGBoost: extreme gradient boosting.

Table 5 and 6 presents the calculated t-values and p-values for the comparisons of the performance mean values for the RFR and XGBoost models. The t-values were compared to a critical t-value of 1.96, obtained from the student's-t distribution at a 5% significance level with 1998 degrees of freedom. All t-values computed for these two models, and across various performance measures were higher than this critical value (see Table 5 and 6). The calculated p-values show that the actual levels of statistical significance are all below 1%.

The data shown in Tables 4, 5 and 6 mean that the differences between the mean values of RFR and XGBoost for the various performance measures are statistically significant. From Table 4, the XGBoost model outperformed the RFR model when the RMSE, R and EV are considered, whiles a higher IA values was observed for the RFR model, during the testing phase. To this end, we recommed the XGBoost model as the optimal model for load level estimation of shear-critical RC beams and slabs. Since the generalization capability of a model is usually assessed by considering how it performs during the testing phase, further comparisons between these two models are drawn.

Table 5. T-values of various model comparisons

Performance Measure	Dataset	Model	t-value			
			SVR	ENR	RFR	XGB
RMSE	Training	SVR		-5.59017	381.3707	290.9295
		ENR	5.59017		387.6952	297.2541
		RFR	-381.371	-387.695		-106.586
		XGB	-290.93	-297.254	106.5859	
	Testing	SVR		-2.62613	32.29876	39.39193
		ENR	2.626129		36.7658	44.72136
		RFR	-32.2988	-36.7658		5.252257
		XGB	-39.3919	-44.7214	-5.25226	
R	Training	SVR		48.44814	-299.2	-248.117
		ENR	-48.4481		-362.446	-311.363
		RFR	299.2001	362.4457		93.91486
		XGB	248.1172	311.3627	-93.9149	
	Testing	SVR		20.76349	-14.0693	-17.3797
		ENR	-20.7635		-35.5871	-38.8975
		RFR	14.0693	35.58705		-3.4401
		XGB	17.37972	38.89748	3.440105	
EV	Training	SVR		13.41641	-344.608	-264.615
		ENR	-13.4164		-361.91	-281.24
		RFR	344.608	361.9105		106.7986
		XGB	264.615	281.2401	-106.799	
	Testing	SVR		7.985022	-31.4975	-35.8615
		ENR	-7.98502		-35.9803	-39.9834
		RFR	31.49748	35.98033		-3.11925
		XGB	35.86152	39.98339	3.119251	
IA	Training	SVR		-35.5995	-305.085	-243.154
		ENR	35.59953		-313.333	-237.6
		RFR	305.0851	313.3333		126.4911
		XGB	243.1545	237.5997	-126.491	
	Testing	SVR		-20.3186	-75.8579	-68.4209
		ENR	20.31856		-51.693	-44.5134
		RFR	75.85792	51.69299		7.985957
		XGB	68.42087	44.51341	-7.98596	

Table 6. P-values of various model comparisons

Performance Measure	Dataset	Model	p-value SVR	ENR	RFR	XGB
RMSE	Training	SVR		2.58E-08	0	0
		ENR	2.58E-08		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		0.008702	1.6E-184	1.2E-251
		ENR	0.008702		1.8E-226	2.9E-303
		RFR	1.6E-184	1.8E-226		1.66E-07
		XGB	1.2E-251	2.9E-303	1.66E-07	
R	Training	SVR		0	0	0
		ENR	0		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		7.19E-87	6.12E-43	4.07E-63
		ENR	7.19E-87		2.9E-215	6.9E-247
		RFR	6.12E-43	2.9E-215		0.000593
		XGB	4.07E-63	6.9E-247	0.000593	
EV	Training	SVR		2.32E-39	0	0
		ENR	2.32E-39		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		2.35E-15	3.8E-177	7.2E-218
		ENR	2.35E-15		5.4E-219	2.4E-257
		RFR	3.8E-177	5.4E-219		0.001839
		XGB	7.2E-218	2.4E-257	0.001839	
IA	Training	SVR		2.2E-215	0	0
		ENR	2.2E-215		0	0
		RFR	0	0		0
		XGB	0	0	0	
	Testing	SVR		1.39E-83	0	0
		ENR	1.39E-83		0	3E-301
		RFR	0	0		2.33E-15
		XGB	0	3E-301	2.33E-15	

In corroborating this finding, a score analysis is conducted. Score analysis basically entails assigning a score to the various values of the performance measures across different models. In this study, with the number of machine learning models being 4, a model that yields the greatest performance is assigned a score of 4, whereas the least performing model is assigned a value of 1. Considering that 1000 runs of the developed models for different training and testing data was conducted, the average score for a particular model is used. Subsequently, a summation of the average scores of the various performance measures for each machine learning model is computed to obtain the total score (see Table 7).

Table 7. Score analysis results of various machine learning models

Data	Algorithm	Score				Total Score
		RMSE	R	EV	IA	
Training	SVR	1.63	1.00	2.00	1.92	6.55
	ENR	1.37	2.00	1.00	1.08	5.45
	RFR	4.00	4.00	4.00	4.00	16.00
	XGBoost	3.00	3.00	3.00	3.00	12.00
Testing	SVR	1.63	1.00	2.25	1.85	6.73
	ENR	1.4.0	2.00	1.02	1.20	5.62
	RFR	3.32	3.88	3.10	3.31	13.61
	XGBoost	3.65	3.12	3.63	3.64	14.04

The model producing the highest total score is deemed to be the best performing model. As seen in Table 7, the RFR and XGBoost models dominated the score analysis by being the best models during the training and testing phase respectively. Nevertheless, since the generalization capability of model can be evaluated by considering its performance on the testing data, the XGBoost model is deemed the optimal model for load-level estimation of shear-critical RC beams and slabs. The attained total score were 13.61 and 14.04 for the RFR and XGboost models respectively, during the testing phase. However, the RFR model tends to outperforms the XGBoost model during the training phase (see Table 7). This observation might imply that there is an inherent overfitting problem with the RFR model. The least performing model was the Elastic-Net Regression (ENR), which yielded total scores of 5.45 and 5.62 during the training and testing phase respectively. This observation also suggests that the linear statistical method of analysis may not be optimal for predicting the load-level of shear-critical beams and slabs using multifractal analysis.

