MDPI

Article

A GIS-Based Artificial Neural Network Model for Flood Susceptibility Assessment

Nanda Khoirunisa, Cheng-Yu Ku * and Chih-Yu Liu *

Department of Harbor and River Engineering, National Taiwan Ocean University, Keelung City 20224, Taiwan;
10786013@mail.ntou.edu.tw
* Correspondence: chkst26@mail.ntou.edu.tw (C.-Y.K.); 20452003@email.ntou.edu.tw (C.-Y.L.)

Abstract: This article presents a geographic information system (GIS)-based artificial neural network (GANN) model for flood susceptibility assessment of Keelung City, Taiwan. Various factors, including elevation, slope angle, slope aspect, flow accumulation, flow direction, topographic wetness index (TWI), drainage density, rainfall, and normalized difference vegetation index, were generated using a digital elevation model and LANDSAT 8 imagery. Historical flood data from 2015 to 2019, including 307 flood events, were adopted for a comparison of flood susceptibility. Using these factors, the GANN model, based on the back-propagation neural network (BPNN), was employed to provide flood susceptibility. The validation results indicate that a satisfactory result, with a correlation coefficient of 0.814, was obtained. A comparison of the GANN model with those from the SOBEK model was conducted. The comparative results demonstrated that the proposed method can provide good accuracy in predicting flood susceptibility. The results of flood susceptibility are categorized into five classes: Very low, low, moderate, high, and very high, with coverage areas of 60.5%, 27.4%, 8.6%, 2.5%, and 1%, respectively. The results demonstrate that nearly 3.5% of the study area, including the core district of the city and an exceedingly populated area including the financial center of the city, can be categorized as high to very high flood susceptibility zones.

Keywords: geographic information system; back-propagation neural network; rainfall; historical flood; prediction

1. Introduction

Coastal areas are vulnerable to climate change, particularly sea-level rises and floods related to extreme rainfall [1,2]. Taiwan is an island which is prone to flood disasters triggered by heavy rainfall and typhoons every year. In Taiwan, extreme weather conditions, such as heavy precipitation and typhoons generated by climate change, strengthen the phenomenon of flood disasters [3]. Keelung City, one of the coastal cities in Northern Taiwan, has become highly urbanized and densely populated in recent years [4]. Flooding events have frequently occurred in the past because typhoons and rainstorms typically sweep over the upstream basins between May and October; this trend is expected to increase in the future [5–7].

Various approaches have been proposed to evaluate flood disaster risk based on the susceptibility of the system and hydrology [8–10]. The artificial neural network (ANN) is one of the most implemented machine-learning techniques in engineering risk assessment [11]. The ANN model is a network of machine learning that is based on the human brain [12]. Nowadays, ANNs and computational intelligence (CI) methods are often used for flood disaster modeling [13]. Machine-learning technologies have been applied for analysis of flood susceptibility assessment, including logistic regression, radial basis function (RBF) neural network, and support vector machine (SVM) [14–16]. For logistic regression, this algorithm is easier to implement and interpret, and more efficient to train [14]. However, it may lead to overfitting. For the RBF neural network, it performs more robustly and

tolerantly than conventional neural networks, especially when dealing with noisy data [15]. For SVMs, it is more effective in high-dimensional spaces [16]. Despite the success of the above machine-learning technologies as effective numerical tools for engineering risk assessment, there is still growing interest in the development of a more accurate predictive risk model. The analysis of the spatial distribution of flood disaster risk plays an important role, especially regarding disasters occurring along coasts and rivers [17–19]. Spatial analysis is applied to define the relationship between flood factors in hazards, vulnerability, and risk through the map, without focusing on complex hydrological modeling [20]. For flood analysis, studies using geographic information system (GIS) technologies, remote sensing, and numerical models, and adopting an artificial neural network approach, have been widely used around the world [21–26].

This article presents a GIS-based artificial neural network (GANN) model for the flood susceptibility assessment of Keelung City, Taiwan. Various factors, including elevation, slope angle, slope aspect, flow accumulation, flow direction, topographic wetness index (TWI), drainage density, rainfall, and normalized difference vegetation index, were generated using a digital elevation model and LANDSAT 8 imagery. Historical flood data from 2015 to 2019, including 307 flood events, were adopted for a comparison of flood susceptibility. Using these factors, the GANN model, based on the back-propagation neural network (BPNN), was employed to assess flood susceptibility. The main contribution of this work is that the proposed method, based on ANN and GIS, may improve the ability to establish further precise flood models, and present the results in a spatial environment. The advantages of the GIS spatial analysis capability were integrated into the artificial neural network model. This work is organized as follows. In Section 2, the methodology is introduced. Results are presented in Section 3. Discussions are presented in Section 4, and key findings of this pioneering work are summarized in this section. Conclusions are made in Section 5.

2. Materials and Methods

2.1. Description of the Study Area

The research area was Keelung City, which is located in the northeastern part of Taiwan. The city area covers 132.7589 km^2, and is divided into 7 districts and 157 villages, as shown in Figure 1. The city is also known as the rainy port for its high frequency of rain, with a yearly rainfall average upwards of 3700 mm. Keelung City is one of the major coastal cities in Northern Taiwan which has become highly urbanized and densely populated in the last few years [4].

2.2. Preparation of Data and Geospatial Layer

Table 1 depicts the source data of the factors. The geomorphologic area of Keelung City and the relevant factors are shown in Figure 2. Factors including elevation, slope angle, slope aspect, flow accumulation, flow direction, and TWI were generated from the digital elevation model (DEM), with a resolution of 20 m. The LANDSAT 8 imagery, with a resolution of 30 m, from the United States Geological Survey, was used to generate the normalized difference vegetation index. Detailed descriptions of the factors are as follows.
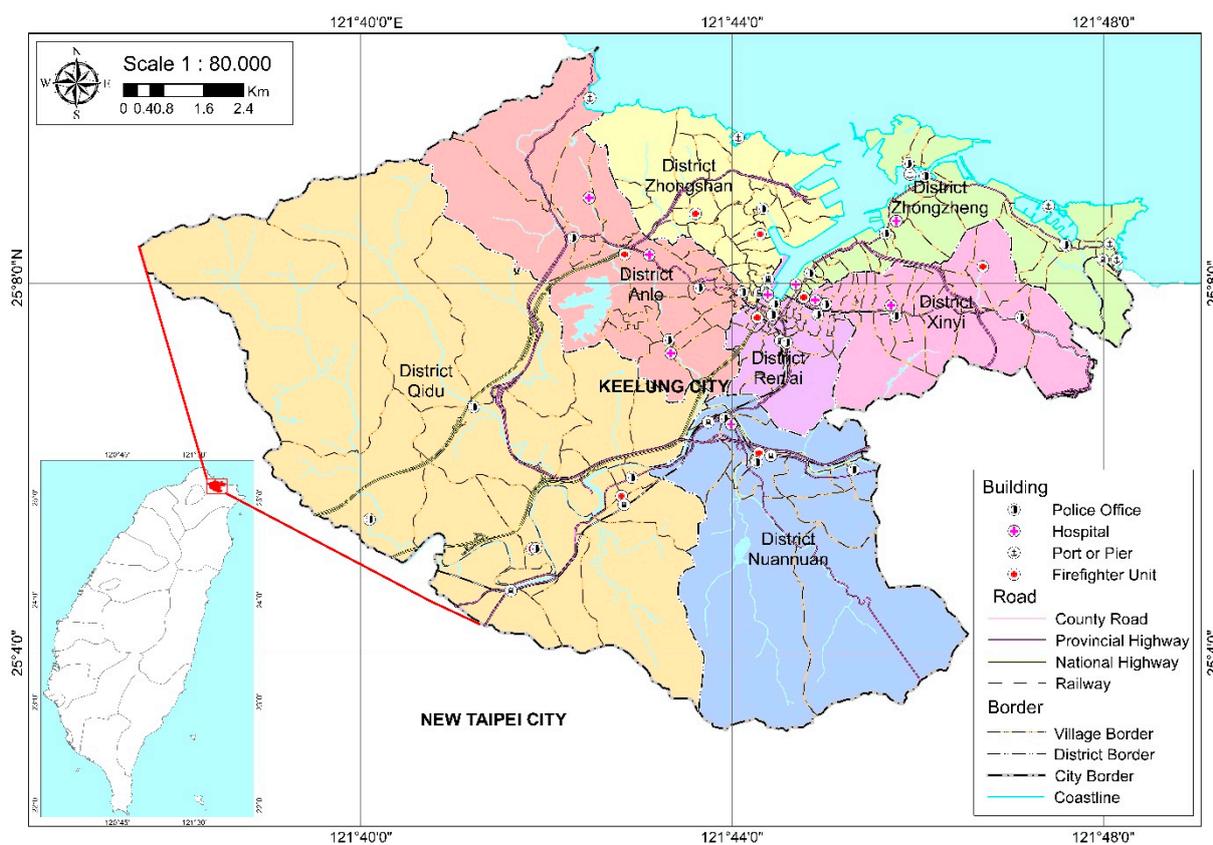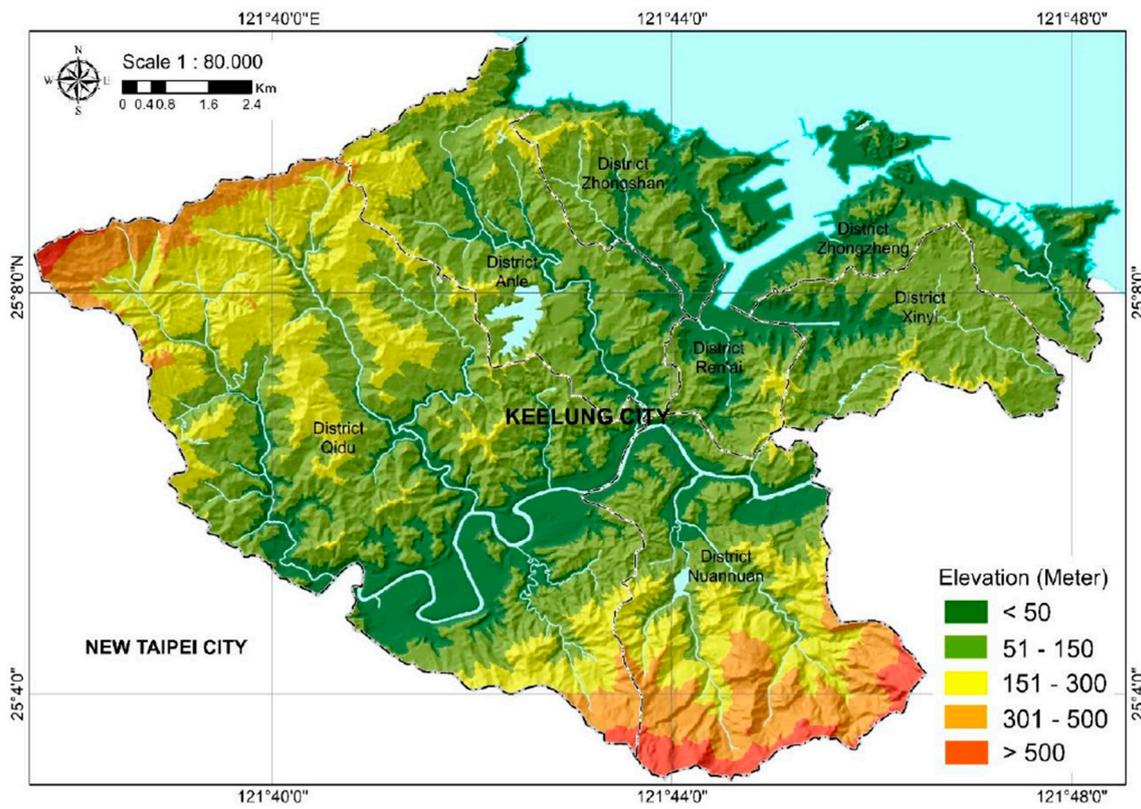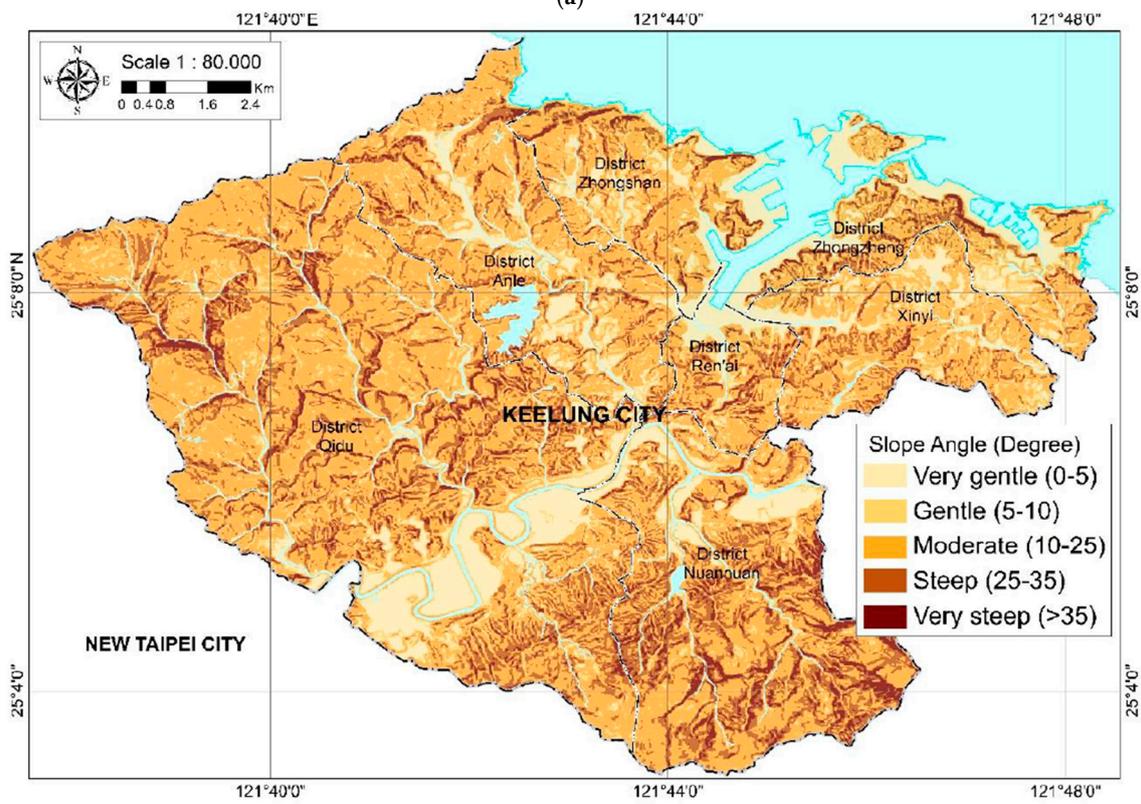
Figure 1. Administration map of Keelung City, Taiwan.

