# An analysis of finding the best strategies of water security for water source areas using an integrated IT2FVIKOR with machine learning

Nurnadiah Zamri[1]*, Wan Nur Amira Wan Azman[1], Mohamad Ammar Pairan[1], Siti Sabariah Abas[1] and Miaomiao Gao[2]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, Besut Terengganu, Malaysia, [2]Faculty of Biology Medicine and Health, School of Medical Sciences, University of Manchester, Manchester, United Kingdom

Worldwide, water security is adversely affected by factors such as population growth, rural–urban migration, climate, hydrological conditions, over-abstraction of groundwater, and increased per-capita water use. Water security modeling is one of the key strategies to better manage water safety and develop appropriate policies to improve security. In view of the growing global demand for safe water, intelligent methods and algorithms must be developed. Therefore, this paper proposes an integrated interval type-2 Fuzzy VIseKriterijumska Optimizcija I Kompromisno Resenje (IT2FVIKOR) with unsupervised machine learning (ML). This includes IT2FVIKOR for ranking and selecting a set of alternatives. Unsupervised machine learning includes hierarchical clustering, self-organizing map, and autoencoder for clustering, silhouette analysis and elbow method to find the most optimal cluster count, and finally Adjusted Rank Index (ARI) to find the best comparison within two clusters. This proposed integrated method can be divided into a two-phase fuzzy-machine learning-based framework to select the best water security strategies and categorize the polluted area using the water datasets from the Terengganu River, one of Malaysia's rivers. Phase 1 focuses on the IT2FVIKOR method to select five different strategies with five different criteria using five decision makers for finding the best water security strategies. Phase 2 continues the unsupervised machine learning where three different clustering algorithms, namely, hierarchical clustering, self-organizing map, and autoencoder, are used to cluster the polluted area in the Terengganu River. Silhouette analysis is applied along with the clustering algorithms to estimate the number of optimal clusters in a dataset. Then, the ARI is applied to find the best comparison within the original data with hierarchical clustering, self-organizing map, and autoencoder. Next, the elbow method is applied to double-confirm the best clusters for each clustering algorithm. Last, lists of polluted areas in each cluster are retrieved. Finally, this 2-phase fuzzy-Machine learning–based framework offers an alternative intelligent model to solve the water security problems and find the most polluted area.

# 1 Introduction

Universally, water security is beneath serious weight because of a number of aspects that include rapid population growth, hydrological conditions, increased per-capita water use, rural–urban migration, pollution of water resources, over-abstraction of groundwater, and climate change and variability (GWP, 2000b; Jones et al., 2009; Vörösmarty et al., 2010; Evengard et al., 2011). Over 1.1 billion individuals need to get to a secure water supply, and 2.6 billion individuals need satisfactory sanitation (WWC, 2000; UNESCO, 2009; Onda et al., 2012). Worldwide, endeavors and resolutions to upgrade water security have been taken by numerous organizations and parties, including the World Water Forum, the Group of Eight (G8), the United Nations, and many researchers (Jones et al., 2009). These researchers have discussed water security issues in many ways and areas, for example, Chapagain et al. (2022) carried out qualified analyses of the outcomes of the indicators and interconnected the strategies, plans, and policies of the cities with the indicators of the combined framework to detect policy gaps and propose steps to enhance urban water management in the five cities. Demerdash et al. (2022) introduced a new theory for water security: the Irrigation Water Security Quality-based index, where irrigation water security was offered based on quality and quantity. In addition, they estimated the impact of water quality parameters on soil deterioration, agriculture productivity, and public health. Veettil et al. (2022) provided an outline of water security assessment by concentrating on several water security indicators and the theory of water footprint. In addition, they also examined the application of various physically based hydrological models, such as variable infiltration capacity (VIC) and Soil and Water Assessment Tool (SWAT), on water security assessment at a provincial-to-continental rainwater level separately. Castro-Pardo et al. (2022) calculated the composite index to assess water security through the application of a DEA approach. The proposed model was used to accumulate nine indicators involving the four dimensions of water security: human health and wellbeing; state of the water environment; functions and responsibility of societies; and sustainability and stability of livelihood.

Apart from these, some studies on water security issues used multi-criteria decision-making (MCDM) for further analysis, for example, Fetanat et al. (2021) expanded methodology based on two multi-criteria decision-making (MCDM) methods called linear assignment and entropy under an intuitionistic fuzzy environment to support the synthetic decision to highlight energy recovery from wastewater treatment technologies. Tu et al. (2021) developed a thorough valuation methodology for measuring the degree of regional water resource coordination (RWRC). First, they created a security–equity–efficiency (SEE) evaluation indicator system. Then, they combined VIKOR and DEMATEL methods to rank six alternatives for the RWRC of the appropriate regions. Namany et al. (2021) integrated an evaluation tool based on the AHP method and the energy, water, and food (EWF) nexus to create a decision-making scheme that guides policymakers in creating national priorities and sectorial strategies. Next, they applied their developed methodology to measure the performance of the EWF sectors in the State of Qatar. Nie et al. (2018) integrated the best–worst method (BWM), DEMATEL, and TOPSIS to construct the water sustainability evaluation and analyzed the valuation results and significant factors toward the development of water security sustainability. However, previous studies are still limited in terms of solving decision problems with conflicting criteria and determining the compromise ranking list and the compromise solution obtained with the initial (given) weights (Qin et al., 2015). Therefore, this study will examine two main research questions. The first question is raised on the most important factor contributing toward the success of water security strategies, and the second question is whether there exists a method that can solve a maximum group utility and a minimum individual regret of the opponent. In response to these problems and research questions, this study aims to carry out a better solution in visualizing the preferences from the group of decision-makers (DMs) and minimum individual DM regret as well as to measure the best strategies that contribute to the success of water security. The method used must be capable of providing all these answers and considering different preferences among DMs. One of the appropriate methods purposely used to capture different preferences among multiple strategies of decision problems is the interval type-2 VIseKriterijumska Optimizacija I Kompromisno Resenje (IT2FVIKOR).

Various authors have discussed the IT2FVIKOR method. Han et al. (2021) proposed a hybrid model combining ANP, DEMATEL, and VIKOR with type-2 fuzzy linguistic variables in considering the fuzzy uncertainty in the procedure of the safety assessment of rail transit operation (SARTO) and the interrelationship and conflict among the criteria. Wan et al. (2021) developed a combined VIKOR and trapezoidal interval type-2 fuzzy (TrIT2F) technique on the best–worst method

(BMW). Then, their proposed method was exhibited with a makeshift (Fangcang) hospital selection example on COVID-19. Gul et al. (2021) improved the Fine–Kinney occupational risk assessment approach with IT2 FVIKOR to provide a useful and solid approach to the occupational health and safety risk assessment. Then, they applied their proposed method in a gun and rifle barrel external surface oxidation and coloring unit of a gun factory. Wang et al. (2019) combined the VIKOR method with IT2FS, where a new signed area function of IT2FSs was identified and the ranking score method based on the signed area function was proposed. Then, they converted the IT2 fuzzy number matrix to a ranking score value matrix and proved some theorems of the ranking score method. Then, they conducted a comparative analysis between the VIKOR method and the signed distance method to show the advantages of their proposed method. Qin and Liu (2019) defined a new distance measure for IT2FS and developed an IT2FVIKOR method based on the prospect theory. Then, they applied a numerical example on high-tech investment evaluation to illustrate the practicality and validity of the proposed method. Wu et al. (2019) proposed an integrated method based on the best–worst method (BMW) and the VIKOR technique in an interval type-2 fuzzy environment. Then, they applied green supply chain management to the proposed integrated method. Soner et al. (2017) integrated the VIKOR technique into the AHP method under an interval type-2 fuzzy environment. Then, they applied their proposed method to the hatch cover design selection problem, one of the structures of bulk carrier ships to protect cargo from external damages and prevent water ingress. Qin et al. (2015) extended the VIKOR method based on the interval type-2 environment and illustrated its applicability toward a high-tech risk evaluation case study. IT2FVIKOR has been proven to help people make decisions according to their preferences, in cases where there is more than one conflicting criterion. As per our knowledge, IT2FVIKOR has not yet been applied in any studies relating to water. However, IT2FVIKOR itself cannot meet the requirement of the growing global water security issue. The impact behind each decision on the growing global water security issue is still lacking and needs to be discussed and applied to the extended algorithms. Therefore, there is a need to develop other methods that can be integrated with IT2FVIKOR to achieve an advanced intelligent method. One of the appropriate methods suitable for the advanced intelligent method is machine learning (ML).

Machine learning (ML) is a branch of artificial intelligence (AI) that focuses on the use of big data and algorithms for predicting and classifying, clustering samples through experience and by the use of data (Tom, 1997). Two major learning in ML are supervised and unsupervised learning which are used in performing, or solving specific tasks, by

learning from experiences and/or the relationships between the data. Both supervised and unsupervised learning are used in various scenarios and with several datasets. The main difference between supervised and unsupervised learning is that supervised learning requires training prelabeled inputs to predict the predetermined outputs. Supervised learning best estimates the relationship between the output and input observed in the data. On the contrary, unsupervised learning deals with unlabeled data, output, and input, and its goal is to discover relationships and patterns between the data. Therefore, supervised learning can be used to predict the classification, and unsupervised learning can be used to find the optimal classification (Dong et al., 2021). Nowadays, research in ML is in demand due to the huge amount of data and the nature of ML itself, easy and cheap computation. Many studies discussed ML in many applications. Ziyadidagen et al. (2022) analyzed the factors that could affect the risk of COVID-19 infection and death using K-means clustering and several classification models. From the K-means clustering, numerous significant attributes of the COVID-19 risk were emphasized. Dong et al. (2021) extracted 16 climatic data variables to estimate the climatic regionalization for pavement infrastructure using both unsupervised and supervised ML, including three unsupervised ML approaches, namely, factor analysis, principal component analysis (PCA), and cluster analysis, and two supervised ML approaches, namely, artificial neural network (ANN) and Fisher's discriminant analysis. Ahmed et al. (2019) compared the supervised and unsupervised approaches to extract traffic-related tweets. They also developed a method to give the probability of level of congestion in real-time of tweets in Chennai.

ML has also been rigorously studied in many FMCDM methods, for example, Umar and Saraswat (2022) proposed a divergence measure for picture fuzzy sets and applied decision-making in ML, which were medical diagnosis, pattern recognition, and clustering using numerical illustration. Then, they justified their proposed method by assessing their results using comparative analysis. Yazici et al. (2022) identified and prioritized individual tacit knowledge criteria using ML methods and FAHP. ML methods including support vector machine (SVM), logistic regression, and artificial neural network were used to extract the variable. Their results show that the SVM and fuzzy analytic hierarchy process (FAHP) defined communication between employees and supervisors, time efficiency of employees, and innovative capability of employees as the most important tacit knowledge criteria. Then, Alazemi et al. (2022) focused on the role of internal factors in small- and medium-scale supply chains, where a three-phase fuzzy-based framework was proposed to address the product completion time problem. These three phases include supervised ML to classify the production alternatives, quantitative research, and fuzzy-based

heuristics to generate production. Mohsin et al. (2022) offered a novel approach for selecting landfill sites and sustainable solid waste management (SWM) in India using integrated FAHP-SVM and FAHP-RF. Their findings provide a guideline to decision-makers (DMs) and planners for optimal landfill site selection in other cities of developing countries. Jain et al. (2021) proposed multi-attribute decision-making (MADM) with ML algorithms for employee churn (ECn) in solving the revenue and brand image of the organization. MADM methods include TOPSIS for quantifying the importance of the employees to perform their class-based categorization. ML algorithms include CatBoost for predicting class-wise ECn. Then, Nilashi et al. (2019) proposed a hybrid method for online review analysis through MCDM, text mining, and predictive learning techniques to find the relative importance of factors affecting travelers' decision-making in selecting green hotels with spa services. ML algorithms include a self-organizing map (SOM) for cluster analysis, latent Dirichlet analysis (LDA) for analyzing textual reviews, TOPSIS for ranking hotel features, and neuro-fuzzy to reveal customer satisfaction levels. Shen and Tzeng (2016) proposed a novel multiple attribute decision-making model to resolve—from ranking/selection to improvement planning—the problems of business analytics in finance, based on the similarity with positive context (rules) and the dissimilarity with negative ones. Their proposed method combines fuzzy-rough machine learning including the dominance-based rough set approach (DRSA), VIKOR, and bipolar decision model. Due to the eligibility of ML with MCDM, our studies enhanced IT2 FVIKOR with ML to solve the water security issue.

Therefore, this paper focuses on integrating IT2FVIKOR with unsupervised ML. IT2FVIKOR has the ability to solve uncertainty issues by combining the solution into a maximum group utility and a minimum individual regret of the opponent. In addition, IT2FVIKOR would discover the best solution and a compromise solution by prioritizing important factors (Moradi et al., 2002). This includes IT2FVIKOR for ranking and selecting a set of alternatives; unsupervised ML includes hierarchical clustering, SOM, and autoencoder for clustering; silhouette analysis and elbow method to find the most optimal number of clusters; and last, Adjusted Rank Index (ARI) to find the best comparing within two clusters. Unsupervised ML, able to determine the relevancy based on data features (El Bouchefry and de Souza, 2020), has no attention to structured semantic relationships, and therefore, is suitable to be applied to heterogeneous data (Jain, 2010). This proposed integrated method can be divided into a two-phase fuzzy-ML-based framework to select the best water quality strategies and categorize the polluted area in the Terengganu River, which is one of the rivers in Malaysia. Phase 1 focuses on the IT2FVIKOR method to select five different strategies with five different criteria using five DMs for finding the best water quality strategies. Phase
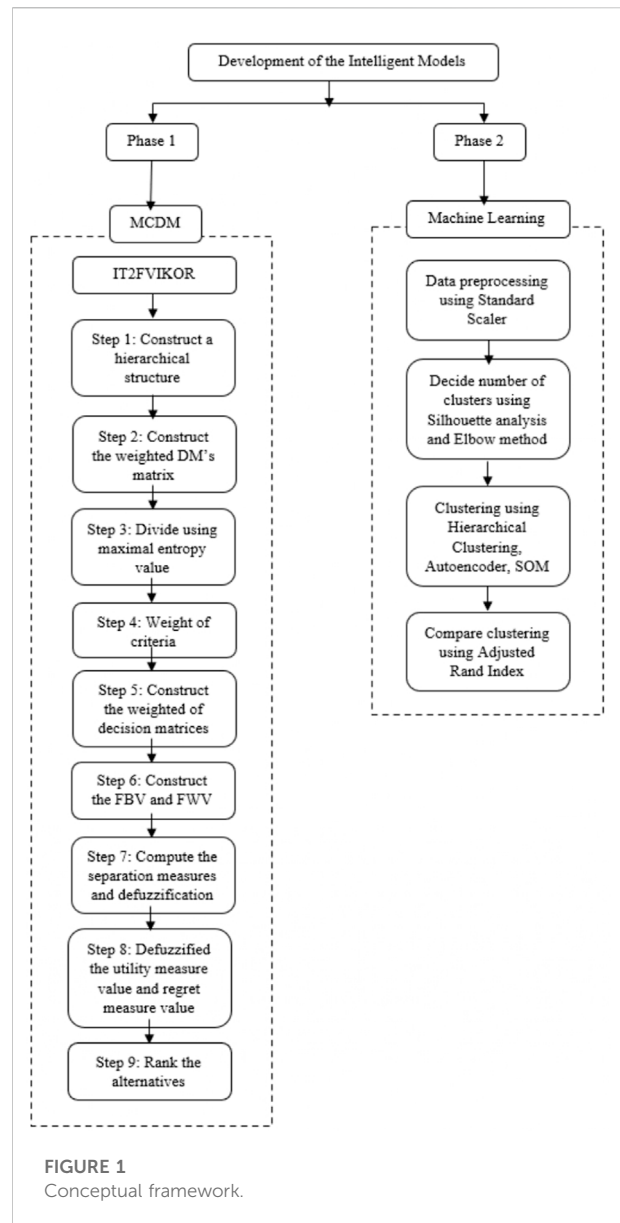


FIGURE 1
Conceptual framework.

2 continues the unsupervised ML, where three different clustering algorithms, namely, hierarchical clustering, SOM, and autoencoder, are used to cluster the polluted area in the Terengganu River. Silhouette analysis is applied along with the clustering algorithms to estimate the number of optimal clusters in a dataset. Then, the ARI is applied to find the best comparison within the original data with hierarchical clustering, SOM, and autoencoder. Next, the elbow method is applied to double-confirm the best clusters for each clustering algorithm. Last, lists of polluted areas in each cluster are retrieved. Finally, this two-phase fuzzy-ML-based framework can be concluded as the best water security strategies to be applied to the most polluted area. The rest of this paper is organized as follows:

Section 2 describes the methodology framework including the concepts of IT2FVIKOR procedures; the unsupervised ML procedures including hierarchical clustering, autoencoder, and SOM; and silhouette analysis and ARI. Section 3 presents an illustrative example to demonstrate the feasibility and consistency of the integrated IT2FVIKOR with ML algorithms. Section 4 presents the conclusion.

# 2 The methodology framework

The conceptual framework is illustrated in Figure 1, which shows the entire fundamental process involved in the study. Generally, the conceptual framework can be divided into two different phases. Phase 1 focuses only on the MCDM method, which is IT2FVIKOR with step-by-step flows. Phase 2 demonstrates the ML part consisting of all the preprocessing steps and unsupervised techniques. The details of each process are described in each subsection.