Graphical presentation of the score analysis is given in a form of a radar chart as shown in Fig. 9, to facilitate interpretation. It is observed that the Random Forest Regression (RFR) model tends to perform well on the training data (see Fig. 9) than any other model across the various performance measures. Similarly, the radar charts indicates the the XGboost model performance better than the RFR model during the testing phase. This suggests that non-linear models, in particular tree-based models such as random forest and the extreme gradient boosting machine, tends to produce better estimates of the load-level of shear-critical concrete beams and slabs using the proposed framework.

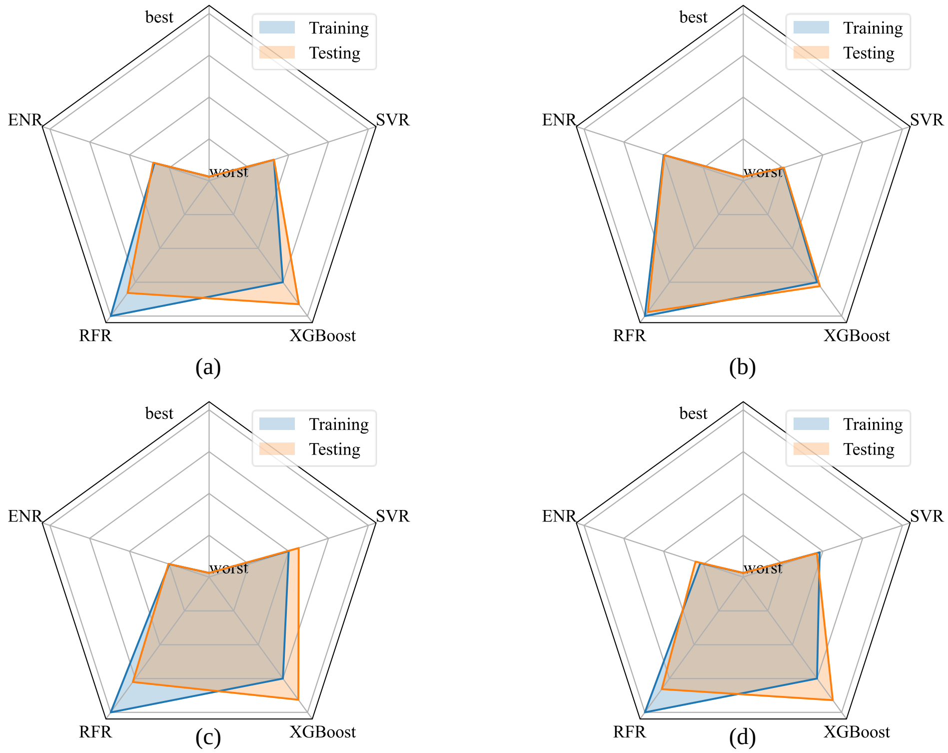


Fig. 9 Radar charts for various performance measures: (a) RMSE; (b) R; (c) EV; (d) IA. (color printed)

6.1.3 Local Level

In order to gain insight into the predictive performance of these models at the local level, Fig. 10 shows a typical scatter plot, to help establish the correlation between predicted and true values of the load-level for each data point in the training and testing dataset. This visualization will also assist in determining which regions across the load-level range, tends to produce better estimates. Evidently, the XGboost produces the lower scatter or deviation with a narrow prediction interval compared to the other models investigated in this study (see Fig. 10). The mean of the predicted-to-tested ratio for this model was 1.04 with a coefficient of variation of 27%. Nonetheless, there seems to be significant error or outliers for some data points, particularly in the testing data. The majority of these data points yielded a prediction of load-level higher than their true values, and hence conservative for damage assessment or design. Although there exist works on estimating the load-level of beams and slabs using fractal analysis and other data-driven machine learning algorithms[2,35,63], fair comparison cannot be generally drawn for most of them due to the disparity in specimens that make up the database as well as its size. Nevertheless, a closely related work that used about 95% of the database in this study is that of Davoudi et al. [2] who provided another alternative to damage assessment of shear-critical concrete beams and slabs using machine vision In their assessment, scatter plot and performance metric values similarly those presented in Fig. 10 and Table 4 were plotted. By comparison, the developed model produced comparable performance measures as against those reported by Davoudi et al. [2]

