Article

A Novel Approach Based on Machine Learning and Public Engagement to Predict Water-Scarcity Risk in Urban Areas

Sadeq Khaleefah Hanoon [1,*] , **Ahmad Fikri Abdullah** [2,3], **Helmi Z. M. Shafri** [1] **and Aimrun Wayayok** [2]

1   Civil Engineering Department, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia
2   Biological and Agricultural Engineering Department, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia
3   International Institute of Aquaculture and Aquatic Sciences (I-AQUAS), Universiti Putra Malaysia, Port Dickson 70150, Malaysia
*   Correspondence: gs58154@student.upm.edu.my

Abstract: Climate change, population growth and urban sprawl have put a strain on water supplies across the world, making it difficult to meet water demand, especially in city regions where more than half of the world's population now reside. Due to the complex urban fabric, conventional techniques should be developed to diagnose water shortage risk (WSR) by engaging crowdsourcing. This study aims to develop a novel approach based on public participation (PP) with a geographic information system coupled with machine learning (ML) in the urban water domain. The approach was used to detect (WSR) in two ways, namely, prediction using ML models directly and using the weighted linear combination (WLC) function in GIS. Five types of ML algorithm, namely, support vector machine (SVM), multilayer perceptron, K-nearest neighbour, random forest and naïve Bayes, were incorporated for this purpose. The Shapley additive explanation model was added to analyse the results. The Water Evolution and Planning system was also used to predict unmet water demand as a relevant criterion, which was aggregated with other criteria. The five algorithms that were used in this work indicated that diagnosing WSR using PP achieved good-to-perfect accuracy. In addition, the findings of the prediction process achieved high accuracy in the two proposed techniques. However, the weights of relevant criteria that were extracted by SVM achieved higher accuracy than the weights of the other four models. Furthermore, the average weights of the five models that were applied in the WLC technique increased the prediction accuracy of WSR. Although the uncertainty ratio was associated with the results, the novel approach interpreted the results clearly, supporting decision makers in the proactive exploration processes of urban WSR, to choose the appropriate alternatives at the right time.

Keywords: public engagement; urban water; water-shortage risk; machine learning; GIS; crowdsourcing

1. Introduction

During the previous century, the world population has grown from 2 billion to 7.7 billion, with nearly 80% living in water-stressed areas [1,2]. Population growth and urban sprawl have key roles in increasing water demand, which becomes incredibly difficult to provide in urban areas given that more than half of the world's population now live in cities [3,4]. Moreover, climate change has influenced water supplies, particularly in heavily populated dry and semi-arid areas; thus, wide areas of human civilisation are facing water shortage risk (WSR) [5–7]. The management of urban water is difficult to understand due to the wide range of measurement indicators, including drinking-water quality and quantity, water governance, hygiene, human well-being, water production, socio-economic factors, environmental and ecological health, WSR and water administration [8]. With the rising pressure on urban water demand, several studies have attempted to explain the interaction of external and internal variables in urban water scarcity [9].

There exist two types of challenge associated to the water crisis facing urban communities: external challenges represented by climate change and national water management regarding dividing quotas water between cities, and internal challenges represented by inequity in water supply between neighbourhoods [10]. On the one hand, climate change, with rising temperatures and decreased precipitation, has increased water demand and reduced the available water [11]. On the other hand, the city's share of the national water has a significant effect on the water delivered to water-treatment projects. A city with a higher priority of water allocation faces fewer water-supply challenges. Despite climate change having a negative effect on the current water supply, other driving factors (e.g., population growth, urbanisation, infrastructure, living standards and spatial inequality) put more pressure on the distribution supply system [12]. The principle of spatial inequality regarding water supply between a city's communities has led to variation in water-supply levels. That is, some sectors of a city do not experience water scarcity even when hydro-climatic circumstances are unfavourable; conversely, other sectors are vulnerable to WSR even when hydro-climatic conditions are conducive. To make progress towards improved water management, water governance must consider human rights alongside other driving forces, such as population increase and climatic conditions [2]. Therefore, urban water management needs more efforts to design scientific approaches that encourage public participation (PP) in different levels of the decision-making process to combat water scarcity [13].

However, most of the previous studies have neglected the theoretical analyses that demonstrate that demand mitigation measures and decreasing water losses must be considered to reduce WSR in urban areas [3]. In this context, including some criteria, such as leaks in the pipeline and water consumption per plot, will boost the accuracy of risk analysis [14]. In addition, the implementation of leakage reduction measures can decrease unmet demand [15]. Moreover, predicting the unmet water demand from water resources is an essential planning element in water risk analysis [16]. Furthermore, the traditional method of estimating water demands based on population projections does not allow for spatial planning, which is critical for water-supply systems (WSS) [17]. Therefore, this study fills those gaps by incorporating water losses and unmet demands within a set of social and geographic factors that can simulate carefully a city's water-supply operation in favourable and unfavourable hydro-climatic conditions. Traditional approaches in this domain, however, must be developed to deal with the suggested driving factors in this study.

According to the literature, the most widely used methods for risk analysis of floods, fire and landslides and selecting the optimal sites of different projects are multiple criteria decision analysis (MCDA) and machine-learning (ML) models [18,19]. MCDA and ML are two of the most significant and rapidly emerging artificial-intelligence fields, which provide approaches that can assist urban planners in decision-making processes [20]. Several forms of MCDA have been developed, such as analytic hierarchy process (AHP; Saaty, 1980) and multi-attribute utility theory, which account for the effects of criteria selection and weights on final decisions [21]. The combination of AHP with GIS constitutes one of the most crucial methods that plays a significant role in spatial analysis [22,23]. Any evaluation method, however, has advantages and limitations.

On the one hand, MCDA has significant disadvantages. The needless complexity of the optimisation process requires extra time, effort, resources and implementation costs [24]. Its results are associated with the uncertainty and subjective bias of experts' opinions. Different from MCDA, ML has a stronger predictive capability; thus, ML models outperform MCDA in terms of prediction performance [18,25]. On the other hand, in comparison with the integration of MCDA and GIS, the primary advantage of integrating GIS and ML approaches is improved performance on the resulting susceptibility maps, as well as faster data-processing time [26]. Although GIS and ML models can predict the WSR for the study area effectively, the real effects of the continuous factors (e.g., proximity to pumping stations (PS), main pipelines and rivers) are difficult to identify

because their effects cannot be obstructed by the neighbourhood's borders across the study area. Furthermore, the conventional ML technique faces challenges of result interpretability, thereby limiting its broad usage in the identification of anomalies of hydrological-affecting elements [27]. Moreover, human perspectives are ignored in ML judgements, although the opinions of people who are exposed to vulnerability or risk are an important factor in decision making.

To acquire appropriate answers to the difficulties that face urban areas, the decision-making process should entail listening to people's thoughts, concerns and wants and then engaging them in active participation [28]. According to some scholars, social participation is critical for the achievement of sustainability because expanding public engagement broadens the spectrum of perspectives, making possible conflict lines more visible [29]. Therefore, citizens' feedback provides a rich supply of updated information, which helps improve the quality of the analysis and leads to different answers than when traditional types of data are used [30].

This study aims to develop a novel approach in the urban water field to overcome the limitations and challenges of traditional approaches and take advantage of the good characteristics of ML, MCDA and GIS, which were mentioned in previous studies. This approach can detect WSR in urban areas, identify the contribution of each driving factor, rank the city districts according to water deficit level and determine the districts that are more affected by WSR than others. The PP of local residents was adopted in this approach as a new technique for overcoming subjectivity bias and ensuring the proactive evaluation of the mitigation process in urban areas that face WSR. Thus, Nasiriyah City in Iraq was chosen as a study area. For this goal, five types of ML algorithms, namely, support vector machine (SVM), multilayer perceptron (MLP), K-nearest neighbour (KNN), random forest (RF) and naïve Bayes (NB), were used to build ensemble algorithms to predict WSR, as well as extract the weights of the conditional factors to use them with the weighted linear combination (WLC) function to define WSR with another method [31]. Furthermore, a sixth model, SHapley Additive exPlanations (SHAP), was included to explain the findings of the aforementioned models. The SHAP technique helps increase the interpretability of ML results [26,27].

2. Methodology

Urban water management is difficult to understand given the wide range of measurement indicators, including drinking-water quality and quantity, water governance, hygiene, human well-being, water production, socioeconomic factors, environmental and ecological health, and WSR [8]. This study was conducted to develop a novel machine-learning approach to predict WSR in urban areas based on PP. However, theoretical approaches face many limitations when implemented in practical cases. Thus, a practical approach based on public participation is proposed, focusing on vulnerable areas while considering physical and socioeconomic contexts. The approach was developed to detect WSR in two ways, namely, direct prediction using ML models and the weighted linear combination (WLC) function in GIS. The trained model was built from independent and dependent variables. Ten driving factors were adopted for the independent variables based on the literature review, local experts' opinions, and data availability, to detect WSR in urban areas. By contrast, the PP results of the evaluation of water shortage risk were adopted as dependent variables, as explained in the following sections. This work was carried out in Nasiriyah City, Iraq.

2.1. Study Area

The study area is Nasiriyah City, Iraq. It is located along the Euphrates River banks between latitudes of 31°10′00″ N and 30°50′00″ N and around longitudes of 46°00′00″ E and 46°24′00″ E, with an altitude of approximately 4 m above M.S.L., as shown in Figure 1. The region has a total area of about 46,000 ha, 4300 ha of which are urban areas divided into 93 neighbourhoods, and 200 ha of which is rural land that is connected to the city

neighbourhoods. Its population is over 700,000 people living in the city, and 20,000 people in the rural area (based on the 2021 local census). The city is the largest urban centre in Dhi-Qar governorate, and it is the fourth governorate in Iraq in terms of population. Although the main Euphrates River passes the Nasiriyah City centre, the water is supplied from the Garaf River, which is about 40 km from the city, due to the high saltiness of water of the Euphrates River in this region. The Garaf River is a small branch of the Tigris River, which supports more than million residents, most of whom mainly rely on farming. Moreover, climate change has affected the environment of the city; for example, in recent years, the temperature in the region has reached 52 °C in the summer [32]. As a result, the study area's residents frequently and severely experience water shortages. Although some neighbourhoods experience water scarcity to a lesser extent, the majority of the neighbourhoods consistently experience this problem due to social, economic and infrastructure distribution inequalities.
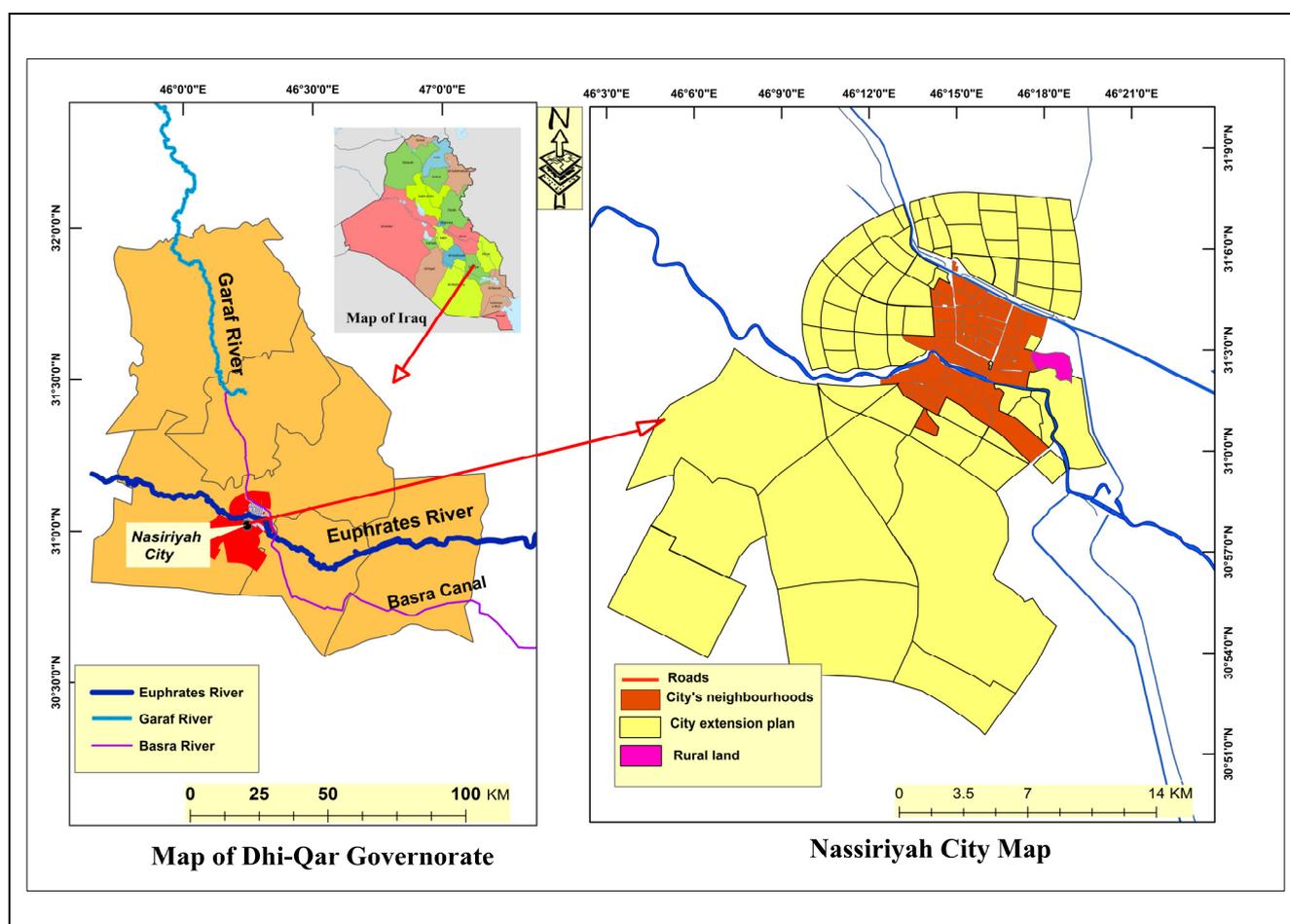
Figure 1. The study area: Nasiriyah city, Iraq.