Table 1. Source data of the factors and outliers. DEM: Digital elevation model.

Source Data	Factors	Resolution (m)	Maximum Outliers
DEM	Elevation	20	704
DEM	Slope	20	60.40
DEM	Slope aspect	20	358.95
DEM	Flow direction	20	128
DEM	Flow accumulation	20	247,638
DEM	Topographic wetness index (TWI)	20	19.74
Stream river	Drainage density	20	4.07
Rainfall data from the Central Weather Bureau (CWB) of Taiwan	Rainfall interpolation	96	102.65
LANDSAT 8 imagery	Normalized difference vegetation index	30	0.41

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

4 of 20

(**a**)

(**b**)

Figure 2. *Cont.*

(**c**)

(**d**)

Figure 2. *Cont.*

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

6 of 20
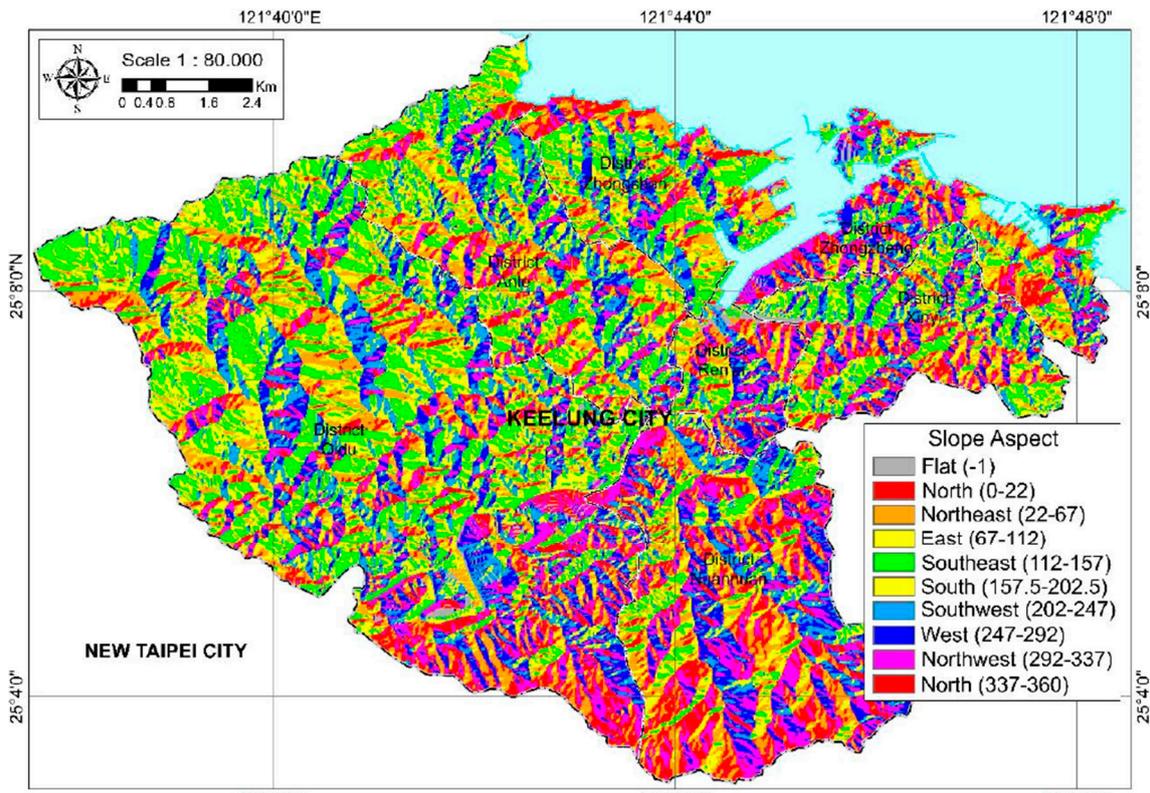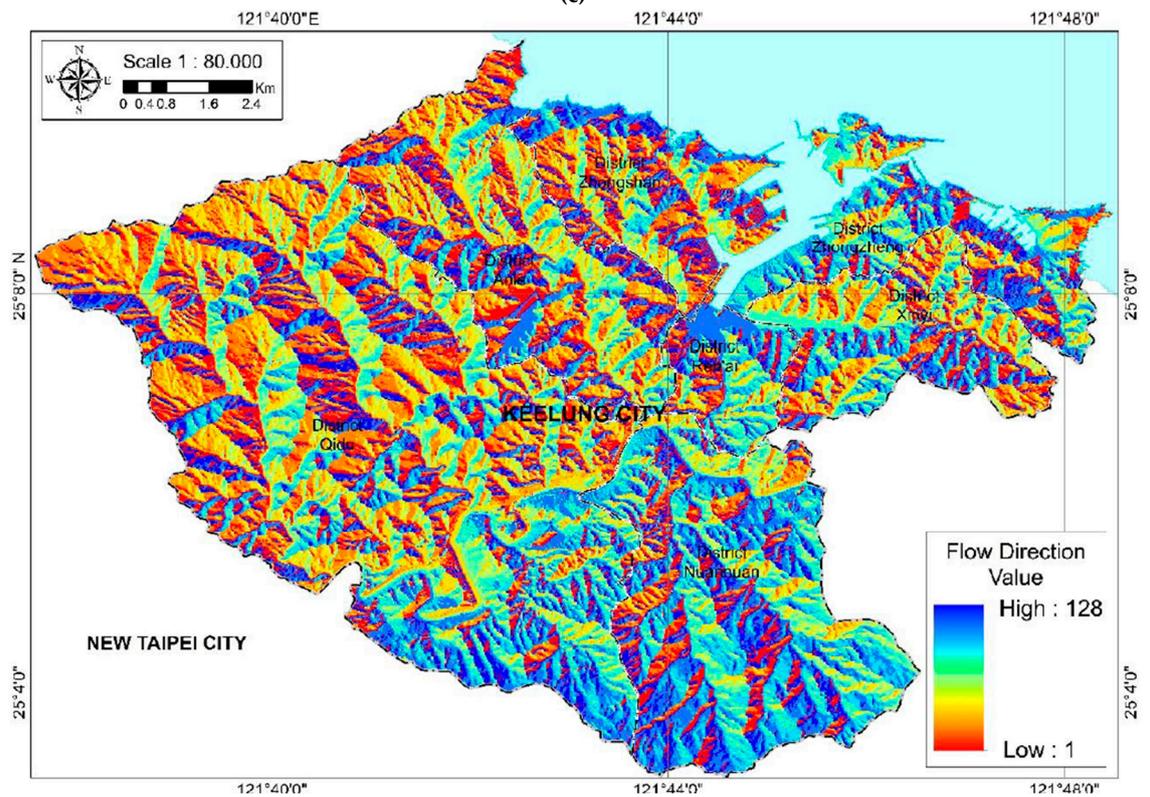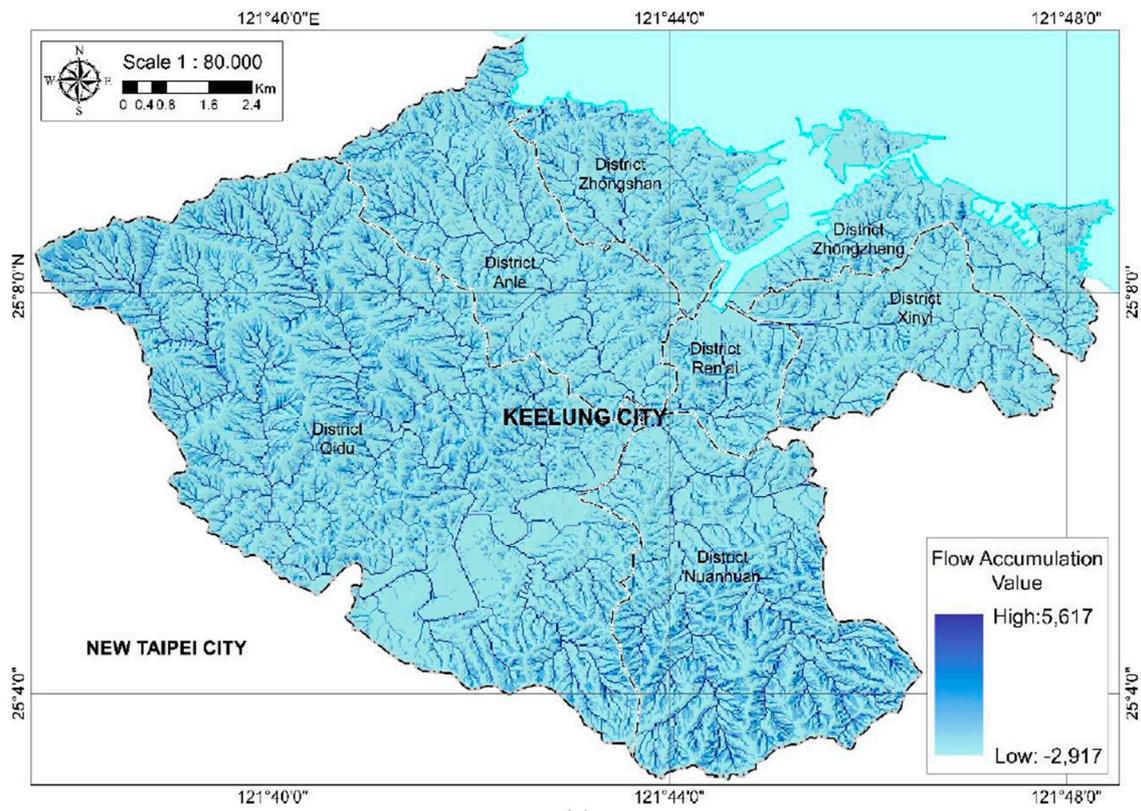
(**e**)
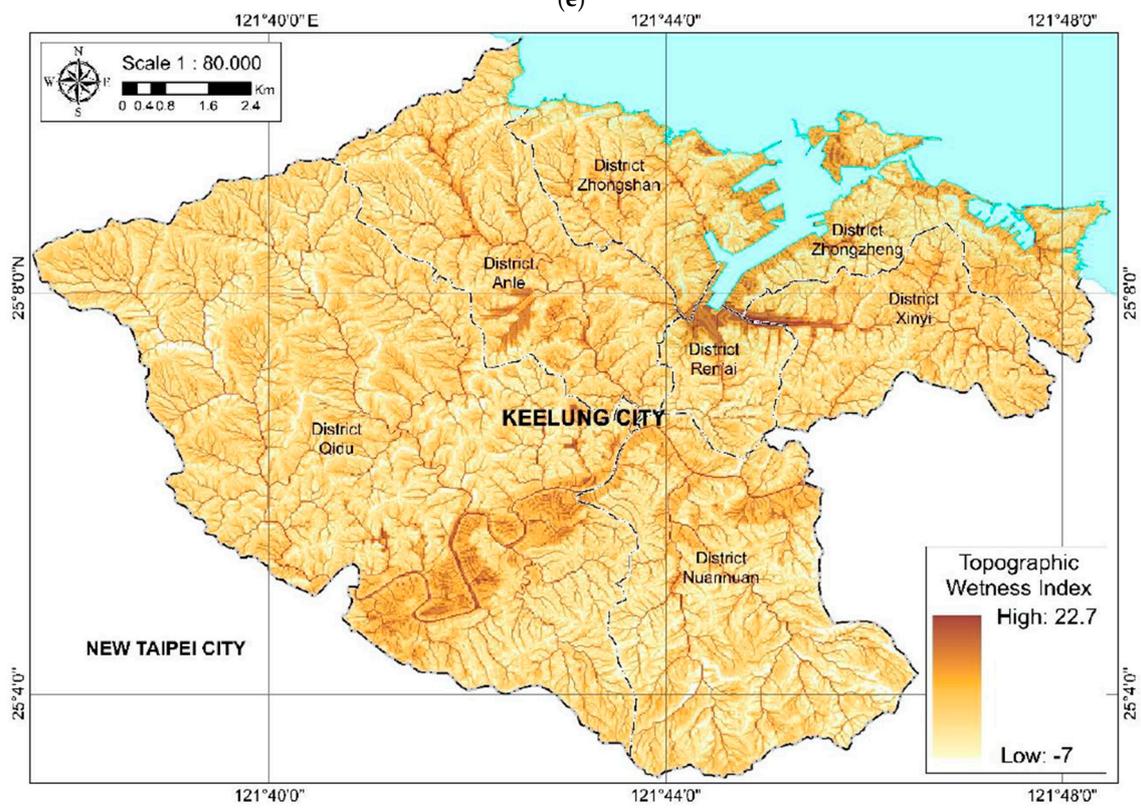
(**f**)

Figure 2. *Cont.*

(**g**)

(**h**)

Figure 2. *Cont.*

(**i**)

(**j**)
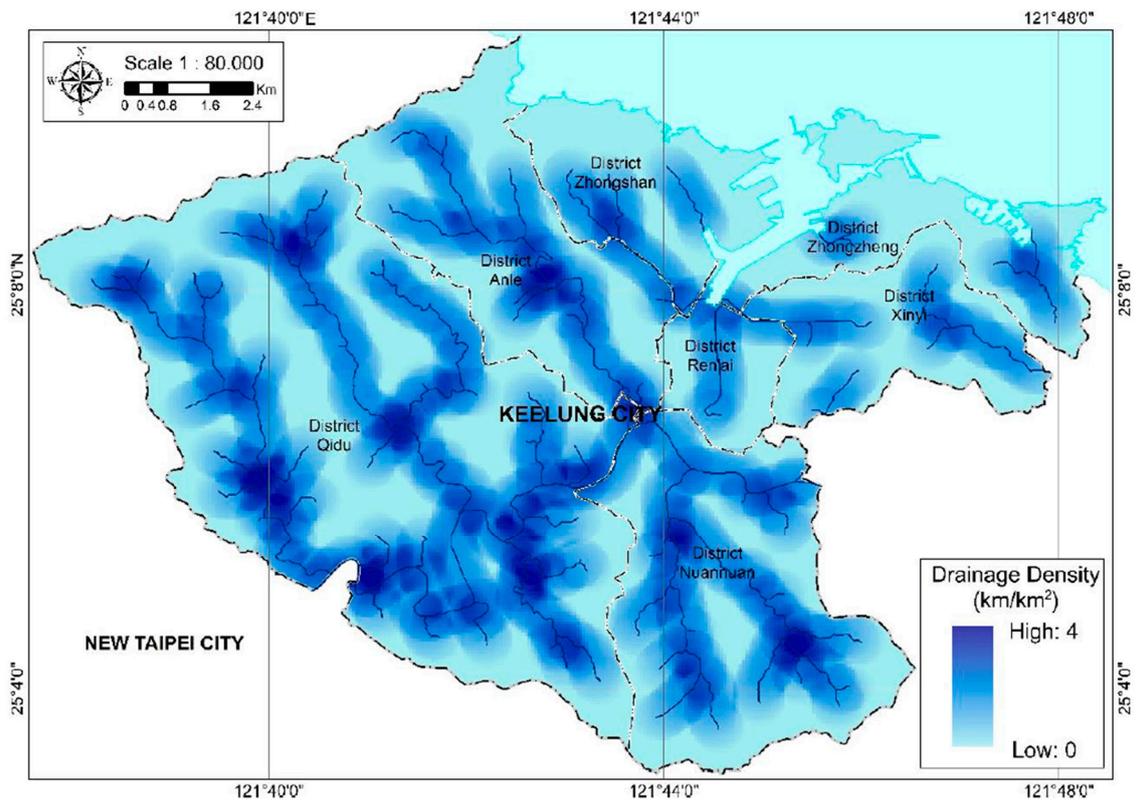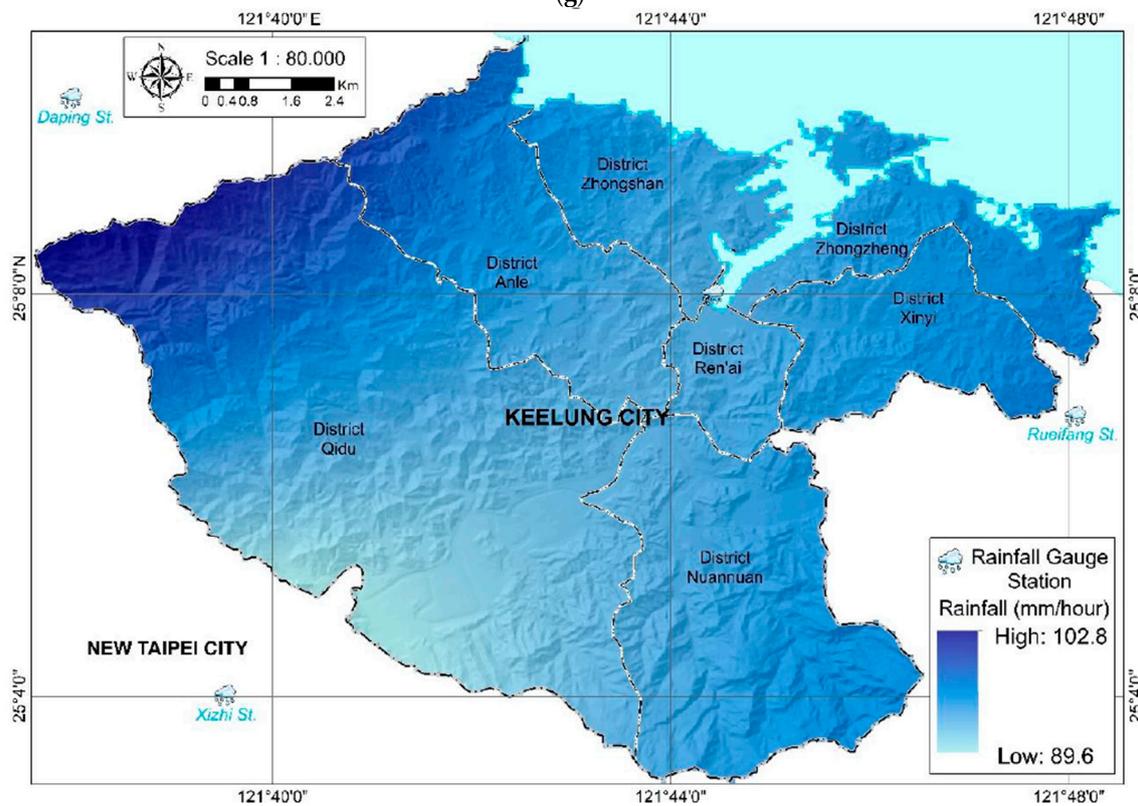
Figure 2. Geomorphologic area of Keelung City and relevant factors: (**a**) Elevation; (**b**) slope; (**c**) slope aspect; (**d**) flow direction; (**e**) flow accumulation; (**f**) TWI; (**g**) drainage density; (**h**) rainfall interpolation; (**i**) normalized difference vegetation index (NDVI); and (**j**) historical flood data.

