

## Review

# Determining Uncertainties in AI Applications in AEC Sector and their Corresponding Mitigation Strategies

Yi An <sup>a</sup>, Haijiang Li <sup>a,\*</sup>, Tengxiang Su <sup>a</sup>, Yitong Wang <sup>b</sup>

<sup>a</sup> Cardiff School of Engineering, Cardiff University, UK

<sup>b</sup> China International Engineering Consulting Corporation, China

## ARTICLE INFO

## Keywords:

AI uncertainty  
Optimization guidelines  
AEC industry  
Artificial neural networks  
Support vector machine  
Genetic algorithm  
Fuzzy logic

## ABSTRACT

The Artificial Intelligence (AI) methodologies and techniques have been used to solve a wide spectrum of engineering problems in Architectural, Engineering and Construction (AEC) industry with the aim of improving overall productivity and optimized decision throughout full project life cycle (planning, design, construction and maintenance). However, many AI applications are facing different limitations and constraints due to the lack of comprehensive understanding about the inherent uncertainty fundamentally and mathematically, hence the use of AI has not achieved a satisfactory level. It requires different actions to tackle different types of uncertainties which varies according to different types of applications. This paper therefore reviews 5 type of popular AI algorithms, including Primary Component Analysis, Multilayer Perceptron, Fuzzy Logic, Support Vector Machine and Genetic Algorithm; then examines how these artificial intelligence techniques can assist the decision-making process by mitigating uncertainty meanwhile achieving the expected high efficiency. The paper reviews each germane technique, mathematical explanation, analysis of reasons causing uncertainty, and concludes a set of guidelines and an application framework for optimizing their informed uncertainty for AEC applications. This work will pave the way for the fundamental understanding and in turn to provide a valuable reference for applying AI techniques in AEC sector properly to achieve better overall performance.

## 1. Introduction

Artificial Intelligence (AI) is a branch of computer science that aims to enable computers to perform human-like tasks [1]. It is a technique that enables machines independently interpret and learn from external data to achieve specific outcomes via flexible adaptation. Generally, the AI techniques are in the form of perception, reasoning, planning, motion and natural language processing, whereof they are categorized into 7 common domains, Symbolic Mathematics, Game Playing, Neutral Networks, Expert Systems, Fuzzy Logic, Robotics and Natural Language Processing [2]. The advances of artificial intelligence have been well and detailed documented in the vast literatures. For instance, it can use sophisticated algorithms to “learn” from “big” data, and the use the knowledge gained to assist industry/practice [3]. Because of this, In the last two decades, AI techniques managed to attract substantial attention within the Architectural, Engineering, and Construction (AEC) industry, which, according to Architecture, Engineering & Construction Management Institute, refers to the collective term for business in three

partner industries (architecture, engineering and construction) and its objectives are the complexities and need of the building spaces [4]. Many scholars did researches about the AI application in the AEC industry on the topic of construction robotics, safety management, supply chain management etc. This has led to an upsurge in the number of research and publications on AI in the AEC industry. This situation presents danger, as the main focus is on the application layer—trying to set the framework of applying all kinds of AI techniques to certain activities and developing the smarter platforms/systems/algorithms/models to assist design and construction process of buildings, posing a major risk of neglecting essential areas—and questions specifically uncertainty related for research and practice improvement. The extensive research papers mostly conducted the uncertainty analysis while developing/applying AI algorithms to AEC tasks without comprehensively and critically analyze the reasons caused uncertainty, which would lead unclear improvement room for the following scholars.

Meanwhile, AI techniques in AEC industry and their applications have been summarized in several review papers. Levitt and Kartam

\* Corresponding author at: Chair in BIM Smart Engineering, Cardiff School of Engineering, Cardiff University, Queen's Building, The Parade, Cardiff CF24 3AA, Wales, UK.

E-mail address: [LiH@cardiff.ac.uk](mailto:LiH@cardiff.ac.uk) (H. Li).

<https://doi.org/10.1016/j.autcon.2021.103883>

Received 15 January 2021; Received in revised form 31 July 2021; Accepted 8 August 2021

Available online 16 August 2021

0926-5805/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

summarized the drawbacks of traditional construction planning tools, then critically reviewed and analysis the AI techniques in terms of construction planning [5]. Irani and Kamal profiled intelligent systems application as well as studies in the construction industries [6]. Faghihi et al. reviewed the artificial intelligence and optimization tools related researches (from 1985 to 2014) on automation in construction scheduling, the review claim the genetic algorithms (GA) is the primary approach comparing with case-based reasoning (CBR), knowledge-based approaches, model-based approaches, genetic algorithms, expert systems (ES), neural networks (NN) and other methods [7]. Bilal et al. unveiled big data analytics in the AEC industry, discussed both pitfalls and opportunities of big data technologies for civil engineering [8]. Yu and Liao made review on intuitionistic fuzzy studies on broad construction activities including planning and robotics [9]. Shukla et al. comprehensively reviewed the engineering applications of artificial intelligence for the 30 years up to 2018 [10]. Sacks et al. reviewed AI from the perspective of construction technology innovations and thoroughly analyze the links between building information modelling and AI [11]. Khallaf et al. conducted systematically review to analyze and classify deep learning applications in construction, the review found that deep learning applications are mostly used in crack detection and most popular algorithm in general is convolution neural networks [12]. Geyer et al., through reviewing the major work about fusing data, engineering

knowledge and artificial intelligence in the built environment, reveal the links between big data technologies and AEC industry [13]. Though in the 9 reviews, they have made valuable contributions. Certain limitations exist, most of these have been qualitative and based on manual appraisals. Hence, they may be significantly impacted by subjective biases, lack of reproducibility, and reduced reliability; the review studies have had narrowed perspectives. To overcome such issue, scholars like Darko et al. undertook a rigorous review, provided a comprehensive survey about the intellectual core and the landscape of the general body of knowledge on AI in AEC industry using a quantitative technique [14]. However, their study focuses on the whole AEC industry while they eliminate subjective biases, and all these studies neglected the review of existence of uncertainty and its caused reasons in the application. Therefore, existing review did not afford a full picture of the state-of-the-art research on future directions for the AI application. As matter of fact, a study that offers the understanding uncertainty of AI application literature in the AEC domain is still missing.

As an attempt to fill in gap and then enhance such AI applications, the present study protrudes, being the first one to undertake a rigorous analysis of AI uncertainty and understand its driven forces in AEC industry from a perspective that considered both the industrial and mathematical context. This study focuses on the uncertainty, its mathematical explanation, and their potential to be improved of AI in AEC

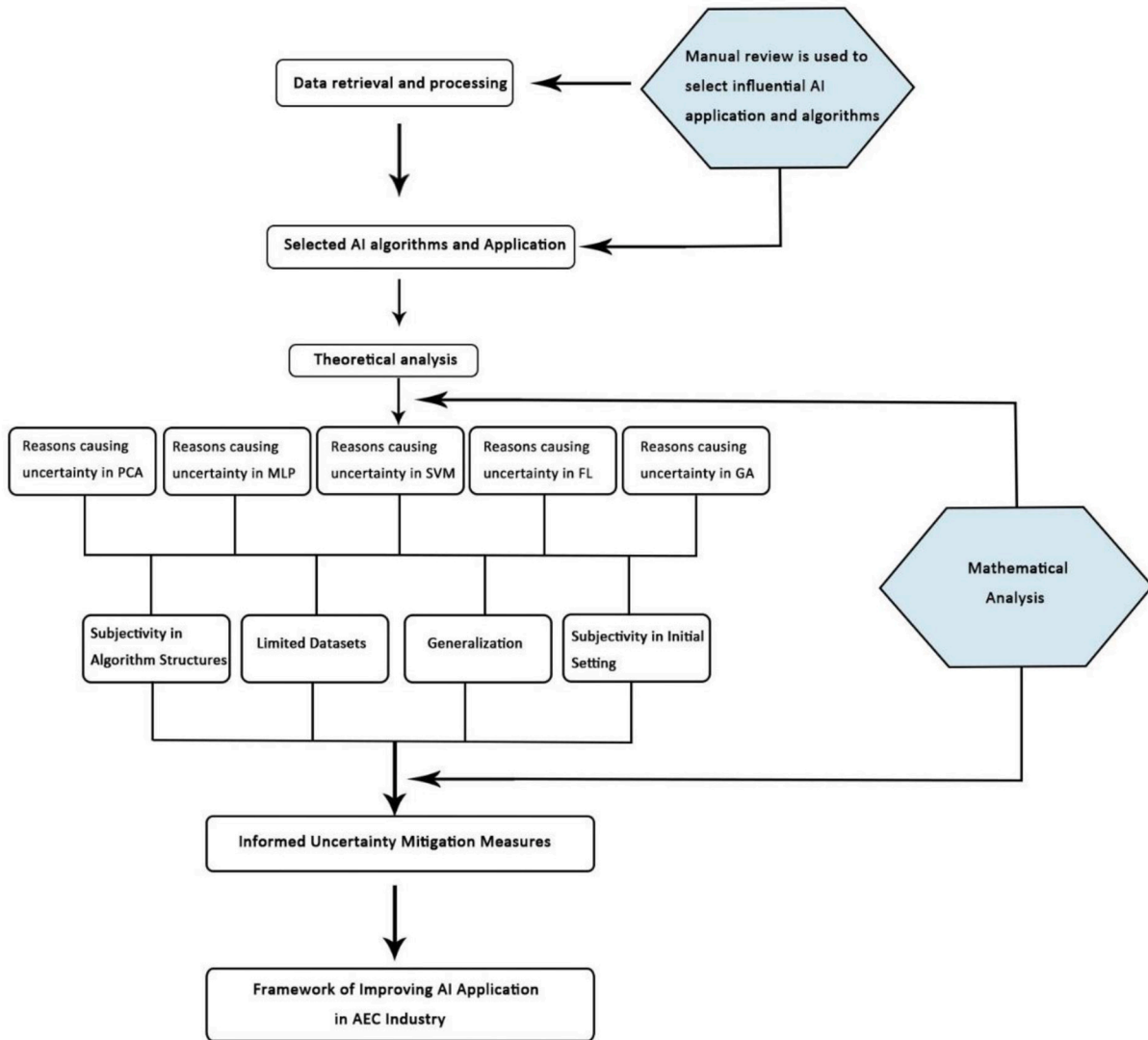


Fig. 1. Outline of research design.

industry. In Section 2, research methodology is presented. In Section 3, 5 most common AI algorithms applied in AEC sector are identified and introduced. Section 4 presents the uncertainty affiliated with popular AI techniques, and highlights the analysis of the explicit explanation from the perspective of mathematics as the results. Section 5 discusses techniques and guidelines of optimization AI application from the 7 general domains as aforementioned in AEC sector associating with target causes to the uncertainty through the analysis of 5 commonly used algorithms. Section 6 concludes the findings as well as prevail challenges of optimizing the AI techniques in the AEC industry, and then proposes the key research directions to be addressed in the future.

## 2. Research methodology

This paper provides an in-depth review and analysis through the combination of mathematical analysis with the critical review of AI algorithm application. Fig. 1 shows the study's research design. The manual review is used to objectively reveal the commonly applied influential AI algorithms in AEC industry while mathematical analysis aims to identify the reasons causing uncertainty within AI application and the corresponding mitigation measures based on the analysis results. Follow upon that, the framework of improving AI application could be developed.

The literatures retrieval started with the major databases including Scopus, Web of Science, ScienceDirect, Google Scholar as the first screening by selecting keywords "Artificial Intelligence", "AEC industry", "machine learning", "expert system", "robotics", "neural networks", "natural language process", "civil engineering". After the first screening, the combination of different literature databases is not applicable due to the generated difficulties in checking and eliminating duplications of publications from the various databases, with the large amount of dataset involved. Hence, the second round of screening stayed with Scopus. The rationale behind this is that compared to the other databases, Scopus possesses a wider range of scientific publication coverage. The literature search in Scopus using the selected search keywords was performed on the title, abstract, and keywords sections of publications; the date range was set from 2009 to present; the "document type" was limited to "article" and "conference paper". The rationale for limiting date range is that both construction industry and artificial intelligence got involved in construction industry and produce significant attempts since 2009. Also, the rationale of limiting document type is that, in the AEC industry, for science referencing purpose, journal articles represent the most influential and reputable research work; and for artificial intelligence, some conference papers represent the most influential and reputable research work. As of Oct 14th, 6719 publications were initially identified. Scopus was used to sort these upon "relevance", and after assessing them and excluding irrelevant publication, e.g., publications in unrelated journals, then considering the citation descent with the chronology descent selecting the more influential ones and manually read the abstracts to exclude those that are not related to the areas of AEC industry. In total, 58 publications are identified in the end.

The mathematical analysis was conducted by, firstly, identifying the proper way of explaining the algorithms, in which, proper way means the more obvious way to express algorithms comprehensively as well as explore uncertainty logically for scholars and engineers of AEC industry. Accordingly, among all the derivatives of algorithms' definition along with expression, the Hotelling's method is used to express Principal Component Analysis (PCA) [15], Rumellhart and McClelland's way to Multilayer Perceptron (MLP) [16], Boser and Guyon' development to Support Vector Machine (SVM) [17], Zadeh's theory to Fuzzy Logic (FL) [18], Goldberg's contribution to Genetic Algorithm (GA) [19]. Secondly, by integrating the application status quo with the mathematical explanation, the detailed reasons caused uncertainty were figured and then summarized into 4 categories. Based on that, mitigation measures to the informed reasons were identified. Finally, all the findings through the

mathematical analysis are emerged as a framework of improving AI application in AEC industry.

## 3. Research model

The studied methods are selected from the existing conclusive researches, those clearly revealed the most frequent applied AI algorithms in AEC sectors, namely Principal Component Analysis, Multilayer Perceptron, Support Vector Machine, Fuzzy Logic and Genetic Algorithm [10,14,20]. This pooling allows us to globally observe the main 5 algorithms working philosophies to discover what reasons are to deliver the uncertainty to application results related to AEC industry. Unlike some outstanding researchers such as Abdolbaghi et al. and Najafi-Marghmaleki et al., the ideology of their research is also extensively reviewing the AI methods, they would focus on more specific algorithms and associated optimizations through the actual applications with statistical quality measure approaches [21,22]. This paper would shift the from application to theoretical analysis, explaining the driven forces underneath the recent AI developments in AEC industry. For each algorithm, a brief development of the principles and of its mathematical model is introduced in this section. Followingly, along with the introduction of algorithm, the literatures about AI application in AEC are reviewed to help understand as well as verify the finds like why the algorithm could be used targeting specific tasks. This explanation serves as a basis for addressing, from a formal and rigorous perspective, the main driven forces, which could easily give an idea of reasons causing uncertain for the final results afterwards. A good understanding of the theoretical basis and mathematical foundations of each method allows to mitigate the uncertainty that may appear during the modelling and exploitation phases.