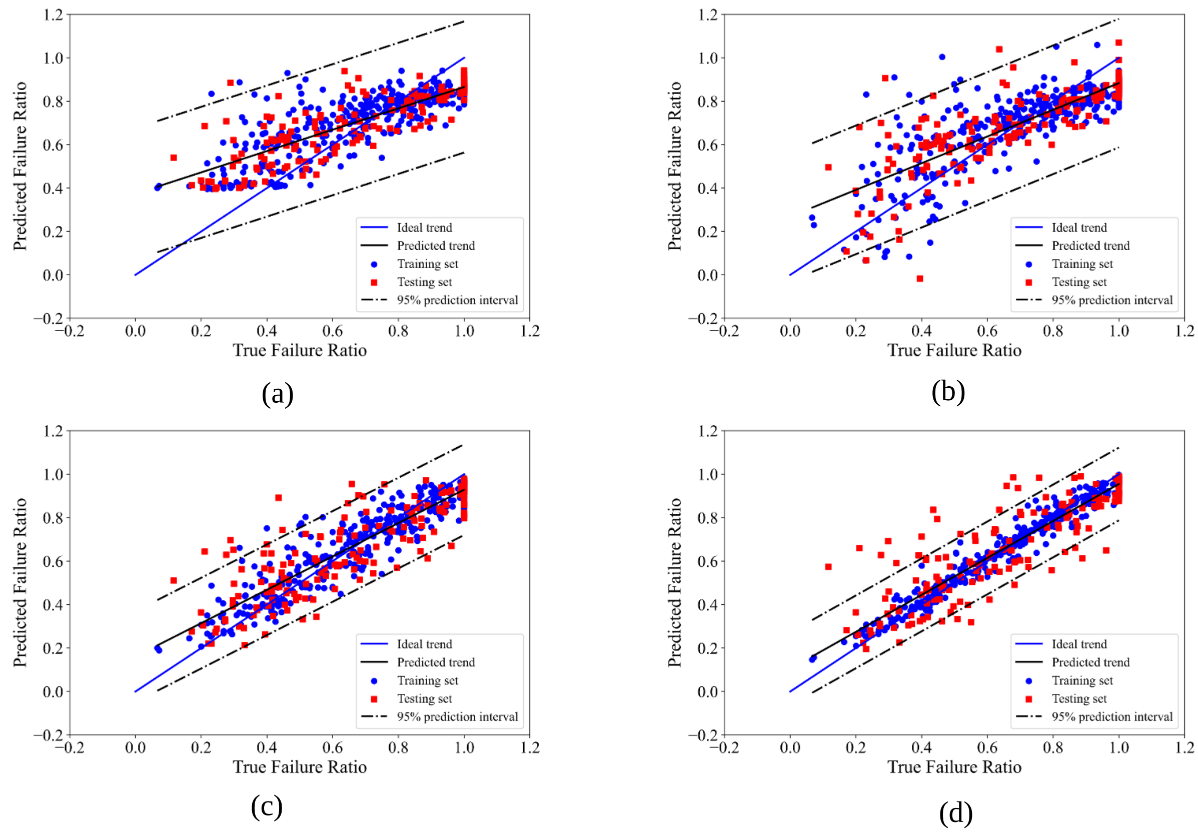


Fig. 10 Scatter plot of load-level results predicted by different machine learning model: (a) SVR; (b) ENR; (c) RFR; (d) XGBoost. (color printed)

A typical regression error characteristic (REC) curve as constructed in Fig. 11 for the various models is used to facilitate model predictability at the local scale. The REC curve is a cumulative distribution function which tends to establish a relationship between the absolute error or deviation (x-axis) as against the proportion of datapoints (y-axis) with absolute error lesser than or equal to the current level. It is analogous to the receiver operating characteristic (ROC) curve in classification problems for model assessment. Whereas the ROC curve uses the area under the curve (AUC) to evaluate performance, it has been widely established that the area over curve (ROC) be used to provide a valid measure for regression problems. The ROC can be simply computed by subtracting the AUC from 1. A regression model is known to perform well if the AOC value of an REC curve is low.

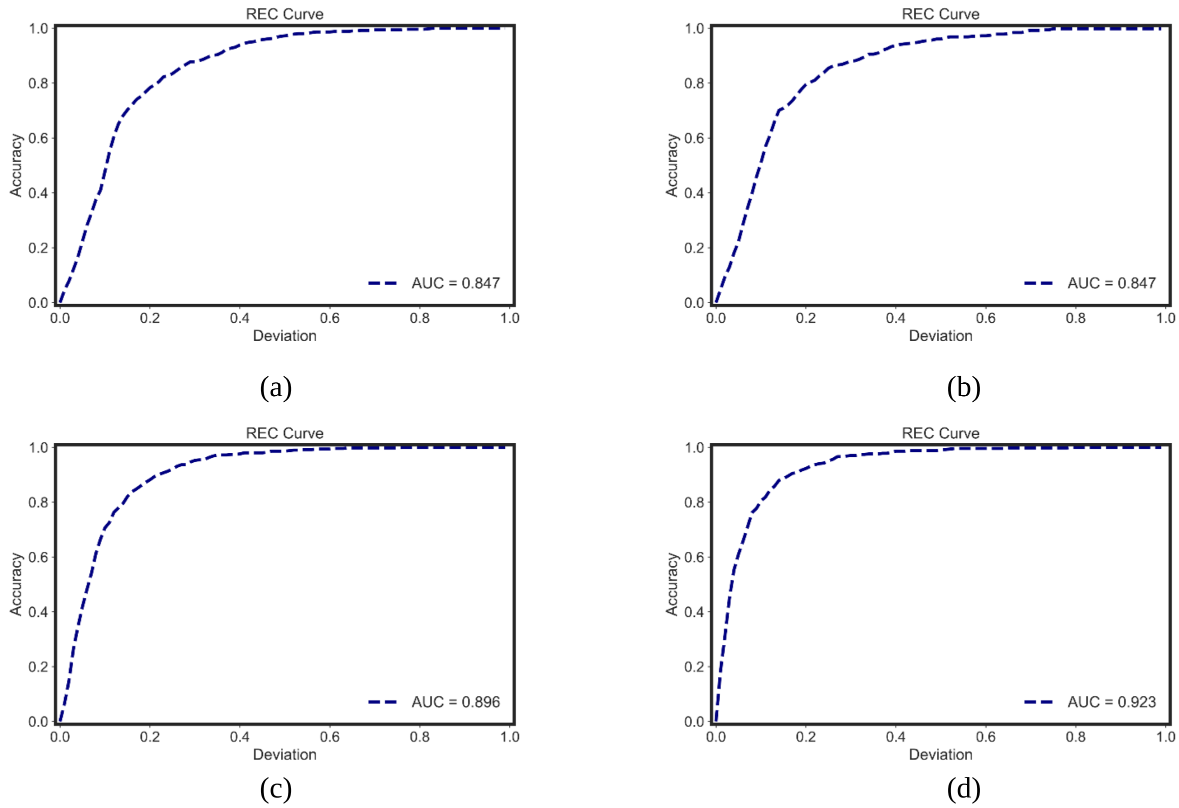


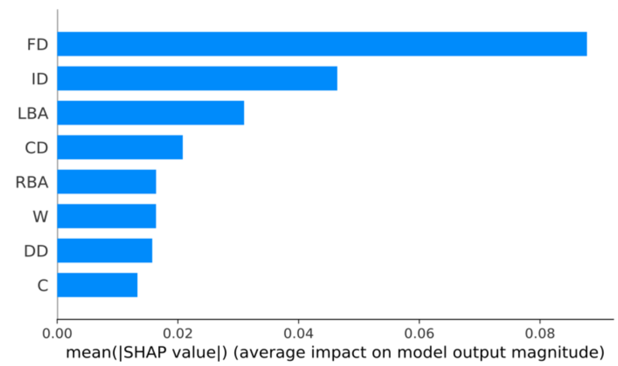
Fig. 11 Regression error characteristic curves for various machine learning model: (a) SVR; (b) ENR; (c) RFR; and (d) XGBoost

From Fig. 11, which shows the REC curve using the full dataset, the XGBoost model produced the lowest ROC of 0.077, hence corroborating findings attained at the global level of assessment. The ROC for both SVR and ENR models were the same, hence suggesting equal performance. 80% of the datapoints produced absolute errors of load-level lesser than 0.1 for the XGBoost model (see Fig. 11d). The RFR, ENR and SVR models yielded predictions of which 80% had absolute errors within 0.17, 0.21 and 0.21 respectively. In this study, the XGBoost model developed remains the optimal model at both local and global levels for estimating the load-level of shear-critical RC beams and slabs.