2.2.1. Elevation

The elevation map is a representation of altitude, with ranges between 0 and 783 m. Elevation is another frequently used parameter and one of the key factors controlling the floods of an area [21,22]. Flood disasters t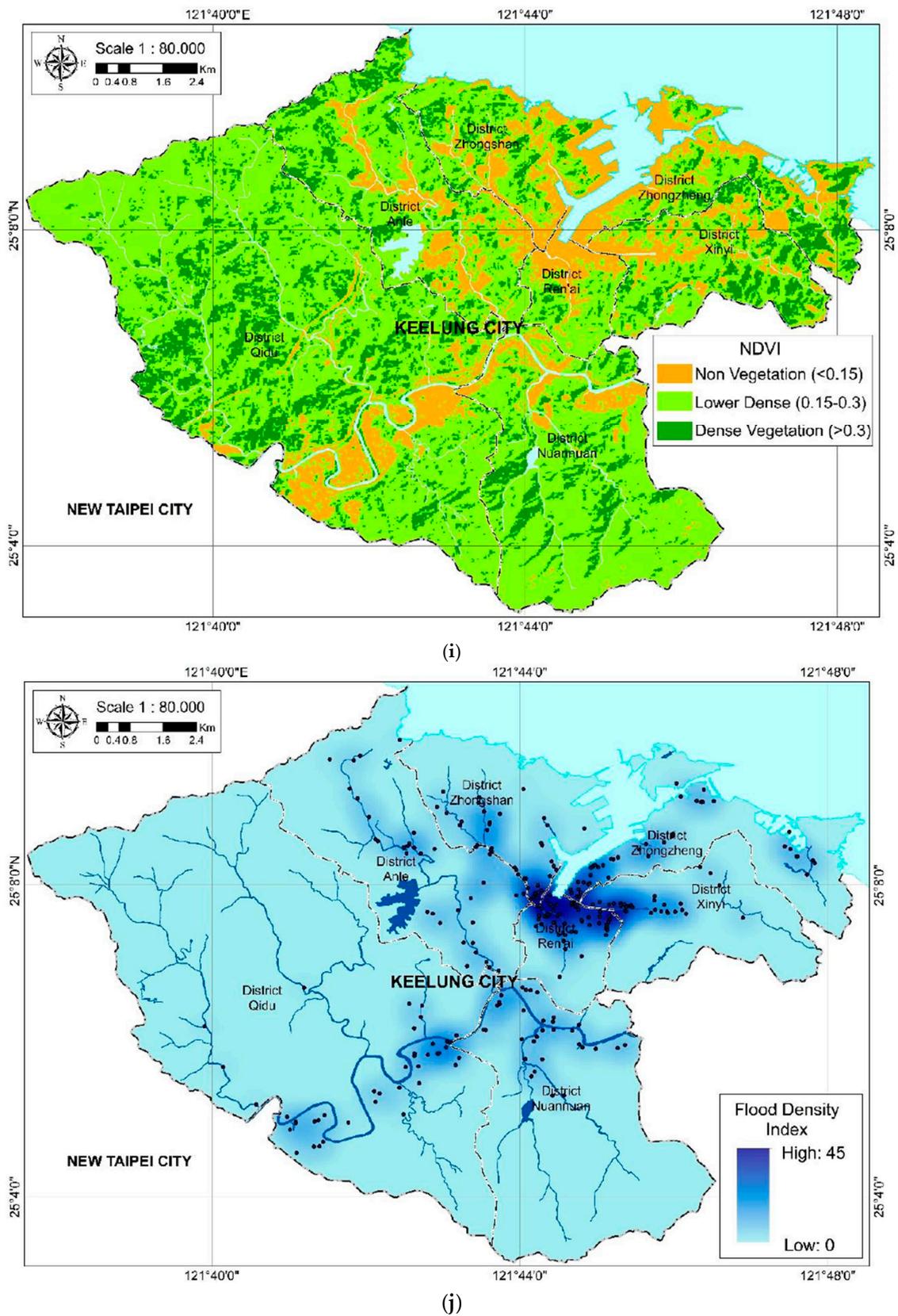end to occur in low altitude areas, compared with landslides that have a tendency to happen at higher altitudes [23]. Generally, water continually flows from higher elevations to lower elevation areas.

2.2.2. Slope

This topographic factor is fundamental in hydrological studies. The relationship with rainfall is likely that the slope is directly influenced by the infiltration of rainfall [24]. The slope of an area and the surface flow velocity could have a positive correlation. The slope is also very closely connected to the flow of runoff directly toward downstream; a higher magnitude of slope in an area might accelerate precipitation-related runoff. The surface runoff increases significantly as the gradient increases; consequently, the infiltration decreases. As an outcome of this, regions with a sudden decrease in the slope have a higher probability of flooding as a massive volume of water becomes stationary, which causes a severe flood disaster situation.

2.2.3. Slope Aspect

This aspect recognizes the downslope direction, and is also thought of as the slope orientation [24]. The aspect of the slope presents the steepness of the surface, and is represented by three groups based on color brightness or saturation. The pixel values in the output aspect–slope raster represent a combination of aspect and slope. The aspect is one of the significant factors in producing flood susceptibility maps [25].

2.2.4. Flow Direction

The hydrologic characteristics of a surface are the capacity to establish the flow direction of each raster cell. The flow direction is a grid whose value indicates the direction of flow for every cell to its steepest downslope neighbor in the DEM [26].

2.2.5. Flow Accumulation

The flow accumulation tool generates accumulated flow as the accumulated weight of all cells flowing into each downslope cell in the output raster. The results of flow accumulation can be used to create a stream network by applying a threshold value to select cells with a high accumulated flow.

2.2.6. Topographic Wetness Index

The TWI is a physical representation of flood inundation areas, which is an important component of a river catchment. The TWI of a catchment indicates two types of measurements: Flat lands and hydrographic positions. The TWI is commonly used to quantify topographic control of hydrological processes. It is expressed as

$$TWI = \ln\left(\frac{\alpha}{\tan \beta}\right), \tag{1}$$

where α is the cumulative upslope contribution area draining through a point (per unit contour length), and $\tan \beta$ is the slope angle at the point. It affects the spatial distribution of soil moisture, and the groundwater flow often follows the surface topography. In this study, TWI is considered another contributing factor. Areas with a high wetness index occur where there is a combination of low slope and high flow accumulation and, therefore, may indicate locations that are at greater flood risk [27].

2.2.7. Drainage Density

The calculation of the drainage system raster was done by using the line density method, with rivers (polyline) as the main input data. The unit of calculated density was

the length per unit of area. A higher probability of flooding is strongly associated with higher drainage density, as it represents greater surface runoff. The drainage density map of Keelung City was calculated from the drainage network map using line density tool in the ArcGIS, and ranges from 0 to 4 km/km^2.

2.2.8. Rainfall Interpolation

Rainfall-induced flooding is associated with tropical storms, hurricanes, tropical depressions, and west trade winds that directly strike the windward side of the highlands. A large number of previous studies in the literature have established a relationship between the rainfall and the flood occurrence of an area [28–30]. Preparation of a rainfall map in this study used the pixel-based highest hourly rainfall data of several decades, spread around four rain gauge stations around Keelung City, with the Kriging interpolation method [31]. The rainfall data were collected from Keelung station, Xizhi station, Ruifang station, and Daping station through the Central Weather Bureau (CWB) of Taiwan.

2.2.9. Normalized Difference Vegetation Index (NDVI)

The NDVI is one of the most extensively adopted vegetation indexes using satellite imagery, and for monitoring of global vegetation cover [32]. The NDVI, developed by Rouse in 1973, is used to monitor vegetation health, and to compare outputs across sensors with slightly different specifications [33]. The NDVI equation is defined as

$$\text{NDVI} = \frac{NIR - RED}{NIR + RED}, \tag{2}$$

where *NIR* is the reflection in the near-infrared spectrum and *RED* is the reflection in the red range of the spectrum. *NIR* and *RED* represent near-infrared ($\lambda \sim 0.8$ μm) regions of the spectrum and surface reflectance averaged over the visible spectrum ($\lambda \sim 0.6$ μm), respectively. The NDVI data source for this layer was the LANDSAT 8 OLI/TIRS C1 Level 1, to match the multispectral bands captured on 13 March 2018 and the area coverage of the study area on path 117 and row 43. The satellite imagery was rendered as NDVI, and colorized for use in the visualization analysis. Specifically, the source of the NDVI layer is the metadata of the imagery, which shows land cloud cover of only 0.8% and scene cloud cover of 1.91%. The NDVI response indicates the effective flood extent as an influential factor for the strength and capacity of an area against flood hazards.