### 3.1. Principal component analysis

Pearson invented PCA in 1901 as an analogue of the principal axis theorem in mechanics [23]. Later, Hotelling's input officialized it as a formal method [15]. Since then, PCA has proved in many ways of forming the basis for multivariate data analysis. It provides an approximation of a data table, a data matrix. Taken from the Hotelling's derivation of PCA, the mathematical definition of this method would be:

For a set of observe  $d$ -dimensional data vectors  $\{t_n\}$ ,  $n \in \{1 \dots N\}$ ;

The  $q$  principal axes  $w_j$ ,  $j \in \{1 \dots q\}$ , are those orthonormal axes onto which the retained variance under projection is maximal.

It can be shown that the vectors  $w_j$  are given by the  $q$  dominant eigenvectors (those with the largest associate eigen values  $\lambda_j$ ) of the sample covariance matrix:

$$S = E[(t - u)(t - u)^T] \quad (1)$$

Such that

$$S w_j = \lambda_j w_j \quad (2)$$

Note: the  $q$  principal components of the observed vector  $t_n$  are given by the vector  $x_n = W^T(t_n - u)$ , where  $W^T = (w_1, w_2, \dots, w_q)^T$ . The variables  $x_j$  are then decorrelated such that the covariance matrix  $E[xx^T]$  is diagonal with elements  $\lambda_j$ .

To the above mathematical definition from Hotelling, based on the property A3 (spectral decomposition of covariance matrix) from Jolliffe [24], another derivation of covariance matrix is given as:

$$S = \lambda_1 \alpha_1 \alpha_1' + \lambda_2 \alpha_2 \alpha_2' + \dots + \lambda_n \alpha_n \alpha_n' \quad (3)$$

it could also be expressed with summation notation as:

$$S = \sum_{i=1}^n \lambda_i \alpha_i \alpha_i' \quad (4)$$

With Eq. (3), The combined variances of all the elements of  $x_n$  are

decomposed into decreased contributions due to each principal component with  $\lambda_n \alpha_n \alpha'_n$ , and this item would be decreasing as  $n$  increases. With the dimension reduction point of view, presumably speaking, the whole phase space is  $n$  dimension, and the desired computation capacity is  $k$  dimension, where  $k$  is integer and  $k < n$ . The Eq. (4) would be rewritten as:

$$S = \sum_{i=1}^k \lambda_i \alpha_i \alpha'_i + \sum_{i=k+1}^n \lambda_i \alpha_i \alpha'_i \quad (5)$$

In the Eq. (5), the former part  $\sum_{i=1}^k \lambda_i \alpha_i \alpha'_i$  is considered as the principal components as well as the principal  $k$  dimensional phase space, the latter part  $\sum_{i=k+1}^n \lambda_i \alpha_i \alpha'_i$  are regarded as non-principal parts due to the smaller values and its minor contributions, hence the according factors of these covariances could be eliminated as the noise in order to improve the efficiency. Based on the Linsker's research, there is another derivation of Eq. (5) which is widely used in the information theory. Linsker suggested that in general PCA-based dimensionality reduction tends to minimize that information loss, under certain signal and noise models [25]. With such assumption, Eq. (5) can be reformed as below:

$$x = s + n \quad (6)$$

that is, from perspective of information theory, the data vector  $x$  is the linear combination of the effective information bearing signal  $s$  and noise signal  $n$ .  $s$  represents the vectors (factors) those link to the first part of Eq. (5)  $\sum_{i=1}^k \lambda_i \alpha_i \alpha'_i$ , and  $n$  stands for the vectors (factors) those link to the last part of Eq. (5)  $\sum_{i=k+1}^n \lambda_i \alpha_i \alpha'_i$ .

The principal component analysis is a method for classification through reducing data dimension. Also, it can be employed to optimize other algorithms by decreasing the sample vectors size. The components with more contributions of the data set are the directions align with greater variance and underlying structures that variance represent. The way in which the variance is distributed in the data cloud gives an idea of the directions in which there is more information. Those in which the samples are more spread out will have more relevance than those that are constant. The principal components allow to summarize the data eliminating the redundant or little differentiating information and highlighting the one that is likely to be more important [26]. The principal component analysis (PCA) is employed for the optimization of neural networks in the case of structural design and engineering, construction techniques [20,27]. Moreover, it is used to promote the expert system in the case of structural design and engineering [28]. The algorithm could help to determine which features are the most important ones during the decision making, therefore, reduce the range of possibilities as well as noise contained in the redundant datasets.

### 3.2. Multilayer perceptron

A multilayer perceptron is a supervised learning algorithm that learns a nonlinear function  $y$  training on a labeled dataset which can be employed to perform classifications and regression [29]. The graphical topology of the perceptron is define as three layers in general: input layer, perceptrons within it receive data from an external source; output layer, perceptrons within it return the results; hidden layer, perceptrons within it are assigned weight factors and cannot communicate with indirect perceptrons [30]. Multilayer perceptron employs backpropagation as learning functions for training.

A typical multi perceptron would be trained to learn a function:

$$Y = f(X) : \mathbb{R}^m \rightarrow \mathbb{R}^k \quad (7)$$

Where  $X = \{x_i | i \in 1 \dots m\}$  is the input vector,  $m$  is the size of the input vector,  $Y = \{y_j | j \in 1 \dots k\}$  is the output vector and  $k$  is the size of the output vector.

Given the  $X$  and  $Y$ , the hidden layer would perform the required calculations with the activation function to approximate the function  $f$

$(\cdot)$ , each node of output layer and hidden layer receives data from the previous perceptrons, calculates a weighted linear summation and returns the result of its non-linear activation function as follow:

$$h\left(\sum_{i=1}^l w_i p_i\right) : \mathbb{R}^l \rightarrow \mathbb{R} \quad (8)$$

where  $h(\cdot)$  is the activation function,  $l$  is the size of input vector for this node,  $W = \{w_i / i \in 1 \dots l\}$  is the weights vector,  $P = \{p_i / i \in 1 \dots l\}$  is the input vector for the node.

The first hidden layer receives the input data from the input layer and the last hidden layer input data to the output layer. The weights vector values  $W$  of all perceptrons are learned by backpropagation algorithm training. Fig. 2 shows a network topology of 5-layer multilayer perceptron to learn  $\mathbb{R}^8 \rightarrow \mathbb{R}^4$  functions.

Rumellhart and McClelland defined propagation rule for multi-layer perceptron as follow [16],

$$r = (d_j - f(w_j^T x)) f'(w_j^T x) = (d_j - o_j) f'(net_j) \quad (9)$$

Where  $d_j$  is expected output,  $o_j$  is perceptron output vector, representing the output values of the last nodes and combines it with the connectivity matrices to produce a net input for each type of input into the previous node. It defines as:

$$O_j = f(w_j^T x) \quad (10)$$

The  $net_j$  means the net input into node  $j$ . The net input is defined as:

$$net_j = w_j^T x \quad (11)$$

Where  $w$  represents the weights matrices among the hidden layers and output layer.  $j$  represents the different iteration process.

Define the square error between the perceptron output and expected output as:

$$E = \frac{1}{2} (d_j - o_j)^2 = \frac{1}{2} (d_j - f(w_j^T x))^2 \quad (12)$$

To optimize the  $E$ , Gradient descent would be employed:

$$\Delta w_j = -\eta \nabla E \quad (13)$$

In which ratio  $\eta$  is a positive constant and  $\nabla E$  is:

$$\nabla E = -(d_j - f(w_j^T x)) f'(w_j^T x) x \quad (14)$$

Hence,

$$\Delta w_j = \eta (d_j - o_j) f'(net_j) x \quad (15)$$

The  $f(\cdot)$  is the differential form of activation function  $f(\cdot)$  which remain consistent with the one in Eq. (7). And there are 5 common ways to determine the activation function, four of them are predefined and one is self-defined.

The first predefined activation function—Sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (16)$$

The second predefined activation function—Tanh function:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

The third predefined activation function—Softsign function:

$$f(x) = \frac{x}{1 + |x|} \quad (18)$$

The fourth predefined activation function—ReLU function:

$$f(x) = \max(0, x) \quad (19)$$

Due to the capability of learning nonlinear models and the reaching

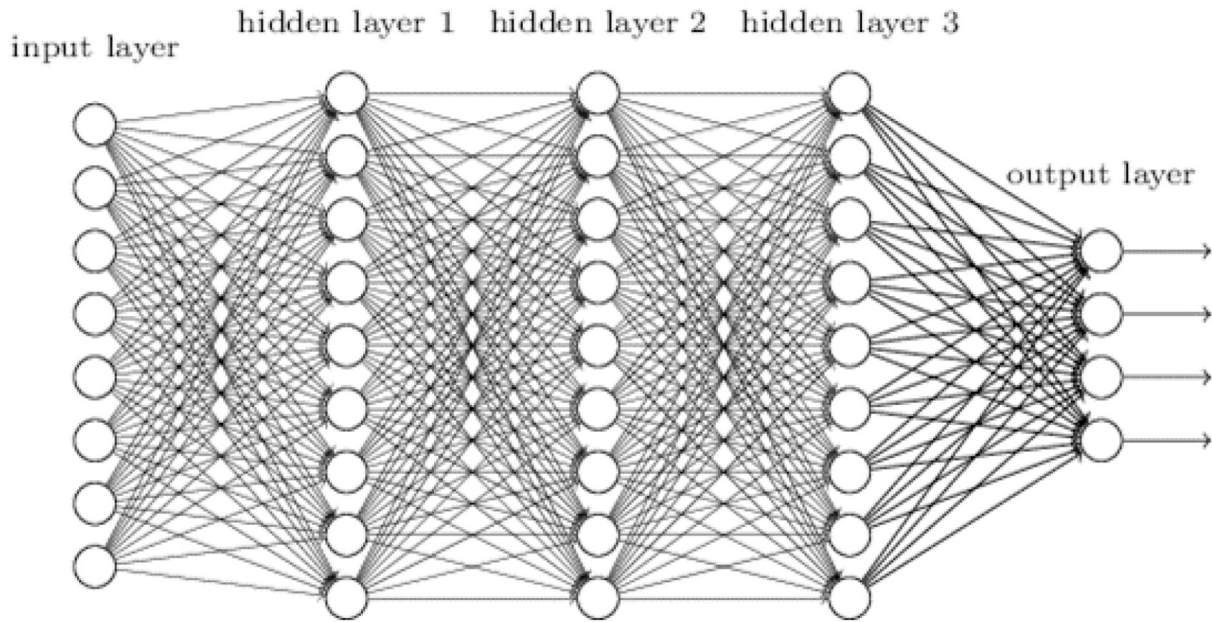


Fig. 2. A network topology of 5-layer multilayer perceptron to learn  $\mathbb{R}^8 \rightarrow \mathbb{R}^4$  functions.

high accuracy with great complexity, the multi-layer perceptron were frequently employed in project planning, technical design and facility management within AEC industry [31–42]. But the applications of Artificial neural network (ANN) covered the limited tasks rather than the wide range as expected, such as Abdolbaghi et al. proposed a new computer based model with multilayer perceptron neural network for prediction of the viscosity of  $\text{CO}_2$ , the new developed model optimized average absolute relative deviation to 0.842% [21]. The main reason of that are insufficient datasets due to lack of Information and Communications Technology (ICT) infrastructures in AEC industry and fuzzy understanding of decision theory for complex activities in AEC industry. In the project planning, execution plan, budget plan and site surveys are the most mentioned topic aligning with neural networks. Since the processable data for neural networks does not necessarily obtain from high-end ICT facilities. And the hierarchy for decision making are not complicated. Taking example of Cheng et al., they research on estimating completion for construction projects using neural networks, the most features of input vectors are simply linear correlated, such as fluctuation, actual cost, planned cost, contract payment, etc. [43]. And these data do not heavily rely on the advanced ICTs to obtain and documented. Hence, the relatively adequate applicable datasets and less complexity of problems enabled neural networks application. The same reason applies to the technical design and facility management.

### 3.3. Support vector machine

Support vector machine is a neural network algorithm with single hidden layer. It was proposed by Vapnik and Chervonekis, and further developed by Boser and Guyon later on. The intention of this development is to better perform data analysis for classification and regression [17]. It is presently one of the best-known classification techniques with computational advantages over their contenders. The support vector machine is able to handle both linear and nonlinear decision boundaries of arbitrary complexity. The hard margin SVM and soft margin SVM are used to deal with linear cases, in which soft margin SVM is more robust since it considered the noise within training data. As for the nonlinear cases, a kernel function would be employed. To set the logic more unequivocal, this subsection is limited to start with soft margin SVM because of its superior to hard margin SVM. The soft margin SVM is described with Cortes variant, the quadratic program is given as follow:

Inputs: Training examples  $\{x_1, x_2, \dots, x_l\}$  and class labels  $\{y_1, y_2, \dots, y_l\}$ .  
Minimize over  $\alpha_k$ :

$$J = \left( \frac{1}{2} \right) \sum_{hk} y_h y_k \alpha_h \alpha_k (x_h \cdot x_k + \lambda \delta_{hk}) - \sum_k \alpha_k \quad (20)$$

Subject to:

$$0 \leq \alpha_k \leq C \text{ and } \sum_k \alpha_k y_k = 0 \quad (21)$$

Output: parameters  $\alpha_k$ .

The summation run over all training patterns  $x_k$  that are  $n$  dimensional feature vectors,  $x_h \cdot x_k$  denotes the scalar product,  $y_k$  encodes the class label as a binary value  $+1$  or  $-1$ ,  $\delta_{hk}$  is the Kronecker symbol, and  $\lambda$  and  $C$  are positive constants (soft margin parameters). The soft margin parameters ensure convergence even when the problem is wrongfully labeled or poorly conditioned. The resulting decision function of an input vector  $x$  is:

$$D(x) = w \cdot x + b \quad (22)$$

With

$$w = \sum_k \alpha_k y_k x_k \quad (23)$$

and

$$b = \langle y_k - w \cdot x_k \rangle \quad (24)$$

The decision boundary (a straight line in the case of a two-dimensional separation) is positioned to leave the largest possible margin on either side. A particularity of SVM is that the weights  $w_i$  of the decision function  $D(x)$  are a function only of a small subset of the training examples—support vectors. Those are closest to the decision boundary and lie on the margin. The weight vector  $w$  is a linear combination of training patterns. Most weights  $\alpha_k$  are zero. The training patterns with non-zero weights are support vectors. Those with weight satisfying the strict inequality  $0 \leq \alpha_k \leq C$  are marginal support vectors. The bias value  $b$  is an average over marginal support vectors [17].

