Integrating machine learning with QFD for selecting functional requirements in construction automation

Edgar C. TAMAYO[1], Yasir Imtiaz KHAN[2], Mohamed AL-HUSSEIN[3], Ahmed Jawad QURESHI[4]

[1] *PhD, Department of Civil and Environmental Engineering, University of Alberta*
[2] *PhD, Lecturer, Coventry University*
[3] *Professor, Department of Civil and Environmental Engineering, University of Alberta*
[4] *Assistant Professor, Department of Mechanical Engineering, University of Alberta*

ABSTRACT

An important aspect of the conceptual design is at the customer requirement definition stage, where an optimal number of functional requirements are specified with the application of quality function deployment. To facilitate a systematic specification of functional requirements, state-of-the-art unsupervised machine learning techniques will be introduced in the feature selection of functional requirements. However, the scarcity of references on unsupervised feature selection in the literature reflects the difficulty associated with this topic. At the customer requirement definition phase, three techniques will be proposed for selecting functional requirements, namely: (a) principal component analysis, (b) forward orthogonal search, and (c) Kohonen self-organizing map neural network. These machine learning feature selection techniques address the limitations of current approaches in systematically determining the minimum functional requirements from the mapping of customer requirements in quality function deployment. When applied to the conceptual design of the transportable automated wood wall framing machine that is under development at the University of Alberta, the proposed feature selection techniques have been observed to be: *(i)* fast, *(ii)* amenable to small quality function deployment dataset, and *(iii)* adequate in realizing design objectives. The results presented in this paper can be easily extended to online determination of customer requirements and functional requirements, project management, contract management, and marketing.

KEYWORDS

Unsupervised Machine Learning, Self-organizing Map, Modular Construction, Quality Function Deployment, Feature Selection

Introduction

Defining customer needs and determining the minimum functional requirements (FRs) constitute the first and crucial step in conceptual design [1]. Quality function deployment (QFD), a methodology that originated in Japan, is used by the product development team in this initial phase of conceptual design to identify customer needs and to evaluate the influence of a product or service in achieving these needs [2]. A QFD matrix, which is a mapping of FRs and customer requirements (CRs), and the optimal or suboptimal selection of FRs are the desired outcomes of the product development team's collaboration in this phase. Each entry to this QFD matrix consists of a weighting factor that signifies the strength of relationship between a FR and a CR.

Once a QFD matrix is built, decision-making tools are then used to determine the minimum number of FRs that will realize a suitable design. However, methods suggested in the literature such as analytical network program and fuzzy logic are difficult to use. The objective of this paper is to present unsupervised machine learning techniques as alternatives in overcoming the difficulties specified in the next section for current approaches.

Shin and Kim [3] have applied factor analysis (FA), an unsupervised classical statistical method, with QFD to reduce the dimension of the FRs mapped from the CRs in QFD. Dimension reduction pertains to reducing the order of a model by grouping features or attributes, whereas feature selection refers to excluding the least important features from the model. As subsequently discussed further below, integrating feature selection with QFD is proposed in this paper to reduce the design complexity that is associated with satisfying a large set of FRs [4]. To address the limitation of FA in feature selection, this paper will leverage unsupervised machine learning techniques in systematically determining the minimum FRs from the mapping of CRs in QFD.

Based on the notion that a product design team must not be constrained to using only one technique at the customer requirement definition phase of conceptual design, a comprehensive set of unsupervised machine learning methods will be described in detail. Respecting this notion, the paper will be structured as follows: section 2 provides a literature review that addresses the limitations of current state-of-the-art techniques through the introduction of unsupervised machine learning methods; section 3 discusses the motivation, and detailed description of the proposed machine techniques, and algorithms; section 4 describes the application of the algorithms to the QFD of an automated wood wall framing machine and the validation of the results, followed by the conclusion in section 5.

Literature Review

The current state-of-the-art rating of the importance of FRs is achieved using techniques such as analytical network process (ANP), FA, fuzzy logic (FL), and principal component analysis (PCA). Karsak et al. [5] have used ANP to rank the FRs and zero-one goal programming to determine the FRs to consider in designing the product. Mazurek and Kiszová [6] highlight two disadvantages of ANP: (1) obtaining the correct network structure is difficult, and (2) forming the supermatrix by pair-wise comparison of all criteria is difficult and unnatural.

To overcome the difficulties associated with the use of ANP, Mazurek and Kiszová [6] suggest the application of FL. This approach that utilizes ANP and fuzzy QFD has been used by Zaim et al. [7] in developing an improved product for systematically ranking FRs on the basis of CRs. Other applications of ANP and FL include: landfill site selection [8], shipyard site location [9], and importance ranking of the factors involved in a SWOT (strengths, weaknesses, opportunities and threats) analysis for the airline industry. However, the requirement prioritization through FL of these applications is prone to error due to its reliance on experts [10].

Shin and Kim [3] have introduced FA to restructure QFD. An interesting method that combines FA and ANP has been described by Zebardast [11] for determining social vulnerability to earthquake hazards. FA is a classical statistical technique that does not have the difficulties encountered in ANP or fuzzy logic, but Subbaiah et al. [2] give preference to PCA over FA due to the weakness of the latter in handling strongly correlated requirements. Weber et al. [12] have explored PCA in features prioritization of an airport, but, for feature selection, they have not used this state-of-the-art machine learning technique to establish the criteria for determining the optimum number of FRs.

PCA is an unsupervised machine learning technique that has been used so far for evaluating the importance of FRs to satisfy customer expectations. Although unsupervised machine learning techniques, such as forward orthogonal search [13] and self-organizing map [14], have been used for feature selection, they have yet to be introduced to the literature for the optimum selection of FRs in QFD. Thus, this paper contributes towards the application and evaluation of state-of-the-art machine learning approaches to a systematic customer requirement definition in the conceptual design of automated construction machines.

Unsupervised machine learning methods circumvent the aforementioned problems associated with ANP, FA, and FL. Moreover, these methods significantly simplify the

FR selection process by directly applying them to QFD, which is nothing but a dataset of FRs and CRs. Suitable choice of learning algorithms in this paper has been mainly reduced to PCA, forward orthogonal search, and self-organizing map neural network due to their reported applicability to feature selection. PCA has been used to evaluate machine defects [15]. Successful application of forward orthogonal search has been reported for datasets involving diagnostic breast cancer, cardiac arrythmia, ionosphere, and forest cover types [13]. Since its introduction by Professor Kohonen in 1982 [16], self-organizing map neural network has been utilized in such feature selection applications as the detection of arrythmias from electrocardiographic (ECG) signals [17] and the discovery of patterns in seismic wavefields [18].

