Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks

Mohammad Ahmadlou1 | A’kif Al-Fugara2 | Abdel Rahman Al-Shabeeb3 | Aman Arora4 | Rida Al-Adamat3 | Quoc Bao Pham5,6 | Nadhir Al-Ansari7 | Nguyen Thi Thuy Linh8,9 | Hedieh Sajedi10

1GIS Department, Geodesy and Geomatics Faculty, K. N. Toosi University of Technology, Tehran, Iran
2Department of Surveying Engineering, Faculty of Engineering, Al al-Bayt University, Mafraq, Jordan
3Department of GIS and Remote Sensing, Institute of Earth and Environmental Sciences, Al al-Bayt University, Mafraq, Jordan
4Department of Geography, Faculty of Natural Sciences, New Delhi, India
5Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam
6Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam
7Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187, Lulea, Sweden
8Institute of Research and Development, Duy Tan University, Danang 550000, Vietnam
9Faculty of Environmental and Chemical Engineering, Duy Tan University, Danang 550000, Vietnam
10Department of Computer Science, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran

Abstract
Floods are one of the most destructive natural disasters causing financial damages and casualties every year worldwide. Recently, the combination of data-driven techniques with remote sensing (RS) and geographical information systems (GIS) has been widely used by researchers for flood susceptibility mapping. This study presents a novel hybrid model combining the multilayer perceptron (MLP) and autoencoder models to produce the susceptibility maps for two study areas located in Iran and India. For two cases, nine, and twelve factors were considered as the predictor variables for flood susceptibility mapping, respectively. The prediction capability of the proposed hybrid model was compared with that of the traditional MLP model through the area under the receiver operating characteristic (AUROC) criterion. The AUROC curve for the MLP and autoencoder-MLP models were, respectively, 75 and 90, 74 and 93% in the training phase and 60 and 91, 81 and 97% in the testing phase, for Iran and India cases, respectively. The results suggested that the hybrid autoencoder-MLP model outperformed the MLP model and, therefore, can be used as a powerful model in other studies for flood susceptibility mapping.
1 | INTRODUCTION

In the recent past decades, the number of climatological disasters such as storms, floods, tsunamis, cyclones, droughts, etc. has increased in a dramatic way (Gaillard, 2007; Shaluf, 2007). Most of the concentration of climatological disasters can be found in the tropical and sub-tropical belts (Diaz, 2006). Many countries, for example, Iran, India, Bangladesh, Egypt, Sudan, are the main victims of these disasters. Floods are one of the most devastating and deadliest natural disasters affecting human life worldwide (Dassanayake, Burzel, & Oumeraci, 2015; Elnazer, Salman, & Asmoay, 2017; Kousky, 2018). Originated by climatic-hydrologic causes, the flood phenomenon refers to a situation where the river flow and water level rise unexpectedly (Judi, Rakowski, Waichler, Feng, & Wigmosta, 2018; Schumm & Lichty, 1965). Flood intensity is dependent on the geographical location and the climatic and geological conditions (Li, Wu, Dai, & Xu, 2012; Testa, Zuccala, Alcrudo, Mulet, & Soares-Frazão, 2007). Every year, huge volumes of floodwater destroy residential buildings and agricultural lands and cause major casualties and financial damages (Abbas, Amjath-Babu, Kächele, & Müller, 2015; Teng, Hsu, Wu, & Chen, 2006; V. Alphen, Van Beek, & Taal, 2006).

Floods are considered one of the most important and highly destructive natural hazards in Iran and their frequency and intensity have increased in the recent years (Ahmadlou et al., 2019; Arabameri, Rezaei, Cerdà, Conoscenti, & Kalantari, 2019; Khosravi et al., 2019; Rahmati, Pourghasemi, & Zeinivand, 2016; Termeh, Kornejady, Pourghasemi, & Keesstra, 2018). Iran has been experiencing floods of different intensities every year as it has a semi-arid to arid climate with little and mostly showery annual precipitations having non-uniform spatial and temporal distributions (Sharifi Garmdareh, Vafakhah, & Eslamian, 2018). For example, the flood occurred in 25 Iranian provinces during the first week of March 2019 left at least 19 people dead and billions of dollars’ worth of damage. The recent observations reveal that the changes in the amount and intensity of the precipitation vary in different regions of the Indian subcontinent due to escalating temperature induced by global warming (Pachauri et al., 2014). The regions which were receiving less precipitation earlier are now getting more amount of water due to climate change and due to increasing discharge in catchments are implying more risks of floods (Field, Barros, Stocker, & Dahe, 2012; Pachauri et al., 2014). Hence, flood hazard prediction and zoning in susceptible regions are highly important and can help reduce the damages caused by this phenomenon. This study examines the flood susceptibility in two study areas in Iran and India. In general, there are two approaches in various studies to model different phenomena such as flood, landslide and so on. The first approach is to build a model and test it in various regions (Shafizadeh-Moghadam, Asghari, Taleai, Helbich, & Tayyebi, 2017). In this approach, to test the performance of the model, it is used in different regions and on different data. In the second approach, various models are tested in one study area and their performances are examined (Ahmadlou et al., 2019). This study uses the first approach for flood susceptibility mapping.

Although the rainfall-runoff models can be used for zoning of flood-prone regions, these models require access to different data that are usually not available at a regional scale (Smith & Ward, 1998). Hence, in recent years, a combination of different sciences such as statistics, machine learning, and expert-based models along with geographical information systems (GIS) and remote sensing (RS) have been used for flood susceptibility mapping by researchers (Chapi et al., 2017; Costache et al., 2020a; Hong et al., 2018a; Hong et al., 2018b; Lee, Kim, Jung, Lee, & Lee, 2017; Yariyan, 2020; Youssef, Pradhan, & Sefry, 2016). In other words, by combining the GIS and RS data with the aforementioned methods, environmental decision-makers and managers have been provided with a powerful tool that helps them to monitor and manage such phenomena (Chao et al., 2018; Costache et al., 2020b; Shafizadeh-Moghadam, Valavi, Shahabi, Chapi, & Shirzadi, 2018; Tehrany, Lee, Pradhan, Jebur, & Lee, 2014; Wang, Zhang, Van Beek, Tian, & Bogaard, 2020). Two main approaches have been commonly and generally taken by researchers in recent years for flood susceptibility mapping: (a) expert-based approaches (Fernández & Lutz, 2010; Khosravi, Nohani, Maroufinia, & Pourghasemi, 2016; Tang, Yi, Wang, & Xiao, 2018), and (b) data-driven approaches (Bui et al., 2020; Cao et al., 2020; Costache & Bui, 2020; Lv & Qiao, 2020; Quan, Hao, Xifeng, & Jingchun, 2020; Yang & Chen, 2019).
In the expert-based approach, the opinion of experts is first used in the form of information layers to determine the factors effective in the flood occurrence (Fernández & Lutz, 2010). The multi criteria decision making (MCDM) methods, such as the AHP, are then used to weight the factors and, finally, these factors are combined with their coefficients (Souissi et al., 2019). These methods mostly rely on the technical knowledge of the experts and are, therefore, prone to errors (Khosravi et al., 2019). In the data-driven approaches, various statistical methods, machine learning, and data mining techniques are used based on the historical floods in the region along with the characteristics of the regions experiencing the same phenomenon such as the topographical, climatic, and geological characteristics (Kia et al., 2012; Wang et al., 2015). In fact, the working mechanism of these methods makes use of the existing data on the location of the flood occurrence in the past and their characteristics. This approach, based on the historical floods, provides the researchers with an accurate tool (Khosravi et al., 2019).