To further, considering the more simplified differentiation, the decision function could be shortened with separated variable  $\frac{1}{2} \|w\|^2$ . Hence, with Eqs. (20)–(24), the objective function could be transformed to:

Minimize over  $w, b$ :

$$J = \frac{1}{2} \|w\|^2 \quad (25)$$

Subject to:

$$y_i(w \cdot x_i + b) \geq 1 \quad (26)$$

The most cases have nonlinear separable features in the real world, so does the AEC industry. The main reason to that is either wrongfully labeled—noise exists in the input or the perfect features with nonlinear separable features per se. For the first reason, the slack variable ( $\varepsilon_i \geq 0$ ) is added to an inequality constraint to transform the problem for excluding outlier. It is given as:

$$y_i(w \cdot x_i + b) \geq 1 - \varepsilon_i \quad (27)$$

With Eq. (26), the Eq. (25) could be reformed as:

Minimize over  $w, b, \varepsilon_i$ :

$$J = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \varepsilon_i \quad (28)$$

$C$  is the penalty parameter. Higher value  $C$  would assign more penalty to the wrongfully label data. With the Eqs. (27) and (28), the objective function is equivalent to:

Minimize over  $w, b$

$$J = \sum_{i=1}^N [1 - y_i(w \cdot x_i + b)] + \lambda \|w\|^2 \quad (29)$$

The first part of Eq. (29) is the hinge loss function, define as:

$$L(y(w \cdot x + b)) = [1 - y(w \cdot x + b)] \quad (30)$$

Considering the  $C$  is the constant (equals to  $\frac{1}{2\lambda}$ ) and the slack variable followed the hinge function, meanwhile integrating Eqs. (30) into (29), the simplified form of Eq. (29) could be rewritten as:

$$J = \sum_{i=1}^N \varepsilon_i + \lambda \|w\|^2 \quad (31)$$

As for the feature itself is nonlinear separable cases, then it assumed that hyperspace exists. Then the features could be transformed to the higher limited dimensional Hilbert space ( $H$ ) for formulating another linear separable feature space from the original feature space. In such cases, the hypersurface is expressed as:

$$w^T \phi(x) + b = 0 \quad (32)$$

In which:  $\phi: x \rightarrow H$  is the mapping function.

With Eq. (32) it could easily cause the curse of dimensionality. To avoid that, the kernel method is employed with kernel function to define the inner product of mapping function. There are 4 common kernel functions, e.g., Polynomial kernel function, Sigmoid Kernel function, Radial basis kernel function, Laplace kernel function.

Polynomial kernel function is defined as:

$$\kappa(x, x_i) = ((x \cdot x_i) + 1)^d \quad (33)$$

Sigmoid Kernel function is defined as:

$$\kappa(x, x_i) = \tanh[a(x^T x_i) - b], a, b > 0 \quad (34)$$

Radial basis kernel function is defined as:

$$\kappa(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (35)$$

Laplace kernel function is defined as:

$$\kappa(x, x_i) = \exp\left(-\frac{\|x - x_i\|}{\sigma}\right) \quad (36)$$

The SVM are widely employed in the AEC industry for its efficiency and accuracy with small samples. There are three reason that, for ① same as the multi-layer perceptrons, the data from these stages could be easily modelled linearly; ② different from the multi-layer perceptrons, the limited applicable data was the disadvantage of MLP but it would not cause much troubles for SVM since the SVM demonstrated its excellence dealing with small samples; ③ The kernel function provides a reliable mathematical foundation for nonlinear problems. Nevertheless, potentials to enhance support vector machine with AEC industry input are huge. Still, the learning samples of AEC industry is limited. Hence, the SVM was frequently discussed in the technical design and facility management [44–50]. Most of discussion emerged a pattern that Najafi-Marghaleki et al. presented earlier, that fuzzy logic, expert system (regarded as a derivation of expert system) and particle swarm optimization hybridizing with SVM so that could produce better performance while conduct prediction [22].

### 3.4. Fuzzy logic

Fuzzy logic theory, introduced by Lotfi Zadeh, provides a means to capture uncertainty. The underlying power of fuzzy set theory is that it uses “linguistic” variables rather than quantitative variables to represent imprecise concepts. It is very promising for its expressivity when decision making process involves human reasoning. It is used widely in applications that do not require precision but depend on intuition [18]. All objects of the universe are subject to set membership. In crisp set, the membership function of objects is binary with 0 and 1. In fuzzy set, the membership function is a series of value between 0 and 1 corresponding to objects in the fuzzy set. Fuzzy logic attempts to provide a mathematical framework to such linguistic statements for further reasoning. When the universe of  $x$  is a continuous interval, a fuzzy set is represented as:

$$A = \left\{ \int \frac{u_A(x)}{x} \right\} \quad (37)$$

Where the integral operator indicates continuous function-theoric union; the horizontal demarcating line separates the membership values and the corresponding points and is in no way related to division. When the universe is a collection of a finite number of ordered discrete points, the corresponding fuzzy set may be represented by:

$$A = \left\{ \sum_{i=1}^n \frac{u_A(x_i)}{x_i} \right\} \quad (38)$$

Where the summation indicates aggregation of elements. Basic operations related to fuzzy subsets  $A$  and  $B$  of  $X$  having membership functions  $u_A(x)$  and  $u_B(x)$  are

1.  $A$  is equal to  $B$ .

$$u_A(x) = u_B(x) \forall x \in X \quad (39)$$

2.  $A$  is a complement of  $B$ .

$$u_A(x) = u_{\bar{B}}(x) = 1 - u_B(x) \forall x \in X \quad (40)$$

3.  $A$  is contained in  $B$  ( $A \subseteq B$ ).

$$u_A(x) \leq u_B(x) \forall x \in X \quad (41)$$

4. The union of  $A$  and  $B$  ( $A \cup B$ ).

$$u_{A \cup B}(x) = \vee(u_A(x), u_B(x)) \forall x \in X \quad (42)$$

Where  $\vee$  denotes maximum

5. The intersection  $A$  and  $B$  ( $A \cap B$ ).

$$u_{A \cap B}(x) = \wedge(u_A(x), u_B(x)) \forall x \in X \quad (43)$$

Where  $\wedge$  denotes minimum.

Fuzzy sets obey all the properties of classical sets, excepting the excluded middle laws, i.e., the union and intersection of a fuzzy set and its complement are not equal to the universe and null set respectively. A fuzzy system is a repository of the fuzzy expert knowledge that can reason data in vague terms instead of precise Boolean logic. The expert knowledge is a collection of fuzzy membership functions and a set of fuzzy rules, known as the rule-base, represented as a set of rules. In the real world, knowledge is often represented as a set of “IF premise (antecedent), THEN conclusion (consequent)” type rules. Fuzzy inferencing is performed based on the fuzzy representation of the antecedents and consequents. The basic structure of a fuzzy system is given below as Fig. 3 [51]:

There are 4 main parts in a fuzzy logic system, a fuzzifier, a fuzzy inference engine, a fuzzy/knowledge rule base and a defuzzifier. The fuzzifier maps a real crisp input to a fuzzy set with membership function; the knowledge base contained the knowledge about the application as well as attendant control goals; inference engine emulates human being's decision-making process; Defuzzifier converts fuzzy control values into crisp quantities.

Though fuzzy logic system was designed to take advantage of uncertain embodied within human being's knowledge. Uncertainty prevails along with employment of fuzzy logic system processing. The input from real world data is from crisp set, it would be mapped to a fuzzy set with membership function. A membership function is designed or chosen based on user's experience, intuition. The common forms of membership function are listed as follow:

Gaussian Membership function:

$$u_{A^i}(x) = \exp\left(-\frac{(C_i - x)^2}{2\sigma_i^2}\right) \quad (44)$$

Where  $c_i$  and  $\sigma_i$  are the center and width of the  $i$ th fuzzy set  $A^i$ .

Generalized Bell Membership function:

$$gbell(x; a, b, c) = \frac{1}{1 + \left|\frac{x-c}{b}\right|^{2b}} \quad (45)$$

where  $a$  is responsible for the bell's width,  $c$  is responsible for bell's center and  $b$  is responsible for bell's slopes.

Besides above 2 functions, there are some other common membership functions, such as sigmoid membership function, trapezoidal membership function, triangular membership function [52].

With transformed fuzzy sets, the fuzzy implication operator as a part of fuzzy modus ponens would composite them into a new composite rule of inference. The main fuzzy implication operators are given as follow:

Zadeh's compositional rule of inference (CRI):

Suppose there are  $n$  fuzzy rules in the rule base of a fuzzy rule base. For a given input  $A'$ , in order to obtain a meaningful inference result  $B'$ ,

$$B' = \bigcup_{w=1}^n B'_w \quad (46)$$

Where  $A_w$  and  $B_w$ ,  $w = 1, 2, \dots, n$ , are fuzzy sets defined in the universe of discourses  $V$  and  $W$ , respectively.  $B'$  is the inference result based on rule  $w$ , i.e.,  $B'_w = A' \circ R_w$ , where  $R_w = A_w \rightarrow B_w$  is the fuzzy implication relation for rule  $w$  and  $\circ$  is the composition within the context of CRI, for example, Sup-min composition, and  $\cup$  is a combination operator, i.e.,  $\cup \in \{S, T\}$ , to be more particular,  $\cup \in \{V, \wedge\}$ .

Considering the cartesian product, we denote elements as  $x, y$ ; the sets they belong are  $A, B$  which are subset of the universal sets  $U_1$  and  $U_2$  respectively. A fuzzy relation on  $A \times B$  denoted by  $R(x, y)$  is defined as the set  $R = \{(x, y), uR(x, y) | (x, y) \in A \times B, uR(x, y) \in [0, 1]\}$ . Where  $uR(x, y)$  is function in two variables called membership function.

Let  $R_1(x, y)$ ,  $(x, y \in A \times B)$  and  $R_2(y, z)$ ,  $(y, z \in B \times C)$  be the two relations. The max-min composition is then the fuzzy set:

$$R_1 \circ R_2 = \left\{ \left[ (x, y), \max_y \{ \min \{ uR_1(x, y), uR_2(y, z) \} \} \right] x \in A, y \in B, z \in C \right\} \quad (47)$$

The fuzzy rule base is given as follows:

IF X is  $A_1$ , THEN Y is  $B_1$

IF X is  $A_2$ , THEN Y is  $B_2$

...

IF X is  $A_n$ , THEN Y is  $B_n$

With the rules from rule base and implication results, the results would be aggregated to a final result. Then through the defuzzifier, the results of aggregation would be transformed to crisp value. There are many defuzzification methods, the common methods are given followingly.

Maximum degree of membership:

$$v_0 = \max_v(u_v), v \in V \quad (48)$$

Centroid method:

$$v_0 = \frac{\int v u_v(v) dv}{\int u_v(v) dv} \quad (49)$$

In general, fuzzy logic is widely used in the AEC industry, the application crossed the strategic definition, preparation and briefing, technical design, manufacturing and construction. Among all these, the preparation and briefing are the most one. The main reason to that, the core tasks within preparation and briefing involved more general, macro-level and qualitative assessment, which artlessly directly relate to experience and intuition of human being. To better convert that, fuzzy logic spontaneously performs finer. Also due to the fuzziness, the ambiguous knowledge of industry could be easily transformed into inference system. In the early stage of a project, the decisions normally involved complicated factors not only in terms of amount but also in terms of level. All different experiences from different disciplines make the decision-making process too complicated to reasonably be described.

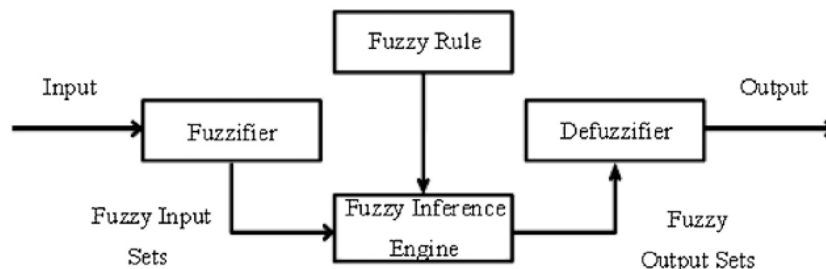


Fig. 3. Fuzzy logic system.

### 3.5. Genetic algorithm

Genetic algorithms are a particular class of evolutionary algorithm based on the mechanics of natural selection and natural genetics [19]. It is used to search an optimum solution from a set of possible solutions that is an array of decision-variable values. This set of possible solutions is called population. There are several populations in a GA run, and each of these populations is called a generation. Generally, at each new generation, better solutions (i.e., decision-variable values) that are closer to the optimum solution as compared to the previous generation are created. In the GA context, the set of possible solutions (array of decision-variable values) is defined as a chromosome, while each decision-variable value present in the chromosome is formed by gene. Population size is the number of chromosomes present in a population [53]. The GA process diagram is briefly presented as follow (Fig. 4):

To launch the process, certain parameters need to be predefined, e.g., type of chromosome representation, population size, selection process, types of crossover and mutation, crossover probabilities and mutation probabilities. The objective function would be evaluating each chromosome in population. Each chromosome should be assigned a fitness value, which is used to select the chromosomes from the current population as the selection. This algorithm is repeated sequentially until the stopping criterion is achieved. The stopping criterion of a GA is governed either by the number of generations or by the rate of change in the objective function value. Fitness values are expected to improve, indicating the creation of better individuals in new generations. Several generations are considered in the GA process until the user-defined termination criteria is reached [53].

It is not easy to directly analyze the uncertainty of genetic algorithm since the optimal solution remains unknown, the GA optimization is trying to search the relatively optimum one. The optimum solution cannot compare with the optimal one, ergo deviation, error, variance and so forth cannot be acquired through the solely optimization result. However, from the other end, the efficiency of GA optimization could be regarded as the uncertainty if it is safe to claim the higher efficiency could produce the better optimum solution that closer to the optimal one. On the basis of this assumption, this section would mainly focus on the mathematical explanation for the possible factors that caused the inefficiency.

The GA operators control the process of Gas, the GA operators including chromosome representation, population size, selection type, crossover and mutation. All these operators affect the efficiency of GA significantly.

### 3.6. Chromosome representation

Physical parameters in the search space constituting the phenotypes are encoded into genotypes. Chromosome representation or encoding is a process of representing the decision variables (phenotypes) in genetic

algorithms to machine readable. The genotype of an individual is the chromosome represents the possible solution. Coding in GA is defined by the type of gene expression, which may be expressed using binary, gray, integers, or real coding. In general, a chromosome (genotype) is presented as

$$(x_1, x_2, \dots, x_n) \text{ such that } x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n \quad (50)$$

Where,  $x_1, x_2, \dots, x_n$  are bits, integers, real numbers or a mixture of these, and  $X_1, X_2, \dots, X_n$  are the respective search spaces for  $x_1, x_2, \dots, x_n$ .

Theoretically, all the character set and coding scheme can be used for chromosome representation. Nowadays, the conventional operator for chromosomes representation is binary coding. In such operator, phenotypes in the parameter set are encoded as a binary string, and it is concatenated to form a chromosome. The length of the binary substring for a variable depends on the size of search space and the number of decimal places required for accuracy of the decoded variable values. If each decision variable is given a string of length  $L$ , and there are  $n$  such variables, then the chromosome will have a total string length of  $nL$  [54]. The search space is divided into  $2^L$  intervals, each having a width equal to  $(x_{i, \max} - x_{i, \min}) / 2^L$  for a binary string of length  $L$ , where  $x_{i, \max}$  is the upper bound of the decision variable, and  $x_{i, \min}$  is the lower bound of the decision variable:

$$d = (x_{i, \max} - x_{i, \min}) / 2^L \text{ defines the solution accuracy} \quad (51)$$

the binary numbers have a base of 2 and use only two characters, 0 and 1. A binary string, therefore, is decoded using:

$$N = a_n 2^n + a_{n-1} 2^{n-1} + \dots + a_1 2^1 + a_0 2^0 \quad (52)$$

Where  $a_i$  is either 0 or 1 (ith bit in the string),  $2^n$  represents the power of 2 of digit  $a_i$ ,  $n$  is the number of bits in binary-coded decision variable,  $N$  is the decoded integer value of the binary string. And the corresponding actual value of the variables is obtained using:

$$x_i = x_{i, \min} + \frac{x_{i, \max} - x_{i, \min}}{2^L - 1} N \quad (53)$$

Population size is another operator matters to the efficiency, it is the number of chromosomes in the population, the population size is application dependent and related to string length. Selection is also another operator namely reproduction operator. It is the survival of the fittest within the GA. It gives a higher priority or preference to better individuals for generating the next population. All chromosomes in the population can undergo the selection process using a selection method. This percentage (generation gap) is defined by the user as an input in genetic algorithms. The most common used selection method is proportional selection method. The probability of selecting a chromosome for reproduction can be expressed as:

$$P = \frac{f_i}{\sum_{i=1}^N f_i} \quad (54)$$

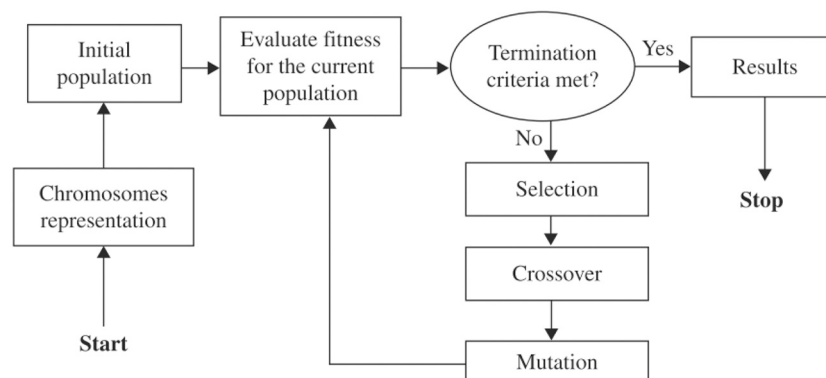


Fig. 4. Genetic algorithm process.



Where  $f_i$  is the fitness value of the  $i$ th chromosome in the current population of size  $N$ , and  $\sum_{i=1}^N f_i$  is the total fitness, which is the sum of fitness values of all chromosomes in the current population.

The crossover operator is used to create new chromosomes for the next generation by combining randomly two selected chromosomes from the current generation. A higher crossover rate encourages better mixing of the chromosomes. For the Real Coding, the BLX- $\alpha$ crossover performed the best in general, it is set as:

Two offspring,  $y_1 = (y_1^1, y_2^1, \dots, y_n^1)$  and  $y_2 = (y_1^2, y_2^2, \dots, y_n^2)$  are generated. Where,  $y_i^k$  is a randomly, uniformly chosen number from the interval  $[x_{min,i} - I\alpha, x_{max,i} + I\alpha]$  and  $x_{min,i}, x_{max,i}$ , and  $I$  are defined as shown here:

$$X_{max} = \max\{x_i^1, x_i^2\} \tag{55}$$

$$X_{min} = \min\{x_i^1, x_i^2\} \tag{56}$$

$$I = X_{max} - X_{min} \tag{57}$$

The last operator is mutation, mutation operator randomly altering a gene value to change the levels of chromosome genes. It is used to prevent genetic algorithm from premature termination.

It can be seen that genetic algorithms are the most popular algorithms for spatial coordination related tasks. Further, unlike the neural networks which is used most times but limited to certain stages, genetic algorithms are employed in almost all stages of projects with a non-negligible frequency [55–64]. From another aspect, the genetic algorithms are the second most used algorithms after neural networks. The main reason is the Eq. (50). The representation of chromosome, it aims at coding set of parameters rather than the parameters. Hence, the search process would not require continuity of objective function, the tasks of AEC industry mostly are not only discrete cases but also produce complicated objective functions as the background theoretical knowledge, individuals' experience and intuitions matters much. Genetic algorithms could efficiently find the optimum solution for the complicated objective functions. And the core tasks such as the site planning, considering costs, time, and risks, it is a solution search process which is discontinuous, multimodal, multivariant and involve extensive noise, this type of problems fits the genetic algorithms. Also, in nature, in the AEC industry, many problems could be modelled as the scheduling problems (travel salesman problem), which are the genetic algorithms' "expertise". Especially the spatial coordination. The core tasks of the stage are typical scheduling problems. Apart from the these, genetic

algorithms could integrate with various fields or algorithms to solve the problems, the ease of use and robustness promote the genetic algorithms being active through life cycle.

#### 4. Analysis results

In view of the variation to the elaboration of the algorithms, appertaining the essential mathematical explanation, 5 according analysis results about reasons causing uncertainty for each algorithm are given followingly. Notably, the analysis is on the basis of literatures those mainly about algorithms development in Section 3, and the additional literatures regarding the specific AI applications in AEC sector are reviewed on the purpose of proving the found points as verification to better align with theoretical analysis perceptibly.

##### 4.1. For primary component analysis

Given the Eqs. (1)–(6), the endogenous reason for uncertainty of PCA algorithm application could be revealed and given in Fig. 5. 1a. from Eqs. (1) and (2), it can be found that the very first uncertainty would be generated from the assumption of this approach, that is, the algorithm is built in the at least a Euclidean space with limited dimension and at most a vector space. In the space, the space regulated certain rules such as the vector addition, scalar multiplication, closure, associative laws and forth. Specific speaking, during the mathematical modelling process, the way of observed feature matters. The better interpretation of engineering features it is, easier to untangle non-linearity to linearity it can be. In some cases, the subjectivity is brought in by assuming the observation data is a linear combination of a certain base, which in many cases, without perfect definition of features /variables/base, it would be linear. Hence, the extremely likely wrongful linear assumption would lead to the generation of uncertainty; 1b. in the Eq. (6), the  $n$  part is deemed as noise, non-principal part. When analyzing data, the noise would be eliminated. In order words, the contribution of underlying structures/components would be removed from the phase space. Hence, the information loss exists. Though the application of PCA in nature is trying to minimize the information loss, it is the loss. The incomplete data sets as input naturally created the possibility of generating inaccurate output. 1c; with the Eq. (5), the subjectivity is imported to the application. Similarly, the non-principal parts of the equation would be removed from the data set, in the Eq. (5), it is  $\sum_{i=k+1}^n \lambda_i \alpha_i \alpha_i^T$ . However, the value of  $k$  depends on the manually input. It varies from the scenario and

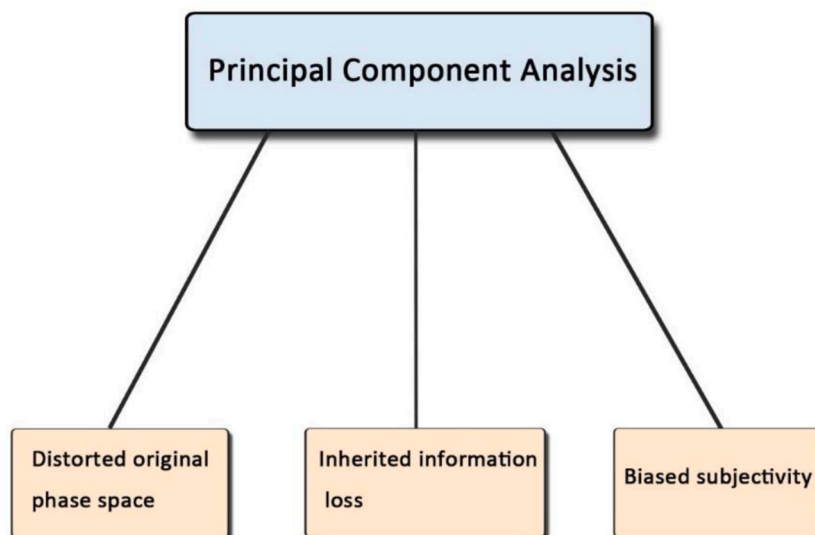


Fig. 5. Reasons causing uncertainty in PCA.

individuals experience. Taken the research of Platon et al., they used the principal component analysis to predict building's electricity consumption, in the research, the  $k$  is assigned with 4. Since when the  $k > 4$ , the  $\lambda_i \alpha_i \alpha'_i < 0.1$ . And the accumulated contribution of first part of Eq. (5)  $\sum_{i=1}^k \lambda_i \alpha_i \alpha'_i$  is 95% [28]. In this case, the 95% of contribution is subjective, it varies from the application. In some cases, the 95% is not enough, it would require higher contribution power. Also, importance factor 0.1 is also subjective. A good example of that is "Artificial neural networks based on principal component analysis for preliminary design of rubble mound breakwaters" from Balas et al., they employed PCA to optimize the hidden layers, the contribution was defined as 70%, and the threshold for importance factor is set as 0.024. Though the  $K$  is coincidentally equal to 4 which is same as Platon's [27]. But it could be easily noticed that the subjectivity prevailed. Hence, the uncertainty raised.

#### 4.2. For multilayer perceptron

Referring the Eqs. (7)–(19), The symbiotic uncertainty of multi-layer perceptron can be concluded as follow (Fig. 6): 2a. though extensive research as well as applications about information and communication technologies in AEC industry, the applicable data sets of buildings/infrastructures are too limited for the neural network training. With limited samples as input, overfitting could cause observably uncertainty variance while the deviation is acceptable [65]. The chance of that gets higher when the  $l$  from  $h\left(\sum_{i=1}^l w_{ip}\right)$  (Eq. (8)) and  $j$  from  $(d_j - o_j)f'(net_j)$  (Eq. (9)) are assigned a high value. Simply put, limited ICT application in AEC industry raised the sensitivity of uncertainty to structure of hidden layers. However, once the value of  $l$  and  $j$  is too low, underfitting problem would arise. In the case, the deviation is unacceptable while variance says otherwise [65]. Ergo, inadequate value of  $l$  and  $j$  lead to greater uncertainty; 2b. considering the activation functions  $f(\cdot)$ , it plays a vital role in the algorithms. However, the principle for choosing the proper predefined function from Eqs. (16)–(19) remained unclear. They all have pro et con. For most time, it is chosen by instinct and experience of data scientist. With the wrong choice, there are 2 negative consequences: i. the  $\Delta w_j \rightarrow 0$ , then the back-propagation process break, the layers after that to the input layer would have no feed. Besides the much longer learning process, the uncertainty could be eminently increased; ii. With wrong choice, such as he first predefined activation function– Sigmoid function (Eq. (16)), it might trigger  $\Delta w_j \rightarrow \infty$ , then the layers near the input layer could not be properly trained even with

large intervals, the error would decrease extremely gradually. Hence, the subjectivity would be imported to the process along with activation function selection. Followingly, the uncertainty is caused. To further, self-defined activation function may result in greater uncertainty compared with common predefined activation function once the ones made mistakes about defining ideal functions; 2c. the driven force to drift the prediction result away from accurate ones is the assumption of the neural network—the training dataset and test dataset must be independent and identically distributed. Thereupon the embedded generalization ability is limited. In a nutshell, there are missing correlated knowledge/patterns those entailed in the training datasets. As a preexisting and widely accepted fact states, a comprehensive decision-making system of AEC industry rely on more than intuition and experience, extensive deterministic theories (white box) are involved and to date act as a major role. The training datasets in the input layer followed the Eqs. (7)–(19) barely overcome the limitation of input layer's confidence interval. Ultimately, this incapability of neural networks causes the uncertainty during the application in AEC industry.

#### 4.3. For support vector machine

Based on Eqs. (28)–(36), the plausible reason for generating uncertainty in the application of SVM to AEC industry are summarized as follow: 3a. the most cases in practice are applying SVM to original features spaces those possess nonlinear separable features. Hence, the soft margin SVM would be applied directly or employ the kernel function first and then back to soft margin SVM. Among the most cases, there are more with the latter one [47]. Since, the soft margin SVM would be applied either way. The penalty parameter  $C$  in Eq. (28) matters, it is the first parameter with subjectivity propagating to the objective function. The value assignment varies from the cases, an example of that is when Mashford et al. used SVM to predict sewer condition grade, the  $C$  is 8000, they took it as an appropriate value [47]. While Paudel et al. applied SVM to predict energy consumption, the  $C$  was tuned from  $\{2^{-5}, 2^{-4}, \dots, 2^5\}$  [50]. Too large  $C$  would cause the overfitting problem, and too close to zero would cause the underfitting problem; 3b. besides the parameter  $C$ , the loss function could also bring in the subjectivity. Even the hinge function Eq. (30) is regarded as the most common form, there are still many other forms, such as ordinary least square loss function, 0–1 loss function, log loss function, etc. The selection strategy to the loss function is subjective during the application. With the poor selection strategy, either the robustness of learning would be undermined, or the accuracy would be impaired; 3c. the chosen kernel

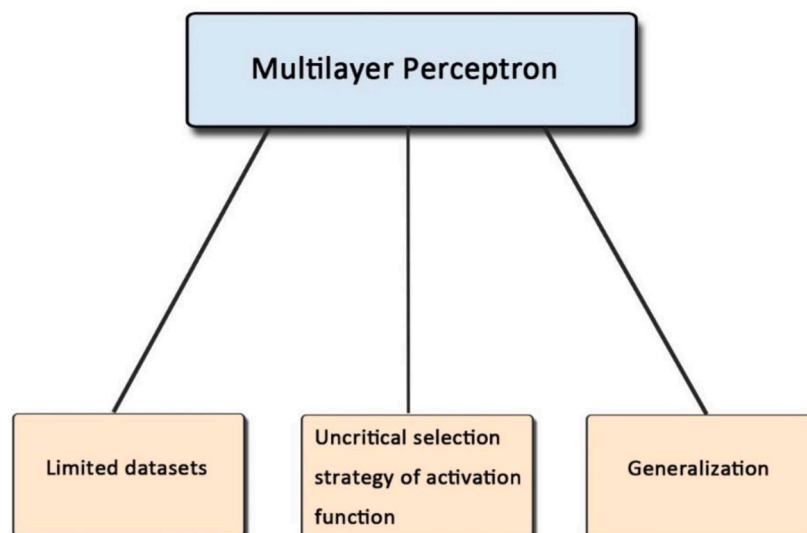


Fig. 6. Reasons causing uncertainty in MLP.

function affects the uncertainty as well. General principle advocates that the prior knowledge could help to decide, then the individual subjectivity rise. A comprehensive decision-making system of AEC industry relied on intuition and experience somehow. The prior knowledge would be hard to keep in the consistent level. Hence, selection from Eqs. (33)–(36) or the way to combine different kernel functions would cause uncertainty; 3d. with the kernel function, the original feature space would be transformed to a higher dimensional Hilbert space. It outstands the generalization problem. Due to the same reason as mentioned in the neural networks – multi-layer perceptrons, the work regarding regulating fundamental form of existing knowledge is far from done. It is even hard to solve the generalization problem with original feature space, not mentioned the higher dimensional feature space. Ultimately, this very incapability of support vector machine causes the uncertainty during the application in AEC industry. For better visualization, the tree diagram is given in Fig. 7 below.

#### 4.4. For fuzzy logic

Pertaining to Eqs. (44)–(49), the plausible reason for generating uncertainty in the application of fuzzy logic to AEC industry are summarized as follow (shown in Fig. 8): 4a. the way of defining membership those could help to transform crisp input to fuzzy sets. The membership value varies from the application, the experts within the field normally assign values to different inputs. It explains the involvement of expert system for fuzzy logic. However, the intuition and experience of experts denotes their own subjectivity to the system. With the same case, it is natural that different experts hold different thinking. In the different cases, it is more diverse. Hence, the uncertainty arises. Apart from the specific value of membership, the membership function like Eqs. (44) and (45) could be another source of uncertainty transformation. Such as the  $c_i$  and  $\sigma_i$  in Eq. (44), it varies from individuals. In Eq. (45), value of  $a$  and value of  $c$  are very likely different due to individual differences. The selection strategy of membership function determined the uncertain level of the later inference. Possibly, the memberships functions do not depend on the existing common membership functions, it could be self-defined as well. In which cases, subjectivity would be more obvious. It is possible to leave more space to generate uncertainty; 4b. another reason for the symbiotic uncertainty is the selection of implication methods. There are many applicable implication methods to use other than only Eqs. (46) and (47). But among all the composition implication operator, finding the most suitable one is vague, though many scholars tried to provide more principle to help selecting. Also, many scholars have been trying to define new operator for more approximate the inferring result. The situation remains the same as first reason, the uncertainty is propagated by users' subjective judgements; 4c. the third factor causes the uncertainty is the rule base. For the same case, number of rules and the aggregation rules could cause different aggregated results. The rule

base is built based on linguistic and semantics of human knowledge. Hence, the diversity of experts' knowledge base could easily cause diversity of rule base. Consequently, the uncertainty exists in the aggregation of result; 4d. defuzzifier process causes the uncertainty as well. As noticed, Eqs. (48) and (49) are the defuzzification methods. They clearly manage to transform the same aggregation of result to different control decision. Besides these two common methods, there are plenty of other defuzzification methods to choose from. Users' selection strategy matters than. Apart from that, users might develop their own methods to produce better decision further accurately.

#### 4.5. For genetic algorithm

Referring Eqs. (50)–(57), the potential driven forces of uncertainty (inefficiency) are listed as below (shown in Fig. 9): 5a. the selection of representation of chromosomes operator affect the efficiency of genetic algorithm. In the Eqs. (50) and (51), it represents the binary coding. While, there are another gene expression. Such as gray coding and real-value coding, the two different coding system improve the efficiency through regulating the distance between any two adjacent binary strings is 1 and removing the genotype-phenotype conversion, respectively. Besides, even within the same binary coding operator, different ranges and accuracies would be caused by different binary substring lengths for different decision variables; 5b. the population size does link to application and string length. The small population size can cause the genetic algorithms to converge prematurely to a suboptimal solution, and for longer chromosomes and challenging optimization problems, larger population sizes were needed to maintain diversity due to possibility of better exploration but the computing effort piling up; 5c. the selection operator assigns more probability to keep chromosomes with better fitness as the population of next generation. It might lead to overlooks of chromosomes with lower fitness value, thus less population diversity and premature convergence. The whole system would be sensitive to the method used in the selection process, taken Eq. (54) as an example, the method is called proportional selection method. There are many other selection methods, i.e., rank selection, tournament selection, elitist selection, generational selection, steady-state selection, hierarchy selection. The selection strategy of these methods as selection operator would propagate the users' subjectivity to the system. Followingly, uncertainty rises; 5d. the crossover operator is a further reason caused uncertainties. Some part of uncertainty source, however, is the derivatives representation of chromosomes and visualizes in the crossover operator. In general, crossover operators for binary coding and real coding are different. Within the same category, there are still many options. Eqs. (55)–(57) is the BLX- $\alpha$ crossover is a crossover operator for real coding, in other cases, geometrical crossover, arithmetic crossover, random crossover and such might be considered; 5e. the mutation operator entitles users the right of intervention by defining the probability. Large

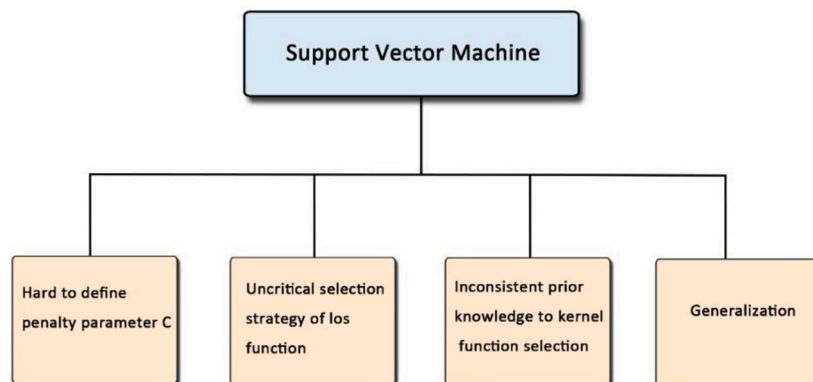


Fig. 7. Reasons causing uncertainty in SVM.

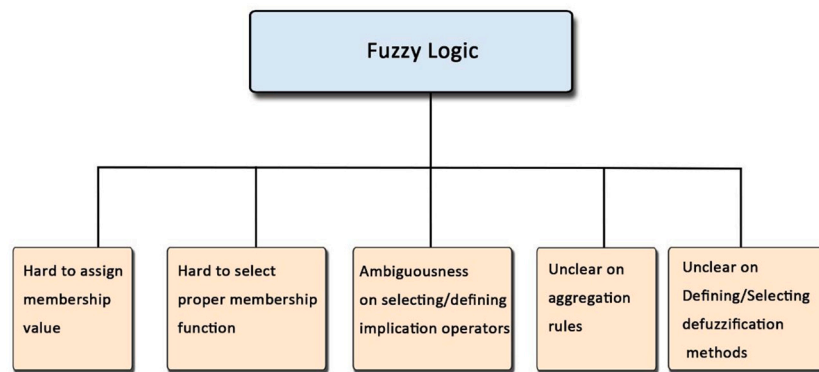


Fig. 8. Reasons causing uncertainty in fuzzy logic.

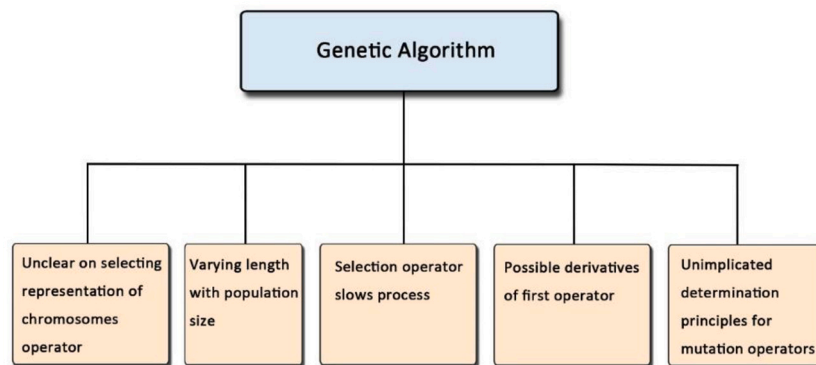


Fig. 9. Reasons causing uncertainty in genetic algorithm.

mutation rates would increase the probability of destroying good chromosomes but prevent premature convergence. The lower mutation rates would easily cause the premature convergence but increase of keeping good chromosomes. To the preset population size, string length, mutation methods would be another driven force to the uncertainty.

## 5. Uncertainty mitigation

The aims of the section are to bring up the mitigation measures to specifically control uncertainty associated with informed reasons those analyzed in the Section 3. The uncertainty mitigation measures does not only entail AI algorithms which would try to take advantage of other algorithms within AI to reduce uncertainty but also import the relevant state-of-art technologies as well as theories in the AEC industry such as Building Information Technology (BIM), Ontology and etc. to optimize the application of AI algorithms. On top of that, the analogy for integrating these technologies and theories with AI algorithms is briefly explained in this section align with associated mitigation measures. In addition, a summative framework is given in the end to conclude the uncertainty mitigation measures attaching to various reasons causing uncertainty named “Framework of bettering AI application in AEC industry” is given.

### 5.1. Explicit uncertainty mitigation for informed algorithm application

#### 5.1.1. For principal component analysis

To optimize the universality of PCA to AEC sector, several actions could be taken. Unlike other industries, the hierarchy of decision-making system for the AEC industry does not purely rely on the white box theory, but also the black box like engineers' experience and intuition. Decisions based on relatively subjective experience and intuition account for large ratio, such as the core task site appraisals under

strategic definition, project execution plan under preparation and briefing. These core tasks are not applicable for the PCA due to the lack of appropriate feature definition that could numericize the intuition as well as experience so that the mathematical model could be built. Without the mathematical model, the assumption of vector space could not be further away. To fit the assumption of Hotelling definition of PCA, the better specific framework for transform according literal engineers' experience and intuition into numericized feature should be developed. There are many existing approaches within expert system domain to promote the framework including multi criteria mapping-analysis, analytical hierarchy process and such. In a word, PCA collaborating with proper expert system techniques enables the universality to multiple working stages of lifecycle. Improving the accuracy of PCA application adumbrates reducing the information loss. Scilicet, the component  $n$  of Eq. (6) keeps minimal. Simultaneously, without affecting the efficiency of algorithm, assigning the value as higher as possible to  $k$  in  $\sum_{i=k+1}^n \lambda_i \alpha_i$  (Eq. (5)) is ideal, which means the higher contribution coefficient would be provided. Hence, the optimization of contribution coefficient under the certain constrain of running efficiency becomes a problem of multi-objective multivariate analysis. The heuristic algorithms within game play domain showed superiority on such problems. With genetic algorithm, Tabu search algorithm, Simulate Anneal Arithmetic and such considering the professionals input on the purpose of transforming intuition and experience to the search parameters, the optimal value for significance coefficient would be ensured without much efficiency tradeoff. It also explained many researchers tried to hybridize the game play with expert system.

#### 5.1.2. For multilayer perceptron

Regarding the inadequate applicable data sets caused limited applied range of multi-layer perceptrons for other stages, the building information modelling and geological information system related technolo-

gies provided wide range of benefits as part of ICT developments to tackle the issues. With more comprehensive and mature innovations of BIM and Geographic information system (GIS) based technologies, there would be higher possibility to digitalize more core tasks in the other stages, which means more applicable as well as organized data stream can be generated so that adequate training data sets would be available. In long term, BIM and GIS based technologies development could lower “the symbiotic uncertainty type 2a” with creating more applicable samples for the input layer aiming at underfitting/overfitting problems.

The optimization strategy for  $l$  from  $\left(\sum_{i=1}^l w_{ip_i}\right)$  (Eq. (8)) and  $j$  from  $(d_j - o_j)f'(net_j)$  (Eq. (9)) could be promoted under the assumption that the sample size is constant. And the strategy based on expert system could be promoted. The expert system would document the similar previous cases, descriptive knowledge, performance assessment criteria and such. Hence, once the similar projects launch, the expert system could efficiently offer the case-based reasoning to the proper values for  $l$  and  $j$  in the time manner. It, to certain degrees, explained the preliminary findings of bibliometric analysis—expert system and neural network fusion. Besides, since the  $l$  and  $j$  define the structure of the networks, it can be modelled as variables. On the other side, the time complexity of algorithm and the accuracy can be modelled as the variables. The others are considered as auxiliary parameters. Then the problem of optimizing neural networks ‘structures turn into a multi-objective multivariate analysis. The feasible area is constrained by overfitting as upper bounds and underfitting as lower bounds. Even though, within the constrained feasible area, the  $l$  and  $j$  are the positive real numbers, it could still accumulate tremendous combinations. Hence, heuristic algorithms are required for the matter.

As for “the symbiotic uncertainty type 2b”, to the selection strategy among common activation functions, expert system could help a lot for the same reason as “type a”. However, common activation functions would not suffice apparently. The self-defined activation functions are supposed to pay more attention to. Integrating symbolic mathematics as well as other philosophy into activation function development, such as compressed sensing based on restricted isometry property, Bayesian thinking, etc. Davoudi et al.’s research on structural load estimation is a good illustration of that, they refined activation function as Gaussian Naïve Bayes, the performance of algorithm was outstandingly improved [66]. And that explains existence of neural networks and symbolic mathematics fusion. Undoubtedly, there is huge potential from this perspective and long way to go. Against the last type of symbiotic uncertainty, equipping generalization ability to neural networks to lower the uncertainty caused by neural networks assumption (independently and identically distributed) is constitutive. To accomplish that, a series of actions need to be done for regulating the fundamentals of AEC industry knowledge. Taking a “tree” as a metaphor for this existing knowledge of AEC industry. Within the AEC, all the knowledge/concepts/facts are connected to each other directly or indirectly like a “tree structure”. The desire to generalize a neural network without the “tree” would be impossible to fulfill. Hence, embedding “tree” to neural network is critical. However, embedding “tree” to neural networks through transmitting “leaf” by “leaf” is utterly irrational. As a deduction, the hierarchy of knowledge itself (structure of tree and mechanisms of tree growing) feels necessity for exploration and clarification. These entails: i. Define the ontology for existing knowledge of AEC, plenty of academics in the Building Information Modelling field have been very active on this topic; ii. Build the freebase of existing AEC knowledge-base, it is similar to an expert base in the early stage, and it would evolve to a knowledge graph later on; iii. Build knowledge base for basic rules those abstract from topological structures, e.g., logics, algebra, grammar. Apart from the task i, the research regarding the other two tasks have not properly initiated yet. Though task i is far away from completion, it already raised enough attention. The scholars in the AEC industry could be suggested to work on latter 2 topics more, which on

the other hand would reform the neural networks.

### 5.1.3. For support vector machine

As stated in “reason 3a”, apparently, establishment of expert system could contribute to lower the subjectivity of determining  $C$ . With the same cases, documenting the  $C$  value and the algorithm performance, followers could choose from based on its circumstances. Once the kernel function is chosen as RBF, it naturally brings in another parameter  $\gamma$ . It defined the distribution of new feature space, larger  $\gamma$  derives less support vectors, smaller  $\gamma$  derives more support vectors. Fitly, the number of support vectors affects training and predicting speed. Hence, with the  $C$  and  $\gamma$ , the problem could be modelled as the multi-objective multivariate model again. Same as others, such problem could be solved efficiently with heuristic algorithms under game play domain. It explained why many SVM worked with game play to optimize the performance. For the “reason 3b”, as any other selection strategy situation, the expert system could always help. Also, when the choices are limited to the small amount, grid search would work as well. Besides, other well-known mathematical methods can be used to define a new loss function, such as the entropy theory, Bayesian model, etc. A typical case would be integrating posteriori probability with support vector machine, utilizing the Sigmoid function map to the outcome of standard SVM to a probabilistic value. Noticeably, the self-defined kernel function must follow the Mercer’s theorem, otherwise it is an inapplicable kernel function. About the “reason 3c” and “reason 3d”, due to the prior knowledge’s importance, there is no doubt for establishing expert system to maximize the use of prior knowledge. And the generalization problem can be solved by the same measures as MLP last guideline on one hand. Differently, the kernel function complicated the problem. Even within the original space, the incapability of generalization exists, it would be much harder to make sense the features of higher dimensional Hilbert space. To that, it would be more advisable to select the right features at the beginning, either it could be linear separable in the lower dimensional space or the linear separable features in the higher dimensional space does have actual meaning align with AEC practice.

### 5.1.4. For fuzzy logic

To the reason 4a—defining membership value or function. The first action can be taken which is also the most common action so far in the real world is to bring in comprehensive expert system. Notwithstanding, when it comes to the membership functions, limited capacity of expert system would not suffice. For determining a membership function, there are 2 unknown parts need to be specified. The structure of function, in which case, is the order of the function and the number of parameters. Normally, the heuristic algorithms can be used to solve such problems. Another unknown part is the coefficient of the functions, to which, the neural-network-based algorithms are famous for tackling this problem. This explains the reviewed outcome, the fuzzy logic appeared in the paper always showed up with neural networks techniques, game play techniques and expert system. For lowering the subjectivity involved in uncertainty 4b and 4c, the very first thing is to build a more comprehensive expert system storing the performance of previous fuzzy logic system, the rules they defined, the value they assigned, the implication factor they refined. Based on that, for the similar cases, the user could refer to them. Another solution would be the further research on game play techniques fusion. With emulation of aggregation rules, the more precise result could be tested against the data. Thus, users could propose better aggregating implications. Since the rules are regulated closely to the human being’s linguistic and semantics, the generalization of knowledge is brought up again. The effort required to make was mentioned in the 4.3, the ontology, the freebase and such need to be developed further. Against the uncertainty problem in the defuzzifier module, the process would be regarded as the inverse mapping of the first step-fuzzifier. Hence, the optimization guidelines applicable to fuzzifier apply here as well. The expert system, neural networks and game play techniques could all help to produce better defuzzification

methods.

5.1.5. For genetic algorithm

Within the genetic algorithms, reforming and refining are the problems throughout from “reason 5a” to “reason 5e”. For a genetic algorithm, these two could barely coexist. But for a complicated system, both reforming and refining would be required. To that, the new genetic strategies are needed. The new strategies are categorized as micro genetic strategy and macro genetic strategy. For micro genetic strategy, the genetic operators, parameters design, population size should be further developed to reach the final goals. For the parameters such as population size, the priori knowledge would contribute. In practice, cases are always modelled with constrains, it is constrained convex optimization. And for that, the constrains abstract from the case would affect the algorithms' efficiency directly. Hence, finding out the better operators and constrains for the specific cases are essential. Consequently, expert system those can cross validate as well as refer for the same/similar type of tasks are crucial. Moreover, for the operator's definition, there are extensive theories could help with, such as the differential equations, Markov Chain theory, sensitivity analysis, chaos theory and etc. This can be classified into either symbolic mathematics or expert systems. For the macro genetic strategy, it focuses on the reforming genetic algorithm process so that reformed macro features, or it employed the genetic algorithms as base, brought in other algorithms to form hybrid genetic algorithms, this is used to improve the capability for searching global optimal solution. This explains the popularity of genetic algorithms fusion with other algorithms. Despite the genetic strategy wise, there are some other perspectives to improve the efficiency of genetic algorithms.

For the genetic operators mentioned in the “reason a” to” reason e”, the operators' selection and definition need to be limited by the no free lunch theorem, same as k-nearest neighbors algorithm (KNN) optimization. On top of that, it already gains the global recognition that parallel genetic algorithm enables the reduction of computational load. To achieve that, the information exchange system that support well exchanging among the different population groups is of great importance. In the AEC industry, the BIM along with the GIS are designed specifically to tackle the issue. However, the awareness of the importance is relatively insufficient within industry. In practice, it is even rarer to detect the exitance of integration of BIM/GIS with genetic algorithms in order to improve the efficiency (reduce uncertainty).

5.2. Framework of improving AI application in AEC industry

This step is to generate framework of bettering AI application in AEC industry based on acknowledged results and findings. To clarify the framework clearer, a well-organized process named RIBA plan of work 2020 is imported firstly to assist representing tasks within AEC sector and then help to format a basic application hierarchy of it, for more information about RIBA plan of work 2020, check Appendix A. Secondly, the results, reasons causing uncertainty are classified into 4 categories, namely limited datasets, generalization, subjectivity in initial settings and subjectivity in algorithm structures, a figure summarized the analysis results is given below (Fig. 10). the 4 categories are colored differently, and the included detailed reasons caused uncertainty are colored consistently with summative reasons, consecutively the involved algorithms are listed surrounding the summative reasons with

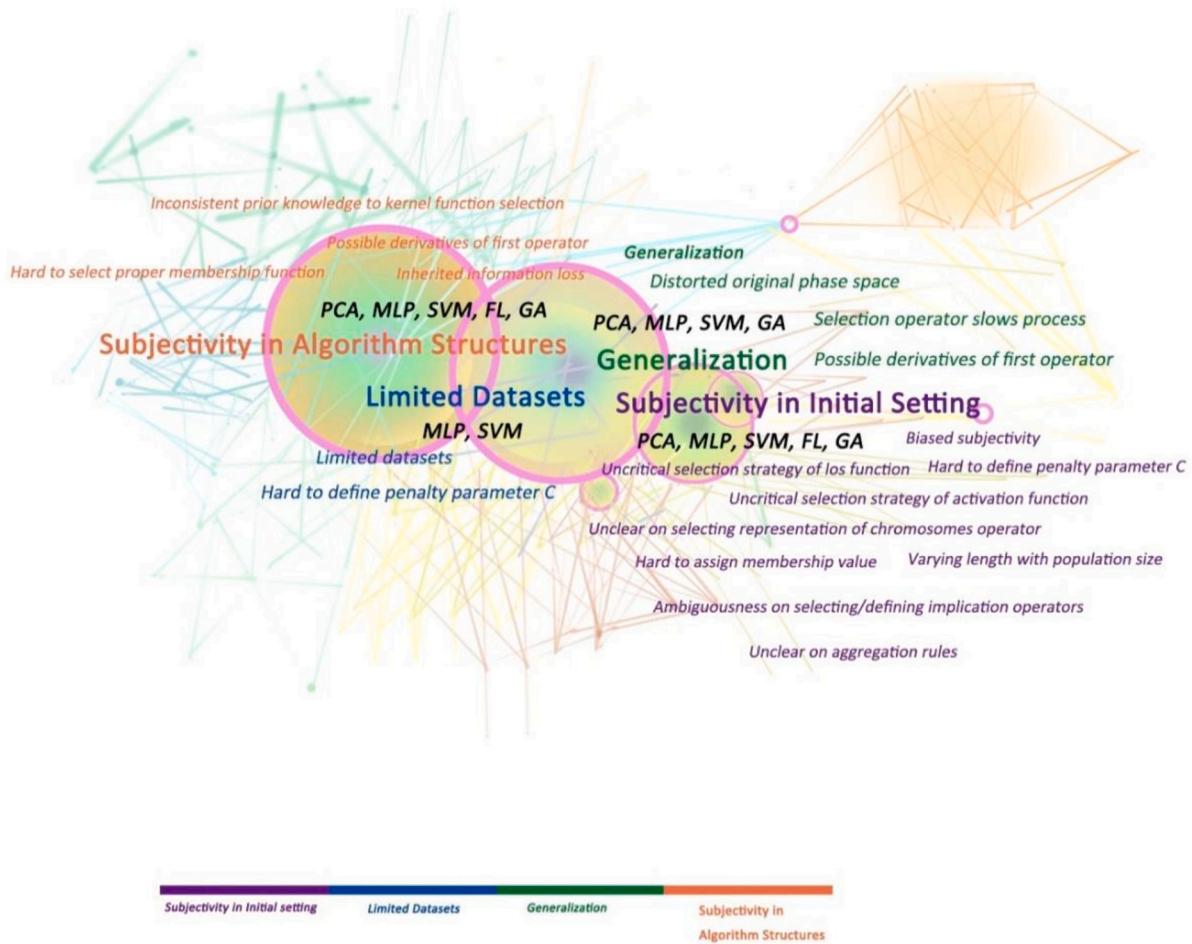

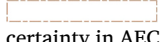









Fig. 10. Summarized analysis results.

black bold font. Taking the summative reason “Subjectivity in Algorithm Structures” an example, it is colored orange, it contains detailed reasons “Inconsistent prior knowledge to kernel function selection”, “Hard to select membership function”, “Inherited information loss”, “Derivative of first operator” those are marked with same color, and the produced uncertainty exists in the application of algorithm principal component analysis, multilayer perceptron, support vector machine, fuzzy logic and genetic algorithm. Last, the uncertainty mitigation measures are listed to sort out according problems caused by certain reasons. In the end of the section, to validate the uncertainty mitigation measures, a few remarkable research outputs from other related papers are discussed, though the topic they discussed may not remain in the scope of AEC, but the methodology as well as thinking stays topologically robust.

For better understanding the framework, notations appear in the figure is given. The  represents the basic domain of “AI application in AEC Industry”, the ICT Infrastructure, RIBA Core Tasks and AI algorithms are 3 essential elements of this domain. Followingly, the  stands for the basic domain of “Reasons Causing Uncertainty in AEC Industry”, the 4 summative reasons are the contents of it. Then, the  appears as the domain of “Optimization Guidelines for AI in AEC Industry”, it contains 2 types of subdomain, the external measures serve as part of ICT infrastructure included in the , and AI algorithms is another type embodied . The symbol  is used to express the virtual container contains datasets. And the symbol  plays the role of “gateway” for controlling how the process flows. Moreover, the dotted arrow  means “links to”. The solid arrow  shows the sequential order.

The framework concludes the AI application in AEC industry, reasons

causing uncertainty and uncertainty mitigation is discussed and visualized as follow (Fig. 11). Within the first section—AI application in AEC industry, blue dotted line frame contains the ICT infrastructure, which is used to digitalize the information. Along with the RIBA plan of work, the required datasets could be extracted and then put into the ‘datasets’ container. The red dotted line frame contains the AI algorithms, which could be applied aligning with RIBA plan of work, the proper algorithms would be selected for latter process the applicable datasets in the container. Based on that, the AI application manage to proceed. Afterwards, evaluated the performance, if good then end. If not, proceed to analyze the reasons. The second section includes the main reasons causing different summative reasons. Taking “generalization” type as an example, once the uncertainty is identified as the “generalization” caused. It could be coped with 3 guidelines, one internal and two externals. By internal, it means uncertainty mitigation measures is within the AI algorithms domain, that is saying one should optimize the algorithm with algorithms those are contained in the red dotted line frame since the algorithms are marked in the same way as prior section does, in this case, Symbolic Mathematics is the one. Two externals are marked with blue dotted line frame since it could be regarded as the ICT infrastructures related domain. In this case, BIM/ GIS related technologies and Ontology/Freebase/Knowledge Graph are the general mitigation measures. After the optimization, the application would be evaluated again to check the performance, then followed the same procedures as before till the evaluation result turn to be good performance.

These mitigation measures are derived from the theoretical analysis in a top-down way. Though it is referential, it would be more inspiring to confirm the findings with more specific and profound AI applications. A very good example of applying heuristic algorithm to mitigate the uncertainty is developing a least square support vector machine optimized

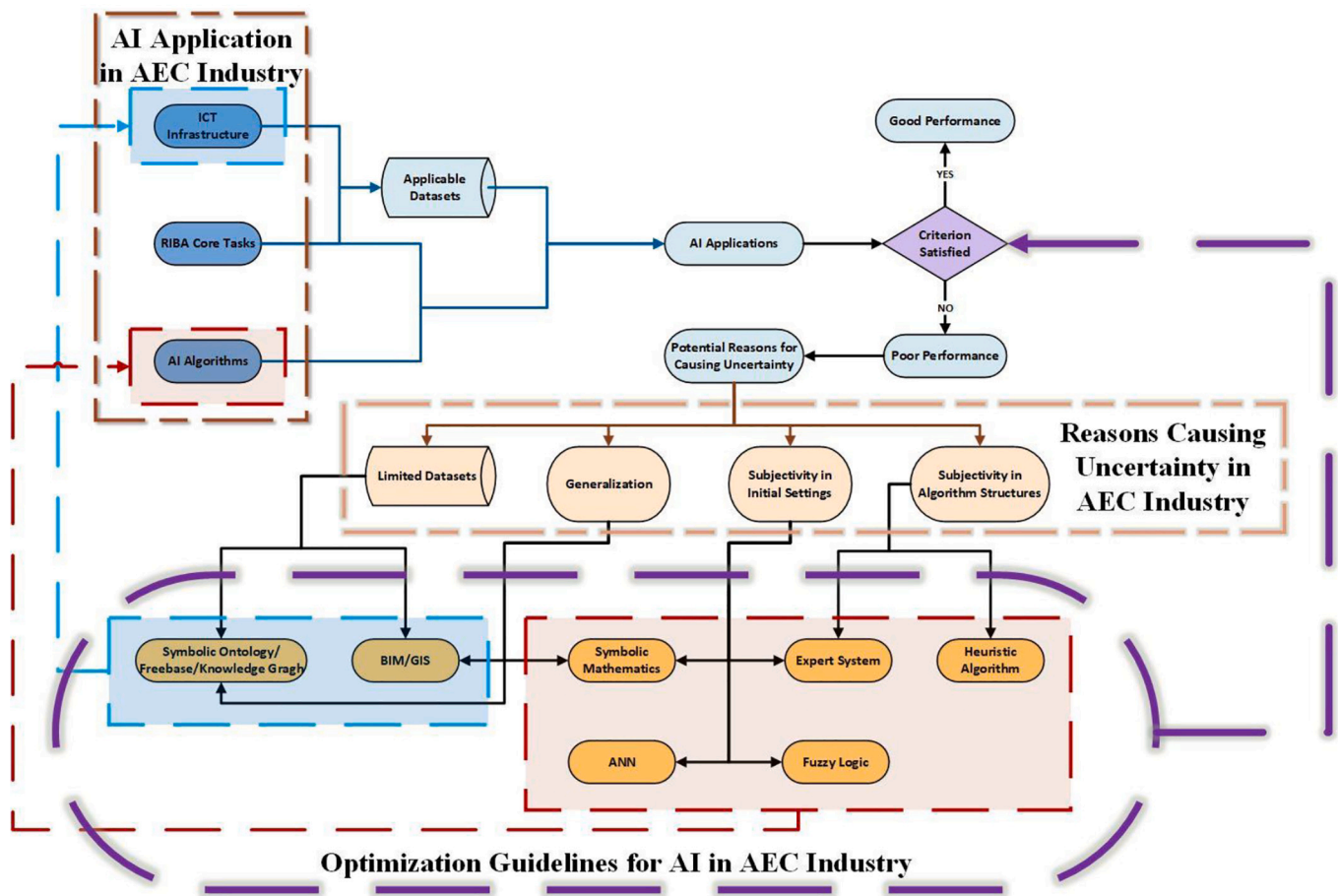


Fig. 11. Framework of Improving AI application in AEC industry.

by particle swarm optimization (PSO-LSSVM) approach to predict experimental VLE data of  $\text{CO}_2 + \text{H}_2$ ,  $\text{CO}_2 + \text{N}_2$  and  $\text{CO}_2 + \text{O}_2$  system [22]. Besides that, Najafi-Marghmaleki et al. address that the fuzzy logic hybridized with neural networks could also produce accurate and dependable prediction of experimental data. More specifically, while using particle swarm optimization to optimize radial basis function neural network which in nature is utilizing heuristic algorithms to optimize neural networks under the guidance of expert system (radial basis function is well-known mathematical function/theory hence categorized as expert system) to predict the viscosity of  $\text{CO}_2$ , the average absolute relative deviation is reduced to 0.351% [21]. From these 2 literatures, the solid applications and research work have been done to verify the assurance of findings from the bottom-up hierarchy. Though, the general categorized domains covered extensive concepts, such as multi-criteria-decision-based data mining algorithms are regarded as expert system optimizing artificial intelligence applications, and firefly optimization is categorized as a type of heuristic algorithm, which as most scholars in many other disciplines might regarded as a soft computing technique, the convincing research outputs of Farzin and Anaraki suggested that most derivatives of these mitigation measures could improve the efficiency and accuracy of AI applications [67–69]. As for the measure—exploiting symbolic mathematics to solve the generalization problems, it is relatively rare in AEC industry. Regardless, scholars from other disciplines have made their point clear, Chaos related thinking is a type of symbolic mathematics since there are many approaches in the chaos research involves defining new symbolic system to mark the features, make sense of features in order to seek the patterns underneath the chaotic data, such as applied symbolic dynamics, renormalization, spectrum analysis, i.e. Zhang et al. present a new activation function design based on the Hermite polynomials for better utilization of spatial representation, analyze the information transfer of deep neural networks, emphasizing the convergence problem caused by the mismatch between input and topological structure [70]; Tian et al. proposed a novel network called a batch-renormalization denoising network which outperforms state-of-the-art image-denoising methods [71]; Premkumar et al. use an improved gradient-based optimization algorithm with chaotic drifts to identify solar photovoltaic [72]. Knowledge graph application is another type of “symbolic mathematics” measure, applying knowledge graph to optimize machine learning are referential [73,74]. All these AI applications in other fields would imply that the mitigation measures proposed in the section are applicable.

## 6. Conclusion

The comprehensive analysis of endogenous reasons causing uncertainty confirmed that an industry like AEC, whose the main body is not only composed of theoretical knowledge but also the engineering experience and intuition of skilled engineers, would highly rely on the expert system and fuzzy logic techniques to ease the uncertainty of certain algorithms' application; and due to the complexity of AEC project, extensive stakeholders and disciplines get involved, the game playing techniques are often employed to optimize the application. Also, the uncertainty analysis provided new insights into the relationship among algorithms, different popular technologies of AEC industry such as BIM and GIS.

Though, the tremendous research outcomes are in line with the hypothesis of AI application and the results might suggest that AI mostly work well with AEC applications. However, based on the findings of uncertainty analysis results, there is still a gap between the real-world knowledge and algorithms original settings. For more detailed algorithm settings including all types of coefficients, parameters, fitness functions and so forth, the most common ways of determining them in current research are either referring the previous researches those failed to provide the explanatory notes or blind trial. Either way, there is no associated engineering implications of them. Hence, the applications of AI still lack enough practice and validation in different fields for the

whole lifecycle. The uncertainty analysis contributed to a clearer understanding of the reasons causing them as well as generic guidelines to optimize accordingly. The findings of this paper have brought convenience for applying AI techniques to subjects of interests and thus have involved many algorithms and been used for extensive tasks for many years. However, the application of AI techniques in AEC industry is still in early stage. Since neither literature analysis result nor uncertainty analysis offered the explicit the framework for the algorithm selection regarding various tasks through the lifecycle.

Notwithstanding, some limitations should be noted. The selected algorithms failed to cover all due to the page limit. Due to that, some findings might be left out. The mathematical analysis and guidelines for optimizing according algorithms are based on writer's civil engineering backgrounds and personal views, biased views exist. Moreover, regarding the uncertainty mitigation for optimizing the algorithms, they are analyzed and summarized theoretically. While study was rigorously conducted, lacking validation would limit the performance potentially though some of existing literatures have proved the guidelines due to the nature of the specific cases.

In summary, it is concluded that the AI techniques application in AEC industry are not mature to properly select right algorithm for the task with low uncertainty, high speed and comprehensive interpretation at the same time. Therefore, more studies need to be conducted to improve both algorithms wise and industry wise for enlarging the application of AI techniques in AEC industry. Based on the findings, analysis and discussion above, the key research directions that have to be addressed in the future are highlighted in following list to enlarge the applications of AI techniques in civil engineering:

- (1) The studies on BIM and GIS need to be promoted more to create applicable datasets to initiate the learning-based algorithms and expert systems. Especially studies regarding frameworks as well as guidelines to link the GIS/BIM techniques with AI require to gain more attention for exploring the potentials.
- (2) Studies regarding the ontology, freebase, knowledge graph of the AEC industry are expected to concentrate on so that the generalization ability of the algorithms could be improved. Hence application of AI techniques would stay in the lower uncertainty level.
- (3) In the findings and the uncertainty analysis, it shows the potential of employing heuristic algorithms to improve the efficiency/lower uncertainty of other algorithms. It is commonly used in many other industries, yet still lack of detail framework in the context of AEC industry. It is commendable that bringing more systematic studies regarding how and what heuristic algorithm could do for an algorithm in a specific case.
- (4) More research about comparative analysis among the algorithms deserved to be conducted on the purpose of evaluating the performances of different algorithms on the same task, though there are some already. They are limited to certain tasks which cannot cover the life-cycle stages.

This paper reviewed 5 popular AI algorithms from the point of applied mathematics then cross-validate the findings with publications of AI applications in the AEC industry. With reviewing classic algorithm developments, the findings vs cross validation is revealed in 3 levels, a. top-down analysis from applied mathematics explains the reason why AI is able to solve the AEC problems and verified it with well-known publications, b. analysis and review provides the reason causing uncertainty and classified them to help AEC professionals to identify the problems, then verified the findings with a few specific cases associated with literatures, c. through analysis, a few uncertainty mitigation measures those are ignored before are developed and attached theoretical proof, some examples are given, for those were proposed by others before, the theoretical reason why they did what they did are given via provided uncertainty analysis those could directly link to applied



mathematics. In virtue of all, this research contributes to the knowledge body of AI application in 3 aspects, i.e. Creating media (uncertainty analysis) to link theory with practice, so that, for those existing AI application developments, it could be understood better such as applying heuristic algorithms to optimize ANN, and for those haven't been widely recognized and fully comprehend, it point out what could help with easing uncertainty and why it could such as BIM/GIS could be used to optimize AI in AEC and so does symbolic mathematics; developing the generic guidelines for AEC industry researchers as well as stakeholders to optimize their existing algorithms' performance with algorithms and other applicable tools/knowledges from industry; and proposing key future research directions to be addressed in the field.

### 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.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.autcon.2021.103883>.

### References

- W.G. Armstrong, A.C. Lorch, A(eye): a review of current applications of artificial intelligence and machine learning in ophthalmology, *Int. Ophthalmol. Clin.* 60 (2020) 57–71, <https://doi.org/10.1097/HO.0000000000000298>.
- tutorialspoint, Learn Artificial Intelligence, Retrieved Jun 18 2020; Available from: [https://www.tutorialspoint.com/artificial\\_intelligence/index.htm](https://www.tutorialspoint.com/artificial_intelligence/index.htm), 2020.
- F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, *Stroke Vasc. Neurol.* 2 (2017) 230–243, <https://doi.org/10.1136/svn-2017-000101>.
- Architecture, Engineering & Construction Management Institute, What is Aec Sector?, Retrieved Oct 23 2021; Available from: <https://aecmi.org/what-is-aec-sector/>, 2021.
- R.E. Levitt, N.A. Kartam, Artificial intelligence techniques for generating construction project plans, *J. Constr. Eng. Manag.* 144 (3) (1988) 329–343, [https://doi.org/10.1061/\(ASCE\)0733-9364\(1988\)114:3\(329\)](https://doi.org/10.1061/(ASCE)0733-9364(1988)114:3(329)).
- Z. Irani, M.M. Kamal, Intelligent systems research in the construction industry, *Expert Syst. Appl.* 41 (4) (2014) 934–950, <https://doi.org/10.1016/j.eswa.2013.06.061>.
- V. Faghihi, A. Nejat, K. Reinschmidt, J.H. Kang, Automation in construction scheduling: a review of the literature, *Int. J. Adv. Manuf. Technol.* 81 (9) (2015) 1845–1856, <https://doi.org/10.1007/s00170-015-7339-0>.
- M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, S.O. Ajayi, O.O. Akinade, H.A. Owolabi, H.A. Alaka, M. Pasha, Big data in the construction industry: a review of present status, opportunities, and future trends, *Adv. Eng. Inform.* 30 (3) (2016) 500–521, <https://doi.org/10.1016/j.aei.2016.07.001>.
- D. Yu, H. Liao, Visualization and quantitative research on intuitionistic fuzzy studies, *J. Intell. Fuzzy Syst.* 30 (6) (2016) 3653–3663, <https://doi.org/10.3233/IFS-162111>.
- A.K. Shukla, M. Janmajaya, A. Abraham, P.K. Muhuri, Engineering applications of artificial intelligence: a bibliometric analysis of 30 years (1988–2018), *Eng. Appl. Artif. Intell.* 85 (2019) 517–532, <https://doi.org/10.1016/j.engappai.2019.06.010>.
- R. Sacks, M. Girolami, I. Brilakis, Building information modelling, artificial intelligence and construction tech, *Dev. Built Environ.* 4 (2020), <https://doi.org/10.1016/j.dibe.2020.100011>.
- R. Khallaf, M. Khallaf, Classification and analysis of deep learning applications in construction: a systematic literature review, *Autom. Constr.* 129 (2021), <https://doi.org/10.1016/j.autcon.2021.103760>.
- P. Geyer, C. Koch, P. Pauwels, Fusing data, engineering knowledge and artificial intelligence for the built environment, *Adv. Eng. Inform.* 48 (2021), <https://doi.org/10.1016/j.aei.2020.101242>.
- A. Darko, A.P.C. Chan, M.A. Adabre, D.J. Edwards, M.R. Hosseini, E.E. Ameyaw, Artificial intelligence in the AEC industry: scientometric analysis and visualization of research activities, *Autom. Constr.* 112 (2020), <https://doi.org/10.1016/j.autcon.2020.103081>.
- H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* 24 (6) (1936) 417–441, <https://doi.org/10.1037/h0071325>.
- David E. Rumelhart, James L. McClelland, A general framework for parallel distributed processing, in: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press, 1987, ISBN 9780262291408, pp. 45–76.
- C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297, <https://doi.org/10.1007/BF00994018>.
- A.K. Deb, Introduction to soft computing techniques: artificial neural networks, fuzzy logic and genetic algorithms, in: A. Majumdar (Ed.), *Soft Computing in Textile Engineering*, Woodhead Publishing, 2011, ISBN 9781845696634, pp. 3–24.
- D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Long Publishing Co., Inc., 1989, ISBN 9780201157673.
- D. Merayo, A. Rodríguez-Prieto, A.M. Camacho, Comparative analysis of artificial intelligence techniques for material selection applied to manufacturing in Industry 4.0, *Procedia Manuf.* 41 (2019) 42–49, <https://doi.org/10.1016/j.promfg.2019.07.027>.
- S. Abdolbaghi, A. Barati-Harooni, A. Najafi-Marghmaleki, Improving the prediction ability of reference correlation for viscosity of carbon dioxide, *J. CO2 Utilization* 31 (2019) 106–114, <https://doi.org/10.1016/j.jcou.2019.02.022>.
- A. Najafi-Marghmaleki, A. Barati-Harooni, A.H. Mohammadi, Impact of gas impurities on CO2 mole fraction: application in carbon capture and storage (CCS) processes, *Int. J. Greenh. Gas Control* 57 (2017) 173–184, <https://doi.org/10.1016/j.ijggc.2016.12.008>.
- K. Pearson, On lines and planes of closest fit to systems of points in space, *London. Dublin Philos. Mag. J. Sci.* 2 (11) (1901) 559–572, <https://doi.org/10.1080/14786440109462720>.
- I.T. Jolliffe, *Principal Component Analysis*, Springer Science & Business Media, 2002, ISBN 0387954422.
- R. Linsker, Self-organization in a perceptual network, *Computer* 21 (3) (1988) 105–117, <https://doi.org/10.1109/2.36>.
- M. Hoffman, D. Blei, F. Bach, Online learning for latent dirichlet allocation, *NIPS'10: Proc. 23rd Int. Conf. Neural Inf. Process. Syst.* 1 (2010) 856–864, <https://doi.org/10.5555/2997189.2997285>.
- C.E. Balas, M.L. Koç, R. Tür, Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters, *Appl. Ocean Res.* 32 (4) (2010) 25–433, <https://doi.org/10.1016/j.apor.2010.09.005>.
- R. Platon, V.R. Dehkordi, J. Martel, Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis, *Energy Build.* 92 (2015) 10–18, <https://doi.org/10.1016/j.enbuild.2015.01.047>.
- J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.* 61 (2015) 85–117, <https://doi.org/10.1016/j.neunet.2014.09.003>.
- A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Netw.* 13 (4) (2020) 411–430, [https://doi.org/10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5).
- W.S. Alaloui, M.S. Liew, N.A.W. Zawawi, B.S. Mohammed, M. Adamu, An artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors, *Cogent Eng.* 5 (2018) 1, <https://doi.org/10.1080/23311916.2018.1507657>.
- K.P. Amber, R. Ahmad, M.W. Aslam, A. Kousar, M. Usman, M.S. Khanm, Intelligent techniques for forecasting electricity consumption of buildings, *Energy* 157 (2018) 886–893, <https://doi.org/10.1016/j.energy.2018.05.155>.
- J.S. Chou, D.S. Tran, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders, *Energy* 165 (2018) 709–726, <https://doi.org/10.1016/j.energy.2018.09.144>.
- J. Ferreira-Cabello, E. Fraile-Garcia, E.M.P. Ascacibar, F.J. Martinez-de-Pison, Metamodel-based design optimization of structural one-way slabs based on deep learning neural networks to reduce environmental impact, *Eng. Struct.* 155 (2018) 91–101, <https://doi.org/10.1016/j.engstruct.2017.11.005>.
- S. Mangalathu, J.S. Jeon, Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques, *Eng. Struct.* 160 (2018) 85–94, <https://doi.org/10.1016/j.engstruct.2018.01.008>.
- S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, *Energy Build.* 158 (2018) 1533–1543, <https://doi.org/10.1016/j.enbuild.2017.11.039>.
- Y. Zhang, H.V. Burton, H. Sun, M. Shokrabadi, A machine learning framework for assessing post-earthquake structural safety, *Struct. Saf.* 72 (2018) 1–16, <https://doi.org/10.1016/j.strusafe.2017.12.001>.
- M.Y. Cheng, D. Prayogo, Y.W. Wu, Prediction of permanent deformation in asphalt pavements using a novel symbiotic organisms search–least squares support vector regression, *Neural Comput. Applic.* 31 (10) (2018) 6261–6273, <https://doi.org/10.1007/s00521-018-3426-0>.
- B. D'Amico, R.J. Myers, J. Sykes, E. Voss, B. Cousins-Jenvey, W. Fawcett, S. Richardson, A. Kermani, F. Pomponi, Machine learning for sustainable structures: a call for data, *Structures* 19 (2019) 1–4, <https://doi.org/10.1016/j.istruc.2018.11.013>.
- M. Ghamari, M. Shoostari, Application of Performance-Based Plastic Design (PBPD) method for 3D steel structures, *Eng. Struct.* 199 (2019), <https://doi.org/10.1016/j.engstruct.2019.109649>.
- M. Kurpínska, L. Kulak, Predicting performance of lightweight concrete with granulated expanded glass and ash aggregate by means of using artificial neural networks, *Materials* 12 (2020) 12, <https://doi.org/10.3390/ma12122002>.
- B. Zhong, X. Pan, P.E.D. Love, L. Ding, W. Fang, Deep learning and network analysis: classifying and visualizing accident narratives in construction, *Autom. Constr.* 113 (2020), <https://doi.org/10.1016/j.autcon.2020.103089>.
- M.-Y. Cheng, H.S. Peng, Y.W. Wu, T.L. Chen, Estimate at completion for construction projects using evolutionary support vector machine inference model, *Autom. Constr.* 19 (5) (2010) 619–629, <https://doi.org/10.1016/j.autcon.2010.02.008>.
- M.Y. Cheng, N.D. Hoang, A.F.V. Roy, Y.W. Wu, A novel time-dependent evolutionary fuzzy SVM inference model for estimating construction project at

- completion, *Eng. Appl. Artif. Intell.* 25 (4) (2012) 744–752, <https://doi.org/10.1016/j.engappai.2011.09.022>.
- [45] M.Y. Cheng, N.D. Hoang, Y.W. Wu, Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: a Taiwan case study, *Autom. Constr.* 35 (2013) 306–313, <https://doi.org/10.1016/j.autcon.2013.05.018>.
- [46] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks, *Energy Convers. Manag.* 50 (1) (2009) 90–96, <https://doi.org/10.1016/j.enconman.2008.08.033>.
- [47] J. Mashford, D. Marlow, D. Tran, R. May, Prediction of sewer condition grade using support vector machines, *J. Comput. Civ. Eng.* 25 (4) (2011) 283–290, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000089](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000089).
- [48] Y.-R. Wang, C.-Y. Yu, H.-H. Chan, Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models, *Int. J. Proj. Manag.* 30 (4) (2012) 470–478, <https://doi.org/10.1016/j.ijproman.2011.09.002>.
- [49] Q. Zhou, H. Zhou, Q. Zhou, F. Yang, L. Luo, T. Li, Structural damage detection based on posteriori probability support vector machine and Dempster-Shafer evidence theory, *Appl. Soft Comput. J.* 36 (2015) 368–374, <https://doi.org/10.1016/j.asoc.2015.06.057>.
- [50] S. Paudel, M. Elmitri, S. Couturier, P. Nguyen, R. Kamphuis, B. Lacarriere, O. L. Corre, A relevant data selection method for energy consumption prediction of low energy building based on support vector machine, *Energy Build.* 138 (2017) 240–256, <https://doi.org/10.1016/j.enbuild.2016.11.009>.
- [51] S. Phomphon, C. So-In, D. Niyato, A hybrid model using fuzzy logic and an extreme learning machine with vector particle swarm optimization for wireless sensor network localization, *Appl. Soft Comput.* 65 (2018) 101–120, <https://doi.org/10.1016/j.asoc.2018.01.004>.
- [52] L.A. Zadeh, Fuzzy sets, *Inf. Control.* 8 (1965) 338–353, [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [53] D. Rani, S.K. Jain, D.K. Srivastava, M. Perumal, Genetic algorithms and their applications to water resources systems, in: X.-S. Yang, S. Gandomi, S. Talatahari, A.H. Alavi (Eds.), *Metaheuristics in Water, Geotechnical and Transport Engineering*, Elsevier, 2013, ISBN 9780123983176, pp. 43–78.
- [54] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1999, ISBN 9783662028308.
- [55] M.-Y. Cheng, Y.-W. Wu, Evolutionary support vector machine inference system for construction management, *Autom. Constr.* 18 (5) (2019) 597–604, <https://doi.org/10.1016/j.autcon.2008.12.002>.
- [56] S. Christodoulou, Construction imitating ants: resource-unconstrained scheduling with artificial ants, *Autom. Constr.* 18 (3) (2009) 285–293, <https://doi.org/10.1016/j.autcon.2008.09.005>.
- [57] A. Filippopolitis, E. Gelenbe, A distributed decision support system for building evacuation, in: *Proceedings of 2009 2nd Conference on Human System Interactions*, 2009, pp. 323–330, <https://doi.org/10.1109/HSI.2009.5091000>.
- [58] M.-Y. Cheng, H.-C. Tsai, E. Sudjono, Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry, *Expert Syst. Appl.* 37 (6) (2010) 4224–4231, <https://doi.org/10.1016/j.eswa.2009.11.080>.
- [59] X. Ning, K.C. Lam, Cost-safety trade-off in unequal-area construction site layout planning, *Autom. Constr.* 32 (2013) 96–103, <https://doi.org/10.1016/j.autcon.2013.01.011>.
- [60] M. Castelli, T. Leonardo, L. Vanneschi, A. Popovic, Prediction of energy performance of residential buildings: a genetic programming approach, *Energy Build.* 102 (2015) 67–74, <https://doi.org/10.1016/j.enbuild.2015.05.013>.
- [61] C. Sivapragasam, K. Kannabiran, G. Karthik, S. Raja, Assessing suitability of GP modeling for groundwater level, *Aquat. Procedia* 4 (2015) 693–699, <https://doi.org/10.1016/j.aqpro.2015.02.089>.
- [62] N. Izadnia, K. Eshghi, A robust mathematical model and ACO solution for multi-floor discrete layout problem with uncertain locations and demands, *Comput. Ind. Eng.* 96 (2016) 237–248, <https://doi.org/10.1016/j.cie.2016.02.026>.
- [63] H. Nguyen, H. Moayedi, W.A.W. Jusoh, A. Sharifi, Proposing a novel predictive technique using M5Rules-PSO model estimating cooling load in energy-efficient building system, *Eng. Comput.* 35 (2) (2020) 1–10, <https://doi.org/10.1007/s00366-019-00735-y>.
- [64] A. Kandiri, E.M. Golafshani, A. Behnood, Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm, *Constr. Build. Mater.* 248 (2020), <https://doi.org/10.1016/j.conbuildmat.2020.118676>.
- [65] B. Everitt, A. Skrondal, *The Cambridge Dictionary of Statistics*, Cambridge University Press, 2010, ISBN 0521766990.
- [66] R. Davoudi, G.R. Miller, J.N. Kutz, Structural load estimation using machine vision and surface crack patterns for shear-critical RC beams and slabs, *J. Comput. Civ. Eng.* 32 (2018) 4, [https://doi.org/10.1061/\(ASCE\)CP.19435487.0000766](https://doi.org/10.1061/(ASCE)CP.19435487.0000766).
- [67] M.V. Anaraki, S. Farzin, S.-Y. Mousavi, H. Karami, Uncertainty analysis of climate change impacts on flood frequency by using hybrid machine learning methods, *Water Resour. Manag.* 35 (1) (2019) 199–223, <https://doi.org/10.1007/s11269-020-02719-w>.
- [68] S. Farzin, F.N. Chianeh, M.V. Anaraki, F. Mahmoudian, Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID), *J. Clean. Prod.* 266 (2020), <https://doi.org/10.1016/j.jclepro.2020.122075>.
- [69] S. Farzin, M.V. Anaraki, Optimal construction of an open channel by considering different conditions and uncertainty: application of evolutionary methods, *Eng. Optim.* 53 (7) (2021) 1173–1191, <https://doi.org/10.1080/0305215X.2020.1775825>.
- [70] G. Zhang, G. Li, W. Zhang, The expressivity and training of deep neural networks: toward the edge of chaos? *Neurocomputing* 386 (2020) 8–17, [arXiv:1910.04970](https://arxiv.org/abs/1910.04970).
- [71] C. Tian, Y. Xu, W. Zuo, Image denoising using deep CNN with batch renormalization, *Neural Netw.* 121 (2020) 461–473, <https://doi.org/10.1016/j.neunet.2019.08.022>.
- [72] M. Premkumar, P. Jangir, C. Ramakrishnan, Nalinipriya, H.H. Alhelou, B.S. Kumar, Identification of solar photovoltaic model parameters using an improved gradient-based optimization algorithm with chaotic drifts, *IEEE Access* 9 (2021) 62347–62379, <https://doi.org/10.1109/ACCESS.2021.3073821>.
- [73] Z. Lei, Y. Sun, Y.A. Nanekaran, S. Yang, M.S. Islam, H. Lei, D. Zhang, A novel data-driven robust framework based on machine learning and knowledge graph for disease classification, *Futur. Gener. Comput. Syst.* 102 (2020) 534–548, <https://doi.org/10.1016/j.future.2019.08.030>.
- [74] Y. Shen, C. Guo, H. Li, J. Chen, Y. Guo, X. Qiu, Financial feature embedding with knowledge representation learning for financial statement fraud detection, *Procedia Comput. Sci.* 187 (2021) 420–425, <https://doi.org/10.1016/j.procs.2021.04.110>.