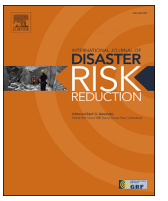
International Journal of Disaster Risk Reduction 108 (2024) 104526

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr



A machine learning approach to evaluate coastal risks related to extreme weather events in the veneto region (Italy)

Maria Katherina Dal Barco [a, b], Margherita Maraschini [a, b], Davide Mauro Ferrario [a, b], Ngoc Diep Nguyen [a, b], Silvia Torresan [a, b], Sebastiano Vascon [a], Andrea Critto [a, b, *]

[a] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy
[b] Risk Assessment and Adaptation Strategies Division, Fondazione Centro Euro-Mediterraneo Sui Cambiamenti Climatici (CMCC), Venice, Italy

ARTICLE INFO

Keywords:
MLP
Risk assessment
Feature importance
Exposure and vulnerability
Impact

ABSTRACT

A significant and unprecedented increase of temperature has been recorded worldwide during the last decades, leading to the occurrence of numerous extreme events. Coastal areas, with their high population density, interconnected economic activities, fragile ecosystems, are particularly vulnerable to climate change impacts. These impacts can be intensified by the interactions of multiple hazards which operate at different spatio-temporal scales and affect exposure and vulnerability patterns. An integrated approach is here proposed to assess the relationship between risk factors and to evaluate the multiplicity of impacts that may affect the coastline. A new path to tackle these multi-risk events is offered by ML algorithms to effectively handle vast amounts of heterogenous data, and model complex non-linear relationships between multiple factors and feedback mechanisms. To assess impacts caused by extreme events (storm surges, extreme precipitation, wind events) in the Veneto coastal municipalities, a ML approach was developed to understand connections between atmospheric and marine hazards and impacts recorded by the Veneto region emergency archive during the 2009–2019 timeframe, identifying the most influencing factors triggering multiple risks. Additionally, the coastal municipalities were clustered considering the intrinsic relationships between impact occurrences, exposure and vulnerability features. Several algorithms were compared to estimate daily risk of impacts to occur providing hazards, exposure and vulnerability information. The MLP algorithm showed satisfactory performances (weighted-F1-score of 0.94) to estimate the relative importance of input features. The proposed algorithm was designed as support tool to increase the understanding of impacts' occurrence in coastal areas, thus helping the adaptation planning process.

1. Introduction

Over the past three decades, the global climate has experienced a significant and unprecedented increase of temperature [1,2]. According to the results of climate models, global surface temperature is projected to further increase up to 4 °C by the end of this century (IPCC, 2021, 2023), exacerbating changes in ecosystem functioning [3,4] and leading to the occurrence of many extreme events worldwide [1,2,5,6]. Both the frequency of high-temperature events (i.e., hot days and nights) and extreme rainfall events (triggering pluvial and fluvial floods) have already increased in many parts of the world [1,2,5,7,8]. Similarly, in many regions, the number of

* Corresponding author. Department of Environmental Sciences, Informatics and Statistics, University Ca' Foscari Venice, Via Torino 155, 30170 Venice, Italy.
E-mail addresses: mariakatherina.dalbarco@cmcc.it (M.K. Dal Barco), margherita.maraschini@cmcc.it (M. Maraschini), davide.ferrario@cmcc.it (D.M. Ferrario), ngocdiep.nguyen@cmcc.it (N.D. Nguyen), silvia.torresan@cmcc.it (S. Torresan), sebastiano.vascon@unive.it (S. Vascon), critto@unive.it (A. Critto).

https://doi.org/10.1016/j.ijdrr.2024.104526
Received 24 October 2023; Received in revised form 29 April 2024; Accepted 30 April 2024
Available online 4 May 2024

rainy days has significantly decreased causing severe water scarcity and drought conditions [4]. As multiple hazards may interact and affect the environment and the society in complex ways, there is the need for improved climate risk methods and tools to better understand and anticipate consecutive, co-occurring or interacting hazards affecting the same territory.

This is particularly important in coastal areas where land, sea and climate-related hazards interact at different spatial and temporal scales, leading to compound and cascade impacts on society ad ecosystems [9]. Given the multi-risk nature of the problem at stake, as well as the non-linear interactions of exposure, vulnerability, and hazard factors, a major scientific effort to enhance current methodologies and capabilities to better assess climate risks from regional to local scales and predict their impacts on society and ecosystems is needed.

In recent years, Machine Learning (ML) approaches have been increasingly applied to model climate change risks, given their capacity to analyse complex interactions and feedback structures, exploiting the availability of big and heterogenous volumes of data [10]. Given the complexity of the coastal system, ML methods started to be implemented to capture the relationships between the natural and anthropic pressures, thus understanding the consequences of such interactions. At first, Bayesian Networks (BN) have been developed as a surrogate of more complex physical-mathematical models, for translating marine offshore hazards into damages at the coastal receptors [11–17]. In particular, Jäger et al. [14] considered four receptors (i.e., residential properties, commercial properties, people, and saltmarshes) affected by different hazards (e.g., flood depth, wave height). Sanuy & Jiménez [17] perfectioned the approach by training the dataset with real storms data, thus decreasing the uncertainty associated to the use of synthetic events. Moreover, by splitting the BN into two parts, the authors were able to consider the variability of the forcing conditions (solving the source–consequence relationships), while examining the spatial distribution of the receptors.

On the other hand, rapid and extreme weather events have been analysed by Park and Lee [18], who developed a risk probability map for evaluating flooding in the coastal areas of South Korea. Different classification ML methods, including k-Nearest Neighbor (kNN), Random Forest (RF), and Support Vector Machine (SVM), were compared to predict the presence or the absence of flooding, in relation to hazard, exposure and vulnerability features (e.g., tide, precipitation, elevation, slope, land-use) [19–21]. Building upon this research area, here a ML model was developed and applied to the coastal municipalities of Veneto region (Italy) in order to estimate the potential risks associated to climate change and extreme weather events in coastal areas, given a set of atmospheric and marine indicators. Specifically, the designed model is able to associate a risk score to each day and municipality in the Veneto costal area given the knowledge of the indicators, where a risk score is a measure of the likelihood that an impact occurs, where impacts includes both damages (e.g., agricultural losses, shoreline erosion, damage to buildings) and services interruption (e.g., blackouts, impaired viability), caused by extreme climate events. Once trained, the model can determine the importance of the different factors that lead to the recorded impacts. For this purpose, several Supervised ML models (i.e., algorithms whose parameters are estimated using data for which both the input features and the output labels are known, such as Random Forest, Neural Networks and Support Vector Machines), have been implemented. These algorithms are able to model interactions between different factors using data from past historical events and cope well with the high number of parameters involved in the risk analysis [22].

Overall, in multi-risk applications, algorithms like Random Forest, Decision Trees and Bayesian Networks are more commonly applied than Deep Learning techniques, mainly due to their simpler implementation and interpretation [10]. Support Vector Machine and Multi-Layer Perceptron algorithms are often included as benchmark for evaluating classification and regression performances of other more complex models, due to their efficiency, together with classical statistical algorithms, like Logistic Regression and Generalised Additive Models (GAM), which instead offer more interpretable frameworks, while being still able to cope with the non-linear nature of the risk drivers' relationships [23]. Deep Learning models instead are more often applied to Early Warning Systems applications, studies that focus on hazard predictions and forecasting or research leveraging earth observations data or social media data, where harnessing big data can really play a big role. Multi-risk applications need to integrate hazard, vulnerability and exposure drivers, often at different spatio-temporal resolution and are constrained by the scarceness and coarseness of impact data, which further limits the possibility to leverage Deep Learning models [24,25].

One of the challenges of the problem at hand is the skewness of the input dataset (i.e., the presence of many more days impact-free compared with the number of days with impacts). This problem is common to several fields [26], including weather predictions [27,28], political science [29], manufacturing failure [30], medical examinations [31], fraud detection [32], finance [33], and social science [34]. This issue was addressed by applying weights that increase the importance of the minority class, as done in most of these papers.