2.2.10. Historical Flood Density

Information on the historical floods in Keelung City was collected from the Emergency Management Information Cloud (EMIC) of Taiwan from 2015 to 2019, and comprised 307 events. The number of events each year is shown in Figure 3. The flood history, also known as the disaster experience, assumes that such areas have higher adaption ability, but also a high probability of flood occurrence in the future [3,34]. The density of point features in the raster cell unit is generated using the kernel density to fit a smoothly tapered surface to each point of the flood history. The kernel density tool in ArcGIS calculates the density of features in a neighborhood around the features, to find the spatial analysis and previous flood history of the area. By calculating the values of all the kernel surfaces, the density of each output raster cell feature is determined.

2.3. Artificial Neural Network

The ANN is considered a quantitative black-box approach that tries to simulate the functional human biological nervous system [12]. Moreover, the environment nonlinearity analysis and the forecast can be studied by applying this effective and affordable machine-learning tool. Even though the ANN has also been effectively applied to flood analysis in previous investigations [21,28], the GANN model is still rarely used for flood susceptibility assessment. Furthermore, the ANN has been considered as an alternative to physically-based models due to its simplicity regarding the minimum requirements for collecting

detailed data [35]. The schematic and conceptual flowchart of this study is depicted in Figure 4. From Figure 4, three major factors, including topography factors, geology and geomorphology factors, and meteorology factors, were considered as a multi-resource aspect of the database in this study. Various factors, including elevation, slope angle, slope aspect, flow accumulation, flow direction, TWI, drainage density, rainfall, and NDVI, were generated using the digital elevation model, as well as LANDSAT 8 imagery. Historical flood data from 2015 to 2019, including 307 flood events, were adopted for the comparison of flood susceptibility. After collecting the possible factors, the proposed GANN model, based on the BPNN, was employed to assess the flood susceptibility. Finally, the accuracy of proposed GANN model predictions was evaluated by calculating the correlation coefficient.
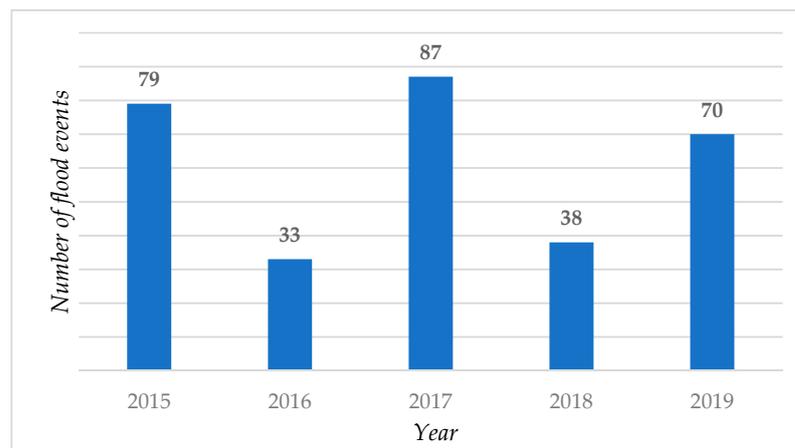

Figure 3. Historical flood data in Keelung City, Taiwan.
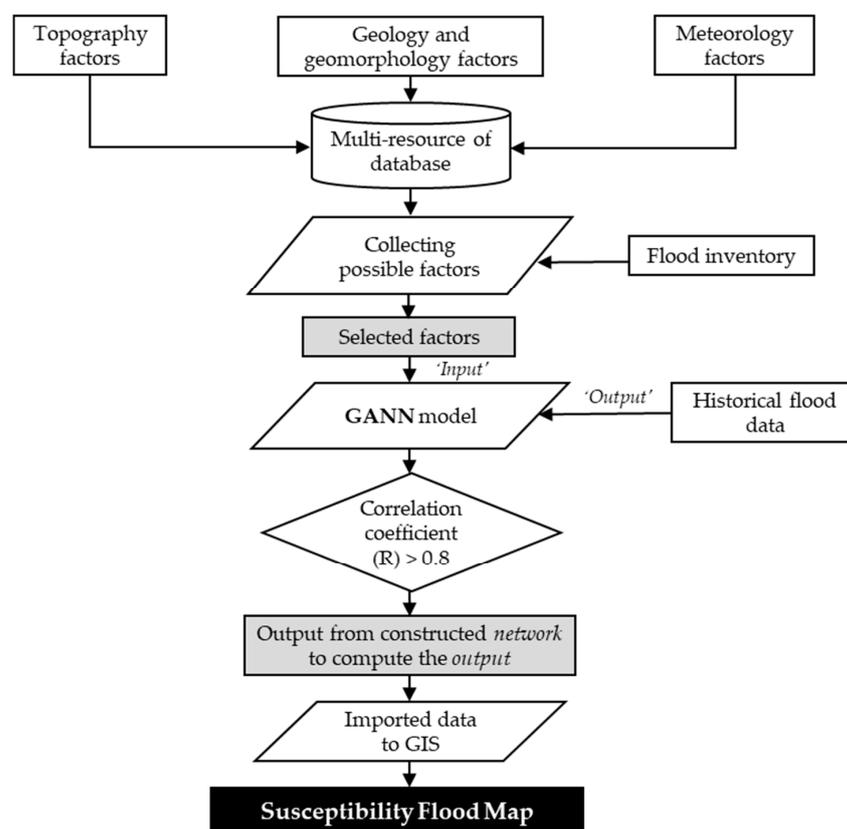

Figure 4. Flow chart of the study.

The GANN was employed to assess the flood susceptibility of Keelung City. This research focuses on a BPNN [36,37]. A typical algorithm of artificial neurons comprises three layers: Input, hidden, and output layers. The BPNN algorithm of feedforward shows an essential feature of the training phase, and the progress result is expressed as follows:

$$y_i = F(X_j) = \left(W_{oj} + \sum_{i=1}^{I} W_{ij} x_i \right), \tag{3}$$

$$Z_k = F(Y_k) = \left(W_{ok} + \sum_{j=1}^{J} W_{kj} y_i \right), \tag{4}$$

where x_i, y_i, and Z_k indicate the input, hidden, and output layers, respectively. The bias weights for setting the threshold values are represented as W_{oj} and W_{ok}. Meanwhile, X_j and Y_k indicate temporarily calculated results before using the activation function, and F is the activation function applied in the hidden and output layers. The F value ranges from 0 to 1. In this study, we adopted the hyperbolic tangent sigmoid function. A sigmoid function is a mathematical function with a characteristic S-shape curve or sigmoid curve. The sigmoid activation function is a widely used activation function for neural networks [38,39]. The positive input value to the function is transformed into a value between 0.0 and 1.0. Inputs that are much larger than 1.0 are transformed to the value 1.0; similarly, values much smaller than 0.0 are snapped to 0.0. The shape of the function for all possible inputs is an S-shape from zero up through 0.5 to 1.0. Since the probability of anything exists only between the range of 0.0 and 1.0, the sigmoid activation function is adopted as the activation function for the proposed neural networks. Thus, the output y_i and Z_k can be expressed as

$$y_i = F(X_j) = F\left(\frac{1}{1 + e^{-X_j}} \right), \tag{5}$$

$$Z_k = F(Y_k) = F\left(\frac{1}{1 + e^{-Y_k}} \right). \tag{6}$$

For the error back-propagation weight training, the error function can be established as

$$E = \frac{1}{2} \sum_{k=1}^{K} \varepsilon_k^2 = \frac{1}{2} \sum_{k=1}^{K} (t_k - z_k)^2, \tag{7}$$

where t_k and ε_k are the target value and error in each output node, respectively. The goal is to minimize E, the error between the actual output values of the network. The weight adjustment in the link between the hidden and output layers can be expressed as

$$\Delta w_{jk} = \eta \times y_j \times \delta_k, \tag{8}$$

where η is the learning rate, with values ranging between 0 and 1. The learning rate values correlate with the speed of convergence of the network of the BPNN. Conversely, a learning rate that is overly large can lead to a widely oscillating network. The new weight herein is updated by the following equation:

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n), \tag{9}$$

where n is the number of iterations in the network. Similarly, the error gradient in links between the input and hidden layers can be derived from the partial derivative with respect to w_{ij}:

$$\frac{\partial E}{\partial w_{ij}} = \left[\sum_{k=1}^{K} \frac{\partial E}{\partial z_k} \frac{\partial z}{\partial Y_k} \frac{\partial_k}{\partial y_j} \right] \times \left(\frac{\partial y_j}{\partial X_j} \right) \times \left(\frac{\partial X_j}{\partial w_{ij}} \right) = -\Delta_j x_i, \tag{10}$$

Int. J. Environ. Res. Public Health **2021**, 18, 1072

13 of 20

$$\Delta_j = F'(X_j) \sum_{k=1}^{K} \delta_k w_{jk}. \tag{11}$$

The new weight in the hidden and input links can be regenerated as

$$\Delta w_{ij} = \eta \times x_i \times \Delta_j , \tag{12}$$

$$w_{ij}(n+1) = w_{ij}(1) + \Delta w_{ij}(n). \tag{13}$$

To evaluate the prediction performance of the proposed GANN model, the correlation coefficient (R) is utilized and expressed as

$$R = \frac{\sum\limits_{i=1}^{n} (t_i - \bar{t})(o_i - \bar{o})}{\sqrt{\sum\limits_{i=1}^{n} (t_i - \bar{t})^2 \sum\limits_{i=1}^{n} (o_i - \bar{o})^2}}, \tag{14}$$

where t_i represents the target value and o_i is the output value. Meanwhile, $\bar{t}$ and $\bar{o}$ are the average values of the target and output values, respectively.