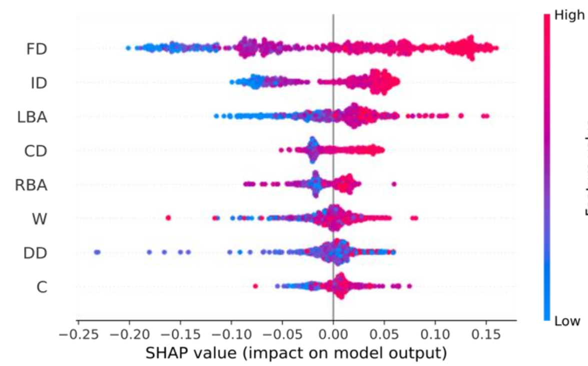
6.2 Model Interpretation

6.2.1 Global Level

A simplified explanation model was developed for the optimal predictive model, i.e, XGBoost, for interpretation using SHapley Additive ExPlanation (SHAP). On the global scale (entire dataset), the relative importance of each feature is given in Fig. 12. It provides the mean of the absolute SHAP values computed for each feature in the full dataset. These mean values are then used to ascertain the impact of each feature on the predictions made.



(b)



(c)

Fig. 12 Global interpretations of XGBoost model: (a) SHAP feature importance; and (b) SHAP summary plot. (color printed)

Generally, it was observed that the so-called generalized dimensions (FD, ID and CD), which were obtained from the multifractal analysis of the crack patterns considered, has significant impact on the estimation of the load-level, as opposed to the other geometric features acquired from the singularity spectrum. For the generalized dimensions, the box-counting fractal dimension (FD) was arguably the most critical parameter (see Fig. 12a). Many of previous works on the application of multifractal analysis for crack damage assessment of RC elements have always considered FD as the most influential feature, with the findings from this study affirming it. The area under the left branch of the singularity spectrum (LBA) tends to contribute the most to the model predictions for the geometric features considered, providing about 35% of that produced by FD. The least contributing feature as seen in Fig. 12a is the capacity (C), whose mean absolute SHAP value was about 17% as important as the most critical feature.

In order to determine how the original values of the features within the dataset affects the model prediction or load-level, Fig. 12b demonstrate a summary plot for such analysis. Each point in the plot shows the SHAP value (x-axis) of a particular feature (y-axis). For each feature, the distribution of SHAP values are shown along the x-axis, which are colour-coded to differentiate between high (red dots) and low (blue dots) values of the original feature. For instance, for high values of the fractal dimension (FD) as seen in the upper right corner of Fig. 12b, there is an expected increase in the load-level of about 16%. Nevertheless, there are instances for which higher values of FD cause a reduction in the load-level (red dots on the left-hand side of the summary plot for FD). To this end, the average value of the distribution of SHAP values is used to ascertain whether a feature impacts the load-level positively or negatively. In general, for the critical features, an increase in the fractal dimension FD, information dimension ID, and correlation dimension CD causes an increase in the load-level. Conversely, the load-level tends to decrease when the area under the left branch (LBA), is low.

6.2.2 Local Level

SHAP also provides interpretation for each individual prediction. In assessing the impact of the various feature at the local level, four RC beams were sampled from the database considered. These samples had load-levels spanning various damage states (low, moderate, near failure and ultimate failure). For the sample exhibiting a lower degree of damage, a simplified explanation model which comprises the aggregation of SHAP values for each feature and a base value to yield a final prediction is given in the second column of Table 8. This sample had a true load-level of 17.1% and a predicted value of 22%. It is worth noting that the base value depicts the default prediction when the attribution from each feature is excluded.

Table 8. Relative SHAP values of features for four selected samples

Feature	Shapley values of selected sample scenarios (%)			
	Low	Moderate	Near failure	Failure
C	-3.6	1.7	0.2	0.6
FD	-18.1	-6.8	1.9	12.2
W	-1.8	0.2	0.7	2.5
LBA	-8.7	3.1	5.6	3.1
RBA	-1.8	-1.5	-3.3	0.9
ID	-8.7	-5.3	1.7	6.3
CD	-2.2	-1.8	-0.2	3.6
DD	-0.8	0.4	5.9	1.3
Base Prediction	67.7	67.7	67.7	67.7
Prediction	**22.0**	**57.7**	**80.2**	**98.2**
True Value	**17.1**	**59.0**	**81.5**	**100**

It is observed that, FD, LBA, ID and C are the most critical features that influence the predictions of RC beams with a low load-level (see Fig. 13a). These features negatively impact the final prediction by reducing the base value. For this particular sample, FD, LBA ID and C caused a reduction in the base value of about 18.1%, 8.7%, 8.7% and 3.6%, respectively.