Following the characterization of the case study area, as well as the collection and the analysis of available data (Section 2), the implemented ML algorithms are described (Section 3), and the main findings resulting from their application in the Veneto region case are summarized, and the strengths and weaknesses of the proposed approach are discussed (Section 4). In the conclusions, suggestions on future model improvements to respond to other EU directives related to water resource assessment and management, including modelled data, are proposed (Section 5).

2. Characterization of the case study

2.1. The coastal municipalities of the veneto region

The Veneto region, located in northeastern Italy along the North-Adriatic Sea, boasts an approximately 169 km coastline [35], spanning eleven municipalities (Fig. 1) across the provinces of Venice (i.e., San Michele al Tagliamento, Caorle, Eraclea, Jesolo, Cavallino-Treporti, Venice, and Chioggia) and Rovigo (i.e., of Rosolina, Porto Viro, Porto Tolle, and Ariano nel Polesine).

The littoral zone of the Veneto region experiences a subcontinental temperate climate [36], characterised by annual mean temperatures (around 14 °C) higher than the average 13 °C of the internal zones [36], along with fewer rainy days, but occasionally heavy

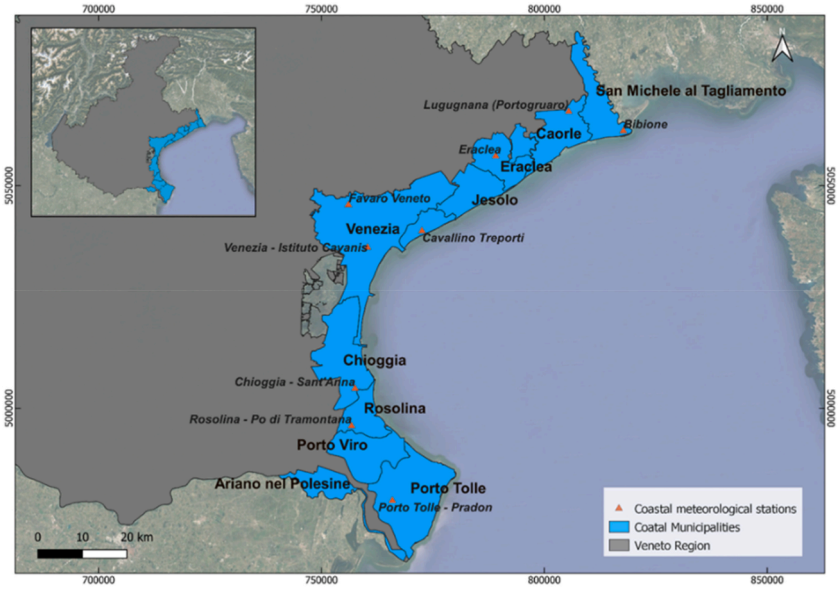


Fig. 1. Coastal municipalities of Veneto Region and their meteorological stations.

precipitation [37]. From a geomorphological point of view, the coast features low-lying sandy beaches, fragmented by the presence of seven river mouths (i.e., Tagliamento, Livenza, Piave, Sile, Brenta, Adige, Po, from north to south), three lagoons (i.e., Caorle lagoon, Venice lagoon, and lagoons of the Po River Delta), barrier beaches, deltas, and spits. Moreover, the sedimentary shore can be subdivided into a northern, central, and southern trait [38–41]. Despite the natural evolution of the coast, over the last century urbanization and anthropic activities have significantly altered the landscape, though approximately 60 km of the overall littoral have conserved their natural status, mainly because they cover lagoonal and fluvial estuary areas, which are difficult to urbanize [42].

Coastal development together with improper management practices (e.g., intensive water withdrawal from rivers, gravel and sand excavation, presence of several dams) has decreased the sedimentary budget for beach and dune accretion, exacerbating coastal erosion exacerbated by sea-level rise and storm surge conditions. Despite efforts, such as beach nourishment interventions and foredune restoration [43], the investigated area continues to face challenges, including relative sea-level rise caused by natural eustacy and subsidence [44,45].

The coastal municipalities of the Veneto region comprehend several natural protected areas, regional parks, and reserves (e.g., Bocche di Po, Valle Averto, Delta Po regional park, Bosco Nordio), and areas included in the European ecological network Natura 2000 [35,39]. The socio-economic capital is mainly related to maritime activities, fisheries, aquaculture, agriculture, industry (e.g., Porto Marghera), offshore activities and tourism [46]. Specifically, the primary sector covers an average share of 7.42 % of the total regional employment, whereas the tertiary sector, comprehending mainly tourism, is particularly associated to the city of Venice, as well as to beach destinations, which are chosen for the high water quality [47].

The coastal area is subjected to multiple natural and anthropic pressures, further intensified by climate change. The average temperature rose by approximately 0.55 °C per decade, with summer and autumn seasons recording the highest increment of 0.7 °C, leading to more frequent intense rainfall with strong wind gusts, flooding, and storm surge, and on the other, the magnification of heatwaves which create health risks for the population and droughty conditions. Extreme sea-level events, driven by a combination of tides, storm surges and wave energy, are expected to rise, posing significant challenges for the region in the coming years [48–50].

2.2. Data collection for ML application

The implementation of the ML model requires information of the occurrence of impacts related to extreme weather events along the coastal municipalities of the Veneto region during the 2009–2019 decade with daily temporal resolution. In particular, for each municipality and for each day, a series of indicators (or features), which constitute the input variables of the algorithm, as well as the output labels (a boolean indicator which describes the presence/absence of impact) were collected. These indicators are a set of spatio-temporal heterogeneous variables that can influence the likelihood of an impact to occur in the investigated area [15,51–54], and are mainly related to hazards, seasonality and location. Moreover, exposure and vulnerability characteristics are collected to be jointly analysed with the ML results.

The collection of the input features related to exposure and vulnerability, hazards, and impacts, are detailed in Table 1.

2.2.1. Exposure and vulnerability characteristics

To support the analysis of the factors that drive the occurrence of impacts and try to understand the underlying multi-risk dynamics, the exposure and vulnerability characteristics of the coastal municipalities of the Veneto region were collected.

Table 1
Summary of input data used for the Machine Learning application in the Veneto coastal area.

Data category	Indicator	Spatial domain	Spatial resolution	Timeframe	Temporal resolution	Source
Exposure and vulnerability	Total area	Veneto region	Municipality	2022	unique value	ARPAV[a]
	Land-use type	Italy/EU	1 km	1990–2020	multi-year	JRC – LUISA platform[b]
	Population					JRC – LUISA population[c]
	Soil characteristics	Veneto region	1:50.000/1:250.000	2022	unique value	ARPAV Geoportal[d]
	Topography	Italy	10 m	2017	unique value	[55,56]
Atmospheric hazards	Total precipitation	Veneto region	Coastal monitoring stations	2009–2019	hourly and daily	ARPAV[1] and ESWD[e]
	Wind speed					
	Humidity					
Oceanographic hazards	Sea surface height	Mediterranean Sea	1/24° (~4 km)	1987–2023	hourly and daily	CMEMS[f]
	Significant wave height					
Impacts	Damages and service interruption	Veneto region	Municipality	2009–2019	daily	Veneto Region[g]

[a] Regional Environmental Protection Agency of the Veneto region (ARPAV): https://www.arpa.veneto.it/dati-ambientali/open-data.
[b] Land Use-based Integrated Sustainability Assessment (LUISA) modelling platform developed by the European Commission's Joint Research Centre (JRC): https://data.jrc.ec.europa.eu/collection/luisa.
[c] Land Use-based Integrated Sustainability Assessment (LUISA) modelling platform for population developed by the European Commission's Joint Research Centre (JRC): https://data.jrc.ec.europa.eu/dataset/jrc-luisa-population-ref-2014.
[d] Geoportal developed by the Regional Environmental Protection Agency of the Veneto region (ARPAV): https://gaia.arpa.veneto.it/.
[e] European Severe Weather Database (ESWD): https://eswd.eu/.
[f] Copernicus Marine Service (CMEMS): https://doi.org/10.25423/CMCC/MEDSEA_MULTIYEAR_PHY_006_004_E3R1.
[g] Veneto Region – Post-Emergency Disaster Events Management Office: https://bur.regione.veneto.it/BurvServices/Pubblica/sommarioDecretiPGR.aspx?expand=19.