## 2.1 The interval type-2 fuzzy VlseKriterijumska Optimizcija I Kompromisno Resenje procedures

The IT2FVIKOR, with an equitable linguistic scale and Z-Numbers method, is developed to achieve a higher rational, systematic decision-making process to discover the best solution and a compromise solution that considers both sides of the scale. Then, entropy weight method is used for weighting values to greatly reduce the decision-making burden in the subjective weighting process and achieves a rather consistent weighting outcome. The IT2FVIKOR procedure is explained step-by-step as follows:

Assume that there is a finite set $X$ of alternatives, where $X = \{x_1, x_2 \ldots, x_n\}$, and assume that there is a finite set $F$ attributes, where $F = \{f_1, f_2 \ldots, f_m\}$. Assume that there are $k$ DMs $D_1, D_2, \ldots$, and $D_k$. The proposed method is now presented as follows:

Step 1: Construction of a hierarchical diagram of the IT2FMCDM problem.

We construct the decision matrix $Y_p$ of the $p$th DM and the average decision matrix $\overline{Y}$, shown as follows:

$$Y_p = \left[\tilde{\tilde{f}}_{ij}^p\right]_{m \times n}$$

$$= \begin{matrix} & \overset{x_1}{} & \overset{x_2}{} & \cdots & \overset{x_n}{} \\ \begin{matrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{matrix} & \left[\begin{matrix} \left(\left(\tilde{\tilde{f}}_{11}^{P-}, \tilde{\tilde{f}}_{11}^{P+}\right); \left(\tilde{\tilde{z}}_{11}^{P-}, \tilde{\tilde{z}}_{11}^{P+}\right)\right) & \left(\left(\tilde{\tilde{f}}_{12}^{P-}, \tilde{\tilde{f}}_{12}^{P+}\right); \left(\tilde{\tilde{z}}_{12}^{P-}, \tilde{\tilde{z}}_{12}^{P+}\right)\right) & \cdots & \left(\left(\tilde{\tilde{f}}_{1n}^{P-}, \tilde{\tilde{f}}_{1n}^{P+}\right); \left(\tilde{\tilde{z}}_{1n}^{P-}, \tilde{\tilde{z}}_{1n}^{P+}\right)\right) \\ \left(\left(\tilde{\tilde{f}}_{21}^{P-}, \tilde{\tilde{f}}_{21}^{P+}\right); \left(\tilde{\tilde{z}}_{21}^{P-}, \tilde{\tilde{z}}_{21}^{P+}\right)\right) & \left(\left(\tilde{\tilde{f}}_{22}^{P-}, \tilde{\tilde{f}}_{22}^{P+}\right); \left(\tilde{\tilde{z}}_{22}^{P-}, \tilde{\tilde{z}}_{22}^{P+}\right)\right) & \cdots & \left(\left(\tilde{\tilde{f}}_{2n}^{P-}, \tilde{\tilde{f}}_{2n}^{P+}\right); \left(\tilde{\tilde{z}}_{2n}^{P-}, \tilde{\tilde{z}}_{2n}^{P+}\right)\right) \\ \left(\left(\tilde{\tilde{f}}_{m1}^{P-}, \tilde{\tilde{f}}_{m1}^{P+}\right); \left(\tilde{\tilde{z}}_{m1}^{P-}, \tilde{\tilde{z}}_{m1}^{P+}\right)\right) & \left(\left(\tilde{\tilde{f}}_{m2}^{P-}, \tilde{\tilde{f}}_{m2}^{P+}\right); \left(\tilde{\tilde{z}}_{m2}^{P-}, \tilde{\tilde{z}}_{m2}^{P+}\right)\right) & \cdots & \left(\left(\tilde{\tilde{f}}_{mn}^{P-}, \tilde{\tilde{f}}_{mn}^{P+}\right); \left(\tilde{\tilde{z}}_{mn}^{P-}, \tilde{\tilde{z}}_{mn}^{P+}\right)\right) \end{matrix}\right] \end{matrix},$$

$$(1)$$

$$\overline{Y} = \left[\tilde{\tilde{f}}_{ij}\right]_{m \times n}$$

where $\tilde{\tilde{f}}_{ij} = \left(\frac{\tilde{\tilde{f}}_{ij}^1 \oplus \tilde{\tilde{f}}_{ij}^2 \oplus \ldots \oplus \tilde{\tilde{f}}_{ij}^k}{k}\right)$, $\left(\tilde{\tilde{f}}_{ij}^{P-}, \tilde{\tilde{f}}_{ij}^{P+}\right)$ is the linguistic scale of restriction of the extended IT2FVIKOR using equitable linguistic scales and Z-Numbers, $\left(\tilde{\tilde{z}}_{ij}^{P-}, \tilde{\tilde{z}}_{ij}^{P+}\right)$ is the linguistic scale of reliability of the extended IT2FVIKOR using equitable linguistic scales and Z-Numbers, $1 \le i \le m$, $1 \le j \le n$, $1 \le p \le k$, and $k$ denotes the number of DMs.

Step 2: Construction of the weighted DM matrix.

The proposed entropy weight method based on IT2FSs is defined as follows:

Let us assign $\tilde{\tilde{f}}_{ij} = \left(\left(\tilde{\tilde{f}}_{ij}^L, \tilde{\tilde{f}}_{ij}^U\right); \left(\tilde{\tilde{z}}_{ij}^L, \tilde{\tilde{z}}_{ij}^U\right)\right)$.

Thus.

$$E\left(\tilde{\tilde{A}}_{ij}^L\right) = \frac{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^{near} - \left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^L\right]}}{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^{far} - \left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^L\right]}}.$$

$$E\left(\tilde{\tilde{A}}_{ij}^U\right) = \frac{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^{near} - \left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^U\right]}}{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^{far} - \left(\tilde{\tilde{f}}_{ij}^-, \tilde{\tilde{f}}_{ij}^+\right)^U\right]}}.$$

$$zE\left(z\tilde{\tilde{A}}_{ij}^L\right) = \frac{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^{near} - \left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^L\right]}}{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^{far} - \left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^L\right]}}.$$

$$zE\left(\tilde{\tilde{A}}_{ij}^U\right) = \frac{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^{near} - \left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^U\right]}}{\sqrt{\sum_{i=1}^n \left[\left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^{far} - \left(\tilde{\tilde{z}}_{ij}^-, \tilde{\tilde{z}}_{ij}^+\right)^U\right]}}, \quad (2)$$

where $E\left(\tilde{\tilde{A}}_{ij}\right) = \left[E\left(\tilde{\tilde{A}}_{ij}^L\right), E\left(\tilde{\tilde{A}}_{ij}^U\right)\right], \left[zE\left(z\tilde{\tilde{A}}_{ij}^L\right), zE\left(\tilde{\tilde{A}}_{ij}^U\right)\right]$.

Step 3: Divide using maximal entropy value.

Then, all the entropy values are divided by using the maximal entropy value, and the value of $ij$ h is used to represent the outcomes of the maximal entropy value. Therefore, it can be defined as follows:

$$\tilde{\tilde{h}}_{ij} = \frac{1}{FOU\left(\tilde{\tilde{h}}_{ij}\right)} = \left[\left(\tilde{\tilde{h}}_{ij}^L, \tilde{\tilde{h}}_{ij}^U\right); \left(z\tilde{\tilde{h}}_{ij}^L, z\tilde{\tilde{h}}_{ij}^U\right)\right],$$

$$\tilde{\tilde{h}}_{i1} = \left[\left(\frac{E\left(\tilde{\tilde{A}}_{i1}^L\right)}{\max\left(E\left(\tilde{\tilde{A}}_{i1}^L\right)\right)}, \frac{E\left(\tilde{\tilde{A}}_{i1}^U\right)}{\max\left(E\left(\tilde{\tilde{A}}_{i1}^U\right)\right)}\right); \left(\frac{zE\left(z\tilde{\tilde{A}}_{i1}^L\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{i1}^L\right)\right)}, \frac{zE\left(z\tilde{\tilde{A}}_{i1}^U\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{i1}^U\right)\right)}\right)\right],$$

$$\tilde{\tilde{h}}_{i2} = \left[\left(\frac{E\left(\tilde{\tilde{A}}_{i2}^L\right)}{\max\left(E\left(\tilde{\tilde{A}}_{i2}^L\right)\right)}, \frac{E\left(\tilde{\tilde{A}}_{i2}^U\right)}{\max\left(E\left(\tilde{\tilde{A}}_{i2}^U\right)\right)}\right); \left(\frac{zE\left(z\tilde{\tilde{A}}_{i2}^L\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{i2}^L\right)\right)}, \frac{zE\left(z\tilde{\tilde{A}}_{i2}^U\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{i2}^U\right)\right)}\right)\right], \ldots$$

$$\tilde{\tilde{h}}_{in} = \left[\left(\frac{E\left(\tilde{\tilde{A}}_{in}^L\right)}{\max\left(E\left(\tilde{\tilde{A}}_{in}^L\right)\right)}, \frac{E\left(\tilde{\tilde{A}}_{in}^U\right)}{\max\left(E\left(\tilde{\tilde{A}}_{in}^U\right)\right)}\right); \left(\frac{zE\left(z\tilde{\tilde{A}}_{in}^L\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{in}^L\right)\right)}, \frac{zE\left(z\tilde{\tilde{A}}_{in}^U\right)}{\max\left(zE\left(z\tilde{\tilde{A}}_{in}^U\right)\right)}\right)\right].$$

$$(3)$$

Then, the decision matrix, $D$, can be expressed as follows:

$$D = \begin{bmatrix} \tilde{\tilde{h}}_{11} & \tilde{\tilde{h}}_{12} & \cdots & \tilde{\tilde{h}}_{1n} \\ \tilde{\tilde{h}}_{21} & \tilde{\tilde{h}}_{22} & \cdots & \tilde{\tilde{h}}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\tilde{h}}_{m1} & \tilde{\tilde{h}}_{m2} & \cdots & \tilde{\tilde{h}}_{mn} \end{bmatrix}. \tag{4}$$

Step 4: Weight of criteria.

The weight of criteria is calculated by using the fuzzy entropy weight-based IT2FS formula. The $W_j$ is used to present the outcome of the weight value of criteria $j$. Thus, it can be defined as follows:

$$\tilde{\tilde{w}}_j = \frac{1}{F(\tilde{\tilde{w}}_j)} = \left[ (\tilde{w}_j^L, \tilde{w}_j^U); (\widetilde{zw}_j^L, \widetilde{zw}_j^U) \right], \tag{5}$$

$$\left[ (\tilde{w}_j^L, \tilde{w}_j^U); (\widetilde{zw}_j^L, \widetilde{zw}_j^U) \right] = \left[ \left( \frac{1 - \tilde{a}_j^L}{\tilde{T}^L}, \frac{1 - \tilde{a}_j^U}{\tilde{T}^U} \right), \left( \frac{1 - z\tilde{a}_j^L}{z\tilde{T}^L}, \frac{1 - z\tilde{a}_j^U}{z\tilde{T}^U} \right) \right], \tag{6}$$

where $(\tilde{\tilde{a}}_j, \tilde{\tilde{za}}_j) = (\frac{\tilde{\tilde{h}}_{i1} + \tilde{\tilde{h}}_{i2} + \ldots + \tilde{\tilde{h}}_{in}}{n}; \frac{z\tilde{\tilde{h}}_{i1} + z\tilde{\tilde{h}}_{i2} + \ldots + z\tilde{\tilde{h}}_{in}}{n})$,

$(\tilde{\tilde{T}}, z\tilde{\tilde{T}}) = (\tilde{\tilde{h}}_{i1} + \tilde{\tilde{h}}_{i2} + \ldots + \tilde{\tilde{h}}_{in}; z\tilde{\tilde{h}}_{i1} + z\tilde{\tilde{h}}_{i2} + \ldots + z\tilde{\tilde{h}}_{in})$   $1 \le j \le n$.

Step 5: Construct the weighted decision matrices.

$x_1 \ x_2 \cdots x_n$

$$\bar{Y}_w = \left[ \tilde{\tilde{v}}_{ij} \right]_{m \times n}$$
$$= \begin{matrix} f_1 \\ f_2 \\ \vdots \\ f_1 \end{matrix} \begin{bmatrix} (\tilde{\tilde{v}}_{11}; \tilde{\tilde{zv}}_{11}) & (\tilde{\tilde{v}}_{12}; \tilde{\tilde{zv}}_{12}) & \cdots & (\tilde{\tilde{v}}_{1n}; \tilde{\tilde{zv}}_{1n}) \\ (\tilde{\tilde{v}}_{21}; \tilde{\tilde{zv}}_{21}) & (\tilde{\tilde{v}}_{22}; \tilde{\tilde{zv}}_{22}) & \cdots & (\tilde{\tilde{v}}_{2n}; \tilde{\tilde{zv}}_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (\tilde{\tilde{v}}_{m1}; \tilde{\tilde{zv}}_{m1}) & (\tilde{\tilde{v}}_{m2}; \tilde{\tilde{zv}}_{m2}) & \cdots & (\tilde{\tilde{v}}_{mn}; \tilde{\tilde{zv}}_{mn}) \end{bmatrix}, \tag{7}$$

where $\tilde{\tilde{v}}_{ij} = \tilde{\tilde{w}}_j \otimes \tilde{\tilde{f}}_{ij}$, $1 \le i \le m$, and $1 \le j \le n$.

Step 6: Construct the fuzzy best value (FBV) and fuzzy worst value (FWV).

Choose the FBV and FWV values using the following equations:

$$\tilde{f}_j^* = \max_i \tilde{\tilde{v}}_{ij}, \ \tilde{f}_j^- = \min_i \tilde{\tilde{v}}_{ij}, \ i = 1, 2, \ldots, m; j = 1, 2, \ldots, n, \tag{8}$$

where $\tilde{f}_j^*$ refers to FBV and $\tilde{f}_j^-$ refers to FWV. Then, the values are defuzzified. Here, $\tilde{f}_j^*$ is the maximum value of $\tilde{\tilde{v}}_{ij}$, which is the maximum majority rule or maximum group utility, whereas $\tilde{f}_j^-$ is the minimum value of $\tilde{\tilde{v}}_{ij}$, which is the minimum individual regret of the opponent.

Step 7: Compute the separation measures and defuzzification.

Evaluate the following values,

$$\frac{W_i\left( \tilde{f}_i^* - \tilde{\tilde{v}}_{ij} \right)}{\tilde{f}_i^* - \tilde{f}_i^-}, \tag{9}$$

$$\tilde{S}_i = \sum_{i=1}^k \frac{\tilde{w}_i\left( \tilde{f}_i^* - \tilde{\tilde{v}}_{ij} \right)}{\tilde{f}_i^* - \tilde{f}_i^-}, \tag{10}$$

$$\tilde{R}_i = \max_j \left[ \frac{\tilde{w}_i\left( \tilde{f}_i^* - \tilde{\tilde{v}}_{ij} \right)}{\tilde{f}_i^* - \tilde{f}_i^-} \right], \tag{11}$$

where $\tilde{S}$ represents the utility measure and $\tilde{R}$ represents the regret measure. $\tilde{S}_i$ and $\tilde{R}_i$ can be calculated using the sum of the FBV distance with respect to all criteria.

Step 8: Defuzzify the utility measure value ($\tilde{S}_i$) and regret measure value ($\tilde{R}_i$) using the formula given as follows:

$$\text{Defuzzified} = \frac{\frac{(u_U - l_U) + (\beta_U \cdot m_{1U} - l_U) + (\propto_U \cdot m_{2U} - l_U)}{4} + l_U + \left[ \frac{(u_L - l_L) + (\beta_L \cdot m_{1L} - l_L) + (\propto_L \cdot m_{2L} - l_L)}{4} + l_L \right]}{2}. \tag{12}$$

Step 9: Rank the alternatives.

Calculate the following values,

$$\tilde{S}^* = \min_i \tilde{S}, \ \tilde{S}^- = \max_i \tilde{S}, \tag{13}$$

where $\tilde{S}^*$ is the value for the maximum group of utility and $\tilde{S}^-$ is the minimum value for the maximum group of utility.

Then,

$$\tilde{R}^* = \min_i \tilde{R}, \ \tilde{R}^- = \max_i \tilde{R}, \tag{14}$$

where $\tilde{R}^*$ is the value of the minimum individual regret of the opponent and $\tilde{R}^-$ is the minimum value.

Next,

$$\tilde{Q}_i = v\frac{\left( \tilde{S}_j - \tilde{S}^* \right)}{\left( \tilde{S}^- - \tilde{S}^* \right)} + (1 - v)\frac{\left( \tilde{R}_j - \tilde{R}^* \right)}{\left( \tilde{R}^- - \tilde{R}^* \right)}, \tag{15}$$

where $\tilde{Q}_i$ is the index for both $\tilde{S}^*$ and $\tilde{R}^*$, whereas $v$ is the weight of the strategy to be used in the maximum group of utility.

$v > 0.5$ refers to the maximum majority of rule, and $v \le 0.5$ refers to the individual regret of the opponent. The normal value is when v is 0.5.

Last, $\tilde{S}^*$, $\tilde{R}^*$, and $\tilde{Q}_i$ are sorted and ranked in a decreasing order. The decreasing order helps decrease the gaps in the criteria, and the best is chosen based on the lowest rank value.

## 2.2 The unsupervised machine learning procedures

Unsupervised machine learning is a type of algorithm that learns patterns from unlabeled data. Through mimicry, the machine is forced to build the same patterns and characteristic representation from the important mode of learning in each situation. Some of the most used algorithms in unsupervised learning include clustering, anomaly detection, and approaches for learning latent variable models. This study only focuses on clustering algorithms with two comparison methods. The details of each algorithm are explained in the next subsection.

## 2.2.1 Hierarchical clustering

Ward studied hierarchical clustering techniques based on minimizing the 'loss of information' from joining two groups (Ward, 1963). This method is usually implemented with a loss of information taken to be an increase in an error sum of squares criterion (ESS). Thus, the ESS is defined as follows:

$$o(t) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \sum_{j=1}^{n} \left(x_{ij} - x_{kj}\right)^2, \quad (16)$$

where $K$ is the number of clusters, $j$ is the sample index $(1, 2, \cdots, n)$, $C_k$ is a k-th cluster, $x_i$ indicates elements in $C_k$, $x_{ij}$ denotes the i-the item in the j-th cluster $C_j$, and $\mu_{kj}$ is a mean vector in $C_k$.