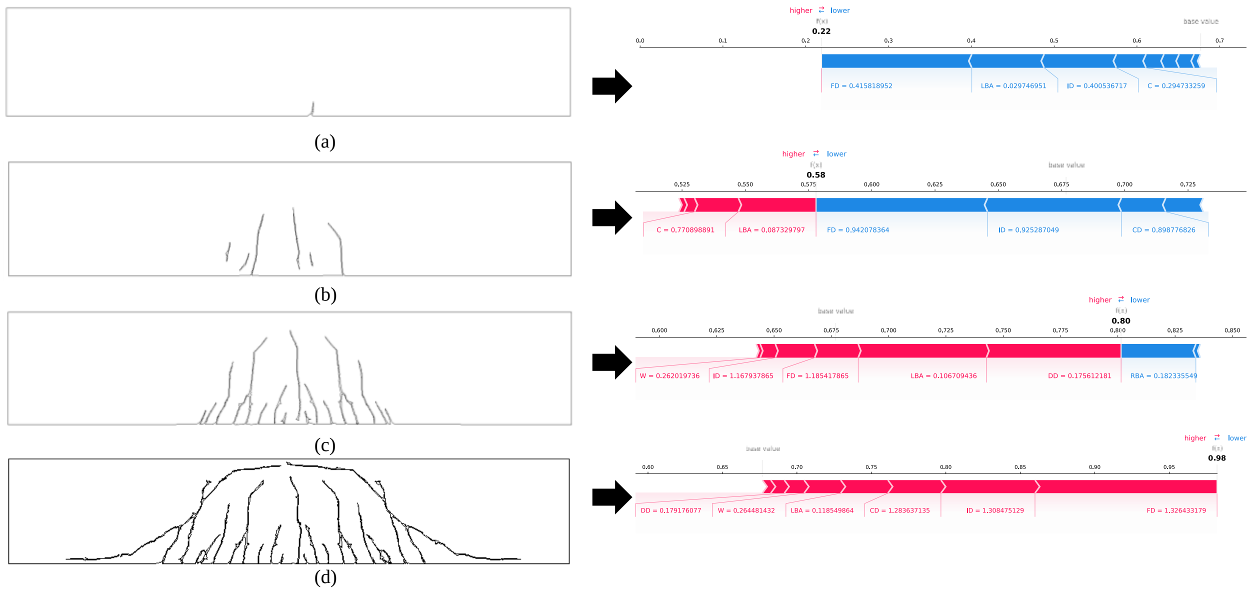


Fig. 13 Local interpretations of selected RC beams with different damage levels: (a) low-Sherwood [54] ; (b) moderate Cao[57] - ; (c) near failure - Cao[57] ; and (d) ultimate failure - Cao[57]. (color printed)

The second sample was selected to depict an instance where the RC beam is moderately damaged. The true and predicted load-level for this sample is 59% and 57.7%, respectively. The SHAP values of each feature for this sample are given in Table 8. Fig. 13b illustrates the critical features that influence the prediction made for this sample. The red bars represent contributions from features that increase the load-level, with the blue bars outlining features that affect the load-level prediction negatively. It is observed that whereas LBA and C reduce the load-level for the slightly damaged beams (Fig. 13a), they rather tend to increase the load-level for moderately damaged RC beams (Fig. 13b). The original values of LBA and C are relatively higher for the moderately damaged beams when compared to the slightly damaged beams, and hence could be a contributing factor to explain this observation (see annotations in Fig. 13a and 13b). As the level of damage of the RC beam increases and approaches failure, the SHAP values for the features assume positive values (Table 8). This is evident in the two other samples which were used to represent near failure and ultimate failure cases (see Table 8 and Fig. 13). The fractal characteristics of these beams produced relatively high values of the original features and hence can partly give a physical reason why the predictions are increased from the base value to the final output. In all cases, FD and ID appears to dominate the most critical features for the four samples considered and either affect the load-level prediction positively or negatively, depending on the level of damage the RC beam in question has sustained.

6.3 Feature Dependency plot

The correlation between SHAP values and features values can give a detailed insight into which scenarios can either cause a decrease or increase in the load-level. Fig.14 shows feature dependency plots to facilitate such analysis. For brevity, the variation of SHAP values for six selected features is presented.

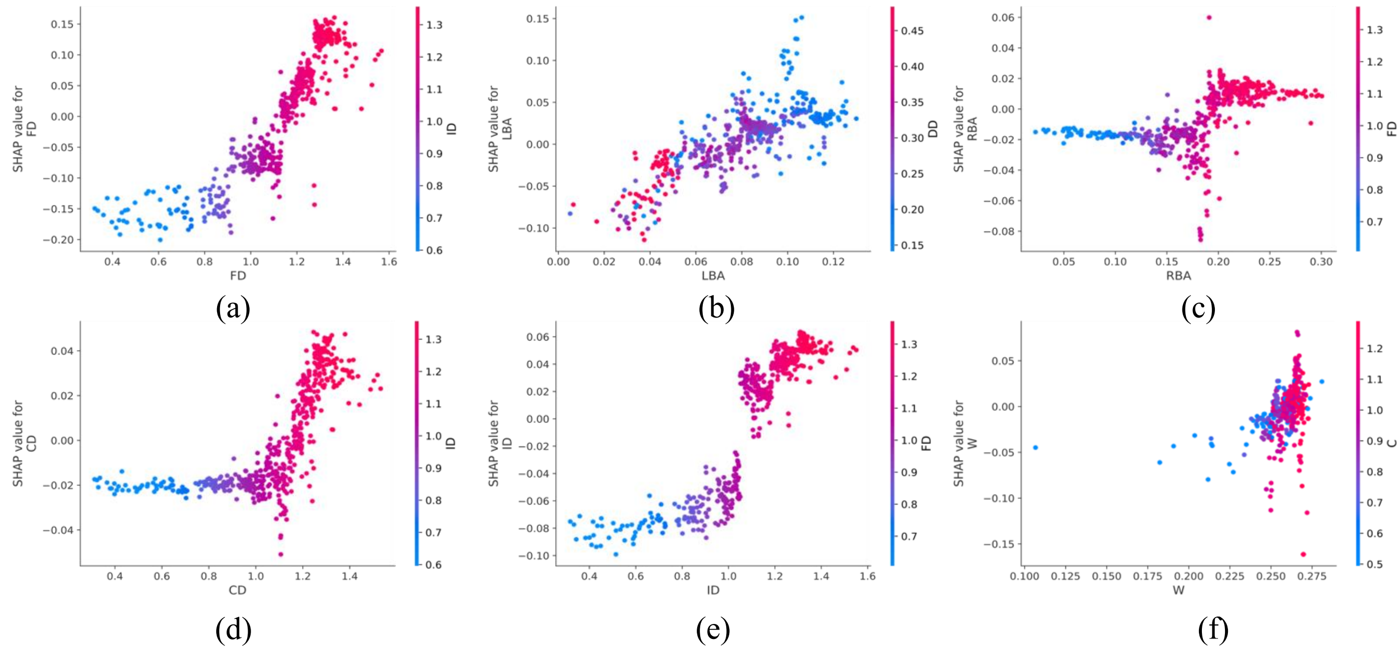


Fig. 14 Plots of feature dependency: (a) FD; (b) LBA; (c) RBA; (d) CD; (e) ID; and (f) W. (color printed)