The total area of each municipality, the land-use types (i.e., anthropic, agriculture, natural, water bodies), soil characteristics (i.e., permeability, dune ridges and lagoon islands, reclaimed lagoon areas artificially drained), population (i.e., total population and density), and topography (i.e., elevation, slope) were studied. In particular, the municipality extension deserves a special mention, as it is reasonable to expect that bigger municipalities are affected by more impacts. However, it is impossible to normalize the number of days with impacts using the municipality size, as it is not recorded if more than one impact happens on a given day. Topography data of the study site were derived directly from the Digital Elevation Model (DEM) of 10 m × 10 m obtained by the National Institute of Geophysics and Volcanology [55,56]. Data on soil types and permeability were retrieved from ARPAV Geoportal[4]. Population and land-use data were collected from the JRC web portal.

As the exposure and vulnerability indicators are almost constant over time, and only one value per municipality is available, in the ML framework their contribution to the creation of the impact is conveyed by the municipality index. In other words, it is possible to estimate the influence that each municipality has in creating the impact, but it was not possible to determine which of the exposure and vulnerability factors are playing a major role. For this reason, the only indicator used as input for the ML model related to exposure and vulnerability is the municipality index. Although exposure and vulnerability features are not directly used as input data for the Machine Learning, the description of these indicators is included in this study to contextualise the interpretation of the results (Section 4.3.2).

2.2.2. Hazard input features

In order to estimate the impacts along the Veneto coastal municipalities, the ML-based tools are trained on hazard-related indicators, such as atmospheric (i.e., precipitation, wind, humidity) and oceanographic features (i.e., sea surface height and significant wave height). The hazard information is related to the day when the impact did/did not occur, as well as to the 3 days just before, as some cascading effects can be delayed [57,58].

The atmospheric data were derived from the monitoring network of the Regional Agency for Environmental Protection and Prevention of the Veneto Region (ARPAV), which acquires precipitation, temperature, humidity and wind data at hourly or daily basis. Each municipality has at least one monitoring station, except for Jesolo, Porto Viro, and Ariano nel Polesine (Fig. 1). Considering the proximity of the municipalities and the low variation in the values of the atmospheric parameters, in accordance with the data provider, the nearest neighbor rule was applied to infer their atmospheric conditions in the missing stations. In particular, in Jesolo and Porto Viro the missing entries are filled with the average of the values measured by two closest stations, while in Ariano nel Polesine, which borders another region, they are replaced with the values measured in the nearest station. On the other hand, Venezia has two stations within its territory, and similar values were recorded. However, since the Favaro Veneto station has more missing entries than the Venezia - Istituto Cavanis, the latter was selected to represent the municipality. Additionally, wind indicator data from the monitoring stations has been integrated with information from the European Severe Weather Database[5], as local events (e.g., tornadoes) may not be detected by the stations.

Oceanographic hazard indicators for the Veneto pilot were collected from the Copernicus Marine Service (CMEMS) provided by the European Union's Earth Observation Programme. The above-mentioned indicators were retrieved from the Mediterranean Sea Physics Reanalysis database (i.e., MEDSEA_MULTIYEAR_PHY_006_004)[6], which is the product of a numerical composed hydrody-

namic model, supplied by the Nucleus for European Modelling of the Ocean (NEMO)[1] and a variational data assimilation scheme (OceanVAR).[2] Specifically, the sea surface height and the wave height data are available on a regular grid of 1/24° (about 4 km) at hourly resolution. The values of the closest pixels to the municipalities' coastline were extracted and aligned with the local data at the *Acqua Alta* platform.

2.2.3. Impact data

The impact data indicator includes both damages (e.g., agricultural losses, shoreline erosion, damage to buildings) and services interruption (e.g., blackouts, impaired viability), caused by extreme climate events within the coastal municipalities of the Veneto region. The list includes events that are severe enough to trigger the 'State of crisis' (i.e., *Stato di crisi*) in some coastal municipalities of the Veneto region, i.e., events that for intensity and extension require an immediate response from the regional authorities. This information is retrieved from Decreto del Presidente della Giunta Regionale reports (DPGR, namely Decree of the President of the Regional Council). For each impact, these documents provide qualitative information on the reported damages, the list of the municipalities affected and the dates when the impact took place. Typologies of the damages reported include physical damages related to urban flooding, agriculture/fisheries, people (e.g., fatalities, injuries, displacements), beaches (e.g., shoreline erosion, debris accumulation), structures/infrastructures, economic activities, and tertiary sector. Ideally, in order to accurately determine the quantitative extent of impacts, more detailed data and reports on the investigated extreme event, including coordinates, typology, monetary cost of restoration would be required. However, detailed quantitative data on impacts and their economic costs are not in the public domain. Therefore, the impact dataset is a set of couples (day, municipality) where an impact occurred.

Unfortunately, the impacts collected in the 'Stato di crisi' reports may be incomplete and misleading: not all impacts that created an economic loss or a service interruption have been filed. On the other hand, each 'Stato di crisi' report lists a series of municipalities and a series of dates but does not explain in which date each municipality was affected. Similarly, the typology of damage, when present, is not associated to a specific date or municipality, making this information impossible to be used.

To identify the correct coupling between dates and municipalities, and therefore reduce the errors in the input dataset, each impact has been checked against local newspapers. Additionally, local newspapers have also been used to identify events that were not reported in the 'Stato di crisi' database, but still led to damages to population or infrastructures. Eventually, the impact dataset comprises a total of 447 days of impacts from extreme weather events over eleven Veneto coastal municipalities during the 2009–2019 timeframe.

3. Method

The objective of this study is to understand the dynamics that generate the occurrence of impacts, with the aid of a tool able to calculate a risk score for each sample (day, municipality), which is a measure of the likelihood that an impact occurs in that municipality on that day, given the corresponding set of atmospheric and marine indicators. The different methodological steps to achieve this goal are depicted in Fig. 2. The first step is related to the data collection (already described in Section 2.2). The collected data are then cleaned and analysed (Section 3.2) to have a preliminary idea of the dynamics involved. In parallel, the performance metric that will be used to measure the quality of the algorithms needs to be chosen (Section 3.3.1). The cleaned data and the performance metric are jointly used to design, train, tune, validate, and test several ML algorithms (Section 3.3.2). These can be used to create a decision support tool for risk assessment, and to estimate the importance of different hazard factors (Section 3.3.3), which is then compared with the results of the data analysis to gain insight into the physical drivers of the phenomena, hence contributing to the implementation of adaptation planning.

3.1. Exposure and vulnerability feature analysis

The analysis of exposure and vulnerability factors is aimed at understanding the intrinsic characteristics of the investigated case study, highlighting the territorial peculiarities that could intensify the effects of extreme weather events.

Once collected (Section 2.2.1), the exposure and vulnerability data were extracted as average value for each municipality and normalized. The resulting output of this analysis is detailed in Section 4.1.

3.2. Data wrangling, cleaning and analysis

The collected input data for ML (i.e., hazard indicators, location, seasonality, impacts) need to be transformed and cleaned from one raw to into a structured format, with the intent of improving data quality and make it more suitable as input for the algorithm (data wrangling): each sample of this dataset is related to a couple (day, municipality), and one and only one value for each indicator is associated to each sample. Each sample in the input dataset also contains the corresponding output label, indicating the presence or absence of impact. All samples, whose output label is 1 (an impact occurred on the corresponding day and municipality), are referred to as positive event and belong to the positive class. Specularly, all samples whose output label is 0 (no impact occurred on the corresponding day and municipality) are referred to as negative event and belong to the negative class.