2.2. Criteria Selection

Scholars have argued about the driving forces of urban WSR. Millington (2018) stated that hydro-climatic conditions and technical, social and inequality issues are significant driving factors for urban WSR [33]. Cordão et al. (2020) stated that WSR in cities is linked not only to population, urbanisation and climatological change but also to water imbalances (supply and water demand), inequitable allocation of infrastructure amongst city areas and economic situations [14]. Similarly, Simukonda, Farmani and Butler (2018) highlighted seven factors that contribute to WSR in urban areas, including social injustice, demographics and economics, increasing water demand, low system pressure, inadequate

electrical supply and water scarcity [34]. Noori et al. (2022) considered different factors, including environmental, geographical, geological, economic, climatic and social conditions, related to WSR [35]. In addition, the water-scarcity issue in a city was summarised by Krueger, Rao and Borchardt (2019) into four basic factors: (1) the allocated amount of water to the city; (2) the facilities for water filtration, storage and delivery; (3) the monetary budget for infrastructure construction and maintenance; and (4) managerial performance for controlling and running the water system [36]. Liu et al. (2022) mentioned that for the short term, the electricity power supply is strongly related to WSR [37]. In brief, urban places that are facing the same conditions of water deficit (a prolonged and extreme drought) might have different degrees of deficit hazard due to other reasons that are not linked to climate condition.

Furthermore, this study highlighted the spatial analysis of factors essential to supplying homes with potable water. This process, which is usually used by most countries worldwide, passes through many stages via a system known as WSS, which runs from water sources to homes. Water reservoirs, water treatment plants (WTP), PS, electricity-supply facilities, pipeline networks, and indoor water storage are the essential components that make up this system. However, it is difficult to dispense with any part of the WSS. In other words, if any part of it breaks, the water supply to houses will be cut off. Therefore, spatial driving factors that reflect the roles of the components of WSS were adopted in this study. Consequently, the ten criteria listed below were chosen based on the literature review, the effect degree of each criterion on the WSS, local experts' opinions and data availability.

2.2.1. Unmet Demand (C1)

The unmet demand is a term used to express the difference between the amount of water required and the amount of water allocated for each demand node along the river based on water availability and demand priorities [38]. The water demand for any city is usually calculated based on its population and water consumption per capita [39]. The water amount should be enough to keep WTPs operating efficiently to produce sufficient water for all shared districts. However, demand will be unmet if the water in the rivers cannot sufficiently meet the city's needs. As a result, increasing the unmet demand will increase WSR. In this work, the Water Evolution and Planning (WEAP) model was applied to calculate unmet water demand as a required criterion which was aggregated with other spatial criteria to produce WSR maps using GIS.

2.2.2. Proximity to the River Basin (C2)

According to the Water Distribution Directorate, one of the most important factors affecting the degree of WSR in urban areas is the distance from the river to the water reservoirs. As the distance increases, the loss increases due to unofficial subscriptions, water leaks and pipe accidents. Therefore, neighbourhoods that draw their water from reservoirs far from rivers are more susceptible to water shortages than other urban areas [40,41].

2.2.3. Proximity to PS (C3)

On the one hand, compared with neighbourhoods farther away from the PS, water leaking is higher in the neighbourhoods close to it due to the high pressure. On the other hand, neighbourhoods farther away from PS are most at risk from a shortage due to the low water pressure. As a result, residents who live in the middle of the city, between the PS and the city's edge, are in a good location for water pressure [42–45].

2.2.4. Proximity to Major Pipes (C4)

The water pressure in the vicinity of the main pipelines is higher than the pressure in the areas close to the pipeline ends. Thus, neighbourhoods at the network's end often have lower pressure and are vulnerable to water scarcity. Consequently, the proximity to the main pipelines and water-supply quantities has a negative linear relationship [46].

2.2.5. Age of the Network Pipes (C5)

The efficiency of the water distribution system in urban areas is largely determined by the system's operating life. The roughness of the pipe tends to increase with age due to the accumulation of various components surrounding the interior surface of the pipes, resulting in negative effects on the performance of the pipeline network system [47]. Therefore, the efficiency of the water supply decreases as the age of the network pipes increases [48].

2.2.6. Capacity of Reservoir (C6)

The capacity of reservoirs is critical for controlling the efficiency of supplying water to the neighbourhood. However, an equilibrium equation between the population and the capacity of reservoirs must be maintained [49]. This equation should be considered in long-term development plans, and policymakers should link population growth and urban sprawl to the capacity of urban water required. Therefore, the risk is reduced by increasing the capacity of the reservoir [50].

2.2.7. Electricity Supply (C7)

To offer uninterrupted service, water systems require an adequate level of electric-power reliability. WTP and PS rely heavily on electricity to keep water flowing into city neighbourhoods. Recurrent power outages cause unexpected cuts in water delivery [37]. Even slight power outages might have major ramifications for some utilities; thus, the redundancy of supply or backup generating capacity reduces the risk [51]. However, WSR and power outages have a positive linear relationship [34].

2.2.8. Population (C8)

Water consumption and population are intricately related; water demand is estimated by multiplying water use per person by population. Increases in population and demand, for example, can have a significant influence on urban water resources [52]. Furthermore, increased populations have exacerbated water shortages in urban areas [53,54]. A neighbourhood with a higher population is more vulnerable to water scarcity than others.

2.2.9. Population Density (C9)

This creates relative pressure on WSS. A positive linear relationship exists between the population density of a neighbourhood and WSR; thus, the higher density districts put higher pressure on WSS than districts with a lower density, even though they have the same population [14,55,56].

2.2.10. Unemployment Ratio (C10)

According to specialists on urban water supply, unemployment affects residential water-storage capacity [57]. On the one hand, houses in wealthy neighbourhoods are frequently provided with enough water reservoirs [14]. Private reservoirs assist people during water shortages. Households in poor neighbourhoods, on the other hand, typically suffer from water shortages when the water supply is cut [58].

2.3. Data Collection

The data collection was the stage in which a long time and more effort were spent compared with the rest of the stages in this study, because the WSS is a complex and overlapping system. In this stage, different types of data (i.e., raster, spatial dataset and statistical data) from different sources were collected to extract the driving factors of the WSR. In addition, surveying work was needed to determine missing data and check the locations of different parts of the WSS. Furthermore, given that climatic data provided by the NASA/Prediction of World Wide Energy Resources (POWER) can be used with reasonable confidence in scientific research [59], the climate data of the study area for period of 1984–2021 were downloaded from NASA/POWER. Table 1 lists the types of data collected for this work.

Table 1. Data collected from the Nasiriyah City and some reliable open data source websites for this study.

No.	Data	Description	Source of Data	Data Accuracy
1	Sentinel 2 image, acquired on 14 November 2021	The images were used to determine the catchment area of the Garaf River Basin and land cover classes	European Union's Earth observation programme (https://scihub.copernicus.eu, (accessed on 5 December 2021))	10 m
3	Land use shapefiles (SHF) of streets, districts borders and class of land uses	Shapefiles were used to extract land use classes, neighbourhoods' borders and street paths	Nasiriyah City Municipality, Iraq	2 m
4	Shapefiles and raster for the WSS (2021)	Data were used to generate network pipeline paths and determine the pipeline diameters and locations of the reservoirs, WTP and PS	Dhi-Qar Water Directorate	2 m
4	Master plan of Nasiriyah City	Shapefiles were used to validate neighbourhoods' borders, streets paths, pipeline paths and land use	The Office of Urban Planning, Nasiriyah City, Iraq	2 m
7	Statistical data on unemployment (2021)	Data were used to extract a conditional factor (C10)	Department of Statistics in Ministry of Planning	-
8	Population census (2021)	Data were used to extract two conditional factors (C8 and C9)	Department of Statistics in Ministry of Planning	-
9	Shapefiles of network river	The shapefiles were used to extract location demand nodes, gauge location, and Garaf River path and its branches	Ministry of Irrigation/Dhi-Qar (Iraq)	5 m
10	Hydrological data (1990–2021)	Garaf River's monthly discharge was used to model the WEAP, and hydrological parameters were used to estimate the conditional factor (C1)	Ministry of Irrigation/Dhi-Qar (Iraq)	±5%
11	Climate data (1984–2021)	Climate data were required for the WEAP model	(NASA/POWER) website, (https://power.larc.nasa.gov, (accessed on 21 April 2022)).	-
12	Site surveying work using GPS	The survey was needed to check locations of PS, reservoirs data, pipeline paths, WTP, piping division, and gauges of rivers	Author	1 m

2.4. Generating Driving Factors

In this study, GIS was implemented to supply inputs, spatial data analysis and WSR visualisation at a neighbourhood scale. GIS techniques can generate driving factors from the primary database based on the relationship between geographic features and WSR. In this study, several spatial-analysis techniques were used to derive the driving factors of WSR, as shown in Figure 2.
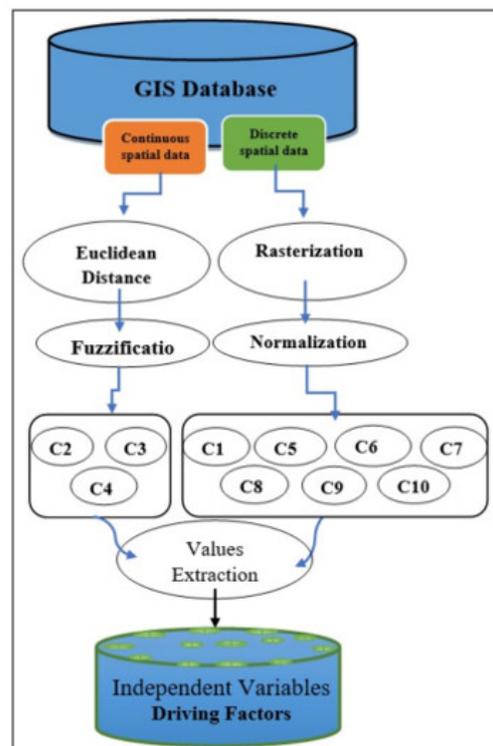
Figure 2. Generating driving factors using spatial-analysis techniques in GIS.

