



Review

Hazard Susceptibility Mapping with Machine and Deep Learning: A Literature Review

Angelly de Jesus Pugliese Vitoria ^{*}, Andrea Folini, Daniela Carrion and Maria Antonia Brovelli

Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milan, Italy; daniela.carrion@polimi.it (D.C.); maria.brovelli@polimi.it (M.A.B.)

* Correspondence: angellyde.pugliese@polimi.it

Abstract: With the increase in climate-change-related hazardous events alongside population concentration in urban centres, it is important to provide resilient cities with tools for understanding and eventually preparing for such events. Machine learning (ML) and deep learning (DL) techniques have increasingly been employed to model susceptibility of hazardous events. This study consists of a systematic review of the ML/DL techniques applied to model the susceptibility of air pollution, urban heat islands, floods, and landslides, with the aim of providing a comprehensive source of reference both for techniques and modelling approaches. A total of 1454 articles published between 2020 and 2023 were systematically selected from the Scopus and Web of Science search engines based on search queries and selection criteria. ML/DL techniques were extracted from the selected articles and categorised using ad hoc classification. Consequently, a general approach for modelling the susceptibility of hazardous events was consolidated, covering the data preprocessing, feature selection, modelling, model interpretation, and susceptibility map validation, along with examples of related global/continental data. The most frequently employed techniques across various hazards include random forest, artificial neural networks, and support vector machines. This review also provides, per hazard, the definition, data requirements, and insights into the ML/DL techniques used, including examples of both state-of-the-art and novel modelling approaches.

Keywords: susceptibility modelling; hazard events; machine learning; deep learning; literature review; air pollution; urban heat island; flood; landslide



Citation: Pugliese Vitoria, A.d.J.; Folini, A.; Carrion, D.; Brovelli, M.A. Hazard Susceptibility Mapping with Machine and Deep Learning: A Literature Review. *Remote Sens.* **2024**, *16*, 3374. <https://doi.org/10.3390/rs16183374>

Academic Editor: Jiaojiao Li

Received: 30 July 2024

Revised: 28 August 2024

Accepted: 9 September 2024

Published: 11 September 2024



Copyright: © 2024 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/>).

1. Introduction

More than 3.3 billion people live in areas highly vulnerable to climate change, which negatively impacts human health, infrastructure, overall well-being, and causes economic losses [1]. In the decade from 2001 to 2010, the global area affected by heatwaves has increased three times compared to the first decade of the 20th century [2]. Furthermore, natural disasters have caused higher losses in 2021 than in the two previous years, particularly with record losses due to extreme flash flood events in Europe [3]. Resilience is the ability to face and recover from the effects of hazardous events in an efficient manner by guaranteeing the preservation, restoration, or improvement of essential services [4]. Among the Sustainable Development Goals (SDGs) proposed by the United Nations, SDG11 calls for inclusive, safe, resilient, and sustainable cities and human settlements. A key part of resilience relies on the tools to understand and prepare for the risks posed by hazardous events.

A hazard is a condition with the potential for damage, e.g., the potential of flooding due to rainfall. Vulnerability refers to the conditions which make a location or asset likely to be affected by a hazard, such as physical, social, economic, and environmental factors [5]. In this context, risk can be defined as a hazardous event impacting an exposed asset, causing consequences which depend on the vulnerability of the asset to that specific hazard [6]. On the other hand, susceptibility is defined as the probability of a hazardous event happening at a specific location. Given the historical occurrences of hazardous events, susceptibility

to such events can be modelled by analysing the existing external factors which could condition such occurrences, namely conditioning factors, e.g., meteorological or land cover factors. These models can be later applied to locations without historical events to compute their susceptibility towards a specific hazardous event.

The use of machine learning (ML) and deep learning (DL) techniques for the tasks of susceptibility modelling and the forecasting of hazardous events has increased in recent years due to the availability of environmental data and computational tools and resources. ML and DL are possible approaches for these tasks among other approaches such as traditional statistical methods, expert knowledge-based systems, and physically based models. The annual scientific production of machine learning applied to climate change risk assessment has been growing exponentially from 2000 to 2020 [7]. Geospatial big data grow by at least 20% every year, including data in different formats and temporal and spatial resolutions [8]. This growth accounts for the substantial contribution of the community to the generation of data and the launch of new satellites for the monitoring of different phenomena. An additional fact is the substantial increase in open source packages (e.g., R, SciKit-Learn, and TensorFlow) and cloud computing platforms (e.g., Google Earth Engine, Google Colab) [9] which allow for the fast implementation and computation of ML/DL models. Therefore, this review focuses on the methodologies, data, and algorithms used to model the susceptibility to hazardous events by means of ML/DL techniques. This focus is due to the ability of these methods to capture temporal and spatial nonlinear relationships and the increasing trend of publications on the matter.

In this paper, we propose a literature review to revise the state-of-the-art and novel approaches related to the analyses and methodologies used for hazard susceptibility modelling and mapping by means of ML/DL techniques, particularly those focused on the hazards of air pollution, urban heat islands, floods, and landslides. The objective is to provide a reference point for the selection of modelling techniques and data in the production of susceptibility maps for each hazard. The work was carried out within the framework of the project HARMONIA (EU-Horizon 2020), which aims to provide stakeholders and urban planners with a decision support system to improve urban resilience and mitigate the effects of climate change.

The sections of this study are described as follows. Section 2, entitled “Background”, defines the terminology used in the present manuscript and clarifies different concepts in the areas of risk and ML/DL models. The methodology of the literature review procedure is described in Section 3. Subsequently, the two following sections, Sections 4 and 5, are devoted to the classification of the algorithms considered in this review and the description of the feature selection methods. Afterwards, Section 6 regards the susceptibility modelling of hazardous events and describes the general workflow and data insights shared among the hazards considered in the present review. Subsequently, Sections 7–10 provide specific information related to the definition, methodology, data acquisition, and insights pertaining to the ML/DL algorithms for each of the four studied hazards. The last sections, namely Sections 11 and 12, provide the conclusion and discussion. Furthermore, there is an appendix containing the details of the ad hoc classification of ML/DL algorithms implemented in the reviewed articles.

2. Background

This section is divided into two subsections. The first is concerned with the definitions of hazard, susceptibility, risk, and other associated concepts. The second is concerned with the concepts of machine learning, deep learning, the difference between them, and the other methods considered in this review. The objective is to clarify the definitions and to establish the wording that is used throughout the manuscript.

2.1. Hazard, Susceptibility and Risk

Although the concepts of hazard, susceptibility, and risk are closely related, they differ in terms of how they model the potential of hazardous events and their associated map representations.

- Hazard is a condition which can potentially cause a consequence [10]. The materialisation of a hazard is the occurrence of a hazardous event at a specific time and location [6].
- Susceptibility models the tendency of the occurrence of hazardous events at a specific location based on its physical and environmental characteristics [11]. The target variable for susceptibility modelling is the occurrence (or non-occurrence) of a hazardous event at a specific location.
- Risk measures the probability and severity of the negative impact on an asset [10]. The risk is the product of the hazard, the exposed assets, and the vulnerability of the assets towards that hazard [6]. In this context, exposure refers to the presence of infrastructure or a population at the event location, whilst vulnerability refers to how an asset can be affected by the hazard considering the physical, social, economic, and environmental factors [5].

The differences between these concepts are also reflected by their spatial representations. The definitions of the hazard, susceptibility, and risk maps are provided as follows:

- Hazard maps express the magnitude of the events and their frequency [6]. For example, flood hazard maps provide the water depth, extent, and return period of floods [11].
- Susceptibility maps depict the areas which are prone to the occurrence of hazardous events. The susceptibility is provided quantitatively in terms of probability or qualitatively in terms of low, medium, or highly susceptible areas. The time frame of susceptibility maps tends to not be considered [10].
- Risk maps show the product of the combination of the hazard, exposure, and vulnerability maps. The parameters of the combination are customised to the case study.

This study focuses on susceptibility modelling and the production of susceptibility maps for the hazards of air pollution, urban heat island, flood, and landslide. The term “hazard susceptibility mapping” refers to the susceptibility mapping of the four aforementioned hazards.

2.2. Machine Learning, Deep Learning, and Other Methods

There are different approaches for hazard susceptibility modelling in the literature. Some of the possibilities, among other methods, are expert-knowledge-based models, physically based models, statistical methods, machine learning (ML), and more recently, deep learning (DL) models [12]. These methods are defined in the context of susceptibility mapping:

- Expert-knowledge-based models rely on questionnaires and the expert’s opinion [12].
- Physically based models are deterministic models which simulate the interactions between different environmental variables. They provide a theoretical framework integrating areas such as hydrology and geomorphology [13]. For example, for landslide susceptibility mapping, a physical model may integrate the soil characteristics, slope gradient, and water density [14].
- Statistical methods are data-driven methods which analyse the relationships between the previous occurrences of hazardous events and environmental variables to understand the correlations among them. For example, statistical methods include frequency ratio, weight of evidence, and analytical hierarchy process.
- Machine learning methods are also data-driven methods which rely on the same relationship as statistical methods. However, ML models learn the characteristics of the environmental variables associated with previous event occurrences to determine

the probability of occurrence at unseen locations. This can be described as analytical model building [15]. ML encompasses several subcategories, even deep learning. In this context, ML includes regressions, decision trees, support vector machines, ensembles, Bayesian-based classifiers and regressors, instance-based classifiers, and artificial neural networks.

- Deep learning methods are part of ML. Deep learning methods are a branch of artificial neural networks, specifically those with multiple hidden layers and specific processing approaches [15]. These tend to perform better with respect to 'traditional' ML [16], due to their capacity to build even more complex relationships. DL includes convolutional neural networks, recurrent neural networks, transformer neural networks, autoencoders, and others.

Each of these approaches has its own limitations. Expert-based modelling tends to be subjective, which could lead to biased results [12]. Physically based models provide acceptable accuracy; however, their usage is limited to small areas of interest due to the need for computational resources and precise geotechnical data [17,18]. Statistical methods rely on certain assumptions about the data, such as linearity, which may limit their effectiveness if the data are not well structured or do not fall into specific patterns. ML and DL models are often 'black-box' models which do not provide insights into the parameters which are more important within the model structure. This limits the interpretability of these models, in particular, DL models [15]. Furthermore, DL requires high-performance computational resources with respect to traditional ML [15].

The advantage of ML and DL methods is their ability to construct complex relationships from the training data, which results in outstanding accuracies and can be used in large areas of interest. Less complex ML methods, such as linear regressions and decision trees, are able to ascertain the importance of the environmental variables inside the model. This is also the case for statistical models. In fact, statistical models tend to be used as an initial step to understand the correlation between the dependent and independent variables.

3. Methodology

The procedure followed to conduct the literature review of hazard susceptibility mapping using machine learning (ML) and deep learning (DL) techniques consisted of a search of two scientific article indexing databases, the retrieval of articles, screening based on inclusion/exclusion criteria, sorting the articles based on the number of citations, and lastly, the extraction of relevant information from each article. The workflow is described in detail in Figure 1.

The search was conducted in two search engines, Scopus (<https://www.scopus.com/> accessed on 23 January 2024) and Web of Science (<http://webofscience.com/> accessed on 23 January 2024), considering articles from 2020 to 2023. Studies performing similar literature reviews considered a longer time span, e.g., 1980–2021 [19], or 1992–2021 [9]; however, these only retrieved statistics on the number of articles and analyses of affiliations, keywords, and methods. A deeper analysis was performed in this review, extracting the ML/DL techniques used to model the susceptibility of the different hazards and other attributes used for that purpose, such as the conditioning factors and the methods for feature selection, outlier removal, and gap filling. Therefore, the selected time span is shorter to include the latest technologies and maintain a reasonable amount of articles to be used for an in-depth analysis.

Four queries were prepared, one per each hazard, to obtain the highest number of relevant articles in both search engines. By relevant, we mean articles which are related to the susceptibility mapping of hazardous events using ML/DL techniques. Before selecting the final keywords to be used, we manually checked the different queries for each hazard and we only kept the ones that provided significant and distinct results with respect to other queries, e.g., for air pollution, the query "'air pollution' susceptibility (machine or deep) learning' provided only 9 results, among which only 1 was relevant; then, the query "'air pollution' mapping (machine OR deep) learning provided 52 results, among which

26 were relevant. This iterative procedure was used to compose the final queries for each hazard. The final queries were used to search articles' titles, abstracts, and author keywords, selecting journals as the document type, and the period of 2020–2023 as the time span:

1. "air pollution" AND (mapping OR prediction OR modelling) AND (deep OR machine) learning.
2. "heat island" AND (prediction OR modelling OR intensity) AND (deep OR machine) learning.
3. Flood AND (mapping OR susceptibility) AND (machine OR deep) learning.
4. Landslide AND (mapping OR susceptibility) AND (machine OR deep) learning.

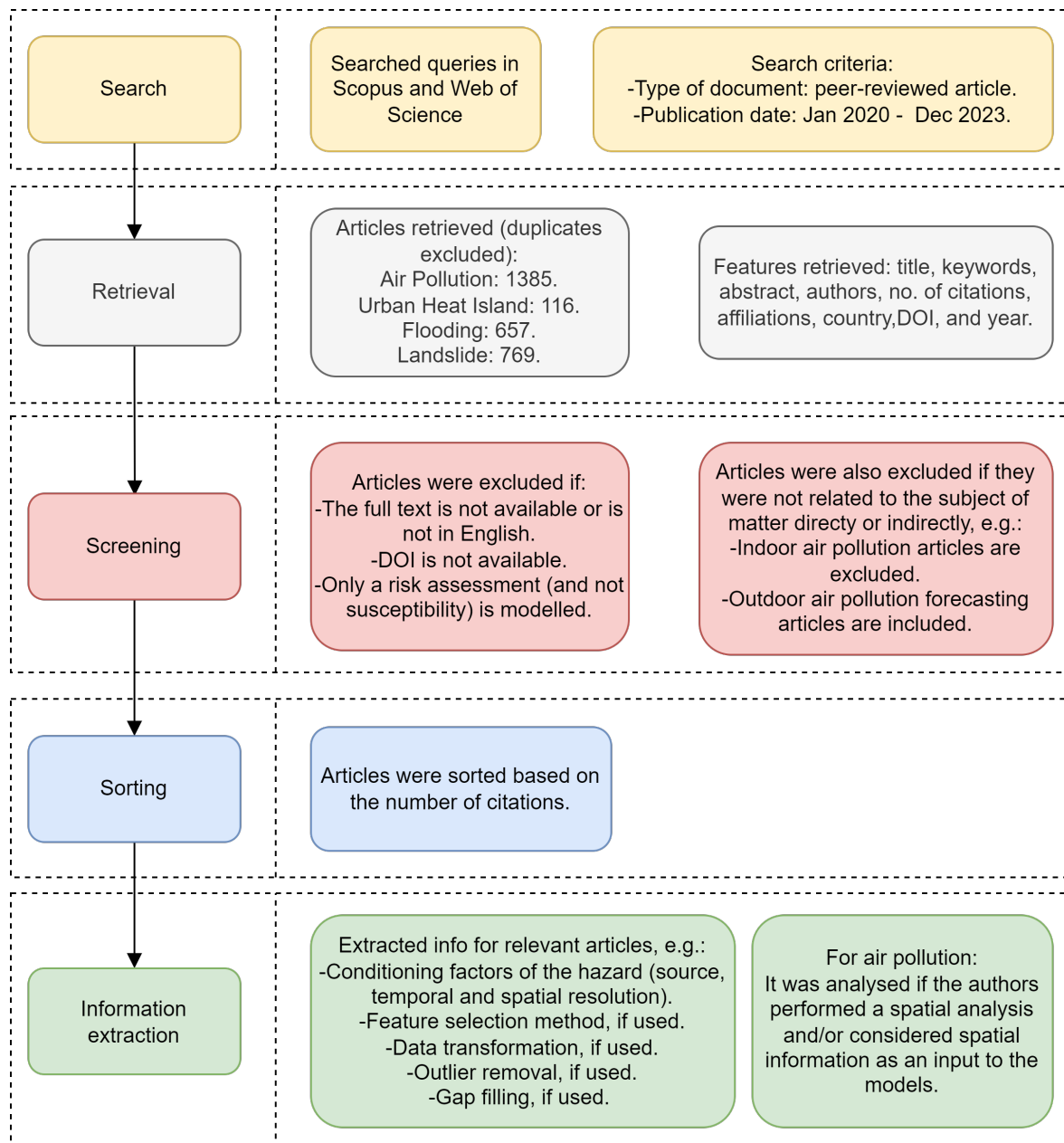


Figure 1. Literature review methodology.

The main features of the articles were retrieved, i.e., title, keywords, abstract, authors, number of citations, affiliations, affiliations' countries, DOI, and year. Duplicates were removed using the articles' DOIs. Afterwards, a manual screening of the articles

was performed to select the relevant ones: this consisted of checking each article and including/excluding it based on certain criteria. The selection criteria were the following:

1. Articles were selected if all the following criteria were met:
 - (a) The publication date was between 1 January 2020 and 31 December 2023.
 - (b) The publication was a peer-reviewed article (i.e., not conference proceedings or other types of text).
 - (c) The publication reported on the use of ML/DL models, algorithms, or techniques in the production of hazard susceptibility mapping, forecasting, or modelling.
 - (d) The publication was related to risk assessment, but the susceptibility of hazards was modelled individually, i.e., the risk assessment could be split into hazard susceptibility modelling and then the risk was assessed based on the modelled hazard susceptibility.
2. Articles were excluded if any of the following criteria were met:
 - (a) The publication was a literature review.
 - (b) The full text was not available.
 - (c) The publication was written in languages other than English and without an English translation.
 - (d) The publication did not have a DOI.

The total number of retrieved articles vs. the number of manually selected articles per each hazard is reported in Table 1 and a yearly breakdown is plotted in Figure 2. Subsequently to the screening, the articles were sorted based on the number of citations per year. This is particularly important when the number of articles to be individually reviewed is large (above 100); in that case, the references considered in the hazard sections were the ones with the highest number of citations or the ones with novel approaches.

Furthermore, some attributes were extracted for each reviewed article, i.e., the ML/DL techniques used, the conditioning factors of the hazard (alongside their source and their spatial and temporal resolution), as well as whether a feature selection, outlier removal, and gap filling methods are used and which ones.

Table 1. Number of retrieved articles vs. number of selected articles.

Hazard	No. of Retrieved Articles	No. of Manually Selected Articles
Air pollution	1385	654
Urban heat island	116	36
Flood	657	253
Landslide	769	511

It is important to mention that the selection criteria, the search queries, and the methodology were defined with the purpose of retrieving the largest number of relevant articles in the defined time span. However, these come with some limitations and biases which are described as follows:

1. Potentially relevant studies published before the year 2020 were excluded; therefore, the information presented in this manuscript is limited to the latest studies.
2. The exclusion of articles which are not in English introduces a language bias by ignoring relevant studies written in other languages.
3. The exclusion of grey literature, such as conference proceedings, tends to result in cutting-edge approaches being overlooked.
4. The articles are sorted based on the number of citations per year; therefore, the sorting is biased towards older articles.

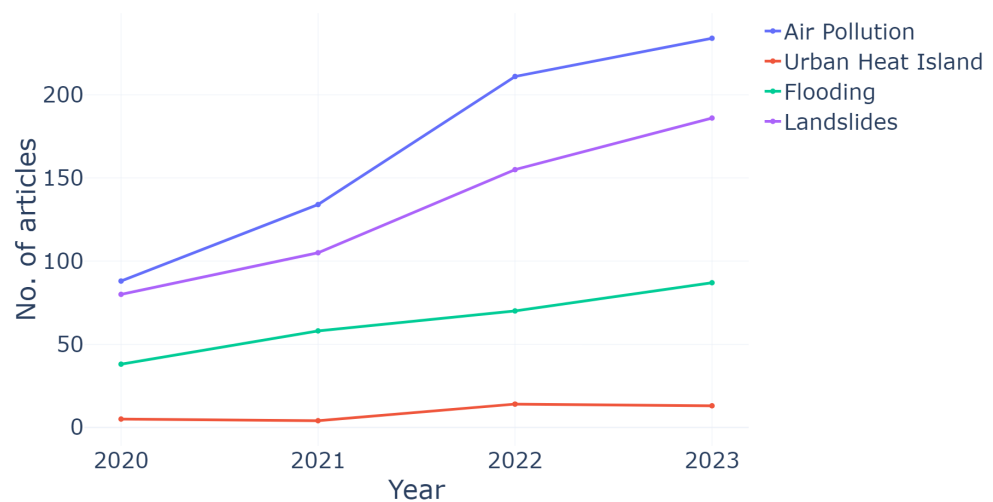


Figure 2. Number of selected articles per year.

Despite the aforementioned biases and limitations, the number of articles selected as relevant is large. Nevertheless, the amount of selected articles is still manageable for extracting information on the data and modelling techniques and methodologies. Furthermore, articles with novel approaches are mentioned regardless of their number of citations.

Finally, the results of the searches and the selected articles have been published in Zenodo [20]. This repository also contains the Jupyter notebooks to reproduce the plots of the individual hazard sections.

4. Algorithms Classification

As the reviewed articles contain a very large number of ML and DL algorithms, in the following, a classification overview of the methods is provided. The most frequent classification of machine and deep learning algorithms is supervised, unsupervised, semi-supervised, and reinforcement learning [21,22]. These four classes differ based on the data used in the training phase and on the learning method.

- Supervised learning algorithms start from a known training dataset with labelled data and produce an inferred function used to make predictions.
- Unsupervised learning algorithms take as input data those which are not labelled or classified a priori. The goal of these models is to find hidden structures or patterns in the training data without explicit feedback or guidance.
- Semi-supervised learning algorithms are a combination of the two previous categories since they use both labelled and unlabelled data for training. The objective is to provide a better prediction with respect to using supervised learning with scarce data [16].
- Reinforcement learning algorithms learn to interact with their surrounding environment to achieve a specific goal. The model takes actions and receives rewards based on the outcome of the actions. Its goal is to learn the set of actions that maximises the cumulative reward. Examples of the use cases for reinforcement learning are autonomous driving and robotics, which are in an environment-driven context [16].

Despite being very common, this classification is not particularly useful for the scope of this review, as almost all the techniques used for susceptibility mapping in the reviewed articles are supervised. Furthermore, as there is no universal taxonomy that groups the algorithms based on the similarity of their functions, we propose a custom classification for the purpose of the review which includes all the algorithms introduced in the articles. For this reason, the classes are not mutually exclusive, and some algorithms could be considered part of multiple classes.

The categorisation considered in this paper is similar to those presented in [23,24], with some slight modifications and additions to better account for the algorithms of the reviewed articles. Specifically, the classes discussed in the following subsections are the ensemble (ENS), neural networks (NNs), decision trees (DTs), support vector machines (SVMs), regression (REG), Bayesian (BAY), instance-based (IB), dimensionality reduction (DR), statistical (STAT), clustering (CLUST), rule-based systems (RUL), and time series (TS). Compared to the previously cited classifications, we added the classes statistical and time series, added the regularisation methods to the regression class, and merged all types of neural networks into a single class. Finally, we separated support vector machines from the instance-based class to further highlight their importance in the susceptibility mapping task. The classification of each particular algorithm in the defined classes can be found in Appendix A.

4.1. Neural Networks

Neural networks are systems composed of multiple units, called artificial neurons, which are connected in a layered structure that resembles the human brain. The neurons are usually aggregated into three different kinds of layers: the input layer, one or more hidden layers, and the output layers. The input layer takes the original data and passes these to the rest of the network, while the output layer stores the result of the network. The hidden layers, instead, perform multiple data transformations and processing and are responsible for the performance and complexity of the network. Based on how the neurons and the layers are connected, it is possible to distinguish a large number of network architectures. In recent years, the use of neural networks in various fields has constantly increased due to their ability to learn complex and nonlinear problems.

The simplest architectures are artificial neural networks (ANNs), which are usually composed of a single hidden layer and the inputs are only processed in the forward direction from the input to the output layer [25], as depicted in Figure 3.

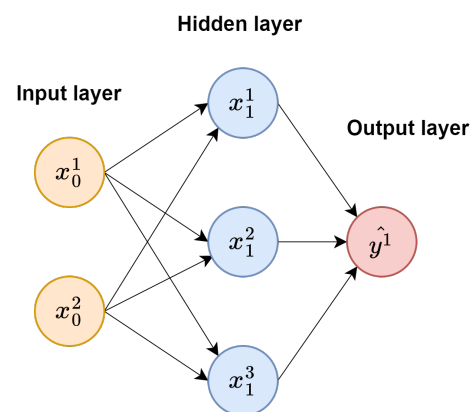


Figure 3. The structure of a fully connected ANN.

Convolutional neural networks (CNNs) have a more complex structure than ANNs and are widely used in image processing, and are thus particularly suited for various geospatial applications. This is due to their ability to automatically extract features and learn spatial patterns with the presence of convolutional and pooling layers in the network [26], which are described below and depicted in Figure 4.

Convolution layers are based on the multiplication of a sliding window of weights (filter or kernel) with the input values. The network automatically determines the weights during the learning process. This is the main advantage of CNNs in comparison to conventional ML methods where user-defined filter weights are needed. Pooling layers, instead, aggregate neighbouring pixels into a single pixel with a maximum or average function, reducing the image's overall dimensions. As the numbers of convolutional and pooling layers increases, the network becomes deeper and can identify more complex structures in

the data [26]. After the input data are processed through these layers, the classification or regression is applied to obtain the final output.

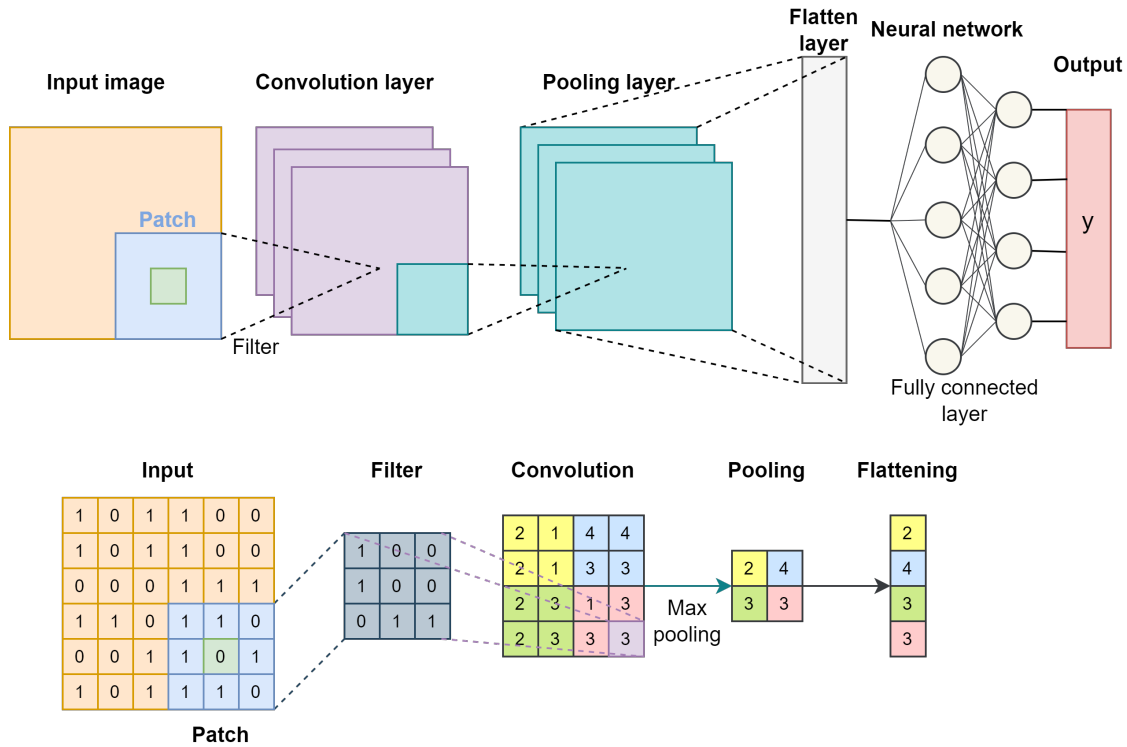


Figure 4. The structure of a CNN and an example of convolution.

Recurrent neural networks (RNNs) are networks that contain recursion connections among the neurons of the hidden layers (Figure 5). The loops and the memory in each neuron allow the networks to capture sequential information in the input data, making them useful when dealing with time series problems such as air quality assessment [27].

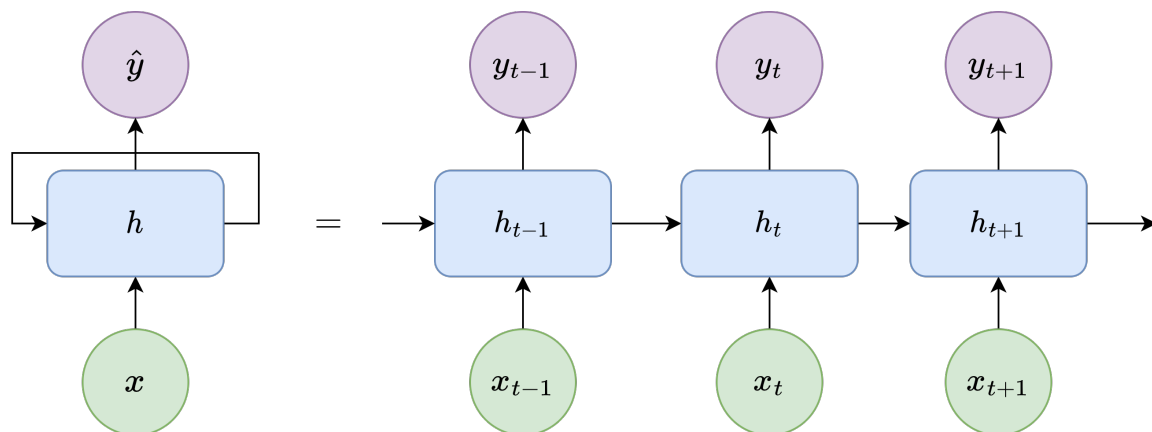


Figure 5. The structure of an RNN.

In the reviewed articles, among the most used types of architectures of RNN are long short-term memory (LSTM) and its variations. A memory cell in an LSTM unit can store data for long periods and the flow of information into and out of the cell is managed by three gates. The ‘forget gate’ determines whether the information from the previous state cell will be memorised or removed, the ‘input gate’ determines which information should enter the cell state, and the ‘output gate’ determines and controls the outputs [15].

CNNs and RNNs, especially LSTM, can be combined in geospatial problems with both spatial and temporal dependencies in the data (Figure 6). This is often the case when dealing with air quality prediction, as the conditioning factors include the time series of the historical pollutant concentrations. A common example in the field [28,29] is to first use the convolution and pooling layers of the CNN to extract the spatial features of the original input data. The obtained features are then flattened into a one-dimensional array and input into the LSTM as time series to analyse the time features of the data. Finally, the result is obtained through the fully connected output layers. In these studies, the hybrid CNN–LSTM architectures provided better accuracy compared to the single CNN and LSTM counterparts.

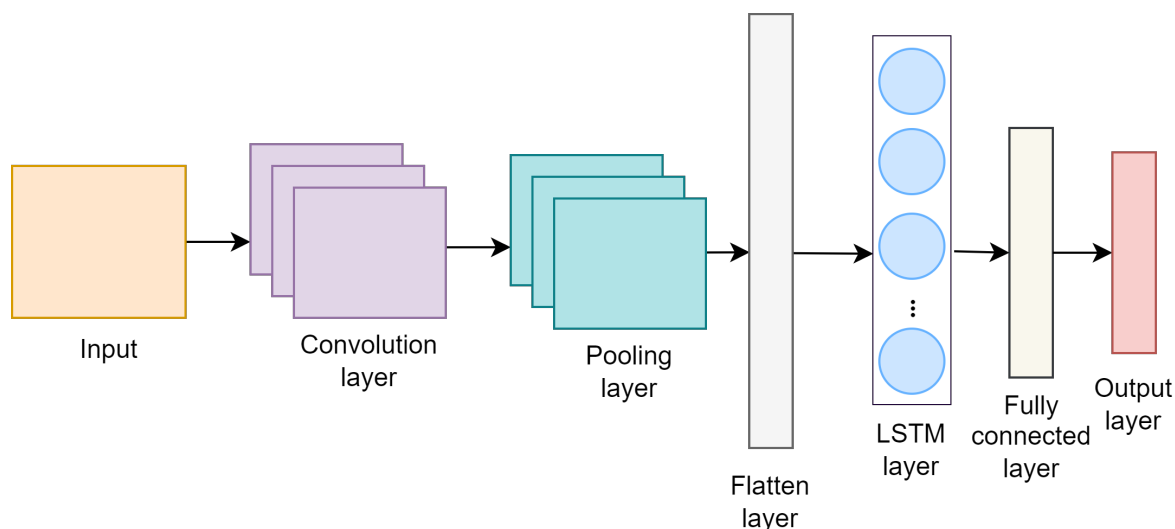


Figure 6. The structure of a CNN–LSTM hybrid.

Furthermore, transformer neural networks are an emerging architecture in the context of hazard susceptibility mapping. Originally proposed for natural language processing [30], this encoder–decoder architecture contains a multi-head self-attention mechanism which allows the encoding of short- and long-range correlations between the words of a sentence. A recent study proposed the use of a spatiotemporal Transformer architecture adding spatial, temporal, and value/variable embeddings and a sparse attention mechanism to the original architecture in the task of predicting PM_{2.5} concentrations in wildfire-prone areas [31].

4.2. Regression

Regression methods are statistical and machine learning algorithms that estimate with a function the relationship between a dependent variable (or multiple variables) and one or more independent variables. The most common regression method is linear regression, which finds the line that better fits the data. To extend the regression to nonlinear problems as well, it is possible to project the data to other spaces with the use of a kernel. Moreover, it is possible to use nonlinear functions instead of linear ones to better fit the data. To reduce the number of dimensions in the data, there are various regularisation techniques that can be applied to regression algorithms, such as lasso, ridge, and elastic net [32]. These methods constrain some of the feature coefficients towards zero, effectively reducing the complexity of the model and the risk of overfitting.

Despite the name, some regression methods can be used in classification problems, and the main example is logistic regression. The logistic regression model is very similar to linear regression, but it is applied to predict the probability of the binary occurrence of an event [33]. This method can be useful in flood and landslide susceptibility mapping [34] where the formulation of the problem is a binary classification, as it directly provides a probability of the hazard occurrence as an output. The probability can then be categorised

with methods like natural breaks or quantiles to produce the final maps. Moreover, in logistic regression, the input variables can be provided as both continuous (e.g., digital terrain model, slope) and discrete (e.g., soil type, land cover), which is often the case for hydrogeological data.

4.3. Decision Trees

Decision trees are simple non-parametric machine learning algorithms that can be used for both classification and regression. They have a tree structure composed of a root node, internal nodes, and leaf nodes. Each internal node is used to make a decision on one of the attributes of the instance, while the leaves represent all possible outcomes in the dataset [35]. One of the main advantages of decision trees is how easy it is to interpret their output since there is the possibility of visualising all of the decisions that lead to it. Based on how the trees are created and modified during the learning phase, a lot of different variations exist in the literature, such as classification and regression trees [36], C5.0 [37], and reduced error pruning trees [38].

4.4. Support Vector Machines

Support vector machines (SVMs) are supervised algorithms that are used for both classification and regression problems, the latter of which is known as a support vector regressor (SVR). The main idea of an SVM is to find the decision boundaries (hyperplanes) that separate the input data points into the predefined classes with the highest margin and the minimum misclassification (Figure 7).

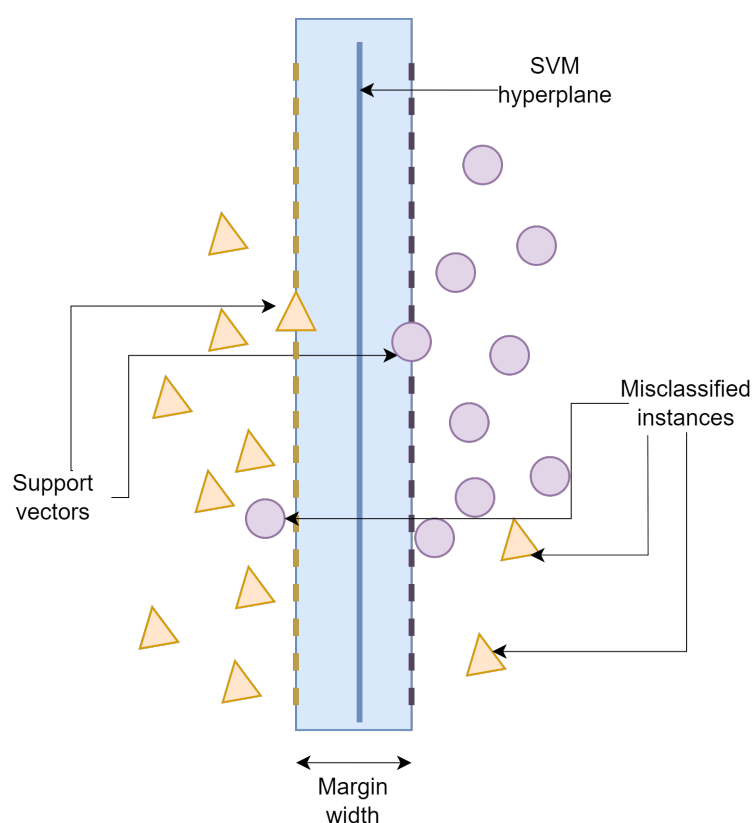


Figure 7. Example of a support vector machine.

The simplest formulation of an SVM is the linear one, where the solution hyperplanes are found in the same space as the input data [39]. In the case where a linear solution does not exist, the input data are mapped onto a higher-dimensional feature space with the use of a kernel function [40]. After this process, the algorithm searches for a solution in the new

space. Multiple variations of SVMs are formulated based on how they deal with the search for the optimal solution and how they treat misclassified points.

The choice of the correct kernel function is essential to obtain high training and prediction accuracy. There are four main groups of kernels that are commonly used with SVMs: linear, polynomial, sigmoid, and radial basis function. In the field of susceptibility mapping, especially for floods and landslides, the latter is widely used [41,42].

4.5. Ensemble Methods

Ensemble learning is a technique that combine multiple base models—or weak learners—into a single model with the purpose of increasing the performance of both regression and classification problems. There are several reasons why ensemble methods improve predictive performance [43]:

- Overfitting avoidance: when the number of data points available for training is limited, learning algorithms tend to fit the training data too closely and perform poorly on unseen instances. Averaging multiple predictions from different models can effectively reduce this limitation and improve the overall predictive performance.
- Local optima avoidance: single machine learning algorithms have the possibility of getting stuck in local optima solutions. This drawback is reduced in ensemble methods with the combination of multiple learners.
- Expansion of search space: the best solution for a problem can be outside the hypothesis space of any single model. However, the combination of different models can expand the search space and increase the chance of finding the best fit for the data.

Several machine or deep learning algorithms can be used as base models, with decision trees being one of the most common. The main ensemble techniques are bagging, boosting, and stacking.

Bagging (Figure 8) trains multiple base learners with different subsamples of the original training dataset and combines the prediction of each learner with averaging, voting, or other methods based on the problem formulation. One of the most widely used examples of bagging is random forest, a tree-based algorithm popular for its high accuracy and the fact that its hyperparameters [44], i.e., the number of trees and the number of features used to split each node, can be easily tuned. In the study of landslide susceptibility mapping, these two hyperparameters are optimised with the use of the grid search techniques [45]. This approach consists of creating a discrete grid with all possible variable combinations and evaluating each of them with validation criteria to determine the optimal solution.

Boosting (Figure 9), instead, sequentially trains the base models by increasing the weights assigned to the observations predicted poorly by the previous iterations. The final prediction of the model is calculated with a weighted sum or vote of all the iterations' results [46]. Popular ensemble algorithms include AdaBoost [47], gradient boosting machine [48], and extreme gradient boosting [49].

Stacking (Figure 10) is a technique that can be used to combine heterogeneous base models and is composed of two steps. In the first step, the base models are trained on the original training dataset to obtain their decisions. These decisions are then used to train another model, called meta learner, which will provide the final prediction [50]. The stacking method has been applied for flash flood susceptibility mapping [51]. In the study, the predictions from the base models are combined to create a new feature set, which is then used by the meta learner for training. The authors chose as base models KNN, logistic regression, SVM, and random forest, while the meta learner is another logistic regression model. The use of weak learners of different natures is often desirable to improve the accuracy, as each of them may be able to learn different concepts from the data. The logistic regression algorithm as a meta learner has the advantages of being simple and making the final output easy to interpret.

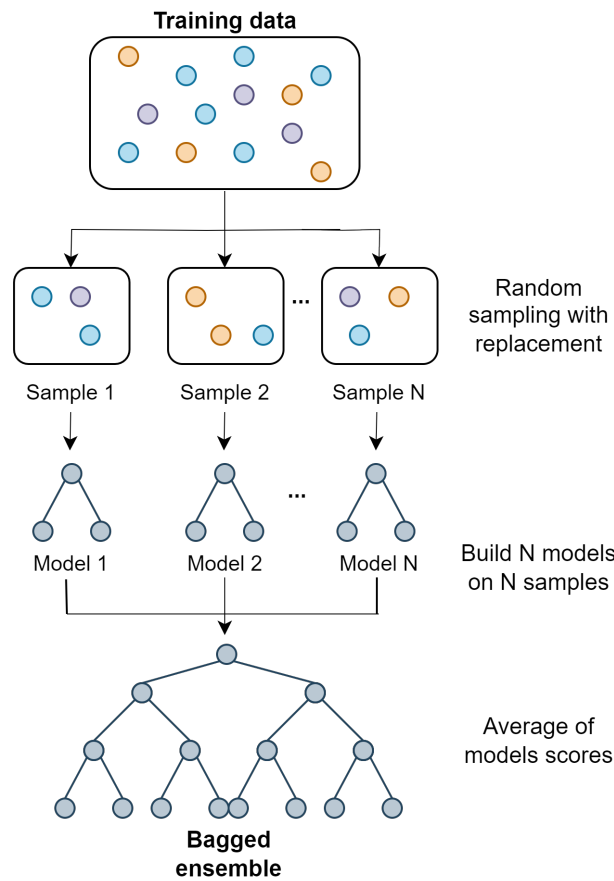


Figure 8. Scheme of the bagging ensemble method.

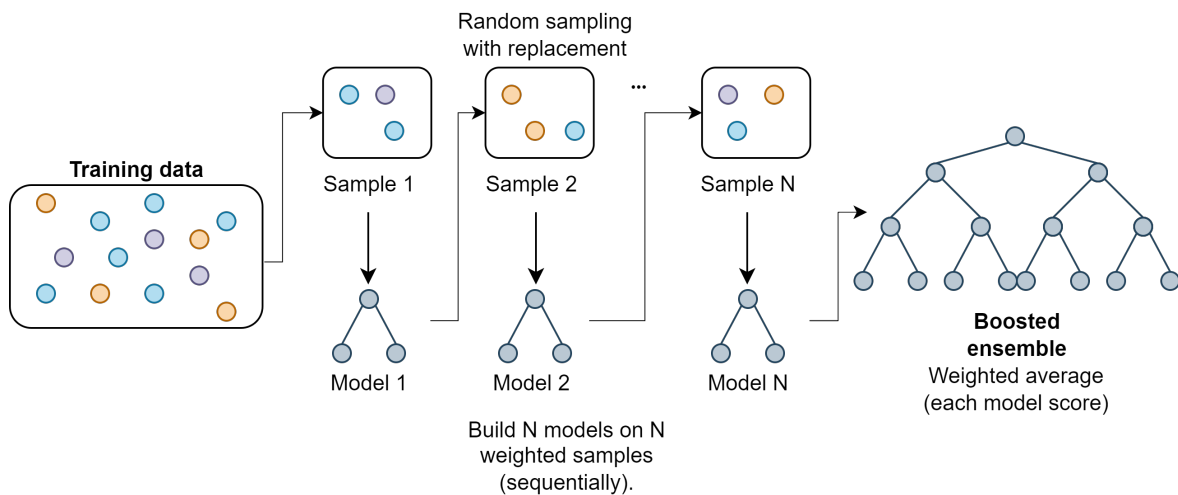


Figure 9. Scheme of the boosting ensemble method.

Ensemble algorithms can be particularly useful in susceptibility mapping as they can mitigate the problem of class imbalances, where one class of data (e.g., “flooded”) is significantly less represented compared to the other class (e.g., “non-flooded”). This situation is frequent for hazards such as landslides and floods since the areas impacted by past events are often smaller than those not affected by the hazard. To address this issue, ensemble methods can be used, e.g., to create a combination of base models that are trained with the balanced subsamples of the data, or to combine undersampling and ensemble techniques [43]. Both approaches ensure that the algorithm is not biased towards the majority class.

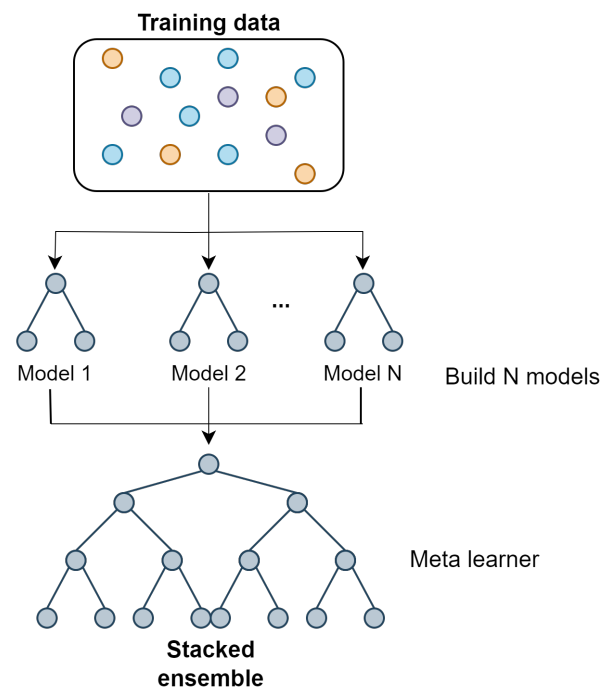


Figure 10. Scheme of the stacking ensemble method.

4.6. Instance Based

Instance-based is a family of algorithms which, instead of creating a parametric model to perform predictions, compare the new observations with the previous instances seen in training, which are stored in memory [52]. A popular algorithm in this class is the K-nearest neighbour [53], which assigns the output to each instance based on the values of its closest neighbours. When using KNN, a distance metric used to measure the proximity between data samples must be defined based on the problem settings, while the K parameter represents the number of neighbours that are used to compute the output of new instances with a majority vote or an average. For example, for landslide susceptibility mapping, the Euclidean and Manhattan distance metrics were chosen, while the possible K values were 3, 5, 11, and 19 [54]. The optimal values were selected with a grid search.

4.7. Bayesian

Bayesian algorithms in ML are methods that are based on the Bayes probability theorem [55]. The theorem is used by the algorithms to model the probability of each class as a conditional probability, which is updated with new evidence from new observations. One of the main advantages of these algorithms is the possibility of introducing a priori information on the problem inside the model, which is not usually possible with other ML algorithms. The most popular Bayesian method is Naive Bayes [56] which is used for classification. Naive Bayes computes the probability of each feature value given a particular class. Once all the feature probabilities are calculated, they are multiplied to obtain the joint probability of a given class. This joint probability is then multiplied by the prior probability of the class to obtain the final probability of the class given the input features. The main disadvantage of naive Bayes is the assumption that the data features are independent, which is often not the case in the geospatial field. Other algorithms in this category such as Bayesian networks avoid this assumption, allowing them to model more complex relationships among the features.

4.8. Time Series

Time series forecasting algorithms are specifically developed to deal with time series data as input. They include the time factor in the model by introducing a dependency

among all the observations, which provides additional information, and the prediction of future values is based on the previous ones [57]. The output of these algorithms is provided as a time series. Widely used time series methods are auto regressive models and prophet, which is an algorithm developed by Meta [58]. For what concerns the articles in this review, they are only used for the air pollution hazard, as it is often based on temporal data.

4.9. Other Classes

Some of the classes contain algorithms that are rarely used in the reviewed articles, but they are reported for completeness.

4.9.1. Dimensionality Reduction

Dimensionality reduction is a preliminary step in machine learning which is used to reduce the number of features, or dimensions, in a dataset. It is possible to distinguish between feature selection and feature extraction methods. The former reduces the dimensions by only keeping a subset of the original features, while the latter projects the original data to a new feature space with different attributes [59]. This class contains algorithms that combine the process of feature extraction with the regression and the classification step, such as principal component regression and linear discriminant analysis. This type of methods can be useful for geospatial problems, which often use a high dimensional dataset. Feature selection algorithms will be further described in Section 5.

4.9.2. Statistical

This class contains algorithms that are not part of machine learning but are pure statistical analysis methods, and are sometimes proposed as a benchmark for comparison with ML and DL algorithms for a case study. The most used algorithm of this kind, particularly for landslide susceptibility, is the frequency ratio algorithm [60]. A good reason to apply these methods is that they estimate how much each conditioning factor influences the target variable, providing useful insights into the problem.

4.9.3. Clustering

Clustering is an unsupervised machine learning task, where the main goal is to partition objects into groups of similar objects (clusters) and to discover hidden structures or patterns in the data. Commonly used algorithms are K-means and hierarchical clustering [61]. These methods are used in the reviewed articles, e.g., using K-Means to define flood susceptibility indices [62], or using K-means and DBSCAN (density-based spatial clustering of application with noise) to produce landslide susceptibility maps [63,64].

4.9.4. Rule-Based Systems

Rule-based systems are machine learning algorithms that have the goal of automatically generating a set of rules which will be used to perform the predictions. Each rule is in the form "IF condition THEN action" and should have two characteristics: the rules allow the machine to implement an optimal strategy towards its environment, and the rules do not contain unnecessary information. Rule-based algorithms are different from decision trees since the rules are not mutually exclusive [65], e.g., the prediction of air pollution concentrations using the cubist rule system method [66].

5. Feature Selection

As previously described in Section 4.9.1, feature selection is the process of selecting the set of data variables that will be used in the prediction model. This is a crucial step of machine learning, especially when dealing with high dimensional problems, which is often the case in the geospatial field. This process will determine the quality and performance of the produced system. Indeed, having fewer features than required will produce a model that is too simple and not capable of predicting the right output or finding the best patterns

in the data; on the other hand, selecting too many features may lead to overfitting and an excessive increase in model complexity.

Feature selection is closely related to a common problem in machine learning first introduced by Bellman [67] as the “curse of dimensionality”. This concept refers to the explosive nature of spatial dimensions and their resulting effects, such as an exponential increase in computational effort, large waste of space, and poor visualisation capabilities.

Because of its nature, some feature selection methods provide the importance of a feature or subset of features inside an ML/DL model. This information is useful not only for selecting the features to be used in the modelling process but through the lifecycle of the model by providing the means for model interpretability [68].

There are different feature selection algorithms that can help in the process of the identification and removal of irrelevant and/or redundant variables and later in model interpretation. They can be separated into four main categories as follows.

5.1. Wrappers

These methods explore different combinations of features to find the best subset. This is achieved by training and evaluating an ML/DL classifier using each feature subset to measure its quality. The selection of features depends on the chosen machine learning algorithm.

The search for the best feature subset can be performed in two ways: sequentially or by means of a heuristic algorithm. In the sequential approach, the process starts with either an empty set or the full set of features and gradually adding or removing features until the best result is achieved based on a defined objective function. These methods are called sequential forward selection and sequential backward elimination, respectively.

On the other hand, heuristic algorithms use a greedy search strategy to find a subset that optimises the objective function.

Sequential and heuristic wrapper methods typically yield better predictive accuracy compared to filter methods. However, they come with a trade-off of higher computational complexity [69].

5.2. Filters

These methods base their selection on a rank of the features’ intrinsic properties. They do not rely on a specific ML/DL algorithm as wrappers do. For this reason, filter methods are computationally less expensive compared to wrappers, but they tend to have a lower prediction performance.

Filter methods establish a ranking criteria and a threshold. The ranking is based on the correlation or dependency between variables. Features that fall below the set threshold are considered less important and are therefore removed from the final feature set. This ranking-threshold-based procedure helps determine the subset of features to be used for subsequent analysis [69]. The methods used for feature selection include Pearson’s correlation, linear discriminant analysis (LDA), analysis of variance (ANOVA), and mutual information (MI) [70].

5.3. Embedded

These are methods that incorporate the feature selection process into the model construction itself. One example of such methods is lasso regression. In this technique, a penalty term is introduced and calculated based on the absolute values of the coefficients. By applying this penalty, some of the coefficients are shrunk to zero, effectively removing the corresponding features from the model. Only the features with non-zero coefficients are retained in the final model.

Embedded methods fall between wrapper and filter methods in terms of computational complexity. They are computationally more intensive than filter methods but less demanding than wrapper methods [69].

5.4. Explainable

Explainable artificial intelligence (XAI) consists of producing human-readable explanations of the feature contributions in a model and/or of the model structure. This helps in a first instance on the feature selection process while building the model for its enhancement, bias detection, and robust model building, and in a second instance, on model interpretation and transparency [68]. There are two categories for model interpretability and explainability [71]: integrated and post hoc explanations.

Integrated explanations refer to transparency-based models, i.e., white-box models with an output that directly allows their interpretation, e.g., linear models, decision trees, and generalised additive models. More recent approaches include interpretable deep neural networks which rely on layer-wise relevance propagation to analyse the correlation between dependent and independent variables [72].

On the other hand, post hoc explanations refer to black-box models for which additional methods have to be used to understand the model and the roles of the features inside the black-box model, e.g., support vector machines and neural networks. Post-hoc model explanations consist of a white-box model built using the black-box predictions as targets in order to mimic its behaviour, providing a description of the model via decision trees or lists of rules [73]. Furthermore, post hoc outcome or prediction explanations are produced by building a model in the vicinity of the observation of interest [73]. Different methods have been used in the context of hazard susceptibility mapping to provide the post hoc prediction explanations of black-box models, e.g., the Shapley additive explanation values (SHAP) [74,75] and the local interpretable model-agnostic explanations (LIME) [76,77].

There is a trade-off between model readability and model performance which makes it so that white-box models are easier to interpret, whilst black-box models allow complex modelling and therefore, tend to have better performance [78].

6. Hazard Susceptibility Modelling

Hazard susceptibility modelling refers to the process of learning from the environmental and socio-economic conditions of the previous occurrences of hazardous events to estimate the hazard susceptibility in unseen locations, considering the similarity of the conditioning factors (CFs). A general methodology for the susceptibility modelling of hazards using ML/DL techniques was consolidated from the reviewed articles, as depicted in Figure 11 and described in the following.

The historical hazard occurrences and the CF are retrieved from different sources, e.g., observations from sensor networks, national or regional authorities datasets, and satellite imagery. Both historical occurrences and CF are preprocessed to improve the input data. The most common approaches are data standardisation or normalisation, outlier removal, temporal and/or spatial gap filling, data transformation (e.g., discretisation), and the signal decomposition of time series data.

Additionally, a geospatial analysis may be performed to extract the geospatial relationships between the data. This is specific to air pollution and urban heat island when modelling discrete data, e.g., in situ stations data. Furthermore, in several studies [79,80], the historical occurrences are considered as positive samples; then, negative samples are extracted from unfeasible hazard locations and the problem is approached as a binary classification.

Feature selection may be performed to reduce the number of CF and to avoid feature correlation. The most used methods in the reviewed articles are Pearson correlation, Spearman's correlation, random forest feature importance, information gain, recursive feature elimination, variance inflation factor and embedded feature selection, i.e., the most important features are directly selected by the model. In most recent articles, XAI has been used for feature selection as well [81].

The ML/DL modelling part consists of training a model with the occurrences and CF data and the direct production of the hazard susceptibility map in a specific area of interest.

However, it may include additional steps such as a feature selection, correlation analysis, model hyperparameter optimisation, and model embedded feature selection.

The hyperparameters of the models are often optimised using grid search and K-fold cross-validation. Furthermore, in several articles, more advanced techniques are used for model optimisation, i.e., attention-based mechanisms, Kalman filter, swarm intelligence algorithms, and genetic algorithms.

There are different accuracy assessment metrics among the reviewed articles which include: overall accuracy, relative absolute error, mean absolute error, root relative squared error, root mean squared error, coefficient of determination (R^2), receiver operating characteristic (ROC), area under the ROC curve (AUC), precision, recall, F-1 score, sensitivity, and specificity.

In the reviewed studies, one or multiple models are trained on the susceptibility task; afterwards, their accuracy is tested to select the best one. The best model is used for the production of the hazard susceptibility map in the entire area of interest. There are three types of approaches for using one or multiple models. The first is that only one state-of-the-art model is used [82]. The second is when multiple state-of-the-art models are tested [83]. The third is when a novel approach is proposed and state-of-the-art models are tested to showcase the accuracy improvement of the proposed approach [84].

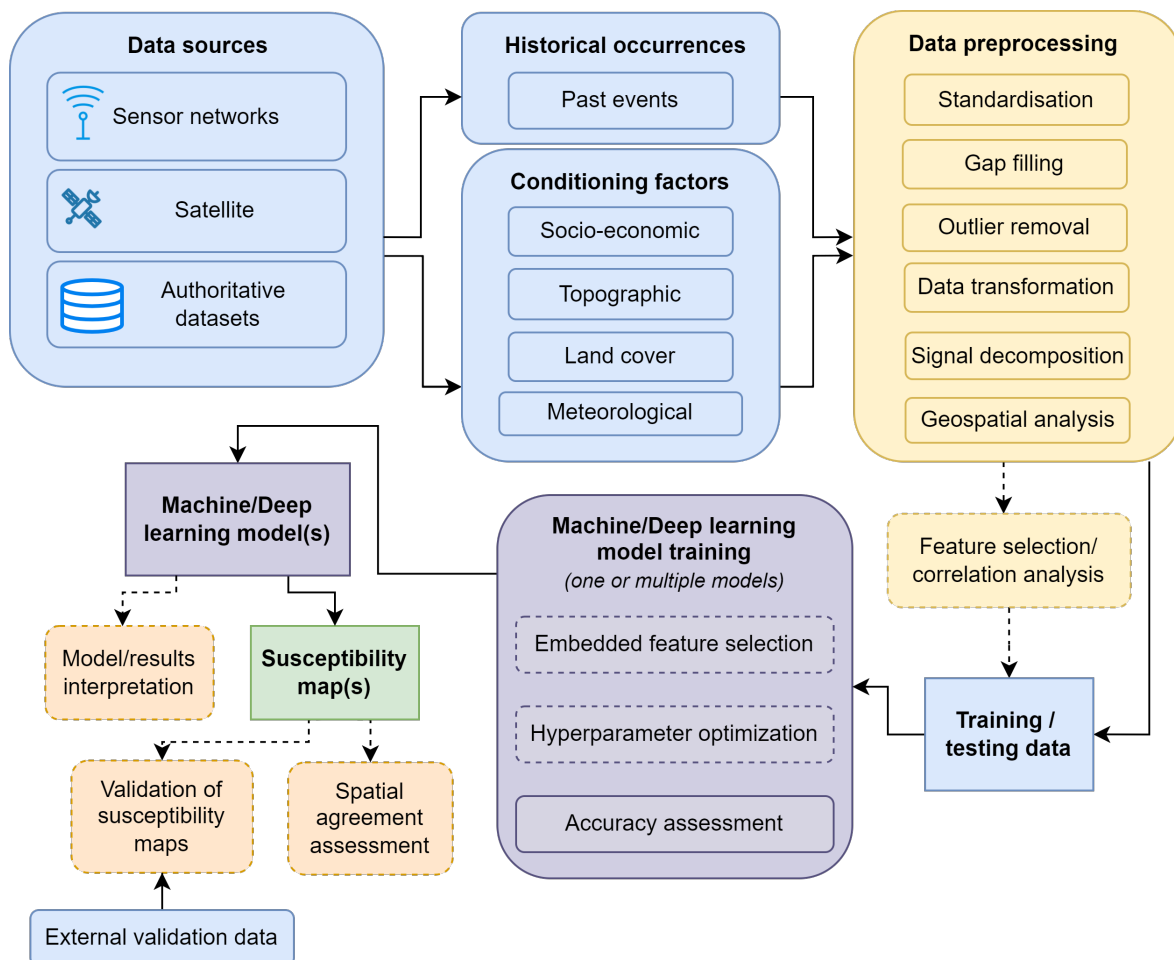


Figure 11. Hazard modelling workflow. * The solid connections suggest that it is a mandatory step, while the dotted connections suggest that it is an optional one. ** The blue colour groups the data, the yellow groups the data processing (preprocessing and feature selection), the purple groups the modelling process and the model, the green groups the susceptibility map as a product, and lastly, in orange groups the final optional steps.

There are three additional (optional) steps which help provide an enhanced post-modelling result. The first is model interpretation, which consists of using post hoc explanation methods to understand the contribution of CFs in the ML/DL model (black-box model). The second is the assessment of the spatial agreement between susceptibility maps. Some studies perform an analysis to understand the spatial agreement of the hazard susceptibility maps produced by different models. It can be the case that the accuracy metrics are high for different models; however, the resulting maps are spatially different. To overcome this, methods such as the McNemar’s test have been used to understand the spatial agreement of hazard susceptibility maps derived from different models [85,86]. Lastly, the susceptibility maps can be validated with external data, such as new event occurrences or authoritative data. The quality of the results as well as the quality of the model should be validated.

The previously described workflow encompasses the individual modelling of the four hazards considered in this review, namely air pollution, urban heat island, flood, and landslide. Additional considerations of the specifics to each hazard are detailed in the corresponding hazard sections.

The individual hazard susceptibility maps can be combined to produce a multi-hazard susceptibility map, showing the interaction between different hazards at specific locations. There is no consensus on how to combine individual susceptibility maps [87]. Most of the reviewed articles dealing with multiple hazards modelled the individual hazard susceptibility maps and produced the multi-hazard susceptibility maps with the univariate combination of different hazards, resulting in a categorical map. For example, when combining flood, landslide, and fire susceptibility, the resulting map is composed of eight classes, three for each single hazard, three for the combination of two hazards, one for all hazards, and one for no hazards [82]. Similar approaches have been followed in other studies [88,89]. The only different approach is to use a Mamdani fuzzy inference system method which consists of a control system using linguistic rules based on experts’ opinions [87]. The resulting multi-hazard susceptibility map is continuous.

The following subsection lists the CFs which may be influential for two or more hazards (see Section 6.1). There are conditioning factors which are specifically related to a single hazard which will be detailed in the specific data acquisition subsections per hazard, including the historical occurrences data.

6.1. Conditioning Factors Common to Multiple Hazards

Table 2 reports the conditioning factors (CFs) considered in the reviewed articles which are available at a global or continental level.

Table 2. Conditioning factors common to multiple hazards and examples of datasets.

Type	Variable	Source	Spatial Resolution	Temporal Resolution	References
Meteorological data	Temperature Wind Rain Solar radiation ...	ERA5 ¹	0.25° × 0.25°	Hourly	[90,91]
Meteorological data	Temperature Atmospheric pressure Wind Humidity ...	Netatmo ²	Discrete	5 min	[92]

Table 2. Cont.

Type	Variable	Source	Spatial Resolution	Temporal Resolution	References
DEM ³	Elevation	ASTER-GDEM ⁴	30 m	N/A	[93–95]
	Slope	SRTM-DEM ⁵	30 m	N/A	[96,97]
	Aspect ...	ALOS-PALSAR-DEM ⁶	12.5 m	N/A	[98,99]
Derived surface indices and land cover ⁷	LULC NDVI NDBI MDWI ...	Landsat-8	30 m	Revisit time: 16 days	[80,94,100]
		Landsat-5	30 m	Revisit time: 16 days	[96,101]
		Sentinel-2	10 m	Revisit time: 5 days (twin satellites)	[90,102]
		MODIS Terra and Aqua	250 m	Revisit time: 2 days	[103]
Land cover		Copernicus CORINE ⁸ Land Cover 2018	100 m	6 years	[34]
		ESA ⁹ WorldCover 2020	10 m	Annual	[104]
Features	Roads Waterways Power plants Points of interest ...	OpenStreetMap (OSM)	vector	Updated by the community	[104,105]

¹ ERA5 = ECMWF (European Centre for Medium-Range Weather Forecasts) Atmospheric Reanalysis V5. ² Netatmo provides crowdsourced data. ³ DEM = Digital elevation model. ⁴ ASTER-GDEM = Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer - Global Digital Elevation Model. ⁵ SRTM-DEM = Shuttle Radar Topography Mission - Digital Elevation Model. ⁶ ALOS-PALSAR-DEM = Advanced Land Observing Satellite - Phased Array type L-band Synthetic Aperture Radar - Digital Elevation Model. ⁷ Datasets which can be produced based on satellite imagery. NDVI = Normalised difference vegetation index, NDBI = Normalised difference built-up index, LULC = Land use–land cover. ⁸ CORINE = Coordination of Information on the Environment. ⁹ ESA = European Space Agency.

The CFs shared among the four studied hazards are meteorological factors (e.g., temperature, precipitation, wind direction, and velocity), topographic factors (e.g., digital elevation model (DEM)), land use and land cover, and socio-economic factors (e.g., power plants, night-time lights).

There are CFs which are provided by local authorities and are specific to a city, region, or country. These local data generally include meteorological and air pollution data from in situ station networks, digital terrain models, land use and land cover maps, building footprints and heights, road traffic data, and others. The local data usually can be, at least, partially replaced by global or continental products [106], such as those reported in Table 2.

7. Air Pollution Susceptibility Modelling

Air pollution is defined as the presence of toxic chemicals or compounds in the air, at levels that pose a health risk [107]. The most common air pollutants which are modelled in the reviewed articles are particulate matter (PM₁₀, PM_{2.5}), ozone (O₃), nitrogen dioxide (NO₂), carbon monoxide (CO), black carbon (BC), and sulphur dioxide (SO₂). Additionally, the Air Quality Index (AQI) is contemplated in several articles. The AQI is an indicator used by government authorities to categorise pollution in terms of its severity [108]. It is based on the weighted combination of different pollutants, and it may vary per country according to their own pollution levels and restrictions; e.g., in China, the AQI considers the concentration of six pollutants (CO, NO₂, O₃, SO₂, PM₁₀, and PM_{2.5}) to obtain the sub-index of each pollutant using a piecewise linear function and take the maximum argument value

of such sub-indexes to reflect the overall air quality [28]. For European countries, the AQI guidelines are provided by the European Environment Agency [109], and the pollutants considered to compute the index are up to five (PM_{10} , $PM_{2.5}$, O_3 , NO_2 , SO_2).

Air pollution susceptibility modelling and mapping consists, as explained in the general workflow, in correlating air pollution concentration events with the data on conditioning factors at specific locations in order to provide an overview on how the pollution levels would be given certain factors at unseen locations. The air pollution events consist of a concentration level threshold being exceeded [79].

Air pollution susceptibility is directly treated in few of the reviewed articles [79,110,111]; however, several studies proposed methods that could potentially contribute to the production of the susceptibility maps. Therefore, these are also included, e.g., enhancing the pollution concentration data and forecasting the pollution concentration levels. An example of this is the production of high-resolution pollution concentration maps at the national [112] or regional [113] levels.

The following subsections provide insights into the data sources, the data preprocessing, and the ML/DL techniques related to this hazard.

7.1. Air Pollution Data Sources and Conditioning Factors

The air pollution monitoring data, i.e., the occurrences data, can be retrieved from satellite imagery, authoritative datasets, or ground stations, either environmental agencies' sensor networks, crowdsourced networks, or 'amateur' sensor networks. Examples of satellites with instruments for measuring the air pollution concentration are Sentinel-5P TROPospheric Monitoring Instrument (TROPOMI) [114,115] and Aura satellite Ozone Monitoring Instrument (OMI) [116]. Authoritative datasets refer to the data provided by trusted organisations which are validated and provided as a higher-level product, e.g., Copernicus Atmosphere Monitoring Service (CAMS) [76], the China High Air Pollutants (CHAP) [117], and the Air Quality e-Reporting database from the European Environmental Agency [105]. Ground station (or in situ) data are derived from sensor networks that measure the pollutant concentrations at specific locations, providing very accurate data but usually just allow a sparse representation of the pollution levels in the area of interest. The sensors' data may be provided by local authorities, e.g., the U.S. Environmental Protection Agency [118], China National Environmental Monitoring Center (CNEMC) [119,120], the Environmental Protection Department of Hong Kong [120], and the Department of Territory and Sustainability of the Catalonia Government [79]. A sensor network could be built as part of the study as well, e.g., creating an Internet of Things (IoT) system [121,122]. Table 3 lists the air pollution sources that provide data at least at the global, continental, or multi-national levels.

Each of these sources can be used individually as occurrences data or they can be combined to produce and use an enhanced dataset, e.g., using CAMS data to fill the gaps of satellite imagery [116].

The air pollution conditioning factors considered in the reviewed studies are meteorological—temperature, dew point, wind speed and direction, atmospheric pressure, relative humidity, and precipitation; socio-economic—population data, traffic data, road density, night-time lights, points of interest (bus stations, gas stations, heat suppliers, polluting factories, and restaurants); topographic—elevation (digital elevation or terrain model), slope, aspect, and surface imperviousness; and land cover and derived vegetation and built areas indices. Additionally, the aerosol optical depth (AOD) is widely used as a conditioning factor [112].

Table 3. Data sources for air pollution monitoring.

Source	Pollutants	Spatial Resolution	Temporal Resolution	References
Sentinel-5P	O ₃ , SO ₂ , NO ₂ , CO, HCHO, and CH ₄	5.5 km × 3.5 km	Revisit time: daily	[114,115,123]
Aura OMI ¹	O ₃ , NO ₂ , SO ₂ , and aerosols	13 km × 25 km	Revisit time: daily	[116,124,125]
Air Quality e-Reporting database	O ₃ , SO ₂ , NO ₂ , CO, HCHO, and CH ₄	discrete	Daily/hourly	[105]
CAMS ² reanalysis	O ₃ , SO ₂ , NO ₂ , CO, HCHO, and CH ₄	10 km × 10 km	Hourly	[76,116]
MERRA-2 ³	PM _{2.5} , BC, and aerosols	0.625° × 0.5°	Hourly	[116,117,126]

¹ OMI = Ozone Monitoring Instrument. ² CAMS = Copernicus Atmosphere Monitoring Service. ³ MERRA-2 = Modern-Era Retrospective Analysis for Research and Applications, version 2.

7.2. Air Pollution Monitoring Data Preprocessing

The preprocessing of air pollution monitoring data consists of gap filling, outlier removal, transformation, decomposition, and an additional geospatial analysis. The latter applies when working with discrete data and is conducted to provide spatial relationships to the model.

Temporal and spatial gaps are filled by means of interpolation, average, fill forward, or estimated using ML models. Some use cases of temporal air pollution data gap filling are linear interpolation for small gaps and the average value for significant ones [127], spatiotemporal interpolation [128], random forest and linear interpolation [129], and light gradient boosting machine model [117]. Examples of spatial gap filling could consist of using complementary sources, e.g., using MODIS Terra and Acqua [112], CAMS data [116], or training ML/DL models to predict the values at missing locations [130].

Outliers or anomalies are a common cause of error in the model estimates, mostly on time series data. However, few authors specify the techniques used to remove them during the preprocessing phase. Some of the techniques for outlier removal are the Hampel identifier which is a triple standard deviation discriminant algorithm [131], the numerical limiting method [129], and the interquartile range method [132,133].

The decomposition method core concept is to decompose the original non-stationary series into several relatively more stable subseries. For each subseries, a predictor is established to achieve the forecasting task. The decompositions proven to be effective in processing pollution levels and AQI time series are Wavelet and multiscale analysis [134] and variational mode decomposition [131]. This applies to time series non-spatial data.

7.3. Insights into Pollution ML/DL Modelling Techniques

The articles included in the following statistics and insights regard air pollution susceptibility modelling and also air pollution modelling studies that can contribute to the production of susceptibility maps, i.e., as previously mentioned, the enhancement of air pollution concentration maps and the air pollution forecast using temporal or spatiotemporal data. Enhancement provides a better product for obtaining more accurate air pollution susceptibility maps and the forecast provides ML/DL techniques and approaches that can be extrapolated for susceptibility modelling. The total number of articles used to produce Figure 12 was 534, which were those with at least one citation the date they were retrieved. The label of less recurring algorithms was removed to improve readability. All the statistics are available in the repository created for this article (see Section 3).

Figure 12 shows the distribution of the ML/DL classes defined in Section 4 and the techniques used for air pollution temporal and/or spatial modelling and forecasting, with a visible label if present in more than 20 articles. The largest used class for dealing with air pollution temporal and spatiotemporal modelling is neural networks, followed by the ensemble (ENS), linear regression (LR), and support vector machine (SVM) classes. Concerning the specific algorithms, more than 117 different ML/DL techniques were used in the reviewed articles, without considering the variations due to the hyperparameter optimisation algorithms. The most used technique was long short-term memory (LSTM), used in a total of 203 articles, followed by random forest (RF), support vector machines

(SVMs), convolutional neural networks (CNN), extreme gradient boost (XGB), and linear regression, which were used in a total of 166, 81, 81, 65, and 62 articles, respectively.

Different variations of LSTM were used, such as graph LSTM [135], spatiotemporal LSTM [27,136], and bidirectional LSTM (BLSTM) [137]. Furthermore, to improve the results, LSTM and its variations are often ensembled with other neural networks (NNs) and optimisation methods. Examples of combinations with other NNs are convolutional neural network (CNN) LSTM–CNN [28,138,139], gated recurrent unit (GRU) [140], or several models like adding graph convolutional network (GCNs), multi-layer perceptron, and Gaussian process regression (GPR) to propose the GCN–LSTM–MLP–GPR model [141], or artificial neural networks (ANNs) and recurrent neural networks (RNNs) to propose the ANN–LSTM, RNN–LSTM, and ANN–RNN–LSTM models [142]. Examples of optimisation algorithms are convolutional block attention module [143], multi-verse optimisation algorithm [144], and Bayesian optimisation [145]. The combination of CNN or other algorithms which are commonly used in image processing alongside LSTM is crucial for the spatiotemporal modelling of air quality.

Several articles deal with the air pollution concentration forecast using air pollution concentration measurements from in situ stations, i.e., point data, or using an additional continuous representation of the air pollution concentration levels from satellite data, both coupled with other auxiliary data (CF). Examples of the first scenario using the previously mentioned approaches are the AQI forecast using CNN, LSTM, and CNN–LSTM [28] and the PM_{2.5} forecast using 1D-CNN, Bi-LSTM, and their combination [146]. Other state-of-the-art approaches include support vector regression (SVR) [147] and extreme learning machine coupled with the multi-objective Harris hawks optimisation [146]. Novel approaches include the implementation of Transformers, e.g., the spatiotemporal transformer model for the prediction of hourly PM_{2.5} [148] and the autocorrelation-error informer model [149], which outperformed other state-of-the-art models based on time series like LSTM, GRU, and ARIMA. Examples of the second scenario are the satellite-based PM_{2.5} prediction using a spatiotemporal deep neural network (DNN) [150] and RF [151].

Examples of articles dealing with the enhancement of air pollution concentration maps are the production of daily 1km PM_{2.5} estimates in China using extremely randomised trees (ERTs) using an AOD product, in situ PM_{2.5} data, and auxiliary data [112]; daily 1 km NO₂ estimates in the United States using the NO₂ column density from OMI Aura satellite, NO₂ from chemical transport models, land cover, and ancillary data as input to the ensemble of neural network, RF, and gradient boosting [116]; and daily maximum 8-hour average O₃ at 0.1° resolution using an XGB model with ground-level ozone monitoring data, MERRA-2 data, and geographical data [124]. This type of study tends to implement tree-based techniques like RF, light gradient boosting machine, and XGB, which usually provide the highest accuracy [123,151–153].

Transfer learning has been implemented in some articles. In this context, it consists of reusing a highly accurate model and fine tuning it with a reduced amount of data from a different time and/or location. It is a solution for predicting or estimating air pollution levels at locations/monitoring stations which lack data. Transfer learning stacked BLSTM was used to train with the data of historical monitoring stations and fine tune with the scarce data of a new station, which resulted in an accuracy improvement of 23.37% with respect to traditional training [154]. Similar results were obtained in transfer learning studies using LSTM [155,156] and Gaussian mixture [93] models. Furthermore, the prediction performance of the air pollution concentration of a city was tested based on a CNN–LSTM model trained on another city; in this case, the transfer learning did not improve the testing accuracy [139]. Moreover, the spatial and temporal transferability of models, without fine tuning, has been tested in some studies [157,158].

An additional fact observed in the reviewed articles is the utilisation of integrated and post hoc explainable artificial intelligence (XAI). Examples of integrated XAI are the proposal of an interpretable DNN [150] for predicting PM_{2.5} and extracting the spatiotemporal features of the model and a second interpretable DNN for understanding the contribution of

urban traffic to air quality [72]. Examples of post hoc XAI are the usage of Shapley additive explanations (SHAP) values to analyse the feature contributions in the model [123,159,160].

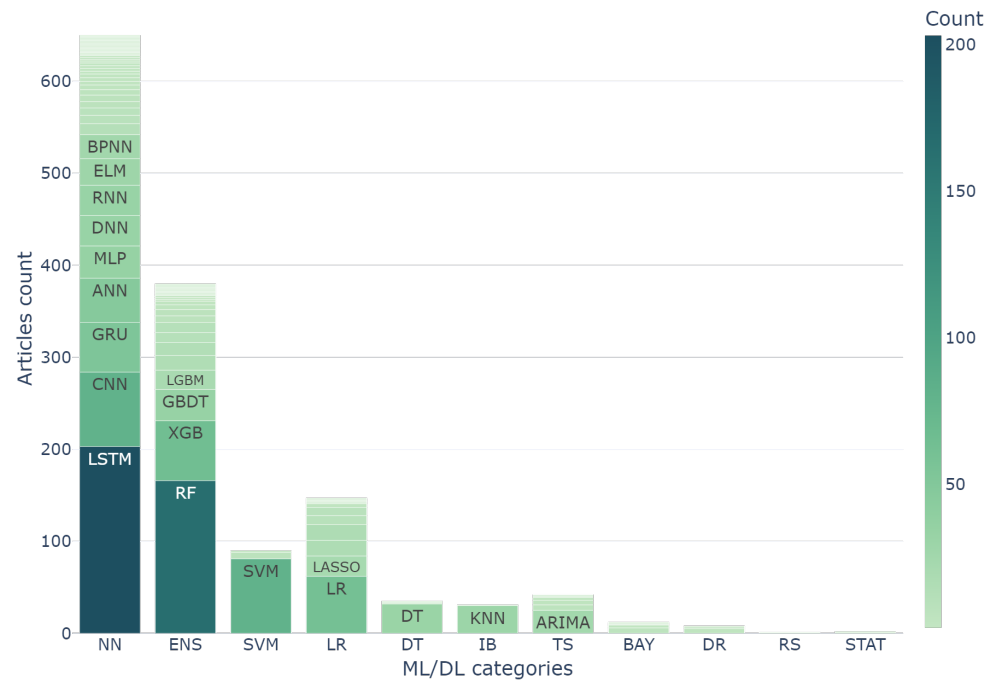


Figure 12. Air pollution algorithm classes (see Section 4) and frequently used methods. Complete dataset available in the repository (see Section 3). * BPNN = Backpropagation Neural Networks, ELM = Extreme Learning Machine, RNN = Recurrent Neural Networks, DNN = Deep Neural networks, MLP = Multi-Layer Perceptron, ANN = Artificial Neural Networks, GRU = Gated Recurrent Unit, CNN = Convolutional Neural Network, LSTM = Long Short-Term Memory, LGBM = Light Gradient Boosting Machine, GBDT = Gradient Boosting Decision Trees, XGB = Extreme Gradient Boosting, RF = Random Forest, SVM = Support Vector Machine, LASSO = Least Absolute Shrinkage and Selection Operator Regression, LR = Linear Regression, DT = Decision Trees, KNN = K-Nearest Neighbours, ARIMA = Autoregressive Integrated Moving Average.

Furthermore, Figure 13 exposes the articles’ spatial distribution based on the affiliation country of the first author. It can be clearly seen that China has the highest number of studies related to air pollution, followed by India and the United States.

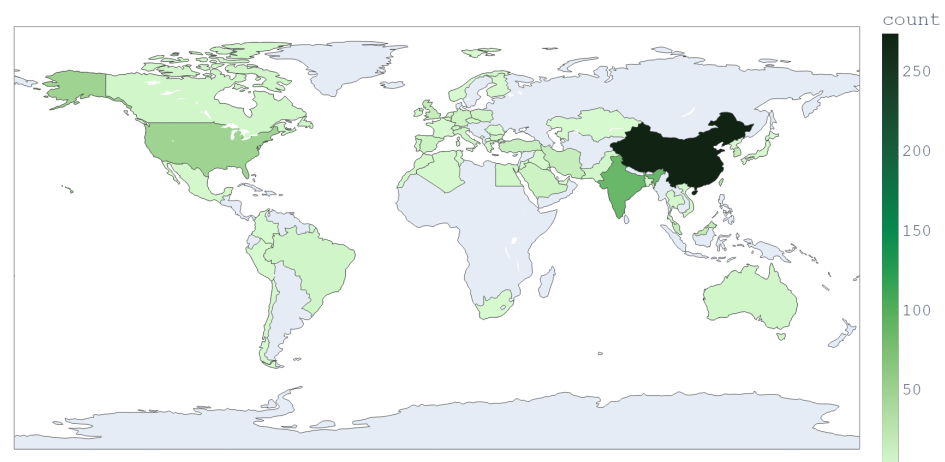


Figure 13. Air pollution articles’ distribution based on the first author’s affiliation.

8. Urban Heat Island Susceptibility Modelling

Urban heat islands (UHIs) are urban areas composed of structures, e.g., buildings, roads, and other infrastructure, that become ‘islands’ of higher temperatures with respect to outlying areas. This happens since infrastructures absorb and re-emit the sun’s heat in a higher ratio than natural landscapes. UHIs are usual when there is a high infrastructure concentration and greenery is limited [161].

The related articles describe the thermal characteristics of an area of interest by modelling the air temperature [95] or the land surface temperature (LST) [162] and deriving indices such as the urban heat island intensity (UHII) [163], the Urban Thermal Field Variance Index (UTFVI) [164], and the Urban Thermal Environment Index (UTEI) [97]. These indices will be described as follows.

The UHII is defined as the difference in surface temperature between the urban centre and the countryside [165]. UHII can be computed as the difference in temperature between a station or a pixel and the reference temperature [90], for discrete or continuous data, respectively.

$$UHII = T_i - T_{countryside} \quad (1)$$

where T_i refers to the temperature at a station or cell i and $T_{countryside}$ refers to the temperature of the reference station, the average temperature of multiple reference stations, or the average temperature of the cells belonging to the countryside area.

The UTFVI expresses the thermal characteristics of the city with respect to the mean temperature of the same space. It can be computed as stated in the following Equation [164].

$$UTFVI = \frac{T_i - T_m}{T_m} \quad (2)$$

where T_i refers to the temperature at station or cell i and T_m refers to the mean temperature of the city or area of interest (AOI).

A similar index is the UTEI which only considers the temperature of the AOI, which is further described in the following Equation [97].

$$UTEI = \frac{\sum_{i=1}^m \sum_{j=1}^n (x_{i,j})}{m \times n} \quad (3)$$

where $x_{i,j}$ stands for a cell value in a continuous representation, where m and n the dimensions of the grid.

The objective of modelling UHI with the UHII index or other similar indices is to explore the spatiotemporal evolution of the thermal characteristics in an AOI. In the reviewed articles, the monitoring of the thermal characteristics is divided into daytime and night-time [90] or seasonally into summer and winter [162].

8.1. Urban Heat Island Data Sources and Conditioning Factors

Among the reviewed articles, land surface temperature (LST), land use–land cover (LULC) [96,101,164], and meteorological variables are combined with other ancillary data to build the models. Important ancillary data include albedo, building height, proximity to water bodies, and anthropogenic heat sources [97,166,167].

Table 4 lists the global sources available for the retrieval of LST, LULC, albedo, and anthropogenic heat flux.

Table 4. Urban heat island conditioning factors and the frequently used sources.

Variable	Source	Spatial Resolution	Temporal Resolution	References
LST ¹	MODIS ² Terra and Aqua	1 km	Revisit time: daily (morning pass Terra + afternoon pass Aqua)	[168,169]
	Landsat 8/5	30 m	Revisit time: 16 days	[162,170]
	Sentinel-3	1 km	Revisit time: 2 days	[171]
AOD ³	MODIS Terra and Aqua	10 km	Revisit time: daily (morning pass Terra + afternoon pass Aqua)	[168]
Albedo	MODIS Terra and Aqua	0.05°	Revisit time: daily (morning pass Terra + afternoon pass Aqua)	[168]
	Landsat 8	30 m	Revisit time: 16 days	[172]
Anthropogenic heat flux	NOAA ⁴ night-time lights	1 km	Daily/monthly	[93]

¹ LST = Land Surface Temperature. ² MODIS = Moderate Resolution Imaging Spectroradiometer. ³ AOD = Aerosol Optical Depth. ⁴ NOAA = National Oceanic and Atmospheric Administration.

Besides the different factors that may condition UHIs (see Figure 11), several environmental indices are considered as well, i.e., the Normalised Difference Vegetation Index (NDVI), the Normalised Difference Built-up Index (NDBI), or the Normalised Difference Water Index (NDWI) [173]. NDBI exposes the impervious surfaces that reduce humidity and increase the environmental temperature. On the other hand, NDVI and NDWI show the correlation between vegetation, water, and water present in vegetation with respect to humidity and the vegetation cooling effect.

Additional data to be considered are the urban growth and population growth [169,174] if the UHI analysis is focused on these variables.

8.2. Insights into the Urban Heat Island ML/DL Modelling Techniques

In the UHI susceptibility modelling and related studies, the most used ML method is ensembles (ENS), followed by neural networks (NNs) and regressions (LRs), as shown in Figure 14, where the specific ML/DL techniques can be appreciated as well. A total of 25 ML/DL algorithms were used to model UHI and related studies. The most used method was random forest regression (RFR), followed by an artificial neural network (ANN), linear regression (LR), random forest (RF), and support vector regression (SVR).

A common approach to studying UHIs is to analyse the LULC and the LST, and subsequently derive the UHII or related indices, e.g., using the ANN to predict the LULC, LST, and UTFVI in 2029 and 2039 during summer and winter based on Landsat imagery [162]. Similarly, forecasting the LULC, LST, UHII, and UTFVI [101].

Other derived information may be useful to analyse the interaction between the LULC landscape and the thermal comfort, such is the case of the morphological spatial pattern analysis (MSPA) [175] of green space or built-up areas. For example, the MSPA of green space can be extracted from land cover maps and through RF deriving the nonlinear relationships between the UHII, the vegetation patterns, and ancillary data like sky view factor, density of green space, mean building height, etc. [163].

Moreover, other studies focus on the direct forecast of the UHII, for example, utilising deep neural networks (DNNs) to forecast the magnitude of UHI in a large city based on temperature measurements derived from weather stations [166]. Additionally, the authors of the previous study proposed a new index called UHI-hours which provides the number of hours in which the UHI phenomenon occurs in a specific location. The aim is to quantify the cumulative effects of UHI.

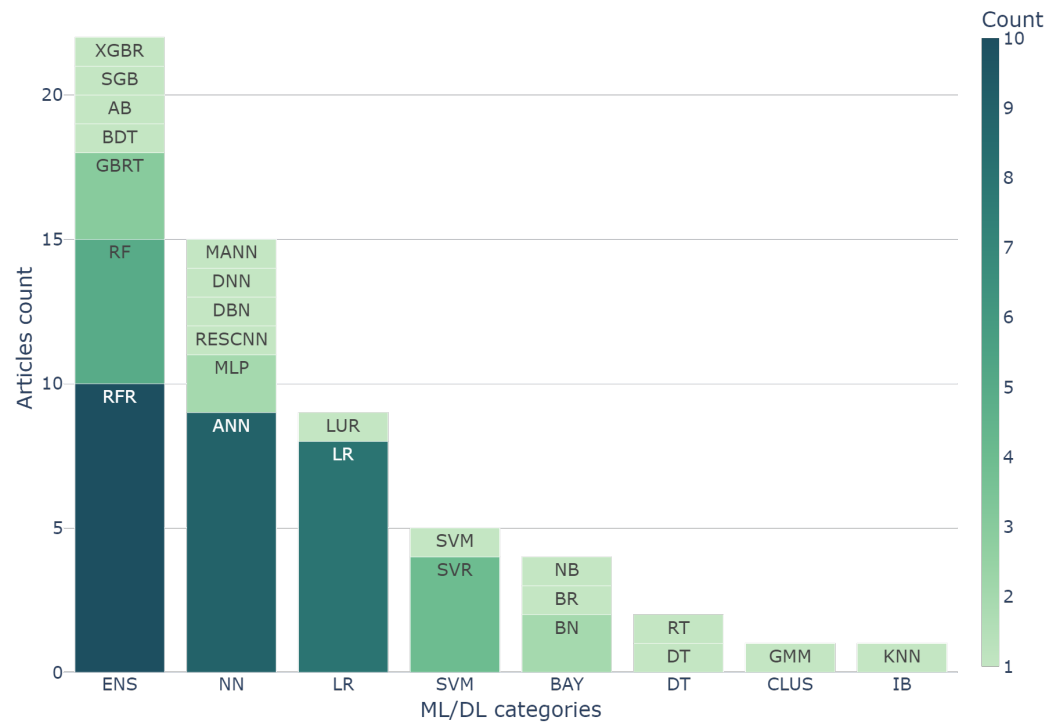


Figure 14. Urban heat island algorithm classes count (see Section 4). Complete dataset available in the repository (see Section 3). * XGBR = Extreme Gradient Boosting Regression, SGB = Stochastic Gradient Boosting, AB = AdaBoost, BDT = Bagging Decision Trees, GBRT = Gradient Boosted Regression Trees, RF = Random Forest, RFR = Random Forest Regression, MANN = Model-Averaged Neural Network, DNN = Deep Neural Network, DBN = Deep Belief Network, RESCNN = Residual Convolutional Neural Network, MLP = Multi-Layer Perceptron, ANN = Artificial Neural Network, LUR = Land Use Regression, LR = Linear Regression, SVM = Support Vector Machine, SVR = Support Vector Regression, NB = Naive Bayes, BR = Bayesian Regression, BN = Bayesian Network, RT = Regression Trees, DT = Decision Trees, GMM = Gaussian Mixture Models, KNN = K-Nearest Neighbours.

Furthermore, Figure 15 exhibits the spatial distribution of the articles considering only the affiliation of the first author. It can be observed that China has the highest number of articles while other countries produced one or two articles.

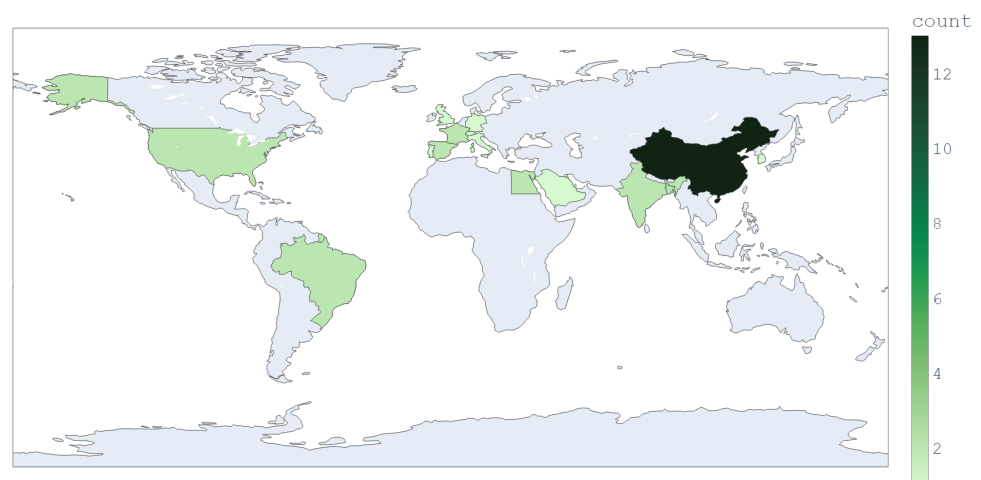


Figure 15. Urban heat island articles’ distribution based on THE first author’s affiliation.

9. Flood Susceptibility Modelling

Floods are defined as an overflowing of water onto land that is normally not covered by water. Different factors can trigger flooding, e.g., heavy rains, ocean waves, snow melting, and dam breaks [176]. Consequently, floods can be classified into different types based on how, why, and where they occur: river floods, flash floods, coastal floods, urban floods, and dam floods [11]. The types of floods included in the review are coastal floods [177], flash floods [178], river floods [179], and urban floods [180].

Flood susceptibility modelling aims to determine the probability of the flood occurrence across the study area. It is treated, in the reviewed literature, as a binary classification problem where the two classes are flooded and non-flooded. The methodology for the production of the susceptibility maps is similar in most of the articles, following the workflow explained in Section 6. ML/DL models are created to provide the probability of being 'flooded' in a specific location starting from the creation of the historical flood occurrence database, the collection of the conditioning factors data, their preprocessing, and feature selection. Furthermore, optimisation methods tend to be used for fine-tuning and improving the accuracy of the models (see Section 9.3). The flood susceptibility maps are produced by the categorisation of the predicted probabilities using natural breaks or equal intervals, e.g., three classes (high, medium, and low) or five classes (very low, low, medium, high, and very high).

Moreover, the studies are carried out at the provincial scale with areas of a few thousands of square meters [80,100], but there are also a few applications of the same methodology to small river basins [181] or considerably larger regions [82].

9.1. Flood Data Sources and Conditioning Factors

For the production of the models, two different datasets are necessary. The first one contains the past flood events in the study area and is used as the dependent variable, which is referred to as "flood inventory" in the literature. The second one combines the conditioning factors which are used to estimate the probability of future flood occurrences.

9.1.1. Flood Inventory

Previous flood events can be retrieved from different sources depending on the spatial and temporal characteristics of the event. In general, the sources are official archives from local authorities, field surveys, or they are derived from aerial photos or satellite imagery.

The sources tend to be combined to perform the delineation of flooded areas or sample the flooded points, e.g., combining historical data sources, fieldwork, the perception of residents, and Google Earth (GE) [100]. GE is used to delineate the areas and then validate them with GNSS field surveys [182]; the Modified Normalised Difference Water Index (MNDWI) derived from Sentinel-2 imagery is used to extract the inundated areas and then validate with field surveys and reports [183]; and Sentinel-1 imagery and GE [184] use information from social media posts [102].

Many studies directly rely on official reports and field surveys for the collection of the flood inventory database [98,185,186]. The coverage of fine-scale areas tends to be limited to ad hoc field surveys [102].

After the flooded areas or points are collected, they are aggregated to a grid of a specific spatial resolution consistent with the other relevant data. According to most of the reviewed articles, a number of 'flooded' points is randomly selected from the grid to obtain the training data. As the problem is approached as a binary classification, the previous samples are complemented by adding the 'non-flooded' locations. These can be randomly selected from areas where floods are unlikely, e.g., high-elevation locations that have not been previously flooded, locations far from water bodies, etc. The selection of actual non-flooded points is very relevant to avoid negative sampling errors (false negatives) which affect the results of the classification [104].

To maintain the balance between the two classes, the general rule is to select an equal number of flooded and non-flooded points which are split into two distinct sets for the

training and validation of the models. Commonly used split ratios are 70%/30% and 80%/20%, respectively.

9.1.2. Flood Conditioning Factors

The conditioning factors that influence flood events include topographic, geological, precipitation, and land cover data. The topographic features (e.g., elevation, slope, topographic wetness index) are computed from the digital terrain model of the study area, which can be retrieved from local authorities datasets [80] or from global products, e.g., ASTER GDEM [184]. Geological data such as lithology and soil type are important as they influence the permeability of the terrain. Geological maps are usually provided by local authorities or research institutes [80]. In flood susceptibility models, precipitation data are key and are generally collected from ground stations and tend to be aggregated, e.g., total annual precipitation or maximum precipitation in a day [100]. Land cover data can be obtained from local authorities datasets, Corine Land Cover in Europe, or by processing satellite imagery (e.g., Landsat-8 or Sentinel-2) [184]. To use the different data sources as input in the machine and deep learning models, all of the datasets must be resampled at the same spatial resolution, which typically ranges between 5 and 30 m in the reviewed articles.

9.2. Flood Data Preprocessing

The independent data may be preprocessed to be prepared as input for the feature selection or classification algorithms, or to remove differences across the factors that could cause biases in the models. The most common transformation is the categorisation of all datasets into discrete classes, which can be determined with methods such as natural breaks, equal intervals, or manual thresholds [80]. In some cases [182], the features are categorised by scaling their values in the 0–1 range to account for the different scales of the data, which could lead the models to overweight or underweight certain features. The precipitation data from meteorological stations must be interpolated to create a grid. The interpolation methods employed in the reviewed paper are inverse distance weighting [183], spline interpolation [187], and kriging [184,188]. The latter usually performs better when the number of available stations is low, which is a common problem when dealing with meteorological variables, e.g., only four stations were used across an area of over 2000 km² [100].

9.3. Insights into the Flood ML/DL Modelling Techniques

Ensembles (ENSs), neural networks (NNs), support vector machines (SVMs), decision trees (DTs), and linear regressions (LRs) were the mostly used classes (see Section 4) in the task of flood susceptibility mapping. Figure 16 shows an overview of the statistics per class alongside the specific models, with a visible label if implemented in more than nine articles. Random forest (RF), artificial neural network (ANN), and support vector machine (SVM) were the most popular techniques used in 105, 75, and 65 articles, respectively. The result is expected given the suitability of the previous models for solving binary classification problems, usually guaranteeing high accuracy in the task.

Due to the nature of the problem, flood susceptibility modelling poses data balancing issues given that the number of flooded samples which are obtained from the flood inventory is less than the number of actual non-flooded samples in an area of interest. Non-flood events happen more frequently than flood events in almost any location of the world [189]. In most of the reviewed articles, the number of non-flooded samples used is the same as the number of flooded samples; however, some authors tackle the imbalance problem, e.g., testing the RF model with different flooded and non-flooded sampling ratios consisting of 1×, 10×, 25×, 50×, and 100× non-flooded samples with respect to the flooded ones [189]. The 1× sampling provided the highest AUC and recall and the 100× sampling provided the highest precision and f-1 score; however, the 50× sampling showed a middle point considering all the metrics, which is also a closer representation of the reality. Some of the techniques used in the articles to deal with the class imbalance problem are random

undersampling and random oversampling [190] and the synthetic minority oversampling technique [191].

The study of landscape patterns, i.e., the shape, composition, proportion, and configuration of the land use classes, is also addressed in the reviewed articles as a nature-based solution for flood resilience [104]. In this case, the authors used the computed landscape patterns in the AOI to compute their influence on the flood susceptibility. The results showed that the susceptibility to floods tend to be higher with the separation of forest patches, water connectivity, the growth of core urban areas, etc.

Despite the fact that flood susceptibility maps are produced with high accuracy, the pixel-by-pixel agreement between the maps produced of different highly accurate models is rarely studied [192]. This problem is referred to in the literature as uncertainty in spatial patterns or spatial agreement. Different approaches have been used to test the spatial agreement of flood susceptibility maps, e.g., Pearson's correlation and a subsequent linear regression to combine the maps produced by different models [192], the Kendall synergy coefficient to analyse the significance difference between the susceptibility indices of the produced maps, and computing the agreement between the models' results with the ground data [193].

Furthermore, various techniques have been used to improve the accuracy of the base ML/DL models. A simple method consists of determining the input weight of each conditioning factor with bivariate statistical methods such as frequency ratio [34] or with operations research, e.g., analytical hierarchy process [194]. Another valid approach is the optimisation of the algorithm's hyperparameters. The tuning can be easily performed with grid search or trial and error [80], or with more complex algorithms such as metaheuristics, e.g., swarm intelligence [84,195] and genetic algorithms [196].

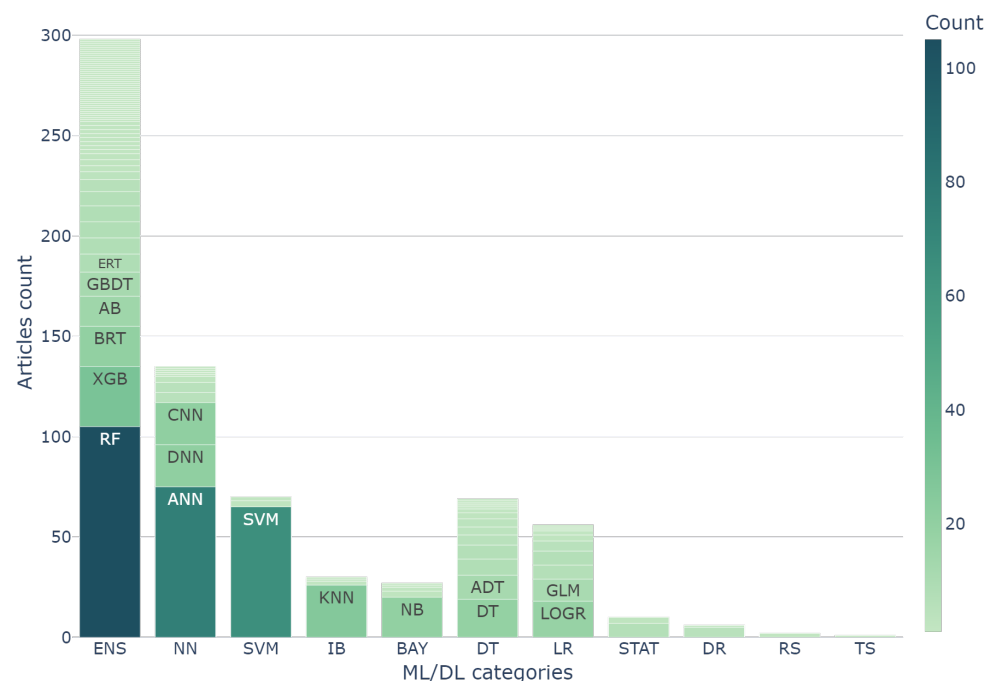


Figure 16. Flooding algorithm classes count (see Section 4). Complete dataset available in the repository (see Section 3). * ERT = Extremely Randomised Trees, GBDT = Gradient Boosted Decision Trees, AB = AdaBoost, BRT = Boosted Regression Tree, XGB = Extreme Gradient Boosting, RF = Random Forest, CNN = Convolutional Neural Network, DNN = Deep Neural Network, ANN = Artificial Neural Network, SVM = Support Vector Machine, KNN = K-Nearest-Neighbours, NB = Naive Bayes, ADT = Alternating Decision Trees, DT = Decision Trees, GLM = General Linear Model, LOGR = Logistic Regression, CB = CatBoost.

Finally, post hoc XAI methods have been used to understand the predictions of flood susceptibility models, e.g., SHAP for explaining convolutional neural networks (CNNs) [197], recurrent neural networks (RNNs) [198], boosting ensemble models [85,199], and extremely randomised trees (ERTs) [200].

Figure 17 presents the distribution per country of the flood susceptibility mapping articles based on the first author's affiliation. Iran is the leading country in this subject, followed by India, China, and Vietnam.

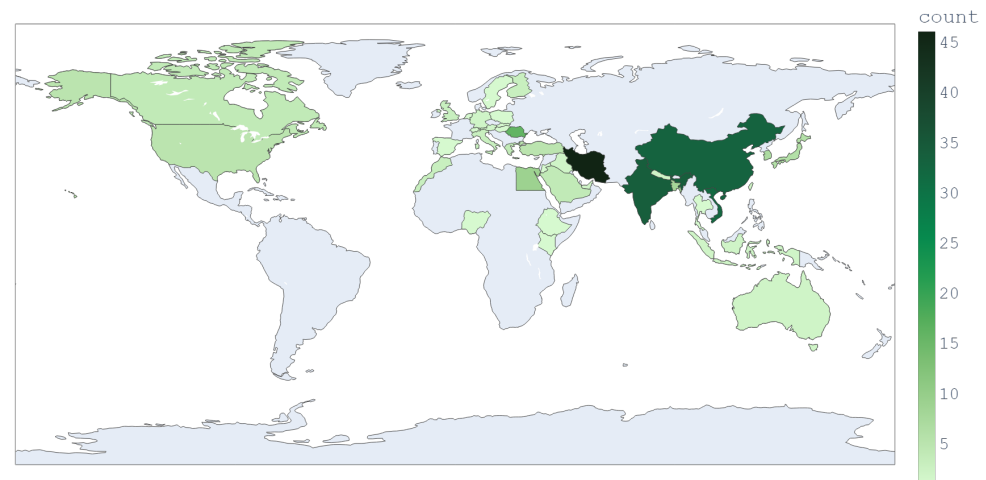


Figure 17. Flooding articles' distribution based on the first author's affiliation.

10. Landslide Susceptibility Modelling

Landslides are defined as the movement of a mass of rock, debris (coarse-grained), or earth (fine-grained) down a slope due to external forces (mainly gravity). The term encompasses five modes of slope movement: falls, topples, slides, spreads, and flows. Furthermore, these can be subdivided based on the type of geologic material: bedrock, debris, or earth [201]. The landslide types in this review include debris flow and rock flow [202–204]. It is important to highlight that, for the purpose of the literature review, no differentiation is made between the ML/DL techniques and the types of landslides in the review statistics for this kind of event.

The workflow followed to model landslide susceptibility in the reviewed articles is the same as that used to model flood susceptibility described in Section 9, with some differences in the algorithms and data used. The problem is still approached as a binary classification, with the classes landslide and non-landslide.

10.1. Landslide Data Sources and Conditioning Factors

Landslide susceptibility is modelled using previous landslide occurrences or events and the conditioning factors at the event locations. These datasets are the landslide inventory and the landslide conditioning factors, respectively, which will be detailed in the following subsections.

10.2. Landslide Inventory

The areas of landslide occurrences can be retrieved from historical records, satellite imagery, aerial photos, and field surveys, e.g., with manual interpretation from aerial photographs and DEM [205]. The inventory may be accompanied by metadata such as size and material type [202]. The landslide inventory is usually retrieved from official data archives or from an enhanced product coming from the combination of different sources, e.g., official data archives, GE, and field surveys [206,207], lidar, and aerial photos [205].

The landslide points can be sampled in different ways from the landslide records, e.g., by taking the landslide scarp centroid, the landslide body centroid, or samples from the landslide body or scarp [205]. The occurrence data generally consist of binary data (land-

slide and non-landslide) with a balanced number of samples; furthermore, the sampling ratio of training and testing points is 70%/30%.

10.3. Landslide Conditioning Factors

The mainly used conditioning factors (CFs) for landslide susceptibility modelling can be grouped into topographic, hydrological, geological, and meteorological CFs. The topographic factors are altitude, slope, aspect, and curvature, and can be derived from a DEM. The hydrological factors are the distance to streams or water bodies, drainage density, and the topographic wetness index. The geological factors are lithology, distance to faults, distance to geological boundaries, and soil type. Meteorological factors refer mostly to precipitation for this hazard. Additional conditioning factors not included in the previous groups are land use/land cover, Normalised Difference Vegetation Index (NDVI), Normalised Difference Built-up Index (NDBI), distance to roads, and points of interest. Furthermore, the landslides may be triggered by earthquakes, and therefore, seismic maps and their derived factors can be considered as well [205]. Additional CFs may be considered according to the details of the case study.

10.4. Landslide Data Preprocessing

In the reviewed articles, the conditioning factors (CFs) datasets are often categorised in discrete classes; afterwards, in several cases, the frequency ratio method is applied as a preliminary analysis [208] or to optimise the models [45,209]. The statistical approach allows the quantification of the relationship between each CF class and the areas affected by landslides. Moreover, it provides details on how the CFs' classes are distributed across the study area.

The datasets related to the different CFs have to be resampled to match a common spatial resolution which tends to be between 5 and 30 m. A different approach consists of computing slope units and conducting the CFs sampling and landslide modelling based on them [210–212].

10.5. Insights into the Landslide ML/DL Modelling Techniques

Ensembles (ENSs), neural networks (NNs), linear regressions (LRs), and support vector machines (SVMs) are the most used classes for landslide susceptibility modelling, as depicted in Figure 18. The figure also provides information about the specific ML/DL algorithms belonging to each class, with a visible label if used in more than 14 articles. The most used techniques are random forest (RF), support vector machine (SVM), logistic regression (LOGR), and artificial neural network (ANN), being implemented in 200, 151, 108, and 106 articles, respectively. The previously mentioned graph is limited to articles with at least one citation, at the time of retrieval, which corresponds to 402 articles out of 511 selected articles for this particular hazard. The popularity of SVMs to model these hazards is justified by the ability of the algorithms to perform well in high dimensional spaces and with class imbalances, compared to other ML methods [26].

The sampling of the dependent dataset (landslide/non-landslide) has a significant impact on the accuracy of the ML/DL models [210]. The selection of the non-landslide samples poses the problems of uncertainty and imbalance.

The uncertainty on the sampling of non-landslide points is posed by the influence of the distribution of non-landslide samples in the model, also considering that the sampled points are truly non-landslide. The non-landslide samples are usually randomly sampled in the area of interest (AOI) after excluding the landslide locations and considering certain conditioning factors (CFs) such as elevation. More advanced methods can be used to improve the quality of the non-landslide samples, e.g., self-organising-map and SVM [213], the Newmark sampling approach which, in the study of coseismic landslides, adds the slope earthquake displacement factor to narrow down the sampling area [214].

Imbalance refers to the misrepresentation of non-landslide samples with respect to the landslide samples, considering that the non-landslide locations are the majority class in

most of the AOIs. The data imbalance problem is assessed with undersampling techniques such as easy ensemble and balance cascade [215]. Another approach for avoiding imbalance is to use positive unlabelled learning [216,217].

Furthermore, the sampling of the landslide points is also relevant. A study tested the difference on model performances due to the use of different sampling types in the landslide areas, i.e., landslide scarp samples, landslide scarp centroids, landslide body samples, and landslide body centroids [205]. The results showed that using the landslide scarps samples produced a more accurate model. Another study tackled the generation of new landslide samples with generative adversarial networks (GANs), improving the performance of the models [18].

The spatial heterogeneity of landslide locations is assessed in the reviewed articles. It can be handled with the GeoSOM method to cluster landslide locations based on their environmental characteristics [218]. Additionally, the spatial heterogeneity of a model can be tested using spatial cross-validation which consists of splitting the AOI into different subregions and analysing the response of the model considering the environmental characteristics of each subregion [211].

Using different models for each landslide type yields higher accuracy than modelling all landslide types together [202,219]. In a similar way, a study divided the AOI into geomorphologically different areas (based on slope, elevation, constituent materials, etc.) to model and assess the landslide susceptibility per area [220].

Explainable artificial intelligence (XAI) is used to understand the dependence of the landslide susceptibility models on the CFs. Post hoc XAI models include the usage of the SHAP method, e.g., for global and individual prediction explanations of NN models [211,221], XGB [222], LightGBM [220]. Integrated XAI models include the generalised additive models with structured interactions (GAMI-net) method [223].

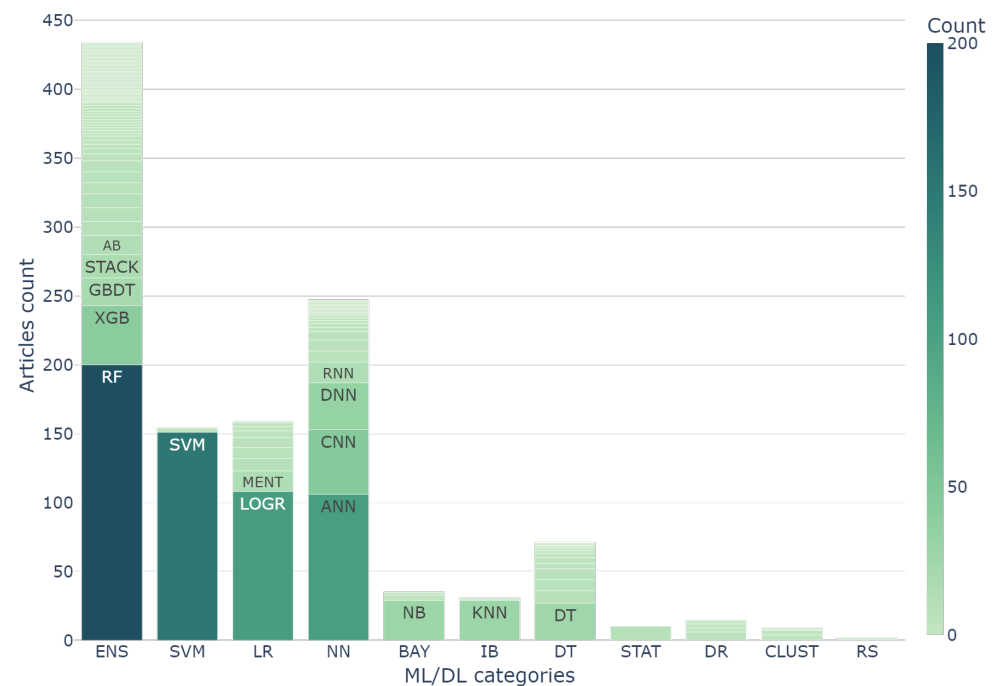


Figure 18. Landslide algorithm classes count (see Section 4). Complete dataset available in the repository (see Section 3). * AB = AdaBoost, STACK = stack of multiple models, GBDT = Gradient Boosted Decision Trees, XGB = Extreme Gradient Boosting, RF = Random Forest, SVM = Support Vector Machine, MENT = Maximum Entropy, LOGR = Logistic Regression, RNN = Recurrent Neural Network, DNN = Deep Neural Network, CNN = Convolutional Neural Network, ANN = Artificial Neural Network, NB = Naive Bayes, KNN = K-Nearest Neighbours, DT = Decision Trees.

The generalisation of models is also studied. It consists of training a model in a specific region and using it to produce the landslide susceptibility map in other regions with different environmental characteristics. A study tested the generalisation of the XGB model and used the SHAP method to explain the difference in the CFs' contributions to the model per region [224]. Another study followed an unsupervised transferable representation learning approach to improve the generalisation performance of a regressor [225]. The generalisation of a model can be further improved with transfer learning. This is key for assessing landslide susceptibility in areas with few landslide samples [226]. For example, an AdaBoost model could be improved with transferred samples from an additional region [226], or more advanced studies could implement transfer-learning strategies such as case-based reasoning and domain adaptation to generalised additive models [227].

The spatial agreement of susceptibility maps produced of different models is assessed in few articles. The agreement between the susceptibility maps produced by different highly accurate models has been tested with the correlation coefficient, ranging from 0.69 to 0.85; afterwards, a combined map was developed with linear regression to integrate the results of multiple models [54]. McNemar's test is also used to evaluate whether the misclassifications between two models are statistically significant [220].

Figure 19 depicts the spatial distribution of landslide susceptibility studies, showing that China is the leading country on the subject, in the frame of this review, followed by India, Vietnam, and Iran.

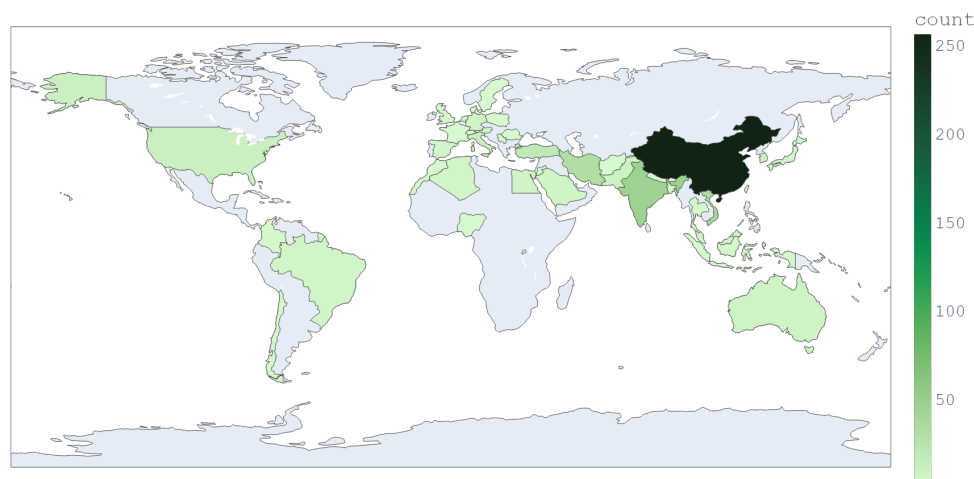


Figure 19. Landslide articles' distribution based on the first author's affiliation.

11. Discussion

Modelling the susceptibility of hazardous events with machine learning (ML) and deep learning (DL) techniques has proven to be a strong approach. However, there are some considerations which arose after analysing the techniques and methodologies utilised in the selected articles. The discussion covers three main topics: the first is the importance of the quality of the training data given the data-driven nature of ML/DL; the second is the model generalisation and potential use of transfer learning; finally, the third regards the spatial agreement of susceptibility maps.

11.1. The Importance of the Quality of Training Data

The quality of training data is key for a good model. In the current context, the quality of the data can be commented on through three aspects. The first is the temporal and spatial variability of the hazardous event occurrence data and its sampling; the second is the precise association in time of conditioning factors (CFs) and the occurrences of hazardous events; and the third is data preprocessing.

The temporal and spatial variability or the heterogeneity of data is important for the production of a robust model which is able to generalise in unseen locations. The

sampling of the training data also has an influence on the modelling results. Let us recall that the susceptibility modelling of hazardous events is based on the positive and negative occurrences of such events, usually approached as a binary classification problem. The binary classification tends to be associated with the data imbalance problem. This is often the case in flood and landslide susceptibility modelling. The majority of the selected articles ignore this issue by selecting the same number of positive and negative samples, without considering the fact that the actual ratio of occurrence vs. non-occurrence locations in the studied areas is not equal [189]. The data imbalance problem is assessed in few articles with undersampling and oversampling techniques [190,191], and positive unlabelled learning [216,217]. New approaches for tackling the class imbalance problem for susceptibility mapping could be analysed in future studies.

The precise association in time of CFs and the occurrences of hazardous events can be referred to as time-consistent modelling. Although the occurrences of hazardous events happen at specific times, the CFs at the event locations tend to remain constant in time, which mostly applies to the flood and landslide studies. Event occurrences in an 11-year time span are associated with the same time-invariant CFs [45]. It may be the case that the time span of the event occurrences is not mentioned and still the CFs are constant, e.g., a total of 243 landslides events were associated with a single Normalised Difference Vegetation Index (NDVI) map [228]. Additionally, areas closer to recent event occurrences tend to be more susceptible than areas closer to older events, even if susceptibility maps tend to be considered time-invariant [229]. The susceptibility maps of hazardous events can be further improved if the hazardous event occurrences are associated with CFs which are close to the event in time, of course, depending on the availability of temporal CF data.

The data provided to the model can be enhanced in the preprocessing phase. Many studies emphasise the feature selection and the analysis of the correlation between CFs and occurrence data. The feature selection methods range from statistical correlations to ML approaches, recently including explainable artificial intelligence.

A final comment on this discussion is about the actual lack of data. ML and mostly DL models need a reasonable amount of data to build accurate models. Some options for dealing with data scarcity are the generation of synthetic data [18] and transfer learning. The second is discussed in the following subsection.

11.2. Model Generalisation and Transfer Learning

Model generalisation (MG) and transfer learning (TL) are closely related concepts. MG refers to the ability to accurately predict susceptibility in unseen locations, even if the conditioning factors (CFs) of the unseen locations are different from those used for training. For example, spatial cross-validation, i.e., training a model in a region and testing it in another region with different environmental characteristics such as slope or temperature [211]. MG is referred to as the 'spatial heterogeneity' of a model in the articles related to landslide susceptibility mapping [218]. On the other hand, TL consists of starting from a previously trained robust model and fine-tuning it with the data of a new region. TL is particularly useful for the assessment of areas with scarce data [226] and applies to both spatial and temporal modelling.

MG was tested in few articles, e.g., by analysing the behaviour of landslide susceptibility in regions with different environmental characteristics [224]. However, the majority of the articles tested the susceptibility models in a single area of interest. TL has been implemented in the studies related to the hazards of air pollution, floods, and landslides. The results of using a TL model were, mostly, more accurate with respect to producing a model from scratch with only the scarce available data.

The combination of MG and TL supports the production of robust and spatially heterogeneous susceptibility models which are not only custom-made for a specific region. This can potentially support the production of susceptibility maps in regions of the world with scarce training data.

11.3. Spatial Agreement of Susceptibility Maps

A significant concern identified is the spatial agreement of susceptibility maps produced by different models. Although accuracy metrics are commonly used to assess the performance of susceptibility models, these metrics do not always correlate with the spatial consistency of the maps produced by different models. The susceptibility maps can exhibit significant spatial differences despite being the product of models with similar accuracies.

Few studies tackled this issue by testing the spatial correlation of the susceptibility maps. The approaches of the revised studies were two. The first uses the McNemar's test and Kendall synergy for evaluating the statistical significance of the misclassification of the maps [192,220]. The second is to test the spatial correlation of the produced maps and their eventual combination with linear regression [54].

Nevertheless, in the majority of the articles, there is a lack of assessment of the spatial agreement among the produced susceptibility maps.

12. Conclusions

There is an increasing trend in the usage of ML/DL techniques for modelling the susceptibility of the four studied hazards which are air pollution, urban heat islands (UHIs), floods, and landslides. This reflects how this type of approach is able to assess complex environmental relationships, usually yielding better results than other approaches.

The majority of the reviewed articles utilised supervised learning techniques. Therefore, for providing a holistic overview, an ad hoc classification of these techniques has been provided, underlining their utilisation across the different hazards. The proposed classification includes ensembles, neural networks, decision trees, support vector machines, and regression techniques, among others.

From the literature, a general hazard susceptibility modelling approach was consolidated into a schema. This workflow encompasses the entire process, beginning with the collection and preprocessing of previous hazardous event occurrences and their associated conditioning factors (CFs), followed by feature selection and correlation analysis, and posterior modelling and validation. Additionally, the optional spatial agreement and model/results interpretation is commented. Furthermore, the data sources of CFs common to multiple hazards were listed if available at the global or continental level, providing the list of the main sources for meteorological data, digital elevation models, land cover, and surface indices.

The concept of multi-hazard susceptibility, which involves the combination of individual susceptibility maps, was explored in the studies. Most articles utilised a univariate combination approach, resulting in categorical maps. However, one study employed a fuzzy inference approach to incorporate experts' opinion, producing a continuous susceptibility map. This suggests the potential value of exploring alternative methods for more comprehensive multi-hazard assessments.

Each hazard was discussed in a dedicated section, covering aspects such as its definition, particular data, modelling approaches, and related considered articles. Each of these sections also provided an overview of the usage of different techniques, and the spatial distribution of studies based on the first author's affiliation. Moreover, we highlighted novel methodologies such as Explainable Artificial Intelligence (XAI), transfer learning, model generalisation, and methods to address data imbalance and spatial agreement.

For the UHI, flood, and landslide hazards, random forest, support vector machines, and artificial neural networks emerged as the overall most frequently used techniques, primarily because they represent the state-of-the-art in the field. Despite this, there is a growing interest in exploring novel ML/DL approaches, particularly neural network and ensemble models with hyperparameter optimisation techniques. On the other hand, air pollution was the hazard which included the most computationally complex approaches due to the nature of the problems modelled, i.e., time series or spatiotemporal forecasting and spatial estimation.

Furthermore, XAI plays a very important role in the modelling of hazard susceptibility with ML/DL techniques due to the mostly black-box nature of this technique. XAI helps one understand the model and the model predictions, providing insights into the relations between a prediction and the contributions of CFs.

To summarise, ML/DL models are a useful tool for the production of hazard susceptibility maps. The aim is to contribute to the improvement of the resilience of cities by providing a means to understand the behaviour and distribution of hazardous events. Ergo, a holistic hazard susceptibility modelling shall not only be accurate but also spatially validated and rely on XAI to provide human-readable explanations of the results.

Author Contributions: Conceptualisation, M.A.B., D.C., A.d.J.P.V. and A.F.; methodology, A.d.J.P.V. and A.F.; software, A.d.J.P.V.; validation, M.A.B. and D.C.; formal analysis, A.d.J.P.V.; investigation, A.d.J.P.V. and A.F.; resources, A.d.J.P.V.; data curation, A.d.J.P.V. and A.F.; writing—original draft preparation, A.d.J.P.V. and A.F.; writing—review and editing, D.C. and A.d.J.P.V.; visualisation, A.d.J.P.V.; supervision, D.C. and M.A.B.; project administration, M.A.B.; funding acquisition, M.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This review has been partially funded by the HARMONIA project 2020-LC-CLA-2018-2019-2020, GA No.101003517.

Data Availability Statement: The data are available in Zenodo [20]. It contains tables with the articles result of the search, the selected articles, the legends of the ML/DL techniques, and the Jupyter notebooks to visualise the interactive version of the plots included in this article.

Acknowledgments: The authors acknowledge the contributions of the anonymous reviewers which helped enhance the background and discussion sections and the manuscript overall.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
AOD	Aerosol Optical Depth
AOI	Area of Interest
AQI	Air Quality Index
AUC	Area under the ROC Curve
BAY	Bayesian
BLSTM	Bidirectional Long Short-Term Memory
CF	Conditioning Factor
CLUST	Clustering
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
DL	Deep Learning
DNN	Deep Neural Network
DR	Dimensionality Reduction
DTs	Decision Trees
ENS	Ensemble
ERA5	ECMWF (European Centre for Medium-Range Weather Forecasts) Atmospheric Reanalysis V5
ERT	Extremely Randomised Trees
GAMI	Generalised Additive Models with structured Interaction
GAN	Generative Adversarial Networks
GCN	Graph Convolutional Network
GE	Google Earth
GPR	Gaussian Process Regression
GRU	Gated Recurrent Unit
IB	Instance-based

IoT	Internet of Things
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model-agnostic Explanations
LR	Linear Regression
LST	Land Surface Temperature
LSTM	Long Short-Term Memory
LULC	Land Use–Land Cover
MG	Model Generalisation
MI	Mutual Information
ML	Machine Learning
MNDWI	Modified Normalised Difference Water Index
MSPA	Morphological Spatial Pattern Analysis
NDBI	Normalised Difference Built-up Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NN	Neural Networks
OSM	OpenStreetMap
REG	Regression
RF	Random Forest
RFR	Random Forest Regression
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RUL	Rule-based Systems
SDG	Sustainable Development Goal
SHAP	Shapley Additive Explanation values
STAT	Statistical
SVM	Support Vector Machine
SVR	Support Vector Regression
TL	Transfer Learning
TS	Time Series
UHI	Urban Heat Island
UHII	Urban Heat Island Intensity
UTEI	Urban Thermal Environment Index
UTFVI	Urban Thermal Field Variance Index
XAI	Explainable Artificial Intelligence
XGB	Extreme Gradient Boost

Appendix A. Algorithms Classification

Table A1. Ad hoc classification of algorithms.

Class	Algorithms
Bayesian	Bayesian General Linear Model (BGLM), Bayesian Logistic Regression (BLOGR), Bayesian Moving Average (BMA), Bayesian Network (BN), Bayesian Regression (BR), Incremental Learning Bayesian Network (ILBN), Naive Bayes (NB).
Clustering	DBScan (DBS), Gaussian Mixture Models (GMM), Hierarchical Clustering (HC), K-Means (KM), Positive Unlabelled Bagging (PUB).
Dimensionality Reduction	Flexible Discriminant Analysis (FLDA), Functional Discriminant Analysis (FUDA), Mixture Discriminant Analysis (MIDA), Multivariate Discriminant Analysis (MDA), Partial Least Square Regression (PLSR), Quadratic Discriminant Analysis (QDA).
Decision Trees	Alternating Decision Trees (ADT), Best First Decision Trees (BFDT), C4.5 Decision Tree (C45DT), C5.0 Decision Trees (C50DT), CHi-squared Automatic Interaction Detection (CHAID), Classification and Regression Trees (CART), Decision Table Classifier (DTC), Decision Trees (DT), Functional Trees (FT), Gradient Boosting Regression Trees (GBRT), Hoeffding Trees (HT), J48 Decision Tree (JDT), Logistic Model Tree (LMT), M5 model trees (M5P), Naive Bayes Trees (NBT), Partial Decision Tree (PDT), Reduced Error Pruning Decision Tree (REPDT), Reduced Error Pruning Tree (REPT), Regression Trees (RT), Spatiotemporal Decision Trees (STDT).
Instance-based	Hyperpipes (HP), K-Nearest-Neighbour (KNN), K-Star (KS), Locally Weighted Linear Regression (LWLR), Subspace K-Nearest-Neighbour (SSKNN), Voting Feature Intervals (VFI).
Rule-based Systems	Cubist (CUB), Genetic Algorithm Rule-Set Production (GARP), Rough Set (RS).
Regression	Additive Regression (ADR), Complete Subset Regression (CRS), Elastic Regression (ER), Elasticnet Classifier (ENC), Gaussian Process Regression (GPR), General Linear Model (GLM), Generalised Additive Model (GAM), Kernel Logistic Regression (KLOGR), Kernel-based Regularised Least Squares (KRLS), LASSO Regression (LASSO), Land Use Regression (LUR), Linear Regression (LR), Logistic Regression (LOGR), Maximum Entropy (MENT), Multivariate Adaptive Regression Spline (MARS), Principal Component Regression (PCR), Ridge Regression (RR), Volterra (VOL).
Support Vector Machine	Least Square Support Vector Machine (LSSVM), Relevance Vector Machine (RVM), Spatiotemporal Support Vector Machine (STSVM), Support Vector Machine (SVM), Support Vector Regression (SVR).
Statistical	Dynamic Conditional Pareto (DCP), Frequency Ratio (FR), Functional Data Analysis (FDA), Response Surface Model (RSM).
Time Series	Autoregressive model (AR), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average (ARMA), Generalised Autoregressive Conditional Heteroskedasticity (GARCH), Moving Average (MA), Prophet Forecasting Model (PFM), Vector Autoregression (VAR).

Table A1. Cont.

Class	Algorithms
Neural Networks	Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Autoencoders (AE), Autocorrelation Error Informer (AEI), Back Propagation Neural Network (BPNN), Bayesian Neural Network (BNN), Bottleneck Transformer Network (BTN), Convolutional Neural Network (CNN), Deep Autoencoder (DAE), Deep Belief Network (DBN), Deep Convolutional Neural Network (DCNN), Deep Neural Network (DNN), Dense Convolutional Networks (DCN), Echo State Neural Network (ESNN), Elman Network (EN), Extreme Learning Adaptive Neuro-Fuzzy Inference System (ELANFIS), Extreme Learning Machine (ELM), Full Connection Layers (FCL), Fuzzy Neural Network (FNN), Gated Recurrent Unit (GRU), General Regression Neural Network (GRNN), Generalised Additive Models with Structured Interactions (GAMI), Generative Adversarial Network (GAN), Graph Convolutional Network (GCN), Graph Long Short-Term Memory (GLSTM), Graph Neural Network (GNN), Hierarchical Neural Network (HNN), Long Short-Term Memory (LSTM), Model Averaged Neural Network (MANN), Multi-Graph Convolution (MGC), Multi-Layer Perceptron (MLP), Multi-Step Ahead Neural Network (MSANN), Passive Aggressive Classifier (PA), Quasi-Recurrent Neural Networks (QRNN), Radial Basis Function (RBF), Recurrent Neural Network (RNN), Residual Convolutional Neural Network (RESCNN), Residual Neural Network (RESNN), Restricted Boltzman Machine (RBM), Self-Adaptive Deep Neural Network (SADNN), Self-Organising Map (SOM), Sequence2Sequence RNN (SEQ2SEQ), Simple Recurrent Unit (SRU), Spatiotemporal Backpropagation Neural Network (STBPNN), Spatiotemporal Dynamic Advection (STDA), Spatiotemporal Extreme Learning Machine (STELM), Spatiotemporal Gated Recurrent Unit (STGRU), Spatiotemporal Informer (STI), Spatiotemporal Long Short Term Memory (STLSTM), Spatiotemporal Multi-Layer Perceptron (STMLP), Spatiotemporal Neural Networks (STNN), Spatiotemporal Orthogonal Cube (STOC), Spatiotemporal Transformer (STT), Temporal Convolutional Neural Network (TCNN), Temporal Difference-based Graph Transformer Networks (TDGTN), Transformer Neural Network (TNN), Variational AutoEncoder (VAE), Vision Transformer (ViT), Wavelet Neural Network (WNN).

Table A1. Cont.

Class	Algorithms
Ensemble	<p>Adaboost (AB): AB Decision Table (ABDTA), AB Alternating Decision Trees (ABADT), AB Backpropagation Neural Network (ABBPNN), AB Classifier (ABC), AB Credal Decision Trees (ABCDT), AB Decision Trees (ABDT), AB Extreme Learning Machine (ABELM), AB Hyperpipes (ABHP), AB Partial Decision Tree (ABPDT), AB Reduced Error Pruning Decision Tree (ABREPDT), AB Rough Set (ABRS), AB Voting Feature Intervals (ABVFI), AB Credal Decision Tree (ABCDT), Real AB Hyperpipes (RABHP), Real AB J48 Decision Tree (RABJDT), Real AB Reduced Error Pruning Tree (RABREPDT); Adaptively Resample and Combine (ARC), Attribute Selected Classifier Artificial Neural Network (ASCANN); Bagging (Ba): Ba Credal Decision Tree (BCDT), Ba Forest Penalising Attribute (BFPA), Ba Functional Trees (BFT), Ba Artificial Neural Network (BANN), Ba Best First Decision Trees (BBFDT), Ba C4.5 Decision Tree (BC45DT), Ba Credal Decision Trees (BCDT), Ba Decision Table (BDTA), Bagging Decision Trees (BDT), Ba Deep Neural Network (BDNN), Ba Functional Tree (BFT), Ba Gaussian Process (BGP), Ba Hyperpipes (BHP), Ba K-Nearest-Neighbour (BKNN), Ba Logistic Model Tree (BLMT), Ba M5 model trees (BM5P), Ba Partial Decision Tree (BPDT), Ba Random Forest (BRF), Ba Random Subspace Naive Bayes Trees (BRSSNBT), Ba Random Trees (BART), Ba Reduced Error Pruning Decision Tree (BREPDT), Ba Rough Set (BRS), Ba Sequential Minimal Optimisation (BSMO), Ba Support Vector Machine (BSVM); Bayesian Additive Decision Trees (BADT); Boosted (Bo): Bo Classification Tree (BCT), Bo Decision Trees (BODT), Bo Generalised Additive Model (BGAM), Bo Generalised Linear Model (BOGLM), Bo Regression Tree (BRT), Bo Regression Trees (BRT), Bo Artificial Neural Network (BOANN), Bo C4.5 Decision Tree (BOC45DT), Bo Decision Trees (BODT), Bo Logistic Model Tree (BOLMT), Bo Support Vector Machine (BOSVM), Explainable Bo Machine (EBM), Extreme Gradient Bo (XGB), Extreme Gradient Bo Regression (XGBR), Generalised Bo Model (GBM), Gradient Bo Decision Trees (GBDT), Gradient Bo Regression Trees (GBRT), Gradient Bo Classifier (GBC), Gradient Bo Extreme Learning Machine (GBELM), Gradient Bo Machine (GBM), Histogram-based Gradient Bo (HGB), Light Gradient Bo Machine (LGBM), Logit Bo (LOGB), Natural Gradient Bo (NGB), Stochastic Gradient Bo (SGB); Bootstrap aggregation (BA), Cascade Generalisation Artificial Neural Network (CGANN), Cascade Random Forest (CRF), Catboost (CB), Conditional Inference Random Forest (CIRF), Cost Sensitive Forest (CSF), Credal Decision Trees (CDT); Dagging (Da): Da Alternating Decision Trees (DADT), Da Artificial Neural Network (DANN), Da Best First Decision Trees (DABFDT), Da Credal Decision Trees (DCDT), Da Decision Trees (DDT), Da Functional Tree (DFT), Da HyperPipes (DHP), Da M5 model trees (DM5P), Da Partial Decision Tree (DPDT), Da Reduced Error Pruning Decision Tree (DAREPDT); Decorate (De): De Best First Decision Trees (DBFDT), De Credal Decision Tree (DECDT), De Decision Trees (DDT), De Forest Penalising Attribute (DFPA), De HyperPipes (DEHP), De Reduced Error Pruning Decision Tree (DREPDT); Deep Forest (DF), Deepboost (DB), Extremely Randomised Trees (ERT), Forest Penalising Attribute (FPA), Geographical Random Forest (GRF), Isolated Forest (IF); Multiboost (MB): MB Adaboost Credal Decision Trees (MBCDT), MB J48 Decision Tree (MBJDT), MB Alternating Decision Trees (MBADT), MB Artificial Neural Network (MBANN), MB Decision Trees (MBDT), MB Voting Feature Intervals (MBVFI), MB Reduced Error Pruning Decision Tree (MBREPDT); Multiple Kernel Learning (MKL), Rotation forest (ROF): ROF Credal Decision Tree (ROFCDT), ROF Random Forest (ROFRF), ROF Functional Tree (ROFFT), ROF Reduced Error Pruning Decision Tree (ROFREPDT); Random Forest (RF): RF Logistic Model Tree (RFLMT), RF Machine (RFM), RF Regression (RFR); Random Naive Bayes (RNB), Random Subspace (RSS): RSS Alternating Decision Trees (RSSADT), RSS Artificial Neural Network (RSSANN), RSS Best First Decision Trees (RSSBFDT), RSS C4.5 Decision Tree (RSSC45DT), RSS Credal Decision Tree (RSSCDT), RSS Decision Trees (RSSDT), RSS Functional Trees (RSSFT), RSS J48 Decision Tree (RSSJDT), RSS Partial Decision Trees (RSSPDT), RSS Random Forest (RSSRF), RSS Reduced Error Pruning Decision Tree (RSSREPDT); Random Trees Classifier (RTC), Spatiotemporal Extreme Gradient Boost (STXGB), Spatiotemporal Extremely Randomised Trees (STERT), Spatiotemporal Gradient Boosted Decision Tree (STGBDT), Spatiotemporal Light Gradient Boosting Machine (STLGBM), Spatiotemporal Random Forest (STRF), Stacking multiple models (STACK), Subspace Discriminant (SSD), SysFor (SF), Ultraboost (UB).</p>

References

1. IPCC. IPCC, 2023: Summary for Policymakers. In *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2023; pp. 1–34. [CrossRef]
2. Zampieri, M.; Russo, S.; di Sabatino, S.; Michetti, M.; Scoccimarro, E.; Gualdi, S. Global assessment of heat wave magnitudes from 1901 to 2010 and implications for the river discharge of the Alps. *Sci. Total Environ.* **2016**, *571*, 1330–1339. [CrossRef] [PubMed]
3. Munich Re. *Hurricanes, Cold Waves, Tornadoes: Weather Disasters in USA Dominate Natural Disaster Losses in 2021*; Munich Re: Munich, Germany, 2022.
4. Asian Development Bank. *Moving from Risk to Resilience: Sustainable Urban Development in the Pacific*; Asian Development Bank: Mandaluyong, Philippines, 2013.
5. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S.; Catani, F.; Eeckhout, M.V.D.; Mavrouli, O.; Agliardi, F.; et al. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [CrossRef]
6. Schmidt, J.; Matcham, I.; Reese, S.; King, A.; Bell, R.; Henderson, R.; Smart, G.; Cousins, J.; Smith, W.; Heron, D. Quantitative multi-risk analysis for natural hazards: A framework for multi-risk modelling. *Nat. Hazards* **2011**, *58*, 1169–1192. [CrossRef]
7. Zennaro, F.; Furlan, E.; Simeoni, C.; Torresan, S.; Aslan, S.; Critto, A.; Marcomini, A. Exploring machine learning potential for climate change risk assessment. *Earth-Sci. Rev.* **2021**, *220*, 103752. [CrossRef]
8. Lee, J.G.; Kang, M. Geospatial Big Data: Challenges and Opportunities. *Big Data Res.* **2015**, *2*, 74–81. [CrossRef]
9. Mehmood, K.; Bao, Y.; Saifullah; Cheng, W.; Khan, M.A.; Siddique, N.; Abrar, M.M.; Soban, A.; Fahad, S.; Naidu, R. Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives. *J. Clean. Prod.* **2022**, *379*, 134656. [CrossRef]
10. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng. Geol.* **2008**, *102*, 85–98. [CrossRef]
11. Bentivoglio, R.; Isufi, E.; Jonkman, S.N.; Taormina, R. Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 4345–4378. [CrossRef]
12. Nhu, V.H.; Zandi, D.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Al-Ansari, N.; Singh, S.K.; Dou, J.; Nguyen, H. Comparison of Support Vector Machine, Bayesian Logistic Regression, and Alternating Decision Tree Algorithms for Shallow Landslide Susceptibility Mapping along a Mountainous Road in the West of Iran. *Appl. Sci.* **2020**, *10*, 5047. [CrossRef]
13. Formetta, G.; Rago, V.; Capparelli, G.; Rigon, R.; Muto, F.; Versace, P. Integrated Physically based System for Modeling Landslide Susceptibility. *Procedia Earth Planet. Sci.* **2014**, *9*, 74–82. [CrossRef]
14. Feng, L.; Guo, M.; Wang, W.; Chen, Y.; Shi, Q.; Guo, W.; Lou, Y.; Kang, H.; Chen, Z.; Zhu, Y. Comparative Analysis of Machine Learning Methods and a Physical Model for Shallow Landslide Risk Modeling. *Sustainability* **2022**, *15*, 6. [CrossRef]
15. Sarker, I. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 420. [CrossRef] [PubMed]
16. Sarker, I. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]
17. Tehrani, F.S.; Calvello, M.; Liu, Z.; Zhang, L.; Lacasse, S. Machine learning and landslide studies: Recent advances and applications. *Nat. Hazards* **2022**, *114*, 1197–1245. [CrossRef]
18. Al-Najjar, H.A.; Pradhan, B. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. *Geosci. Front.* **2021**, *12*, 625–637. [CrossRef]
19. Guo, Q.; Ren, M.; Wu, S.; Sun, Y.; Wang, J.; Wang, Q.; Ma, Y.; Song, X.; Chen, Y. Applications of artificial intelligence in the field of air pollution: A bibliometric analysis. *Front. Public Health* **2022**, *10*, 2972. [CrossRef]
20. Pugliese-Viloria, A. Hazard Susceptibility Mapping with Machine and Deep Learning: A Literature Review—Data and Software, V1.0.0; Zenodo: 2024. Available online: <https://zenodo.org/records/13386422> (accessed on 29 July 2024).
21. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **2021**, *31*, 685–695. [CrossRef]
22. Chauhan, N.K.; Singh, K. A review on conventional machine learning vs. deep learning. In Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018, Greater Noida, Uttar Pradesh, India, 28–29 September 2018; pp. 347–352. [CrossRef]
23. Patgiri, R. Taxonomy of Big Data: A Survey. *arXiv* **2018**, arXiv:1808.08474. [CrossRef]
24. Preeti, K.S.; Dhankar, A. A review on Machine Learning Techniques. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 778–782. [CrossRef]
25. Grossi, E.; Buscema, M. Introduction to artificial neural networks. *Eur. J. Gastroenterol. Hepatol.* **2008**, *19*, 1046–1054. [CrossRef]
26. Nikparvar, B.; Thill, J.C. Machine Learning of Spatial Data. *ISPRS Int. J. -Geo-Inf.* **2021**, *10*, 600. [CrossRef]
27. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Deep Air Quality Forecasting Using Hybrid Deep Learning Framework. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 2412–2424. [CrossRef]
28. Yan, R.; Liao, J.; Yang, J.; Sun, W.; Nong, M.; Li, F. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* **2021**, *169*, 114513. [CrossRef]
29. Li, T.; Hua, M.; Wu, X. A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5). *IEEE Access* **2020**, *8*, 26933–26940. [CrossRef]

30. Vaswani, A.; Brain, G.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
31. Yu, H.; Pei, W.; Zhang, J.; Chen, G. Landslide Susceptibility Mapping and Driving Mechanisms in a Vulnerable Region Based on Multiple Machine Learning Models. *Remote Sens.* **2023**, *15*, 1886. [[CrossRef](#)]
32. Doreswamy; Vastrad, C.M. Performance Analysis Of Regularized Linear Regression Models For Oxazolines And Oxazoles Derivative Descriptor Dataset. *arXiv* **2013**, arXiv:1312.2789. [[CrossRef](#)]
33. Peng, J.; Lee, K.; Ingersoll, G. An Introduction to Logistic Regression Analysis and Reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [[CrossRef](#)]
34. Ali, S.A.; Parvin, F.; Pham, Q.B.; Vojtek, M.; Vojteková, J.; Costache, R.; Linh, N.T.T.; Nguyen, H.Q.; Ahmad, A.; Ghorbani, M.A. 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 Topľa basin, Slovakia. *Ecol. Indic.* **2020**, *117*, 106620. [[CrossRef](#)]
35. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
36. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1984.
37. Quinlan, J. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Francisco, CA, USA, 1993.
38. Elomaa, T.; Kääriäinen, M. An Analysis of Reduced Error Pruning. *arXiv* **2011**, arXiv:1106.0668. [[CrossRef](#)]
39. Hearst, M.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
40. Cristianini, N.; Scholkopf, B. Support Vector Machines and Kernel Methods: The New Generation of Learning Machines. *AI Mag.* **2002**, *23*, 31. [[CrossRef](#)]
41. Costache, R.; Hong, H.; Pham, Q.B. Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Sci. Total Environ.* **2020**, *711*, 134514. [[CrossRef](#)] [[PubMed](#)]
42. 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](#)]
43. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
45. Bui, D.T.; Tsangaratos, P.; Nguyen, V.T.; Liem, N.V.; Trinh, P.T. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *CATENA* **2020**, *188*, 104426. [[CrossRef](#)]
46. Freund, Y.; Schapire, R.E. A Short Introduction to Boosting. In Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI'99, Stockholm, Sweden, 31 July–6 August 1999; Volume 2, pp. 1401–1406.
47. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
48. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
49. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 16, New York, NY, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
50. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
51. Yao, J.; Zhang, X.; Luo, W.; Liu, C.; Ren, L. Applications of Stacking/Blending ensemble learning approaches for evaluating flash flood susceptibility. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102932. [[CrossRef](#)]
52. Aha, W.; Kibler, D.; Albert, M. Instance-Based Learning Algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [[CrossRef](#)]
53. Cunningham, P.; Delany, S. k-Nearest neighbour classifiers. *Mult. Classif. Syst.* **2007**, *54*, 1–25. [[CrossRef](#)]
54. Adnan, M.S.G.; Rahman, M.S.; Ahmed, N.; Ahmed, B.; Rabbi, M.F.; Rahman, R.M. Improving Spatial Agreement in Machine Learning-Based Landslide Susceptibility Mapping. *Remote Sens.* **2020**, *12*, 3347. [[CrossRef](#)]
55. Zhu, J.; Chen, J.; Hu, W. Big Learning with Bayesian Methods. *arXiv* **2014**, arXiv:1411.6370. [[CrossRef](#)]
56. Vikramkumar; Vijaykumar, B.; Trilochan. Bayes and Naive Bayes Classifier. *arXiv* **2014**, arXiv:1404.0933. [[CrossRef](#)]
57. Liu, Z.; Zhu, Z.; Gao, J.; Xu, C. Forecast Methods for Time Series Data: A Survey. *IEEE Access* **2021**, *9*, 91896–91912. [[CrossRef](#)]
58. Taylor, S.; Letham, B. Forecasting at Scale. *Am. Stat.* **2017**, *72*, 37–45. [[CrossRef](#)]
59. Jia, W.; Sun, M.; Lian, J.; Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **2022**, *8*, 2663–2693. [[CrossRef](#)]
60. Wang, Y.; Sun, D.; Wen, H.; Zhang, H.; Zhang, F. Comparison of Random Forest Model and Frequency Ratio Model for Landslide Susceptibility Mapping (LSM) in Yunyang County (Chongqing, China). *Int. J. Environ. Res. Public Health* **2020**, *17*, 4206. [[CrossRef](#)] [[PubMed](#)]
61. Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [[CrossRef](#)]
62. Lin, J.; Sreng, C.; Oare, E.; Batarseh, F.A. NeuralFlood: An AI-driven flood susceptibility index. *Front. Water* **2023**, *5*. [[CrossRef](#)]
63. Huang, F.; Cao, Z.; Jiang, S.H.; Zhou, C.; Huang, J.; Guo, Z. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. *Landslides* **2020**, *17*, 2919–2930. [[CrossRef](#)]
64. Mao, Y.; Mwakapesa, D.S.; Li, Y.; Xu, K.; Nanekaran, Y.A.; Zhang, M. Assessment of landslide susceptibility using DBSCAN-AHD and LD-EV methods. *J. Mt. Sci.* **2022**, *19*, 184–197. [[CrossRef](#)]

65. Sette, S.; Boullart, L. An implementation of genetic algorithms for rule based machine learning. *Eng. Appl. Artif. Intell.* **2000**, *13*, 381–390. [[CrossRef](#)]
66. Li, Z.; Yim, S.H.L.; Ho, K.F. High temporal resolution prediction of street-level PM2.5 and NOx concentrations using machine learning approach. *J. Clean. Prod.* **2020**, *268*, 121975. [[CrossRef](#)]
67. Bellman, R. Dynamic programming. *Science* **1966**, *153*, 34–37. [[CrossRef](#)]
68. Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G.; et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* **2023**, *55*, 1–33. [[CrossRef](#)]
69. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
70. Venkatesh, B.; Anuradha, J. A review of Feature Selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [[CrossRef](#)]
71. Dosilovic, F.K.; Brcic, M.; Hlupic, N. Explainable artificial intelligence: A survey. In Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018—Proceedings, Opatija, Croatia, 21–25 May 2018; pp. 210–215. [[CrossRef](#)]
72. Du, W.; Chen, L.; Wang, H.; Shan, Z.; Zhou, Z.; Li, W.; Wang, Y. Deciphering urban traffic impacts on air quality by deep learning and emission inventory. *J. Environ. Sci.* **2023**, *124*, 745–757. [[CrossRef](#)] [[PubMed](#)]
73. Henckaerts, R.; Antonio, K.; Côté, M.P. When stakes are high: Balancing accuracy and transparency with Model-Agnostic Interpretable Data-driven suRRogates. *Expert Syst. Appl.* **2022**, *202*, 117230. [[CrossRef](#)]
74. Shao, Y.; Zhao, W.; Liu, R.; Yang, J.; Liu, M.; Fang, W.; Hu, L.; Adams, M.; Bi, J.; Ma, Z. Estimation of daily NO₂ with explainable machine learning model in China, 2007–2020. *Atmos. Environ.* **2023**, *314*, 120111. [[CrossRef](#)]
75. Lalonde, M.; Oudin, L.; Bastin, S. Urban effects on precipitation: Do the diversity of research strategies and urban characteristics preclude general conclusions? *Urban Clim.* **2023**, *51*, 101605. [[CrossRef](#)]
76. Nabavi, S.O.; Nolscher, A.C.; Samimi, C.; Thomas, C.; Haimberger, L.; Luers, J.; Held, A. Site-scale modeling of surface ozone in Northern Bavaria using machine learning algorithms, regional dynamic models, and a hybrid model. *Environ. Pollut.* **2021**, *268*, 115736. [[CrossRef](#)]
77. Sun, D.; Gu, Q.; Wen, H.; Xu, J.; Zhang, Y.; Shi, S.; Xue, M.; Zhou, X. Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization. *Gondwana Res.* **2022**, *123*, 89–106. [[CrossRef](#)]
78. Martens, D.; Vanthienen, J.; Verbeke, W.; Baesens, B. Performance of classification models from a user perspective. *Decis. Support Syst.* **2011**, *51*, 782–793. [[CrossRef](#)]
79. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; Ghamisi, P. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Total Environ.* **2020**, *701*, 134474. [[CrossRef](#)]
80. Bui, D.T.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2020**, *701*, 134413. [[CrossRef](#)]
81. Lei, T.M.T.; Ng, S.C.W.; Siu, S.W.I. Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau. *Sustainability* **2023**, *15*, 5341. [[CrossRef](#)]
82. Pourghasemi, H.R.; Kariminejad, N.; Amiri, M.; Edalat, M.; Zarafshar, M.; Blaschke, T.; Cerda, A. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Sci. Rep.* **2020**, *10*, 3203. [[CrossRef](#)] [[PubMed](#)]
83. Ghosh, S.; Saha, S.; Bera, B. Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. *Nat. Hazards Res.* **2022**, *2*, 363–374. [[CrossRef](#)]
84. Bui, Q.T.; Nguyen, Q.H.; Nguyen, X.L.; Pham, V.D.; Nguyen, H.D.; Pham, V.M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *J. Hydrol.* **2020**, *581*, 124379. [[CrossRef](#)]
85. Aydin, H.E.; Iban, M.C. Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. *Nat. Hazards* **2023**, *116*, 2957–2991. [[CrossRef](#)]
86. Ozdemir, H.; Koçyiğit, M.B.; Akay, D. Flood susceptibility mapping with ensemble machine learning: A case of Eastern Mediterranean basin, Türkiye. *Stoch. Environ. Res. Risk Assess.* **2023**, *37*, 4273–4290. [[CrossRef](#)]
87. Karakas, G.; Kocaman, S.; Gokceoglu, C. A Hybrid Multi-Hazard Susceptibility Assessment Model for a Basin in Elazig Province, Türkiye. *Int. J. Disaster Risk Sci.* **2023**, *14*, 326–341. [[CrossRef](#)]
88. Pourghasemi, H.R.; Gayen, A.; Edalat, M.; Zarafshar, M.; Tiefenbacher, J.P. Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management? *Geosci. Front.* **2020**, *11*, 1203–1217. [[CrossRef](#)]
89. Yousefi, S.; Pourghasemi, H.R.; Emami, S.N.; Pouyan, S.; Eskandari, S.; Tiefenbacher, J.P. A machine learning framework for multi-hazards modeling and mapping in a mountainous area. *Sci. Rep.* **2020**, *10*, 12144. [[CrossRef](#)]
90. Oukawa, G.Y.; Krecl, P.; Targino, A.C. Fine-scale modeling of the urban heat island: A comparison of multiple linear regression and random forest approaches. *Sci. Total Environ.* **2022**, *815*, 152836. [[CrossRef](#)]
91. Zhang, X.; Huang, T.; Gulakhmadov, A.; Song, Y.; Gu, X.; Zeng, J.; Huang, S.; Nam, W.H.; Chen, N.; Niyogi, D. Deep Learning-Based 500 m Spatio-Temporally Continuous Air Temperature Generation by Fusing Multi-Source Data. *Remote Sens.* **2022**, *14*, 3536. [[CrossRef](#)]
92. Vulova, S.; Meier, F.; Fenner, D.; Nouri, H.; Kleinschmit, B. Summer Nights in Berlin, Germany: Modeling Air Temperature Spatially With Remote Sensing, Crowdsourced Weather Data, and Machine Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5074–5087. [[CrossRef](#)]

93. Chen, S.; Xu, Z.; Wang, X.; Zhang, C. Ambient air pollutants concentration prediction during the COVID-19: A method based on transfer learning. *Knowl.-Based Syst.* **2022**, *258*, 109996. [[CrossRef](#)] [[PubMed](#)]
94. Oliveira, A.; Lopes, A.; Niza, S.; Soares, A. An urban energy balance-guided machine learning approach for synthetic nocturnal surface Urban Heat Island prediction: A heatwave event in Naples. *Sci. Total Environ.* **2022**, *805*, 150130. [[CrossRef](#)] [[PubMed](#)]
95. dos Santos, R.S. Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102066. [[CrossRef](#)]
96. Mohammad, P.; Goswami, A.; Chauhan, S.; Nayak, S. Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. *Urban Clim.* **2022**, *42*, 101116. [[CrossRef](#)]
97. Wang, Y.; Liang, Z.; Ding, J.; Shen, J.; Wei, F.; Li, S. Prediction of Urban Thermal Environment Based on Multi-Dimensional Nature and Urban Form Factors. *Atmosphere* **2022**, *13*, 1493. [[CrossRef](#)]
98. Pham, B.T.; Avand, M.; Janizadeh, S.; Phong, T.V.; Al-Ansari, N.; Ho, L.S.; Das, S.; Le, H.V.; Amini, A.; Bozchaloei, S.K.; et al. GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment. *Water* **2020**, *12*, 683. [[CrossRef](#)]
99. Nhu, V.H.; Ngo, P.T.T.; Pham, T.D.; Dou, J.; Song, X.; Hoang, N.D.; Tran, D.A.; Cao, D.P.; Aydilek, İ.B.; Amiri, M.; et al. A New Hybrid Firefly-PSO Optimized Random Subspace Tree Intelligence for Torrential Rainfall-Induced Flash Flood Susceptible Mapping. *Remote Sens.* **2020**, *12*, 2688. [[CrossRef](#)]
100. Islam, A.R.M.T.; Talukdar, S.; Mahato, S.; Kundu, S.; Eibek, K.U.; Pham, Q.B.; Kuriqi, A.; Linh, N.T.T. Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci. Front.* **2021**, *12*, 101075. [[CrossRef](#)]
101. AlDousari, A.E.; Kafy, A.A.; Saha, M.; Fattah, M.A.; Almulhim, A.I.; Abdullah-Al-Faisal; Rakib, A.A.; Jahir, D.M.A.; Rahaman, Z.A.; Bakshi, A.; et al. Modelling the impacts of land use/land cover changing pattern on urban thermal characteristics in Kuwait. *Sustain. Cities Soc.* **2022**, *86*, 104107. [[CrossRef](#)]
102. Li, Y.; Osei, F.B.; Hu, T.; Stein, A. Urban flood susceptibility mapping based on social media data in Chengdu city, China. *Sustain. Cities Soc.* **2023**, *88*, 104307. [[CrossRef](#)]
103. He, W.; Zhang, S.; Meng, H.; Han, J.; Zhou, G.; Song, H.; Zhou, S.; Zheng, H. Full-Coverage PM2.5 Mapping and Variation Assessment during the Three-Year Blue-Sky Action Plan Based on a Daily Adaptive Modeling Approach. *Remote Sens.* **2022**, *14*, 3571. [[CrossRef](#)]
104. Luo, Z.; Tian, J.; Zeng, J.; Pilla, F. Resilient landscape pattern for reducing coastal flood susceptibility. *Sci. Total Environ.* **2023**, *856*, 159087. [[CrossRef](#)]
105. Shen, Y.; de Hoogh, K.; Schmitz, O.; Clinton, N.; Tuxen-Bettman, K.; Brandt, J.; Christensen, J.H.; Frohn, L.M.; Geels, C.; Karssenber, D.; et al. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. *Environ. Int.* **2022**, *168*, 107485. [[CrossRef](#)] [[PubMed](#)]
106. Vavassori, A.; Vilorio, A.D.J.P.; Brovelli, M.A. Using open data to reveal factors of urban susceptibility to natural hazards and human-made hazards: Case of Milan and Sofia. *GeoScape* **2022**, *16*, 93–107. [[CrossRef](#)]
107. Environmental Pollution Centers. *What Is Air Pollution*; Environmental Pollution Centers: Nairobi, Kenya, 2022.
108. Khan, M.A.; Kim, H.; Park, H. Leveraging Machine Learning for Fault-Tolerant Air Pollutants Monitoring for a Smart City Design. *Electronics* **2022**, *11*, 3122. [[CrossRef](#)]
109. European Environment Agency. *European Air Quality Index*; European Environment Agency: Copenhagen, Denmark, 2021.
110. El-Magd, S.A.; Soliman, G.; Morsy, M.; Kharbish, S. Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing. *Int. J. Environ. Sci. Technol.* **2022**, *20*, 6103–6116. [[CrossRef](#)]
111. Shogrkhodaei, S.Z.; Razavi-Termeh, S.V.; Fathnia, A. Spatio-temporal modeling of PM2.5 risk mapping using three machine learning algorithms. *Environ. Pollut.* **2021**, *289*, 117859. [[CrossRef](#)]
112. Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; et al. Improved 1km resolution PM2.5 estimates across China using enhanced space-time extremely randomized trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289. [[CrossRef](#)]
113. Zhao, C.; Wang, Q.; Ban, J.; Liu, Z.; Zhang, Y.; Ma, R.; Li, S.; Li, T. Estimating the daily PM2.5 concentration in the Beijing-Tianjin-Hebei region using a random forest model with a $0.01^\circ \times 0.01^\circ$ spatial resolution. *Environ. Int.* **2020**, *134*, 105297. [[CrossRef](#)] [[PubMed](#)]
114. Long, S.; Wei, X.; Zhang, F.; Zhang, R.; Xu, J.; Wu, K.; Li, Q.; Li, W. Estimating daily ground-level NO₂ concentrations over China based on TROPOMI observations and machine learning approach. *Atmos. Environ.* **2022**, *289*, 119310. [[CrossRef](#)]
115. Zhang, C.; Liu, C.; Li, B.; Zhao, F.; Zhao, C. Spatiotemporal neural network for estimating surface NO₂ concentrations over north China and their human health impact. *Environ. Pollut.* **2022**, *307*, 119510. [[CrossRef](#)]
116. Di, Q.; Amini, H.; Shi, L.; Kloog, I.; Silvern, R.; Kelly, J.; Sabath, M.B.; Choirat, C.; Koutrakis, P.; Lyapustin, A.; et al. Assessing NO₂ Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. *Environ. Sci. Technol.* **2020**, *54*, 1372–1384. [[CrossRef](#)] [[PubMed](#)]
117. Chu, W.; Zhang, C.; Zhao, Y.; Li, R.; Wu, P. Spatiotemporally Continuous Reconstruction of Retrieved PM2.5 Data Using an Autogeoi-Stacking Model in the Beijing-Tianjin-Hebei Region, China. *Remote Sens.* **2022**, *14*, 4432. [[CrossRef](#)]
118. Just, A.C.; Arfer, K.B.; Rush, J.; Dorman, M.; Shtein, A.; Lyapustin, A.; Kloog, I. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. *Atmos. Environ.* **2020**, *239*, 117649. [[CrossRef](#)]

119. Yu, Y.; Li, H.; Sun, S.; Li, Y. PM_{2.5} concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy. *Sustain. Cities Soc.* **2022**, *85*, 104049. [[CrossRef](#)]
120. Liu, X.; Li, C.; Liu, D.; Grieneisen, M.L.; Yang, F.; Chen, C.; Zhan, Y. Hybrid deep learning models for mapping surface NO₂ across China: One complicated model, many simple models, or many complicated models? *Atmos. Res.* **2022**, *278*, 106339. [[CrossRef](#)]
121. Moursi, A.S.; El-Fishawy, N.; Djahel, S.; Shouman, M.A. An IoT enabled system for enhanced air quality monitoring and prediction on the edge. *Complex Intell. Syst.* **2021**, *7*, 2923–2947. [[CrossRef](#)]
122. Ram, R.S.; Venkatachalam, K.; Masud, M.; Abouhawwash, M. Air Pollution Prediction Using Dual Graph Convolution LSTM Technique. *Intell. Autom. Soft Comput.* **2022**, *33*, 1639–1652. [[CrossRef](#)]
123. Kang, Y.; Choi, H.; Im, J.; Park, S.; Shin, M.; Song, C.K.; Kim, S. Estimation of surface-level NO₂ and O₃ concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* **2021**, *288*, 117711. [[CrossRef](#)] [[PubMed](#)]
124. Liu, R.; Ma, Z.; Liu, Y.; Shao, Y.; Zhao, W.; Bi, J. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environ. Int.* **2020**, *142*, 105823. [[CrossRef](#)]
125. Huang, C.; Sun, K.; Hu, J.; Xue, T.; Xu, H.; Wang, M. Estimating 2013–2019 NO₂ exposure with high spatiotemporal resolution in China using an ensemble model. *Environ. Pollut.* **2022**, *292*, 118285. [[CrossRef](#)]
126. Xiao, Q.; Zheng, Y.; Geng, G.; Chen, C.; Huang, X.; Che, H.; Zhang, X.; He, K.; Zhang, Q. Separating emission and meteorological contributions to long-term PM_{2.5} trends over eastern China during 2000–2018. *Atmos. Chem. Phys.* **2021**, *21*, 9475–9496. [[CrossRef](#)]
127. Heidari, A.A.; Akhoondzadeh, M.; Chen, H. A Wavelet PM_{2.5} Prediction System Using Optimized Kernel Extreme Learning with Boruta-XGBoost Feature Selection. *Mathematics* **2022**, *10*, 3566. [[CrossRef](#)]
128. Zhang, B.; Zou, G.; Qin, D.; Ni, Q.; Mao, H.; Li, M. RCL-Learning: ResNet and convolutional long short-term memory-based spatiotemporal air pollutant concentration prediction model. *Expert Syst. Appl.* **2022**, *207*, 118017. [[CrossRef](#)]
129. Zhang, Y.; Zhang, R.; Ma, Q.; Wang, Y.; Wang, Q.; Huang, Z.; Huang, L. A feature selection and multi-model fusion-based approach of predicting air quality. *ISA Trans.* **2020**, *100*, 210–220. [[CrossRef](#)]
130. Huang, C.; Hu, J.; Xue, T.; Xu, H.; Wang, M. High-Resolution Spatiotemporal Modeling for Ambient PM_{2.5} Exposure Assessment in China from 2013 to 2019. *Environ. Sci. Technol.* **2021**, *55*, 2152–2162. [[CrossRef](#)]
131. Wang, D.; Wang, H.W.; Li, C.; Lu, K.F.; Peng, Z.R.; Zhao, J.; Fu, Q.; Pan, J. Roadside Air Quality Forecasting in Shanghai with a Novel Sequence-to-Sequence Model. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9471. [[CrossRef](#)]
132. Faraji, M.; Nadi, S.; Ghaffarparand, O.; Homayoni, S.; Downey, K. An integrated 3D CNN-GRU deep learning method for short-term prediction of PM_{2.5} concentration in urban environment. *Sci. Total Environ.* **2022**, *834*, 155324. [[CrossRef](#)]
133. Arowosegbe, O.O.; Roosli, M.; Kunzli, N.; Saucy, A.; Adebayo-Ojo, T.C.; Schwartz, J.; Kebalepile, M.; Jeebhay, M.F.; Dalvie, M.A.; de Hoogh, K. Ensemble averaging using remote sensing data to model spatiotemporal PM₁₀ concentrations in sparsely monitored South Africa. *Environ. Pollut.* **2022**, *310*, 119883. [[CrossRef](#)] [[PubMed](#)]
134. Liu, H.; Chen, C. Prediction of outdoor PM_{2.5} concentrations based on a three-stage hybrid neural network model. *Atmos. Pollut. Res.* **2020**, *11*, 469–481. [[CrossRef](#)]
135. Mao, Y.; Mwakapesa, D.S.; Wang, G.; Nanekaran, Y.; Zhang, M. Landslide susceptibility modelling based on AHC-OLID clustering algorithm. *Adv. Space Res.* **2021**, *68*, 301–316. [[CrossRef](#)]
136. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **2020**, *699*, 133561. [[CrossRef](#)] [[PubMed](#)]
137. Zhang, Z.; Zeng, Y.; Yan, K. A hybrid deep learning technology for PM_{2.5} air quality forecasting. *Environ. Sci. Pollut. Res.* **2021**, *28*, 39409–39422. [[CrossRef](#)]
138. Li, S.; Xie, G.; Ren, J.; Guo, L.; Yang, Y.; Xu, X. Urban PM_{2.5} Concentration Prediction via Attention-Based CNN-LSTM. *Appl. Sci.* **2020**, *10*, 1953. [[CrossRef](#)]
139. Gilik, A.; Ogrenci, A.S.; Ozmen, A. Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environ. Sci. Pollut. Res.* **2022**, *29*, 11920–11938. [[CrossRef](#)]
140. Sarkar, N.; Gupta, R.; Keserwani, P.K.; Govil, M.C. Air Quality Index prediction using an effective hybrid deep learning model. *Environ. Pollut.* **2022**, *315*, 120404. [[CrossRef](#)]
141. Ehteram, M.; Ahmed, A.N.; Khozani, Z.S.; El-Shafie, A. Graph convolutional network—Long short term memory neural network—multi layer perceptron- Gaussian process regression model: A new deep learning model for predicting ozone concentration. *Atmos. Pollut. Res.* **2023**, *14*, 101766. [[CrossRef](#)]
142. Guo, C.; Liu, G.; Chen, C.H. Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 1–13. [[CrossRef](#)]
143. Li, D.; Liu, J.; Zhao, Y. Prediction of Multi-Site PM_{2.5} Concentrations in Beijing Using CNN-Bi LSTM with CBAM. *Atmosphere* **2022**, *13*, 1719. [[CrossRef](#)]
144. Heydari, A.; Nezhad, M.M.; Garcia, D.A.; Keynia, F.; Santoli, L.D. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technol. Environ. Policy* **2022**, *24*, 607–621. [[CrossRef](#)]
145. Ma, J.; Ding, Y.; Cheng, J.C.; Jiang, F.; Gan, V.J.; Xu, Z. A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM_{2.5} prediction. *Sustain. Cities Soc.* **2020**, *60*, 102237. [[CrossRef](#)]
146. Du, P.; Wang, J.; Hao, Y.; Niu, T.; Yang, W. A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM_{2.5} and PM₁₀ forecasting. *Appl. Soft Comput.* **2020**, *96*, 106620. [[CrossRef](#)]

147. Castelli, M.; Clemente, F.M.; Popovic, A.; Silva, S.; Vanneschi, L. A Machine Learning Approach to Predict Air Quality in California. *Complexity* **2020**, *2020*, 1–23. [[CrossRef](#)]
148. Yu, M.; Masrur, A.; Blaszczyk-Boxe, C. Predicting hourly PM_{2.5} concentrations in wildfire-prone areas using a SpatioTemporal Transformer model. *Sci. Total. Environ.* **2023**, *860*, 160446. [[CrossRef](#)]
149. Cai, K.; Zhang, X.; Zhang, M.; Ge, Q.; Li, S.; Qiao, B.; Liu, Y. Improving air pollutant prediction in Henan Province, China, by enhancing the concentration prediction accuracy using autocorrelation errors and an Informer deep learning model. *Sustain. Environ. Res.* **2023**, *33*, 13. [[CrossRef](#)]
150. Yan, X.; Zang, Z.; Jiang, Y.; Shi, W.; Guo, Y.; Li, D.; Zhao, C.; Husi, L. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM_{2.5}. *Environ. Pollut.* **2021**, *273*, 116459. [[CrossRef](#)]
151. Guo, B.; Zhang, D.; Pei, L.; Su, Y.; Wang, X.; Bian, Y.; Zhang, D.; Yao, W.; Zhou, Z.; Guo, L. Estimating PM_{2.5} concentrations via random forest method using satellite, auxiliary, and ground-level station dataset at multiple temporal scales across China in 2017. *Sci. Total. Environ.* **2021**, *778*, 146288. [[CrossRef](#)]
152. Ren, X.; Mi, Z.; Georgopoulos, P.G. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. *Environ. Int.* **2020**, *142*, 105827. [[CrossRef](#)]
153. Wong, P.Y.; Lee, H.Y.; Chen, Y.C.; Zeng, Y.T.; Chern, Y.R.; Chen, N.T.; Lung, S.C.C.; Su, H.J.; Wu, C.D. Using a land use regression model with machine learning to estimate ground level PM_{2.5}. *Environ. Pollut.* **2021**, *277*, 116846. [[CrossRef](#)] [[PubMed](#)]
154. Ma, J.; Li, Z.; Cheng, J.C.; Ding, Y.; Lin, C.; Xu, Z. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total. Environ.* **2020**, *705*, 135771. [[CrossRef](#)]
155. Ma, W.; Yuan, Z.; Lau, A.K.; Wang, L.; Liao, C.; Zhang, Y. Optimized neural network for daily-scale ozone prediction based on transfer learning. *Sci. Total. Environ.* **2022**, *827*, 154279. [[CrossRef](#)]
156. Parthiban, S.; Amudha, P.; Sivakumari, S.P. Exploitation of Advanced Deep Learning Methods and Feature Modeling for Air Quality Prediction. *Rev. Dintelligence Artif.* **2022**, *36*, 959–967. [[CrossRef](#)]
157. Zhang, L.; Na, J.; Zhu, J.; Shi, Z.; Zou, C.; Yang, L. Spatiotemporal causal convolutional network for forecasting hourly PM_{2.5} concentrations in Beijing, China. *Comput. Geosci.* **2021**, *155*, 104869. [[CrossRef](#)]
158. Wang, Z.; Hu, B.; Huang, B.; Ma, Z.; Biswas, A.; Jiang, Y.; Shi, Z. Predicting annual PM_{2.5} in mainland China from 2014 to 2020 using multi temporal satellite product: An improved deep learning approach with spatial generalization ability. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 141–158. [[CrossRef](#)]
159. Kim, M.; Brunner, D.; Kuhlmann, G. Importance of satellite observations for high-resolution mapping of near-surface NO₂ by machine learning. *Remote Sens. Environ.* **2021**, *264*, 112573. [[CrossRef](#)]
160. Ren, X.; Mi, Z.; Cai, T.; Nolte, C.G.; Georgopoulos, P.G. Flexible Bayesian Ensemble Machine Learning Framework for Predicting Local Ozone Concentrations. *Environ. Sci. Technol.* **2022**, *56*, 3871–3883. [[CrossRef](#)]
161. United States Environmental Protection Agency. *Learn About Heat Islands*; US EPA: Washington, DC, USA, 2022.
162. Kafy, A.A.; Abdullah-Al-Faisal; Rahman, M.S.; Islam, M.; Rakib, A.A.; Islam, M.A.; Khan, M.H.H.; Sikdar, M.S.; Sarker, M.H.S.; Mawa, J.; et al. Prediction of seasonal urban thermal field variance index using machine learning algorithms in Cumilla, Bangladesh. *Sustain. Cities Soc.* **2021**, *64*, 102542. [[CrossRef](#)]
163. Lin, J.; Qiu, S.; Tan, X.; Zhuang, Y. Measuring the relationship between morphological spatial pattern of green space and urban heat island using machine learning methods. *Build. Environ.* **2023**, *228*, 109910. [[CrossRef](#)]
164. Kafy, A.A.; Saha, M.; Abdullah-Al-Faisal; Rahaman, Z.A.; Rahman, M.T.; Liu, D.; Fattah, M.A.; Rakib, A.A.; AlDousari, A.E.; Rahaman, S.N.; et al. Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms. *Build. Environ.* **2022**, *217*, 109066. [[CrossRef](#)]
165. Jiang, J.; Zhou, Y.; Guo, X.; Qu, T. Calculation and Expression of the Urban Heat Island Indices Based on GeoSOT Grid. *Sustainability* **2022**, *14*, 2588. [[CrossRef](#)]
166. Oh, J.W.; Ngarambe, J.; Duhirwe, P.N.; Yun, G.Y.; Santamouris, M. Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea. *Sci. Rep.* **2020**, *10*, 3559. [[CrossRef](#)]
167. Jato-Espino, D.; Machado, C.; Roldán-Valcarce, A.; Moscardó, V. ArcUHI: A GIS add-in for automated modelling of the Urban Heat Island effect through machine learning. *Urban Clim.* **2022**, *44*, 101203. [[CrossRef](#)]
168. Lai, J.; Zhan, W.; Quan, J.; Bechtel, B.; Wang, K.; Zhou, J.; Huang, F.; Chakraborty, T.; Liu, Z.; Lee, X. Statistical estimation of next-day nighttime surface urban heat islands. *ISPRS J. Photogramm. Remote Sens.* **2021**, *176*, 182–195. [[CrossRef](#)]
169. Lan, T.; Peng, J.; Liu, Y.; Zhao, Y.; Dong, J.; Jiang, S.; Cheng, X.; Corcoran, J. The future of Chinas urban heat island effects: A machine learning based scenario analysis on climatic-socioeconomic policies. *Urban Clim.* **2023**, *49*, 101463. [[CrossRef](#)]
170. Choudhury, U.; Singh, S.K.; Kumar, A.; Meraj, G.; Kumar, P.; Kanga, S. Assessing Land Use/Land Cover Changes and Urban Heat Island Intensification: A Case Study of Kamrup Metropolitan District, Northeast India (2000–2032). *Earth* **2023**, *4*, 503–521. [[CrossRef](#)]
171. Abou, S.; Rasha, M. Investigating and mapping day-night urban heat island and its driving factors using Sentinel/MODIS data and Google Earth Engine. Case study: Greater Cairo, Egypt. *Urban Clim.* **2023**, *52*, 101729. [[CrossRef](#)]
172. Han, L.; Zhao, J.; Gao, Y.; Gu, Z. Prediction and evaluation of spatial distributions of ozone and urban heat island using a machine learning modified land use regression method. *Sustain. Cities Soc.* **2022**, *78*, 103643. [[CrossRef](#)]

173. Garzón, J.; Molina, I.; Velasco, J.; Calabia, A. A Remote Sensing Approach for Surface Urban Heat Island Modeling in a Tropical Colombian City Using Regression Analysis and Machine Learning Algorithms. *Remote Sens.* **2021**, *13*, 4256. [[CrossRef](#)]
174. Waleed, M.; Sajjad, M.; Acheampong, A.O.; Alam, M.T. Towards Sustainable and Livable Cities: Leveraging Remote Sensing, Machine Learning, and Geo-Information Modelling to Explore and Predict Thermal Field Variance in Response to Urban Growth. *Sustainability* **2023**, *15*, 1416. [[CrossRef](#)]
175. Soille, P.; Vogt, P. Morphological segmentation of binary patterns. *Pattern Recognit. Lett.* **2009**, *30*, 456–459. [[CrossRef](#)]
176. NOAA National Severe Storms Laboratory. *Severe Weather 101: Flood Basics*; NOAA National Severe Storms Laboratory: Norman, OK, USA, 2022.
177. Park, S.J.; Lee, D.K. Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. *Environ. Res. Lett.* **2020**, *15*, 094052. [[CrossRef](#)]
178. El-Magd, S.A.A.; Maged, A.; Farhat, H.I. Hybrid-based Bayesian algorithm and hydrologic indices for flash flood vulnerability assessment in coastal regions: Machine learning, risk prediction, and environmental impact. *Environ. Sci. Pollut. Res.* **2022**, *29*, 57345–57356. [[CrossRef](#)]
179. Costache, R.; Arabameri, A.; Costache, I.; Crăciun, A.; Pham, B.T. New Machine Learning Ensemble for Flood Susceptibility Estimation. *Water Resour. Manag.* **2022**, *36*, 4765–4783. [[CrossRef](#)]
180. Parvin, F.; Ali, S.A.; Calka, B.; Bielecka, E.; Linh, N.T.T.; Pham, Q.B. Urban flood vulnerability assessment in a densely urbanized city using multi-factor analysis and machine learning algorithms. *Theor. Appl. Climatol.* **2022**, *149*, 639–659. [[CrossRef](#)]
181. Yariyan, P.; Janizadeh, S.; Phong, T.V.; Nguyen, H.D.; Costache, R.; Le, H.V.; Pham, B.T.; Pradhan, B.; Tiefenbacher, J.P. Improvement of Best First Decision Trees Using Bagging and Dagging Ensembles for Flood Probability Mapping. *Water Resour. Manag.* **2020**, *34*, 3037–3053. [[CrossRef](#)]
182. Costache, R.; Pham, Q.B.; Sharifi, E.; Linh, N.T.T.; Abba, S.; Vojtek, M.; Vojteková, J.; Nhi, P.T.T.; Khoi, D.N. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. *Remote Sens.* **2019**, *12*, 106. [[CrossRef](#)]
183. Hosseini, F.S.; Choubin, B.; Mosavi, A.; Nabipour, N.; Shamshirband, S.; Darabi, H.; Haghighi, A.T. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Sci. Total. Environ.* **2020**, *711*, 135161. [[CrossRef](#)]
184. Shahabi, H.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J.J.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S.; et al. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. *Remote Sens.* **2020**, *12*, 266. [[CrossRef](#)]
185. Ma, M.; Zhao, G.; He, B.; Li, Q.; Dong, H.; Wang, S.; Wang, Z. XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **2021**, *598*, 126382. [[CrossRef](#)]
186. Dodangeh, E.; Choubin, B.; Eigdir, A.N.; Nabipour, N.; Panahi, M.; Shamshirband, S.; Mosavi, A. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci. Total. Environ.* **2020**, *705*, 135983. [[CrossRef](#)] [[PubMed](#)]
187. Costache, R.; Pham, Q.B.; Avand, M.; Linh, N.T.T.; Vojtek, M.; Vojteková, J.; Lee, S.; Khoi, D.N.; Nhi, P.T.T.; Dung, T.D. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. *J. Environ. Manag.* **2020**, *265*, 110485. [[CrossRef](#)] [[PubMed](#)]
188. Talukdar, S.; Ghose, B.; Shahfahad.; Salam, R.; Mahato, S.; Pham, Q.B.; Linh, N.T.T.; Costache, R.; Avand, M. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 2277–2300. [[CrossRef](#)]
189. Ekmekcioğlu, O.; Koc, K.; ozger, M.; Işık, Z. Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States. *J. Hydrol.* **2022**, *610*, 127877. [[CrossRef](#)]
190. Abu-Salih, B.; Wongthongtham, P.; Coutinho, K.; Qaddoura, R.; Alshaweesh, O.; Wedyan, M. The development of a road network flood risk detection model using optimised ensemble learning. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106081. [[CrossRef](#)]
191. Priscillia, S.; Schillaci, C.; Lipani, A. Flood susceptibility assessment using artificial neural networks in Indonesia. *Artif. Intell. Geosci.* **2021**, *2*, 215–222. [[CrossRef](#)]
192. Adnan, M.S.G.; Siam, Z.S.; Kabir, I.; Kabir, Z.; Ahmed, M.R.; Hassan, Q.K.; Rahman, R.M.; Dewan, A. A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. *J. Environ. Manag.* **2023**, *326*, 116813. [[CrossRef](#)]
193. Meliho, M.; Khattabi, A.; Asinyo, J. Spatial modeling of flood susceptibility using machine learning algorithms. *Arab. J. Geosci.* **2021**, *14*, 2243. [[CrossRef](#)]
194. Costache, R.; Bui, D.T. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci. Total. Environ.* **2020**, *712*, 136492. [[CrossRef](#)]
195. Al-Areeq, A.M.; Saleh, R.A.A.; Ghanim, A.A.J.; Ghaleb, M.; Al-Areeq, N.M.; Al-Wajih, E. Flood hazard assessment in Yemen using a novel hybrid approach of Grey Wolf and Levenberg Marquardt optimizers. *Geocarto Int.* **2023**, *38*, 2243884. [[CrossRef](#)]
196. Razavi-Termeh, S.V.; Sadeghi-Niaraki, A.; Seo, M.; Choi, S.M. Application of genetic algorithm in optimization parallel ensemble-based machine learning algorithms to flood susceptibility mapping using radar satellite imagery. *Sci. Total Environ.* **2023**, *873*, 162285. [[CrossRef](#)]

197. Pradhan, B.; Lee, S.; Dikshit, A.; Kim, H. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. *Geosci. Front.* **2023**, *14*, 101625. [[CrossRef](#)]
198. Liu, J.; Liu, K.; Wang, M. A Residual Neural Network Integrated with a Hydrological Model for Global Flood Susceptibility Mapping Based on Remote Sensing Datasets. *Remote Sens.* **2023**, *15*, 2447. [[CrossRef](#)]
199. Kaspi, M.; Kuleshov, Y. Flood Hazard Assessment in Australian Tropical Cyclone-Prone Regions. *Climate* **2023**, *11*, 229. [[CrossRef](#)]
200. omer Ekmekcioğlu.; Koc, K. Explainable step-wise binary classification for the susceptibility assessment of geo-hydrological hazards. *CATENA* **2022**, *216*, 106379. [[CrossRef](#)]
201. United States Geological Survey. *What is a Landslide and What Causes One?* U.S. Geological Survey: Reston, VA, USA, 2022.
202. Bera, S.; Upadhyay, V.K.; Guru, B.; Oommen, T. Landslide inventory and susceptibility models considering the landslide typology using deep learning: Himalayas, India. *Nat. Hazards* **2021**, *108*, 1257–1289. [[CrossRef](#)]
203. Chang, L.; Zhang, R.; Wang, C. Evaluation and Prediction of Landslide Susceptibility in Yichang Section of Yangtze River Basin Based on Integrated Deep Learning Algorithm. *Remote Sens.* **2022**, *14*, 2717. [[CrossRef](#)]
204. Sun, X.; Yu, C.; Li, Y.; Rene, N.N. Susceptibility Mapping of Typical Geological Hazards in Helong City Affected by Volcanic Activity of Changbai Mountain, Northeastern China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 344. [[CrossRef](#)]
205. Dou, J.; Yunus, A.P.; Merghadi, A.; Shirzadi, A.; Nguyen, H.; Hussain, Y.; Avtar, R.; Chen, Y.; Pham, B.T.; Yamagishi, H. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Sci. Total. Environ.* **2020**, *720*, 137320. [[CrossRef](#)] [[PubMed](#)]
206. Dao, D.V.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayed, H.; Phong, T.V.; Ly, H.B.; Le, T.T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA* **2020**, *188*, 104451. [[CrossRef](#)]
207. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.W.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2020**, *17*, 641–658. [[CrossRef](#)]
208. Huang, F.; Cao, Z.; Guo, J.; Jiang, S.H.; Li, S.; Guo, Z. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *CATENA* **2020**, *191*, 104580. [[CrossRef](#)]
209. Huang, F.; Zhang, J.; Zhou, C.; Wang, Y.; Huang, J.; Zhu, L. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* **2020**, *17*, 217–229. [[CrossRef](#)]
210. Chang, Z.; Huang, J.; Huang, F.; Bhuyan, K.; Meena, S.R.; Catani, F. Uncertainty analysis of non-landslide sample selection in landslide susceptibility prediction using slope unit-based machine learning models. *Gondwana Res.* **2023**, *117*, 307–320. [[CrossRef](#)]
211. Dahal, A.; Lombardo, L. Explainable artificial intelligence in geoscience: A glimpse into the future of landslide susceptibility modeling. *Comput. Geosci.* **2023**, *176*, 105364. [[CrossRef](#)]
212. Zhao, Z.; Liu, Z.; Xu, C. Slope Unit-Based Landslide Susceptibility Mapping Using Certainty Factor, Support Vector Machine, Random Forest, CF-SVM and CF-RF Models. *Front. Earth Sci.* **2021**, *9*, 589630. [[CrossRef](#)]
213. Ye, C.; Tang, R.; Wei, R.; Guo, Z.; Zhang, H. Generating accurate negative samples for landslide susceptibility mapping: A combined self-organizing-map and one-class SVM method. *Front. Earth Sci.* **2023**, *10*, 1054027. [[CrossRef](#)]
214. Xi, C.; Han, M.; Hu, X.; Liu, B.; He, K.; Luo, G.; Cao, X. Effectiveness of Newmark-based sampling strategy for coseismic landslide susceptibility mapping using deep learning, support vector machine, and logistic regression. *Bull. Eng. Geol. Environ.* **2022**, *81*, 174. [[CrossRef](#)]
215. Gupta, S.K.; Shukla, D.P. Handling data imbalance in machine learning based landslide susceptibility mapping: A case study of Mandakini River Basin, North-Western Himalayas. *Landslides* **2023**, *20*, 933–949. [[CrossRef](#)]
216. Fang, Z.; Wang, Y.; Niu, R.; Peng, L. Landslide Susceptibility Prediction Based on Positive Unlabeled Learning Coupled With Adaptive Sampling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11581–11592. [[CrossRef](#)]
217. Wu, B.; Qiu, W.; Jia, J.; Liu, N. Landslide Susceptibility Modeling Using Bagging-Based Positive-Unlabeled Learning. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 766–770. [[CrossRef](#)]
218. Wang, Y.; Feng, L.; Li, S.; Ren, F.; Du, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *CATENA* **2020**, *188*, 104425. [[CrossRef](#)]
219. Huang, F.; Xiong, H.; Yao, C.; Catani, F.; Zhou, C.; Huang, J. Uncertainties of landslide susceptibility prediction considering different landslide types. *J. Rock Mech. Geotech. Eng.* **2023**, *15*, 2954–2972. [[CrossRef](#)]
220. Sun, D.; Chen, D.; Zhang, J.; Mi, C.; Gu, Q.; Wen, H. Landslide Susceptibility Mapping Based on Interpretable Machine Learning from the Perspective of Geomorphological Differentiation. *Land* **2023**, *12*, 1018. [[CrossRef](#)]
221. Pradhan, B.; Dikshit, A.; Lee, S.; Kim, H. An explainable AI (XAI) model for landslide susceptibility modeling. *Appl. Soft Comput.* **2023**, *142*, 110324. [[CrossRef](#)]
222. Collini, E.; Palesi, L.A.I.; Nesi, P.; Pantaleo, G.; Nocentini, N.; Rosi, A. Predicting and Understanding Landslide Events With Explainable AI. *IEEE Access* **2022**, *10*, 31175–31189. [[CrossRef](#)]
223. Fang, H.; Shao, Y.; Xie, C.; Tian, B.; Shen, C.; Zhu, Y.; Guo, Y.; Yang, Y.; Chen, G.; Zhang, M. A New Approach to Spatial Landslide Susceptibility Prediction in Karst Mining Areas Based on Explainable Artificial Intelligence. *Sustainability* **2023**, *15*, 3094. [[CrossRef](#)]
224. Zhang, J.; Ma, X.; Zhang, J.; Sun, D.; Zhou, X.; Mi, C.; Wen, H. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manag.* **2023**, *332*, 117357. [[CrossRef](#)]

225. Zhu, Q.; Chen, L.; Hu, H.; Pirasteh, S.; Li, H.; Xie, X. Unsupervised Feature Learning to Improve Transferability of Landslide Susceptibility Representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3917–3930. [[CrossRef](#)]
226. Zhiyong, F.; Changdong, L.; Wenmin, Y. Landslide susceptibility assessment through TrAdaBoost transfer learning models using two landslide inventories. *CATENA* **2023**, *222*, 106799. [[CrossRef](#)]
227. Wang, Z.; Goetz, J.; Brenning, A. Transfer learning for landslide susceptibility modeling using domain adaptation and case-based reasoning. *Geosci. Model Dev.* **2022**, *15*, 8765–8784. [[CrossRef](#)]
228. Youssef, A.M.; Pourghasemi, H.R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **2021**, *12*, 639–655. [[CrossRef](#)]
229. Chang, Z.; Huang, F.; Huang, J.; Jiang, S.H.; Liu, Y.; Meena, S.R.; Catani, F. An updating of landslide susceptibility prediction from the perspective of space and time. *Geosci. Front.* **2023**, *14*, 101619. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.