At first, each cluster comprises a single item. At each step in the analysis, the union of every possible pair of clusters is considered, and the two clusters whose combination results in the smallest increase in ESS (a minimum loss of information) are joined. When all the clusters are combined in a single group of N items, the clustering analysis is carried out. The results of Ward's method can be displayed as a dendrogram. The vertical axis gives the values of ESS at which the merges occur (Richard and Dean, 2002).

## 2.2.2 Autoencoder algorithms

An autoencoder is an unsupervised feed-forward multi-layer neural network where the desired output is the input itself (Silberer and Lapata, 2014; Zhou and Paffenroth, 2017). The autoencoder is used to discover clusters of similar instances in an unlabeled dataset and consists of two parts (Delgado, 2019):

Encoder $f_\theta$ maps an input vector to a latent representation $y^{(i)} = f_{\theta'}(x_i) = s(Wx_i + b)$, with $s$ being a nonlinear activation function.

Decoder $g_{\theta'}$ aims to reconstruct the input $x_i$ from $y_i$. For that purpose, $x_i = g_{\theta'}(y_i) = s(W'y_i + b')$ is applied.

In this training, we need to determine $\hat{\theta} = \{W, b\}$ and $\hat{\theta}' = \{W', b'\}$ to minimize the reconstruction error of input vectors $x_i$.

$$\hat{\theta}, \hat{\theta}' = \arg_{\theta,\theta'}^{\min} \frac{1}{n} \sum_{i=1}^{n} \left(L\left(x_i, g_{\theta'}(f_\theta(x_i))\right)\right), \quad (17)$$

where $L$ is a loss function and $\theta$ and $\theta'$ can be optimized by gradient descent methods.

Moreover, all the layers used rectified linear unit (ReLU) activation function because it is easier to train and often achieves better performance. The optimizer selected to train the neural network is Adam, and the metric that is going to be minimized is the mean squared error (MSE) for loss function, which can be described as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2, \quad (18)$$

where $Y_i$ is the observed values and $\widehat{Y}_i$ is the predicted values.

## 2.2.3 Self-organizing map algorithms

The self-organizing map (SOM) is a prominent unsupervised neural network, considered a 2D mapping of the data group (Kohonen, 1984). In other words, the SOM converts a non-linear projection of n-dimensional multivariate data onto a low-dimensional display. The SOM net is made up of a number of nodes usually organized in a rectangular or hexagonal grid. Each neuron in the SOM array is linked to the input vector $x_i$, allocated with a weight vector $w_i$. Moreover, the nodes placed adjacent to the cells are constantly updated *via* the neighborhood function. The SOM algorithm contains two main phases: the competitive and the cooperative (Hoffmann, 2005; Liu et al., 2008). The former attempts to select the best matching neuron, while the latter adjusts the weight of the winner and its immediate lattice neighbors. During the SOM training procedure, a sample vector $x$ is randomly selected. Then, the Euclidean distances are calculated between the input vector and all neurons. Based on the minimization of the distance between a sample and other weighted vectors, the best matching unit is computed as

$$\|x - m_c\| = \min_i \|x - m_i\|, \quad (19)$$

where $c$ represents the winning neuron and $m_c$ is the closest vector to $x$ on the map.

The weight vector $m_c$ related to the best matching unit is updated in order to attune the sample $x$ even closer. Moreover, the nodes around the best matching unit are updated as well. The updated version of Eq. 1 may be reformulated as

$$m_i(t + 1) = m_i(t) + h_{ci}(t)^* (x - m_i(t)), \quad (20)$$

where $h_{ci}(t)$ is the neighborhood kernel centered on the winner unit and $t$ denotes the time.

$$h_{ci}(t) = a(t)^* \exp\left(-\frac{\|r_c - r_i\|^2}{2o(t)^2}\right), \quad (21)$$

where $a(t)$ indicates the learning rate considered the degree of matching, and it lessens monotically with regression steps (time). $o(t)$ is the radius of expansion of the neighborhood function. $r_c$, $r_i$ are locations of neurons $c$ and $i$ on the SOM grid, respectively.

## 2.2.4 Silhouette analysis

Silhouette is used to evaluate how well clustering results are clearly separated. In order to construct a silhouette, we only need two things: the partition and collection of all proximities between objects. For each object $i$, we need to introduce a certain value $s(i)$ and take any object $i$ in the dataset and denote by A the cluster to which it has been assigned. Also, $a(i)$ is the average Euclidean distance of object $i$ to all other objects of A, and $b(i)$ is the minimum of the average Euclidean distance between $i$ and the objects in other clusters to which the object $i$ does not belong. Then, $s(i)$ is obtained by combining $a(i)$ and $b(i)$ as follows in Eq. 2. Finally, the silhouette index takes the average of $s(i)$ for all

samples to evaluate the clustering result, as in Eq. 10 (Rousseeuw, 1987). The silhouette index ranges from −1 to 1, where a high value indicates better clustering accuracy.

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}.$$

$$\text{Silhouette} = \frac{1}{n} \sum_{i=1}^{n} s(i). \tag{22}$$

Silhouette analysis can be used for two purposes: 1) to check the validity of the obtained clustering solution and 2) to use along with a clustering algorithm to estimate the number of clusters in a dataset.

### 2.2.5 Adjusted Rank Index algorithms

Adjusted Rank Index (ARI) is the best-performing method for comparing two partitions. The ARI obtains a value of unity when the two partitions perfectly agree (in the current situation, when two nodes are connected to the same set of other nodes) and a value of zero when there is only chance agreement between the partitions. At face value, while the ARI might be known to outperform other methods of partition agreement, it still suffers from the problem of establishing an arbitrary threshold to decide if a pair of nodes should be linked or remain unlinked.

Fortunately, we can take advantage of a clever proof given by Warrens (2008) that demonstrates that the ARI is equivalent to Cohen's K (1960) when there are two partitions, and the cluster labels are known *a priori*. Both of those conditions are met as the two partitions are defined by the edge sets of nodes $n_i$ and $n_j$, while the labels are static across all nodes as either "connected" or "not connected," avoiding the problem of unknown cluster labels discussed in Steinley (2004). Under these two constraints, the ARI (and K) can be computed as

$$\text{ARI}_{ij} = \frac{2(ad - bc)}{(a+b)(b+d) + (a+c)(c+d)}. \tag{23}$$

In this application, the ARI gives each potential link a score between -1 and 1, where a score greater than 0 indicates that the probability of a link being there is greater than random chance. Using 0 as a defined cutoff value to defect links is an inherent benefit of ARI over other methods where the cutoff must be determined ahead of time.

This section successfully discusses the method formulation involving step-by-step systematic procedures of the IT2FVIKOR. Then, it follows with the unsupervised machine learning algorithms. Later, the next section discusses the evaluation of these algorithms with real application.

## 3 Experiment and results

This section relies on MCDM which is IT2FVIKOR and unsupervised ML algorithms, hierarchical clustering, autoencoder, and SOM. The schematic diagram of the proposed method is shown in (Figure 2). In phase 1, we used IT2FVIKOR to discover the best water security strategies using the data in the form of linguistic variables collected from DMs. All the human and non-human factors that are normally linked with water security strategies are considered for the best ranking. Then, in phase 2, using the data from the Department of Environment, Malaysia, we applied three different clustering algorithms along with silhouette analysis to obtain a list of clusters and the best optimal number of clusters. The ARI is used to compare the extent of agreement between the assignment of clusters for different algorithms. This method calculates the extent of agreement between two sets of groupings and outputs a representative score between 1 and −1. The elbow method is applied to double-confirm the best clusters for each clustering algorithm. Last, lists of polluted areas in each cluster are retrieved.

## 3.1 Application of water security using interval type-2 fuzzy VlseKriterijumska Optimizcija I Kompromisno Resenje

A real application for searching the best strategies to enhance water supply security is used to illustrate the procedures and feasibility of the IT2FVIKOR. Six alternatives ($A_1$, $A_2$, $A_3$, $A_4$, $A_5$, and $A_6$), namely, strengthening the protection of water source areas ($A_1$), improving infrastructure to safeguard urban and rural water security ($A_2$), developing a water-saving system ($A_3$), fully implementing the river chief system ($A_4$), reinforcing groundwater monitoring and protection ($A_5$), and strengthening the policy on water security ($A_6$), are proposed according to five criteria, namely, household water security ($C_1$), economic water security ($C_2$), urban water security ($C_3$), environmental water security ($C_4$), and resilience to water-related disasters ($C_5$) to select the most suitable alternative for enhance water security. The five decision makers ($DM_1$, $DM_2$, $DM_3$, $DM_4$, and $DM_5$) used the linguistic scales of restriction and reliability (from Table 1 and Table 2) to evaluate the rating of alternatives with respect to each criterion in the form of a decision matrix.

Next, the computational procedure is summarized as follows:

Step 1: Construction of a hierarchical structure of weight factors associated with the strategies to enhance water supply security.

The hierarchical structure of evaluating the best strategies to enhance water supply security in Malaysia is given in Figure 3.

The comparison results involve six strategies as the alternatives and five criteria based on water supply security. The rating of each alternative for each criterion is presented by DMs.
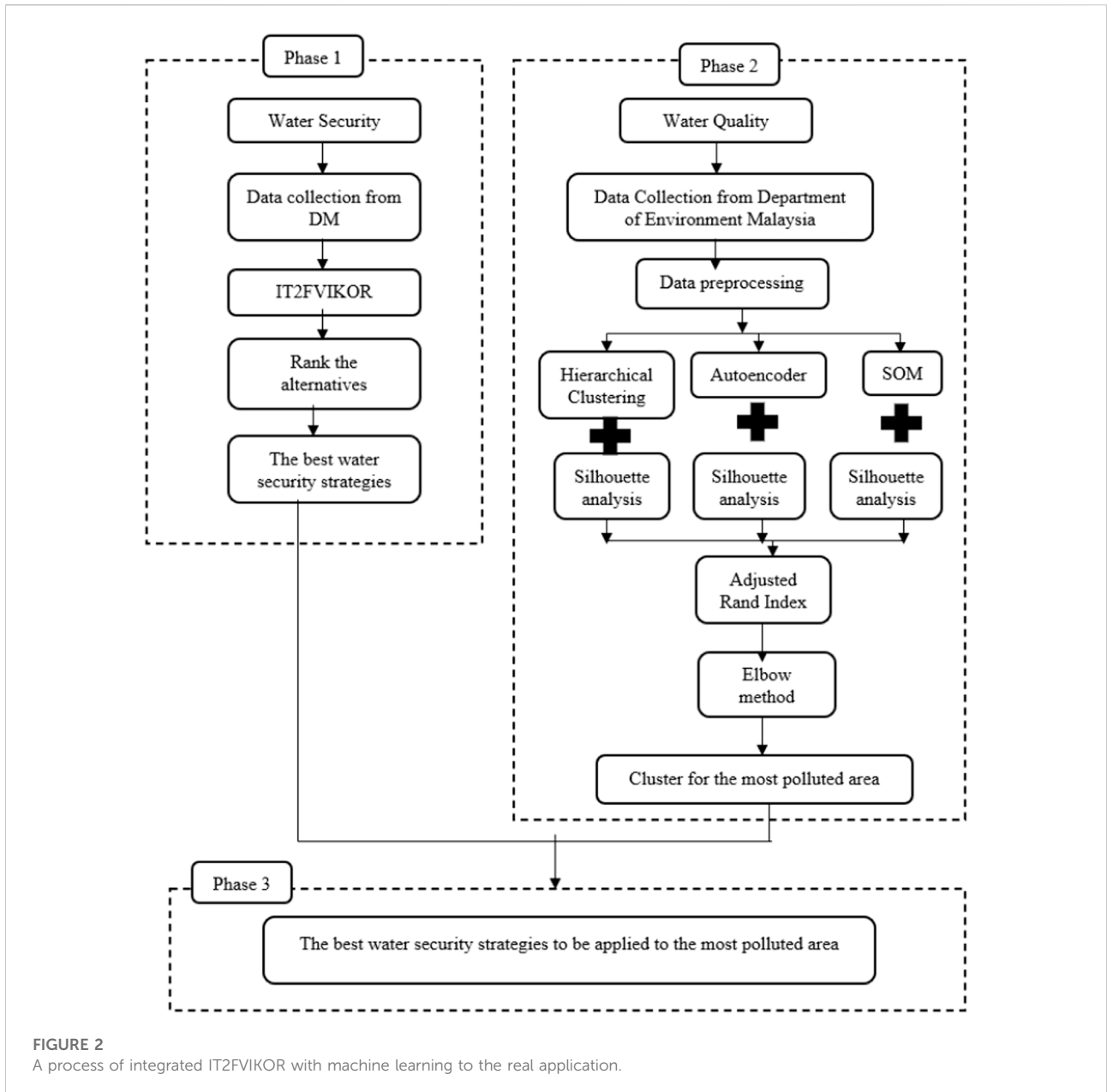
**FIGURE 2**
A process of integrated IT2FVIKOR with machine learning to the real application.

The data that consider negative and positive sides with restriction and reliability and its conversion (using Table 1 and Table 2) are referred in order to construct a matrix of attributes. Therefore, let us take an example of calculating the $\tilde{\tilde{f}}_{11}$. The average for $\tilde{\tilde{f}}_{11}$ ;

(VG, L) = ((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)),

((0.4, 0.5, 0.7, 0.8; 0.8, 0.8), (0.3, 0.5, 0.7, 0.9; 1, 1))

(G, SWL) = ((0.4, 0.5, 0.7, 0.8; 0.8, 0.8), (0.3, 0.5, 0.7, 0.9; 1, 1)),

((0.1, 0.2, 0.4, 0.5; 0.8, 0.8), (0, 0.2, 0.4, 0.6; 1, 1))

(VG, SLL) = ((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)),

((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1))

(VG, SLL) = ((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)),

((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1))

(VG, SLL) = ((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)),

((0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1))

Then, the average for (VG, L), (G, SWL), (VG, SLL), (VG, SLL), and (VG, SLL) is

((0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)).

((0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.9; 1, 1)).

**TABLE 1 Linguistic scale of restriction for the extended IT2FVIKOR.**

| Linguistic terms | Linguistic scale |
|---|---|
| Very poor (VP) | [(−1.0, −0.9, −0.8, −0.7; 0.8, 0.8), (−1.0, −1.0, −0.8, −0.6; 1, 1)] |
| Poor (P) | [(−0.8, −0.7, −0.5, −0.4; 0.8, 0.8), (−0.9, −0.7, −0.5, −0.3; 1, 1)] |
| Medium poor (MP) | [(−0.5, −0.4, −0.2, −0.1; 0.8, 0.8), (−0.6, −0.4, −0.2, 0; 1, 1)] |
| Medium (M) | [(−0.2, −0.1, 0.1, 0.2; 0.8, 0.8), (−0.3, −0.2, 0.2, 0.3; 1, 1)] |
| Medium good (MG) | [(0.1, 0.2, 0.4, 0.5; 0.8, 0.8), (0, 0.2, 0.4, 0.6; 1, 1)] |
| Good (G) | [(0.4, 0.5, 0.7, 0.8; 0.8, 0.8), (0.3, 0.5, 0.7, 0.9; 1, 1)] |
| Very good (VG) | [(0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)] |

**TABLE 2 Linguistic scale of reliability for the extended IT2FVIKOR.**

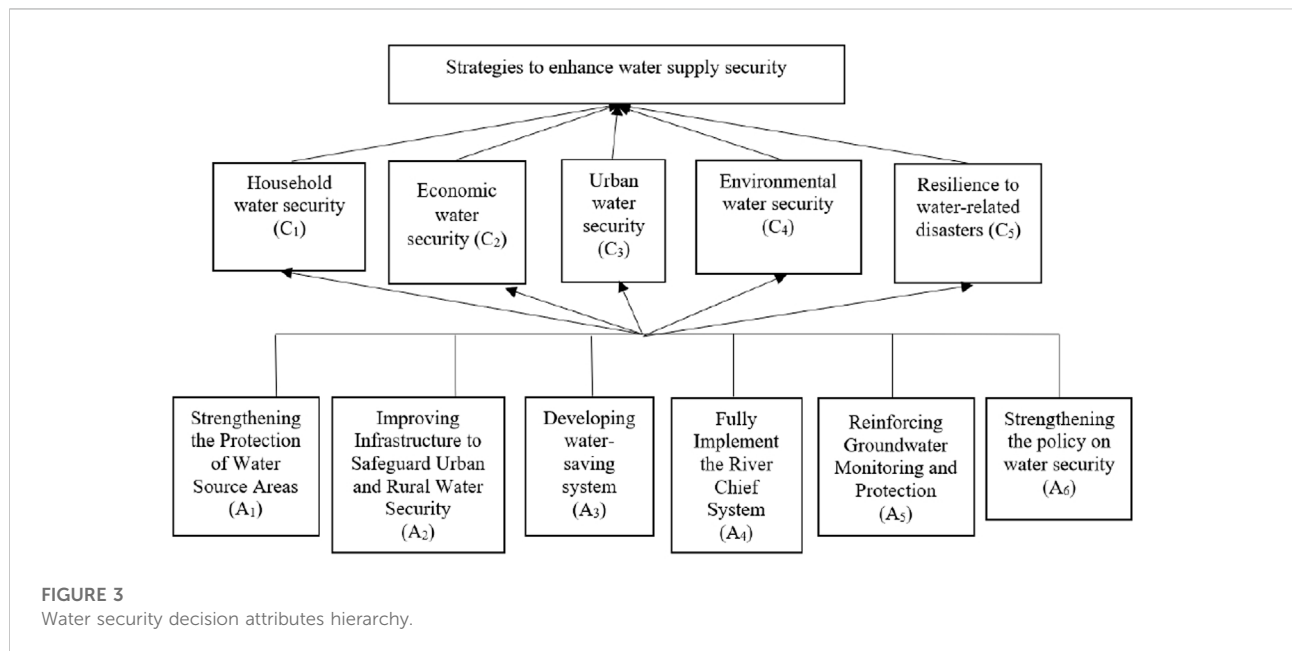| Linguistic terms | Linguistic scale |
|---|---|
| Strongly unlikely (SU) | [(−1.0, −0.9, −0.8, −0.7; 0.8, 0.8), (−1.0, −1.0, −0.8, −0.6; 1, 1)] |
| Unlikely (U) | [(−0.8, −0.7, −0.5, −0.4; 0.8, 0.8), (−0.9, −0.7, −0.5, −0.3; 1, 1)] |
| Somewhat unlikely (SWU) | [(−0.5, −0.4, −0.2, −0.1; 0.8, 0.8), (−0.6, −0.4, −0.2, 0; 1, 1)] |
| Neutral (N) | [(−0.2, −0.1, 0.1, 0.2; 0.8, 0.8), (−0.3, −0.2, 0.2, 0.3; 1, 1)] |
| Somewhat likely (SWL) | [(0.1, 0.2, 0.4, 0.5; 0.8, 0.8), (0, 0.2, 0.4, 0.6; 1, 1)] |
| Likely (L) | [(0.4, 0.5, 0.7, 0.8; 0.8, 0.8), (0.3, 0.5, 0.7, 0.9; 1, 1)] |
| Strongly likely (SLL) | [(0.7, 0.8, 1.0, 1.0; 0.8, 0.8), (0.6, 0.8, 1.0, 1.0; 1, 1)] |



FIGURE 3
Water security decision attributes hierarchy.