2.4.1. Euclidean Distance and Rasterization

In a GIS project, rasterization and Euclidean distance are typically used to convert vector geometries (points, lines, and polygons) into raster images. Rasterization is used on discrete spatial data, whereas Euclidean distance is used on continuous geographical data [60]. As a result, seven driving factors were constructed from the collected data using the rasterization tool based on specified data: population, population density, unemployment ratio, unmet water demand, electric supply, age of the network pipes and capacity of reservoirs. These driving factors had distinct units, and they were normalised to a common range (from 0 to 1) in the modelling using the linear interpolation Equation (1) [61]. The normalisation technique allows for expressive comparisons, with 1 representing 'high risk' and 0 representing 'low risk'. By contrast, the Euclidean distance is a useful function based on the Pythagorean theorem, which calculates the Euclidean distance to the nearest source for each cell in a raster dataset using Equation (2) [62]. Using this technique, spatial data of the main pipes, PS, the river and reservoirs were transformed into raster form, indicating the existing distances from those geographic features to the remaining buffer region.

$$yi = \frac{(xi - \text{x.min})}{(\text{x.max} - \text{x.min})} \tag{1}$$

where $0 \leq yi \geq 1$, and xi is the value of any pixel in the raster.

$$\text{d} = \sqrt{\left[(X2 - X1)^2 + (y2 - y1)^2\right]} \tag{2}$$

where d is the distance between the specific feature (the major pipes or PS) and the remaining sites.
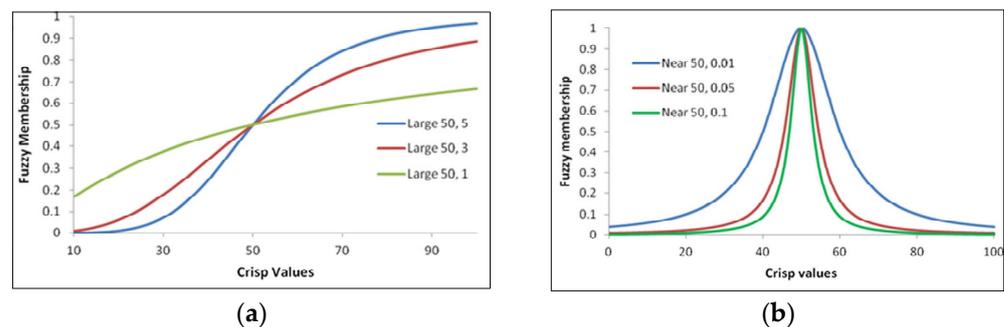
2.4.2. Fuzzy Large and Near Membership

Fuzzy large membership and fuzzy near membership are two major spatial analytic functions used in this study [63]. On the one hand, when larger input values are more

probable to be part of the set, the fuzzy large function is employed, where values higher than the midpoint have a higher probability of membership, and values less than the midway have a declining membership [64]. As a consequence, this function was applied against the raster dataset of two relevant criteria (proximity to the river basin and proximity to the main pipelines) based on the relationship in which the WSR degree increases with the distance from the main pipelines or river basin to householders or WTP based on Equation (3), as shown in Figure 3a. The fuzzy near function, on the other hand, is most useful when membership is close to a given value [65]. The function is defined by a midpoint that defines the set's centre [66]. On the basis of Equation (4) and as shown in Figure 3b, membership reduces from 1 to 0 as values travel away from the midpoint in positive and negative directions [67]. As a result, the criterion of proximity to PS was determined using fuzzy near membership. This criterion defined the neighbourhoods in the study area that are near to the centroids of the distances between the PS and the edges of the pipeline network. Conversely, the area near the PS has high water pressure, which causes pipes to leak; the area at the pipeline network's edges has a lower water pressure.

$$u(x) = \frac{1}{1 + \left(\frac{x}{f2}\right)^{-f1}} \tag{3}$$

$$u(x) = \frac{1}{1 + f1(x - f2)^2} \tag{4}$$

where $f1$ is the spread of the function, $f2$ is the midpoint and $u(x)$ is the membership value of category x.

Figure 3. (**a**) Fuzzy large membership; (**b**) fuzzy near membership.

2.5. Public Participation Process

Risk levels in urban areas that are evaluated by stakeholders or experts are often associated with uncertainty and subjective bias in experts' opinions. In addition, human perspectives are usually ignored in ML judgments, although the opinions of people exposed to vulnerability or risk are an important factor in decision-making. Therefore, participation of residents was adopted in this approach as a new technique for overcoming subjectivity bias and ensuring the proactive evaluation of the mitigation process in urban areas that face WSR. However, the safety requirements associated with COVID-19 presented complex challenges to conducting interviews with the public. Therefore, in this study, mobile phones were used to connect with local residents with the help of the local administration of Nasiriyah City (the study area). The advantage of mobile technology for data acquisition is relevant when local-authority resources are lacking and when more expensive data-collection methods cannot be used [68]. Fifty neighbourhoods were selected as the sample for the city. At least 10 people were contacted as representatives of each neighbourhood to evaluate WSR. Local experts reviewed the collected data and determined that 35 neighbourhoods were vulnerable to water shortage risk.

2.6. Data Model in ML

Nasiriyah City in Iraq, with a total of 94 neighbourhoods and the 10 criteria related to WSR, was selected as an input dataset of the ML model to predict WSR and extract the weights of relevant criteria to water shortage in urban areas. The data model and independent and dependent variables were first organised. Feature-reduction process was conducted, and a collection of six ML models were then constructed to predict uncertainty in the dependent variable. The independent variables were the values of the 10 criteria, which were extracted by a specific geospatial tool into the GIS environment. By contrast, the dependent variables were the evaluation of the WSR (high risk, low risk) of the city neighbourhoods by their residents' participation through PP. This assessment adopted the result of PP, which was organised by the local municipal council of Nasiriyah City (the study area) to evaluate the WSS in the city. Figure 4 exhibits the methodology employed in this paper.
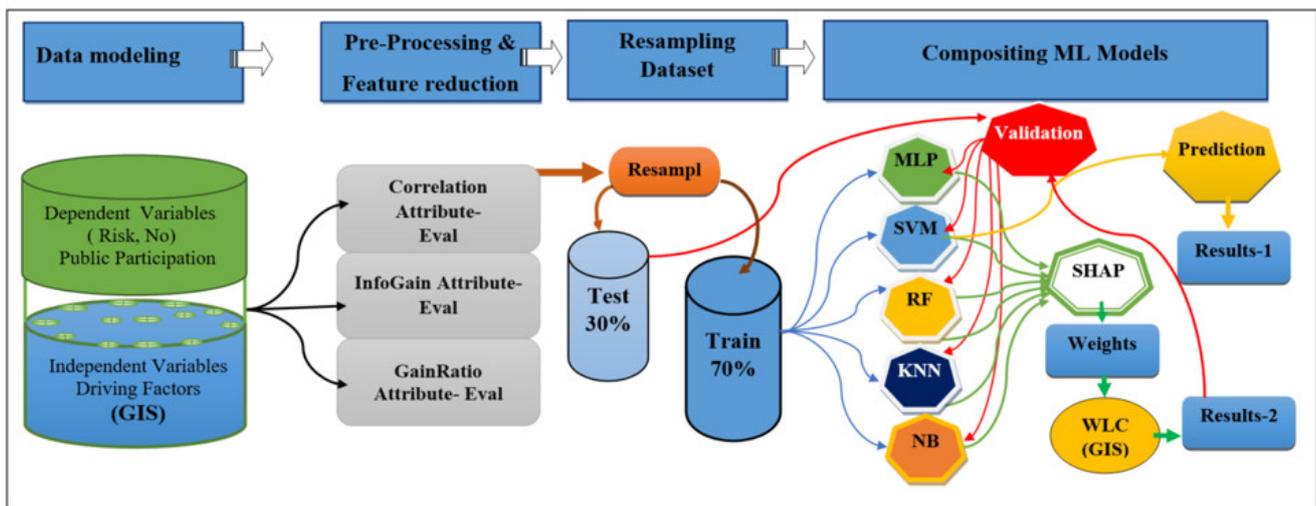

Figure 4. Methodology of the data modelling in this work.

2.7. Constructing ML Models

Several ML models were selected to predict WSR, with the aid of two common ML applications, Weka® and Orange®. On the one hand, data pre-processing was conducted using Weka software [69]. Data pre-processing refers to the preparation of the original data for usage by an ML model, which includes data integration, reduction, cleaning and transformation. On the other hand, the ensemble of ML models was built using Orange software, including MLP, SVM, KNN, RF, NB and SHAP models.

2.7.1. Feature Reduction Using Weka Software

The large magnitude of the input attributes hinders the ML performance and has a negative influence on the classification results. From this point forward, feature reduction is a required pre-processing strategy before employing ML models to remove irrelevant criteria and apply the best prediction model [70]. The quality of the models' execution is often improved by using an appropriate feature-reduction strategy before the training process [71]. In this study, three types of attribute evaluator were used: InfoGainAttributeEval, GainRatioAttributeEval and CorrelationAttributeEval, which were provided by Waikato Environment for Knowledge Analysis, Weka software, version 3.8.6. InfoGainAttributeEval is a single-attribute examiner that employs the ranker algorithm, which measures the information gain of the selected attributes according to the stated classes using a discretisation approach [72]. GainRatioAttributeEval is another evaluator, which identifies the factor value by measuring the gain ratio in relation to the output [73]. CorrelationAttributeEval calculates the Pearson correlation coefficient between the output and input of attributes [74].

In addition, the algorithm's 10-fold cross-validation approach was used to improve the performance, decrease error rates and boost efficiency [75]. Each attribute evaluator, however, evaluates the independent variables based on their mean value expressed as a number between 0 and 1 [76].

2.7.2. Modelling with Orange Software Application

MLP, SVM, NBC, KNN, and RF were used to build the prediction models using Orange data-mining software. A sixth model, SHAP, was also added to explain the findings of those models. MLP is a feedforward artificial neural network made up of three key components: an input layer, hidden layers (one layer or more), and an output layer. Feedforward allows the data to transmit in one direction from the input layer through hidden layers to the output layer [77]. Neural networks with one, and rarely two, hidden layers are widely used in practice and have shown excellent performance [78]. A large number of hidden layers leads to increased computation time and the danger of over-fitting [79]. In addition, the number of neurons in the hidden layer can be found by trial and error [80]. However, in this study the hidden layer was set to one layer with 6 hidden neurons based on Equation (5) [81].

$$n = (i + o)/2, \tag{5}$$

where i is the input layer, and n represents the number of neurons in the hidden layer (s).

SVM is a collection of supervised ML models with a related kernel function that may be used to accomplish non-linear classification (SVC) and regression (SVR) [82]. It is utilised in various fields, including time series, regression, character recognition and classification (clustering). SVM may be modified to convert non-linear input into linear input using a kernel function and then solve problems (e.g., linear problems) [83]. Nevertheless, SVM can handle weighted instances and perform incremental–decremental learning based on Equation (6) [75,84].

$$f(x) = w\mathrm{T}\,\varphi(x) + b \tag{6}$$

where b denotes the deviation vector, w is the weight and $\varphi(x)$ is the function of non-linear mapping.

NBC can also forecast and produce data over a range of classes as opposed to merely providing the most likely class with most approaches. Although NBC is comparable to the widely used supervised grouping algorithms, such as SVM and MLP, it has a number of important parameters that must be carefully recognised and set for better results [85]. Moreover, strong non-linearity in NBC, as opposed to SVM, makes it an excellent choice for the classification of urban vulnerability [86]. KNN is a well-known, straightforward and non-parametric ML technique that has been employed in several pattern-recognition applications; thus, many academics have suggested using it for different categorisation problems because of its simplicity and efficacy [87]. KNN-based methods regularly calculate the similarity between a training model and a testing model using Euclidean distance; then, KNN is identified based on the calculated similarities, thereby ultimately determining the output label of the testing model, that is, the highest number of votes amongst the k neighbours [88,89]. RF, in contrast to SVM and KNN, can analyse highly dimensional data and handle various issues, including classification, feature selection and regression [90]. The model combines the outcomes from each decision tree in accordance with a specific combination technique to create a classifier with good generalisation abilities [89]. This technique involves reusing samples from the first training samples to create a number of subgroups. These subsets train several base classifiers, and the base classifiers' vote determines the best classification outcomes [90]. Therefore, recent research has shown that choosing RF features may significantly increase prediction accuracy [91].