Data-driven approaches have been employed in various studies to prepare flood susceptibility maps (Bui et al., 2020; Bui et al., 2016; Costache et al., 2020c; Khosravi et al., 2019; Kia et al., 2012). However, with the advancements in machine learning and data mining techniques, more advanced models are put forth in this field every day, enabling researchers to combine them with GIS and RS for zoning and detection of susceptible regions (Bui et al., 2020). Artificial neural networks (ANNs) are among the most widely used algorithms in various disciplines (Costache & Bui, 2019; Kia et al., 2012; Shi, Wang, Tang, & Zhong, 2020; Yang et al., 2019). This model is a highly powerful tool and has been reported to provide appropriate results in various studies (Costache et al., 2019; Kia et al., 2012; Pradhan, 2010). However, the traditional ANN can get trapped in local optima through random initialization, which can be prevented by using a deep-learning algorithm, autoencoder neural networks, based on the MLP neural network to obtain a better initialization (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010). In fact, an autoencoder is used to improve the accuracy and efficiency of MLP neural networks through a nonlinear mapping that both reduces the dimension of the problem and serves as a feature extraction procedure (Hernández, Sanchez-Anguix, Julian, Palanca, & duque, 2016, Oliveira et al., oliveira, barbar, & soares, 2014). The MLP network is then used for prediction and estimation. Hence, the main objective of this study is to obtain flood susceptibility maps using a model combining autoencoders and MLP neural networks.

2 | STUDY AREAS AND DATA SET

2.1 | The first study area

The first study region is located in Golestan Province in Iran extending between latitudes 36° 27′ and 38° 14′N and between longitudes 53° 40′ and 56° 30′E (Figure 1). The study area has an area of 12,050 km², altitude range of −147-3,348 m above sea level, and precipitation range of 180–880 mm. The northern parts of the region have less rainfall intensity than the southern parts. The north of this region is surrounded by agricultural lands while the south is surrounded by forest areas. Part of the Alborz mountain range is also located in the south of the region. The roughness of the region is such that they can be clearly divided into plains and mountains. In this study area, the slope of the land decreases from the heights to the plains. At the confluence of the plains and foothills of northern Alborz, due to the severity of erosion and alluvial density, part of the old roughness is covered by newer sediments and only in some places has appeared as hills. Deadly floods occurred in this province in 2001, 2002, 2005, and 2019.

2.2 | The second study area

The second study area (Figure 2) which shares Upper and Lower Ganga basins of Ganga River Basin (GRB), is holding one of the densest populated region in the Indian territory (Singh, 1971). This area has altitude range of 45–96 m above sea level, and precipitation range of 1,001–1,281 mm. It faces most devastating floods every year during monsoon period in influence of South-West monsoon rainfall (Vittal et al., 2016). The unparalleled distribution of the population in the region put the lives in danger in unprecedented situation during flood. On an average hundreds of lives lost or get missing every year in India due to flood and mostly in GRB only. The colossal loss of properties and agricultural products happen in the recurring period is recorded in various government reports for this study area. The study area is the confluence zone of major rivers- Ghaghara, Gandak, Ganga, Son, Kosi, and other minor tributaries of Ganga (Arora, Pandey, Siddiqui, Hong, & Mishra, 2019). Hence, the risk of inundation during monsoon period is much higher than other regions of India.

The region experiences four seasons- summer season (April–May), monsoon season (June–September), post-monsoon season (October–December) and winter season (January–March) (Dimri et al., 2019). The sub-tropical-humid region experiences the highest temperature from...
April to July. During these months the maximum temperature recorded between 35 to 45°C. The lowest temperature recorded in December and January, where the downfall of the lowest temperature is recorded up to 03–04°C. On arrival of monsoon, in late June and early July, the high intensity rainfall devoured the upper catchments of GRB, resultant the lower basins (study area) experiences an unprecedented situation during late August and September. The increment in discharge in the rivers has been noticed 50 to 100 times greater than average discharge (Shukla and Singh, 2004) and cause the flood.

2.3 | Data set

The flood inventory map and flood conditioning factors are required for flood susceptibility mapping using data-driven methods. In fact, the flood inventory map acts as the target variable to be modelled, and the flood conditioning factors represent the independent variables (predictors) used for modelling the target variable.

The flood inventory map contains the location of past floods (147 and 300 flood events for Iran and India, respectively). Various methods are available to determine these points including field observations and satellite images and Google Earth imagery. One hundred and forty seven flood points in Golestan Province were recorded by the Golestan Water resources organisation. For GRB, Landsat 5 multispectral scanner (MSS) and shuttle radar topography mission (SRTM) version 4 digital elevation model (DEM) satellite images were used to create flood zones (Table 1). Then, 300 flood points were generated by creating random points tool in GIS environment. To identify non-flood points, random sampling was first performed in ArcGIS 10.4 software and finally, from the generated points, 300 non-flood points where flooding is not able to occur were selected using field surveys, topography maps and Google Earth software. Such a process was used to generate 147 non-flood points in Golestan Province. During the modelling, 70% of the flood and non-flood points were used for training and 30% for testing purposes. Flood occurrence in a region is affected by various factors (Kourgialas & Karatzas, 2017;
Talukdar et al., 2020). For Golestan Province, nine factors including altitude, slope, aspect, plan curvature, topographic wetness index (TWI), lithology, distance to drainage, rainfall, and land use. For GRB, 12 factors including altitude, slope, aspect, plan curvature, distance from the river, rainfall, river density, TWI, land use land cover, distance from roads, soil type, and geomorphology factors were selected based on previous studies and data availability. Tables 2 and 3 show source of input data, original format of source data (vector and raster), original map scale or spatial resolution of source data and derived map (factor). Altitude is considered one of the important factors in most studies related to flood susceptibility mapping (Costache, 2019; Janizadeh et al., 2019). In high-altitude regions, flood occurrence is highly unlikely, whereas flat regions have a high potential for flooding (Janizadeh et al., 2019). This factor can be prepared using DEM. Some topographic factors such as slope, aspect, plan curvature and TWI are also extracted from DEM. The slope map, due to its direct effect on surface runoff, is another factor influencing flood occurrence so that an increase in the slope reduces the time for surface infiltration, hence allowing a larger volume of water to enter the river bed causing flooding (Tehrany, Pradhan, Mansor, & Ahmad, 2015). Aspect and curvature are two other height factors considered in this study. Moreover, Equation 1 is used to calculate the TWI as a water-related factor highly important in flood occurrence (Pourghasemi, Pradhan, Gokceoglu, & Moezzi, 2012; Ali et al., 2020):

\[
TWI = \ln\left(\frac{A}{\tan\alpha}\right)
\]  

**FIGURE 2** Location of the first study area and flood inventories (Iran)

**TABLE 1** The satellite and DEM data characteristic details used in GRB

| S. No. | Satellite | Duration | Acquisition date | Spatial reference (projected) |
|--------|-----------|----------|------------------|-------------------------------|
| 1      | Landsat 5 MSS | Preflood | 28 May 2008 | Projection: UTMDatum & spheroid: WGS84Zone: 44 N |
| 2      |          | Postflood | 19 October 2008 | |
| 3      |          | During flood | 9 January 2008 | |
| 4      | SRTM v4 DEM | – | 11–22 Feb 2000 | |