Apply the same calculation as $\tilde{\tilde{f}}_{11}$; thus, the whole results for the matrix of alternatives are summarized (Table 3) as follows:

Step 2: Construct the weighted DM matrix.

Use the fuzzy entropy with IT2FS formulas (Eq. 2) to calculate each of the entropy values in the decision matrix. Therefore, the entropy value for $E(\tilde{\tilde{A}}_{11})$ is represented as follows:

**TABLE 3 Judgement matrix.**

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $A_1$ | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.90; 1, 1)] | [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)], [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)] | [(0.46, 0.56, 0.76, 0.80; 0.8, 0.8) (0.36, 0.56, 0.76, 0.84; 1, 1)], [(0.40, 0.50, 0.70, 0.76; 0.8, 0.8) (0.30, 0.50, 0.70, 0.82; 1, 1)] |
| $A_2$ | [(0.40, 0.50, 0.70, 0.78; 0.8, 0.8) (0.30, 0.50, 0.70, 0.86; 1, 1)], [(0.40, 0.50, 0.70, 0.74; 0.8, 0.8) (0.30, 0.50, 0.70, 0.78; 1, 1)] | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)] | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.90; 1, 1)] | [(0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.90; 1, 1)], [(0.34, 0.44, 0.64, 0.70; 0.8, 0.8) (0.24, 0.42, 0.66, 0.76; 1, 1)] | [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)], [(0.52, 0.62, 0.82, 0.88; 0.8, 0.8) (0.42, 0.62, 0.82, 0.94; 1, 1)] |
| $A_3$ | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.90; 1, 1)] | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.90; 1, 1)], [(0.40, 0.50, 0.70, 0.76; 0.8, 0.8) (0.30, 0.50, 0.70, 0.82; 1, 1)] | [(0.40, 0.50, 0.70, 0.74; 0.8, 0.8) (0.30, 0.48, 0.72, 0.78; 1, 1)], [(0.28, 0.38, 0.58, 0.64; 0.8, 0.8) (0.18, 0.34, 0.62, 0.70; 1, 1)] |
| $A_4$ | [(0.34, 0.44, 0.64, 0.72; 0.8, 0.8) (0.24, 0.42, 0.66, 0.8; 1, 1)], [(0.04, 0.14, 0.34, 0.42; 0.8, 0.8) (−0.06, 0.14, 0.34, 0.50; 1, 1)] | [(0.24, 0.34, 0.52, 0.58; 0.8, 0.8) (0.16, 0.32, 0.52, 0.64; 1, 1)], [(0.52, 0.62, 0.82, 0.88; 0.8, 0.8) (0.42, 0.62, 0.82, 0.94; 1, 1)] | [(0.12, 0.22, 0.40, 0.46; 0.8, 0.8) (0.04, 0.20, 0.40, 0.52; 1, 1)], [(0.34, 0.44, 0.64, 0.70; 0.8, 0.8) (0.24, 0.42, 0.66, 0.76; 1, 1)] | [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)], [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)] | [(0.46, 0.56, 0.76, 0.82; 0.8, 0.8) (0.36, 0.56, 0.76, 0.88; 1, 1)], [(0.40, 0.50, 0.70, 0.76; 0.8, 0.8) (0.30, 0.50, 0.70, 0.82; 1, 1)] |
| $A_5$ | [(0.64, 0.72, 0.88, 0.88; 0.8, 0.8) (0.56, 0.72, 0.88, 0.88; 1, 1)], [(0.64, 0.72, 0.88, 0.88; 0.8, 0.8) (0.56, 0.72, 0.88, 0.88; 1, 1)] | [(0.64, 0.72, 0.88, 0.88; 0.8, 0.8) (0.56, 0.72, 0.88, 0.88; 1, 1)], [(0.64, 0.72, 0.88, 0.88; 0.8, 0.8) (0.56, 0.72, 0.88, 0.88; 1, 1)] | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)] | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)], [(0.58, 0.68, 0.88, 0.92; 0.8, 0.8) (0.48, 0.68, 0.88, 0.96; 1, 1)] |
| $A_6$ | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)] | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)] | [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)], [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)] | [(0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)], [(0.58, 0.68, 0.88, 0.90; 0.8, 0.8) (0.48, 0.68, 0.88, 0.92; 1, 1)] | [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)], [(0.70, 0.80, 1.00, 1.00; 0.8, 0.8) (0.60, 0.80, 1.00, 1.00; 1, 1)] |

$$E\left(\tilde{\tilde{A}}_{11}^{L}\right) = \frac{\sqrt{(1-0.64)^2 + (1-0.74)^2 + (1-0.94)^2 + (1-0.96)^2}}{\sqrt{(0-0.64)^2 + (0-0.74)^2 + (0-0.94)^2 + (0-0.96)^2}}$$
$$= 0.27$$

$$E\left(\tilde{\tilde{A}}_{11}^{U}\right) = \frac{\sqrt{(1-0.54)^2 + (1-0.74)^2 + (1-0.94)^2 + (1-0.98)^2}}{\sqrt{(0-0.54)^2 + (0-0.74)^2 + (0-0.94)^2 + (0-0.98)^2}}$$
$$= 0.32$$

$$zE\left(\tilde{z}\tilde{A}_{11}^{L}\right) = \frac{\sqrt{(1-0.52)^2 + (1-0.62)^2 + (1-0.82)^2 + (1-0.86)^2}}{\sqrt{(0-0.52)^2 + (0-0.62)^2 + (0-0.82)^2 + (0-0.86)^2}}$$
$$= 0.45$$

$$zE\left(\tilde{z}\tilde{A}_{11}^{U}\right) = \frac{\sqrt{(0-0.42)^2 + (1-0.62)^2 + (1-0.82)^2 + (1-0.9)^2}}{\sqrt{(1-0.42)^2 + (0-0.62)^2 + (0-0.82)^2 + (0-0.9)^2}}$$
$$= 0.41$$

$$E\left(\tilde{\tilde{A}}_{11}\right) = ((0.27; 0.8, 0.8)(0.32; 1, 1)), ((0.45; 0.8, 0.8)(0.41; 1, 1)).$$

Using a similar calculation, the entropy value for other attributes is calculated and listed in Table 4.

Step 3: Divide using the maximal entropy value.

All the entropy values are divided by using the maximal entropy value followed by Eq. 3.

$$E\left(\tilde{\tilde{A}}_{11}\right) = ((0.27; 0.8, 0.8)(0.32; 1, 1)), ((0.45; 0.8, 0.8)(0.41; 1, 1))$$

$$E\left(\tilde{\tilde{A}}_{12}\right) = ((0.20; 0.8, 0.8)(0.26; 1, 1)), ((0.27; 0.8, 0.8)(0.32; 1, 1))$$

$$E\left(\tilde{\tilde{A}}_{13}\right) = ((0.20; 0.8, 0.8)(0.26; 1, 1)), ((0.27; 0.8, 0.8)(0.32; 1, 1))$$

$$E\left(\tilde{\tilde{A}}_{14}\right) = ((0.20; 0.8, 0.8)(0.26; 1, 1)), ((0.35; 0.8, 0.8)(0.38; 1, 1))$$

$$E\left(\tilde{\tilde{A}}_{15}\right) = ((0.53; 0.8, 0.8)(0.45; 1, 1)), ((0.58; 0.8, 0.8)(0.49; 1, 1)).$$

Let us take $E\left(\tilde{\tilde{A}}_{1j}\right)$ as an example on calculating the maximal entropy value, where the maximal value for $E\left(\tilde{\tilde{A}}_{1j}\right)$ is

max $\left(E\left(\tilde{\tilde{A}}_{15}\right)\right) = [(0.53; 0.8, 0.8) (0.45; 1, 1)], [(0.58; 0.8, 0.8) (0.49; 1, 1)].$

Thus,

$$\tilde{\tilde{h}}_{11} = \left[\left(\frac{\tilde{\tilde{E}}(\tilde{\tilde{A}}_{11}^{L})}{\max(\tilde{\tilde{E}}(\tilde{\tilde{A}}_{15}^{L}))}\right), \left(\frac{\tilde{\tilde{E}}(\tilde{\tilde{A}}_{11}^{U})}{\max(\tilde{\tilde{E}}(\tilde{\tilde{A}}_{15}^{U}))}\right)\right], \left[\left(\frac{\tilde{\tilde{E}}(\tilde{\tilde{Z}}_{11}^{L})}{\max(\tilde{\tilde{E}}(\tilde{\tilde{Z}}_{15}^{L}))}\right), \left(\frac{\tilde{\tilde{E}}(\tilde{\tilde{Z}}_{11}^{U})}{\max(\tilde{\tilde{E}}(\tilde{\tilde{Z}}_{15}^{U}))}\right)\right]$$

$$\tilde{\tilde{h}}_{11} = \left[ \left( \frac{0.27; 0.8, 0.8}{0.53; 0.8, 0.8} \right), \left( \frac{0.32; 1, 1}{0.45; 1, 1} \right) \right], \left[ \left( \frac{0.45; 0.8, 0.8}{0.58; 0.8, 0.8} \right), \left( \frac{0.41; 1, 1}{0.49; 1, 1} \right) \right]$$

$$\tilde{\tilde{h}}_{11} = ((0.52; 0.8, 0.8)(0.72; 1, 1)), (0.79; 0.8, 0.8)(0.82; 1, 1)).$$

Using the same calculation, the maximal entropy value is shown (Table 5) as follows:

Step 4: Weight of criteria.

Next, calculate the weight of the criteria by using the weight formula from Eqs. 5, 6. The whole entropy-based weights are listed as follows:

$$\tilde{\tilde{w}}_1 = ((0.04; 0.8, 0.8)(0.02; 1, 1)), ((0.03; 0.8, 0.8)(0.02; 1, 1))$$

$$\tilde{\tilde{w}}_2 = ((0.09; 0.8, 0.8)(0.04; 1, 1)), ((0.08; 0.8, 0.8)(0.05; 1, 1))$$

$$\tilde{\tilde{w}}_3 = ((0.07; 0.8, 0.8)(0.04; 1, 1)), ((0.05; 0.8, 0.8)(0.03; 1, 1))$$

$$\tilde{\tilde{w}}_4 = ((0.09; 0.8, 0.8)(0.06; 1, 1)), ((0.04; 0.8, 0.8)(0.02; 1, 1))$$

$$\tilde{\tilde{w}}_5 = ((0.03; 0.8, 0.8)(0.02; 1, 1)), ((0.02; 0.8, 0.8)(0.02; 1, 1)).$$

Step 5: Construct the weighted decision matrices.

The weighted DM matrix with respect to aggregated matrix comparison of each criterion and alternatives is constructed by using Eq. 7.

For example, let the value of the weighted matrix for $\tilde{\tilde{v}}_{11}$ be

TABLE 4 IT2 fuzzy entropy.

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $A_1$ | [(0.27; 0.8, 0.8) | [(0.20; 0.8, 0.8) | [(0.20; 0.8, 0.8) | [(0.20; 0.8, 0.8) | [(0.53; 0.8, 0.8) |
| | (0.32; 1, 1)] | (0.26; 1, 1)] | (0.26; 1, 1)] | (0.26; 1, 1)] | (0.45; 1, 1)] |
| | [(0.45; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.35; 0.8, 0.8) | [(0.58; 0.8, 0.8) |
| | (0.41; 1, 1)] | (0.32; 1, 1)] | (0.32; 1, 1)] | (0.38; 1, 1)] | (0.49; 1, 1)] |
| $A_2$ | [(0.57; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.45; 0.8, 0.8) | [(0.59; 0.8, 0.8) |
| | (0.48; 1, 1)] | (0.32; 1, 1)] | (0.32; 1, 1)] | (0.41; 1, 1)] | (0.49; 1, 1)] |
| | [(0.59; 0.8, 0.8) | [(0.35; 0.8, 0.8) | [(0.45; 0.8, 0.8) | [(0.58; 0.8, 0.8) | [(0.57; 0.8, 0.8) |
| | (0.51; 1, 1)] | (0.38; 1, 1)] | (0.41; 1, 1)] | (0.49; 1, 1)] | (0.44; 1, 1)] |
| $A_3$ | [(0.36; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.45; 0.8, 0.8) | [(0.59; 0.8, 0.8) |
| | (0.39; 1, 1)] | (0.39; 1, 1)] | (0.39; 1, 1)] | (0.41; 1, 1)] | (0.49; 1, 1)] |
| | [(0.45; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.58; 0.8, 0.8) | [(0.57; 0.8, 0.8) |
| | (0.41; 1, 1)] | (0.32; 1, 1)] | (0.32; 1, 1)] | (0.49; 1, 1)] | (0.44; 1, 1)] |
| $A_4$ | [(0.56; 0.8, 0.8) | [(0.60; 0.8, 0.8) | [(0.46; 0.8, 0.8) | [(0.20; 0.8, 0.8) | [(0.52; 0.8, 0.8) |
| | (0.44; 1, 1)] | (0.51; 1, 1)] | (0.44; 1, 1)] | (0.26; 1, 1)] | (0.44; 1, 1)] |
| | [(0.36; 0.8, 0.8) | [(0.45; 0.8, 0.8) | [(0.57; 0.8, 0.8) | [(0.35; 0.8, 0.8) | [(0.58; 0.8, 0.8) |
| | (0.39; 1, 1)] | (0.40; 1, 1)] | (0.46; 1, 1)] | (0.38; 1, 1)] | (0.49; 1, 1)] |
| $A_5$ | [(0.31; 0.8, 0.8) | [(0.31; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.35; 0.8, 0.8) |
| | (0.36; 1, 1)] | (0.36; 1, 1)] | (0.39; 1, 1)] | (0.32; 1, 1)] | (0.38; 1, 1)] |
| | [(0.31; 0.8, 0.8) | [(0.31; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.35; 0.8, 0.8) |
| | (0.36; 1, 1)] | (0.36; 1, 1)] | (0.39; 1, 1)] | (0.32; 1, 1)] | (0.38; 1, 1)] |
| $A_6$ | [(0.36; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.20; 0.8, 0.8) |
| | (0.39; 1, 1)] | (0.32; 1, 1)] | (0.39; 1, 1)] | (0.32,1,1)] | (0.26; 1, 1)] |
| | [(0.36; 0.8, 0.8) | [(0.27; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.36; 0.8, 0.8) | [(0.20; 0.8, 0.8) |
| | (0.39; 1, 1)] | (0.32; 1, 1)] | (0.39; 1, 1)] | (0.39; 1, 1)] | (0.26; 1, 1)] |