The input data were the training model, that is, 70% of the dataset. It consisted of the independent variables, which were the nine related criteria endorsed by the attribute evaluator functions, whereas the dependent variables were the evaluation of WSR in the city neighbourhoods by their residents' opinion (high risk, non-risk). In addition, 30% of the dataset was used as the testing model to evaluate the training model using six indicators (i.e., precision, sensitivity (recall), specificity, F1-score, classification accuracy (CA) and area under the curve (AUC) measures) to assess how well the models performed. On the one hand, after the team of ML models was trained using the training model, it was able to predict WSR for the study area, as well as any other urban area. On the other hand, the SHAP model provided the weights of independent variables (nine criteria) and illustrated the contribution of each factor visually. Those weights were entered in WLC in GIS to produce WSR maps of the study area with another method. The average weight of the five models was used to produce the final map.

2.8. Model Validation

Validation is the most crucial step in any ML modelling process. In this study, a portion of the 30% test dataset was used to evaluate the performance of the five models, and validation results were compared using several validation indicators, such as precision, sensitivity (recall), specificity, F1-score, CA and AUC measures based on Equations (7)–(12) [92]. Generally, precision indicates the proportion of relevant instances amongst the retrieved instances; sensitivity (recall) denotes how effectively a model can detect positive instances; specificity represents how effectively it discovers negative instances; and accuracy is expected to denote how well it identifies both categories [79]. The F1-score is the harmonic mean of accuracy and recall calculated using Equation (12) [93]. AUC presents the area under the receiver operating characteristic (ROC) curve, which is determined using Equation (13). It is a useful parameter for assessing the performance of a classifier and measuring a classifier's ability to avoid false categorisation [94]. In addition, the overall performance of the models was confirmed using the ROC accuracy evaluation parameter [95].

$$\text{Precision} = \frac{Tp}{Tp + FP} \tag{7}$$

$$\text{Sensitivity (Recall)} = \frac{Tp}{TP + FN} \tag{8}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{9}$$

$$Accuracy = \frac{Tp + TN}{Tp + TN + FP + FN} \tag{10}$$

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

$$AUC = \frac{1}{2}\left(\frac{Tp}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{12}$$

T, *F*, *P* and *N* represent true, false, positive and negative. *TP, TN, FP* and *FN* are true positive, true negative, false positive and false negative.

2.9. Production of WSR Maps Using GIS

In this study, WSR maps were produced in two ways in GIS. The first way is a simple traditional method to present the prediction results of ML models directly. Thus, five maps were produced based on the result of the five algorithms used in this study. The second way is using the WLC function. It was applied six times in total, yielding six urban WSR maps. WLC aggregated the layers of criteria and their associated weights which were obtained individually from each model to produce five maps; the sixth map (the final map) was produced using the average weights based on Equation (13) [96]. Furthermore, using

the Jenks natural breaks (JNB) classification method in GIS, the maps were divided into five classifications of urban water vulnerability (very high, high, moderate, low and very low). Finally, the evaluation process between the two ways was conducted using the evaluation indicators to validate the results.

$$\text{WSR} = \sum_{i=0}^{n} c_i w_i \, (q_i), \tag{13}$$

where n represents the number of criteria, w_i represents the weight criterion (c_i), q_i is the value of a criterion (c_i) and WSR is the risk of water shortage.

3. Results

3.1. Results of Feature Reduction

The results of using the three ranking algorithms, InfoGainAttributeEval, GainRatioAttributeEval and CorrelationAttributeEval, provided insight into the degree of importance of each criterion of the 10 criteria to the output class, as shown in Table 2. However, the factors are ranked based on their average merit, which is given as a value between 0 and 1. The attributes with low merit (a threshold of 0.1) were discarded. Furthermore, the evaluation process applied, which was conducted by NB, confirmed that the performance of CorrelationAttributeEval was more accurate than the other two. As result, only one independent variable (population) was discarded.

Table 2. Performance of the three ranking algorithms using 10-fold cross-validation.

No	Attributes	InfoGain AttributeEval	GainRatio AttributeEval	Correlation AttributeEval
1	Capacity of reservoir	0.615	0.621	0.739
2	Proximity to river	0.339	0.346	0.641
3	Unmet demand	0.339	0.346	0.652
4	Proximity to pipelines	0.196	0.298	0.241
5	Age of the network pipe	0.258	0.193	0.283
6	Electricity supply	0.191	0.294	0.223
7	Proximity to PS	0.149	0.262	0.364
8	Unemployment ratio	0.011	0.024	0.312
9	Population	0	0	0.1
10	Population density	0	0	0.126

3.2. Performance of ML Models

The accuracy of the training models was 0.923, 0.923, 0.923, 0.885 and 0.846 based on SVM, MLP, RF, KNN and NB algorithms, respectively. On the one hand, the dataset (conditional factors) was entered into the models to predict the WSR level of all study areas directly. On the other hand, the SHAP model gathered the weights of conditional factors to use them in the WLC methods to produce WSR maps with another technique (Figure 5).
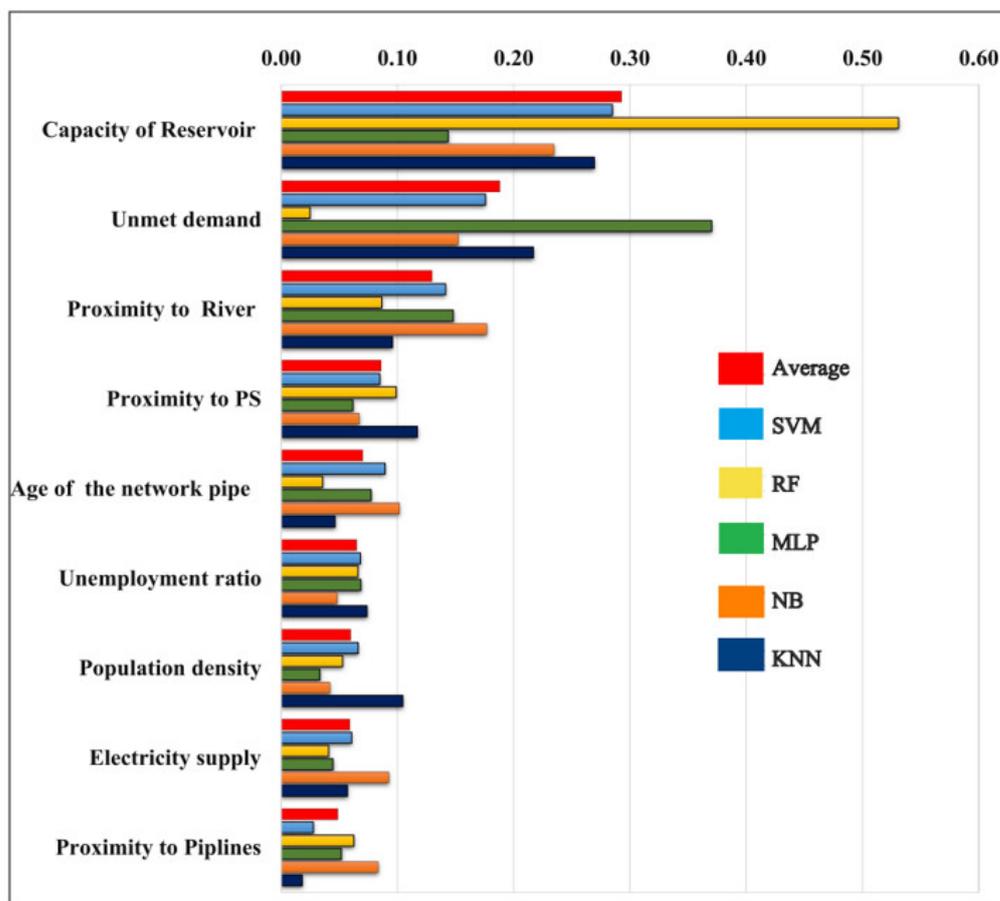
Figure 5. Weights of criteria according to model type and the average weights.

3.3. GIS Techniques Results

The results of the prediction process were directly represented using the GIS application; thus, five WSR maps were produced based on the ML model used, as illustrated in the left side of Figures 6–10. Furthermore, the WLC technique was applied five times in the GIS environment to generate five WSR maps for Nasiriyah City with another technique, as shown in right side of Figures 6–10. In these processes, the set weights of the criterion were changed according to the ML model each time. Furthermore, the classification of each map produced by WLC was evaluated using the corresponding ML algorithm. The CA was 78%, 84%, 85%, 90% and 96% for the produced maps from weights of NB, RF, KNN, MLP and SVM models, respectively (Table 3). To reach high accuracy, the average weights produced by the five ML models were applied to the WLC function to create the final map (Figure 11), the accuracy of which was higher than the others, as shown in Figure 14. According to the final map, 26% of the city's population, who are living in 17 neighbourhoods, are under very high WSR, 13%, who are living in 18 neighbourhoods are under high WSR, and the rest are under moderate to very low risk, as shown in Figure 12.

(**a**) (**b**)
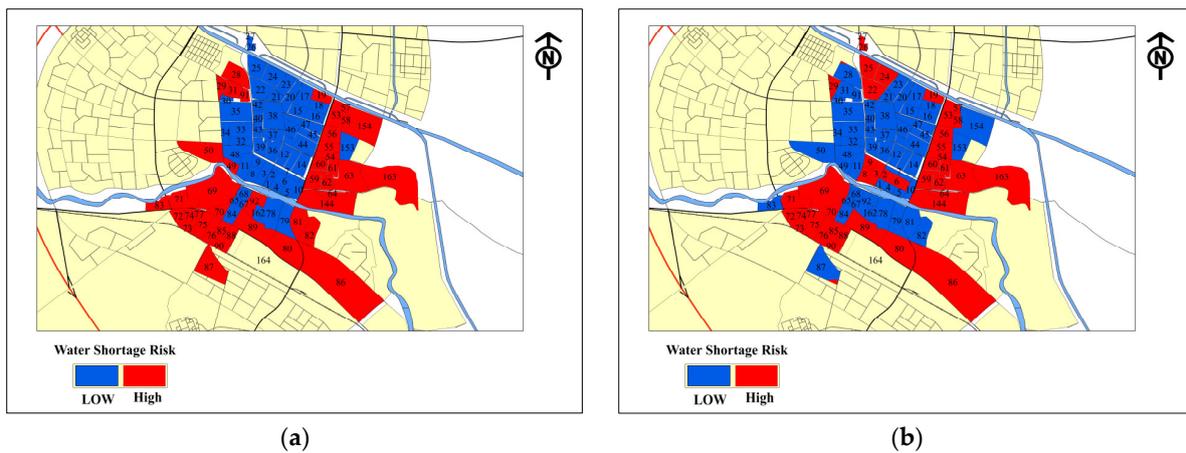
Figure 6. Map (**a**) was created by the direct representative method in GIS, using the prediction results of NB, whereas map (**b**) was produced by the WLC technique of GIS, using the criteria weights gained from NB algorithm.

(**a**) (**b**)
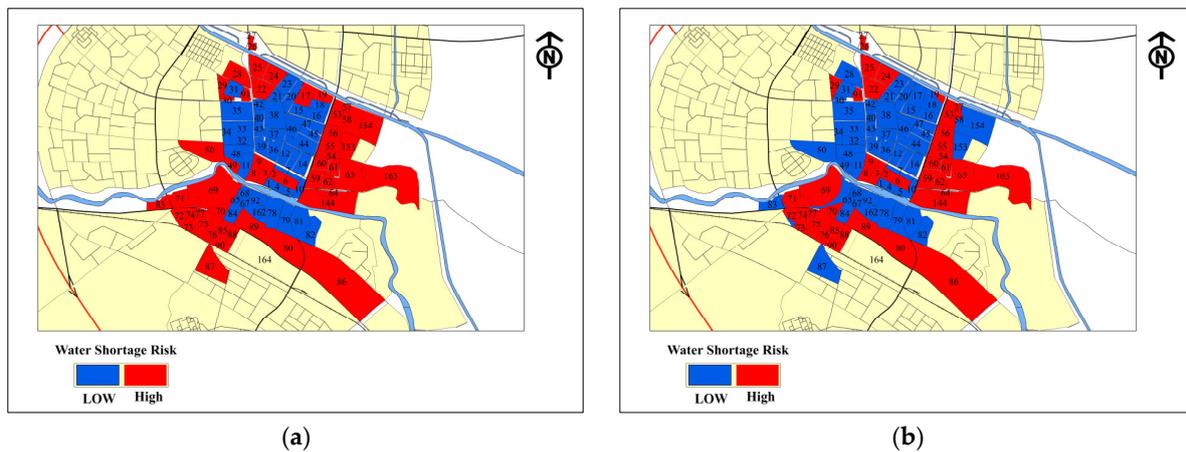
Figure 7. Map (**a**) was created by the direct representative method in GIS using the prediction results of RF, whereas map (**b**) was produced by the WLC technique of GIS, using the criteria weights extracted by RF algorithm.