Feature Selection Using Machine Learning Techniques

Motivation

Conceptual design is initiated by identifying CRs, a necessary step prior to any design methodology such as that of the integrated conceptual design described in a study by Tamayo et al. [19]. This is the most critical stage of conceptual design since correctly identifying the CRs defines the FRs that make up the design intent of the project. QFD is the methodology used by the product design team to evaluate the significance of FRs that achieve these CRs. Degrees of significance are assigned values such as 1 (weak significance), 3 (medium significance), and 9 (strong significance) [3]. An example of a QFD matrix is illustrated in Table 1, where m and n indicate the total number of CRs and FRs, respectively. This m×n QFD matrix will be denoted as the dataset matrix X consisting of n columns of features or FRs.

The complexity of the system increases with the increase in the number of FRs as expressed in terms of the probability of achieving the highest FRs, which diminish with larger n [4]:

$$P = 1/n! \qquad (1)$$

Equation (1) suggests the use of minimum number of FRs to increase the probability of finding the right DPs to satisfy n FRs at the axiomatic design (AD) stage and thus reduce the complexity of the design. Selecting a near optimum number of FRs is accomplished at the CR definition stage using the feature selection techniques discussed below.

Table 1. Typical QFD

	FR_1	FR_2	FR_3	FR_4	FR_5	.	FR_n
CR_1	9		1		9		
CR_2		3					
CR_3	1		9	3	3		
CR_4			3	3	1		
CR_5			3	3	1		
CR_6	9		1		9		
CR_7					9		
CR_8	3						9
⋮	...	...	...	...	...		...
CR_m			1	9	3		

Selecting the minimum number of FRs is accomplished by retaining the p number of FRs that adequately satisfy the CRs in the QFD matrix and exclude the least significant FRs. Approaches to determining p include [20]: (i) filter, and (ii) wrapper models. A simple method of determining p is by setting a threshold in the cumulative of the percent explained variances versus the retained number of FRs.

Figure 1 describes the general block diagram of an unsupervised learning process for feature selection, which takes a full QFD matrix in as the dataset $X \in \mathbb{R}^{m \times n}$ and outputs a reduced QFD matrix $\hat{X} \in \mathbb{R}^{m \times p}$. A description of the process block in Figure 1 is illustrated in detail for each of the proposed unsupervised machine learning algorithms in the following sections.
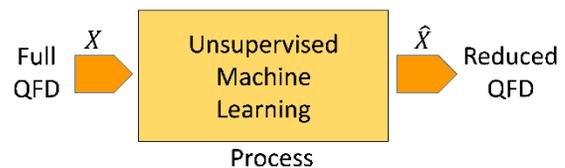

Figure 1. General input-output block diagram of unsupervised machine learning process

Principal Component Analysis

Figure 1 is redrawn in Figure 2 to represent the unsupervised feature selection that makes use of PCA. Notations commonly used in chemometrics will be adopted in the following derivations [21]. PCA is a model of the form:

$$X = \hat{X} + E$$
$$= TP^T + E \tag{2}$$

where $\hat{X}$ is the $m \times n$ reconstructed dataset that approximates X and that is expressed as the product of the scores matrix, $T \in \mathbb{R}^{m \times q}$. and the loadings matrix, $P \in \mathbb{R}^{\{q \times n\}}$. Scores matrix T projections of the original features onto the principal components. E is the reconstruction error introduced in approximating the dataset.
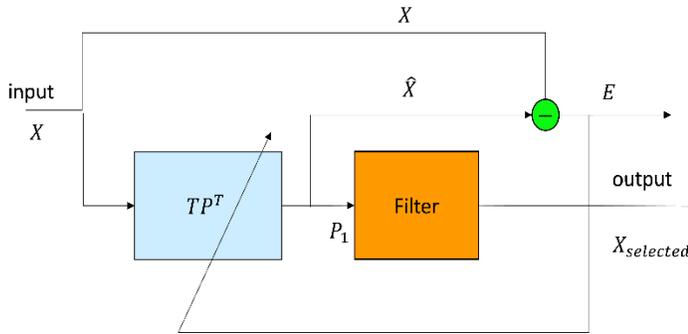

Figure 2. Feature selection by Principal Component Analysis

In multi-dimensional data, the solution to minimizing the reconstruction error is obtained from Equation (4) [22].

$$T = \underset{T \in \mathbb{R}^n}{arg\ min} \|X - TP^T\|^2 \tag{3}$$

$$= XP$$

Forming the singular decomposition of X with respect to a diagonal matrix Σ, and orthogonal matrices U and V such that $UU^T = I$ and $VV^T = I$:

$$X = U\Sigma V^T \tag{4}$$

Loadings matrix P is determined by minimizing the reconstruction error as in the study by Hastie et al. (2009):

$$P = \underset{P \in \mathbb{R}^n}{arg\ min} \|X - XPP^T\|^2 \tag{5}$$

Equation (5) indicates that $P = V$ since PP^T is the orthogonal projection of X onto the subspace spanned by the columns of V. Thus, the loadings, scores, and reconstructed dataset are derived through the singular value decomposition of X as follows:

$$P = V \tag{6}$$

$$T = U\Sigma \tag{7}$$

$$\hat{X} = XVV^T \tag{8}$$

The expected value of the reconstruction error or mean squares error (MSE) is calculated as the sum of the minor eigenvalues:

$$MSE(q) = \sum_{i=q+1}^{n} \lambda_i \tag{9}$$

The MSE for different values of q principal components and setting a threshold MSE is what determines the optimum number of principal components q^*.

PCA is normally used for feature extraction. However, Song et al. [24] have shown that PCA can also serve the purpose of feature selection. Since PCs are automatically ranked when singular value decomposition (SVD) is used, for feature selection, the q^*- dimensional P and T matrices are derived by determining the eigenvectors and eigenvalues from the covariance matrix:

$$C_X = \frac{1}{m-1} XX^T \tag{10}$$

$$C_X p_i = \lambda_i p_i \tag{11}$$

where p_i is the ith column eigenvector of the loadings matrix P and λ_i is the associated eigenvalue. Importance of each feature is obtained by the magnitude of each column eigenvector where the eigenvalue is maximum, i.e., each element of the first principal component, through Equation (12) [15]. Xu et al. [25] have observed that Equation (12) is valid when the maximum eigenvalue, λ_1, is significantly larger than the rest of the eigenvalues.

$$c_i = |p_i|\ for\ i = 1:n \tag{12}$$

A summary of the steps required to develop Algorithm 1 [24] for feature selection using principal component analysis is presented below.