SHAP values increase with increasing values of FD, LBA, CD and ID. This indicates that FD, LBA, CD and ID are positively correlated with load-level estimation of shear-critical RC beams. From Fig. 14a, RC beams with FD greater than 1.05 tend to cause an increase in load-level. Note that many of these RC beams tend to have values of ID greater than 1.1 (Fig. 14e). Nevertheless, whereas the maximum increase in load-level considering ID is about 6%, FD can contribute an increase of about 16% in load-level (Fig. 14a and 14e). Beams with LBA values greater than 0.05 and DD less than 0.3, do cause an increase in load-level (Fig. 14b). Even though RC beams with CD greater than 1.1 tend to cause an increase in load-level, its contribution is not so significant with a maximum increase of about 4.5%. For W and RBA, the pattern is inconclusive and hence insignificantly affect load-level estimates. Findings from this analysis can be used to develop closed form solutions to load-level estimation for damage assessment of shear-critical RC beams and slabs.

7. Conclusions

This paper explored the application of multifractal analysis to shear-critical RC beams and slabs for load-level estimation. A database of 508 RC beams and slabs were used for model training (70%) and testing (30%). Multifractal analysis was first conducted on images of crack patterns of these beams, with critical features extracted from the singularity and generalized dimension spectra to form the design input matrix in the model development phase, whereas the load-level for each specimen served as the output. The efficiency of four regression-like machine learning models (elastic-net regression (ENR), support vector regression (SVR), random forest regression (RFR) and extreme gradient boosting (XGBoost)) were explored on the dataset. Hyperparameter optimization was conducted for these models using a random search algorithm. For performance measures (root-mean squared error (RMSE), correlation (R), explained variance (EV) and index of agreement (IA)) were used to facilitate model evaluation and selection. Shapley additive explanations (SHAP) was later used for model interpretation. The primary findings from this study are listed below:

- The XGBoost model was the most effective model for estimating the load-level of shear-critical RC beams and slabs. The mean of the predicted-to-tested ratio was 1.04 with coefficient of variation of 27%.
- Upon comparing the XGBoost model with the other models, it was found out that tree-based methods perform significantly better than linear and non-linear methods of regression.
- For model interpretation at the global level, it was revealed by SHAP that the so-called generalized dimensions (fractal dimension (FD), information dimension (ID) and correlation dimension (CD)) which was obtained from the multifractal analysis of the crack patterns considered, had significant impact on the estimation of the load-level, as opposed to the other geometric features acquired from the singularity spectrum. The fractal dimension (FD) was arguably the most critical feature whereas the capacity (C) was the least influential.
- Shear-critical RC beams with FD greater than 1.05 tend to cause an increase in load-level, which can be as high as high as 16%. Even though RC beams with CD greater than 1.1 tend to cause an increase in load-level, its contribution is not so significant with a maximum increase of about 4.5%.
- It was observed that depending on how high or low the original values of the multifractal features are, which is heavily related to the level of damage, the obtained SHAP values will either increase or decrease the load-level estimates. For instance, whereas the area under the left branch (LBA) and C reduce the load-level for the slightly damaged beams (Fig. 14a), they rather tend to increase the load-level for moderately damaged RC beams (Fig. 14b).

To facilitate the practical application of the developed model as well as reproducibility, the source code and database will be made available to the public on a GitHub account. Users may use the proposed model to either get a firsthand insight on the level of damage sustained by such structural elements in service, before another sophisticated framework can be applied.

8. Limitations

Despite the successful development of the structural load estimation model based on multifractal features, some limitations have been identified. The present study only considers RC beams and slabs that have been designed to exhibit shear dominant failure. In order words, the developed model is not generally applicable, as it cannot be utilized for other structural failure phenomena. Future studies should continuously explore the combined application of machine-learning and multifractal analysis to other modes of structural failure, type of RC element and loading conditions. This could assist in the development of a unified model for structural load level estimation for a wide variety of RC structural elements. Secondly, valuable damage parameters on crack patterns such concrete spalling and crack width were not considered in the present study. The generalization error of the develop model can be improved if information relating these parameters are provided and well documented. Therefore, future experimental testing programs should grant the research community access to raw data if possible. Despite these limitations, findings from this research have revealed the need for continuous research in the application of machine-learning based multifractal analysis of reinforced concrete structures for structural load-level assessment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

9. References

[1] Athanasiou A, Ebrahimkhanlou A, Zaborac J, Hrynyk T, Salamone S. A machine learning approach based on multifractal features for crack assessment of reinforced concrete shells. Comput-Aided Civ Infrastruct Eng 2020;35:565–78. https://doi.org/10.1111/mice.12509.

[2] Davoudi R, Miller GR, Kutz JN. Structural Load Estimation Using Machine Vision and Surface Crack Patterns for Shear-Critical RC Beams and Slabs. J Comput Civ Eng 2018;32:04018024. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000766.

[3] Rens KL, Wipf TJ, Klaiber FW. Review of Nondestructive Evaluation Techniques of Civil Infrastructure. J Perform Constr Facil 1997;11:152–60. https://doi.org/10.1061/(ASCE)0887-3828(1997)11:4(152).

[4] Cha Y-J, Choi W, Büyüköztürk O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Comput-Aided Civ Infrastruct Eng 2017;32:361–78. https://doi.org/10.1111/mice.12263.

[5] Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, et al. Automated Crack Detection on Concrete Bridges. IEEE Trans Autom Sci Eng 2016;13:591–9. https://doi.org/10.1109/TASE.2014.2354314.

[6] Guidebook on Non-destructive Testing of Concrete Structures 2019. https://www.iaea.org/publications/6347/guidebook-on-non-destructive-testing-of-concrete-structures (accessed March 16, 2022).

[7] Maffei J, Comartin CD, Kehoe B, Kingsley GR, Lizundia B. Evaluation of Earthquake-Damaged Concrete and Masonry Wall Buildings. Earthq Spectra 2000;16:263–83. https://doi.org/10.1193/1.1586111.