(**a**) (**b**)
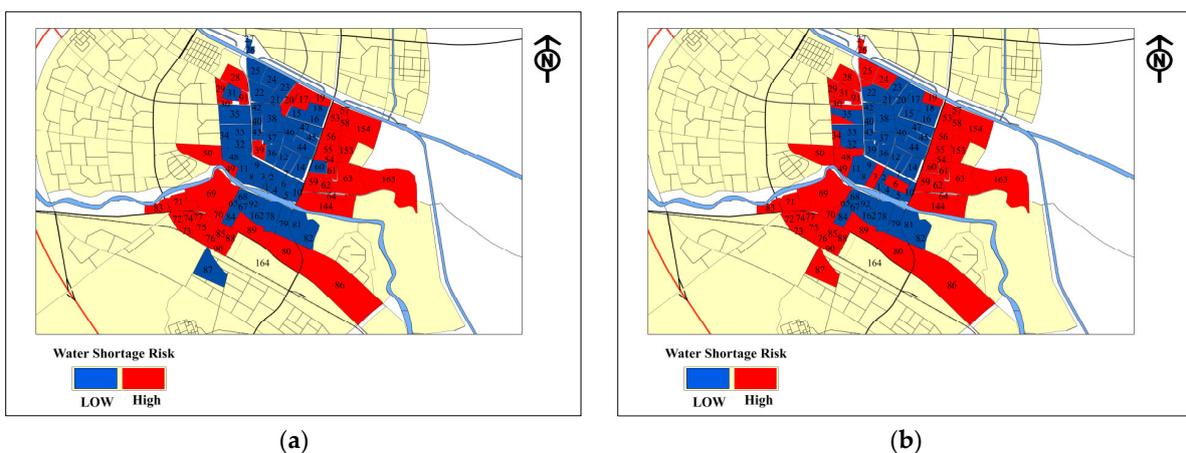
Figure 8. Map (**a**) was created by the direct representative method in GIS using the prediction results of KNN algorithm, whereas map (**b**) was produced by the WLC technique of GIS using the criteria weights gained from KNN algorithm.
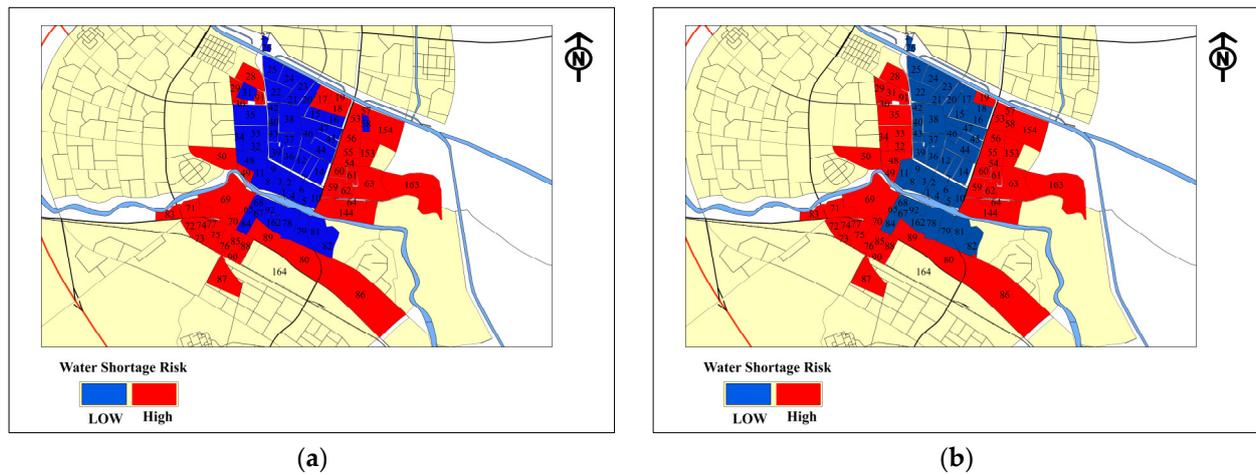
Figure 9. Map (**a**) was created by the direct representative method in GIS, using the prediction results of MLP algorithm, whereas map (**b**) was produced by the WLC technique of GIS using the criteria weights gained from MLP algorithm.
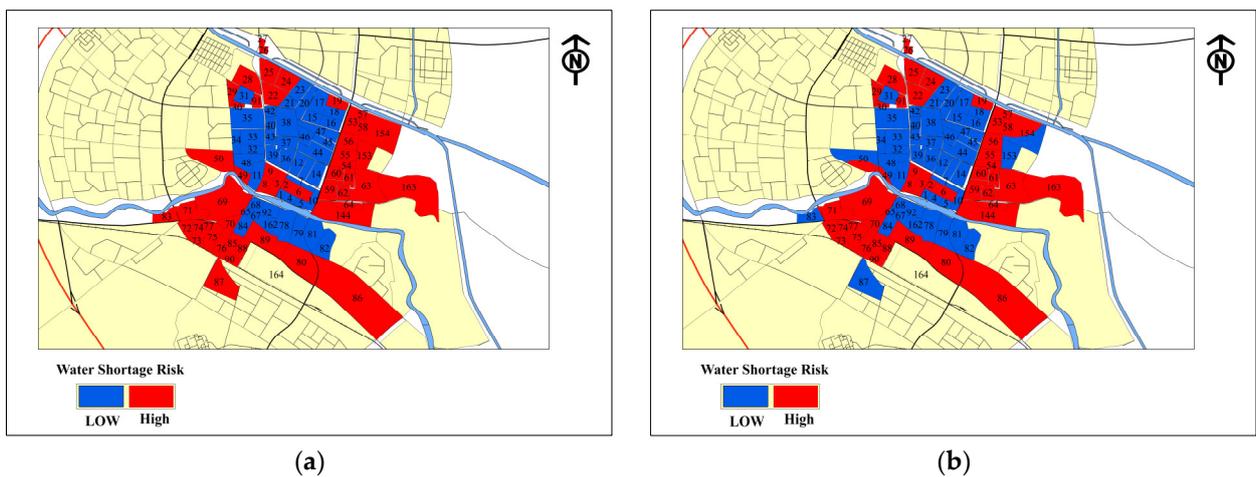

Figure 10. Map (**a**) was created by the direct representative method in GIS, whereas map (**b**) was produced by using the criteria weights gained from SVM algorithm in the WLC technique of GIS.

Table 3. Evaluation of classification of map (b) in Figures 6–10 using the corresponding ML algorithm.

Model	AUC	CA	F1	Precision	Recall
NB	0.875	0.777	0.747	0.738	0.756
RF	0.889	0.84	0.831	0.755	0.925
KNN	0.864	0.851	0.844	0.927	0.776
MLP	0.954	0.904	0.897	0.951	0.848
SVM	0.962	0.957	0.958	0.92	1
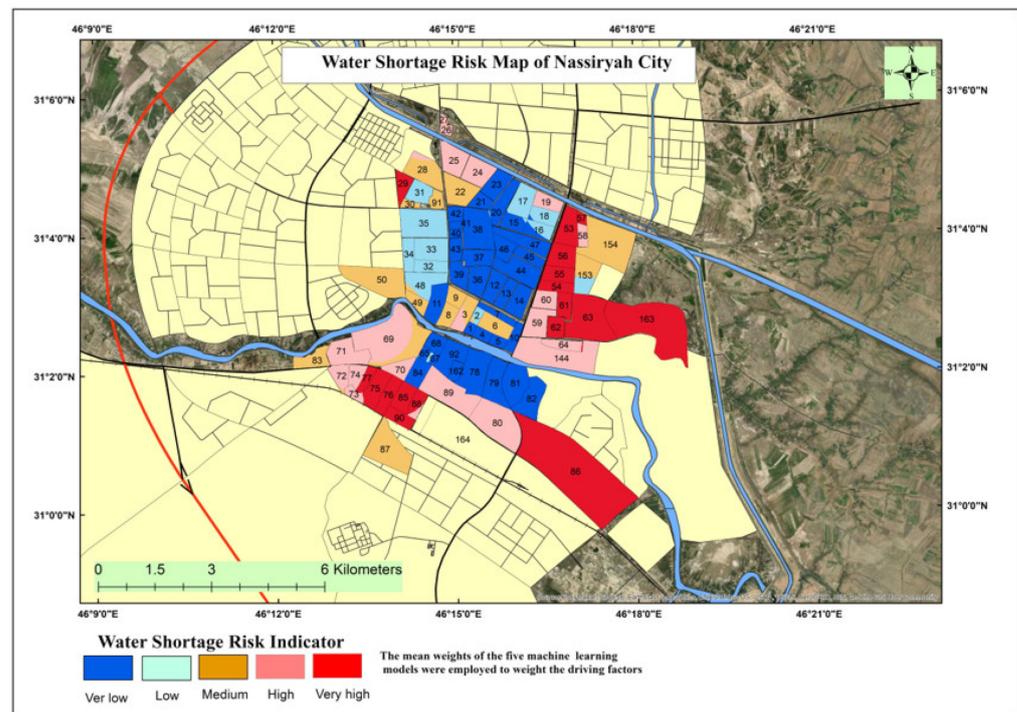
Figure 11. Final WSR map classified into five categories (Very high, High, Medium, Low and Very low).
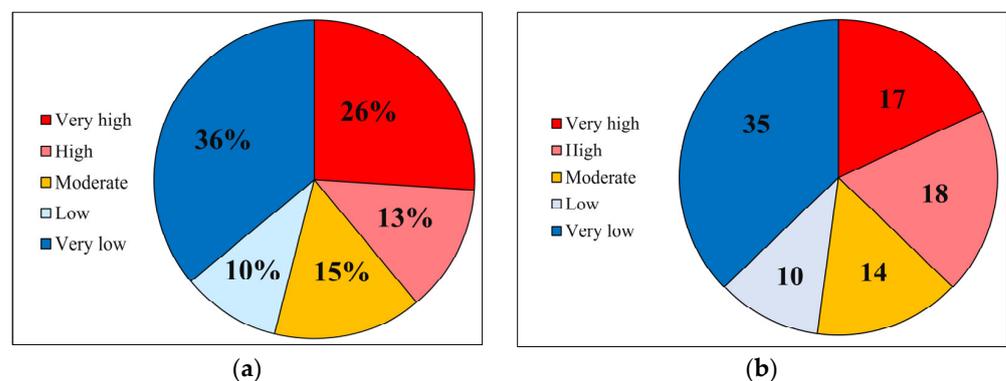
Figure 12. Statistical graphics of WSR for Nasiriyah City. (**a**) Distribution of population ratio based on WSR classes. (**b**) Distribution of city neighbourhoods based on WSR classes.
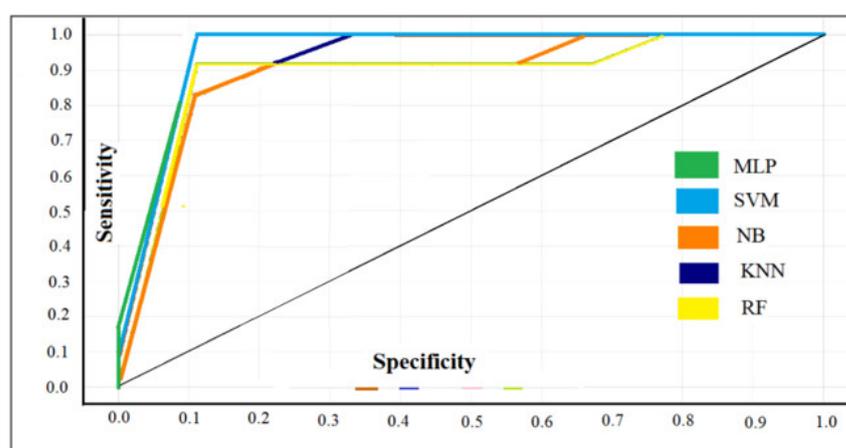
4. Validation