The BPNN algorithm is used in the feedforward GANN, and the structure of the neural network model is displayed in Figure 5. The selected input, containing nine quantitative input variables and one quantitative output variable, includes elevation, slope, TWI, rainfall, NDVI, flow accumulation, drainage density, flow direction, and slope aspect, as shown in Figure 5. The flood history data from 2015 to 2019 in Keelung City were the only output of this model. All the input data in the GANN model were normalized in the range 0–1, with the initial weights automatically assigned to random values.
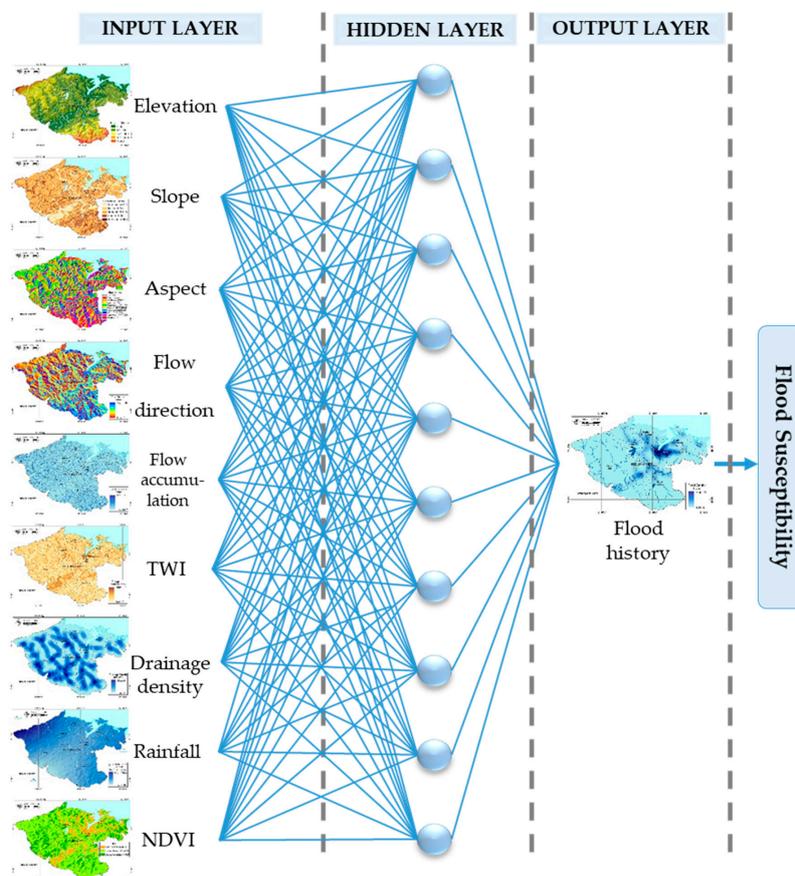

Figure 5. Structure of a typical three-layer model for flood susceptibility.

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

14 of 20

3. Results

3.1. Results of the Artificial Neural Network Model

The network architecture was designed to determine the flood susceptibility of Keelung City; an output layer of the flood history was required for the architecture of the model. This was used to evaluate the prediction performance of the train validation and test the model using the mean squared normalized error performance function, which measures the network's performance function according to the mean squared error (MSE).

The GANN study needed to classify the variation data in the training process. Kia et al. [21] indicated that 60% should be used for training, 20% for validation, and 20% as testing data, while Aziz et al. [29] and Latt and Wittenberg [40] used a combination of 80% and 20% for training and the testing process, respectively. In this study, the flood data were divided into three groups: 70% for training the network, 15% for validating the model, and 15% for testing the data to completely independently test the network generalization. The chosen training algorithm of this research was Levenberg–Marquardt (LM), which has the fastest training compared to Bayesian regulation back-propagation. This latter method takes longer, but may be better for challenging problems, while scaled conjugate gradient back-propagation is suitable for low-memory situations and was not used here [41]. The LM algorithm was selected as the training function, which combines the Gauss–Newton method and the gradient descent method. The LM algorithm was used to solve non-linear least squares problems and for its fault tolerance and fast convergence ability.

The number of hidden layers, along with neurons inside the model, is frequently defined by trial and error. The number of neurons in the output layers is fixed by the application, and is represented by the class being processed. The GANN model of flood susceptibility used 10 hidden layers, as at that point the model starts to reach the minimum requirement for the correlation coefficient, according to some research [8,28,42].

Figure 6 shows that the correlation coefficient value is very low, with only single nodes, and increases rapidly in five hidden layers, finally gradually stabilizing in 10 hidden layers. This is despite the fact that several researchers used only one hidden layer in their ANN architecture, or a small value of hidden layers from 1 to 7 [29,40]. Some studies show that using more hidden layers obtains the best result, such as Campolo et al. [43], who trained the variables with 20, 25, 30, 35, and 40 hidden layers, and Islam [35] as a comparison, with 15, 20, 25, 30, and 35 hidden layers. There are no strict rules for assigning the number of hidden layers and neurons in the literature [21]. The best design for GANN architecture depends on the problem type under investigation. In this study, the GANN model of flood susceptibility used 9 input layers, 10 hidden layers, and 1 output layer.
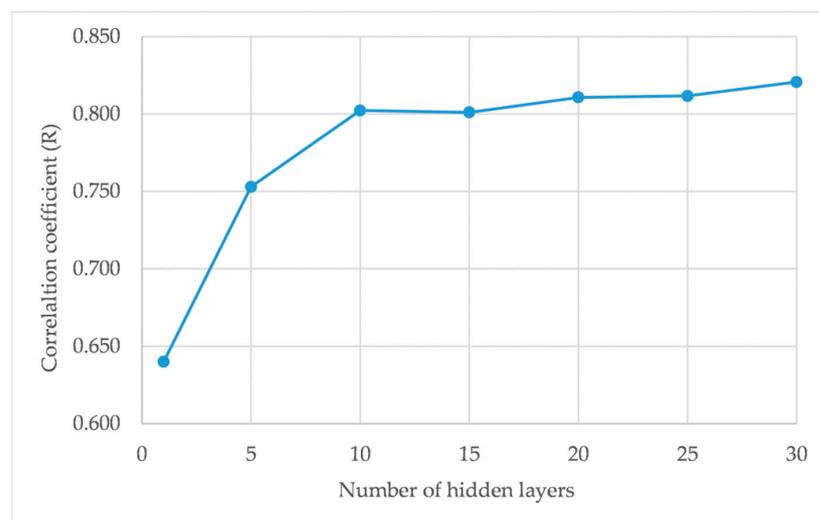

Figure 6. Number of hidden layers versus the value of R.

The training process reduced the MSE value from 10^3 to 9.27 in 161 iterations. However, the best performance, at epoch 155, was 9.7189. The performance declined sharply in the first 20 epochs, and then gradually decreased until epoch 161. Figure 7 presents the linear regression for targets relative to the output of the different sub-divisions after the training process was completed.
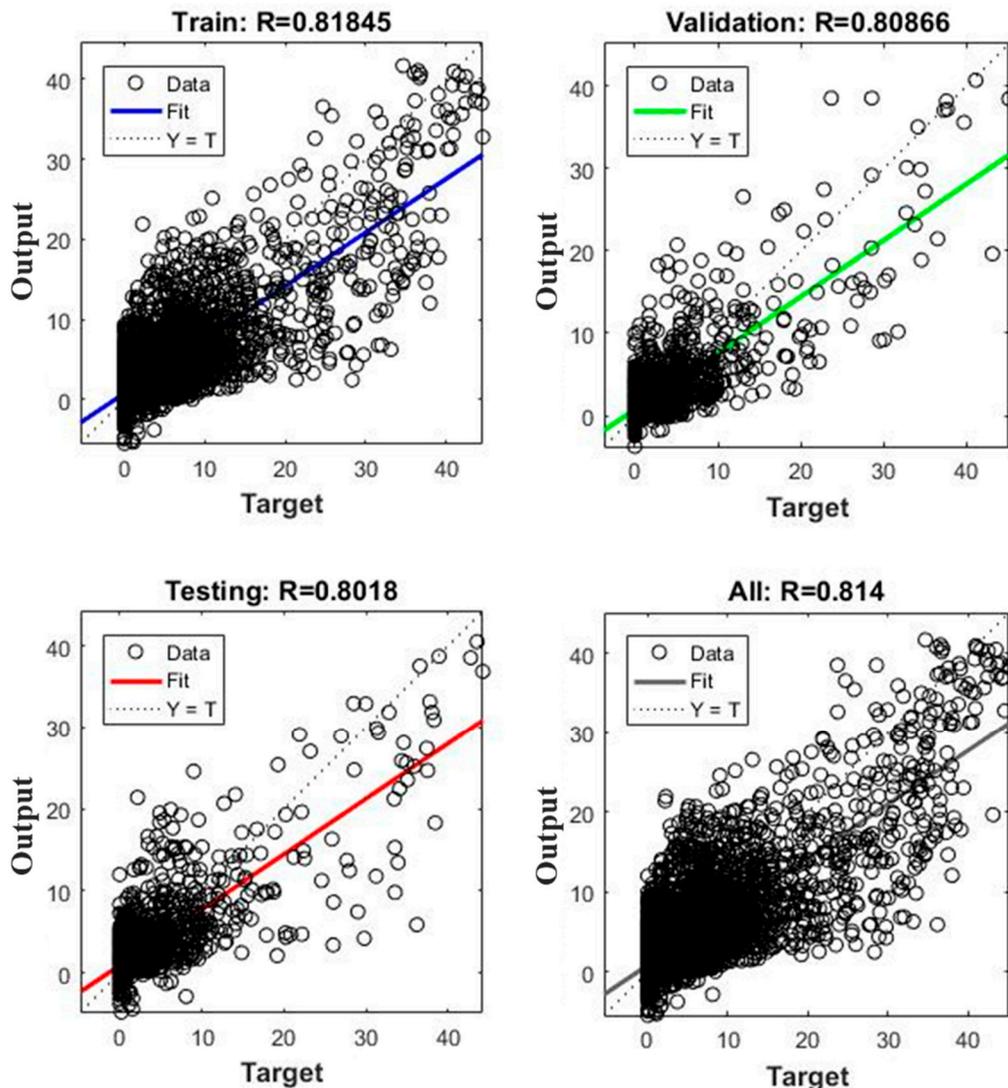

Figure 7. Training process for the correlation coefficient.

The efficiency of training the network is represented by the correlation coefficient (R). The cumulative R value was equal to 0.814, which reached the minimum standard of the study requirement (R > 0.8). The training, validation, and testing data sub-division were R = 0.818, 0.808, and 0.801, respectively. The results show that there is a good correlation between the historical flood data and those predicted by the proposed GANN model. According to previous studies [44,45], correlation coefficients whose magnitudes are between 0.7 and 0.9 demonstrate variables which can be considered highly correlated. These results highlight the efficiency of the constructed neural network during the training process and in forecasting flood susceptibility. Our results show that the proposed GANN model could efficiently predict flood susceptibility.