Abbreviations: DEM, digital elevation model; GRB, Ganga River Basin.
In Equation 1, A is the catchment area, and \( \alpha \) is the slope angle. Rainfall is another influential factor in flood occurrence (Bracken, Cox, & Shannon, 2008). Floods can occur when the amount of water flowing from a catchment exceeds the capacity of its drains. However, flood occurrence due to rainfall is also dependent on other factors such as land use and land cover, soil type, and characteristics of waterways such as size and shape (Ahmadlou et al., 2019). The geology factor was also used in the modelling due to its direct effect on infiltration and surface runoff. The activities associated with land use (e.g., urban development or deforestation) are one of the most important human factors affecting flood occurrence. The effect of this factor can vary from one land use type to another as well as at small, medium, or large scales. For example, lack of vegetation and/or urban growth in a region can lead to floods. Land use/land cover (LULC) of Golestan Province was prepared by the maximum likelihood (ML) supervised classification technique using Landsat 8 Operational Land Imager (OLI) satellite image. Also, for GRB LULC map, the ML method was used in the Climate Change Initiative (CCI) LULC 2008 dataset and the study area part has been extracted from the ready-to-use dataset and used in the work as an LULC conditioning factor. Distance to drainage is another

| TABLE 2 | Details of input data and derived data parameters for Golestan Province |
|---------|---------------------------------------------------------------------|
| Data layers | Input data format (original) | Scale/resolution | Derived data format | Scale/resolution | Source of the data |
| Altitude, distance from river, curvature, aspect, slope, TWI | Raster | 1 arc sec global 30×30 m spatial resolution | Raster | 30 m | SRTM DEM, United States Geological Survey (USGS) |
| Rainfall | - | | | | Golestan County Meteorological bureau |
| Land use | - | | | | Landsat 8 OLI |
| Lithology | Vector | 1:100000 | | | Geological survey and mineral Exploration of Iran |

| TABLE 3 | Details of input data and derived data parameters for GRB |
|---------|---------------------------------------------------------------------|
| Data layers | Input data format (original) | Scale/resolution | Derived data format | Scale/resolution | Source of the data |
| Altitude, distance from river, curvature, slope aspect, slope angle, TWI | Raster | 1 arc sec global 30×30 m spatial resolution | Raster | 30 m | SRTM DEM, (USGS) |
| Distance from road | - | | | | Open street map |
| Rainfall distribution | - | | | | Climate forecast system reanalysis (CFSR) 2008 |
| Land use/land cover | - | | | | CCI LULC 2008 Map |
| Soil | - | | | | Food and Agriculture Organization (FAO) |
| Geomorphology; river density | Vector | - | | | Google earth |
| Flood Inventory Training & Validation Data | - | | | | Landsat 8 OLI 10.08.2008 |

Abbreviations: DEM, digital elevation model; GRB, Ganga River Basin, TWI, topographic wetness index.
important factor affecting the flood occurrence (Tehrany, Pradhan, & Jebur, 2013; Wang et al., 2015). Figures 3 and 4 show the factors along with their categories that influence flood occurrence for Iran and India cases, respectively. In this study, ArcGIS and Environment for Visualising Images (ENVI) softwares were used to prepare of conditioning factors. The modelling process was programmed in the MATLAB software.

3 | METHODS

Figure 5 shows the different stages of the research using the employed models. After preparing the flood inventory map and flood conditioning factors, the frequency ratio (FR) model was used to determine the correlation between flood occurrence and the considered variables. In the next step, two models, namely the MLP and autoencoder-MLP, were used for the preparation of susceptibility maps, and
then the results were compared and assessed using ROC. The FR, MLP, and autoencoder-MLP are, respectively, discussed in Sections 3.1, 3.2, and 3.3.

### 3.1 | Frequency ratio

The FR determines the quantitative correlation between flood occurrence and the various factors affecting it (Oh, Kim, Choi, Park, & Lee, 2011). For each class of variables, FR is equal to the occurrence percentage of floods in that class to the percentage of area covered by that class (Lee & Sambath, 2006). Hence, the Equation 2 is developed to determine the FR value for each class of the variables (Lee & Sambath, 2006):

\[
FR = \frac{N_{pix}(S_i)}{\sum_{i=1}^{n} N_{pix}(S_i)} \quad \frac{N_{pix}(N_i)}{\sum_{i=1}^{n} N_{pix}(N_i)}
\]

where \( n \) is the number of classes for the considered variable, \( N_{pix}(S_i) \) is the number of pixels containing floods in the \( i^{th} \) class of the considered variable, and \( N_{pix}(N_i) \) is the number of all pixels for that class. It can be observed that higher FR indicate a more powerful correlation between flood occurrence and the respective variable and, conversely, lower ratios suggest a weaker correlation.

### 3.2 | Multilayer perceptron

Considered as one of the most widely used and most accurate machine learning techniques in various fields, ANNs are highly capable in modelling nonlinear relationships between target variable and explanatory variables (Kia et al., 2012). An MLP neural network is composed of a single input layer, multiple hidden layers, and a single output layer. Each of these layers is made of several neurons as the smallest information processing units (Jain, Mao, & Mohiuddin, 1996, Zurada, 1992). In these networks, the output of the first layer (input layer) is used as the inputs to the next layer (hidden layer). This trend continues in the following layers up to a certain number of layers until the outputs of the last hidden layer are fed to the output layer as the inputs. The MLP includes a set of weights that should be tuned for the training stages of the neural network. The back-propagation (BP) method is common in the training of MLP networks (Jain et al., 1996). This algorithm randomly selects the initial weights, biases and compares the output computed through the network with the real values. The difference between the computed and real outputs is obtained using the criteria such as the root-mean-square error (RMSE) or mean square error (MSE), after which the
Network weights are updated based on the delta rule. Hence, the overall network error is distributed among the various nodes in the network (Jain et al., 1996).

This process continues until the error reaches a stable level. The MLP model specifications in this study are as follows:

A total of 4 fully connected layers were used in this sequential layer, such that any given neuron in each layer is connected to all neurons in the next layer (for example, Golestan Province in Figure 6). Of these 4 layers, 3 were used for data processing, and the last layer was used for prediction. A total of 15, 10, 5, and 1 neurons were considered in the first hidden layer, the second layer, the third layer, and fourth or the output layer, respectively. The rectified linear unit (ReLU; (Nair & hinton, 2010)) was applied to all 4 layers as the activation function after processing. The ReLU formula is as Equation 3 (Nair & hinton, 2010):

\[ R(z) = \begin{cases} 
  z & z > 0 \\ 
  0 & z = 0 
\end{cases} \] (3)

This function is not linear and provides the same benefits as Sigmoid but with better performance (Zeiler

**FIGURE 4** Flood conditioning factors for GRB. GRB, Ganga River Basin
et al., 2013). After finishing the training run, this MLP network is applied to the test data to assess its accuracy.

3.3 | Autoencoder-MLP

This model is composed of two structures, namely the autoencoder neural network (Chicco, Sadowski, & baldi, 2014; Sun et al., 2016) and the MLP neural network. Instead of feeding the input data directly to the MLP for prediction, the autoencoder neural network is initially used for feature extraction, after which the results are provided to the MLP neural network for prediction.