**TABLE 5 Maximal entropy value.**

|  | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $A_1$ | [(0.52; 0.8, 0.8) | [(0.39; 0.8, 0.8) | [(0.39; 0.8, 0.8) | [(0.39; 0.8, 0.8) | [(1.00; 0.8, 0.8) |
|  | (0.72; 1, 1)] | (0.57; 1, 1)] | (0.57; 1, 1)] | (0.57; 1, 1)] | (1.00; 1, 1)] |
|  | [(0.79; 0.8, 0.8) | [(0.47; 0.8, 0.8) | [(0.47; 0.8, 0.8) | [(0.61; 0.8, 0.8) | [(1.00; 0.8, 0.8) |
|  | (0.82; 1, 1)] | (0.66; 1, 1)] | (0.66; 1, 1)] | (0.77; 1, 1)] | (1.00; 1, 1)] |
| $A_2$ | [(1.00; 0.8, 0.8) | [(0.48; 0.8, 0.8) | [(0.48; 0.8, 0.8) | [(0.80; 0.8, 0.8) | [(0.62; 0.8, 0.8) |
|  | (1.00; 1, 1)] | (0.68; 1, 1)] | (0.68; 1, 1)] | (0.85; 1, 1)] | (0.79; 1, 1)] |
|  | [(1.00; 0.8, 0.8) | [(0.60; 0.8, 0.8) | [(0.77; 0.8, 0.8) | [(0.96; 0.8, 0.8) | [(0.76; 0.8, 0.8) |
|  | (1.00; 1, 1)] | (0.74; 1, 1)] | (0.80; 1, 1)] | (0.90; 1, 1)] | (0.78; 1, 1)] |
| $A_3$ | [(0.61; 0.8, 0.8) | [(0.61; 0.8, 0.8) | [(0.61; 0.8, 0.8) | [(0.77; 0.8, 0.8) | [(1.00; 0.8, 0.8) |
|  | (0.80; 1, 1)] | (0.80; 1, 1)] | (0.80; 1, 1)] | (0.83; 1, 1)] | (1.00; 1, 1) |
|  | [(0.79; 0.8, 0.8) | [(0.47; 0.8, 0.8) | [(0.47; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.98; 0.8, 0.8) |
|  | (0.82; 1, 1)] | (0.66; 1, 1)] | (0.66; 1, 1)] | (1.00; 1, 1)] | (0.89; 1, 1)] |
| $A_4$ | [(0.93; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.77; 0.8, 0.8) | [(0.34; 0.8, 0.8) | [(0.86; 0.8, 0.8) |
|  | (0.86; 1, 1)] | (1.00; 1, 1)] | (0.86; 1, 1)] | (0.50; 1, 1)] | (0.85; 1, 1)] |
|  | [(0.62; 0.8, 0.8) | [(0.78; 0.8, 0.8) | [(0.98; 0.8, 0.8) | [(0.61; 0.8, 0.8) | [(1.00; 0.8, 0.8) |
|  | (0.79; 1, 1)] | (0.80; 1, 1)] | (0.93; 1, 1)] | (0.77; 1, 1)] | (1.00; 1, 1)] |
| $A_5$ | [(0.87; 0.8, 0.8) | [(0.87; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.76; 0.8, 0.8) | [(0.99; 0.8, 0.8) |
|  | (0.92; 1, 1)] | (0.92; 1, 1)] | (1.00; 1, 1)] | [((0.84; 1, 1)] | (0.98; 1, 1)] |
|  | [(0.87; 0.8, 0.8) | [(0.87; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.76; 0.8, 0.8) | [(0.99; 0.8, 0.8) |
|  | (0.92; 1, 1)] | (0.92; 1, 1)] | (1.00; 1, 1)] | [((0.84; 1, 1)] | (0.98; 1, 1)] |
| $A_6$ | [(1.00; 0.8, 0.8) | [(0.76; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.76; 0.8, 0.8) | [(0.57; 0.8, 0.8) |
|  | (1.00; 1, 1)] | (0.84; 1, 1)] | (1.00; 1, 1)] | (0.84; 1, 1)] | (0.67; 1, 1)] |
|  | [(1.00; 0.8, 0.8) | [(0.76; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(1.00; 0.8, 0.8) | [(0.57; 0.8, 0.8) |
|  | (1.00; 1, 1)] | (0.84; 1, 1)] | (1.00; 1, 1)] | (1.00; 1, 1)] | (0.67; 1, 1)] |

$$\tilde{\tilde{v}}_{11} = \tilde{\tilde{f}}_{11} \times \tilde{w}_1$$

$\tilde{\tilde{v}}_{11} = ( (0.64, 0.74, 0.94, 0.96; 0.8, 0.8) (0.54, 0.74, 0.94, 0.98; 1, 1)),$
$\quad ( (0.52, 0.62, 0.82, 0.86; 0.8, 0.8) (0.42, 0.62, 0.82, 0.9; 1, 1))$
$\quad \times ( (0.04; 0.8, 0.8) (0.02; 1, 1)), [ (0.03; 0.8, 0.8) (0.02; 1, 1) ]$
$= ( (0.02, 0.03, 0.03, 0.04; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1; 1)),$
$\quad ( (0.02, 0.02, 0.02, 0.03; 0.8, 0.8) (0.08, 0.01, 0.02, 0.02; 1, 1)).$

Thus, the remaining values of weighted DM's matrix are shown (Table 6) as follows:

Step 6: Construct the fuzzy best value (FBV) and fuzzy worst value (FWV).

Values for FBV and FWV are chosen using Eq. 8 and combined (Table 7) as follows:

Step 7: Compute the separation measures and defuzzification.

The utility measure $\tilde{S}_i$ and regret measure $\tilde{R}_i$ are calculated using the sum of the FBV distance with respect to all criteria based on Eq. 9 till Eq. 11. The result of calculation is listed in Table 8.

Step 8: Defuzzification of the utility measure value ($\tilde{S}_i$) and regret measure value ($\tilde{R}_i$)

This defuzzification step is needed to further proceed with the next calculation in Step 9. Therefore, each utility measure value ($\tilde{S}_i$) and regret measure value ($\tilde{R}_i$) from Table 8 is defuzzified using Eq. 12. Let us take the example for $\tilde{S}_1$ upper boundary:

$$\tilde{S}_1 = \frac{\frac{(0.05 - 0.07) + (0.8 \times (0.05 - 0.07)) + (0.8 \times (0.06 - 0.07))}{4} + 0.07 + \left[ \frac{(0.04 - 0.05) + (1 \times (0.03 - 0.05)) + (1 \times (0.04 - 0.05))}{4} + 0.05 \right]}{2}$$

$$\tilde{S}1 = 0.05.$$

**TABLE 6 Weighted decision matrix.**

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $A_1$ | [(0.02, 0.03, 0.03, 0.04; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.06, 0.06, 0.08, 0.08; 0.8, 0.8) (0.03, 0.04, 0.05, 0.05; 1, 1)] | [(0.05, 0.06, 0.07, 0.07; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.03, 0.03, 0.04, 0.04; 0.8, 0.8) (0.02, 0.02, 0.03, 0.03; 1, 1)] | [(0.07, 0.08, 0.09, 0.09; 0.8, 0.8) (0.04, 0.05, 0.06, 0.06; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)] | [(0.01, 0.02, 0.02, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)], [(0.01, 0.01, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $A_2$ | [(0.01, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)], [(0.01, 0.02, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.01, 0.02; 1, 1)] | [(0.05, 0.06, 0.07, 0.07; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) (0.02, 0.03, 0.04, 0.05; 1, 1)] | [(0.04, 0.05, 0.06, 0.07; 0.8, 0.8) (0.02, 0.03, 0.03, 0.04; 1, 1)], [(0.02, 0.03, 0.04, 0.04; 0.8, 0.8) (0.01, 0.02, 0.03, 0.03; 1, 1)] | [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) (0.02, 0.04, 0.05, 0.05; 1, 1)], [(0.01, 0.02, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.01, 0.01, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $A_3$ | [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.04, 0.05, 0.07, 0.07; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.06, 0.06, 0.08, 0.08; 0.8, 0.8) (0.03, 0.04, 0.05, 0.05; 1, 1)] | [(0.04, 0.05, 0.06, 0.07; 0.8, 0.8) (0.02, 0.03, 0.03, 0.03; 1, 1)], [(0.03, 0.03, 0.04, 0.04; 0.8, 0.8) (0.01, 0.02, 0.03, 0.03; 1, 1)] | [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) (0.02, 0.04, 0.05, 0.05; 1, 1)], [(0.01, 0.02, 0.02, 0.03; 0.8, 0.8) (0.01, 0.01, 0.01, 0.02; 1, 1)] | [(0.01, 0.02, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.03; 1, 1)], [(0.01, 0.01, 0.01, 0.01; 0.8, 0.8) (0.00, 0.01, 0.01, 0.02; 1, 1)] |
| $A_4$ | [(0.01, 0.02, 0.02, 0.03; 0.8, 0.8) (0.01, 0.01, 0.01, 0.02; 1, 1)], [(0.00, 0.00, 0.01, 0.01; 0.8, 0.8) (0.00, 0.00, 0.01, 0.01; 1, 1)] | [(0.02, 0.03, 0.04, 0.04; 0.8, 0.8) (0.01, 0.01, 0.02, 0.03; 1, 1)], [(0.05, 0.05, 0.07, 0.08; 0,8,0.8) (0.02, 0.03, 0.04, 0.05; 1, 1)] | [(0.01, 0.02, 0.03, 0.03; 0.8, 0.8) (0.00, 0.01, 0.01, 0.02; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.07, 0.08, 0.09, 0.09; 0.8, 0.8) (0.04, 0.05, 0.06, 0.06; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)] | [(0.01, 0.02, 0.02, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)], [(0.01, 0.01, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $A_5$ | [(0.02, 0.03, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.05, 0.06, 0.07, 0.07; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.06, 0.06, 0.08, 0.08; 0.8, 0.8) (0.03, 0.04, 0.04, 0.04; 1, 1)] | [(0.04, 0.05, 0.06, 0.06; 0.8, 0.8) (0.02, 0.03, 0.03, 0.03; 1, 1)], [(0.03, 0.03, 0.04, 0.04; 0.8, 0.8) (0.02, 0.02, 0.03, 0.03; 1, 1)] | [(0.06, 0.07, 0.09, 0.09; 0.8, 0.8) (0.03, 0.04, 0.06, 0.06; 1, 1)], [(0.02, 0.03, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)] | [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.01, 0.01, 0.02, 0.02; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $A_6$ | [(0.02, 0.02, 0.03, 0.03; 0,8,0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.01, 0.02, 0.02; 1, 1)] | [(0.05, 0.06, 0.07, 0.07; 0.8, 0.8) (0.02, 0.03, 0.04, 0.04; 1, 1)], [(0.06, 0.06, 0.08, 0.08,8,0.8) (0.03, 0.04, 0.05, 0.05; 1, 1)] | [(0.04, 0.05, 0.06, 0.06; 0.8, 0.8) (0.02, 0.03, 0.03, 0.03; 1, 1)], [(0.03, 0.03, 0.04, 0.04; 0.8, 0.8) (0.02, 0.02, 0.03, 0.03; 1, 1)] | [(0.06, 0.07, 0.09, 0.09; 0.8, 0.8) (0.03, 0.04, 0.06, 0.06; 1, 1)], [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)] | [(0.02, 0.03, 0.03, 0.03; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)], [(0.02, 0.02, 0.02, 0.02; 0.8, 0.8) (0.01, 0.02, 0.02, 0.02; 1, 1)] |

Thus, the remaining defuzzification results for $\tilde{S}_i$ and $\tilde{R}_i$ are listed in Table 9:

Step 9: Rank the alternatives.

Here, we listed all the values for 1) the maximum group of utility $\tilde{S}^*$, 2) the minimum value for the maximum group of utility $\tilde{S}$, 3) the minimum individual regret of the opponent $\tilde{R}^*$, and 4) the minimum value $\tilde{R}$ using Eq. 15.

$$\tilde{S}^* = (0)(0)$$
$$\tilde{S}^- = (0.12)(0.09)$$
$$\tilde{R}^* = (0)(0)$$
$$\tilde{R}^- = (0.07)(0.06)$$

The final indexes for both $\tilde{S}^*$ and $\tilde{R}^*$, which is $Q_i$, are $A_1 = 0.38$, $A_2 = 0.88$, $A_3 = 0.63$, $A_4 = 0.70$, $A_5 = 0.56$, and $A_6 = 0$. That is, the best alternative is $A_6$, and the ranking order of the alternatives is $A_6 > A_1 > A_5 > A_3 > A_4 > A_2$, which is listed as in Table 10.

Thus, after taking into account the five criteria, six alternatives, and the opinion from the five DMs, a single measurement for the best strategies to enhance water supply security in Malaysia is obtained. The results showed that the first rank is strengthening the policy on water security at 0, followed by strengthening the protection of water source areas, reinforcing groundwater monitoring and protection, developing a water-saving system, and fully implementing the river chief system. Improving infrastructure to safeguard urban and rural water security is ranked last. This ranking proved that the IT2FVIKOR method can carry out a better solution in visualizing the preferences from the group of DMs and can measure the best strategies that contribute to the success of water security. Next, this study was further continued with the identification of the polluted areas to implement the water security strategies. Usually, studies on IT2FVIKOR only focus on the analysis of various available choices in certain situations or research areas to find the most preferred alternatives (our target or objectives) and present them in ranking order. If the results from IT2FVIKOR can be addressed for a specific polluted area, it will bring huge benefits to public health, economic growth, and environmental sustainability, and at the same time reduce disaster risk. The identification of specific polluted areas is a very important issue and needs intelligent techniques that are more convincing. Therefore, this paper offers further investigation on IT2FVIKOR problems with integrated ML techniques.

**TABLE 7 Fuzzy best value and fuzzy worst value.**

|  | Fuzzy best value | Fuzzy worst value |
|---|---|---|
| $C_1$ | [(0.02, 0.03, 0.03, 0.04; 0.8, 0.8) | [(0.01, 0.02, 0.02, 0.03; 0.8, 0.8) |
|  | (0.01, 0.02, 0.02, 0.02; 1, 1)] | (0.01, 0.01, 0.01, 0.02; 1, 1)] |
|  | [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) | [(0.00, 0.00, 0.01, 0.01; 0.8, 0.8) |
|  | (0.01, 0.01, 0.02, 0.02; 1, 1)] | (−0.00, 0.00, 0.01, 0.01; 1, 1)] |
| $C_2$ | [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) | [(0.02, 0.03, 0.04, 0.04; 0.8, 0.8) |
|  | (0.02, 0.03, 0.04, 0.04; 1, 1)] | (0.01, 0.01, 0.02, 0.03; 1, 1)] |
|  | [(0.06, 0.06, 0.08, 0.08; 0.8, 0.8) | [(0.05, 0.05, 0.07, 0.08; 0.8, 0.8) |
|  | (0.03, 0.04, 0.05, 0.05; 1, 1)] | (0.02, 0.03, 0.04, 0.04; 1, 1)] |
| $C_3$ | [(0.05, 0.06, 0.07, 0.07; 0.8, 0.8) | [(0.01, 0.02, 0.03, 0.03; 0.8, 0.8) |
|  | (0.02, 0.03, 0.04, 0.04; 1, 1)] | (0.00, 0.01, 0.01, 0.02; 1, 1)] |
|  | [(0.03, 0.03, 0.04, 0.04; 0.8, 0.8) | [(0.02, 0.02, 0.03, 0.03; 0.8, 0.8) |
|  | (0.02, 0.02, 0.03, 0.03; 1, 1)] | (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $C_4$ | [(0.07, 0.08, 0.09, 0.09; 0.8, 0.8) | [(0.05, 0.06, 0.08, 0.08; 0.8, 0.8) |
|  | (0.04, 0.05, 0.06, 0.06; 1, 1)] | (0.02, 0.04, 0.05, 0.05; 1, 1)] |
|  | [(0.02, 0.03, 0.03, 0.03; 0.8, 0.8) | [(0.01, 0.02, 0.02, 0.02; 0.8, 0.8) |
|  | (0.01, 0.02, 0.02, 0.02; 1, 1)] | (0.01, 0.01, 0.02, 0.02; 1, 1)] |
| $C_5$ | [(0.02, 0.03, 0.03, 0.03; 0.8, 0.8) | [(0.01, 0.02, 0.02, 0.02; 0.8, 0.8) |
|  | (0.01, 0.02, 0.02, 0.02; 1, 1)] | (0.01, 0.01, 0.02, 0.02; 1, 1)] |
|  | [(0.02, 0.02, 0.02, 0.02; 0.8, 0.8) | [(0.01, 0.01, 0.01, 0.01; 0.8, 0.8) |
|  | (0.01, 0.02, 0.02, 0.02; 1, 1)] | (0.00, 0.01, 0.01, 0.02; 1, 1)] |

**TABLE 8 $\tilde{S}_i$ and $\tilde{R}_i$.**

|  | $\tilde{S}_i$ | $\tilde{R}_i$ |
|---|---|---|
| $A_1$ | ((0.05,0.05,0.06,0.08;0.8,0.8) (0.04,0.03,0.04,0.05;1,1)), ((0.05,0.05,0.04,0.04;0.8,0.8) (0.04,0.03,0.03,0.02;1,1)) | ((0.04,0.04,0.04,0.04;0.8,0.8) (0.02,0.02,0.02;1,1)), ((0.03,0.03,0.03,0.03;0.8,0.8) (0.02,0.02,0.02,0.02;1,1)) |
| $A_2$ | ((0.11,0.12,0.13,0.13;0.8,0.8) (0.06,0.06,0.07,0.07;1,1)), ((0.09,0.10,0.13,0.17;0.8,0.8) (0.08,0.06,0.07,0.07;1,1)) | ((0.08,0.08,0.08,0.08;0.8,0.8) (0.04,0.04,0.04,0.04;1,1)), ((0.09,0.09,0.09,0.09;0.8,0.8) (0.05,0.05,0.05,0.05;1,1)) |
| $A_3$ | ((0.11,0.11,0.11,0.11;0.8,0.8) (0.06,0.06,0.06,0.06;1,1)), ((0.06,0.06,0.06,0.07;0.8,0.8) (0.04,0.04,0.05,0.05;1,1)) | ((0.07,0.07,0.07,0.07;0.8,0.8) (0.04,0.04,0.04,0.04;1,1)), ((0.05,0.05,0.05,0.05;0.8,0.8) (0.03,0.03,0.03,0.03;1,1)) |
| $A_4$ | ((0.16,0.16,0.16,0.15;0.8,0.8) (0.10,0.10,0.10,0.08;0.8,0.8)), ((0.06,0.06,0.06,0.05;0.8,0.8) (0.04,0.03,0.04,0.03;1,1)) | ((0.09,0.09,0.09,0.09;0.8,0.8) (0.06,0.06,0.06,0.06;1,1)), ((0.03,0.03,0.03,0.03;0.8,0.8) (0.02,0.02,0.02,0.02;1,1)) |
| $A_5$ | ((0.12,0.12,0.12,0.11;0.8,0.8) (0.09,0.08,0.09,0.07;1,1)), ((0.08,0.08,0.08,0.08;0.8,0.8) (0.07,0.07,0.08,0.06;1,1)) | ((0.03,0.03,0.03,0.03;0.8,0.8) (0.02,0.02,0.02,0.02;1,1)), ((0.02,0.02,0.02,0.03;0.8,0.8) (0.02,0.02,0.02,0.02;1,1)) |
| $A_6$ | ((0,0,0,0;0.8,0.8) (0,0,0,0;1,1)), ((0,0,0,0;0.8,0.8) (0,0,0,0;1,1)) | ((0,0,0,0;0.8,0.8) (0,0,0,0;1,1)), ((0,0,0,0;0.8,0.8) (0,0,0,0;1,1)) |

**TABLE 9 Defuzzified values of $\tilde{S}_i$ and $\tilde{R}_i$.**

|  | $\tilde{S}_i$ | $\tilde{R}_i$ |
|---|---|---|
| $A_1$ | (0.05) (0.04) | (0.03) (0.02) |
| $A_2$ | (0.09) (0.09) | (0.06) (0.06) |
| $A_3$ | (0.08) (0.05) | (0.05) (0.04) |
| $A_4$ | (0.12) (0.04) | (0.07) (0.02) |
| $A_5$ | (0.09) (0.07) | (0.03) (0.02) |
| $A_6$ | (0) (0) | (0) (0) |

**TABLE 10 Rating of $Q_i$ and rank of each alternative.**

| m | $\tilde{Q}_i$ | $Q_i$ | Rank |
|---|---|---|---|
| $A_1$ | (0.38) (0.38) | 0.38 | 2 |
| $A_2$ | (0.76) (1) | 0.88 | 6 |
| $A_3$ | (0.68) (0.57) | 0.63 | 4 |
| $A_4$ | (1) (0.40) | 0.70 | 5 |
| $A_5$ | (0.58) (0.54) | 0.56 | 3 |
| $A_6$ | (0) (0) | 0 | 1 |

## 3.2 Application of the pollution using machine learning

This section focuses on continuing the investigation on river water problems from the previous section and selecting a set of polluted areas based on the Terengganu River dataset to be implemented with the results of the water security strategies from Section 4. This section includes the description of the research location in Subsection 3.2.1, preprocessing of the datasets in Subsection 3.2.2, and applying the machine learning algorithms which includes three different unsupervised learning in Subsection 3.2.3.