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

16 of 20

3.2. Flood Susceptibility Map

The flood susceptibility map is helpful for disaster planning, in addition to being useful during an actual emergency response to floods. Mapping flood susceptibility could be the first step toward mitigating flooding, because flood susceptibility identifies the most vulnerable locations and provides sufficient lead time for an individual to respond to flooding in an anticipatory rather than reactive manner [46]. The computation of the weight of the factors and artificial neural network modeling was performed in MATLAB; the outputs were exported to GIS for map production and visual interpretation. The flood susceptibility map was analyzed qualitatively using natural breaks (jenks) classification schemes [47,48]. The classification of flood susceptibility mapping in Keelung City could be categorized into five classes [23,27,30] based on the value of GANN prediction of the output: Very low, low, moderate, high, and very high. The flood susceptibility index of the five classes is shown in Table 2, and the flood sustainability of Keelung City is displayed in Figure 8. In Figure 8, the results show that a total of 12.1% of the Keelung City area is classified as flood-prone (very high, high, and moderate), according to the GANN model, with a coverage area of 16.14 km^2. Nearly 3.5% of the study area is in high to very high flood susceptibility zones, as shown in Table 2. The highest susceptibility area was only 1% of the total, and should be prioritized for flood management. To prevent urban inundation in the study, the Water Resources Agency, Ministry of Economic Affairs, ROC has developed a two-dimensional flood forecasting system using the Delft-FEWS platform to integrate the SOBEK models and the precipitation data from the Central Weather Bureau [49]. The flood sustainability obtained from the SOBEK model is adopted for comparison. The comparative results can be seen in Figure 9. It is demonstrated that the overall flood-prone areas from both approaches agree with each other. However, it was also found that the flood sustainability from the SOBEK model was much more conservative than our proposed GANN model.

Table 2. Classification of the flood susceptibility area in Keelung City.

Classification	Flood Susceptibility Index	Area (km^2)	Percentage (%)
Very low	−3.0–1.7	80.30	60.5
Low	1.7–5.3	36.32	27.4
Moderate	5.3–11.6	11.48	8.6
High	11.6–22.5	3.31	2.5
Very high	22.5–38.2	1.35	1
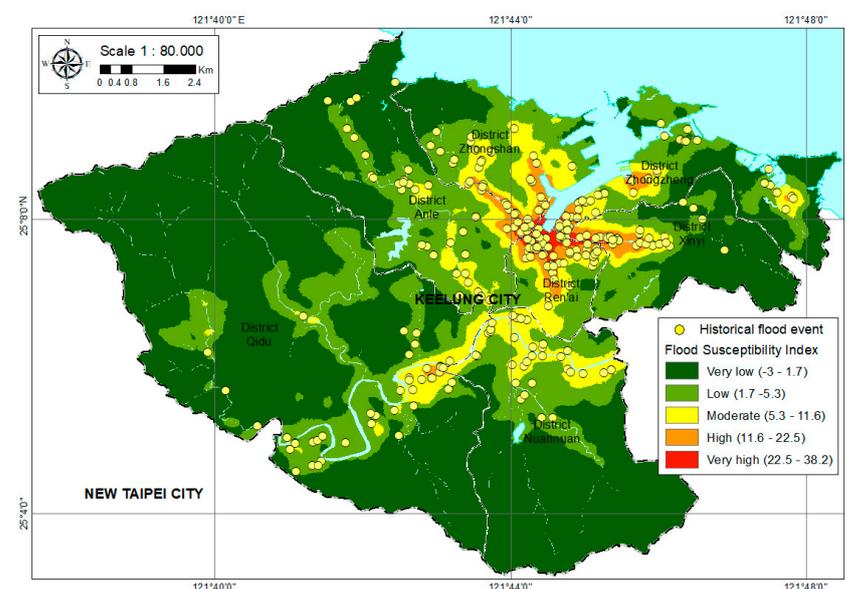
Figure 8. Flood sustainability map of Keelung City.

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

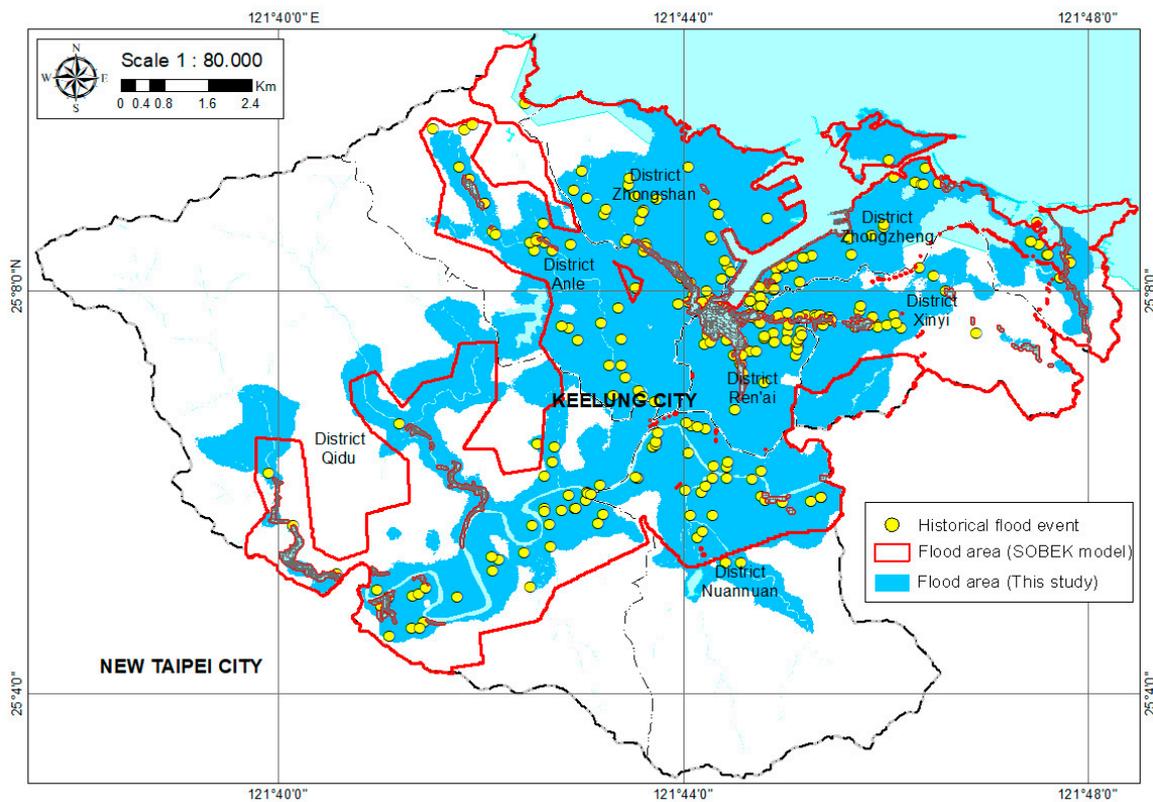17 of 20

Figure 9. Comparison of the flood area of Keelung City.

The proposed method, based on ANN and GIS, may improve the ability to establish further precise flood models, and present results in a spatial environment. The outcomes of the research could be used to help local authorities to develop appropriate new infrastructure to protect lives and property in Keelung City.

4. Discussion

In this research, high elevations occurred in the western part of Qidu District and the southeastern part of Nuannuan District. Meanwhile, the lowland area covered the seashore, especially in the north of Keelung City and the central part, near the Keelung river. The flats area was located in the center of Keelung City, near the flood plain, and a very steep slope is located in the Nuannuan and Qidu District. Based on the GANN results, the most common floods were predicted to occur on the flats, and the northwest (NW), south (S), and southwest (SW) facing slopes. The floods were particularly rare on the north (N) and southeast (SE) facing slopes. Generally, the slopes of the study area are most commonly oriented to the north (N) and southeast (SE) quadrants. To produce a map of the vegetation index using the NDVI method, data from the LANDSAT 8 and high-density vegetation located in the southwest of Qidu District and eastern part of Xinyi District were used. Meanwhile, the non-vegetation area was the built-up area in the downtown region of Keelung City, especially the port area in the Ren'ai District, Xinyi District, Zhongzheng District, Zhongshan District, and southeast of Qidu District.

Finally, the high and very high prone areas were located in the northern part of the city, as well as in the central part of Keelung City alongside the river. Ren'ai District has the highest susceptibility, followed by Xinyi, Zhongzheng, and Zhongshan District. However, this very high susceptibility area is a core district of the city, which is highly populated, and the economic center of the city. The areas with a classification of very low and low potential coverage were, cumulatively, 87.9% or 116.62 km^2. These are located at high elevation, especially in the western and southeastern parts of Keelung City, which includes the western part of Qidu and Anle District, as well as the southeast of Nuannuan District.

Int. J. Environ. Res. Public Health **2021**, *18*, 1072

18 of 20

The novel aspect of this work was to develop a GANN model for flood susceptibility assessment of Keelung, Taiwan. The main contribution of this work is that the proposed method, based on ANN and GIS, may improve the ability to establish further precise flood models, and present results in a spatial environment. The advantages of the GIS spatial analysis capability were integrated into the artificial neural network model. Accordingly, the proposed methodology may represent spatial continuity and the influence of parameters on flood-generating mechanisms. In addition, historical flood data from 2015 to 2019, including 307 flood events, were adopted for the comparison of the flood susceptibility on the regional spatial scale using the GIS. The finding observed in this work may provide a fundamental contribution to environmental protection engineering for flood in areas with higher occurrence and vulnerability to extreme precipitation.

5. Conclusions

In this research, the GANN model was developed using 10 flood causative factors. The thematic layers, including elevation, slope angle, slope aspect, flow direction, flow accumulation, TWI, NDVI, and drainage density, were generated using GIS. Rainfall data were also used as input, while historical flood events from 2015 to 2019 were the output of the GANN model.

The proposed GANN model produced satisfactory results, with a coefficient of correlation of 0.81. The susceptibility was categorized into five classes: Very low, low, moderate, high, and very high, with coverage areas of 60.5%, 27.4%, 8.6%, 2.5%, and 1%, respectively. Just 3.5% of the study area was included in the high to very high flood susceptibility zones; however, this area is a core district of the city, with a dense population, and the economic center of the city.

Furthermore, mapping flood susceptibility is crucial to mitigating flood disasters, since flood susceptibility can identify the most vulnerable areas and predict the potential locations of susceptibility, which can provide authorities responsible for emergency response and evacuation procedures with more information for planning and responses in very high susceptibility areas.

Author Contributions: Conceptualization and method, C.-Y.K.; verification, N.K. and C.-Y.L.; writing—original manuscript, N.K. and C.-Y.K.; data curation, C.-Y.L.; revising the manuscript, C.-Y.K. and C.-Y.L.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The datasets generated during the current study are available from the corresponding authors on reasonable request.

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

References

1. Merkens, J.-L.; Reimann, L.; Hinkel, J.; Vafeidis, A.T. Gridded population projections for the coastal zone under the shared socioeconomic pathways. *Glob. Planet. Chang.* **2016**, *145*, 57–66. [CrossRef]
2. Zanetti, V.; de Sousa Junior, W.; De Freitas, D. A climate change vulnerability index and case study in a Brazilian Coastal City. *Sustainability* **2016**, *8*, 811. [CrossRef]
3. Chang, H.-S.; Chen, T.-L. Spatial heterogeneity of local flood vulnerability indicators within flood-prone areas in Taiwan. *Environ. Earth Sci.* **2016**, *75*, 1484. [CrossRef]
4. Hsieh, L.-S.; Hsu, M.-H.; Li, M.-H. An assessment of structural measures for flood-prone lowlands with high population density along the Keelung River in Taiwan. *Nat. Hazards* **2006**, *37*, 133–152. [CrossRef]
5. Hirabayashi, Y.; Mahendran, R.; Koirala, S.; Konoshima, L.; Yamazaki, D.; Watanabe, S.; Kim, H.; Kanae, S. Global flood risk under climate change. *Nat. Clim. Chang.* **2013**, *3*, 816–821. [CrossRef]
6. Lee, K.T.; Ho, Y.-H.; Chyan, Y.-J. Bridge blockage and overbank flow simulations using HEC–RAS in the Keelung River during the 2001 Nari Typhoon. *J. Hydraul. Eng. ASCE* **2006**, *132*, 319–323. [CrossRef]

7. Wu, S.-J.; Yang, J.-C.; Tung, Y.-K. Risk analysis for flood-control structure under consideration of uncertainties in design flood. *Nat. Hazards* **2010**, *58*, 117–140. [CrossRef]

8. Abou Rjeily, Y.; Abbas, O.; Sadek, M.; Shahrour, I.; Hage Chehade, F. Flood forecasting within urban drainage systems using NARX neural network. *Water Sci. Technol.* **2017**, *76*, 2401–2412. [CrossRef]

9. Hsu, T.-W.; Shih, D.-S.; Li, C.-Y.; Lan, Y.-J.; Lin, Y.-C. A study on coastal flooding and risk assessment under climate change in the Mid-Western Coast of Taiwan. *Water* **2017**, *9*, 390. [CrossRef]

10. Jung, Y.; Shin, Y.; Jang, C.H.; Kum, D.; Kim, Y.S.; Lim, K.J.; Kim, H.B.; Park, T.S.; Lee, S.O. Estimation of flood risk index considering the regional flood characteristics: A case of South Korea. *Paddy Water Environ.* **2014**, *12*, 41–49. [CrossRef]

11. Qiu, J.; Wu, Q.; Ding, G.; Xu, Y.; Feng, S. A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.* **2016**, *2016*, 67. [CrossRef]

12. Thomas, A. Memristor-based neural networks. *J. Phys. D. Appl. Phys.* **2013**, *46*, 093001. [CrossRef]

13. Fotovatikhah, F.; Herrera, M.; Shamshirband, S.; Chau, K.-w.; Faizollahzadeh Ardabili, S.; Piran, M.J. Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. *Eng. Appl. Comp. Fluid Mech.* **2018**, *12*, 411–437. [CrossRef]

14. Zhao, G.; Pang, B.; Xu, Z.; Peng, D.; Xu, L. Assessment of urban flood susceptibility using semi-supervised machine learning model. *Sci. Total Environ.* **2019**, *659*, 940–949. [CrossRef]

15. Chen, J.; Li, Q.; Wang, H.; Deng, M. A machine earning ensemble epproach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze River Delta, China. *Int. J. Environ. Res. Public Health* **2019**, *17*, 49. [CrossRef] [PubMed]

16. Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. *J. Hydrol.* **2015**, *527*, 1130–1141. [CrossRef]

17. Zhang, J.; Chen, Y. Risk assessment of flood disaster induced by typhoon rainstorms in Guangdong Province, China. *Sustainability* **2019**, *11*, 2738. [CrossRef]

18. Moel, H.D.; Alphen, J.V.; Aerts, J.C.J.H. Flood maps in Europe—Methods, availability and use. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 289–301. [CrossRef]

19. Vos, F.; Rodriguez, J.; Below, R.; Guha-Sapir, D. *Annual Disaster Statistical Review 2009: The Numbers and Trends*; Center of Research on the Epidemiology of Disasters (CRED): Brussels, Belgium, 2010.

20. Zhang, Q.; Zhang, J.; Jiang, L.; Liu, X.; Tong, Z. Flood disaster risk assessment of rural housings—A case study of Kouqian Town in China. *Int. J. Environ. Res. Public Health* **2014**, *11*, 3787–3802. [CrossRef]

21. Kia, M.B.; Pirasteh, S.; Pradhan, B.; Mahmud, A.R.; Sulaiman, W.N.A.; Moradi, A. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environ. Earth Sci.* **2011**, *67*, 251–264. [CrossRef]

22. Eakin, H.; Lerner, A.M.; Murtinho, F. Adaptive capacity in evolving peri-urban spaces: Responses to flood risk in the Upper Lerma River Valley, Mexico. *Glob. Planet. Chang.* **2010**, *20*, 14–22. [CrossRef]

23. Dou, J.; Yamagishi, H.; Pourghasemi, H.R.; Yunus, A.P.; Song, X.; Xu, Y.; Zhu, Z. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Nat. Hazards* **2015**, *78*, 1749–1776. [CrossRef]

24. Rahmati, O.; Falah, F.; Naghibi, S.A.; Biggs, T.; Soltani, M.; Deo, R.C.; Cerda, A.; Mohammadi, F.; Tien Bui, D. Land subsidence modelling using tree-based machine learning algorithms. *Sci. Total Environ.* **2019**, *672*, 239–252. [CrossRef]

25. Bajabaa, S.; Masoud, M.; Al-Amri, N. Flash flood hazard mapping based on quantitative hydrology, geomorphology and GIS techniques (case study of Wadi Al Lith, Saudi Arabia). *Arab. J. Geosci.* **2013**, *7*, 2469–2481. [CrossRef]

26. Sanyal, J.; Lu, X.X. Application of remote sensing in flood management with special reference to monsoon Asia: A review. *Nat. Hazards* **2004**, *33*, 283–301. [CrossRef]

27. Nandi, A.; Mandal, A.; Wilson, M.; Smith, D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environ. Earth Sci.* **2016**, *75*, 465. [CrossRef]

28. Abdellatif, M.; Atherton, W.; Alkhaddar, R.; Osman, Y. Flood risk assessment for urban water system in a changing climate using artificial neural network. *Nat. Hazards* **2015**, *79*, 1059–1077. [CrossRef]

29. Aziz, K.; Rahman, A.; Fang, G.; Shrestha, S. Application of artificial neural networks in regional flood frequency analysis: A case study for Australia. *Stoch. Environ. Res. Risk Assess.* **2013**, *28*, 541–554. [CrossRef]

30. Das, S. Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas Basin, India. *Remote Sens. Appl. Soc. Environ.* **2019**, *14*, 60–74. [CrossRef]

31. Mair, A.; Fares, A. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *J. Hydrol. Eng.* **2011**, *16*, 371–383. [CrossRef]

32. Jiang, Z.; Huete, A.R.; Chen, J.; Chen, Y.; Li, J.; Yan, G.; Zhang, X. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens. Environ.* **2006**, *101*, 366–378. [CrossRef]

33. Fayne, J.V.; Bolten, J.D.; Doyle, C.S.; Fuhrmann, S.; Rice, M.T.; Houser, P.R.; Lakshmi, V. Flood mapping in the lower Mekong River Basin using daily MODIS observations. *Int. J. Remote Sens.* **2017**, *38*, 1737–1757. [CrossRef]

34. Zong, Y.; Tooley, M.J. A historical record of coastal floods in Britain frequencies and associated storm tracks. *Nat. Hazards* **2003**, *29*, 13–36. [CrossRef]

35. Islam, A.S. Improving flood forecasting in Bangladesh using an artificial neural network. *J. Hydroinform.* **2010**, *12*, 351–364. [CrossRef]

36. Buscema, M. Back propagation neural networks. *Subst. Use Misuse* **1998**, *33*, 233–270. [CrossRef]
37. Wang, L.; Zeng, Y.; Chen, T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Syst. Appl.* **2015**, *42*, 855–863. [CrossRef]
38. Karlik, B.; Olgac, A.V. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int. J. Artif. Intell. Expert Syst.* **2011**, *1*, 111–122.
39. Sibi, P.; Jones, S.A.; Siddarth, P. Analysis of different activation functions using back propagation neural networks. *J. Theor. Appl. Inf. Technol.* **2013**, *47*, 1264–1268.
40. Latt, Z.Z.; Wittenberg, H. Improving flood forecasting in a developing country: A comparative study of stepwise multiple linear regression and artificial neural network. *Water Resour. Manag.* **2014**, *28*, 2109–2128. [CrossRef]
41. Ciaburro, G. *Matlab for Machine Learning*; Packt: Birmingham, UK, 2017.
42. Dang, V.-H.; Dieu, T.B.; Tran, X.-L.; Hoang, N.-D. Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier. *Bull. Eng. Geol. Environ.* **2018**, *78*, 2835–2849. [CrossRef]
43. Campolo, M.; Soldati, A.; Andreussi, P. Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* **2003**, *48*, 381–398. [CrossRef]
44. Taylor, R. Interpretation of the correlation coefficient: A basic review. *J. Diagn. Med. Sonogr.* **1990**, *6*, 35–39. [CrossRef]
45. Ratner, B. The correlation coefficient: Its values range between+1/−1, or do they? *J. Target. Meas. Anal. Mark.* **2009**, *17*, 139–142. [CrossRef]
46. Zhao, G.; Pang, B.; Xu, Z.; Yue, J.; Tu, T. Mapping flood susceptibility in mountainous areas on a national scale in China. *Sci. Total Environ.* **2018**, *615*, 1133–1142. [CrossRef] [PubMed]
47. Papaioannou, G.; Vasiliades, L.; Loukas, A. Multi-criteria analysis framework for potential flood prone areas mapping. *Water Resour. Manag.* **2014**, *29*, 399–418. [CrossRef]
48. Liu, Y.; Feng, G.; Xue, Y.; Zhang, H.; Wang, R. Small-scale natural disaster risk scenario analysis: A case study from the town of Shuitou, Pingyang County, Wenzhou, China. *Nat. Hazards* **2014**, *75*, 2167–2183. [CrossRef]
49. Chang, C.-H. *Integrated Platform for Application of High-Performance 2D Inundation Simulation*; Water Resources Planning Institute: Taichung, Taiwan, 2016.