Autoencoders are generally neural networks capable of learning to produce an output layer similar to the input layer (Chicco et al., 2014, Sun et al., 2016). This process is carried out in two stages by an encoder and a decoder. In the first stage, the input data are compressed in the hidden layer by the encoder, after which they are reconstructed by the decoder using the hidden layer (Chen, Shi, Zhang, wu, & guizani, 2017). In this model, the objective is not to train the autoencoders to produce the decoder output but to use the hidden layer produced...
by the encoder. This hidden layer is, in fact, a compressed representation of the data and, as a result, the hidden layer of the autoencoder contains suitable low-volume features of the initial data positively affecting the prediction results (Chen et al., 2017; Sun et al., 2016). As a key capability for making correction predictions, the autoencoders can also discover the nonlinear relationships between variables (Chen et al., 2017). Hence, autoencoders are used for two reasons, namely compressing the data and extracting nonlinear relationships between variables.

The stack autoencoder (SAE) was used in this study (Shin, Orton, Collins, Doran, & Leach, 2012; Vincent et al., 2010). Figure 7 shows the SAE architecture. This encoder is a neural network composed of several layers of autoencoders, such that the outputs of each autoencoder are fed to the next autoencoder as the input (Shin et al., 2012, Vincent et al., 2010). As mentioned earlier, two stages are involved in the combined autoencoder-MLP model. In the first stage, the features are extracted through multiple layers of encoders using the SAE. In the second stage, the features extracted from the last SAE layer are given to the MLP layer as the input for prediction.

The autoencoder-MLP model used in this study includes a total of 5 layers, the 4 first of which are associated with the autoencoder, and the last layer belongs to the MLP neural network (Figure 7). A total of 5 neurons were considered in the first autoencoder layer, 15 in the second layer, 10 in the third layer, 5 in the fourth layer, and 1 in the last layer (Figure 7). The activation function was applied to all layers after preprocessing. A linear function was used for the third layer, whereas the ReLU was used for the rest of layers. The processing was performed as described earlier, during which the output of each layer is fed to the next layer as the input. Hence, after the extraction of features in the first stage by the autoencoders, the MLP in the second stage performs the prediction process and completes the model.

Moreover, the first encoder of the SAE is shown in Figure 8. Once the training stage is finished, the autoencoder-MLP model is applied to the test data to investigate its accuracy.

$$Z^{(l)} = W^{(l)}h^{(l-1)} + b^{(l)},$$

where \(w, b\) are the weight and bias vectors, respectively. In Equation 4, \(l\) is the number of hidden layers and \(h^{(l-1)}\) is the \((l-1)th\) hidden layer whose values is taken from the previous hidden layer \(l\). Therefore, in the first stage of the model, the features are extracted through multiple layers of encoders using the SAE. In the second stage, the features extracted from the last SAE layer are given to the MLP layer as the input for prediction.

$$f(z) = \begin{cases} \text{relu} : f(z) = \max(0,z) \\ \text{linear} : f(z) = w \cdot z + b \end{cases}$$

where \(w, b\) are the weight and bias vectors, respectively. In Equation 4, \(l\) is the number of hidden layers and \(h^{(l-1)}\) is the \((l-1)th\) hidden layer whose values is taken from the previous hidden layer \(l\). Therefore, in the first stage of the model, the features are extracted through multiple layers of encoders using the SAE. In the second stage, the features extracted from the last SAE layer are given to the MLP layer as the input for prediction.

The autoencoder-MLP model used in this study includes a total of 5 layers, the 4 first of which are associated with the autoencoder, and the last layer belongs to the MLP neural network (Figure 7). A total of 5 neurons were considered in the first autoencoder layer, 15 in the second layer, 10 in the third layer, 5 in the fourth layer, and 1 in the last layer (Figure 7). The activation function was applied to all layers after preprocessing. A linear function was used for the third layer, whereas the ReLU was used for the rest of layers. The processing was performed as described earlier, during which the output of each layer is fed to the next layer as the input. Hence, after the extraction of features in the first stage by the autoencoders, the MLP in the second stage performs the prediction process and completes the model.

Moreover, the first encoder of the SAE is shown in Figure 8. Once the training stage is finished, the autoencoder-MLP model is applied to the test data to investigate its accuracy.
4 | RESULTS AND DISCUSSION

4.1 | The role of conditioning factors on flood occurrence

The FR was used to determine the correlation between each class of variables and floods. The results are presented in Tables 1 and 2 for both cases. As shown in Table 4, the 45–270 m height class, the 0–3° slope class, the flat aspect class, and the flat class in the plane curvature factor were among the most important classes that were assigned the highest weights by the FR method. On the contrary, the altitude above 1,260 m class received the lowest weight and, therefore, this class plays the least important role in flood occurrence. Moreover, the 500–1,000 m class in the distance to drainage factor, the Proterozoic class associated with lithology factor, the water use class, and the 600–800 mm precipitation class had the highest effect on flood occurrence. Other important classes can be seen in Table 5 for other factors.

4.2 | Application of MLP and autoencoder-MLP in flood susceptibility modelling

After conducting the correlation analysis and determining the weight of each class of variables, the MLP and autoencoder-MLP models were implemented in Python. Seventy percent of the datasets were used as the train data, and the remaining 30% were used to test the models.

After training of 200 iterations, all cells in the two regions were entered into the MLP and autoencoder-MLP models and their flood susceptibility index was
calculated. Figures 9 and 10 show the flood susceptibility maps for the two models for Iran and India cases, respectively. After making the prediction outputs of the two models for the entire region, the natural break classification method was used to classify these maps into five classes including very low, low, moderate, high, and very high. Natural break classification is one of the most common methods in natural hazard mapping to classify the various classes of conditioning factors as well as susceptibility maps. This method identifies real classes within the data. This is useful because it creates maps that have accurate representations of trends in the data (Baz, Geymen, & Er, 2009). For example, for a map with different values, this method finds areas that have close values. Geometric interval or quantiles are other methods of splitting, which do not create the best division. For example, quantiles divides only ranges into classes with equal distances. These two methods are easy and fast, but they do not produce the desired output. For Golestan Province, these five classes cover, respectively, 33.76, 6.78, 6.71, 6.68, and 46.07% of the total study area for MLP model, and 19.52, 13.85, 15.28, 14.09, and 37.26% for the autoencoder-MLP model. The results indicate that 52.75 and 51.35% of the entire region falls into the high and very high flood susceptibility classes in the MLP and the autoencoder-MLP models, respectively. By investigating the characteristics of the cells which were classified into the high flood susceptibility class in the MLP model, it can be clearly observed that the majority of these cells are in the 45–270 m height class, in the Cenozoic geological layer class, and in the agricultural lands in the region. For GRB, these 5 classes (very low, low, moderate, high and very high) cover, respectively, 26.24, 29.26, 22.17, 16.02, and 6.31% of the entire region for the MLP model, and 10.74, 21.97, 35.37, 23.74, and 8.18% for the autoencoder-MLP model. The results indicate that 22.33 and 31.92% of the entire region falls into the high and very high flood susceptibility classes in the MLP and the autoencoder-MLP models, respectively.