Initially the collected hazard indicators list included a wide range of variables: daily precipitation (maximum in 1 h and cumulative during a day), maximum number of consecutive dry days, number of days with cumulative precipitation exceeding that of the 95th percentile, maximum cumulative precipitation in 1 day, maximum and minimum temperature, number of heat waves, heat wave

[1] Nucleus for European Modelling of the Ocean (NEMO) model: https://www.nemo-ocean.eu/.

[2] Variational ocean (OceanVar) data assimilation system developed at the Euro-Mediterranean Center on Climate Change (CMCC): https://www.cmcc.it/models/oceanvar.

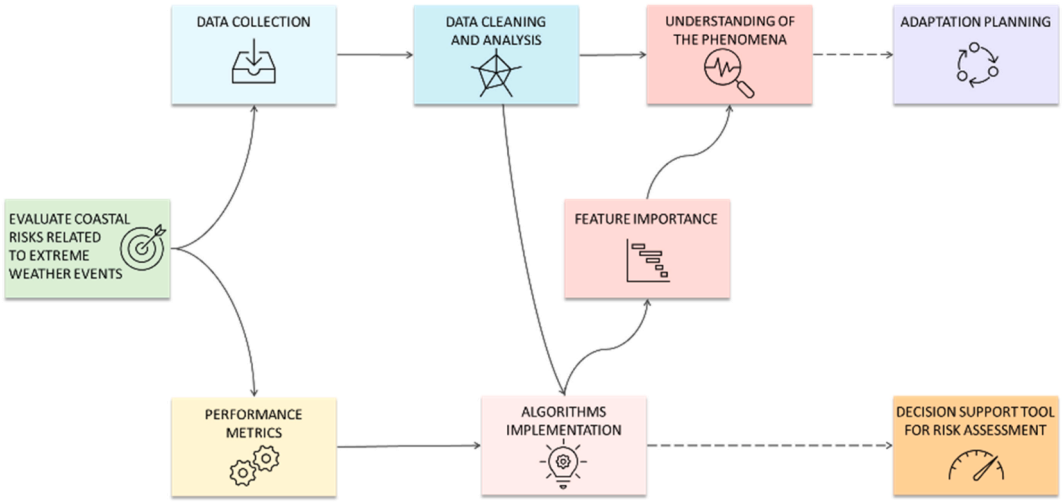


Fig. 2. Operative steps adopted to estimate the risk score and the most influencing features in the reference timeframe (2009–2019) for the Veneto coastal municipalities.

temperature, number of tropical nights, number of hot days, humidity, wind velocity (maximum and average) and direction, sea water velocity and direction, wind wave height, maximum wind wave height, wind wave direction, sea surface wind wave mean period, mean and maximum sea surface height. This list of indicators contains redundant information, as indicators are not independent from each other. Moreover, the low quantity and quality of impact data suggested to reduce the number of input indicators to avoid the risk of overfitting.

In order to reduce the number of input variables without affecting the quality of results an iterative procedure was set in place: first, all collected hazard indicators were used to train the algorithm, and the algorithm performance was evaluated on the validation dataset and the feature importance was calculated; this process was repeated several times in order to have an average performance of the algorithm. Then, the indicators that scored lower in the feature importance were gradually removed from the input indicators, the algorithm was retrained, and the performance validated: if the new performance was better or equal to the performance obtained before removing the indicators, the indicators were considered superfluous. This process was repeated until only indicators whose presence improved the performance were left. The final list of input indicators is summarized in Table 2.

The following data cleaning process consists in filling out missing data and identifying potential errors; this step is necessary, as its correct execution has major effects on the performances of the model.

Finally, the conditional distributions of the indicators given the presence/absence of impacts were analysed. This analysis is meaningful as every supervised learning algorithm relies, although in different ways, on the differences of these conditional distributions. The indicators are not independent from one another, especially those describing the same phenomenon, hence there is a lot of redundancy in the information they provide. However, each indicator adds useful information to the analysis. The results of the hazard data analysis are presented in Section 4.2.

3.3. Development of machine learning algorithms

AI and ML algorithms offer a new path to address the analysis of multiple environmental hazards due to their ability to model complex feedbacks and non-linear interactions between different factors without the need for an explicit modelling. In the frame of this work, several supervised ML methods for binary classification, whose goal is learning a function that maps input features to their associated classes [59], were tested. During training, each of these algorithms takes as input the values of the hazard-related indica-

Table 2
List of input indicators to characterize the Veneto coastal case study.

Acronym	Definition
Month	*Month related to the impact*
Municipality	*Location where the impact occurred*
MSSH	*Maximum hourly sea surface height in a day*
MWIH	*Maximum hourly significant wave height in a day*
WIH	*Daily average significant wave height*
PRCMAX	*Maximum hourly precipitation in a day*
PRCTOT	*Total daily precipitation*
PRCTOT_TOT_MAX_3	*Maximum daily total precipitation in the previous 3 days*
RX-1day	*Maximum daily precipitation in a month*
URmax	*Maximum hourly relative humidity in a day*
VRFDd	*Maximum hourly wind velocity in a day*

tors (e.g., the atmospheric and oceanographic indicators) and as output the corresponding value of the risk score, which, for labelled data, can only be 1 for samples associated with impacts and 0 otherwise. During testing and application, the outputs of the algorithms applied to new samples are the corresponding estimated risk scores.

3.3.1. Performance metrics

The following step of the ML design consists in the choice of the metric, which is a scalar function that takes as input the estimated and the observed output for the whole dataset and calculates the error, a single real number that describes the distance of the algorithm result from the target: the lower the error value, the better the algorithm performance (Fig. 3). The metric is then used to optimize the values of the algorithm parameters and to compare the performances of different algorithms.

The choice of the metric is extremely important as it would set the objective of the analysis and defines which errors are worse than others, influencing the results more than any of the other choice. To define the metric for this specific case study (i.e., a binary classification problem of an extremely skewed dataset), we need to address two issues: (i) how to calculate the error of each sample? (ii) which weight to attribute to each class?

Considering the estimation of the error for a single sample first, most algorithms for binary classification (including the one implemented in this study) calculate a probability p_i that a sample i is associated to the positive class. Consequently, it is possible to associate a sample i to the positive class if this estimate p_i is above a threshold T and to the negative class otherwise. A metric that estimates the error for each sample can be based on the difference between the observed and the estimated class label (Fig. 4-a), or on the difference between the observed class label and the estimated probability value (Fig. 4-b). The main difference between these approaches is that, when using a metric based on the number of samples that fell into each class (e.g., accuracy, precision, recall, F1 score), the same loss can be attributed to events with very different values of risk score (points P1 and P2 in Fig. 4-a have the same loss, while their risk score is very different). On the other hand, if a metric such as the log-loss is used, two elements belonging to the same class but with different predicted probabilities have a different loss (points P1 and P2 in Fig. 4-b have the different losses).

The risk score-based metric 'Cross Entropy Loss' has been used because this function can better represent the uncertainty of the problem and allows for more flexibility in the use of the results. However, as the knowledge of the class (i.e., if a sample is likely associated to an impact) is useful in most applications, the weighted F1 score is also used to analyse the results. The threshold is defined by maximizing the weighted F1 score on the test dataset.

In rare event estimation most classification algorithms can sharply underestimate the probability events belonging to the minority class [60]: possible solutions of the unbalance problem can be the application of weights, the under-sampling of the majority class or the oversampling of the minority one [61–63].

In light of these considerations, the explicit formula of our chosen metric, the (binary) cross entropy with weights inversely proportional to the cardinality of each class, is:

$$loss = \frac{1}{n}\sum_{i=1}^{n} w[y_i \ ln(f(x_i))\ ] + (1 - y_i)\ ln\big(1 - f(x_i)\big),$$

where w is the ratio between the number of elements in the negative class and in the positive class and $f(x_i)$ is the predicted value of the risk score for sample x_i. *It* is important to note that the risk score calculated by the algorithm is not the probability of having an impact on a given day because the weights are not 1.