### 3.2.1 Description of the research location

The Terengganu River originating from the Terengganu state, one of the states in Peninsular Malaysia, is chosen as the main research location for this study. This river flows through the Kenyir Dam (Empangan Tasik Kenyir) located in Hulu Terengganu and ends in Kuala Terengganu (downstream part) toward the South China Sea. This river basin is situated between a latitude of 4°40′N-5°20′N and a longitude of 102°30′E-

103°09′E in the North Eastern coastal region of Peninsular Malaysia. There are sixteen main tributaries of the Terengganu River Basin, with a total catchment area of about

**FIGURE 4**
Terengganu river map (Source: https://uniqorne.wordpress.com/2009/05/16/uniquely-marang/).
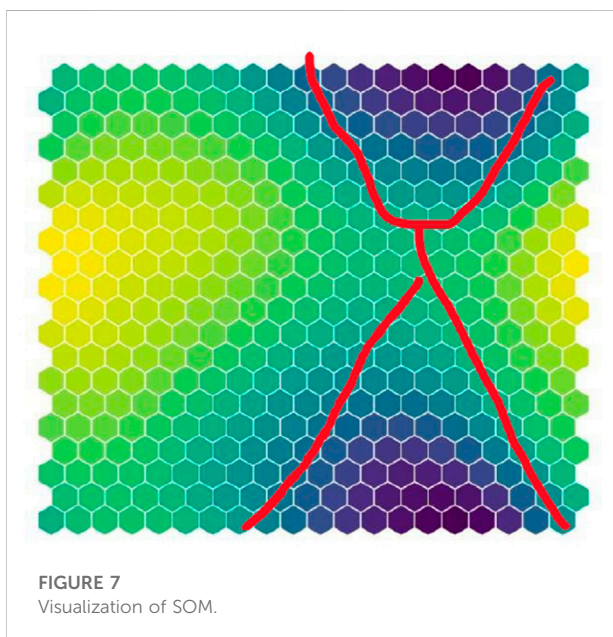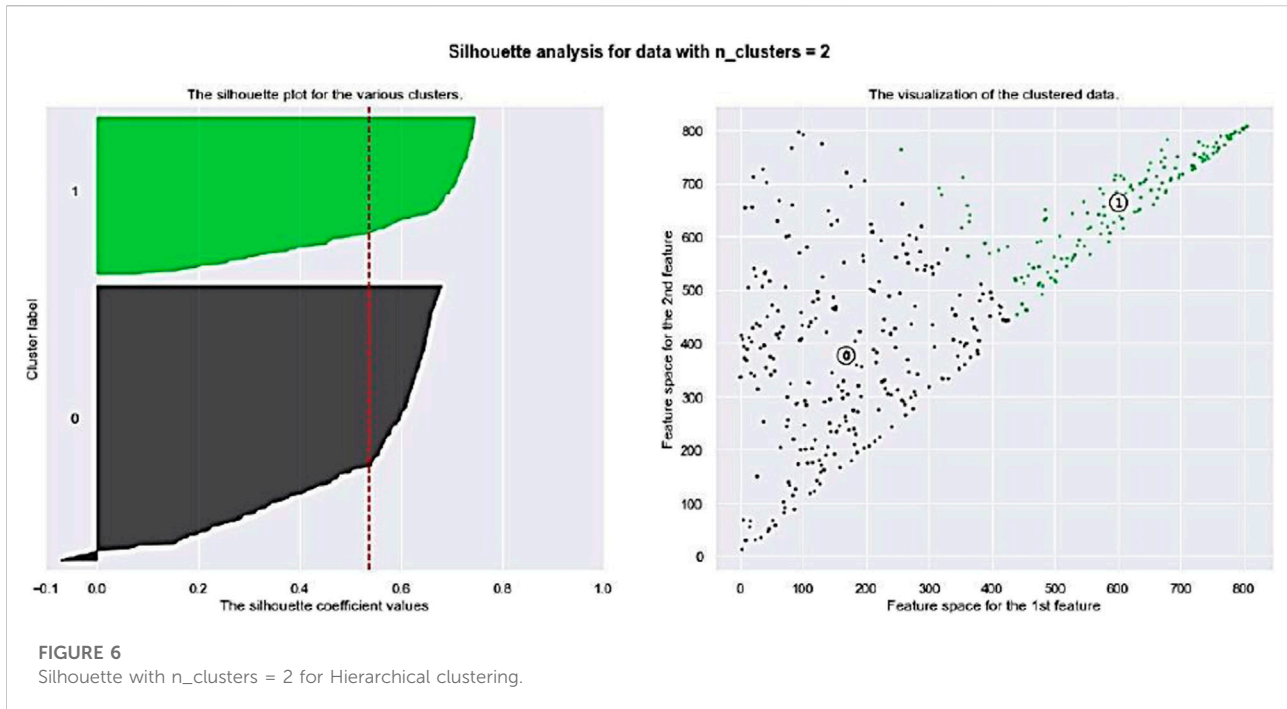
**FIGURE 5**
Hierarchical maps.
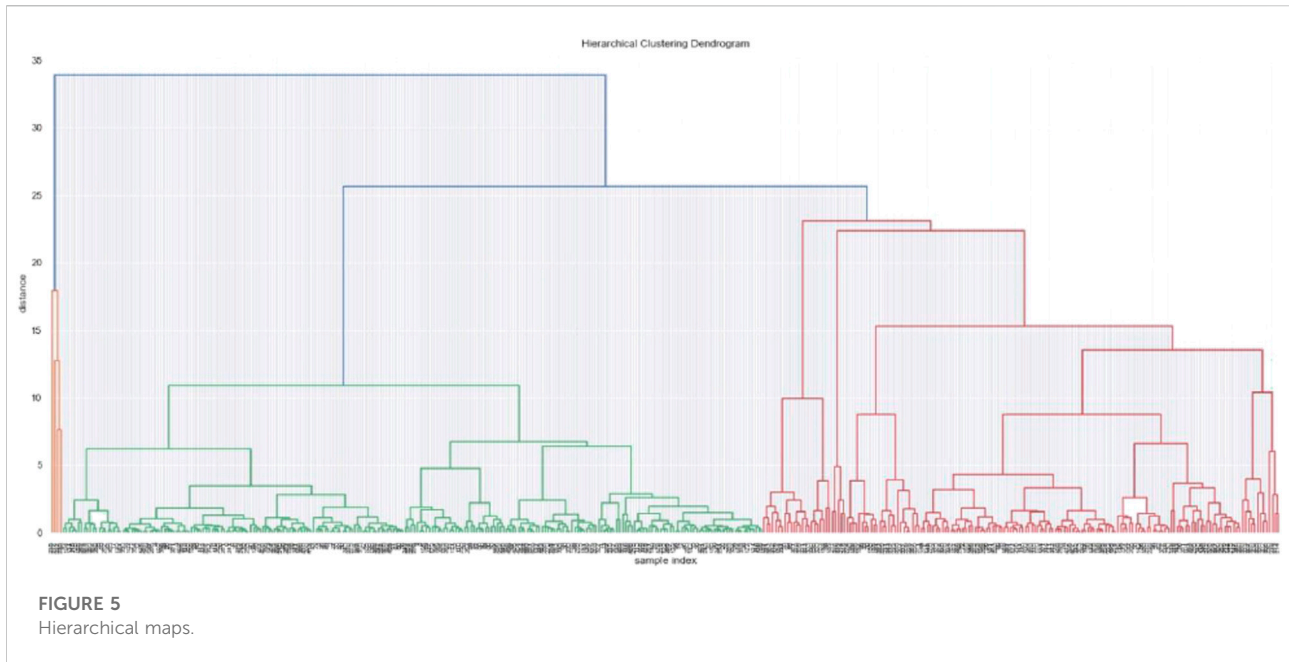
5,000 km² (11, 46, 47) (Wahab et al., 2019). The largest basin among the sixteen basins is the Nerus River Basin. All rivers pass through different socio-economic activity zones such as aquacultures, agricultural, commercial industries, farming, urban and rural settlements, tourism, reserves, and forests (Taman Negara). This study involves 14 main sampling stations with 24 water quality parameters, including 405 water samples. These 405 water samples were taken from all around the Terengganu River Basins, whereas the 24 water quality parameters include dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, ammonia nitrates ($NH_3$-N), temperature (TEMP), nitrogen nitrates ($NO_3$), chloride (CI), phosphate ($PO_4$), arsenic (As), Mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb), zinc (Zn), calcium (Ca), iron (Fe), potassium (K), sodium (Na), oganesson (OG), surfactant (MBAS), E-coli, and total coliform. Detailed description of the research location is summarized in Figure 4.

### 3.2.2 Preparing for datasets and tools

The discovery of novel groupings within a dataset requires unsupervised methods. This study used unsupervised ML, which does not require class labels associated with pre-defined groups to allocate targets to. This unsupervised ML can be achieved through data clustering. Before applying the clustering algorithms, these datasets went through a process called preprocessing. The 24 water quality parameters have a variety of data types with different scales. Direct usage of data without preprocessing leads to poor performance of the learning

**TABLE 11** $\tilde{S}_i$ and $\tilde{R}_i$.

| | $\tilde{S}_i$ | $\tilde{R}_i$ |
|---|---|---|
| $A_1$ | [(0.05, 0.05, 0.06, 0.08; 0.8, 0.8) | [(0.04, 0.04, 0.04, 0.04; 0.8, 0.8) |
| | (0.04, 0.03, 0.04, 0.05; 1, 1)] | (0.02, 0.02, 0.02; 1, 1)] |
| | [(0.05, 0.05, 0.04, 0.04; 0.8, 0.8) | [(0.03, 0.03, 0.03, 0.03; 0.8, 0.8) |
| | (0.04, 0.03, 0.03, 0.02; 1, 1)] | (0.02, 0.02, 0.02, 0.02; 1, 1)] |
| $A_2$ | [(0.11, 0.12, 0.13, 0.13; 0.8, 0.8) | [(0.08, 0.08, 0.08, 0.08; 0.8, 0.8) |
| | (0.06, 0.06, 0.07, 0.07; 1, 1)] | (0.04, 0.04, 0.04, 0.04; 1, 1)] |
| | [(0.09, 0.10, 0.13, 0.17; 0.8, 0.8) | [(0.09, 0.09, 0.09, 0.09; 0.8, 0.8) |
| | (0.08, 0.06, 0.07, 0.07; 1, 1)] | (0.05, 0.05, 0.05, 0.05; 1, 1)] |
| $A_3$ | [(0.11, 0.11, 0.11, 0.11; 0.8, 0.8) | [(0.07, 0.07, 0.07, 0.07; 0.8, 0.8) |
| | (0.06, 0.06, 0.06, 0.06; 1, 1)] | (0.04, 0.04, 0.04, 0.04; 1, 1)] |
| | [(0.06, 0.06, 0.06, 0.07; 0.8, 0.8) | [(0.05, 0.05, 0.05, 0.05; 0.8, 0.8) |
| | (0.04, 0.04, 0.05, 0.05; 1, 1)] | (0.03, 0.03, 0.03, 0.03; 1, 1)] |
| $A_4$ | [(0.16, 0.16, 0.16, 0.15; 0.8, 0.8) | [(0.09, 0.09, 0.09, 0.09; 0.8, 0.8) |
| | (0.10, 0.10, 0.10, 0.08; 0.8, 0.8)] | (0.06, 0.06, 0.06, 0.06; 1, 1)] |
| | [(0.06, 0.06, 0.06, 0.05; 0.8, 0.8) | [(0.03, 0.03, 0.03, 0.03; 0.8, 0.8) |
| | (0.04, 0.03, 0.04, 0.03; 1, 1)] | (0.02, 0.02, 0.02, 0.02; 1, 1)] |
| $A_5$ | [(0.12, 0.12, 0.12, 0.11; 0.8, 0.8) | [(0.03, 0.03, 0.03, 0.03; 0.8, 0.8) |
| | (0.09, 0.08, 0.09, 0.07; 1, 1)] | (0.02, 0.02, 0.02, 0.02; 1, 1)] |
| | [(0.08, 0.08, 0.08, 0.08; 0.8, 0.8) | [(0.02, 0.02, 0.02, 0.03; 0.8, 0.8) |
| | (0.07, 0.07, 0.08, 0.06; 1, 1)] | (0.02, 0.02, 0.02, 0.02; 1, 1)] |

**FIGURE 6**
Silhouette with n_clusters = 2 for Hierarchical clustering.



**FIGURE 7**
Visualization of SOM.

methods. In this study, we used the StandardScaler method to standardize all the parameters. The StandardScaler is used to standardize a feature by subtracting the mean and then scaling to unit variance. Unit variance means dividing all the values by the standard deviation.
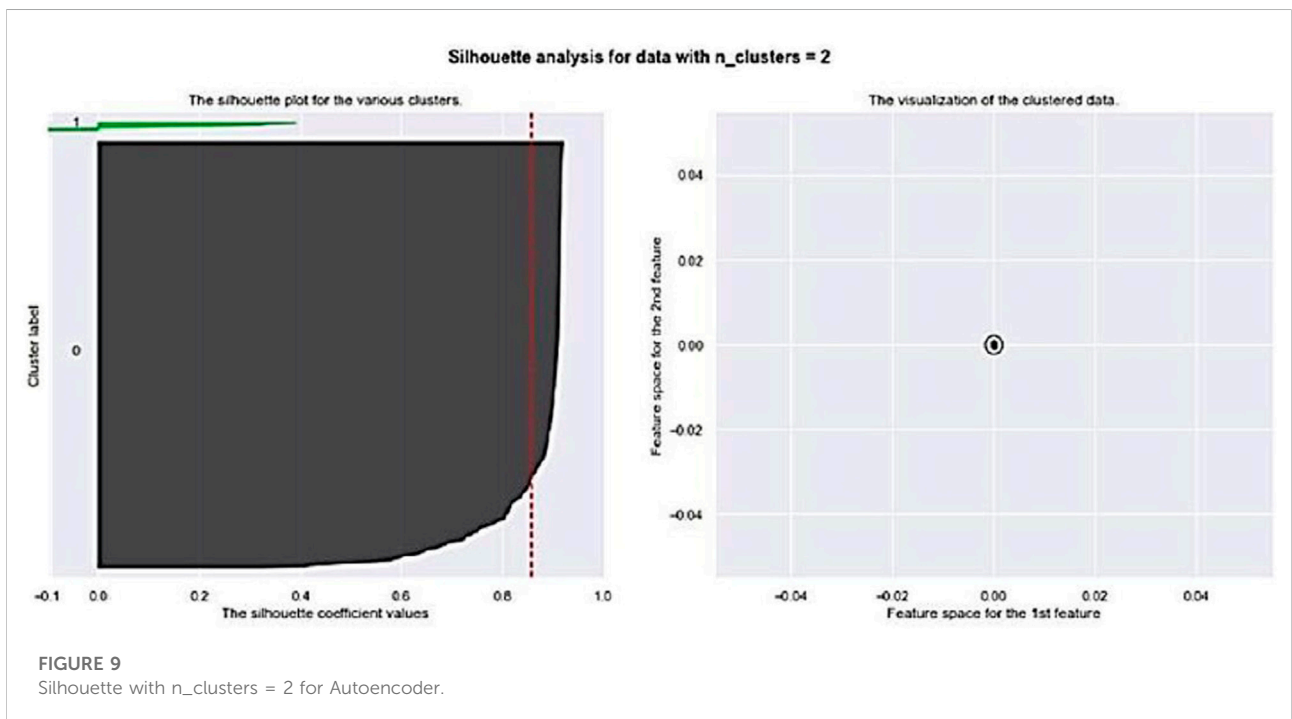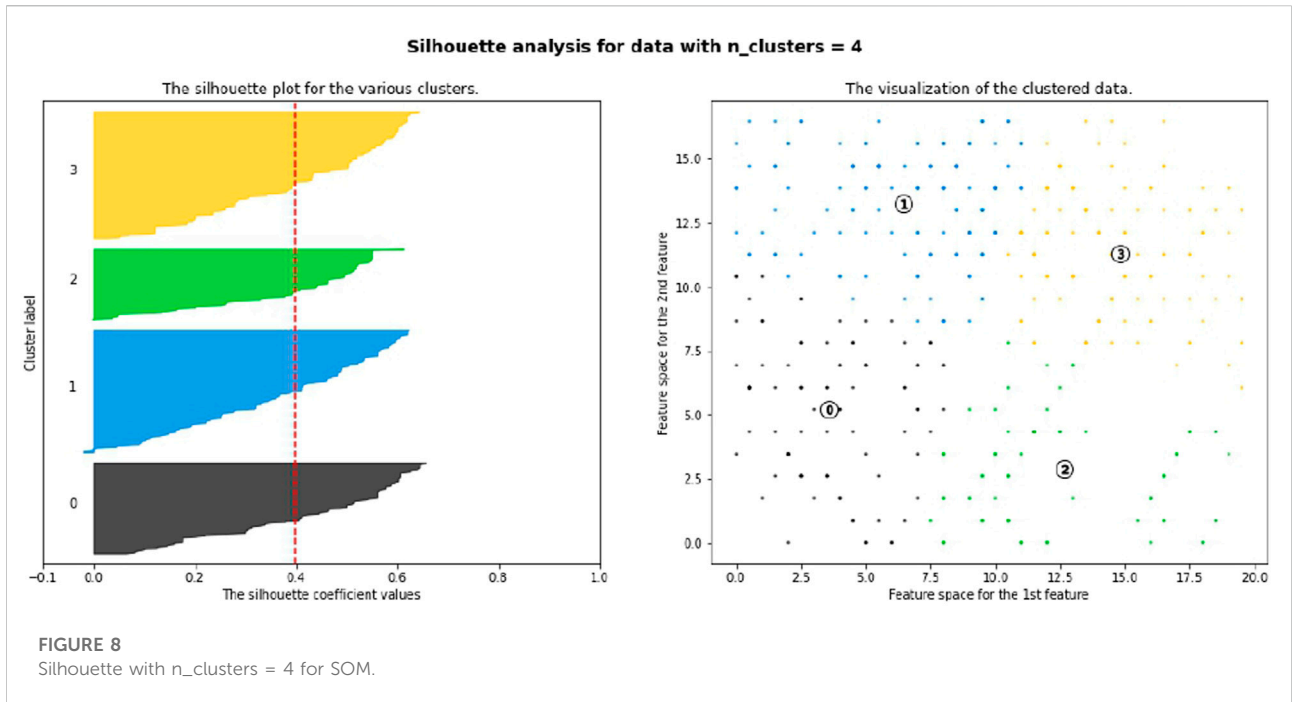
All the algorithms are executed using the Python programming language. ML algorithms and preprocessing