An evaluation process was applied in each stage of this work. Firstly, the NB algorithm was applied with the method of 10-fold cross-validation to validate the feature-redaction stage. The result demonstrated that the performance accuracy of CorrelationAttributeEval was about 93% while the accuracy of InfoGainAttributeEval and GainRati-oAt-tributeEval were 91.67% and 91.25%, respectively. Then, the trained model was evaluated using the five ML models (NB, RF, KNN, SVM and MLP) using a testing dataset via a confusion matrix. Six validation indicators were calculated: AUC, ACC, F1, precision, recall and specificity, as shown in Table 4 and Figure 13. Despite the fact that the trained model evaluated by MLP was the best in terms of AUC values, the ACC values of the five ML models proved that the dataset was almost perfect. In the next stage, the six maps produced by entering the gained weight from the ML models into the WLC technique were evaluated by applying the corresponding ML model. The CA was 78%, 84%, 85%, 90% and 96% for the produced maps from weights of NB, RF, KNN, MLP and SVM models, respectively (Table 3). Furthermore, the final map that applied the average weights of the five models
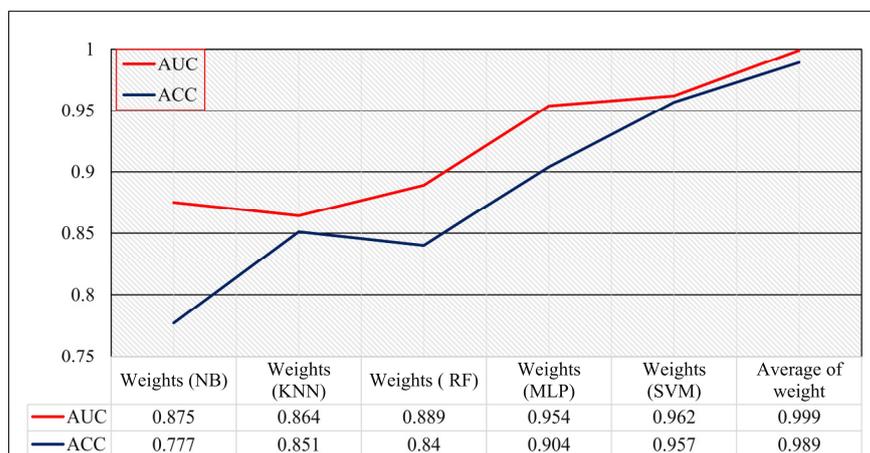
and the WLC technique achieved higher accuracy than the other five weights, as is seen in Figure 14.

Table 4. Trained model evaluation.

Model	AUC	CA	F1	Precision	Recall	Specificity
MLP	0.940	0.923	0.917	0.917	0.917	0.929
SVM	0.934	0.923	0.917	0.917	0.917	0.929
KNN	0.9116	0.885	0.869	0.909	0.833	0.929
RF	0.893	0.923	0.917	0.917	0.917	0.929
NB	0.875	0.846	0.846	0.786	0.917	0.786

Figure 13. The area under ROC curve (AUC) for the five models.

	Weights (NB)	Weights (KNN)	Weights (RF)	Weights (MLP)	Weights (SVM)	Average of weight
AUC	0.875	0.864	0.889	0.954	0.962	0.999
ACC	0.777	0.851	0.84	0.904	0.957	0.989

Figure 14. A comparative evaluation of the WSR classifications produced by the W.L.C technique in GIS using criteria weights extracted by the five models.

5. Discussion

This study highlights the water supply problem in urban areas by analysing the external and internal challenges that cause WSR. In urban areas, the external challenges are represented by climate change and quota city shares of water, whereas the internal challenges are represented by inequity in water management amongst city neighbourhoods.

Although the first challenge can be understood using traditional hydrological models, obtaining a deep understanding of the second challenge by conventional methods that

adopt experts' opinions to determine the vulnerable urban areas is difficult due to subjectivity bias. Therefore, a research question is proposed in this study, that is, Are the opinions of people who are vulnerable to WSR nearer to reality than the opinions of experts? To address this problem, this study used PP to present a novel approach in detecting urban areas that are vulnerable to WSR by integrating ML and GIS. The novel approach presented in this study demonstrated that people's opinions matched the aggregation result of the driving forces of WSR.

The results of this work revealed that the novel approach of incorporating PP with ML and GIS is an effective technique for exploring urban risks, in general, and WSR, especially. All five ML models that were used in this work indicated that the accuracy of the dataset was almost perfect. The accuracy of the dataset was 0.923, 0.923, 0.885, 0.923 and 0.846 by performing MLP, SVM, KNN, RF and NBC, respectively. Although the ML models predicted the WSR for the study area effectively, determining the real effects of the continuous factors, proximity to PS, proximity to main pipelines and proximity to rivers of the study area is difficult due to the effects of continuous factors that could not be obstructed by borders of neighbourhoods. Therefore, this study suggested that the average weights that were extracted by multiple ML models could resolve this issue [31].

This suggestion would be convincing if the result of comparison between classifications based on the ML prediction method and the WLC, which uses the average weights, were validated. However, the CA of the five maps of WSR that were produced individually from the weights of NB, RF, KNN, MLP and SVM algorithms was 78%, 84%, 85%, 90% and 96%, respectively, compared with the classification that was produced directly by model prediction. Although the CA of the map produced using the weights of the SVM model achieved a high accuracy (96%), the final WSR map using the average weights of the five models achieved 99%, as shown in Figure 11. Furthermore, this approach successfully displayed the effects of the continuous factors that could not be obstructed by the neighbourhood's borders across the study area, as shown in Figure 15. Therefore, the results endorsed the capability of the novel approach to detecting WSR with high accuracy.
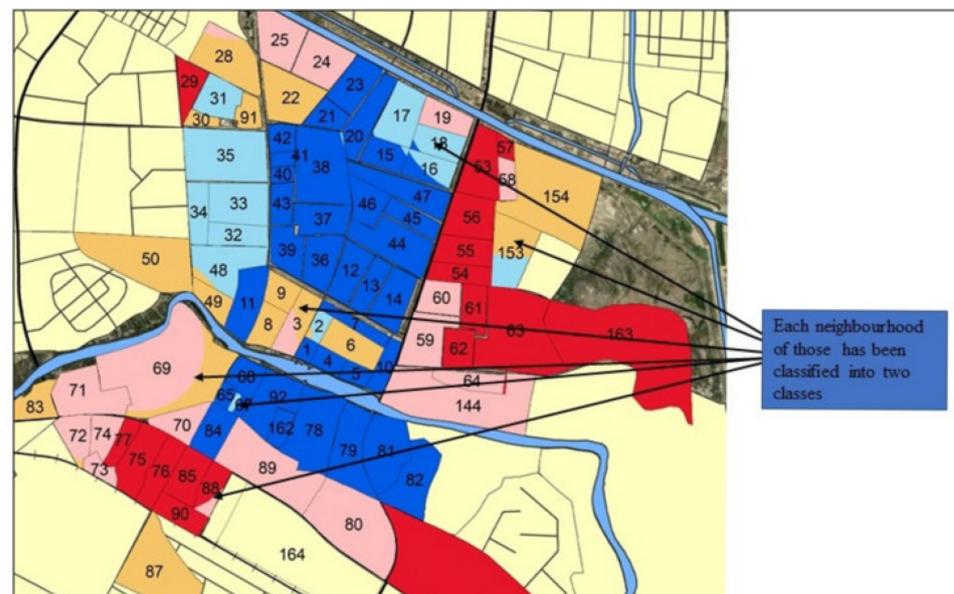
Figure 15. Zoomed-in final map, showing the effects of continuous factors on the classification of several neighbourhoods; each neighbourhood has been classified into two classes.

The analysis process of the driving factors conducted by the SHAP model showed that the criteria of the capacity of reservoirs, unmet water demand and proximity to the river gained the highest weights (29%, 19% and 12%, respectively) amongst the other criteria, as shown in Figure 5. Thus, 60% of the decision depended on only these three criteria,

whereas all six criteria contributed to only 40% of the decision. Therefore, the decision maker should consider this conclusion when the mitigation processes of WSR are taking place. However, the results demonstrated that a variation in water supply existed in this single city due to the spatial distribution of the urban WSS. The final map showed that 26% of the city's population, who are living in 17 neighbourhoods, were under very high WSR; 13%, who are living in 18 neighbourhoods, were under high WSR and the rest are under moderate to very low risk, as shown in Figure 12. This result endorsed the hypothesis of this study: that the inequality in water distribution management within the same urban area is a significant challenge leading to WSR.

Different from traditional approaches, this novel approach considered all contributing factors for supplying water, beginning from the river to the house, to analyse the actual reasons for WSR in urban areas [11,12,14]. In addition, compared with recent research, this study adopted the theoretical analysis of the mitigation of WSR by decreasing water losses; therefore, this analysis was adopted in the deriving process of the relevant criteria, such as the proximity to PS and the proximity to the river [3,17]. Furthermore, modern techniques were applied (i.e., fuzzy large membership and fuzzy near membership) in the derivation process [93,97]. However, in the risk-management field, machine-learning models are often assumed to be applicable to similar spatial environments of interest. Therefore, the proposed training model was built from the spatial analysis of essential components of a conventional water supply system (WSS) which is usually used by most countries worldwide. The training model included several independent variables, namely, 10 spatial driving factors that reflected the roles of WSS components, which run from water sources to homes, and a dependent variable, that is, the water situation of each sector of an urban area. In addition, the quality of the model's execution was improved using three types of attribute evaluators to remove irrelevant driving factors and apply the best prediction model. Therefore, the model was carefully organized to achieve high accuracy in other water-scarcity risk-estimation studies in similar spatial environments.

The proposed approach can be generalised to other fields to support decision makers in reducing the degree of vulnerability in urban areas in developed and developing countries. However, the accuracy of the required data is difficult to guarantee in this study. On the one hand, the WSS within a city is complex. On the other hand, the PP approach in developing countries faces many limitations in collecting people's opinions regarding the main service provided by the government. Although the uncertainty ratio was associated with the results due to the accuracy of collected data, the novel approach interpreted the results in a clear form, which allows decision makers to support the proactive exploration of urban WSR and select appropriate alternatives at the right time.

6. Conclusions

Understanding urban water security using traditional methods is difficult due to the wide range of measurement indicators, including drinking-water quality and quantity, human needs, socio-economic indicators and water-linked risks. Therefore, the key aim of the present work was to develop a novel approach based on PP and ML in the domain of urban water management. For this purpose, GIS coupled with six typical ML models, namely, SVM, MLP, KNN, RF, NB and SHAP, were used. The findings confirmed that, in general, the novel approach is effective in detecting WSR in urban area. This experiment approved the significant role of PP in the mitigation process regarding WSR in urban communities. Furthermore, the findings of this research suggest that the average weights of conditional factors extracted by the five ML models are the best in terms of time, cost and accuracy compared with traditional methods that adopt MCDA and ML in the domain of urban risks. Therefore, this study adds a novel approach to the urban water management domain, which can be used to support proactive decisions to mitigate risks. However, the access to the required data is a significant challenge for this study due to the restricted regulations of the local management of WSS. Further study could develop this approach to determine WSR at a street scale instead of a neighbourhood level. Water considerably

impacts life maintenance, environmental health, and even cultural innovation; thus, people worldwide are invited to exert extraordinary efforts to address water scarcity. Continued efforts are also required from governments to engage the public in making decisions about risk prevention.