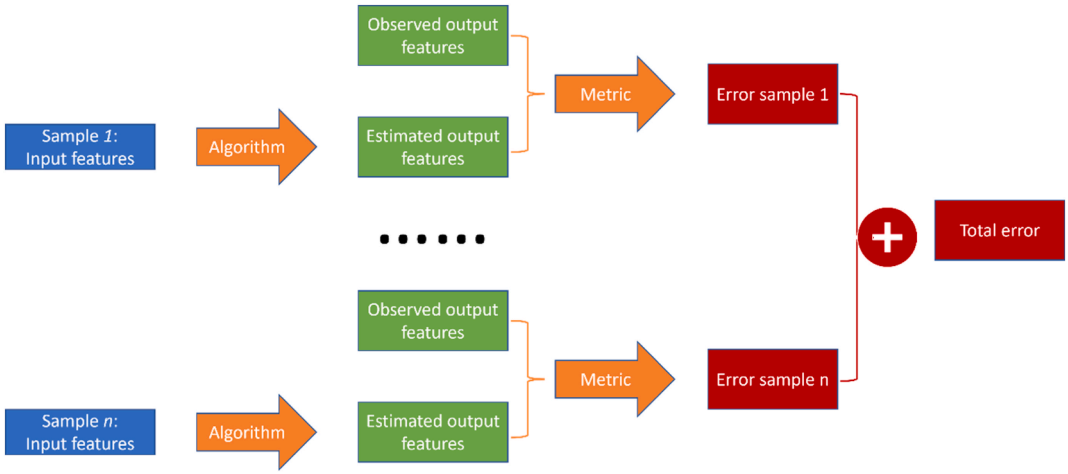


Fig. 3. Scheme of the calculation of the error for each algorithm: for each sample the algorithm calculates the estimated output features. The distance between the estimated and the observed output features for each sample is calculated by the metric, and the weighted sum of the sample error over the dataset represents the performance of the algorithm on the dataset.

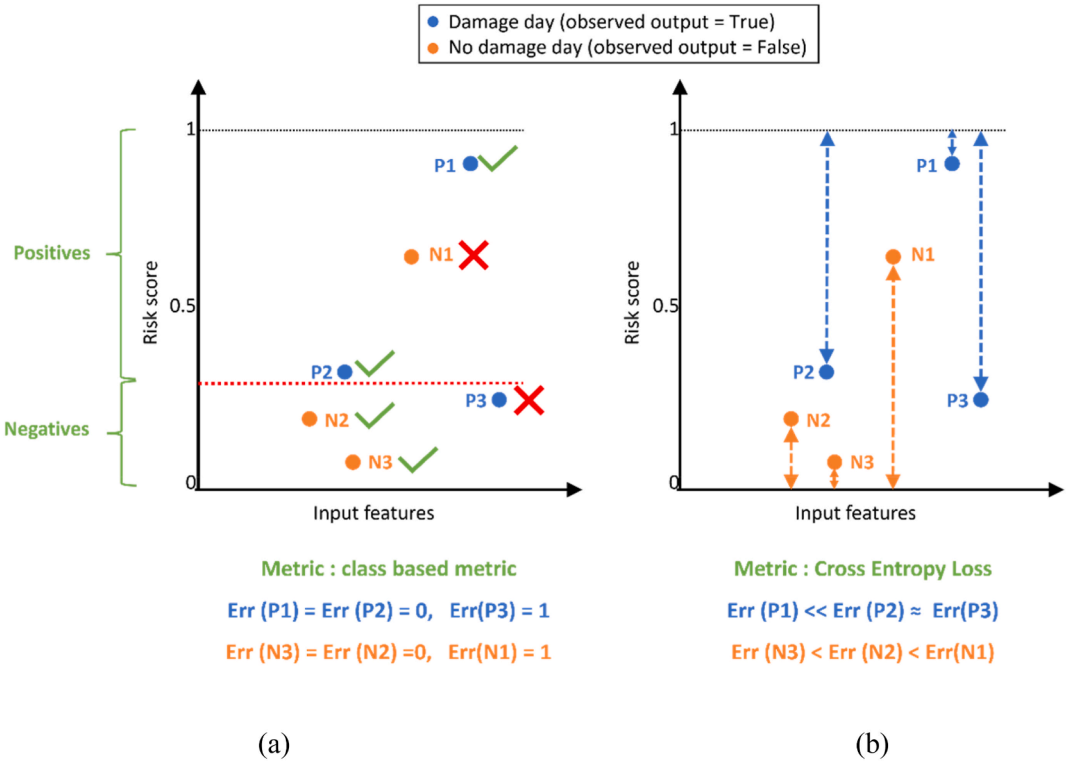


Fig. 4. Schematization of a class-based metric (a) versus a risk score-based metric (b). In (a) the risk score value estimated for each sample is only used to be compared against a threshold (red dotted line). All samples with risk score higher than the threshold are considered positives, and the others are considered negatives, irrespectively of the value of the risk score inside the range. In (b) the error of each sample is associated with the risk score value of the sample, and no threshold is considered. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.3.2. Algorithm

To design a ML model for the analysis of impacts related to atmospheric and marine hazards, the available data must be divided into training/validation and test sets. In this work, the data is split chronologically [64] i.e., the train and validation dataset correspond to the years from 2009 to 2016, and test dataset corresponds to the years after 2016. This choice, although not ideal from a ML point of view, is justified by the non-independence of data samples: some indicators are monthly or yearly, and others are based on the values that the indicator assumed in the previous days, hence indicators of adjacent days are correlated.

Then, several ML algorithms were coded using ScikitLean [65], trained, tuned and compared: *(i)* Multi-Layer Perceptron, which is a type of neural network (MLP henceforth; [66]), *(ii)* a Random Forest (RF henceforth; [67]), and *(iii)* a Support Vector Classifier (SVC henceforth; [68]). Regularization coefficients were used in the MLP and SVC algorithms to reduce the risk of overfitting [69]. Other options were also considered, including autoencoders [70] and isolation forest [71] for anomaly detection, but as the preliminary results were not promising, these options were dropped in the early stages.

Given the relatively small amount of data available, to reduce the high redundancy of clusters of features (e.g., precipitation features, sea features) and the risk of overfitting, during the ML analysis the number of indicators were iteratively decreased, gradually removing the indicators that, accordingly to the feature importance, did not play a key role in the estimate of the result. After each step of this procedure, the algorithm performances were checked. Moreover, an error analysis was performed iteratively on the training set [72], and outliers were identified and cross-checked.

The final parameters of the algorithms are listed in Table 3.

In order to increase the weight of positive events, oversampling of the minority class in the training dataset is performed before training. Oversampling was performed using the SMOTE tool implemented in the imblearn package [73].

The algorithms are then applied to the test dataset, and results are summarized in Section 4.2.

Table 3
Parameters for the implemented ML algorithms – i.e., Multi-Layer Perceptron (MLP), Random Forest (RF) and Support Vector Classifier (SVC).

RF		MLP		SVC	
Number of Trees	200	**Maximum # of iterations**	2000	**Kernel**	rbf
Max depth	4	**# hidden layers**	2	**C**	0.01
Criterion	Log Loss	**Hidden layer sizes**	5	**Probability**	True
Minimum sample split	5	**Alpha**	3	**Degree**	3

3.3.3. Feature importance for sensitivity evaluation

A sensitivity analysis was performed to identify the most influential factors in the impact generation, which is important for the impact prevention and management. This sensitivity analysis was carried out using the permutation method for the estimation of the feature importance, which was defined by Breiman [67] as the decrease of the metric value when the values of a single feature are randomly shuffled. Features are important if they contain valuable information, hence the more important a feature is, the more the prediction results will be affected by its incorrect values; this decrease of performances may be a measure of the importance of that specific feature.