It is noteworthy that the first study area, Golestan province, covers more susceptible lands, 52%, in terms of combined share of high and very high susceptible lands in comparison to the second study area, Middle Ganga Plain, where the share stands for the same category is 27% (average values of MLP and autoencoder-MLP outputs). The main reason behind this odd share, for both of the study areas with same model, is the altitude and slope of the region. In Golestan province the crescent shaped...
upper part covers the low altitude regions ranging from −147 m to 270 m (Figure 3(c)) and low slope (0–3°C); Figure 3(a)) is also recorded for the same place. Also, the high rainfall is received in the upper catchment of the study area which provides surplus water to the lower catchments (low altitude part) of the region. Ultimately, this part of the Golestan province having low altitude and low slope characteristics receive more water during and after rainfall. These are the main reasons behind 54% share of high & very high susceptible lands to flood. Whereas, in the second study area, Middle Ganga Plain, the complete region characterised with low altitude zones and low slope (Figure 4) and it’s a part of the lower catchment of Gang River Basin, India. Therefore, the high and very high lands, 24%, are only visible along to rivers and low depressions only.

Based on the maps produced by both models, in Golestan Province the areas in the very low to moderate susceptibility classes are mainly located in the southern and southwestern parts of the region where the Alborz mountain range acts as a barrier preventing the entry of humidity derived from the Caspian Sea into these regions. Consequently, these areas have low rainfall and a dry climate. As a result, the probability of flood occurrence in these parts of the study area is low. The areas with high flood susceptibility are located in the northern and northwestern parts of the region. Evaporation of the Caspian Sea increases the humidity in these areas giving rise to heavy precipitations that can lead to floods. The proximity of the water table to the ground surface, as well as the saturated soil in these areas, can increase the intensity of floods.

In the India case study, the low altitude floodplains (<50 m) of the region have recorded high and very high flood susceptible zones in produced maps from Autoencoder-MLP (Figure 10). The major concentration of high to very high susceptible zones can be observed in the complete eastern MGP where the major concentration of total annual average rainfall (>1,100 mm) is being recorded. The monsoon rainfall hits the area in the last of June and early July, submerged the low altitude basins first and causes an unprecedented situation (Arora et al., 2019).

During the monsoon period, the high volume of discharge of water from upstream influxes the downstream catchments and flood water spread over the region in the eastern parts (Bhatt & Rao, 2016). From the early flood records, it has been also observed that the sudden rainfall in the post-monsoon on already wet areas, flooded due to monsoon rainfall, brings more disaster in August and create havoc situation. Apart from both major factors, the river density plays a crucial role to distinguish the more and least flooded regions, high dense regions formed in permeable soil with low relief (altitude) regions (Gajbhiye, Mishra, & Pandey, 2014). The high dense streams’ regions in the central northern and north eastern parts account for high flood susceptible zones.

The high dense and higher amount of rainfall is being recorded in the confluence zones of rivers, which provide a higher probability of flood than other parts in the study area. It has been observed in the earlier studies that the flood probability is higher in the confluence point (Kadam & Sen, 2012). The study area is having the confluence zone in the eastern margin where four rivers (i.e., Ganga, Ghaghara, Son and Rapti rivers) meet and cause more discharge of flooded water in low relief basin in the eastern parts.

The area under the ROC curve was used for assessing the accuracy of the results from the MLP and autoencoder-MLP models. As shown in Figures 11 and 12, the area under the curve for the MLP and autoencoder-MLP models in Golestan Province were 79 and 97% in the training and 82 and 96% in the testing phases, and for GRB were 74 and 93% in the training and 81 and 97% in the testing phases, respectively, indicating that the autoencoder-MLP model outperforms the MLP model in terms of accuracy in both study areas. The
| Conditioning factors          | Classes                  | No. of pixels | No. of flood | FR  |
|------------------------------|--------------------------|---------------|--------------|-----|
| **Attitude (m)**             | -147–45                  | 3,664,414     | 16           | 0.19|
|                              | 45–270                   | 3,922,916     | 49           | 0.46|
|                              | 270–680                  | 3,889,090     | 22           | 0.23|
|                              | 680–1,260                | 3,859,869     | 11           | 0.11|
|                              | 1,260 <                  | 3,797,639     | 2            | 0.01|
| **Aspect**                   | Flat                     | 376,359       | 4            | 0.18|
|                              | North                    | 2,130,667     | 9            | 0.09|
|                              | Northeast                 | 3,193,858     | 22           | 0.13|
|                              | East                      | 2,145,605     | 9            | 0.09|
|                              | Southeast                 | 1,687,343     | 5            | 0.06|
|                              | South                     | 2,905,840     | 14           | 0.10|
|                              | Southwest                 | 1,832,706     | 12           | 0.13|
|                              | West                      | 1,964,231     | 12           | 0.12|
|                              | Northwest                 | 2,897,319     | 13           | 0.10|
| **Slope**                    | 0–3                      | 4,850,994     | 48           | 0.40|
|                              | 43.165                   | 2,959,892     | 20           | 0.27|
|                              | 43.264                   | 4,326,352     | 15           | 0.13|
|                              | 13–21                    | 3,983,431     | 10           | 0.10|
|                              | 21 <                     | 3,013,259     | 7            | 0.10|
| **Plan curvature**           | Convex                   | 8,350,014     | 40           | 0.27|
|                              | Flat                     | 2,529,811     | 22           | 0.48|
| **Distance of river (m)**   | 0–500                    | 2,423,954     | 28           | 0.28|
|                              | 500–1,000                | 2,162,674     | 37           | 0.42|
|                              | 1,000–2000               | 3,408,686     | 22           | 0.16|
|                              | 2000–3,000               | 2,079,079     | 4            | 0.05|
|                              | 3,000–4,000              | 2,305,451     | 9            | 0.10|
| **TWI**                      | 0–6.4                    | 7,564,764     | 29           | 0.15|
|                              | 6.4–9.2                  | 7,211,735     | 35           | 0.19|
|                              | 9.2–12                   | 2,588,480     | 19           | 0.29|
|                              | 12 <                     | 1,768,949     | 17           | 0.37|
| **Lithology**                | CENOZOIC                 | 7,109,996     | 80           | 0.35|
|                              | MESOZOIC                 | 2,992,033     | 8            | 0.09|
|                              | PALEOZOIC                | 1,862,050     | 6            | 0.10|
|                              | PROTEROZOIC              | 415,765       | 6            | 0.46|
| **Land use**                 | Forest                   | 4,620,888     | 22           | 0.11|
|                              | Agriculture              | 5,842,380     | 71           | 0.18|
|                              | Other                    | 137,086       | 3            | 0.26|
|                              | Water                    | 96,454        | 4            | 0.38|
|                              | Range                    | 1,680,991     | 4            | 0.07|
| **Rainfall (mm)**            | 400–200                  | 4,718,034     | 16           | 0.16|
|                              | 400–600                  | 7,384,134     | 41           | 0.26|
|                              | 600–800                  | 5,972,241     | 37           | 0.30|
|                              | 800–1,000                | 1,059,519     | 6            | 0.27|