[8] Hartle RA, Ryan TW, Mann E, Danovich LJ, Sosko WB, Bouscher JW. Bridge Inspector's Reference Manual: Volume 1 and Volume 2. 2002.

[9] Yumpu.com. ATC-20-1 Field Manual: Postearthquake Safety Evaluation of ... YumpuCom n.d. https://www.yumpu.com/en/document/view/47765334/atc-20-1-field-manual-postearthquake-safety-evaluation-of- (accessed March 16, 2022).

[10] Ayub M. Investigation of March 15, 2018 Pedestrian Bridge Collapse at Florida International University, Miami, FL. Washington, D.C.: US Department of Labour; 2019.

[11] ACI PRC-201.1-08 Guide for Conducting a Visual Inspection of Concrete in Service n.d. https://www.concrete.org/store/productdetail.aspx?ItemID=201108&Language=English&Units=US_AND_METRIC (accessed March 23, 2022).

[12] Manual for Bridge Element Inspection (2nd Edition) - Vanderbilt University n.d. https://catalog.library.vanderbilt.edu/discovery/fulldisplay/alma991043715690603276/01VAN_INST:vanui (accessed March 23, 2022).
[13] Madani HM, Dolatshahi KM. Strength and stiffness estimation of damaged reinforced concrete shear walls using crack patterns. Struct Control Health Monit 2020;27:e2494. https://doi.org/10.1002/stc.2494.
[14] Tan X, Bao Y. Measuring crack width using a distributed fiber optic sensor based on optical frequency domain reflectometry. Measurement 2021;172:108945. https://doi.org/10.1016/j.measurement.2020.108945.
[15] Felice MV, Velichko A, Wilcox PD. Accurate depth measurement of small surface-breaking cracks using an ultrasonic array post-processing technique. NDT E Int 2014;68:105–12. https://doi.org/10.1016/j.ndteint.2014.08.004.
[16] Hamia R, Cordier C, Dolabdjian C. Eddy-current non-destructive testing system for the determination of crack orientation. NDT E Int 2014;61:24–8. https://doi.org/10.1016/j.ndteint.2013.09.005.
[17] Aggelis DG, Kordatos EZ, Soulioti DV, Matikas TE. Combined use of thermography and ultrasound for the characterization of subsurface cracks in concrete. Constr Build Mater 2010;24:1888–97. https://doi.org/10.1016/j.conbuildmat.2010.04.014.
[18] Huang J, Liu W, Sun X. A Pavement Crack Detection Method Combining 2D with 3D Information Based on Dempster-Shafer Theory. Comput-Aided Civ Infrastruct Eng 2014;29.
[19] Park SE, Eem S-H, Jeon H. Concrete crack detection and quantification using deep learning and structured light. Constr Build Mater 2020;252:119096. https://doi.org/10.1016/j.conbuildmat.2020.119096.
[20] Yamaguchi T, Hashimoto S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. Mach Vis Appl 2010;21:797–809. https://doi.org/10.1007/s00138-009-0189-8.
[21] Yeum CM, Dyke SJ. Vision-Based Automated Crack Detection for Bridge Inspection. Comput-Aided Civ Infrastruct Eng 2015;30:759–70. https://doi.org/10.1111/mice.12141.
[22] Kong X, Li J. Vision-Based Fatigue Crack Detection of Steel Structures Using Video Feature Tracking. Comput-Aided Civ Infrastruct Eng 2018;33:783–99. https://doi.org/10.1111/mice.12353.
[23] Wu R-T, Singla A, Jahanshahi MR, Bertino E, Ko BJ, Verma D. Pruning deep convolutional neural networks for efficient edge computing in condition assessment of infrastructures. Comput-Aided Civ Infrastruct Eng 2019;34:774–89. https://doi.org/10.1111/mice.12449.
[24] Destrebecq J-F, Toussaint E, Ferrier E. Analysis of Cracks and Deformations in a Full Scale Reinforced Concrete Beam Using a Digital Image Correlation Technique. Exp Mech 2011;51:879–90. https://doi.org/10.1007/s11340-010-9384-9.
[25] Ebrahimkhanlou A, Salamone S. A probabilistic model for visual inspection of concrete shear walls. Sens. Smart Struct. Technol. Civ. Mech. Aerosp. Syst. 2017, vol. 10168, SPIE; 2017, p. 247–53. https://doi.org/10.1117/12.2258614.
[26] Sohn H-G, Lim Y-M, Yun K-H, Kim G-H. Monitoring Crack Changes in Concrete Structures. Comput-Aided Civ Infrastruct Eng 2005;20:52–61. https://doi.org/10.1111/j.1467-8667.2005.00376.x.
[27] Lin Y, Nie Z, Ma H. Structural Damage Detection with Automatic Feature-Extraction through Deep Learning. Comput-Aided Civ Infrastruct Eng 2017;32:1025–46. https://doi.org/10.1111/mice.12313.
[28] Jahanshahi MR, Kelly JS, Masri SF, Sukhatme GS. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. Struct Infrastruct Eng 2009;5:455–86. https://doi.org/10.1080/15732470801945930.
[29] Koch C, Georgieva K, Kasireddy V, Akinci B, Fieguth P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv Eng Inform 2015;29:196–210. https://doi.org/10.1016/j.aei.2015.01.008.
[30] Ito A, Aoki Y, Hashimoto S. Accurate extraction and measurement of fine cracks from concrete block surface image. IEEE 2002 28th Annu. Conf. Ind. Electron. Soc. IECON 02, vol. 3, 2002, p. 2202–7 vol.3. https://doi.org/10.1109/IECON.2002.1185314.

[31] Fathalla E, Tanaka Y, Maekawa K. Remaining fatigue life assessment of in-service road bridge decks based upon artificial neural networks. Eng Struct 2018;171:602–16. https://doi.org/10.1016/j.engstruct.2018.05.122.

[32] Davoudi R, Miller GR, Kutz JN. Data-driven vision-based inspection for reinforced concrete beams and slabs: Quantitative damage and load estimation. Autom Constr 2018;96:292–309. https://doi.org/10.1016/j.autcon.2018.09.024.

[33] Farhidzadeh A, Dehghan-Niri E, Moustafa A, Salamone S, Whittaker A. Damage Assessment of Reinforced Concrete Structures Using Fractal Analysis of Residual Crack Patterns. Exp Mech 2013;53:1607–19. https://doi.org/10.1007/s11340-013-9769-7.