The implementation of the method consists of randomizing one input feature at a time, calculating the corresponding metric value and repeating the process several times to reduce the sensitivity to the randomized set.

The contribution of the atmospheric, oceanographic, temporal, and geographical factors to impacts was explored. In addition, to better understand the contribution of exposure and vulnerability to the dynamics of risks, the feature importance was performed on groups of municipalities sharing similar patterns, as detailed in Section 3.1. The results of these analyses are illustrated in Section 4.3.2.

4. Results

As described in Section 3, the proposed study was organized into a preliminary data analysis and the following implementation of the ML algorithms. Specifically, the results of the exposure and vulnerability analysis are presented in Section 4.1, whereas Section 4.2 highlights the outcomes of the conditional distribution of the input data for the algorithm. Finally, Section 4.3 outlines the result of the ML model application along the coastal municipalities of the Veneto region, as well as the results of the feature importance for the whole Veneto region and the group of municipalities.

4.1. Exposure and vulnerability results

Exposure and vulnerability factors have been analysed to outline the territorial characteristics of the investigated case study, which may intensify the effects of extreme weather events.

The coastal municipalities of the Veneto region have been divided into four groups which share similar features (Fig. 5). The main criteria used to create these groups are the geographical position and geomorphology. The municipalities belonging to each group are adjacent to each other; the only exception is the municipality of Venice, which was extracted from *Group 2* due to its peculiar vulnerabilities related to its cultural heritage and to its high population density (Fig. 6). Municipalities belonging to the same group share the same orientation and similar geomorphological characteristics (e.g., river valley, lagoon, river delta; Fig. 6). It can be noted that municipalities in each group are also affected by similar number of impacts.

The resulting outputs of the data analysis for each group of municipalities are shown in Fig. 6, and the characteristics of each group are described below, from North to South.

Group 1 consists of the municipalities of San Michele al Tagliamento, Caorle and Eraclea. The territory is mainly covered by agricultural fields, which are former lagoon areas that have been reclaimed and artificially drained, lowering the soil permeability coefficient.

The lagoonal municipalities of Jesolo, Cavallino-Treporti and Chioggia have been assembled in *Group 2*. Although the dunes have been reinforced to protect the mainland in all these municipalities, this group has the second highest number of impacts. However, the high number of impacts may be due the presence of tourist facilities that increases the exposure of the area.

The municipality of Venice, identified as *Group 3*, is by far the biggest, the most densely populated and urbanized municipality, covering the highest extension of water (i.e., lagoon and coast), which increases the inundation risks. It has recorded the most impacts, probably because of its size, population and high vulnerability, especially related to its historic centre, classified as UNESCO World Heritage site. Accordingly, due to its peculiarity, the municipality of Venice forms a group by itself.

Finally, the municipalities located in the province of Rovigo (i.e., Rosolina, Porto Viro, Porto Tolle and Ariano nel Polesine – *Group 4*) are the least developed, and the area is covered by rural fields and natural parks (e.g., River Po Delta Regional Park).

4.2. Hazard data analysis results

In order to understand the relationships between hazards and impacts, the conditional distribution of each hazard when an impact and when no impact is recorded are analysed (Fig. 7).

The indicators related to precipitation (i.e., PRCTOT, PRCMAX, PRCTOT_TOT_MAX_3, and RX-1day), sea surface height (MSSH), wave height (MWIH and WIH) and wind velocity (VRFDd) show that the orange curve has a higher average and variance than the blue one. This means that impacts are more likely to occur if the values of precipitation, sea surface height, wave height or wind velocity are higher, as expected. However, for most indicators the range of values in which the distribution overlap is quite wide, indicating that a single feature analysis would not be enough to estimate the presence of an impact, hence suggesting the need to multi-features approaches (e.g., ML). On the other hand, the difference in the distribution of maximum humidity over the previous 3 days (URmax) is more subtle, suggesting that this indicator is not strongly correlated with the presence of impacts.

Finally, the histogram in Fig. 8 shows three peaks of recorded impacts throughout the year. In particular, the highest values belong to the autumn months (i.e., October and November), followed by February, and the months with increasing heat (i.e., May, June, and July). This result suggests that the month indicator has meaningful relevance for machine learning models.

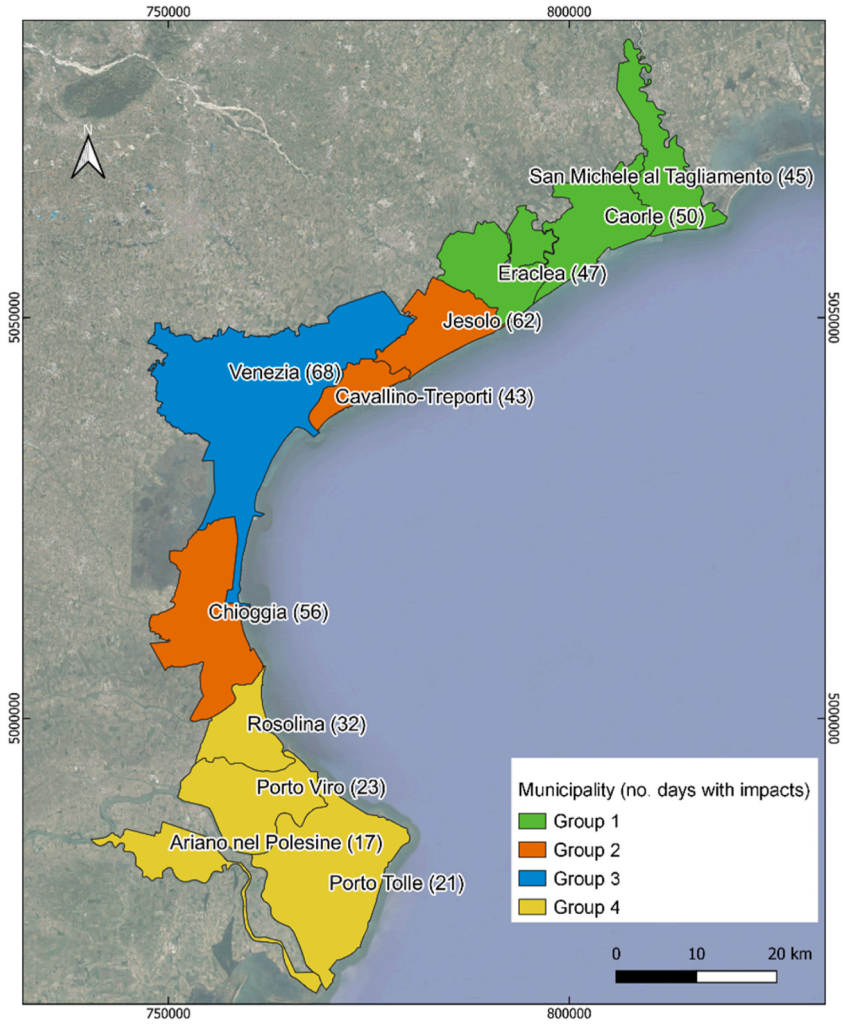


Fig. 5. Classification of the coastal municipalities according to the occurrence of the impacts, as well as on the characteristics of exposure and vulnerability.

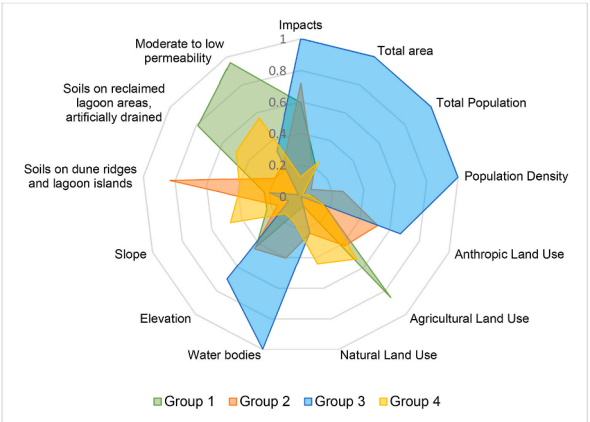


Fig. 6. Visualisation of the analysis combining the historical impacts and territorial features carried out to characterize the coastal municipalities of the Veneto region.

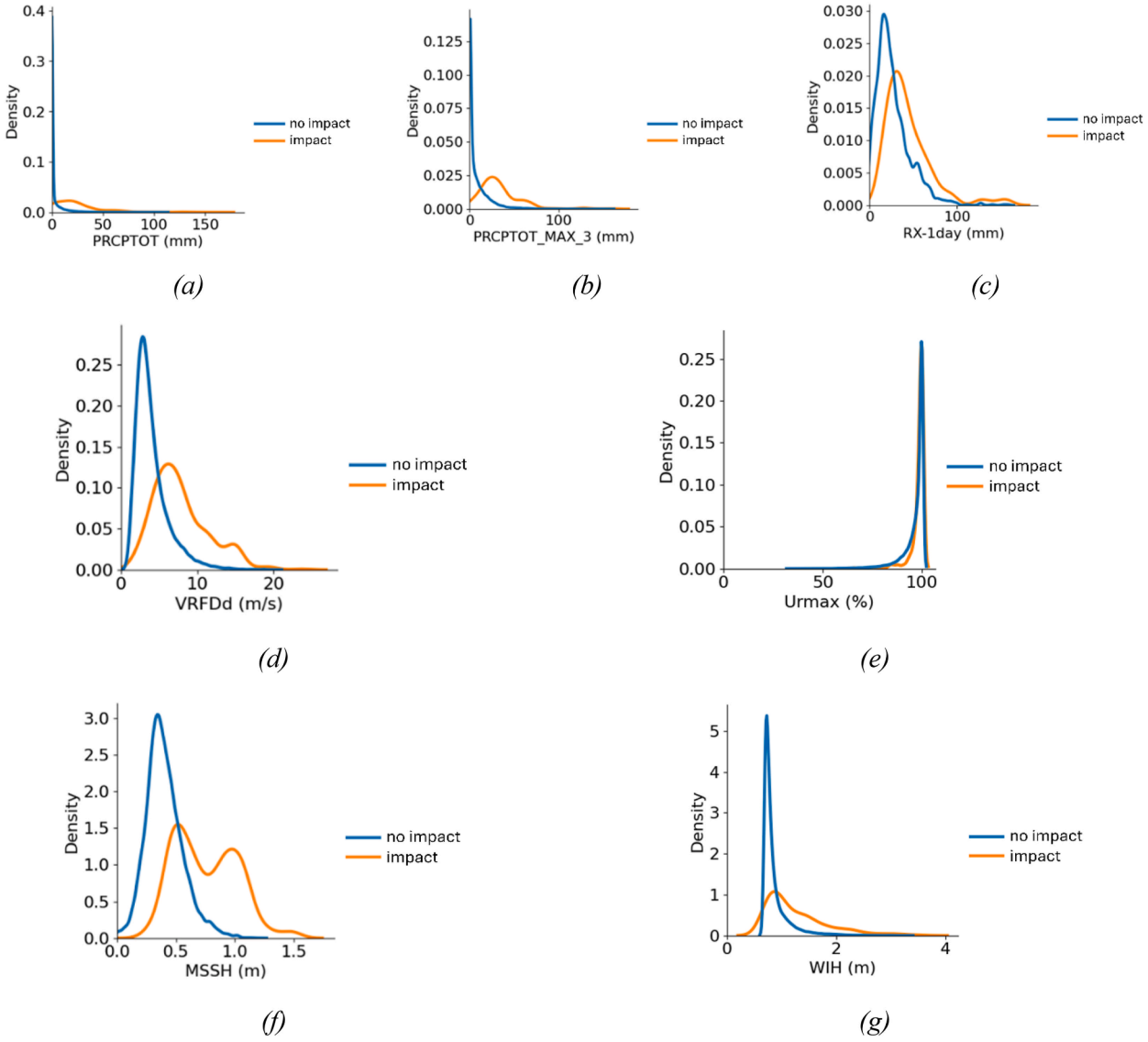


Fig. 7. Distribution of the indicators (introduced in Table 1) of samples with and without impacts: the blue curves represent the distribution of hazards in samples with no impact recorded, while the orange curves represent the distribution of hazards in samples with impact recorded. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.3. Machine learning results

In this section, the results of the application of the ML algorithms and the corresponding feature importance are illustrated. In Section 4.3.1 the results of the binary classification are shown. The goal of this paragraph is to show that the algorithms work properly, and hence can be used to estimate the risk score of new data. Section 4.3.2 on the other hand describes the feature importance obtained by the application of the best performing algorithm, with the aim of understanding the phenomena that generates the occurrence of impacts in different areas of the Veneto coast.

4.3.1. Classification results

The objective of the ML application is the estimate of a risk score for each sample; the risk score is calculated by the implemented ML algorithms using the climate-relate hazard indicators (i.e., atmospheric, and marine drivers of impact) as inputs.

The results obtained by the three algorithms are shown in Fig. 9, where each point corresponds to a sample in the input data (i.e., a couple of date and municipality). In particular, blue points represent the samples associated to an impact and orange points samples associated to no impact. The vertical position of the point represents the risk score: if a point has a high risk-score, it is more likely to be affected by an impact.

In the training process, each algorithm determines its parameter by minimizing the error between the calculated and the measured outcome (i.e., by trying to move all the blue points to the top of the plot, and all the orange points to the bottom), while in the testing

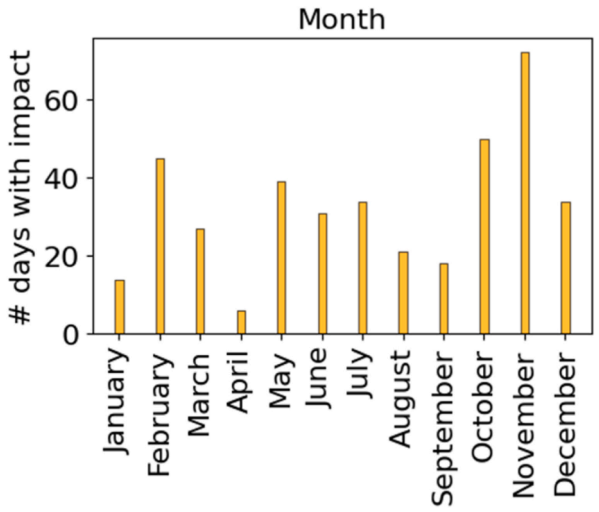


Fig. 8. Histogram representing the number samples associated with impacts per calendar month during the 2009–2019 timeframe.

phase the information on the impact is only used to measure the quality of the results. Ideally, blue points in both training and in testing should be placed as close as possible to 1 and all orange points to 0. The results of Support Vector Classifier (SVC) and Multi-Layer Perceptron (MLP) are quite similar, with most orange points very close to 0 and most blue points on the upper side of the plot. Compared to these algorithms, in the Random Forest (RF) results the separation between blue and orange points is less sharp.

The calculated weighted log losses for the train and test dataset respectively are 0.33 and 0.4 for the MLP algorithm, 0.42 and 0.59 for the RF, 0.36 and 0.45 for SVC. As the quality of the results is similar between the training and the test, the algorithm is not overfitting the data.

For some applications (e.g., early warning systems) it may be useful to decide, based on the value of the risk score, if the sample is at high risk of impact or not. To do that, we need to set a threshold, such that all the samples that have a risk score higher than this value is considered at high risk, while samples with a risk score lower than this threshold are not. The value of this threshold can be chosen by the end-user. If a conservative measure is needed (i.e., if it is important to flag every sample that can be reasonably associated to an impact) a low value for the threshold will be chosen. The drawback of this choice will be a high number of false positives (i.e., samples flagged as high risk that did not present any impact). On the other hand, if some false negatives are acceptable (i.e., days not flagged as high risk that were eventually affected by an impact), the number of false positives can be reduced.