Algorithm 1 Feature Selection using Principal Component Analysis

1: Dataset: $X \leftarrow \{x_i \mid x_i \in \mathbb{R}^m \wedge i \in [1, n]\}$;
2: Mean-center and scale dataset: $X \leftarrow$ normalized(X);
3: Perform SVD: $X \leftarrow U\Sigma V^T$, Eq. 4;
4: Evaluate loadings matrix: $P \leftarrow V$, Eq. 6;
5: Rank first column of P: $c_i \leftarrow |p_{1i}|$. Eq. 10;
    ▷ Ranking of c_i corresponds to the ranking of the features

---

Forward Orthogonal Search

Because of its simplicity and its suitability to highly correlated QFD dataset, PCA is the first feature selection method to use. Two other techniques will be introduced that conform to a specific structure of the QFD matrix. Forward orthogonal search (FOS) [13], depicted in Figure 3, is similar to PCA but it does not have the maximum eigenvalue restriction since its feature selection criterion is based on maximum squared-correlation coefficient and

average sum of error reduction ratio. Equation (13) describes the model that approximates the original QFD matrix.

$$X = \hat{X}\theta + E$$

$$= QR\theta + E \tag{13}$$

where the selected number of features is $q < n$; $\hat{X} \in \mathbb{R}^{m \times q}$ is orthogonally decomposed into orthogonal matrix $Q \in \mathbb{R}^{m \times q}$; upper triangular matrix $R \in \mathbb{R}^{q \times q}$; and parameter matrix $\tau \in \mathbb{R}^{m \times n}$. $\theta \in \mathbb{R}^q$ is the parameter vector.
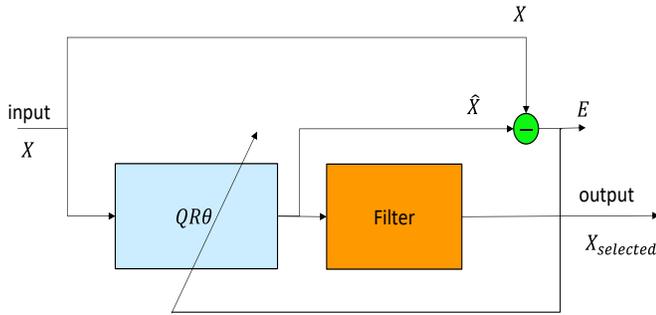

Figure 3. Forward orthogonal search unsupervised feature selection method

Algorithm 2 reflects the equations developed by Wei and Billings [13]. Thus, the function, $sc(a_i, b_j)$, that calculates the index of the maximum squared-correlation coefficients, $\gamma \in \mathbb{R}^n$, of two vectors a_i and b_j is determined as follows:

$$\gamma_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\left(a_i^T b_j\right)^2}{\left(a_i^T a_i\right)\left(b_j^T b_j\right)} \tag{14}$$

$$l = \underset{j \in [1,n]}{arg\ max}\{\gamma_j\} \tag{15}$$

Taking the column vectors of the normalized dataset, X, for both a_i and b_j in the sc function provides the first selected feature vector, x_{l_1}, in $\hat{X}$. $\tilde{X}$ represents the dataset that results from the removal of the feature vector, x_{l_1}, from X. In the subsequent steps, however, $\hat{X}$ and $\tilde{X}$ are updated through $sc(x_i, q_j)$, where q_j is obtained by applying the Gram-Schmidt orthogonalization function $ortho(\tilde{x}_j, q_k)$, or

$$q_j = \tilde{x}_j - \sum_{k=1}^{r-1} \frac{(\tilde{x}_j^T q_k)}{q_k^T q_k} q_k \tag{16}$$

The significance or importance of the selected feature is evaluated through the function $perf(x_j, q_k)$, such that:

$$SE = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{r} \frac{\left(x_j^T q_k\right)^2}{\left(x_i^T x_j\right)\left(q_k^T q_k\right)} \times 100 \tag{17}$$

A summary of the steps required to develop Algorithm 2 for feature selection using forward orthogonal search is presented below (see pseudo-code below).

Self-organizing map neural network

PCA and FOS are appropriate for linear feature selection; however, Kohonen self-organizing map (SOM) can be utilized for a QFD matrix exhibiting nonlinear relationships [26]. SOM has been used for clustering, data visualization, dimensionality reduction, and nonlinear data mapping, and its variants are too many to list [26]. Figure 4 illustrates the Kohonen two-dimensional SOM neural network that is used to model the QFD matrix at the CR definition stage. A strategy developed by Faro et al. [27] is adopted in this paper to automate the process of determining the number of classes with the use of a single layer SOM. A more detailed description of the general block diagram of Figure 1 is depicted in Figure 4 for SOM feature selection method that uses genetic algorithm.
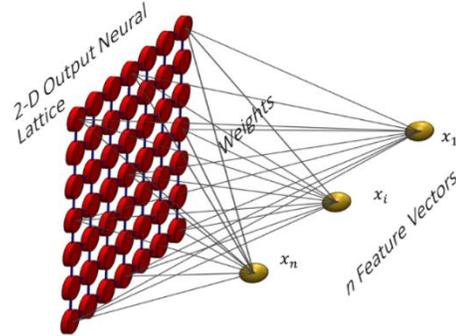

Figure 4. Kohonen Self-Organizing Map Neural Network

Similarity matrix, $S \in \mathbb{R}^{m \times m}$, is used to avoid the difficulties associated with raw data in clustering [28]. In a study by Faro et al. [27], similarity matrix is calculated in Equation (18) then normalized before using it for classification. This equation describes the function $sim(X)$ in the algorithm while Equation (19) describes the random weights initialization that is uniformly distributed between 0 and 1 in the function $initw$.