[34] Carrillo J, Dominguez D, Garcia NP. Damage Index Based on Fractal Dimension of Cracking on Thin Reinforced Concrete Walls. Struct J 2017;114. https://doi.org/10.14359/51700919.

[35] Liu Y, Dai K, Li D, Luo M, Liu Y, Shi Y, et al. Structural performance assessment of concrete components based on fractal information of cracks. J Build Eng 2021;43:103177. https://doi.org/10.1016/j.jobe.2021.103177.

[36] Mandelbrot B. Les objets fractals. 4th ed. Paris: ChampsFlamarion,: 1995.

[37] Teles S, Lopes AR, Ribeiro MB. Fractal Analysis of the UltraVISTA Galaxy Survey. Phys Lett B 2021;813:136034. https://doi.org/10.1016/j.physletb.2020.136034.

[38] Evertsz CJ g. Fractal geometry of financial time series. Fractals 1995;03:609–16. https://doi.org/10.1142/S0218348X95000539.

[39] Multifractal analysis of crack patterns in reinforced concrete shear walls - Arvin Ebrahimkhanlou, Alireza Farhidzadeh, Salvatore Salamone, 2016 n.d. https://journals.sagepub.com/doi/abs/10.1177/1475921715624502 (accessed March 16, 2022).

[40] Ebrahimkhanlou A, Athanasiou A, Hrynyk TD, Bayrak O, Salamone S. Fractal and Multifractal Analysis of Crack Patterns in Prestressed Concrete Girders. J Bridge Eng 2019;24:04019059. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001427.

[41] Raghavendra BS, Narayana Dutt D. A note on fractal dimensions of biomedical waveforms. Comput Biol Med 2009;39:1006–12. https://doi.org/10.1016/j.compbiomed.2009.08.001.

[42] Heymans O, Fissette J, Vico P, Blacher S, Masset D, Brouers F. Is fractal geometry useful in medicine and biomedical sciences? Med Hypotheses 2000;54:360–6. https://doi.org/10.1054/mehy.1999.0848.

[43] Yao B, Imani F, Sakpal AS, Reutzel EW, Yang H. Multifractal Analysis of Image Profiles for the Characterization and Detection of Defects in Additive Manufacturing. J Manuf Sci Eng 2018;140. https://doi.org/10.1115/1.4037891.

[44] Badii R, Politi A. Hausdorff Dimension and Uniformity Factor of Strange Attractors. Phys Rev Lett 1984;52:1661–4. https://doi.org/10.1103/PhysRevLett.52.1661.

[45] Mandelbrot B. Les objets fractals: Forme, hasard et dimension. Paris, France:Flammarion: 1975.

[46] Clarke KC. Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. Comput Geosci 1986;12:713–22. https://doi.org/10.1016/0098-3004(86)90047-6.

[47] Thomas TR, Rosén B-G. Implementation of Whitehouse's method for calculating properties of self-affine fractal profiles. Proc Inst Mech Eng Part C J Mech Eng Sci 2008;222:1547–50. https://doi.org/10.1243/09544062JMES938.

[48] Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI. Fractal measures and their singularities: The characterization of strange sets. Nucl Phys B - Proc Suppl 1987;2:501–11. https://doi.org/10.1016/0920-5632(87)90036-3.

[49] Chhabra null, Jensen null. Direct determination of the f(alpha) singularity spectrum. Phys Rev Lett 1989;62:1327–30. https://doi.org/10.1103/PhysRevLett.62.1327.

[50] Lopes R, Betrouni N. Fractal and multifractal analysis: A review. Med Image Anal 2009;13:634–49. https://doi.org/10.1016/j.media.2009.05.003.

[51] Ebrahimkhanlou A, Farhidzadeh A, Salamone S. Multifractal analysis of two-dimensional images for damage assessment of reinforced concrete structures. In: Lynch JP, editor., San Diego, California, United States: 2015, p. 94351A. https://doi.org/10.1117/12.2084052.

[52] Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition. n.d. https://hastie.su.domains/ElemStatLearn/ (accessed April 18, 2022).
[53] Mangalathu S, Jeon J-S, DesRoches R. Critical uncertainty parameters influencing seismic performance of bridges using Lasso regression. Earthq Eng Struct Dyn 2018;47:784–801. https://doi.org/10.1002/eqe.2991.
[54] Hastie T, Tibshirani R, Friedman J. Linear Methods for Regression. In: Hastie T, Tibshirani R, Friedman J, editors. Elem. Stat. Learn. Data Min. Inference Predict., New York, NY: Springer; 2009, p. 43–99. https://doi.org/10.1007/978-0-387-84858-7_3.
[55] Hastie T, Tibshirani R, Friedman J. Support Vector Machines and Flexible Discriminants. In: Hastie T, Tibshirani R, Friedman J, editors. Elem. Stat. Learn. Data Min. Inference Predict., New York, NY: Springer; 2009, p. 417–58. https://doi.org/10.1007/978-0-387-84858-7_12.
[56] Feng D-C, Wang W-J, Mangalathu S, Taciroglu E. Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. J Struct Eng 2021;147:04021173. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115.
[57] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., New York, NY, USA: Association for Computing Machinery; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.
[58] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier 2016. https://doi.org/10.48550/arXiv.1602.04938.
[59] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 2014;41:647–65. https://doi.org/10.1007/s10115-013-0679-x.
[60] Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. ArXiv170507874 Cs Stat 2017.
[61] Rice JA. Mathematical Statistics and Data Analysis. Cengage Learning; 2006.
[62] Adom-Asamoah M, Banahene JO. Nonlinear seismic analysis of a super 13-element reinforced concrete beam-column joint model. Earthq Struct 2016;11:905–24. https://doi.org/10.12989/EAS.2016.11.5.905.
[63] Dai K, Li D, Luo M, Shi Y, Zhang S, Huang Z. Machine vision-based concrete beam crack pattern identification using fractal theory. Nondestruct. Charact. Monit. Adv. Mater. Aerosp. Civ. Infrastruct. Transp. XIII, vol. 10971, SPIE; 2019, p. 232–7. https://doi.org/10.1117/12.2514124.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: