

## Article

# Assessing Surface Water Flood Risks in Urban Areas Using Machine Learning

Zhufeng Li <sup>1</sup> , Haixing Liu <sup>2,\*</sup>, Chunbo Luo <sup>1</sup> and Guangtao Fu <sup>1</sup>

<sup>1</sup> College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, Devon, UK; zl435@exeter.ac.uk (Z.L.); C.Luo@exeter.ac.uk (C.L.); G.Fu@exeter.ac.uk (G.F.)

<sup>2</sup> School of Hydraulic Engineering, Dalian University of Technology, Dalian 116023, China

\* Correspondence: hliu@dlut.edu.cn

**Abstract:** Urban flooding is a devastating natural hazard for cities around the world. Flood risk mapping is a key tool in flood management. However, it is computationally expensive to produce flood risk maps using hydrodynamic models. To this end, this paper investigates the use of machine learning for the assessment of surface water flood risks in urban areas. The factors that are considered in machine learning models include coordinates, elevation, slope gradient, imperviousness, land use, land cover, soil type, substrate, distance to river, distance to road, and normalized difference vegetation index. The machine learning models are tested using the case study of Exeter, UK. The performance of machine learning algorithms, including naïve Bayes, perceptron, artificial neural networks (ANNs), and convolutional neural networks (CNNs), is compared based on a spectrum of indicators, e.g., accuracy, F-beta score, and receiver operating characteristic curve. The results obtained from the case study show that the flood risk maps can be accurately generated by the machine learning models. The performance of models on the 30-year flood event is better than 100-year and 1000-year flood events. The CNNs and ANNs outperform the other machine learning algorithms tested. This study shows that machine learning can help provide rapid flood mapping, and contribute to urban flood risk assessment and management.



**Citation:** Li, Z.; Liu, H.; Luo, C.; Fu, G. Assessing Surface Water Flood Risks in Urban Areas Using Machine Learning. *Water* **2021**, *13*, 3520. <https://doi.org/10.3390/w13243520>

Academic Editor: Maria Mimikou

Received: 10 November 2021

Accepted: 7 December 2021

Published: 9 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** CNNs; deep learning; flood management; machine learning; surface water flooding

## 1. Introduction

Surface water flooding in urban areas is caused by heavy rainfall that exceeds the capacity of drainage systems, and results in logged water on streets [1]. It poses a severe threat to residents, properties, and economies. Flood risk is likely to increase with the growth of urban population, land use change, and climate change in many cities around the world. In England, surface water flooding affects about 3 million properties, more than those affected by river flooding and coastal flooding (2.7 million) [2]. Thus, it is important to assess the risks of surface water flooding in urban areas to support informed decision-making for flood management and risk mitigation.

There are two types of flood risk assessment approaches, i.e., physically based modelling and data driven modelling [3]. Physically based models are widely applied for flood prediction considering various hydrological processes, such as precipitation, evaporation, and geomorphology factors in catchments [4]. Data-driven models have the capacity to predict floods by identifying relationships in various datasets [5,6], so the in-depth knowledge of physical processes in catchments is not highly required [7]. Machine learning, as a type of data-driven models, can discover the patterns directly from data without pre-defined rules. However, there are two main challenges in machine learning. First, although in-depth knowledge in hydrology is not required, choosing input data (e.g., catchment features) is key to generate a high-quality model [8]. Second, machine learning normally has a generalization problem. It may not perform well with a different data set [8]. Compared to traditional machine learning algorithms, deep learning algorithms have a significantly

improved capacity in detecting low-level features, and learning high-level features after being trained with a large dataset [9].

Machine learning has been used for surface water flood prediction for several decades [8]. Naïve Bayes (NB), perceptron, XGBoost, classification and regression trees [10], random forest [11], support vector machine [12], and artificial neural networks (ANNs) [13] are commonly used machine learning algorithms for classification problems. Among them, ANNs are the most popular benchmark algorithm applied in the literature [8]. In recent years, hybrid models [14] and deep learning [15] (e.g., convolutional neural networks (CNNs) [16,17], long short-term memory) are emerging with a high performance on assessments of flood events [8] as a representation of the state of the art in machine learning. We chose the NB, perceptron, ANNs, and CNNs with an aim to investigate the capacity of CNNs compared to more conventional machine learning algorithms. Further, there is lack of understanding of their performance on high-resolution flood assessments based on the criteria of F-beta score and area under the curve (AUC).

This study conducts an experiment which builds 12 models of the chosen algorithms on flood risks for rainfall events of 30-year, 100-year, and 1000-year return periods using the case study of Exeter, UK. The study demonstrates that CNN models are capable of providing accurate high-resolution flood risk predictions for rain events of specific return periods in urban areas using the static features, which include a range of urban features, such as roads and buildings. This implies that flood risk maps can be produced using machine learning, such as CNNs, for urban areas which are lacking in data for hydrodynamic modelling.

## 2. Methodology

### 2.1. Features Extraction

#### 2.1.1. Urban Hydrological Features

For assessing surface water flood risks in urban areas, 11 catchment features are chosen for constructing the models. These include coordinates, elevation, slope gradient, imperviousness, land use, soil type, substrate, distance to river, normalized difference vegetation index (NDVI), land cover, and distance to road. These features are all static features.

The coordinates indicate the location of each grid and the adjacent grids often show similar characteristic in continuous areas. The elevation and slope data are generated from digital terrain model (DTM) data. Slope is an important feature. It determines the overland-flow rate, because the steeper the slope is, the larger the runoff velocity will be [18]. Imperviousness is an index that measures the extent of imperious areas in a catchment (0~100%) [19]. Imperviousness areas can be roofs and paved areas. The water is easy to accumulate in areas with higher imperviousness. Soil type and substrate determine the infiltration rate. Land cover and land use indicate the way that the land is used. They directly affect the hydrologic characteristics of a catchment regarding the volume and rate of runoff, infiltration, and groundwater recharge [20]. Distance to river and distance to road refer to the distance between one location and the nearest river and road, respectively. The rainfall will accumulate into the river, and increase the water level of the river. The velocity of flood could be fast when it happens on flat roads with fewer obstructions, and the roads are often at a lower location than its sides, which may result in accumulation of water. NDVI measures the existence of green vegetations in a certain area by calculating the reflectance of near-infrared light, and it is an important feature for flood assessment [21]. It can be expressed as [22]:

$$NDVI = \frac{NIR - Vis}{NIR + Vis} \quad (1)$$

where *Vis* represents the visible light, and *NIR* represents the near-infrared light.

### 2.1.2. Feature Selection

For machine learning algorithms, feature selection can simplify the model, improve the performance, and shorten the training time. In this study, variance inflation factor (VIF) is used for feature selection.

VIF represents the multicollinearity of features, which means a feature can be predicted by other features [23]. Generally, the features with  $VIF > 10$  are of high multicollinearity, and should be removed. The formula of VIF for the  $i$ th feature [23]:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (2)$$

where  $R_i^2$  is the determination coefficient of all other features.

## 2.2. Algorithms

### 2.2.1. Naïve Bayes

The naïve Bayes classifier is a probabilistic classification method based on the Bayes' theorem, with the assumption of no dependency between features [24]. It classifies the samples by calculating the probability that a sample belongs to each class, and chooses the largest one. Therefore, naïve Bayes is a stable algorithm with a strong mathematical theory foundation. In addition, it is not sensitive to missing data, and needs less parameters.

### 2.2.2. Perceptron

The perceptron is the simplest ANN with a single layer [25]. The principle is that perceptron sums the input values by their weight, and uses activation function to generate the output. Though perceptron is a binary linear classifier, it can be generalized for multi-classification problems by continuing to classify binary results. The general idea is to train  $k$  binary classifiers.

It is a simple model, but suitable for large-scale learning. It has three advantages: First, it does not require a learning rate. Second, it does not require a regular term as a penalty. Third, it only updates the model when it is wrong [26], which means its training is fast.

### 2.2.3. Multilayer Perceptron

Multilayer perceptron (MLP) is a kind of feedforward ANNs. The theory foundation is derived from modern neuroscience research. It tries to simulate the structure of neural networks in the human brain [27]. An MLP model consists of at least three layers, including one input layer, hidden layer(s), and one output layer. Each node is a neuron which uses activation function to generate the output signal. The model is trained by backpropagation under the supervised learning framework. MLP is distinct from perceptron for its multiple layers and non-linear activation function, which enables MLP to solve non-linear separable problems [28].

### 2.2.4. 1D Convolutional Neural Networks

CNNs are one of the representative algorithms of deep learning, and have been inspired by the process where the visual cortex in the brain receives visual signals [29]. It is a kind of deep-structured feedforward neural network that includes convolution calculation. A CNN model includes the input layer, a convolutional layer, pooling layers, fully connected layers, and the output layer. Compared with traditional machine learning algorithms, it generally has higher accuracy on classification, but it needs more data and computing time for training.

A modified version of CNNs is 1D CNNs. Compared with CNNs, they accept 1D data as input, so they need less computing power [16] and less data [17] to get fitted, and can support city-scale study. 1D CNNs are mainly used to deal with image and signal processing problems [30] in previous studies. Recently, it has been applied to land cover classification and crop yield prediction [17,30]. These studies are similar to flood

risk assessment, because the problems are non-linear with complex features, and they all do spatial analysis by extracting features from satellite images and using environmental features. Therefore, 1D CNNs can be an effective algorithm for flood assessment in urban areas.

### 2.3. Performance Measures

Statistical measures, such as accuracy and F-beta score, are used for algorithm comparison. They are calculated by true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which represent the number of cells that were correctly classified as positive and negative, and incorrectly classified as positive and negative, respectively.

Accuracy is the basic criterion of the performance which indicates the proportion that samples are correctly predicted, calculated as below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

However, accuracy is misleading for imbalanced data sets. Other criteria are essential.

Precision is also called true positive rate, and indicates the proportion of the positive samples predicted correctly to the number of samples that are predicted as positive:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Recall is also called sensitivity, and indicates the proportion that positive samples are correctly predicted compared to the observed positive samples:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

*F1 score* is an overall assessment on precision and recall which are equally weighted:

$$F1 \text{ score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

*F-beta score* is an improvement of *F1 score*. In real-world cases, precision and recall are not equally important, and a model cannot guarantee a precision and recall in the meantime. Therefore, the parameter of  $\beta$  is introduced into *F1 score*. When  $\beta < 1$ , precision is more important than recall; it is opposite when  $\beta > 1$ . When  $\beta = 1$ , precision, and recall are equally important, and *F-beta score* is the same as *F1 score*, *F-beta score* is calculated using:

$$F - \text{beta score} = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall} \quad (7)$$

The receiver operating characteristic (ROC) curve is a standard technique to evaluate the performance of models [31]. It is plotted by the situations with different threshold values of the models. The vertical axis represents the true positive rate, and the horizontal axis represents the false positive rate. The AUC measures the area underneath the ROC which has a range from 0 to 1. The larger the area, the better the model is.

### 2.4. Oversampling and Undersampling

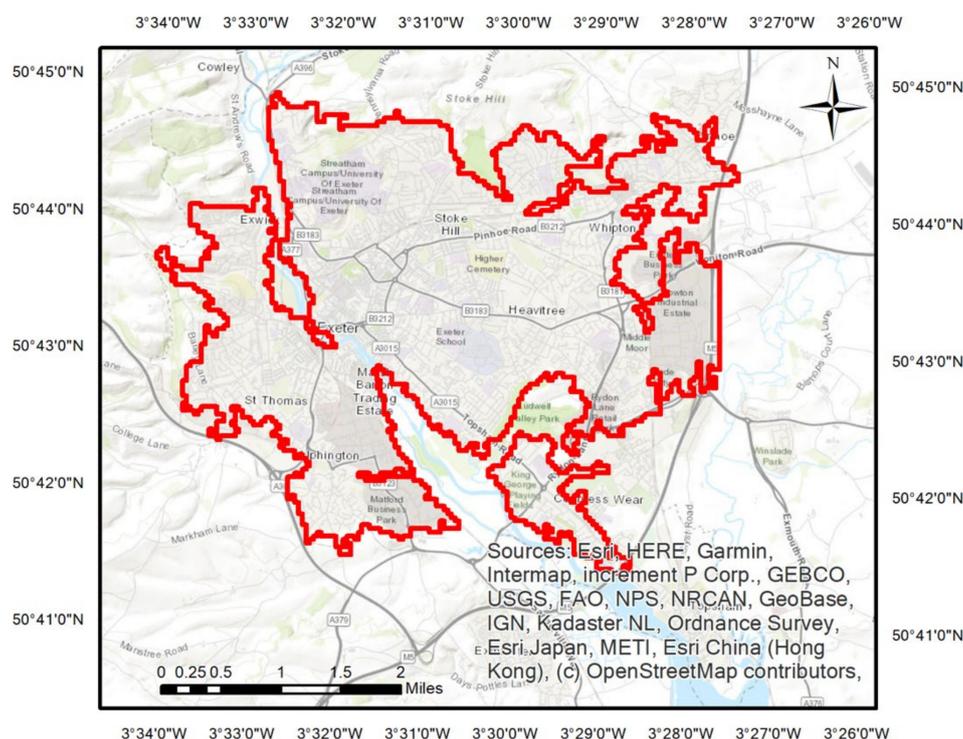
The imbalance of data is the result of the uneven distribution of different classes in real-world problems. The imbalance of data makes models tend to classify a sample to the majority class, resulting in poor performance on the minority class [32]. Previous research shows that the combination of oversampling and undersampling methods can effectively mitigate the imbalance issue [33]. Therefore, this study adopted the combination of synthetic minority oversampling technique (SMOTE) and Tomek link to alleviate the influence of the imbalance of data. Tomek link is defined as an undersampling method based on the improvement of condensed nearest neighbor, and has the ability to eliminate the boundary

in which the samples are considered to have more chance of being misclassified [32]. With this technique, the models gain two benefits: First, it can eliminate noisy and redundant instances. Second, this can balance the number of various classes of samples. However, the drawback is that it increases the variance of independent variables. SMOTE is based on the improvement of KNN, and is used for increasing the number of samples that belong to the minority classes. It only focuses on the minority classes. For each sample in the minority classes, it finds the  $k$  nearest samples, and generates the synthetic samples along the line segments [34]. The benefit is that, compared with the undersampling method, it increases the number of data, which is important for data-driven algorithms. However, the increased data will lead to a low variance of variables. In addition, if the samples of minority are scattered, the synthetic values can make the boundary vague.

### 3. Case Study

#### 3.1. Study Area

The city of Exeter is situated in southwestern England (Figure 1). There are about 129,000 residents [35], and the city covers a total area of 47.04 km<sup>2</sup>. The climate in Exeter is warm and temperate due to the warm Atlantic Gulf Stream. The annual average temperature is 10.7 °C, and the annual rainfall is 825 mm [36]. Exeter lies in the downstream of the River Exe joined by River Creedy, and the geomorphology shows a ridge of land including a steep slope side and a wide, gentle floodplain and estuary side. The geology of Exeter is mainly sandstone and conglomerate [37], and the soils are stony and well-drained sandy silt loamy or clay loams [38].



**Figure 1.** Study area of Exeter in Devon, England.

Flooding is the main natural disaster in Exeter since the 13th century. The recent devastating flood events happened in 1960s and 1970s. Though the flood defense systems were recently upgraded, the city still faces risks of extreme floods [39]. Thus, we chose Exeter as the study area to validate the methods of flood risk mapping.

### 3.2. Flood Inventory Map

In this study, the flood inventory maps used data of 30-year, 100-year, and 1000-year flood events in Exeter provided by the Department for Environment, Food & Rural Affairs [40]. According to the UK Environment Agency, predicting flood risks of 30-, 100-, and 1000-year events plays a significant role in the strategic overview in flood management [41], so these flood events are discussed in this study.

The Environment Agency has produced flood risk maps for pluvial and fluvial flooding, and, in this study, we have used the map from pluvial flooding in order to capture the urban features, such as land uses, road networks, and buildings. The flood risks are represented by a flood hazard indicator used by the Environment Agency considering both water depth and water velocity [41]:

$$HR = d * (v + 0.5) + DF \quad (8)$$

where  $HR$  is the flooding hazard rating;  $d$  is the depth of flooding (m);  $v$  is the velocity of floodwaters (m/s). The  $DF$  (Debris Factor) ranges in 0, 1 or 2 based on the probability that debris will lead to higher hazard level. In this study, the  $HR$ , i.e., flood risks are classified into five levels: no risk (0~0.5); level 1 risk (0.5~0.75); level 2 risk (0.75~1.25); level 3 risk (1.25~2); level 4 risk (2 and above) [41].

### 3.3. Data Preprocessing

The experiment applied the models of NB, perceptron, ANNs, and CNNs for assessing the flood risk of 30-year, 100-year, and 1000-year flood events. To generate high-quality models, we chose the features according to the literature. In particular, we included the VIF of features to check their multicollinearity, the distance to road as an urban feature affecting stormwater flow paths. In addition, the input data were substantially improved through the experiments described in the section of preprocessing and training (e.g., choosing high resolution images, removing no value area, unifying the formats and coordinates, checking multicollinearity, oversampling and undersampling, testing the structures).

The original data are in different sizes, formats, coordination reference systems (CRS), and raster resolutions. Using QGIS, the CRS of all maps were projected into an EPSG:27700 projected coordinate system for the United Kingdom. We converted vector maps into raster maps, and extracted the clip of maps by mask layer based on the boundary of Exeter from an administrative division map of England. Then we resampled and set the resolution as 1 m. Then, data were converted to numerical data from raster images with a grid size of 10 m using the Rasterio tool in numerical experiments.

After preprocessing, the geo-environmental data of Exeter were visualized, as seen in Figure 2. The city is characterized by the hilly topography with some flat land in the southwest (Figure 2a). The terrain is generally undulating, and is steeper in the northwest (Figure 2b). The soil types in Exeter are mainly gravel and clay (Figure 2c). The substrate types are bedrock in the north, west, and east, and superficial in the south and central areas (Figure 2d). Most of the land is urban and suburban, and the improved grass land is scattered in the city (Figure 2e,f). The majority of areas in Exeter are buildings and roads, and there are two railways in the north and southeast (Figure 2g). The imperviousness percentage is larger in the southeast and around the railways in the north (Figure 2h). Two main rivers flow through Exeter in the southwest and central east, and the distance between the river and each pixel is calculated in Figure 2i. Figure 2j shows the distance of a pixel to the nearest road, and the road network is sparse in several areas in the east and southwest.



**Table 1.** Multicollinearity diagnosis test: the VIF of 30-year, 100-year, and 1000-year flood events in Exeter.

	30-Year	100-Year	1000-Year
Latitude	1.02	1.03	1.1
Longitude	1.02	1.03	1.1
Elevation	1.0	1.01	1.02
Slope	1.0	1.01	1.01
Imperviousness	1.01	1.03	1.09
NDVI	1.01	1.02	1.06
Distance to River	1.01	1.02	1.05
Land Cover	1.01	1.03	1.1
Soil	1.02	1.03	1.09
Substrate	1.02	1.04	1.13
Land Use	1.03	1.06	1.16
Distance to Road	1.0	1.01	1.0

The statistics of data help in understanding the general distribution of flood risks. As shown in Table 2, the distribution of flood risks is extremely imbalanced. The non-flood area of the dataset accounts for about 98%, 95%, and 90% for 30-year, 100-year, and 1000-year flood events. The level 4 risk areas only account for 0.01% and 0.06% for 30-year and 100-year flood events, respectively. This can influence models on detecting the pattern of minority classes during the training stage, so oversampling and undersampling methods were introduced to balance the data.

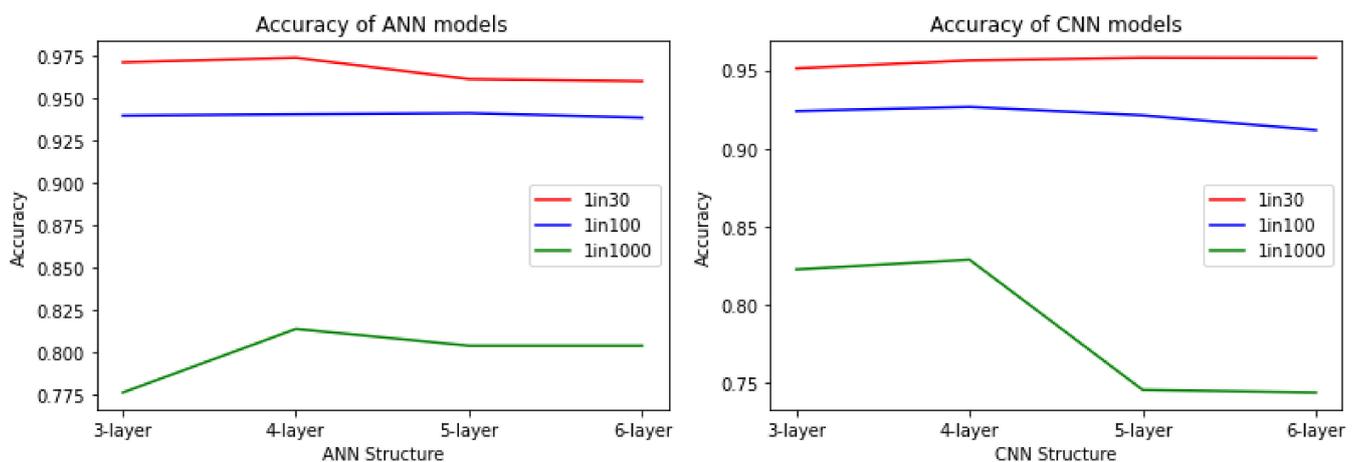
**Table 2.** The distribution of risk levels.

Risk Level	30-Year		100-Year		1000-Year	
	Count	Percent	Count	Percent	Count	Percent
0	279,009	98.10	272,568	95.84	253,219	89.03
1	2915	1.02	6355	2.23	15,034	5.29
2	1424	0.50	2302	0.81	7190	2.53
3	999	0.35	2994	1.05	7868	2.77
4	32	0.01	160	0.06	1068	0.38

### 3.4. Model Training

The process of machine learning includes three stages, i.e., model training, validation, and testing. There were about a total of 260,000 samples after preprocessing, and the dataset was divided into the training and test sets with a ratio of 8 to 2. In the training set, 10% of data were used for validation. In the training stage, each of the four algorithms was trained for assessing 30-year, 100-year, and 1000-year flood events, and, therefore, 12 models were built up. The NB and perceptron models were constructed using the scikit-learn tool [26], whereas ANN and CNN models were built in the Keras and Tensorflow framework [42].

For ANNs and CNNs, we tested several models with different structures, and made a comparison to choose a model that can perform better in enhancing and recognizing the features. For all network structures, the Adam optimizer was used for gradient descent, and the learning rate was 0.0001. The loss function was categorical cross-entropy. The activation function was ReLU in dense layers for ANNs, and convolutional layers for CNNs. The activation function was SoftMax in the output layer for both ANNs and CNNs. For training ANNs, 3-layer, 6-layer, 9-layer, and 12-layer models were tested; for training CNNs, 3-layer, 4-layer, 5-layer, and 6-layer models were tested. Sixty epochs were run for each model for all flood events. Based on accuracy results (Figure 3), the 6-layer ANN model (22,661 parameters) was chosen for 30-year and 1000-year flood events, and the 9-layer ANN model (22,501 parameters) was chosen for the 100-year flood event; the 4-layer CNN model (109,285 parameters) was chosen for all flood events. F-beta score and AUC were introduced to decide the number of epochs to stop training.



**Figure 3.** The accuracy of ANN (left) and CNN (right) models.

The validation stage happens during the training, and the validation set checked the performance of models at the end of each epoch. In the testing stage, the test set was used for testing the models, and the discussion was based on the obtained results from the test set.

#### 4. Results

The accuracy of the assessments of machine learning algorithms on the flood events is shown in Table 3. CNNs has the highest accuracy on all flood events. Naïve Bayes and perceptron show a lower accuracy compared with other algorithms. ANNs shows higher accuracy on 30-year and 100-year flood events than CNNs. However, higher accuracy does not necessarily mean an algorithm is better. The majority of Exeter is no-risk area, and ANNs partially sacrificed precision (i.e., underestimating the high-risk area) to reach higher accuracy, even after applying oversampling techniques.

**Table 3.** Accuracy of models for the assessment of 30-year, 100-year, and 1000-year flood events in Exeter.

	30-Year (%)	100-Year (%)	1000-Year (%)
Naïve Bayes	77.36	77.39	69.61
Perceptron	72.10	69.96	78.27
ANNs	93.61	94.44	81.34
CNNs	92.59	92.92	83.34

Underestimating the flood risks can generally cause more serious consequences than overestimating the flood risks, so we focused more on precision than accuracy, and chose a value of 0.5 for  $\beta$ . Table 4 shows the F-beta score of all flood events. ANNs perform the best on 30-year and 100-year events because the pattern of flood risks is simple to recognize. However, the pattern of flood risks of the 1000-year event is more complex than the 30-year and 100-year events, because the number of samples of high-risk areas are larger (Table 1), and the entropy becomes larger. Since CNNs can detect the low-level features, it can perform well in recognizing such complex patterns. This is the reason that CNNs can outperform ANNs in the 1000-year flood event, and also the reason that the performance of all models on the 30-year flood event is better than the 100-year and 1000-year flood events.

**Table 4.** F-beta score of models for the assessment of 30-year, 100-year, and 1000-year flood events in Exeter ( $\beta = 0.5$ ).

	30-Year (%)	100-Year (%)	1000-Year (%)
Naïve Bayes	92.41	89.91	80.92
Perceptron	90.89	87.66	81.41
ANNs	96.53	93.51	83.98
CNNs	95.71	92.49	84.11

Figure 4 demonstrates the ROC curves and AUC of flood events. CNNs perform the best on 100-year and 1000-year flood events. This situation of the 100-year flood event is slightly different from the rank of F-beta score. The two criteria both focus on recall, but are slightly different in detail. ROC curves and AUC give an importance on specificity (i.e., true negative rate), whereas F-beta score gives an importance on precision. This means when a model is applied in real-world cases, the model with higher AUC is more conservative, and avoids misclassifying; the model with higher F-beta score is more active, and endeavors to find all positive samples. This explains the different ranks on F-beta score and AUC. In addition, with the increment of return period, it becomes more difficult to recognize all flood patterns due to various factors and processes involved in larger urban areas affected (Table 2). However, the performance of CNNs becomes better than ANNs, illustrating the increased capacity of CNNs in detecting complex flood patterns.

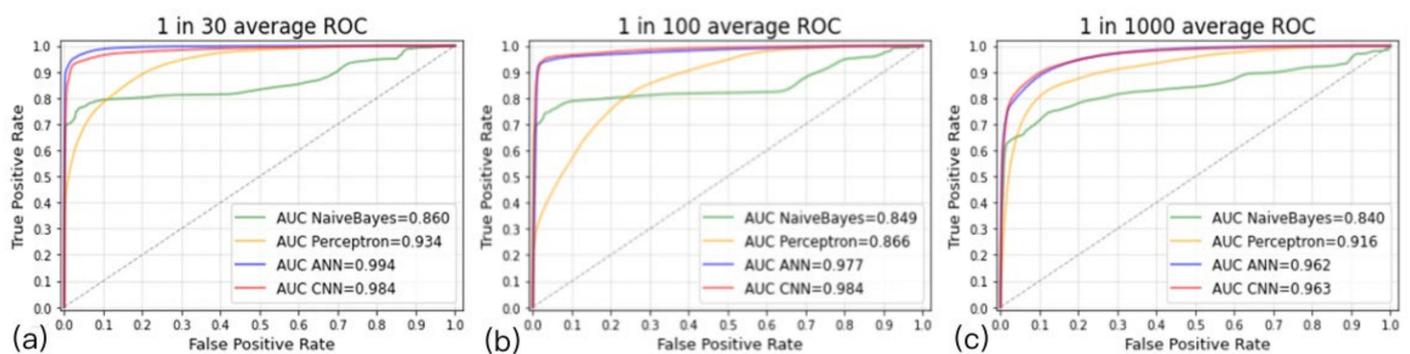
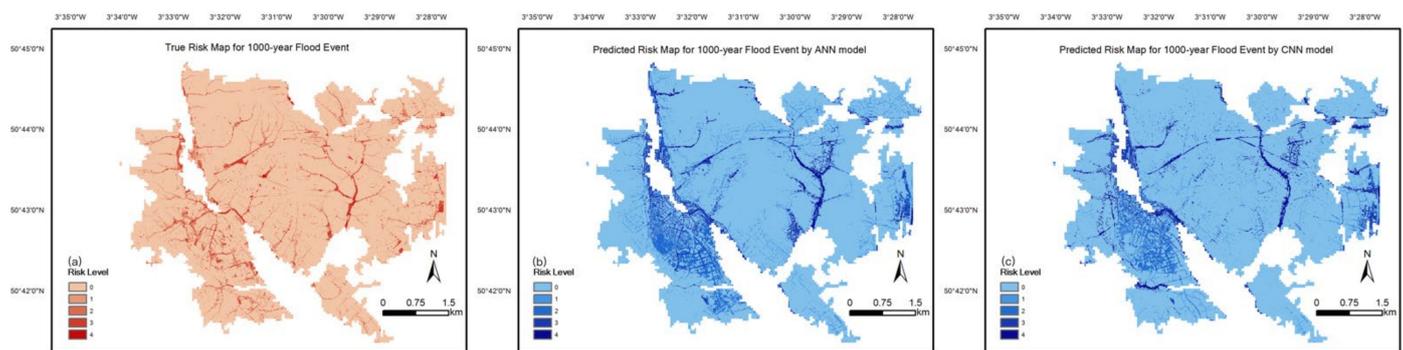
**Figure 4.** The ROC curve and AUC of models for the assessment of (a) 30-year, (b) 100-year, and (c) 1000-year flood events in Exeter (micro average).

Figure 5 visualizes the contrast of true risk areas and predicted results of the 1000-year flood event. The level 4 risk areas are mainly distributed along the two rivers, and both CNN and ANN models provide correct predictions. The Lower Hoopern Valley and the railway in the north also have intensive areas that suffer from level 3 and level 4 risks, and both models made correct predictions. However, the ANN model misclassified more areas as level 2 or level 3 compared with the CNN model. The southwest of Exeter has flat terrain, and most area of the St. Thomas Station and Alphington Road suffer flood risks from level 1 to level 3. The CNN model performed better than ANNs in these areas. ANNs seriously overestimated the flood risks on Cowick Barton Playing Fields located in the southwest of Exeter, whereas CNNs gave a correct prediction. For the remaining area of Exeter, especially the central part, the flood risks are mainly level 0 and 1. CNNs could correctly predict the level 0 area, but underestimated the areas of Wonford and St Loyes, located in the central south of Exeter. ANNs tended to overestimate level 0 areas, but successfully predicted the flood risk of Wonford and St Loyes. ANNs and CNNs could catch the general knowledge that the lower altitude area is likely to have high-level flood risks; the roads and railways are usually affected by flooding; and the areas covered by green vegetation seldom get affected by flooding.



**Figure 5.** 1000-year flood event: (a) ground-truth risk map; (b) predicted risk map by ANNs; (c) predicted risk map by CNNs.

The main limitation of the study is that the prediction on level 1 risk areas is not accurate enough. All models tend to classify level 1 risk areas as a non-risk area in central Exeter, and as a level 2 risk area in the southwest. By comparing the value of input features, two reasons were found. The first one is the resolution is not high enough, and this made the attributes of the grid ambiguous. For example, a 10 m grid can only represent one attribute in raster, but it may contain both a building and road in a vector map. The second reason is that, in the central area, the values of the features of the level 1 risk grid are nearly the same with its neighbors, so the algorithms cannot recognize the pattern. This means additional features that help recognize the pattern may exist. The future work could focus two aspects. To begin with, the study can divide the area into the grid size of 1 m to improve the models, if higher resolution data map can be acquired for all layers. Moreover, more features could be introduced for training. We can manually select additional features with geographical knowledge. As an alternative, 2D CNNs could be used, because 2D CNNs perform well on detecting low-level features, and Landsat pictures contain more low-level features than 1D data.

## 5. Conclusions

This study applied machine learning algorithms to assess the surface water flood risks in urban areas. We constructed NB, perceptron, ANN, and CNN models to assess the 30-year, 100-year, and 1000-year flood events. ANNs and CNNs are more complicated algorithms, and are good at detecting low-level features on non-linear problems (i.e., flood risk assessment). The results show that the performance of ANNs and CNNs are better than NB and perceptron. Besides, as the affected areas become larger, CNNs perform better than the others. In addition, flood risks maps can be generated based on the predicted result. By comparing predicted risk maps and true risk maps, we find the models can learn the general patterns of floods, and the prediction is similar to the real-world case. The main limitation is the lower accuracy of prediction on level 1 flood risk in central and southeast areas. Improvements can be done in two aspects in future. Input data in higher resolution may improve the performance of models, and 2D CNNs could be introduced in the study as an alternative.

The goal of flood risk management is to reduce flood risk, but as the first step, this study focuses on the assessment of flood risks using machine learning algorithms. The flood risks for training did not consider the impact of flood prevention measures. The machine learning models can be easily trained with updated flood risks when flood prevention measures are considered. In future, various flood prevention measures will be considered in hydrodynamic models to access their impact in practice.

This study reveals that machine learning models can use static data to assess the flood risk for a specific storm event without rain-driven data. They are generic models that can assess the flood risks of similar areas without generating additional models for the same type of flood event. Therefore, CNNs can be used for rapid flood mapping. This study

shows that machine learning is a useful tool for flood management, as it can help identify priority areas for risk reduction.

**Author Contributions:** Conceptualization, G.F. and H.L.; Methodology, G.F., C.L. and Z.L.; Software, Z.L.; Validation, Z.L. and H.L.; Formal Analysis, G.F. and Z.L.; Investigation, Z.L. and C.L.; Resources, G.F. and Z.L.; Data Curation, G.F., H.L. and Z.L.; Writing—Original Draft Preparation, Z.L.; Writing—Review and Editing, G.F., H.L. and C.L.; Visualization, Z.L.; Supervision, G.F., H.L. and C.L.; Project Administration, G.F.; Funding Acquisition, G.F., H.L. and C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the British Council grant number 2019-RLWK11-10585, the Royal Society under the Industry Fellowship Scheme grant number IF160108, the Alan Turing Institute under the EPSRC Grant grant number EP/N510129/1, and the National Natural Science Foundation of China grant number 52079016, 52122901. And the APC was funded by the British Council.

**Institutional Review Board Statement:** The authors declare no ethical issues. No human or animal subjects are involved, neither has personal data of human subjects been processed. Also, no security or safety critical activities have been carried out.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** This work was partially supported by the British Council (Ref: 2019-RLWK11-10585), the Royal Society under the Industry Fellowship Scheme (Ref: IF160108), the Alan Turing Institute under the EPSRC Grant (Ref: EP/N510129/1), and the National Natural Science Foundation of China (52079016, 52122901).

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

## References

1. Falconer, R.H.; Cobby, D.; Smyth, P.; Astle, G.; Dent, J.; Golding, B. Pluvial flooding: New approaches in flood warning, mapping and risk management. *J. Flood Risk Manag.* **2009**, *2*, 198–208. [[CrossRef](#)]
2. Bevan, J. Surface water: The biggest flood risk of all. In Proceedings of the CIWEM Surface Water Management Conference, London, UK, 17 October 2018.
3. Sidrane, C.; Fitzpatrick, D.J.; Annex, A.; O'Donoghue, D.; Gal, Y.; Biliński, P. Machine Learning for Generalizable Prediction of Flood Susceptibility. *arXiv* **2019**, arXiv:1910.06521.
4. Grayson, R.B.; Moore, I.D.; McMahon, T.A. Physically based hydrologic modeling: 2. Is the concept realistic? *Water Resour. Res.* **1992**, *28*, 2659–2666. [[CrossRef](#)]
5. Towe, R.; Dean, G.; Edwards, L.; Nundloll, V.; Blair, G.; Lamb, R.; Hankin, B.; Manson, S. Rethinking data—Driven decision support in flood risk management for a big data age. *J. Flood Risk Manag.* **2020**, *13*, e12652. [[CrossRef](#)]
6. Tayfur, G.; Singh, V.P.; Moramarco, T.; Barbetta, S. Flood Hydrograph Prediction Using Machine Learning Methods. *Water* **2018**, *10*, 968. [[CrossRef](#)]
7. Maspo, N.-A.; Bin Harun, A.N.; Goto, M.; Cheros, F.; Haron, N.A.; Nawawi, M.N. M Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *479*, 012038. [[CrossRef](#)]
8. Mosavi, A.; Ozturk, P.; Chau, K.-W. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [[CrossRef](#)]
9. Pham, B.T.; Luu, C.; Vanphong, T.; Trinh, P.T.; Clague, J. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *J. Hydrol.* **2021**, *592*, 125615. [[CrossRef](#)]
10. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [[CrossRef](#)] [[PubMed](#)]
11. Paul, G.C.; Saha, S.; Hembram, T.K. Application of the GIS-Based Probabilistic Models for Mapping the Flood Susceptibility in Bansloi Sub-basin of Ganga-Bhagirathi River and Their Comparison. *Remote Sens. Earth Syst. Sci.* **2019**, *2*, 120–146. [[CrossRef](#)]
12. Khosravi, K.; Melesse, A.M.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hong, H. Chapter 33—Flood susceptibility mapping at Ningdu catchment, China using bivariate and data mining techniques. In *Extreme Hydrology and Climate Variability*; Melesse, A.M., Abtew, W., Senay, G., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 419–434. [[CrossRef](#)]