Author Contributions: Conceptualization, Sadeq Khaleefah Hanoon and Ahmad Fikri Abdullah; methodology, Sadeq Khaleefah Hanoon; software, Sadeq Khaleefah Hanoon; validation, Sadeq Khaleefah Hanoon, Ahmad Fikri Abdullah, Helmi Z. M. Shafri and Aimrun Wayayok; formal analysis, Sadeq Khaleefah Hanoon; investigation, Sadeq Khaleefah Hanoon and Ahmad Fikri Abdullah; resources, Sadeq Khaleefah Hanoon; data curation, Sadeq Khaleefah Hanoon; writing—original draft preparation, Sadeq Khaleefah Hanoon; writing—review and editing, Sadeq Khaleefah Hanoon and Ahmad Fikri Abdullah; supervision, Sadeq Khaleefah Hanoon, Ahmad Fikri Abdullah, Helmi Z. M. Shafri and Aimrun Wayayok. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are reported in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Johannessen, O.M.; Shalina, E.V. Population increase impacts the climate, using the sensitive Arctic as an example. *Atmospheric Ocean. Sci. Lett.* **2022**, *15*, 100192. [CrossRef]
2. Gesualdo, G.C.; Sone, J.S.; Galvão, C.D.O.; Martins, E.S.; Montenegro, S.M.G.L.; Tomasella, J.; Mendiondo, E.M. Unveiling water security in Brazil: Current challenges and future perspectives. *Hydrol. Sci. J.* **2021**, *66*, 759–768. [CrossRef]
3. Ahmadi, M.S.; Sušnik, J.; Veerbeek, W.; Zevenbergen, C. Towards a global day zero? Assessment of current and future water supply and demand in 12 rapidly developing megacities. *Sustain. Cities Soc.* **2020**, *61*, 102295. [CrossRef]
4. Luo, P.; Sun, Y.; Wang, S.; Wang, S.; Lyu, J.; Zhou, M.; Nakagami, K.; Takara, K.; Nover, D. Historical assessment and future sustainability challenges of Egyptian water resources management. *J. Clean. Prod.* **2020**, *263*, 121154. [CrossRef]
5. Rafiei, F.; Gharechelou, S.; Golian, S.; Johnson, B.A. Aquifer and Land Subsidence Interaction Assessment Using Sentinel-1 Data and DInSAR Technique. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 495. [CrossRef]
6. Albarqouni, M.M.Y.; Yagmur, N.; Balcik, F.B.; Sekertekin, A. Assessment of Spatio-Temporal Changes in Water Surface Extents and Lake Surface Temperatures Using Google Earth Engine for Lakes Region, Türkiye. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 407. [CrossRef]
7. Hanoon, S.K.; Abdullah, A.F.; Shafri, H.Z.M.; Wayayok, A. Using Supervised Classification technique to monitor hydrological systems of Mesopotamia marshes in Dhi-Qar province (Iraq). In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 6189–6192.
8. Chapagain, K.; Aboelnga, H.T.; Babel, M.S.; Ribbe, L.; Shinde, V.R.; Sharma, D.; Dang, N.M. Urban water security: A comparative assessment and policy analysis of five cities in diverse developing countries of Asia. *Environ. Dev.* **2022**, *43*, 100713. [CrossRef]
9. Asghar, A.; Iqbal, J.; Amin, A.; Ribbe, L. Integrated hydrological modeling for assessment of water demand and supply under socio-economic and IPCC climate change scenarios using WEAP in Central Indus Basin. *J. Water Supply Res. Technol.* **2019**, *68*, 136–148. [CrossRef]
10. Yousuf, M.A.; Rapantova, N.; Younis, J.H. Sustainable Water Management in Iraq (Kurdistan) as a Challenge for Governmental Responsibility. *Water* **2018**, *10*, 1651. [CrossRef]
11. Ougougdal, H.A.; Khebiza, M.Y.; Messouli, M.; Lachir, A. Assessment of futurewater demand and supply under IPCC climate change and socio-economic scenarios, using a combination of models in Ourika watershed, High Atlas, Morocco. *Water* **2020**, *12*, 1751. [CrossRef]
12. Amin, A.; Iqbal, J.; Asghar, A.; Ribbe, L. Analysis of current and futurewater demands in the Upper Indus Basin under IPCC climate and socio-economic scenarios using a hydro-economic WEAP Model. *Water* **2018**, *10*, 537. [CrossRef]
13. Alwan, I.A.; Aziz, N.A.; Hamoodi, M.N. Potential Water Harvesting Sites Identification Using Spatial Multi-Criteria Evaluation in Maysan Province, Iraq. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 235. [CrossRef]
14. de Sousa Cordão, M.J.; Rufino, I.A.A.; Barros Ramalho Alves, P.; Barros Filho, M.N.M. Water shortage risk mapping: A GIS-MCDA approach for a medium-sized city in the Brazilian semi-arid region. *Urban Water J.* **2020**, *17*, 642–655. [CrossRef]
15. Al-Juaidi, A.E.; Al-Shotairy, A.S. Evaluation of municipal water supply system options using water evaluation and planning system (WEAP): Jeddah case study. *Desalination Water Treat.* **2020**, *176*, 317–323. [CrossRef]

16. Chini, C.M.; Stillwell, A.S. One Model Does Not Fit All: Bottom-Up Indicators of Residential Water Use Provide Limited Explanation of Urban Water Fluxes. *J. Sustain. Water Built Environ.* **2020**, *6*, 04020011. [CrossRef]

17. Rufino, I.; Djordjević, S.; de Brito, H.C.; Alves, P.B.R. Multi-temporal built-up grids of brazilian cities: How trends and dynamic modelling could help on resilience challenges? *Sustainability* **2021**, *13*, 748. [CrossRef]

18. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamowski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L.; et al. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *J. Hydrol.* **2019**, *573*, 311–323. [CrossRef]

19. Sachit, M.S.; Shafri, H.Z.M.; Abdullah, A.F.; Rafie, A.S.M.; Gibril, M.B.A. Global Spatial Suitability Mapping of Wind and Solar Systems Using an Explainable AI-Based Approach. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 422. [CrossRef]

20. Martyn, K.; Kadziński, M. Deep preference learning for multiple criteria decision analysis. *Eur. J. Oper. Res.* **2023**, *305*, 781–805. [CrossRef]

21. Roy, A.; Kar, B. A multicriteria decision analysis framework to measure equitable healthcare access during COVID-19. *J. Transp. Heal.* **2022**, *24*, 101331. [CrossRef]

22. Hanoon, S.K.; Abdullah, A.F.; Shafri, H.Z.M.; Wayayok, A. Using scenario modelling for adapting to urbanization and water scarcity: Towards a sustainable city in semi-arid areas. *Period. Eng. Nat. Sci. (PEN)* **2021**, *10*, 518–532. [CrossRef]

23. Almansi, K.Y.; Shariff, A.R.M.; Abdullah, A.F.; Ismail, S.N.S. Hospital Site Assessment Using Three Machine Learning Approaches: Evidence from the Gaza Strip in Palestine. *Appl. Sci.* **2021**, *11*, 11054. [CrossRef]

24. Kumar, P.; Sharma, R.; Bhaumik, S. MCDA techniques used in optimization of weights and ratings of DRASTIC model for groundwater vulnerability assessment. *Data Sci. Manag.* **2022**, *5*, 28–41. [CrossRef]

25. Nachappa, T.G.; Piralilou, S.T.; Gholamnia, K.; Ghorbanzadeh, O.; Rahmati, O.; Blaschke, T. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *J. Hydrol.* **2020**, *590*, 125275. [CrossRef]

26. Iban, M.C.; Sekertekin, A. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecol. Inform.* **2022**, *69*, 101647. [CrossRef]

27. Wang, S.; Peng, H.; Hu, Q.; Jiang, M. Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101139. [CrossRef]

28. Haltofova, B. Using Crowdsourcing To Support Civic Engagement in Strategic Urban Development Planning: A Case Study of Ostrava, Czech Republic. *J. Competitiveness* **2018**, *10*, 85–103. [CrossRef]

29. Aubert, A.H.; Esculier, F.; Lienert, J. Recommendations for online elicitation of swing weights from citizens in environmental decision-making. *Oper. Res. Perspect.* **2020**, *7*, 100156. [CrossRef]

30. Bugs, G.; Granell, C.; Fonts, O.; Huerta, J.; Painho, M. An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil. *Cities* **2010**, *27*, 172–181. [CrossRef]

31. Li, W.; Shi, Y.; Huang, F.; Hong, H.; Song, G. Uncertainties of Collapse Susceptibility Prediction Based on Remote Sensing and GIS: Effects of Different Machine Learning Models. *Front. Earth Sci.* **2021**, *9*, 731058. [CrossRef]

32. Sachit, M.S.; Shafri, H.Z.M.; Abdullah, A.F.; Rafie, A.S.M. Combining Re-Analyzed Climate Data and Landcover Products to Assess the Temporal Complementarity of Wind and Solar Resources in Iraq. *Sustainability* **2022**, *14*, 388. [CrossRef]

33. Millington, N. Producing water scarcity in São Paulo, Brazil: The 2014-2015 water crisis and the binding politics of infrastructure. *Politi-Geogr.* **2018**, *65*, 26–34. [CrossRef]

34. Simukonda, K.; Farmani, R.; Butler, D. Intermittent water supply systems: Causal factors, problems and solution options. *Urban Water J.* **2018**, *15*, 488–500. [CrossRef]

35. Noori, A.; Bonakdari, H.; Hassaninia, M.; Morovati, K.; Khorshidi, I.; Noori, A.; Gharabaghi, B. A reliable GIS-based FAHP-FTOPSIS model to prioritize urban water supply management scenarios: A case study in semi-arid climate. *Sustain. Cities Soc.* **2022**, *81*, 103846. [CrossRef]

36. Krueger, E.H.; Rao, P.S.C.; Borchardt, D. Quantifying urban water supply security under global change. *Glob. Environ. Chang.* **2019**, *56*, 66–74. [CrossRef]

37. Liu, Y.; Tan, Q.; Zhang, X.; Han, J.; Guo, M. How does electricity supply mode affect energy-water-emissions nexus in urban energy system? Evidence from energy transformation in Beijing, China. *J. Clean. Prod.* **2022**, *366*, 132892. [CrossRef]

38. Dallison, R.J.; Patil, S.D.; Williams, A.P. Impacts of climate change on future water availability for hydropower and public water supply in Wales, UK. *J. Hydrol. Reg. Stud.* **2021**, *36*, 100866. [CrossRef]

39. Mena, D.; Solera, A.; Restrepo, L.; Pimiento, M.; Cañón, M.; Duarte, F. An analysis of unmet water demand under climate change scenarios in the Gualí river basin, Colombia, through the implementation of hydro-bid and weap hydrological modeling tools. *J. Water Clim. Chang.* **2021**, *12*, 185–200. [CrossRef]

40. Savari, M.; Mombeni, A.S.; Izadi, H. Socio-psychological determinants of Iranian rural households' adoption of water consumption curtailment behaviors. *Sci. Rep.* **2022**, *12*, 1–12. [CrossRef]

41. Wee, S.Y.; Aris, A.Z.; Yusoff, F.M.; Praveena, S.M.; Harun, R. Drinking water consumption and association between actual and perceived risks of endocrine disrupting compounds. *npj Clean Water* **2022**, *5*, 1–10. [CrossRef]

42. Zhuang, J.; Sela, L. Impact of Emerging Water Savings Scenarios on Performance of Urban Water Networks. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04019063. [CrossRef]

43. Eryiğit, M.; Sulaiman, S.O. Specifying optimum water resources based on cost-benefit relationship for settlements by artificial immune systems: Case study of Rutba City, Iraq. *Water Supply* **2022**, *22*, 5873–5881. [CrossRef]

44. Vicente, D.J.; Garrote, L.; Sánchez, R.; Santillán, D. Pressure Management in Water Distribution Systems: Current Status, Proposals, and Future Trends. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04015061. [CrossRef]

45. Zyoud, S.H.; Kaufmann, L.G.; Shaheen, H.; Samhan, S.; Fuchs-Hanusch, D. A framework for water loss management in developing countries under fuzzy environment: Integration of Fuzzy AHP with Fuzzy TOPSIS. *Expert Syst. Appl.* **2016**, *61*, 86–105. [CrossRef]

46. Hashemi, S.; Filion, Y.; Speight, V.; Long, A. Effect of Pipe Size and Location on Water-Main Head Loss in Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2020**, *146*, 06020006. [CrossRef]

47. Latif, J.; Shakir, M.Z.; Edwards, N.; Jaszczykowski, M.; Ramzan, N.; Edwards, V. Review on condition monitoring techniques for water pipelines. *Measurement* **2022**, *193*, 110895. [CrossRef]

48. Zielina, M.; Bielski, A.; Młyńska, A. Leaching of chromium and lead from the cement mortar lining into the flowing drinking water shortly after pipeline rehabilitation. *J. Clean. Prod.* **2022**, *362*, 132512. [CrossRef]

49. Moossa, B.; Trivedi, P.; Saleem, H.; Zaidi, S.J. Desalination in the GCC countries- a review. *J. Clean. Prod.* **2022**, *357*, 131717. [CrossRef]