modules are executed using scikit-learn libraries (Pedregosa et al., 2011; Komer et al., 2014). Meanwhile, all the images are created using Matplotlib modules (Hunter, 2007).

### 3.2.3 The pollution area using machine learning

The purpose of this investigation is to identify clusters of polluted areas using the Terengganu River data. Our investigation is extended to the IT2FVIKOR, which discovers the best water security strategy and applies it to the polluted area clusters. After data preprocessing, the discovery of novel groupings within a dataset requires unsupervised algorithms. These do not require class labels associated with pre-defined groups to allocate items to. This can be achieved through data clustering. In this study, we focus on three different clustering algorithms which are hierarchical clustering, autoencoder, and SOMs. These three methods are the most popular unsupervised ML techniques that can solve different types of applications.

Hierarchical clustering methods are the simplest and most fundamental methods and famously play an important role in statistical data analysis (Johnson, 1967). They are relatively fast and easy to understand and implement (Zhu et al., 2022). The autoencoder has the ability to reduce the dimension of the input in a hierarchical way, leading to high-quality reconstructions of data (Tewari et al., 2017; Zhao et al., 2017). SOMs are frequently used to reveal complex correlations between samples that can be conveyed by a single map. It simplifies the analysis and enables multi-variate exploratory comparisons between samples *via* direct visual separation (Wongravee et al., 2020). Next, K-means was run in each of the clustering algorithms in line with the silhouette analysis to

**FIGURE 8**
Silhouette with n_clusters = 4 for SOM.



**FIGURE 9**
Silhouette with n_clusters = 2 for Autoencoder.

find the optimal number of clusters. The silhouette method measures the average distance between the data points of a cluster's members compared to those of alternative clusters (Rousseeuw, 1987). This silhouette method is measured differently for each clustering method. In this analysis, we set the range of clusters as [2, 3, 4, . . . , 10]. The details are explained as follows.

Using the preprocessing data, we carried out hierarchical clustering to visualize the intrinsic clusters in the Terengganu River data and determine which areas are polluted in which clusters. The algorithm of hierarchical clustering works as follows. Initially, the algorithm considers each data point belonging to its own cluster. In this case, clusters are the

**FIGURE 10**
Adjusted rand index for all clusters.
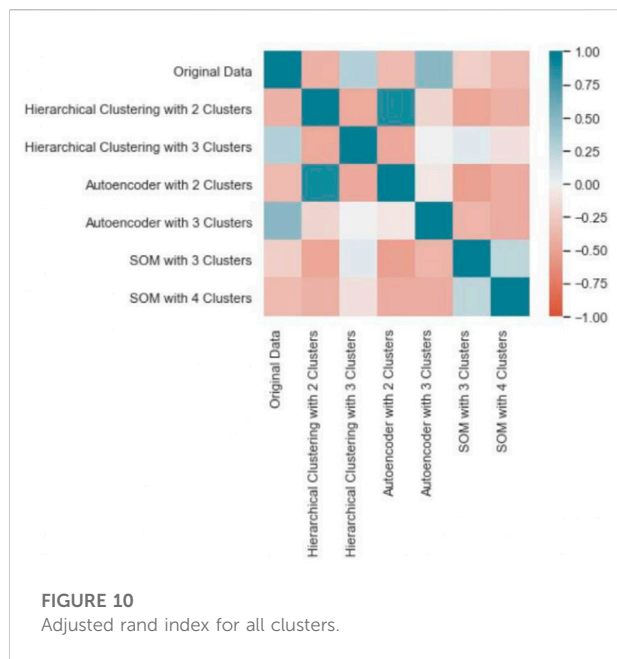
polluted areas. At each step, the algorithm merges the most appropriate pair of clusters based on their level of pollution. The iterative merging process continues until there is only one cluster. The appropriateness of a pair merging is decided by a criterion. Hierarchical clustering has many variants which differ from one another in the merge criterion employed. For this hierarchical clustering, we used the Ward method. According to the framework, one can use an objective function of searching for the most polluted area that reflects the chosen criterion for selecting the pair of polluted area clusters for merging. The pair which leads to the maximal value for the objective function is selected for merging at each step. Ward suggested using the total within-cluster sum of squared errors (TWSSE) as an objective function.

Initially, as each data point is considered belonging to its own polluted area cluster, the cluster mean point coincides with the data point of the cluster for each cluster. So, the TWSSE at this stage is 0. The TWSSE increases at each subsequent step when two polluted area clusters are merged, leading to the minimum increase in the TWSSE, for actual merging. Figure 5 shows the hierarchical maps for 405 water samples. These 405 water samples are well-separated into three groups (based on their different colors). For the best number of clusters, we next run K-means with silhouette analysis.

Then, K-means was set to run ten times setting k = 10 to get ten clusters. For each k, the clustering solution which has the maximum average silhouette analysis was selected. From among these ten clustering, the one with the maximum average silhouette analysis was taken as the final clustering solution. Based on Table 11 and Figure 6, silhouette analysis suggested two

clusters for polluted areas with hierarchical clustering with 0.5349.

Kohonen self-organizing maps (SOMs) are a type of unsupervised ANN (Kohonen, 1984). SOMs work by using a 2D grid of map units, which are each represented by a prototype vector, and are connected to their neighboring units by the neighborhood relation (Vesanto and Alhoniemi, 2000). The map is built up iteratively, so, for each sample which is a polluted area, the distance between the input vector (water quality parameters) and each of these prototype vectors is computed to identify the best matching unit (BMU). Then, the BMU and its neighboring units are adjusted to be closer to the input vector. The resulting grid of map units includes the samples assigned to them. When there is a high number of these map units (MUs), clustering can be performed on the MUs themselves (Vesanto and Alhoniemi, 2000). As the samples are mapped to MUs, the clustering of the samples can be extracted from this. The optimal number of clusters was identified as 4 (Figure 7), covering all 405 water samples.

Then, K-means was set to run ten times, setting k = 10 to get ten clusters. For each k, the clustering solution which has the maximum average silhouette analysis was selected. From among these ten clustering, the one with the maximum average silhouette analysis was taken as the final clustering solution. Based on Figure 8, silhouette analysis suggested four clusters for polluted areas with SOM of 0.3971.

Autoencoders are another form of ANN, operating in a three-layer structure. The first layer consists of the input data which undergo an encoding step. Thus, as where the original set of features is converted into a smaller set, called a latent feature representation of the dataset, there is necessarily data loss in this process, followed by a decoding step which converts this reduced feature set back into a dataset with the same number of features as the original, the reconstructed layer. By training the network to minimize the data loss between the input and reconstructed layers, the maximum information can be encapsulated in this latent feature set. Once the autoencoder has been trained on a subset of the dataset, it can be used to convert the full dataset into its latent features. The results of this feature extraction can then be clustered. These latent features are non-linear combinations of the original measurement, so they cannot be used to make inferences about them. The optimal clustering of a dataset divides it into groups that maximize the similarity within the clusters, while minimizing the similarity between them (Vesanto and Alhoniemi, 2000).

Then, K-means is set to run ten times, setting k = 10 to get ten clusters. For each k, the clustering solution which has the maximum average silhouette analysis is selected. From among these ten clustering, the one with the maximum average silhouette analysis is taken as the final clustering solution. Based on Figure 9, silhouette analysis suggested two clusters for polluted areas with autoencoder with 0.8578.

**FIGURE 11**
Optimal k using Elbow method.
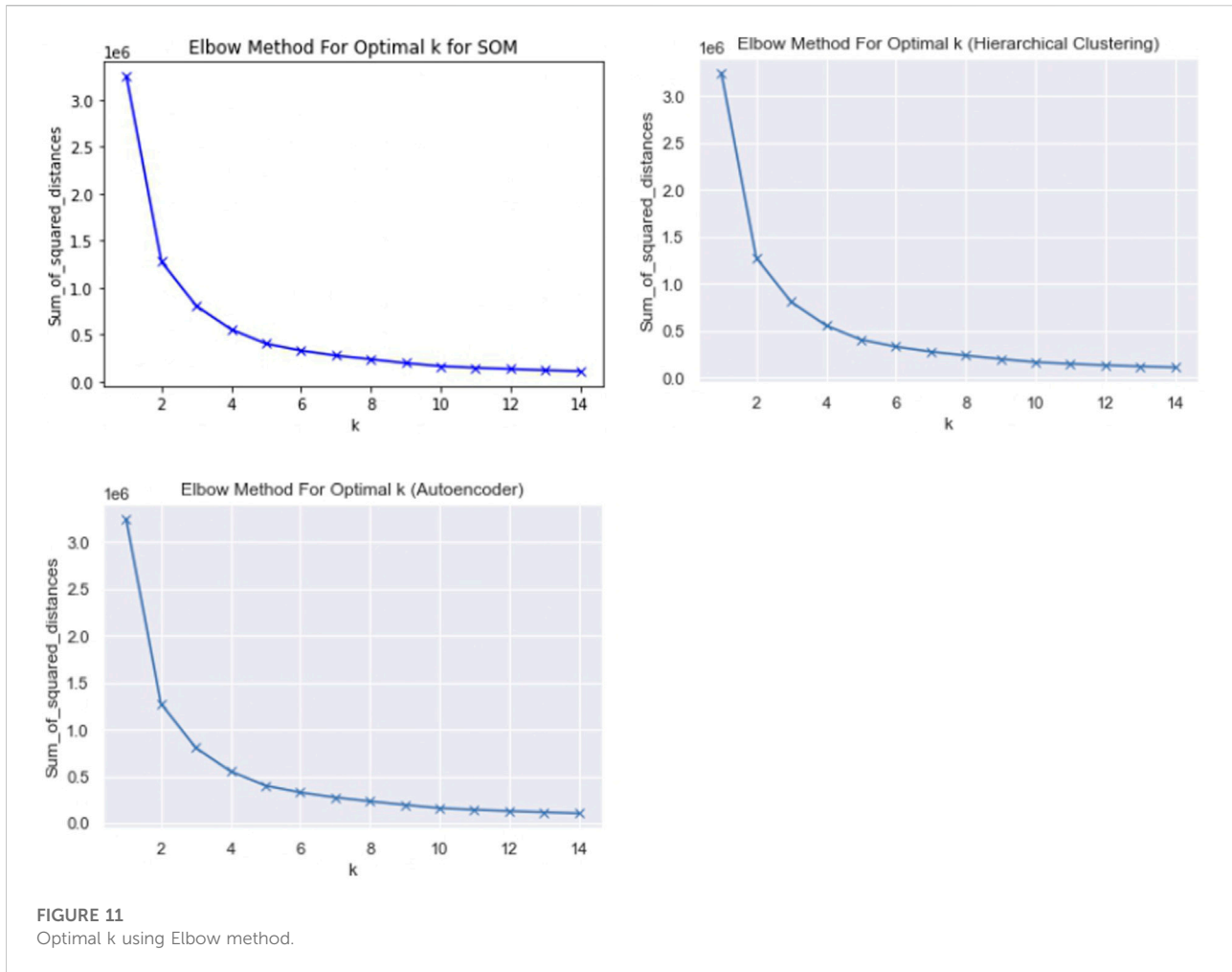
The original Terengganu River dataset proposedly divided this water polluted area into three different clusters which were clean, slightly polluted, and polluted, based on the Water Quality Index (WQI) analysis. This original data cluster information is used for the performance analysis. Therefore, an external cluster validity index is employed for assessing the quality of the clustering analysis. Specifically, we used the ARI. The ARI is a commonly used index proposed by Hubert and Arabie (1985). This method calculates the extent of agreement between two sets of groupings and outputs a representative score between 1 and −1. On this scale, 1 is identical, with 0 being the equivalent of a random association (Gates and Ahn, 2017). The highest achievable ARI score is 1. This score is achieved when the cluster membership computed by the clustering algorithm is the same as the true agreement of the clustering result with the true cluster memberships. Negative scores indicate against clustering.

Although silhouette analysis achieved a high index with two clusters for hierarchical clustering, four clusters for SOM and two clusters for autoencoder, this study suggested running the ARI for

six different numbers of clusters for all the algorithms. These six clusters for the algorithms include hierarchical clustering with two clusters, hierarchical clustering with three clusters, SOM with three clusters, SOM with four clusters, autoencoder with two clusters, and autoencoder with three clusters. We chose six different number clusters due to suggested clusters from the original data.

The ARI scores show that (original data and hierarchical clustering with two clusters: 0.0370) is lower than between (original data and hierarchical clustering with three clusters: 0.3147). Next, (original data and autoencoder with two clusters: 0.0793) is lower than between (original data and autoencoder with three clusters 0.4094). Last, (original data and SOM with three clusters: 0.0950) are higher than between (original data and SOM with four clusters: 0.0359). These results are also illustrated in Figure 10.

Based on the observed samples (water polluted area) and features (water quality parameters), we would expect our algorithms to be equivalent with the original data. Comparing these algorithms with different numbers of clusters with the original data, our results are contradicted with the silhouette analysis' results. Water polluted area samples in hierarchical
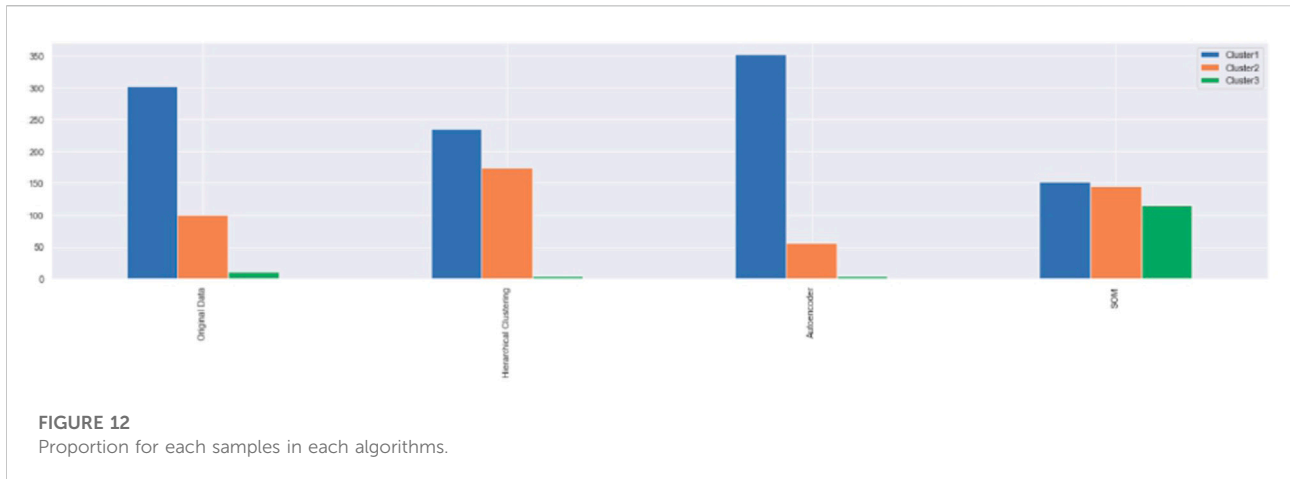
**FIGURE 12**
Proportion for each samples in each algorithms.

clustering with three clusters, autoencoder with three clusters, and SOM with three clusters are more similar with the water polluted area samples from the original data.