$$s_{ij} = \frac{1}{n} \sum_{k=1}^{n} min(x_{ik}, x_{jk}) \tag{18}$$

$$\omega_{j\kappa} = U(0,1), \kappa \in [1, nu] \tag{19}$$

$$d_\kappa = \|S_i - \omega_\kappa\| \tag{20}$$

$$lu = \underset{\kappa \in [1,nu]}{arg\ min}\{d_\kappa\} \tag{21}$$

Algorithm 2 Feature Selection using Forward Orthogonalized Search

1: Dataset: $X \leftarrow \{x_1, \cdots, x_i, \cdots, x_n\} \in \mathbb{R}^{m \times n}$;
2: Mean-center and scale dataset: $X \leftarrow normalized(X)$;
3: Initialization at $i = 1$:
4:     $\gamma \leftarrow sc(x_i, x_j)$ ▷ Calculate average squared coefficients
5:     $l_1 \leftarrow arg\,max(\gamma)$ ▷ Determine 1st index of maximum coefficient
6:     $\hat{X} \leftarrow x_{l_1}$ ▷ Approximate model dataset
7:     $\tilde{X} \leftarrow X^{-l_1}$ ▷ Dataset $\tilde{X}$ with l_1^{th} column of X removed
8:     $Q \leftarrow x_{l_1} \in \mathbb{R}^{m \times 1}$ ▷ Or $q_1 \leftarrow x_{l_1}$ for the 1st step
9:     $SE_1 \leftarrow 100 \times perf(x_j, q_1)$ ▷ Evaluate performance
10:     $Importance_1 \leftarrow (100 - SE_1)$
11: **for** $(i \in [2, n])$ **do**
12:     $q \leftarrow ortho(\tilde{x}, q)$ ▷ Gram-Shmidt Orthogonalization of
                                     ▷ $\tilde{x}$ to the previous $q \in \mathbb{R}^{r-1}$
13:     $\gamma \leftarrow sc(x, q)$
14:     $l \leftarrow arg\,max(\gamma)$
15:     $\hat{X} \leftarrow \tilde{x}_l$ ▷ Append l^{th} column of previous $\tilde{X}$ into $\hat{X}$
16:     $\tilde{X} \leftarrow \tilde{X}^{-l}$ ▷ Remove l^{th} column from previous $\tilde{X}$
17:     $Q \leftarrow q_j$ ▷ Append q_j. $Q \in \mathbb{R}^{m \times r}$
18:     $SE_i \leftarrow min(100 \times perf(x, q), 100)$
19:     $Importance_i \leftarrow (100 - SE_i)$
20: **end for**
21: Rank selected features in $\hat{X}$ according to Importance
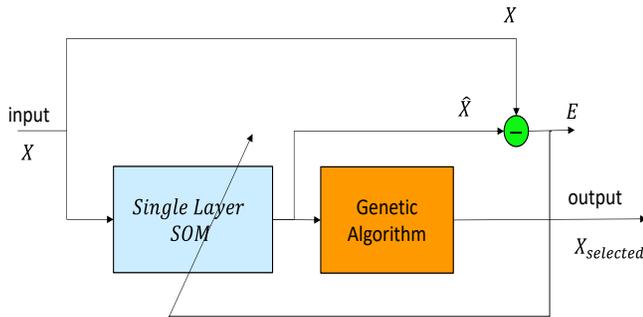

Figure 5. Single layer SOM unsupervised feature selection method

With the use of the above equations, a function for updating the weights does the following: *(i)* calculates and orders the Euclidean distance vector, $d \in \mathbb{R}^{nu}$, between neurons and row vector of the similarity matrix where the maximum number of neurons, *nu*, is 3; *(ii)* calculates the index of neuron with minimum distance; *(iii)* updates the weights, $\omega(t)$, at a user-configurable learning rate α; and *(iv)* evaluates the indices, *ind*, of the order of d, and the neighborhood function h in the following equations [27].

$$ind = order(d) \tag{22}$$

$$h(t) = \frac{1}{ind_i^2} \tag{23}$$

$$\omega_{j\kappa}(t+1) = \omega_{j\kappa}(t) + \alpha h(t)(s_{ij} - \omega_{j\kappa}(t)) \tag{24}$$

Adopting the definitions around the concept of linking energy, *E*, from a study by Faro et al. [29] for evaluating whether two classes are similar or not. Linking energy, *E*, is the inverse of the average mean distance of all the classes. Aside from *E*, classification is evaluated by the mean distance *D* between the centroids, *CoGs*, of two classes. Thus,

$$E = \frac{1}{L} \tag{25}$$

$$\lambda_j = \frac{1}{C} \sum_{i \in [1,n]} \sum_{r \in [j,C], j \notin r} \sqrt{\omega_{ir}^2 + \left(1 - \omega_{ij}\right)^2} \tag{26}$$

$$le = \underset{j \in [1,C]}{arg\,max}\{\lambda_j\} \tag{27}$$

$$L = \sum_{j \in [1,C]} \lambda_j \tag{28}$$

$$CoG_j = \frac{1}{n} \sum_{i \in [1,m], j \leq C-1} \omega_{ij} \tag{29}$$

$$D = \sqrt{\sum_{j \in [1,C-1]} \left(CoG_j - CoG_{j+1}\right)^2} \tag{30}$$

With the above equations, the optimum number of classes, *C*, is determined [27]. Once *C* is obtained, SOM feature selection can then be evaluated using genetic algorithm. This approach requires the function *somlearn(X,w,C,α)* that updates the following: *(i)* the SOM weight matrix, ω, according to Equation (24); *(ii)* the index, *lu*, of the minimum element in the distance vector *d*, as called for by Equations (20) and (21); *(iii)* the

cardinality of class c, that counts the number of hits of a neuron due to the occurrence of lu; *(iv)* the quantization error, qe_p with $p \in [1, n]$ and that $qe_1 = 0$; *(v)* rate of error reduction ∇e_p. Thus, the equation can be written with ω_{ij} taken to be the weights of the winner unit as:

$$qe_p = \sum_{i \in [1,m]} \sum_{j \in [1,C]} \sqrt{\left(X_{pi} - \omega_{ij}\right)^2} \qquad (31)$$

$$\nabla e_p = \frac{qe_p - qe_{p+1}}{qe_p} \qquad (32)$$

Algorithm 3 describes the SOM genetic algorithm (SOMGA) feature selection. Genetic algorithm is the optimization routine chosen for this approach, which provides a framework for expressing binary chromosomes to represent the inclusion and non-inclusion of selected features as 1s and 0s, respectively. Table 2 illustrates such representation of each row, $b \in P_k$ in a population of N chromosomes, $P \in \mathbb{R}^{N \times n}$.

Table 2. A chromosome of functional requirements, $FR_{j \in [1,n]}$

	FR$_j$												
P$_k$	1	1	0	1	0	1	0	0	0	0	1	0	1

Randomly initialized chromosomes form the first set of N parents in the population. At $t = 1$, the 0s in the chromosomes determine which features are excluded to form the selected feature matrix, $\hat{X}$, for training the SOM neural network. As in the previous algorithm, the quantization error is evaluated using the functions *initw* and *somlearn*. In randomly selecting the parents for mating, the following fitness function and selection probability are used [30]:

$$\underset{k \in [1,N]}{fitness_k} = \frac{1}{qe_k} \qquad (33)$$

$$P_k = \frac{fitness_k}{\sum_{k \in [1,N]} fitness_k} \qquad (34)$$

All of the unsupervised feature selection algorithms in this chapter can then be implemented on any high-level programming language, such as Python, R, or MATLAB. An application of these algorithms to the customer requirement definition phase of the conceptual design of a modular automated wood wall-framing machine is described in the next section. Graphs describing the results of the application have been generated in R (RStudio Team 2015).

A summary of the steps required to develop Algorithm 3 is presented below. This algorithm is a modification of the clustering technique by Kuo et al. [30] for SOM feature selection using genetic algorithm.

Algorithm 3 SOM feature selection using genetic algorithm

1: Initialization, $P_k \leftarrow \{b_{k1}, \cdots, b_{ki}, \cdots, b_{kn}\}, b_{ki} \in \{1, 0\}, k \in [1, N]$
      ▷ Population corresponds to N Chromosomes
2: Normalize dataset: $X \leftarrow$ normalized(X);
3: **while** $t \leq NG$ **do**   ▷ NG is the number of generations
4:     $\tilde{P} \leftarrow P$
5:     **for** $(k \in [1, N])$ **do**
6:         $\hat{X}_k \leftarrow X_k^{-lz}$ ▷ features corresponding to columns
                            ▷ with non-zero elements in P_k
7:         $w \leftarrow initw(\hat{X}_k, nu)$ ▷ Initialize weights of normalized dataset
8:         $qe_k \leftarrow somlearn(\hat{X}_k, w, C, \alpha)$ ▷ Evaluate quantization error
9:         $fitness_k \leftarrow \frac{1}{qe_k}$ ▷ $fitness_k \propto selection$
10:     **end for**
11:     $P \leftarrow crossover(P)$ ▷ Next generation after crossover
12:     $P \leftarrow mutate(P)$ ▷ Mutate if required
13:     $P \leftarrow elilism(\tilde{P})$ ▷ Retain 1 or 2 most fit chromosomes from $\tilde{P}$
14: **end while**
15: $\omega \leftarrow revorder(fitness)$ ▷ indices of the descending values in $fitness$
16: $f_{sel} \leftarrow P_{\omega_1}$ ▷ The 1st element of ω points to the maximum
                          ▷ $fitness$ and to the row of P where the desired
                          ▷ features can be retrieved

Application to modular automated wood wall framing machine

Prototypes of automated steel wall framing and wood wall framing machines are under development at the University of Alberta. Tamayo et al. have applied QFD in the conceptual design of these machines and their controllers and control panels [19, 31, 32]. A sketch of a completely designed and operational prototype of a fixed automated wood wall framing machine is depicted in Figure 6. This machine consists of nailing, drilling, and sawing stations, dragging mechanisms and a table that has a non-stationary side to accommodate different widths of wood wall panels. Aside from handling different widths, this machine will be capable of making 4×4 and 4×6 wood panels with studs only, with studs and window, and with studs and door at the construction site. Thus, it is important that this machine be modular.
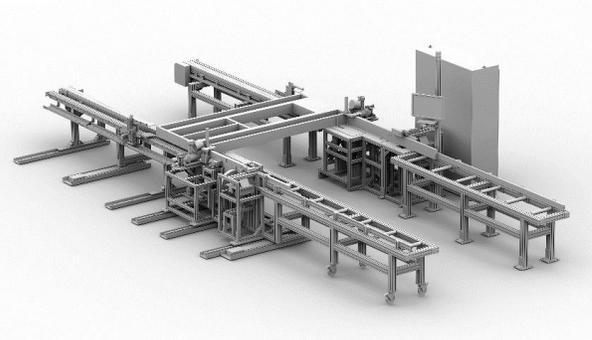

Figure 6. Automated wood wall framing machine

An improvement has been considered for the upcoming version of a fully modular wood wall framing machine, which can be transported to the construction site to minimize production cost. A transportable machine has fewer parts, high level of automation, and easier to assemble than its stationary counterpart.

Table 1. Quality Function Deployment of the modular automated wall framing machine

	FR_1: Operational tolerance	FR_2: Footprint of machine	FR_3: Wall frame production time	FR_4: Power usage	FR_5: Level of automation	FR_6: Supply chain cost	FR_7: Cost of installation and	FR_8: Component lifetime	FR_9: Moving parts	FR_{10}: Human intervention	FR_{11}: Modularity	FR_{12}: Dust	FR_{13}: Ease of training
CR_1	9		1		9	3							
CR_2		3											
CR_3	1		9	3	3	3			1	9			
CR_4			3	3	1				3	9			
CR_5	9		1		9	9	9	3	3		1		
CR_6					9	3		9	3			3	
CR_7	3				9	1	3		1		9		
CR_8		1			3	3	1		3	9			
CR_9			1		3					3			9
CR_{10}			1	9	3				3				

The QFD developed by the product design team at the customer requirement definition phase for the conceptual design of a transportable machine is shown in Table 3. This QFD matrix defines the input dataset, $X \in \mathbb{R}^{10 \times 13}$, for all of the feature selection methods previously discussed. It should be noted that the CRs are the m observations of X, where $m = 10$. The list of customer requirements associated with this QFD includes: (CR_1) quality, (CR_2) small, (CR_3) fast, (CR_4) cheap to operate, (CR_5) cheap to purchase, (CR_6) less maintenance, (CR_7) adaptability, (CR_8) safe, (CR_9) user friendly, and (CR_{10}) energy efficient.

PCA and Forward Orthogonal Search
Applying the PCA and FOS techniques to the QFD data ranking of the importance of the features is shown in Figure 7. It is apparent from the figure that the ranking of feature importance differs for PCA and FOS feature selection techniques. This contrast of ranking features is due to the difference in the selection criteria of these two linear methods [34]. PCA is a dimension reduction technique that has been modified by Equation (12) to be adopted for feature selection. Whereas FOS feature selection directly classifies the features according to Equation (17). These two linear feature selection techniques are considered filter methods and thus are simple, easy to understand, and not subject to overfitting [35]. Since the loadings corresponding to the last few features obtained through PCA are small, FR_{13}, FR_9, and FR_{12} can be ignored. Likewise, the insignificant features FR_5, FR_9, FR_1, and FR_3 can be ignored from the set of features selected using FOS.
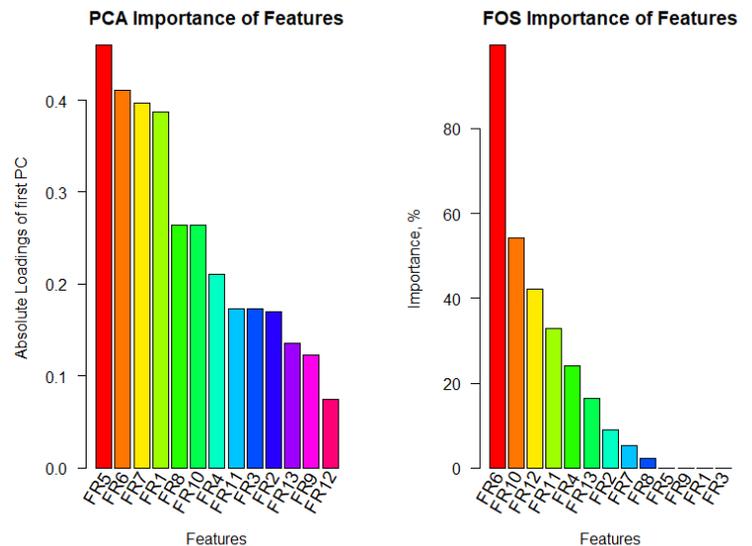
Figure 7. PCA and FOS feature importance

Which method is preferred cannot be determined at this point since the comparison has to be evaluated by the fitness criterion, which is invoked by the application of the SOM genetic algorithm. However, it is clear that PCA is not an appropriate feature selection technique to apply for the QFD data of Table 3 because the maximum eigenvalue is not significantly larger than the rest of the eigenvalues as the scree plot of Figure 8 indicates.
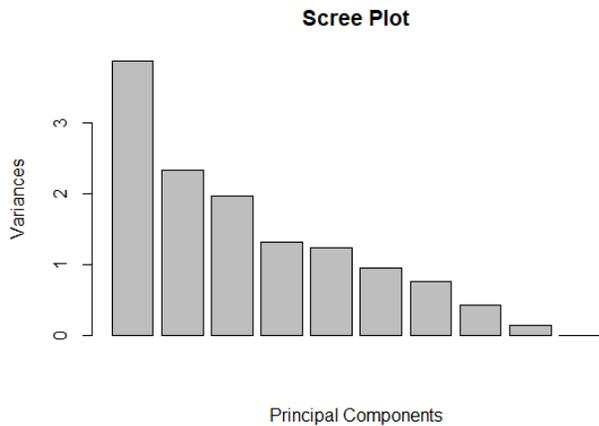
Scree Plot

Figure 8. PCA scree plot for the QFD of the modular automated wall framing machine

Automatic SOM Classification

Figure 9. Optimum number of classes, $C = 2$, detected in the 2nd iteration

SOMGA feature selection

SOMGA uses a nonlinear model and a probabilistic optimization technique that is applied for selecting the significant features of the modular automated wall framing machine QFD. Known as a wrapper approach, SOMGA is susceptible to data overfitting [35]. SOMGA, however, as an unsupervised feature selection method is less prone to overfitting than a supervised approach [36]. Moreover, overfitting is avoided by early stopping once a sufficiently complex structure is found [37]. Thus, to overcome overfitting the QFD data SOMGA uses single-layer network and adaptive classification.

Before applying the SOM model, the optimum number of classes, C, must be determined. Figure 9 reveals an optimum value of $C = 2$ after the application of automatic SOM classification. When SOMGA is applied to the QFD data of Table 3, the selected features are FR_4, FR_5, FR_9, FR_{11}, FR_{12}, and FR_{13} after ten iterations or 25.87 seconds at the crossover and mutation probabilities of 0.8 and 0.1 respectively. Due to its sensitivity to initial conditions, the genetic algorithm must be applied several times to arrive at the fitness value that realizes the desired modularity, level of automation, and cost objectives of the transportable wood wall framing machine.

Validation of results

As it is true with the ANN and FL methods used at the customer requirement definition stage of conceptual design, cross validation does not apply to the unsupervised feature selection techniques since the observations of the QFD dataset constitute the entire population of customer requirements. However, the product design team can easily verify the results of any feature selection methods used. For the QFD of Table 3, the product design team has verified those FRs selected by SOMGA to be the most acceptable since they are the least number of FRs that truly represent the objectives of a transportable automated wood wall framing machine.

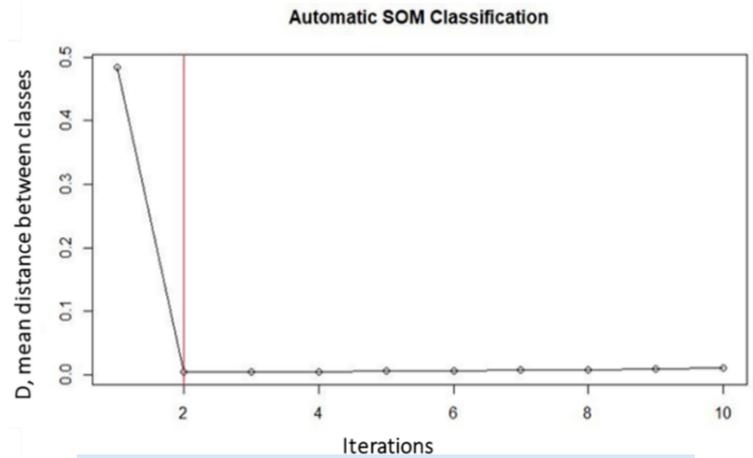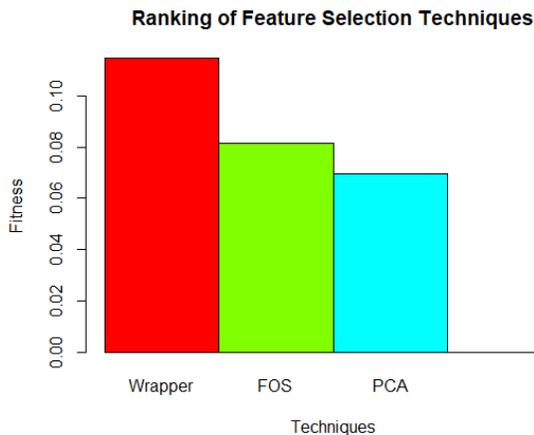A methodology that confirms the decision of the product design team has been developed by ranking the performance of PCA, FOS and SOMGA methods. This step efficiently facilitates the process of employing SOM model that provides the selected features in terms of chromosomes shown in Table 4.

Figure 10 shows the ranking according to fitness values of the selected FRs after applying the *somlearn* function of the SOM genetic algorithm. Ignoring the insignificant features selected, as previously noted, the corresponding set of features selected for the PCA and FOS methods are also represented as chromosomes in the table.

Table 4 presents the set of FRs selected for each of the three feature selection methods. Setting up the results in terms of chromosomes facilitates the ranking of these methods and the evaluation of their performance using the functions provided in Algorithm 3. Executing the *somlearn* and fitness functions for each of the chromosomes provides the comparison of the three techniques in Figure 10. Table 4 replaces the selection, crossover and mutation when Algorithm 3 is used as a mechanism for the ranking shown in Figure 10. This ranking provides a fair comparison of the application of the techniques since Table 4 is a mask applied to the same dataset, the same SOM network, and the same fitness criterion. An interesting observation can be derived from this figure. As to be expected, the SOMGA model tends to a global optimum due to its use of genetic algorithm, a numerical optimization routine. Not only does SOMGA provide the least number of FRs, but it is the only method that identifies modularity as a FR, which delivers the intent of designing an automated wood wall framing machine that can be used at the construction site. Non-exhaustive search algorithms, such as the FOS and SOMGA methods, do not guarantee an optimal solution [13]. However, a feature selection technique based on global search, SOMGA tends towards global optimum [38]. Although computationally expensive, wrappers, such a SOMGA, usually perform better than filters [38, 39].

Table 2. Chromosomes of selected features from the four methods

Name	FR_1: Operational tolerance	FR_2: Footprint of machine	FR_3: Wall frame production time	FR_4: Power usage	FR_5: Level of automation	FR_6: Supply chain cost	FR_7: Cost of installation and training	FR_8: Component lifetime	FR_9: Moving parts	FR_{10}: Human intervention	FR_{11}: Modularity	FR_{12}: Dust	FR_{13}: Ease of training
SOM Genetic Algorithm	0	0	0	1	1	1	1	0	0	0	1	0	1
FOS	0	1	0	1	0	1	1	1	0	1	1	1	1
PCA	1	1	1	1	1	1	1	1	0	1	1	0	0

Figure 10. Ranking by fitness of the different feature selection techniques

Since the PCA model fails to satisfy the requirement of a significantly large eigenvalue as Xu et al. [25] have indicated, it has the lowest performance. PCA, however, ought to be used as an alternative tool whenever possible due to its simplicity and efficiency in handling huge QFD matrices. It can also be used for weight initialization to speed up SOM learning [41].

Among the medium to low performers are the FOS and PCA models. These tools can also serve as alternative tools for obtaining fast preliminary results in applications where the use of SOMGA method can be computationally expensive and slow [42].

An exhaustive search, which maintains the SOM architecture, applied to the QFD dataset results in the optimum selected features of FR_5 and FR_{11} as shown in Table 5. These features are optimum, with reference to maximum fitness value, but are not adequate in fulfilling the design objectives of a transportable machine. Although suboptimal, the features selected by the SOMGA technique, adequately represent the desired objectives. Moreover, the six FRs selected through SOMGA are a significant reduction from the 13 FRs initially obtained at the customer requirement definition phase. Thus, on account of Equation (1), the selected FRs will result in reducing the design complexity of the transportable wood wall framing machine. It is apparent from Table 5, that genetic algorithm is the preferred feature selection technique since it is significantly faster than the exhaustive search approach.

Table 5. Results using genetic algorithm and exhaustive search wrapper methods

Feature Selection Technique	Features Selected	Fitness	Execution time, mins.
SOM Exhaustive Search	FR_5, FR_{11}	0.36862807	30.00
SOM Genetic Algorithm	FR_4, FR_5, FR_6, FR_7, FR_{11}, FR_{13}	0.11456863	0.43

Conclusion

Feature selection techniques play an important role in reducing the design complexity of a product or service by reducing the number of FRs at the customer requirement definition phase. As expected, the proposed unsupervised machine learning feature selection techniques produce suboptimal results because they are non-exhaustive. When these techniques are applied to the QFD of the transportable wood wall framing machine, however, it has been observed that these techniques are adequate in realizing the design objectives and in reducing the design complexity. An optimum number of two FRs have been obtained using an exhaustive search algorithm that is based on the SOM architecture, but these features do not fully capture the design objectives of the transportable machine. Thus, in general, the use of exhaustive search is not necessary especially in applications that involve huge QFD datasets requiring expensive computation.

A comprehensive set of unsupervised machine learning tools has been presented and applied to conceptual design of a transportable wood wall framing machine that is under development at the University of Alberta. These proposed machine learning approaches that are the main contribution of this paper address the limitations, indicated in the literature, of the existing methods in systematically selecting FRs from a QFD matrix [6, 7]. In place of these difficulties, however, users of the proposed tools have to contend with specifying the learning rate α and the thresholds L_{th} and D_{th} to preclude a priori knowledge of the number of classes, thus automating the classification process in SOM learning. Quantization error is used to measure the fitness in the SOMGA algorithm. Other measures such as entropy, Davies-Bouldin Index, and Gini Index can be explored as to the optimal selection of FRs and the automated integrated conceptual design package for future research. Another consideration for future research should include extending the application of the proposed feature selection methods to online CR or FR identification, design alternatives selection, project management, contract management, and marketing. Along with this consideration, a software interface should also be developed.

References

[1] Suh, N. P. (1998). "Axiomatic design theory for systems." *Research in Engineering Design*, 10(4), 189–209. https://doi.org/10.1007/s001639870001

[2] Subbaiah, K. V., Sai, K. Y., and Suresh, C. (2016). "QFD–ANP approach for the conceptual design of research vessels: A case study." *Journal of The Institution of Engineers (India): Series C*, 97(4), 539–546. https://doi.org/10.1007/s40032-016-0321-2

[3] Shin, J., and Kim, K. (2007). "Restructuring a house of quality using factor analysis." *Quality Engineering,* 9(4). https://doi.org/10.1080/08982119708919095

[4] Suh, N. P. (1999). "A Theory of complexity , periodicity and the design axioms." *Research in Engineering Design*, 11(2), 116–132. https://doi.org/10.1007/PL00003883

[5] Karsak, E. E., Sozer, S., and Alptekin, S. E. (2003). "Product planning in quality function deployment using a combined analytic network process and goal programming approach." *Computers & Industrial Engineering*, 44(1), 171–190. https://doi.org/10.1016/S0360-8352(02)00191-2

[6] Kiszová, Z., and Mazurek, J. (2012). "Modeling dependence and feedback in ANP with fuzzy cognitive maps." *Proceedings, 30th International Conference Mathematical Methods in Economics.*, pp. 558–563,. Karviná, Czech Republic, 11-13 September 2012

[7] Zaim, S., Sevkli, M., Camgöz-Akdağ, H., Demirel, O. F., Yesim Yayla, A., and Delen, D. (2014). "Use of ANP weighted crisp and fuzzy QFD for product development." *Expert Systems with Applications*, 41(9), 4464–4474,. https://doi.org/10.1016/j.eswa.2014.01.008

[8] Isalou, A. A., Zamani, V., Shahmoradi, B., and Alizadeh, H. (2013). "Landfill site selection using integrated fuzzy logic and analytic network process (F-ANP)." *Environmental Earth Sciences*, 68(6), 1745–1755. https://doi.org/10.1007/s12665-012-1865-y

[9] Guneri, A. F., Cengiz, M., and Seker, S. (2009). "A fuzzy ANP approach to shipyard location selection." *Expert Systems with Applications*, 36(4), 7992–7999. https://doi.org/10.1016/j.eswa.2008.10.059

[10] Achimugu, P., Selamat, A., Ibrahim, R., and Mahrin, M. N. R. (2014). "A systematic literature review of software requirements prioritization research." *Information and Software Technology*, 56(6), 568–585. https://doi.org/10.1016/j.infsof.2014.02.001

[11] Zebardast, E. (2013). "Constructing a social vulnerability index to earthquake hazards using a hybrid factor analysis and analytic network process (F'ANP) model." *Natural Hazards*, 65(3), 1331–1359. https://doi.org/10.1007/s11069-012-0412-1

[12] Weber, F. D., Mancuso, A. C. B., Senna, L. A. D. S., and Echeveste, M. E. S. (2013). "Quality Function Deployment in Airport Terminals: The Airport of Porto Alegre Case." *Journal of Traffic and Logistics Engineering*, 1(2), 222–227. https://doi.org/10.1109/TPAMI.2007.250607

[13] Wei, H. L., and Billings, S. A. (2007). "Feature subset selection and ranking for data dimensionality reduction." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(1), 162–166.

[14] Jacak, W., Pröll, K., and Winkler, S. (2014). "Neural Networks Based Feature Selection in Biological

Data Analysis." *Advanced Methods and Applications in Computational Intelligence,* Topics in Intelligent Engineering and Informatics, Springer, Heidelberg, 79–94. https://doi.org/10.1007/978-3-319-01436-4_5

[15] Malhi, A., and Gao, R. X. (2004). "PCA-based feature selection scheme for machine defect classification." *IEEE Transactions on Instrumentation and Measurement*, 53(6), 1517–1525. https://doi.org/10.1109/TIM.2004.834070

[16] Kohonen, T. (1982). "Self-organized formation of topologically correct feature maps." *Biological Cybernetics*, 43(1), 59–69. https://doi.org/10.1007/BF00337288

[17] Kutlu, Y., and Kuntalp, D. (2009). "Feature reduction method using self organizing maps." *Proceedings*, *International Conference on Electrical and Electronics Engineering - ELECO 2009,* Bursa, Turkey, pp. 129–132. https://doi.org/10.1109/ELECO.2009.5355238

[18] Köhler, A., Ohrnberger, M., Riggelsen, C., and Scherbaum, F. (2008). "Unsupervised Feature Selection for Pattern Discovery in Seismic Wavefields." *Proceedings, Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery at ECML/PKDD*, Antwerp, Belgium, pp. 106-121.

[19] Tamayo, E. C., Khan, Y. I., Qureshi, A. J., and Al-Hussein, M. (2018). "Design automation of control panels for automated modular construction machines." *Procedia CIRP*, 70, 404–409. https://doi.org/10.1016/j.procir.2018.02.004

[20] Kung, S. Y. (2014). *Kernel Methods and Machine Learning*. Cambridge University Press, UK.

[21] Cordella, C. (2012). "PCA: The Basic Building Block of Chemometrics." *Analytical Chemistry*, InTech, Croatia, 1–46. http://dx/doi.org/10.5772/3086

[22] Ghodsi, A. (2006). *Dimensionality Reduction A Short Tutorial*. Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada, vol. 37 no. 38. http://www.stats.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf

[23] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

[24] Song, F., Guo, Z., and Mei, D. (2010). "Feature Selection Using Principal Component Analysis." *Proceedings*, *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization,* Yichang, Hubei China, pp. 27–30. https://doi.org/10.1109/ICSEM.2010.14

[25] Xu, J., Xu, B., Zhang, W., and Cui Z. (2008).

"Principal Component Analysis Based Feature Selection For Clustering Clustering." *Proceedings, International Conference on Machine Learning and Cybernetics*, 12-15 July 2008, Kunming, China, pp. 12–15. https://doi.org/10.1109/ICMLC.2008.4620449

[26] Yin, H. (2008). "The Self-Organizing Maps: Background, Theories, Extensions and Applications. " *Computational Intelligence: A Compendium. Studies in Computational Intelligence, vol 115*. Springer, Berlin, Heidelberg, 715–762. https://doi.org/10.1007/978-3-540-78293-3_17

[27] Faro, A., Giordano, D., and Maiorana, F. (2007). "Discovering complex regularities by adaptive Self Organizing classification." *Enformatika I*, 27–30.

[28] Ng, A. Y., Jordan, M., and Weiss, Y. (2003). "On Spectral Clustering : Analysis and an algorithm." *Proceedings, 2002 Conference, Advances in Neural Information Processing Systems*, British Columbia, Canada, Dec. 9–14, pp. 849–856. https://papers.nips.cc/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf

[29] Faro, A., Giordano, D., and Maiorana, F. (2005). "Discovering Complex Regularities: from Tree to Semi-Lattice Classifications." *International Journal of Computer and Information Engineering*, 2(1), 34–39.

[30] Kuo, R. J., An, Y. L., Wang, H. S., and Chung W. J. (2006). "Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation." *Expert Systems with Applications*, 30(2), 313–324. https://doi.org/10.1016/j.eswa.2005.07.036

[31] RStudio Team (2015), "RStudio: Integrated Development for R." *RStudio, Inc., Boston, MA*. http://www.rstudio.com/

[32] Tamayo, E. C., Khan, Y. I., Qureshi, A. J., and Al-Hussein M. (2019). "Conceptual design of an automated steel wall framing assembly using axiomatic design and integrated function model." *Construction Robotics*, 3, 83–101. https://doi.org/10.1007/s41693-019-00022-8

[33] Tamayo, E., Bardwell, M., Qureshi, A., and Al-Hussein M. (2017). "Automation of a steel wall framing assembly." *Proceedings*, *ISEC 2017 - 9th International Structural Engineering and Construction Conference: Resilient Structures and Sustainable Construction*, Valencia, Spain, Jul. 24–29, pp. 1–6.

[34] Ilangovan, S., Antonykumar, V., and Balamurugan S. A. (2015). "An Empirical Study on Different Ranking Methods for Effective Data Classification." *Journal of Modern Applied Statistical Methods,* 14(2), 35–52. http://doi.org/10.22237/jmasm/1446350760

[35] Solorio-Fernández, S., Carrasco-Ochoa, J. A., and

Martínez-Trinidad, J. F. (2020). "A review of unsupervised feature selection methods." *Artificial Intelligence Review*, 53(2), 907–948. https://doi.org/10.1007/s10462-019-09682-y

[36] Iguyon I., and Elisseeff, A. (2003). "An introduction to variable and feature selection." *Journal of Machine Learning Research*, 3, 1157–1182.

[37] Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, Third Edition. Morgan Kaufmann Publishers.

[38] Caldas L. G., and Norford, L. K. (2002). "A design optimization tool based on a genetic algorithm." *Automation in Construction*, 11(2), 173–184. https://doi.org/10.1016/S0926-5805(00)00096-0

[39] Loughrey, J., and Cunningham, P. (2005). "Overfitting in Wrapper-Based Feature Subset Selection: The Harder You Try the Worse it Gets." *Procceedings, International Conference on Innovative Techniques and Applications of Artificial Intelligence, Research and Development in Intelligent Systems XXI. SGAI 2004*, pp. 33–43. https://doi.org/10.1007/1-84628-102-4_3

[40] Kumar, V. (2014). "Feature Selection: A literature Review." *Smart Computing Review*, 4(3), 221–229.

[41] Kinouchi, M., Takada, N., Kudo, Y., and Ikemura, T. (2002). "Quick learning for batch-learning self-organizing map." *Genome Informatics*, 267, 266–267. https://doi.org/10.11234/gi1990.13.266

[42] Mahrooghy, M., Younan, N. H., Anantharaj, V. G., Aanstoos, J., and Yarahmadian, S. (2012). "On the use of the genetic algorithm filter-based feature selection technique for satellite precipitation estimation." *IEEE Geoscience and Remote Sensing Letters*, 9(5), 963–967. https://doi.org/10.1109/LGRS.2012.2187513