50. Pissarra, T.C.T.; Fernandes, L.F.S.; Pacheco, F.A.L. Production of clean water in agriculture headwater catchments: A model based on the payment for environmental services. *Sci. Total Environ.* **2021**, *785*, 147331. [CrossRef] [PubMed]

51. Yan, X.; Jie, W.; Minjun, S.; Shouyang, W.; Zhuoying, Z. China's regional imbalance in electricity demand, power and water pricing—From the perspective of electricity-related virtual water transmission. *Energy* **2022**, *257*, 42–50. [CrossRef]

52. Bakhtiari, P.H.; Nikoo, M.R.; Izady, A.; Talebbeydokhti, N. A coupled agent-based risk-based optimization model for integrated urban water management. *Sustain. Cities Soc.* **2020**, *53*, 101922. [CrossRef]

53. Wang, Z.-H.; von Gnechten, R.; Sampson, D.A.; White, D.D. Wastewater Reclamation Holds a Key for Water Sustainability in Future Urban Development of Phoenix Metropolitan Area. *Sustainability* **2019**, *11*, 3537. [CrossRef]

54. Chandra, G.; Chakraborty, M.; Sinha, A. Framework for Assessing Efficient Water Consumption Attributes and Their Relative Importance in Office Complexes. *Asian J. Water, Environ. Pollut.* **2017**, *14*, 65–74. [CrossRef]

55. Diao, K.; Sitzenfrei, R.; Rauch, W. The Impacts of Spatially Variable Demand Patterns on Water Distribution System Design and Operation. *Water* **2019**, *11*, 567. [CrossRef]

56. Chen, X.; Li, F.; Li, X.; Hu, Y.; Hu, P. Evaluating and mapping water supply and demand for sustainable urban ecosystem management in Shenzhen, China. *J. Clean. Prod.* **2019**, *251*, 119754. [CrossRef]

57. Van Leeuwen, C.J. Water governance and the quality of water services in the city of Melbourne. *Urban Water J.* **2017**, *14*, 247–254. [CrossRef]

58. Valencia, A.; Qiu, J.; Chang, N.-B. Integrating sustainability indicators and governance structures via clustering analysis and multicriteria decision making for an urban agriculture network. *Ecol. Indic.* **2022**, *142*, 109237. [CrossRef]

59. Kuntiyawichai, K.; Wongsasri, S. Assessment of Drought Severity and Vulnerability in the Lam Phaniang River Basin, Thailand. *Water* **2021**, *13*, 2743. [CrossRef]

60. Wang, L.; Jin, G.; Xiong, X.; Zhang, H.; Wu, K. Object-Based Automatic Mapping of Winter Wheat Based on Temporal Phenology Patterns Derived from Multitemporal Sentinel-1 and Sentinel-2 Imagery. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 424. [CrossRef]

61. Oliver, M.A.; Webster, R. Kriging: A method of interpolation for geographical information systems. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 312–332. [CrossRef]

62. Li, Z.; Chen, H.; Yan, W. Exploring Spatial Distribution of Urban Park Service Areas in Shanghai Based on Travel Time Estimation: A Method Combining Multi-Source Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 608. [CrossRef]

63. Baalousha, H.; Tawabini, B.; Seers, T. Fuzzy or Non-Fuzzy? A Comparison between Fuzzy Logic-Based Vulnerability Mapping and DRASTIC Approach Using a Numerical Model. A Case Study from Qatar. *Water* **2021**, *13*, 1288. [CrossRef]

64. Arab, S.T.; Ahamed, T. Land Suitability Analysis for Potential Vineyards Extension in Afghanistan at Regional Scale Using Remote Sensing Datasets. *Remote Sens.* **2022**, *14*, 4450. [CrossRef]

65. Liu, J.; Li, Y.; Xiao, B.; Jiao, J. Coupling Fuzzy Multi-Criteria Decision-Making and Clustering Algorithm for MSW Landfill Site Selection (Case Study: Lanzhou, China). *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 403. [CrossRef]

66. Kritikos, T.; Davies, T. Assessment of rainfall-generated shallow landslide/debris-flow susceptibility and runout using a GIS-based approach: Application to western Southern Alps of New Zealand. *Landslides* **2015**, *12*, 1051–1075. [CrossRef]

67. Rahman, M.; Szabó, G. Sustainable Urban Land-Use Optimization Using GIS-Based Multicriteria Decision-Making (GIS-MCDM) Approach. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 313. [CrossRef]

68. Salvati, P.; Ardizzone, F.; Cardinali, M.; Fiorucci, F.; Fugnoli, F.; Guzzetti, F.; Marchesini, I.; Rinaldi, G.; Rossi, M.; Santangelo, M.; et al. Acquiring vulnerability indicators to geo-hydrological hazards: An example of mobile phone-based data collection. *Int. J. Disaster Risk Reduct.* **2021**, *55*, 102087. [CrossRef]

69. Chou, J.-S.; Truong, D.-N.; Tsai, C.-F. Solving Regression Problems with Intelligent Machine Learner for Engineering Informatics. *Mathematics* **2021**, *9*, 686. [CrossRef]

70. Güneş, S.S.; Yeşil, Ç.; Gurdal, E.E.; Korkmaz, E.E.; Yarım, M.; Aydın, A.; Sipahi, H. Primum non nocere: In silico prediction of adverse drug reactions of antidepressant drugs. *Comput. Toxicol.* **2021**, *18*, 100165. [CrossRef]

71. Gao, W.; Alsarraf, J.; Moayedi, H.; Shahsavar, A.; Nguyen, H. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. *Appl. Soft Comput.* **2019**, *84*, 105748. [CrossRef]
72. Albagmi, F.M.; Alansari, A.; Al Shawan, D.S.; AlNujaidi, H.Y.; Olatunji, S.O. Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach. *Inform. Med. Unlocked* **2022**, *28*, 100854. [CrossRef] [PubMed]
73. Sindhu, S.S.S.; Geetha, S.; Kannan, A. Decision tree based light weight intrusion detection using a wrapper approach. *Expert Syst. Appl.* **2012**, *39*, 129–141. [CrossRef]
74. Amirruddin, A.D.; Muharam, F.M.; Ismail, M.H.; Tan, N.P. Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (*Elaeis guineensis*) using spectroradiometers and unmanned aerial vehicles. *Comput. Electron. Agric.* **2022**, *193*, 106646. [CrossRef]
75. Otchere, D.A.; Ganat, T.O.A.; Gholami, R.; Ridha, S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *J. Pet. Sci. Eng.* **2021**, *200*, 108182. [CrossRef]
76. Almeida, V.G.; Vieira, J.; Santos, P.; Pereira, T.; Pereira, H.C.; Correia, C.; Pego, M.; Cardoso, J. Machine Learning Techniques for Arterial Pressure Waveform Analysis. *J. Pers. Med.* **2013**, *3*, 82–101. [CrossRef] [PubMed]
77. Nguyen, D.D.; Roussis, P.C.; Pham, B.T.; Ferentinou, M.; Mamou, A.; Vu, D.Q.; Bui, Q.-A.T.; Trong, D.K.; Asteris, P.G. Bagging and Multilayer Perceptron Hybrid Intelligence Models Predicting the Swelling Potential of Soil. *Transp. Geotech.* **2022**, *36*, 100797. [CrossRef]
78. Bansal, D.; Chhikara, R.; Khanna, K.; Gupta, P. Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia. *Procedia Comput. Sci.* **2018**, *132*, 1497–1502. [CrossRef]
79. Ajdadi, F.R.; Gilandeh, Y.A.; Mollazade, K.; Hasanzadeh, R.P. Application of machine vision for classification of soil aggregate size. *Soil Tillage Res.* **2016**, *162*, 8–17. [CrossRef]
80. Zarei, R.; He, J.; Siuly, S.; Zhang, Y. A PCA aided cross-covariance scheme for discriminative feature extraction from EEG signals. *Comput. Methods Programs Biomed.* **2017**, *146*, 47–57. [CrossRef]
81. Azoff, E.M. Reducing error in neural network time series forecasting. *Neural Comput. Appl.* **1993**, *1*, 240–247. [CrossRef]
82. Mehmood, Z.; Asghar, S. Customizing SVM as a base learner with AdaBoost ensemble to learn from multi-class problems: A hybrid approach AdaBoost-MSVM. *Knowl.-Based Syst.* **2021**, *217*, 106845. [CrossRef]
83. Güven, I.; Şimşir, F. Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Comput. Ind. Eng.* **2020**, *147*, 106678. [CrossRef]
84. Gâlmeanu, H.; Andonie, R. Weighted Incremental–Decremental Support Vector Machines for concept drift with shifting window. *Neural Networks* **2022**, *152*, 528–541. [CrossRef] [PubMed]
85. Balaji, V.; Suganthi, S.; Rajadevi, R.; Kumar, V.K.; Balaji, B.S.; Pandiyan, S. Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier. *Measurement* **2020**, *163*, 107922. [CrossRef]
86. Yao, J.; Ye, Y. The effect of image recognition traffic prediction method under deep learning and naive Bayes algorithm on freeway traffic safety. *Image Vis. Comput.* **2020**, *103*, 103971. [CrossRef]
87. Sun, X.; Opulencia, M.J.C.; Alexandrovich, T.P.; Khan, A.; Algarni, M.; Abdelrahman, A. Modeling and optimization of vegetable oil biodiesel production with heterogeneous nano catalytic process: Multi-layer perceptron, decision regression tree, and K-Nearest Neighbor methods. *Environ. Technol. Innov.* **2022**, *27*, 102794. [CrossRef]
88. Gou, J.; Sun, L.; Du, L.; Ma, H.; Xiong, T.; Ou, W.; Zhan, Y. A representation coefficient-based k-nearest centroid neighbor classifier. *Expert Syst. Appl.* **2022**, *194*, 116529. [CrossRef]
89. Feng, T.; Wang, C.; Zhang, J.; Wang, B.; Jin, Y.-F. An improved artificial bee colony-random forest (IABC-RF) model for predicting the tunnel deformation due to an adjacent foundation pit excavation. *Undergr. Space* **2022**, *7*, 514–527. [CrossRef]
90. Dai, Y.; Wang, Y.; Leng, M.; Yang, X.; Zhou, Q. LOWESS smoothing and Random Forest based GRU model: A short-term photovoltaic power generation forecasting method. *Energy* **2022**, *256*, 124661. [CrossRef]
91. Yagoub, M.M.; Tesfaldet, Y.T.; Elmubarak, M.G.; Al Hosani, N. Extraction of Urban Quality of Life Indicators Using Remote Sensing and Machine Learning: The Case of Al Ain City, United Arab Emirates (UAE). *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 458. [CrossRef]
92. Onah, J.O.; Abdulhamid, S.M.; Abdullahi, M.; Hassan, I.H.; Al-Ghusham, A. Genetic Algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment. *Mach. Learn. Appl.* **2021**, *6*, 100156. [CrossRef]
93. Kordnoori, S.; Mostafaei, H.; Rostamy-Malkhalifeh, M.; Ostadrahimi, M. Diagnosis of Heart Disease Using Feature Selection Methods Based on Recurrent Fuzzy Neural Networks. *IPTEK J. Technol. Sci.* **2021**, *32*, 64. [CrossRef]
94. Hashem, A.; Awad, A.; Shousha, H.; Alakel, W.; Salama, A.; Awad, T.; Mabrouk, M. Validation of a machine learning approach using FIB-4 and APRI scores assessed by the metavir scoring system: A cohort study. *Arab J. Gastroenterol.* **2021**, *22*, 88–92. [CrossRef]
95. Lee, D.-H.; Kim, Y.-T.; Lee, S.-R. Shallow Landslide Susceptibility Models Based on Artificial Neural Networks Considering the Factor Selection Method and Various Non-Linear Activation Functions. *Remote Sens.* **2020**, *12*, 1194. [CrossRef]
96. Malczewski, J. GIS-based land-use suitability analysis: A critical overview. *Prog. Plan.* **2004**, *62*, 3–65. [CrossRef]
97. Hanoon, S.K.; Abdullah, A.F.; Shafri, H.Z.M.; Wayayok, A. Comprehensive Vulnerability Assessment of Urban Areas Using an Integration of Fuzzy Logic Functions: Case Study of Nasiriyah City in South Iraq. *Earth* **2022**, *3*, 699–732. [CrossRef]