Abbreviations: FR, frequency ratio; TWI, topographic wetness index.
| Conditioning factors | Classes                  | No. of pixels | No. of flood | FR  |
|----------------------|--------------------------|---------------|--------------|-----|
| Attitude (m)         | < 45                     | 381,933       | 23           | 3.06|
|                      | 45–50                    | 1,887,520     | 72           | 1.94|
|                      | 50–53.5                  | 1,648,521     | 35           | 1.08|
|                      | 53.5–58.0                | 3,335,707     | 52           | 0.79|
|                      | 58.0–61.6                | 1,468,939     | 19           | 0.66|
|                      | 61.6–65.7                | 1,386,021     | 8            | 0.29|
|                      | > 65.7                   | 574,727       | 1            | 0.09|
| Aspect               | Flat                     | 569,949       | 23           | 2.05|
|                      | North                    | 1,277,468     | 22           | 0.88|
|                      | Northeast                | 1,239,643     | 25           | 1.03|
|                      | East                     | 1,276,267     | 12           | 0.48|
|                      | Southeast                | 1,280,832     | 24           | 0.95|
|                      | South                    | 1,285,409     | 30           | 1.19|
|                      | Southwest                | 1,237,955     | 27           | 1.11|
|                      | West                     | 1,256,733     | 22           | 0.89|
|                      | Northwest                | 1,259,112     | 25           | 1.01|
| Slope                | 0–1                      | 3,190,780     | 71           | 1.13|
|                      | 1–3                      | 6,087,154     | 115          | 0.96|
|                      | 3–5                      | 1,162,388     | 19           | 0.83|
|                      | 5–7                      | 173,245       | 2            | 0.59|
|                      | > 7                      | 69,801        | 3            | 2.19|
| Plan curvature       | Convex                   | 4,265,700     | 81           | 0.97|
|                      | Flat                     | 2,190,592     | 47           | 1.09|
|                      | Concave                  | 4,227,076     | 82           | 0.99|
| Distance to river    | 0–600                    | 3,121,471     | 105          | 1.77|
|                      | 600–1,200                | 2,793,996     | 52           | 0.98|
|                      | 1,200–1,800              | 2,294,963     | 33           | 0.76|
|                      | 1,800–2,400              | 1,706,751     | 17           | 0.52|
|                      | > 2,400–3,000            | 885,701       | 3            | 0.18|
|                      | > 3,000                  | 227,394       | 0            | 0   |
| TWI                  | 7.33–10.90               | 3,614,449     | 55           | 0.77|
|                      | 10.90–12.33              | 2,997,964     | 54           | 0.91|
|                      | 12.33–14.06              | 2,065,668     | 46           | 1.13|
|                      | 14.06–16.27              | 1,044,900     | 25           | 1.21|
|                      | 16.27–18.77              | 487,132       | 7            | 0.73|
|                      | 18.77–22.32              | 398,168       | 21           | 2.68|
|                      | 22.32–31.84              | 44,616        | 2            | 2.27|
| Geomorphology        | FluOri - active flood plain | 1,122,878   | 35           | 1.59|
|                      | Meander scar             | 118,000       | 4            | 1.72|
|                      | Braid Bar                | 87,618        | 4            | 2.32|
|                      | Lateral Bar              | 113,306       | 3            | 1.35|
|                      | Marsh                    | 171,814       | 4            | 1.18|
|                      | Meander scar             | 9,665         | 0            | 0   |
| Conditioning factors                  | Classes                      | No. of pixels | No. of flood | FR  |
|--------------------------------------|------------------------------|---------------|--------------|-----|
| Channel Island                       |                              | 507,630       | 18           | 1.80|
| Palaeochannel                        |                              | 25,970        | 0            | 0   |
| WatBod - pond                        |                              | 8,569         | 1            | 5.94|
| FluOri - older flood plain           |                              | 4,080,832     | 66           | 0.82|
| Abandoned Channel                    |                              | 20,361        | 0            | 0   |
| Point Bar                            |                              | 102,313       | 4            | 1.99|
| FluOri - older alluvial plain        |                              | 1,552,977     | 5            | 0.16|
| Channel Bar                          |                              | 101,318       | 7            | 3.51|
| WatBod - river                       |                              | 39,116        | 0            | 0   |
| Cut-off meander                      |                              | 12,262        | 0            | 0   |
| Back swamp                           |                              | 131,051       | 6            | 2.32|
| Valley fill                          |                              | 400           | 0            | 0   |
| Oxbow Lake                           |                              | 9,952         | 3            | 15.33|
| Natural levee                        |                              | 27,848        | 0            | 0   |
| FluOri - younger alluvial plain      |                              | 2,092,527     | 26           | 0.63|
| WatBod - others                      |                              | 346,872       | 24           | 3.52|
| Land use                              |                              |               |              |     |
| Cropland                             |                              | 9,198,831     | 132          | 0.73|
| Vegetation                           |                              | 146,467       | 6            | 2.08|
| Settlement                           |                              | 164,958       | 1            | 0.31|
| Water                                |                              | 1,147,193     | 71           | 3.14|
| Rainfall (mm)                        |                              |               |              |     |
| 1,001–1,073                          |                              | 760,246       | 17           | 1.13|
| 1,074–1,123                          |                              | 2,323,774     | 48           | 1.05|
| 1,124–1,165                          |                              | 3,942,972     | 63           | 0.81|
| 1,166–1,212                          |                              | 2,677,011     | 55           | 1.04|
| 1,213–1,281                          |                              | 946,078       | 27           | 1.45|
| River density                        |                              |               |              |     |
| 0–2.55                               |                              | 614,232       | 3            | 0.25|
| 2.56–5.12                            |                              | 1,704,079     | 13           | 0.39|
| 5.13–7.68                            |                              | 2,765,879     | 46           | 0.84|
| 7.69–10.20                           |                              | 4,268,183     | 110          | 1.31|
| 10.21–12.79                          |                              | 1,197,072     | 32           | 1.36|
| 12.80–15.35                          |                              | 100,636       | 6            | 3.02|
| Distance to roads                    |                              |               |              |     |
| 0–500                                |                              | 4,668,876     | 53           | 0.58|
| 501–1,000                            |                              | 1,563,959     | 36           | 1.17|
| 1,001–2,000                          |                              | 2,325,422     | 51           | 1.11|
| 2,001–3,000                          |                              | 1,119,349     | 34           | 1.54|
| 3,001–4,000                          |                              | 588,356       | 21           | 1.81|
| 4,001–5,000                          |                              | 298,949       | 13           | 2.21|
| 5,001–7,045                          |                              | 93,274        | 2            | 1.09|
| Soil type                             |                              |               |              |     |
| CM-Cambisols                         |                              | 131,419       | 0            | 0   |
| CM-Cambisols                         |                              | 1,198,791     | 14           | 0.59|
| CL-Calcisols                         |                              | 572,852       | 2            | 0.18|
| CL-Calcisols                         |                              | 2,937,569     | 30           | 0.52|
| FL-Fluvisols                         |                              | 3,652,801     | 108          | 1.50|
| LX-Lixisols                          |                              | 2,161,693     | 56           | 1.31|

Abbreviations: DEM, digital elevation model; GRB, Ganga River Basin, TWI, topographic wetness index.
**FIGURE 9** The flood susceptibility maps of (a) MLP and (b) autoencoder-MLP models for Golestan Province. MLP, multilayer perceptron

**FIGURE 10** The flood susceptibility maps of (a) MLP and (b) autoencoder-MLP models for GRB. GRB, Ganga River Basin; MLP, multilayer perceptron

**FIGURE 11** ROC curves for MLP and autoencoder-MLP models in (a) training and (b) testing runs for Golestan Province. GRB, Ganga River Basin; MLP, multilayer perceptron; ROC, receiver operating characteristic
reason can be attributed to the extraction of effective features and the elimination of the co-linearity between the effective factors by the hybrid model.

Although MLP is one of the most famous and widely-used machine learning models, it has been used in few studies for flood susceptibility mapping. (Janizadeh et al., 2019) compared standalone MLP with alternating decision tree (ADT), functional tree (FT), kernel logistic regression (KLR), and quadratic discriminant analysis (QDA) models. In their study, this model achieved poorer results than that of ADT and KLR. However, using the hybrid model of autoencoder-MLP can achieve better results.