13. Falah, F.; Rahmati, O.; Rostami, M.; Ahmadisharaf, E.; Daliakopoulos, I.N.; Pourghasemi, H.R. 14—Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Pourghasemi, H.R., Gokceoglu, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 323–336. [[CrossRef](#)]
14. Nourani, V.; Hosseini Baghanam, A.; Adamowski, J.; Kisi, O. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *J. Hydrol.* **2014**, *514*, 358–377. [[CrossRef](#)]
15. Guo, Z.; Leitão, J.P.; Simões, N.E.; Moosavi, V. Data-driven Flood Emulation: Speeding up Urban Flood Predictions by Deep Convolutional Neural Networks. *J. Flood Risk Manag.* **2021**, *14*, e12684. [[CrossRef](#)]
16. Jeong, S.; Ko, J.; Yeom, J.-M. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Sci. Total Environ.* **2022**, *802*, 149726. [[CrossRef](#)]
17. Song, Y.; Zhang, Z.; Baghbaderani, R.K.; Wang, F.; Qu, Y.; Stuttsy, C.; Qi, H. Land Cover Classification for Satellite Images Through 1D CNN. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019; pp. 1–5. [[CrossRef](#)]
18. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* **2015**, *125*, 91–101. [[CrossRef](#)]
19. Copernicus Team at EEA. Imperviousness. 2018. Available online: <https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness> (accessed on 5 August 2021).
20. Jack, A. Low Impact Development (LID) Siting Methodology: A Guide to Siting LID Projects Using a GIS and AHP. Ph.D. Thesis, California State University, East Bay, CA, USA, 2012.
21. Khosravi, K.; Nohani, E.; Maroufinia, E.; Pourghasemi, H.R. 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. *Nat. Hazards* **2016**, *83*, 947–987. [[CrossRef](#)]
22. NASA. Measuring Vegetation (NDVI & EVI). 2000. Available online: [https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring\\_vegetation\\_2.php](https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php) (accessed on 5 August 2021).
23. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [[CrossRef](#)]
24. Pham, B.T.; Tien Bui, D.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [[CrossRef](#)]
25. Freund, Y.; Schapire, R.E. Large Margin Classification Using the Perceptron Algorithm. *Mach. Learn.* **1999**, *37*, 277–296. [[CrossRef](#)]
26. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
27. Zhu, W.; Yang, C.; Huang, B.; Guo, Y.; Xie, L.; Zhang, Y.; Wang, J. Predicting and Optimizing Coupling Effect in Magnetolectric Multi-Phase Composites Based on Machine Learning Algorithm. *Compos. Struct.* **2021**, *271*, 114175. [[CrossRef](#)]
28. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **1989**, *2*, 303–314. [[CrossRef](#)]
29. Fukushima, K.; Miyake, S. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition. In Proceedings of the Competition and Cooperation in Neural Nets, Kyoto, Japan, 15–19 February 1982; pp. 267–285. [[CrossRef](#)]
30. Wolanin, A.; Mateo-García, G.; Camps-Valls, G.; Gómez-Chova, L.; Meroni, M.; Duveiller, G.; Liangzhi, Y.; Guanter, L. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environ. Res. Lett.* **2020**, *15*, 024019. [[CrossRef](#)]
31. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [[CrossRef](#)]
32. Devi, D.; Biswas, S.k.; Purkayastha, B. Redundancy-driven modified Tomek-link based undersampling: A solution to class imbalance. *Pattern Recognit. Lett.* **2017**, *93*, 3–12. [[CrossRef](#)]
33. Batista, G.E.A.P.A.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
34. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
35. Exeter City Council. Exeter Economy in Numbers. 2019. Available online: <https://exeter.gov.uk/business/relocating-to-exeter/exeter-economy-in-numbers/> (accessed on 5 August 2021).
36. Climate Exeter. Available online: <https://en.climate-data.org/europe/united-kingdom/england/exeter-52/> (accessed on 5 August 2021).
37. Department for Environment, Food & Rural Affairs. Southwest EDRP Geographical Area and Physical Context. 2008. Available online: <https://webarchive.nationalarchives.gov.uk/ukgwa/20081112091202/http://www.defra.gov.uk/erdp/docs/swchapter/section11/topography.htm> (accessed on 5 August 2021).
38. Land Information System of Cranfield University. The Soils Guide. Available online: <http://www.landis.org.uk/services/soilsguide/index.cfm> (accessed on 5 August 2021).

39. Environment Agency. Exeter Flood Defence Scheme. 2018. Available online: <https://www.gov.uk/government/publications/exeter-flood-defence-scheme/exeter-flood-defence-scheme> (accessed on 5 August 2021).
40. Environment Agency. Flood Risk Maps 2019. 2019. Available online: <https://www.gov.uk/government/publications/flood-risk-maps-2019> (accessed on 5 August 2021).
41. Environment Agency. What Is the Risk of Flooding from Surface Water Map? 2019. Available online: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/842485/What-is-the-Risk-of-Flooding-from-Surface-Water-Map.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842485/What-is-the-Risk-of-Flooding-from-Surface-Water-Map.pdf) (accessed on 5 August 2021).
42. Chollet, F. Keras. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 5 August 2021).