Determining the optimal cluster number is always a difficult part, especially for a dataset with different types of parameters. In addition, estimating the potential optimal cluster number for the analyzed dataset is a fundamental issue in clustering algorithms. Therefore, to double-confirm the best clusters for each method (based on our ARI results), the elbow method is used. The elbow method (part of the silhouette analysis) can determine the number of clusters in a data set. This method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use. Based on the subplot of elbow in Figure 11, we know that the estimated potential optimal cluster number obtained by the elbow method for each algorithm is three, which is consistent with the real cluster number from the original data and in line with the ARI results for all the algorithms. Figure 11 suggests that the optimal number of clusters for the Terengganu River dataset is three for each clustering algorithm. This result seems in line with that of the original Terengganu River dataset that suggested three different clusters for polluted areas, which are clean, slightly polluted, and polluted.

Last, the proportion of samples with the final presence of summary water polluted area identification is plotted in Figure 12. It shows that the original data are of 303 water samples in cluster 1, 101 water samples in cluster 2, and 11 water samples in cluster 3. Hierarchical clustering has 236 water samples in cluster 1, 174 water samples in cluster 2, and four water samples in cluster 3. The autoencoder has 356 water samples in cluster 1, 56 water samples in cluster 2, and four water samples in cluster 3. Last, SOM has 152 water samples for cluster 1, 147 water samples for cluster 2, and 116 water samples for cluster 3. Thus, cluster 1 can be stated

as clean, cluster 2 as slightly polluted, and cluster 3 as polluted.

Therefore, these three ML algorithms (hierarchical clustering, autoencoder, and SOM) can become the alternative algorithms to offer extensions to further solve decision-making problems. Previous studies on MCDM usually focused on the analysis of various available choices in certain situations or research areas to find the most preferred alternatives (our target or objectives) and present them in ranking order. This paper offers further investigation on the MCDM problems with the integrated ML techniques.

# 4 Conclusion and discussion

The main objective of this study is to develop an intelligent model for finding the best strategies for water security. The intelligent model includes integrating the IT2FVIKOR method with ML algorithms. IT2FVIKOR can capture different preferences among DMs and multiple strategies for decision problems. However, IT2FVIKOR itself only focuses on ranking and selecting from a set of alternatives only. Due to that, unsupervised ML was introduced to further analyze and solve the decision problems. This paper presented a two-phase fuzzy-ML–based framework, where in phase 1, we carried out experiments on solving the water security problems in Malaysia using IT2FVIKOR to evaluate six different strategies to find the best ways to enhance water supply security in Malaysia toward the five main criteria. It was found that strengthening the protection of water source areas is the best strategy for water security, reinforcing groundwater monitoring and protection in the second place, developing a water-saving system in the third place, and fully implementing the river chief system in the fourth place. Improving infrastructure to safeguard urban and rural water security was ranked last. Further steps after retrieving the

ranking and selecting the best alternatives were continued in the second phase.

Phase 2 continued with the unsupervised ML, where three different clustering algorithms, namely, hierarchical clustering, SOM, and autoencoder, were used to cluster the polluted area in the Terengganu River. Silhouette analysis was applied along with the clustering algorithms to estimate the number of optimal clusters in a dataset. Silhouette analysis suggested two clusters for polluted areas with hierarchical clustering with 0.5349, four clusters for polluted areas with SOM with 0.3971, and two clusters for polluted areas with autoencoder with 0.8578. Then, the ARI was applied to find the best comparison within original data with hierarchical clustering, SOM, and autoencoder. The ARI scores show that (original data and hierarchical clustering with two clusters: 0.0370) was lower than between (original data and hierarchical clustering with three clusters: 0.3147). Next, (original data and autoencoder with two clusters: 0.0793) was lower than between (original data and autoencoder with three clusters 0.4094). Last, (original data and SOM with three clusters: 0.0950) were higher than between (original data and SOM with four clusters: 0.0359). Based on the observed samples (water polluted area) and features (water quality parameters), we would expect our algorithms to be equivalent with the original data. Comparing these algorithms with different numbers of clusters with the original data, our results contradicted the silhouette analysis' results.

Thus, the elbow method was applied to double-confirm the best clusters for each clustering algorithm. The elbow method suggested that water-polluted area samples in hierarchical clustering with three clusters, autoencoder with three clusters, and SOM with three clusters were more similar to the water-polluted area samples from the original data. This result seems in line with that of the original Terengganu River dataset that suggested three different clusters for polluted areas which were clean, slightly polluted, and polluted. Last, lists of polluted areas in each cluster were retrieved. It shows that the original data are of 303 water samples in cluster 1, 101 water samples in cluster 2, and 11 water samples in cluster 3. Hierarchical clustering has 236 water samples in cluster 1, 174 water samples in cluster 2, and four water samples in cluster 3. Autoencoder has 356 water samples in cluster 1, 56 water samples in cluster 2, and four water samples in cluster 3. Last, SOM has 152 water samples for cluster 1, 147 water samples for cluster 2, and 116 water samples for cluster 3. It can be concluded that cluster 1 can be stated as clean, cluster 2 as slightly polluted, and cluster 3 as polluted. Finally, the two-phase fuzzy-ML-based framework can be concluded as the best water security strategies to be applied to the most polluted area. We can also conclude that integrated IT2FVIKOR with unsupervised ML offers a new alternative

to solving decision-making problems. For future work, we aim to integrate steps in IT2FVIKOR with ML to reach higher levels of uncertainty in the decision-making system. In addition, we also aim to integrate steps in IT2FVIKOR with ML using a large group of DM opinions. Moreover, it is suggested that it might be possible to closely look into different unsupervised ML techniques, develop a better technique that can retrieve better clustering, retrieve the best number of optimal clusters, reduce the noise or unnecessary data, and later improve the generalizability of the model.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

## Author contributions

NZ was the grant leader, who designed and conducted the study. WA and MP collected and analyzed data. SA conducted the study and reviewed the analysis. MG reviewed the programming and analysis. All authors were involved in writing and reviewing the paper and had final approval of the submitted and published versions.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Ahmed, M. F., Vanajakshi, L., and Suriyanarayanan, R. (2019). Real-time traffic congestion information from tweets using supervised and unsupervised machine learning techniques. *Transp. Dev. Econ.* 5, 20. doi:10.1007/s40890-019-0088-2

Alazemi, F. K. A. O. H., Ariffin, M. K. A. M., Mustapha, F., and Supeni, E. E. (2022). A new Fuzzy TOPSIS-Based Machine Learning framework for minimizing completion time in supply chains. *Int. J. Fuzzy Syst.*

Castro-Pardo, M. D., Martínez, P. F., and Zabaleta, A. P. (2022). An initial assessment of water security in Europe using a DEA approach. *Sustain. Technol. Entrepreneursh.* 1, 100002. doi:10.1016/j.stae.2022.100002

Chapagain, K., Aboelnga, H. T., Babel, M. S., Ribbe, L., Shinde, V. R., Sharma, D., et al. (2022). Urban water security: A comparative assessment and policy analysis of five cities in diverse developing countries of asia. *Environ. Dev.* 43, 100713. doi:10.1016/j.envdev.2022.100713

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20 (1), 37–46. doi:10.1177/001316446002000104

Delgado, P. (2019). *Development of a deep learning based attack detection system for smart grids*. Polytechnic De Madrid.

Demerdash, D. E., Omar, M. E. D., El-Din, M. N., El-Badry, H., Aly, E., and El-Molla, D. A. (2022). Development of a quality-based irrigation water security index. *Ain Shams Eng. J.* 13, 101735. doi:10.1016/j.asej.2022.101735

Dong, Q., Chen, X., Dong, S., and Zhang, J. (2021). Classification of pavement climatic regions through unsupervised and supervised machine learnings. *J. Infrastruct. Preserv. Resil.* 2 (5), 5–15. doi:10.1186/s43065-021-00020-7

El Bouchefry, K., and de Souza, R. S. (2020). "Chapter 12 - learning in big data: Introduction to machine learning," in *Knowledge discovery in big data from astronomy and Earth observation*. Editors P. Skoda, ō, and F. Adam (Elsevier), 225.

Evengard, B., Berner, J., Brubaker, M., Mulvard, G., and Revich, B. (2011). Climate change and water security with a focus on the Arctic. *Glob. Health Action* 4, 8449–8454. doi:10.3402/gha.v4i0.8449

Fetanat, A., Tayebi, M., and Mofid, H. (2021). Water-energy-food security nexus based selection of energy recovery from wastewater treatment technologies: An extended decision making framework under intuitionistic fuzzy environment. *Sustain. Energy Technol. Assessments* 43, 100937. doi:10.1016/j.seta.2020.100937

Gates, A. J., and Ahn, Y-Y. (2017). The impact of random models on clustering similarity. *J. Mach. Learn. Res.*

Gul, M., Mete, S., Serin, F., and Celik, E. (2021). "Fine–kinney-based occupational risk assessment using interval type-2 fuzzy VIKOR," in *Fine–kinney-based fuzzy multi-criteria occupational risk assessment* (Cham: Studies in Fuzziness and Soft Computing Springer), Vol. 398.

Gwp (2000b). *Towards water security: A framework for action*. Stockholm: Global Water Partnership.

Han, S., Wang, W., and Liu, X. (2021). A new type-2 fuzzy multi-criteria hybrid method for rail transit operation safety assessment. *Appl. Soft Comput.* 113, 107927. Part A. doi:10.1016/j.asoc.2021.107927

Hoffmann, M. (2005). Numerical control of Kohonen neural network for scattered data approximation. *Numer. Algor.* 39, 175–186. doi:10.1007/s11075-004-3628-7

Hubert, L., and Arabic, P. (1985). Comparing partitions. *J. Classif.* 2, 193–218. doi:10.1007/bf01908075

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9 (3), 90–95. doi:10.1109/mcse.2007.55

Jain, A. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* 31, 651–666. doi:10.1016/j.patrec.2009.09.011

Jain, N., Tomar, A., and Jana, P. K. (2021). A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. *J. Intell. Inf. Syst.* 56, 279–302. doi:10.1007/s10844-020-00614-9

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika* 43 (1), 241–254. doi:10.1007/bf02289588

Jones, J. A. A., Vardianian, T. G., and Hakopian, C. (2009). *Threats to global water security: Population growth, terrorism, climate change or communication*. Netherlands: Springer.

Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer.

Komer, B., Bergstra, J., and Eliasmith, C. (2014). "Hyperopt-sklearn: Automatic hyperparameter configuration for scikit-learn," in *Proceedings of the 13th Python in science conference* (SCIPY), 33

Liu, S., Ouyang, H., and Peng, F. (2008). Spatial data mining and visualization based on self-organizing map. *J. Commun. Comput.* 5 (12), 55.

Mohsin, M., Ali, S. A., Shamim, S. K., and Ahmad, A. (2022). A GIS-based novel approach for suitable sanitary landfill site selection using integrated fuzzy analytic hierarchy process and machine learning algorithms. *Environ. Sci. Pollut. Res.* 29, 31511–31540. doi:10.1007/s11356-021-17961-x

Moradi, J., Ghorbanzad, Y., and Beig, M. (2002). Identifying and prioritizing innovation criteria of projects in science and technology parks using fuzzy VIKOR. *Manag. Sci. Lett.* 2 (2), 587

Namany, S., Govindan, R., Martino, M. D., Pistikopoulos, E. N., Linke, P., Avraamidou, S., et al. (2021). An energy-water-food nexus-based decision-making framework to guide national priorities in Qatar. *Sustain. Cities Soc.* 75, 103342. doi:10.1016/j.scs.2021.103342

Nie, R-x., Tian, Z-p., Wang, J-q., Zhang, H-y., and Wang, T-l. (2018). Water security sustainability evaluation: Applying a multistage decision support framework in industrial region. *J. Clean. Prod.* 196, 1681–1704. doi:10.1016/j.jclepro.2018.06.144

Nilashi, M., Mardani, A., Liao, H., Ahmadi, H., Manaf, A. A., and Almukadi, W. (2019). A hybrid method with TOPSIS and machine learning techniques for sustainable development of green hotels considering online reviews. *Sustainability* 11, 6013. doi:10.3390/su11216013

Onda, K., LoBuglio, J., and Bartram, J. (2012). Global access to safe water: Accounting for water quality and the resulting impact on MDG progress. *Int. J. Environ. Res. Public Health* 9, 880–894. doi:10.3390/ijerph9030880

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12, 2825

Qin, J., and Liu, X. (2019). "An integrated interval type-2 fuzzy decision making based on VIKOR and prospect theory," in *Type-2 fuzzy decision-making theories, methodologies and applications. Uncertainty and operations research* (Singapore: Springer).

Qin, J., Liu, X., and Pedrycz, W. (2015). An extended VIKOR method based on prospect theory for multiple attribute decision making under interval type-2 fuzzy environment. *Knowledge-Based Syst.* 86, 116–130. doi:10.1016/j.knosys.2015.05.025

Richard, A. J., and Dean, W. W. (2002). *Applied multivariate statistical analysis*. London: Prenticee Hall, 265.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Shen, K-Y., and Tzeng, G-H. (2016). Contextual improvement planning by fuzzy-rough machine learning: A novel bipolar approach for business analytics. *Int. J. Fuzzy Syst.* 18 (6), 940–955. doi:10.1007/s40815-016-0215-8

Silberer, C., and Lapata, M. (2014). *Learning grounded meaning representations with autoencoders*, 721.

Soner, O., Celik, E., and Akyuz, E. (2017). Application of AHP and VIKOR methods under interval type 2 fuzzy environment in maritime transportation. *Ocean. Eng.* 129, 107–116. doi:10.1016/j.oceaneng.2016.11.010

Steinley, D. (2004). Properties of the hubert-arable adjusted Rand index. *Psychol. Methods* 9 (3), 386–396. doi:10.1037/1082-989x.9.3.386

Tewari, A., Zollhofer, M., Kim, H., Garrido, P., Bernard, F., and Perez, P. (2017). "Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction," in *Proceedings of the IEEE international conference on computer vision workshops*, 1274.

Tom, M. (1997). *Machine learning*. New York: McGraw-Hill.

Tu, Y., Wang, H., Zhou, X., Shen, W., and Lev, B. (2021). Comprehensive evaluation of security, equity, and efficiency on regional water resources coordination using a hybrid multi-criteria decision-making method with different hesitant fuzzy linguistic term sets. *J. Clean. Prod.* 310, 127447. doi:10.1016/j.jclepro.2021.127447

Umar, A., and Saraswat, R. N. (2022). Decision-making in machine learning using novel picture fuzzy divergence measure. *Neural Comput. Applic.* 34 (34), 457–475. doi:10.1007/s00521-021-06353-4

UNESCO (2009). *Water in a changing World*. Paris: Earthscan.

Veettil, A. V., Mishra, A. K., and Green, T. R. (2022). Explaining water security indicators using hydrologic and agricultural systems models. *J. Hydrology* 607, 127463. doi:10.1016/j.jhydrol.2022.127463

Vesanto, J., and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* 11, 586–600. doi:10.1109/72.846731

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature* 467, 555–561. doi:10.1038/nature09440

Wan, S. P., Chen, Z-H., and Dong, J-Y. (2021). An integrated interval type-2 fuzzy technique for democratic–autocratic multi-criteria decision making. *Knowledge-Based Syst.* 214, 106735. doi:10.1016/j.knosys.2020.106735

Wang, H., Pan, X., and He, S. (2019). A new interval type-2 fuzzy VIKOR method for multi-attribute decision making. *Int. J. Fuzzy Syst.* 21 (1), 145–156. doi:10.1007/s40815-018-0527-y

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58 (301), 236–244. doi:10.1080/01621459.1963.10500845

Warrens, M. J. (2008). On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index. *J. Classif.* 25 (2), 177–183. doi:10.1007/s00357-008-9023-7

Wongravee, K., Ishigaki, M., and Ozaki, Y. (2020). "Chemometrics as a green analytical tool," in *Challenges in green analytical chemistry*, 277

Wu, Q., Zhoua, L., Chena, Y., Chena, Y., and Chen, H. (2019). An integrated approach to green supplier selection based on the interval type-2 fuzzy best-worst and extended VIKOR methods. *Inf. Sci.* 502, 394–417. doi:10.1016/j.ins.2019.06.049

WWC (2000). *World water vision*. Paris: Earthscan.

Yazici, I., Beyca, O. F., Gurcan, O. F., Zaim, H., Delen, D., and Zaim, S. (2022). A comparative analysis of machine learning techniques and fuzzy analytic hierarchy process to determine the tacit knowledge criteria. *Ann. Oper. Res.* 308, 753–776. doi:10.1007/s10479-020-03697-3

Zhao, F., Feng, J., Zhao, J., Yang, W., and Yan, S. (2017). Robust lstm-autoencoders for face de-occlusion in the wild. *IEEE Trans. Image Process.* 27 (2), 778–790. doi:10.1109/tip.2017.2771408

Zhou, C., and Paffenroth, R. (2017). *Anomaly detection with robust deep autoencoders*, 665.

Zhu, Y., Ting, K. M., Jin, Y., and Angelova, M. (2022). Hierarchical clustering that takes advantage of both density-peak and density-connectivity. *Inf. Syst.* 103, 101871. doi:10.1016/j.is.2021.101871

Ziyadidegan, S., Razavi, M., Pesarakli, H., Javid, A. H., and Erraguntla, M. (2022). Factors affecting the COVID-19 risk in the US counties: An innovative approach by combining unsupervised and supervised learning. *Stoch. Environ. Res. Risk Assess.* 11, 1469–1484. doi:10.1007/s00477-021-02148-0