One of the limitations of the autoencoder-MLP hybrid model is that the results are different in each run of the model. This is due to the different initial weights assigned to the input variables. To overcome this limitation, the model can be run several times and the model with the highest accuracy is selected as the final model. Another limitation is related to sampling technique used for training, as well as, testing of the model. Every time random sampling is used, different training and testing datasets are generated. Therefore, models made with these datasets can be different. To solve this problem, the random sampling method can be repeated several times and the best model can be selected.

In this study, a hybrid model composed of the MLP and autoencoder models was constructed to prepare the FSM for two study areas in Iran and India. For Golestan Province, nine factors including altitude, aspect, slope, plan curvature, TWI, lithology, distance to drainage, land use, and rainfall. For GRB, 12 factors including altitude, slope, aspect, plan curvature, distance from the river, rainfall, river density, TWI, land use land cover, distance from roads, soil type, and geomorphology were considered as the effective factors in flood occurrence. The hybrid autoencoder-MLP uses the capabilities of the MLP neural networks as one of the most powerful machine learning techniques and autoencoder neural networks. In this hybrid model, autoencoder was used to reduce the number of features and eliminate the ineffective ones from the modelling process. The results showed that the autoencoder-MLP model provided considerably better results compared to the MLP model in both study areas.

CONFLICT OF INTEREST
The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES
Abbas, A., Amjath-Babu, T., Kächele, H., & Müller, K. (2015). Non-structural flood risk mitigation under developing country conditions: An analysis on the determinants of willingness to pay for flood insurance in rural Pakistan. Natural Hazards, 75, 2119–2135.
Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahahi, H., & Panahi, M. (2019). Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto International*, 34, 1252–1272.

Ali, S. A., Parvin, F., Pham, Q. B., Vojtek, M., Vojteková, J., Costache, R., ... Ghorbani, M. A. (2020). GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topfa basin, Slovakia. *Ecological Indicators*, 117, 106620. http://dx.doi.org/10.1016/j.ecolind.2020.106620.

Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C., & Kalantari, Z. (2019). A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in northern Iran. *Science of the Total Environment*, 660, 443–458.

Arora, A., Pandey, M., Siddiqui, M. A., Hong, H., & Mishra, V. N. (2019). Spatial flood susceptibility prediction in middle ganga plain: Comparison of frequency ratio and Shannon’s entropy models. *Geocarto International*, 1–32. https://doi.org/10.1080/10106049.2019.1687594.

Baz, I., Geymen, A., & Er, S. N. (2009). Development and application of GIS-based analysis/synthesis modeling techniques for urban planning of Istanbul metropolitan area. *Advances in Engineering Software*, 40, 128–140.

Bhatt, C., & Rao, G. (2016). Ganga floods of 2010 in Uttar Pradesh, North India: A perspective analysis using satellite remote sensing data. *Geomatics, Natural Hazards and Risk*, 7, 747–763.

Bracken, L., Cox, N., & Shannon, J. (2008). The relationship between rainfall inputs and flood generation in south–East Spain. *Hydrological Processes: An International Journal*, 22, 683–696.

Bui, D. T., Hoang, N.-D., Martínez-Álvarez, F., Ngo, P.-T. T., Hoa, P. V., Pham, T. D., ... Costache, R. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of the Total Environment*, 701, 134413.

Bui, D. T., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., & Nguyen, Q.-P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, 540, 317–330.

Cao, B., Zhao, J., Lv, Z., Gu, Y., Yang, P., & Halgamuge, S. K. (2020). Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction. *IEEE Transactions on Fuzzy Systems*, 28, 939–952.

Chao, L., Zhang, K., Li, Z., Zou, Y., Wang, J., & Yu, Z. (2018). Geographically weighted regression based methods for merging satellite and gauge precipitation. *Journal of Hydrology*, 558, 275–289.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software*, 95, 229–245.

Chen, M., Shi, X., Zhang, Y., wu, D., & guizani, M. (2017). Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*, 1, 1–1. https://doi.org/10.1109/TBDATA.2017.2717439.

Chicco, D., Sadowski, P. & baldi, P. Deep autoencoder neural networks for gene ontology annotation predictions. Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics; 2014, 533–540.

Costache, R., Hong, H., & Pham, Q. B. (2020). Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Science of The Total Environment*, 711, 134514. http://dx.doi.org/10.1016/j.scitotenv.2019.134514.

Costache, R., Popa, M. C., Tien Bui, D., Diaconu, D. C., Ciubotaru, N., Minea, G., & Pham, Q. B. (2020). Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *Journal of Hydrology*, 585, 124808. http://dx.doi.org/10.1016/j.jhydrol.2020.124808.

Costache, R., & Tien, B. D. (2020). Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Science of The Total Environment*, 712, 136492. http://dx.doi.org/10.1016/j.scitotenv.2019.136492.

Costache, R., Pham, Q. B., Avand, M., Thuy Linh, N. T., Vojtek, M., Vojteková, J., ... Dung, T. D. (2020). Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. *Journal of Environmental Management*, 265, 110485. http://dx.doi.org/10.1016/j.jenvman.2020.110485.

Costache, R. (2019). Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Science of The Total Environment*, 659, 1115–1134. http://dx.doi.org/10.1016/j.scitotenv.2018.12.397.

Costache, R., & Tien, B. D. (2019). Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Science of The Total Environment*, 691, 1098–1118. http://dx.doi.org/10.1016/j.scitotenv.2019.07.197.

Costache, R., Hong, H., & Wang, Y. (2019). Identification of torrential valleys using GIS and a novel hybrid integration of artificial intelligence, machine learning and bivariate statistics. *CATENA*, 183, 104179. http://dx.doi.org/10.1016/j.catena.2019.104179.

Dassanayake, D. R., Burzel, A., & Oumeraci, H. (2015). Methods for the evaluation of intangible flood losses and their integration in flood risk analysis. *Coastal Engineering Journal*, 57, 1540007.

Diaz, J. H. (2006). Global climate changes, natural disasters, and travel health risks. *Journal of Travel Medicine*, 13, 361–372.

Elnazer, A. A., Salman, S. A., & Asmoay, A. S. (2017). Flash flood hazard affected Ras Ghariib city, Red Sea, Egypt: A proposed flash flood channel. *Natural Hazards*, 89, 1389–1400.

Fernández, D., & Lutz, M. (2010). Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology*, 111, 90–98.

Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel...
on climate change. Cambridge, England: Cambridge University Press.

Gaillard, J. C. (2007). Resilience of traditional societies in facing natural hazards. *Disaster Prevention and Management: An International Journal, 16*, 522–544.

Gajbhiye, S., Mishra, S., & Pandey, A. (2014). Prioritizing erosion-prone area through morphometric analysis: An RS and GIS perspective. *Applied Water Science, 4*, 51–61.

Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J. & duque, N. Rainfall prediction: A deep learning approach. Presented at: International Conference on Hybrid Artificial Intelligence Systems; Springer; 2016. 151–162.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.-X., & Kazakis, N. (2018a). Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of the Total Environment, 621*, 1124–1141.

Hong, H., Tsangaratos, P., Illia, I., Liu, J., Zhu, A.-X., & Chen, W. (2018b). Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Science of the Total Environment, 625*, 575–588.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer, 29*, 31–44.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., & Lee, S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability, 11*, 5426.

Judi, D., Rakowski, C., Waichler, S., Feng, Y., & Wigmosta, M. (2018). Integrated modeling approach for the development of climate-informed, actionable information. *Water, 10*, 775.

Kadam, P., & Sen, D. (2012). Flood inundation simulation in Ajoy River using MIKE-FLOOD. *ISH Journal of Hydraulic Engineering, 18*, 129–141.

Khosravi, K., Nohani, E., Maroufzadeh, E., & Pourghasemi, H. R. (2016). A GIS-based flood susceptibility assessment and its mapping in Iran: A comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural Hazards, 83*, 947–987.

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., & Ko, H. L. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology, 573*, 311–323.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River basin, Malaysia. *Environmental Earth Sciences, 67*, 251–264.

Kourgialas, N. N., & Karatzas, G. P. (2017). A national scale flood hazard mapping methodology: The case of Greece—protection and adaptation policy approaches. *Science of the Total Environment, 601*, 441–452.

Kousky, C. (2018). Financing flood losses: A discussion of the National Flood Insurance Program. *Risk Management and Insurance Review, 21*, 11–32.

Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk, 8*, 1185–1203.

Lee, S., & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology, 50*, 847–855.

Li, K., Wu, S., Dai, E., & Xu, Z. (2012). Flood loss analysis and quantitative risk assessment in China. *Natural Hazards, 63*, 737–760.

Lv, Z., & Qiao, L. (2020). Deep belief network and linear perceptron based cognitive computing for collaborative robots. *Applied Soft Computing, 92*, 106300.

Nair, V. & hinton, G. E. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10); 2010. 807–814.

Oh, H.-J., Kim, Y.-S., Choi, J.-K., Park, E., & Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology, 399*, 158–172.

oliveira, T. P., barbar, J. S. & soares, A. S. Multilayer perceptron and stacked autoencoder for internet traffic prediction. IFIP International Conference on Network and Parallel Computing. Springer; 2014, 61–71.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q. & Dasgupta, P. 2014. Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change, Ipcc.

Pourghasemi, H., Pradhan, B., Gokceoglu, C., & Moezzi, K. D. (2012). Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz watershed, Iran. *In Terrigenous mass movements*. Berlin, Germany: Springer.

Pradhan, B. (2010). Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology, 9*, 1–18.

Quan, Q., Hao, Z., Xifeng, H., & Jingchun, L. (2020). Research on water temperature prediction based on improved support vector regression. *Neural Computing and Applications, 1*, 1–10.

Rahmati, O., Pourghasemi, H. R., & Zeinivand, H. (2016). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golistan Province, Iran. *Geocarto International, 31*, 42–70.

Schumm, S. A., & Lichty, R. W. (1965). Time, space, and causality in geomorphology. *American Journal of Science, 263*, 110–119.

Shafizadeh-Moghadam, H., Asghari, A., Taleai, M., Helbich, M., & Tayyebi, A. (2017). Sensitivity analysis and accuracy assessment of the land transformation model using cellular automata. *GISScience & Remote Sensing, 54*, 639–656.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapli, K., & Shirzadi, A. (2018). Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *Journal of Environmental Management, 217*, 1–11.

Shaluf, I. M. (2007). Disaster types. *Disaster Prevention and Management: An International Journal, 16*(5), 704–717.

Sharifi Garmdareh, E., Vafakhah, M., & Eslamian, S. S. (2018). Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. *Hydrological Sciences Journal, 63*, 426–440.
Shi, K., Wang, J., Tang, Y., & Zhong, S. (2020). Reliable asynchronous sampled-data filtering of T-S fuzzy uncertain delayed neural networks with stochastic switched topologies. *Fuzzy Sets and Systems*, 381, 1–25.

Shin, H.-C., Orton, M. R., Collins, D. J., Doran, S. J., & Leach, M. O. (2012). Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1930–1943.

Smith, K., & Ward, R. (1998). *Floods: Physical processes and human impacts*. New Jersey: John Wiley and Sons Ltd.

Souissi, D., Zouhri, L., Hammami, S., Msaddek, M. H., Zghibi, A., & Dlala, M. (2019). GIS-based MCDM-AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia. *Geocarto International*, 35, 1–25.

Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., & Chen, X. (2016). A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*, 89, 171–178.

Tang, Z., Yi, S., Wang, C., & Xiao, Y. (2018). Incorporating probabilistic approach into local multi-criteria decision analysis for flood susceptibility assessment. *Stochastic Environmental Research and Risk Assessment*, 32, 701–714.

Talukdar, S., Ghose, B., Shahfahad, Salam, R., Mahato, S., Pham, Q. B., ... Avand, M. (2020). Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stochastic Environmental Research and Risk Assessment*, 34(12), 2277–2300. https://doi.org/10.1007/s00477-020-01862-5.

Tehrany, M. S., Lee, M.-J., Pradhan, B., Jebur, M. N., & Lee, S. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences*, 72, 4001–4015.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, 504, 69–79.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125, 91–101.

Teng, W.-H., Hsu, M.-H., Wu, C.-H., & Chen, A. S. (2006). Impact of flood disasters on Taiwan in the last quarter century. *Natural Hazards*, 37, 191–207.

Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R., & Keesstra, S. (2018). Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and meta-heuristic algorithms. *Science of the Total Environment*, 615, 438–451.

Testa, G., Zuccala, D., Alcrudo, F., Mulet, J., & Soares-Frazão, S. (2007). Flash flood flow experiment in a simplified urban district. *Journal of Hydraulic Research*, 45, 37–44.

Van Alphen, J., Van Beek, E., & Taal, M. (2006). From flood defence to flood management–prerequisites for sustainable flood management. In *Floods, from defence to management*. Florida: CRC Press.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371–3408.

Wang, S., Zhang, K., Van Beek, L. P., Tian, X., & Bogaard, T. A. (2020). Physically-based landslide prediction over a large region: Scaling low-resolution hydrological model results for high-resolution slope stability assessment. *Environmental Modelling & Software*, 124, 104607.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130–1141.

Yang, L., & Chen, H. (2019). Fault diagnosis of gearbox based on RBF-PF and particle swarm optimization wavelet neural network. *Neural Computing and Applications*, 31, 4463–4478.

Yang, S., Deng, B., Wang, J., Li, H., Lu, M., Che, Y., ... Loparo, K. A. (2019). Scalable digital neuromorphic architecture for large-scale biophysically meaningful neural network with multi-compartment neurons. *IEEE Transactions on Neural Networks and Learning Systems*, 31, 148–162.

Yariyan, P., Janizadeh, S., Van Phong, T., Nguyen, H. D., Costache, R., Van Le, H., ... Tiefenbacher, J. P. (2020). Improvement of Best First Decision Trees Using Bagging and DaggEnsembles for Flood Probability Mapping. *Water Resources Management*, 34(9), 3037–3053. http://dx.doi.org/10.1007/s11269-020-02603-7.

Youssef, A. M., Pradhan, B., & Sefry, S. A. (2016). Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences*, 75, 12.

Zeiler, M. D., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q. V., Nguyen, P., Senior, A., Vanhoucke, V. & Dean, J. On rectified linear units for speech processing. Presented at: 2013 IEEE International conference on acoustics, Speech and Signal Processing. 2013. 3517–3521.

zurada, J. M. (1992). *Introduction to artificial neural systems*. St. Paul: West publishing company.

How to cite this article: Ahmadlou M, Al-Fugara A, Al-Shabeeb AR, et al. Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. *J Flood Risk Management*. 2021;14:e12683. https://doi.org/10.1111/jfr3.12683