In the frame of this study a threshold was selected by maximizing the value of the weighted F1 score of the MLP algorithm, which is a standard metric used on binary classification. The weighted F1 score is maximum and presents a plateau between 0.4 and 0.6, i.e., its value is almost constant for this range of threshold values. The value of 0.5 was set as threshold, as it is located in the middle of the plateau.

The confusion matrix, calculated using the threshold of 0.5 on the test dataset (Fig. 10), illustrates the performances of the algorithms. The confusion matrix describes how the samples are divided between True Negatives (samples correctly labelled as low risk), False Positives (samples incorrectly labelled as high risk). False Negatives (samples incorrectly labelled as high risk), True Positives (samples correctly labelled as high risk). In each box the total number of samples of the test set that fell in each category is reported, together with the percentage of samples over the total number of samples, and the percentage of samples over the total samples with the same true label.

As a result, MLP and SVC algorithms shows only 8 False Negatives from a total of 161 days with impacts, against the 14 False Negatives detected by RF. Moreover, all algorithms incorrectly label approximately 9 % of negative samples as high risk (False Positives), and between 5 and 9 % of positive samples as low risk.

Table 4 summarizes the values of the classic metrics obtained on the test set of the threshold of 0.5. The first columns report the values of the non-weighted metrics, where the importance of positive and negative samples in the calculation is the same; as expected, the precision, and consequently the F1 score calculated are small, due to the huge number of false positives previously discussed. In the second half of the table, the values of weighted metrics, calculated by weighting each sample with a weight inversely proportional to the cardinality of its class (calculated on the test set), show satisfactory results for all the metrics. The difference between weighted and non-weighted values of precision and F1 score is due to the skewness of the input data, and to the choice of giving more importance to positive samples, as they are outnumbered by the negative samples. The use of weighted metrics is more coherent with the choice of the loss function described in Section 3.3.1 and with the goal of the application.

These results confirm a similar good performance for SVC and MLP, and a slightly worse performance of RF.

In terms of computational cost, the three algorithms show very different performances. To train the algorithm the MLP algorithm takes 6s, the RF algorithm takes 12s, and the SVC algorithm takes 289s on a laptop computer (processor: Intel(R) Core(TM) i7-10875H CPU @ 2.30 GHz 2.30 GHz; RAM 160 GB). It must be noted, however, that computational cost was not an issue given, and no

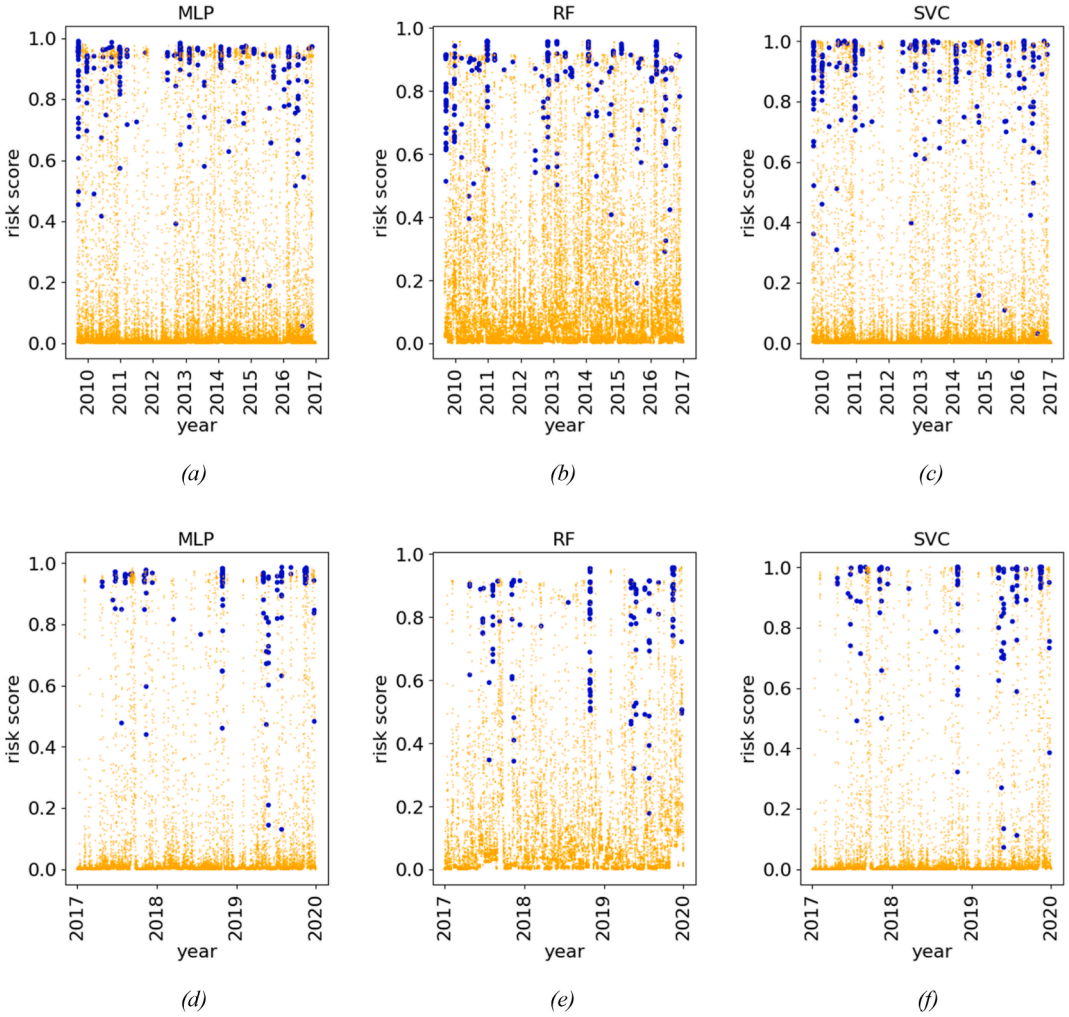


Fig. 9. Predicted risk score on the train dataset for MLP (a), RF (b), SVC (c) algorithms and predicted risk score on the test dataset for MLP (d), RF (e), SVC (f) algorithms. Blue point represents samples associated with impacts, and orange point represents samples not associated with impacts. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

effort was devoted to reducing it, or to find faster implementations of the routines, and the computational cost are here reported for completeness, without the assumption that these considerations can be generalised.

4.3.2. Feature importance

The feature importance, introduced in Section 3.3.3, is aimed at identifying the variables that influence the occurrence of impacts along the coastal municipalities of the Veneto region. It consists in calculating the metric function when a feature is randomly shuffled: the more the value of the metric increases, the more important the shuffled feature is. It can consider as a sensitivity analysis, as it evaluates the impact that the correct value of each parameter has on the quality of the result.

In Section 4.3.1 it was shown that SVC and MLP algorithms produced similar results (with the MLP algorithm being more computational effective) which outperformed the RF algorithm. For this reason, of the three ML algorithms trained, validated and tested (i.e., RF, SVC, and MLP) within the 2010–2019 timeframe, only the results for the MLP algorithm are shown in Fig. 11-a.

The most important factors triggering impacts in the investigated case study are indicators of *maximum sea surface height* (i.e., MSSH), *total precipitation* (i.e., PRCPTOT), and *wind intensity* (i.e., VRFDd). These results agree with the analysis of conditional distributions, detailed in Fig. 7. In particular, Bora and Sirocco wind jets expose the coastal area of the Veneto region to strong winds, heavy rainfall, strong sea storms and flooding, of low frequency but extremely intense [74]. Moreover, the presence of several indicators of precipitation leads to a high level of redundancy, thus reducing the importance of each feature. If a feature shows different conditional distributions according to the presence or absence of impacts, the same feature will have high importance in the corresponding machine learning algorithm.

Furthermore, in accordance with the characterization of the coastal municipalities presented in Section 4.1, the feature importance for each cluster of municipalities has been calculated to understand how exposure and vulnerability play a key role in determining the drivers of impacts.

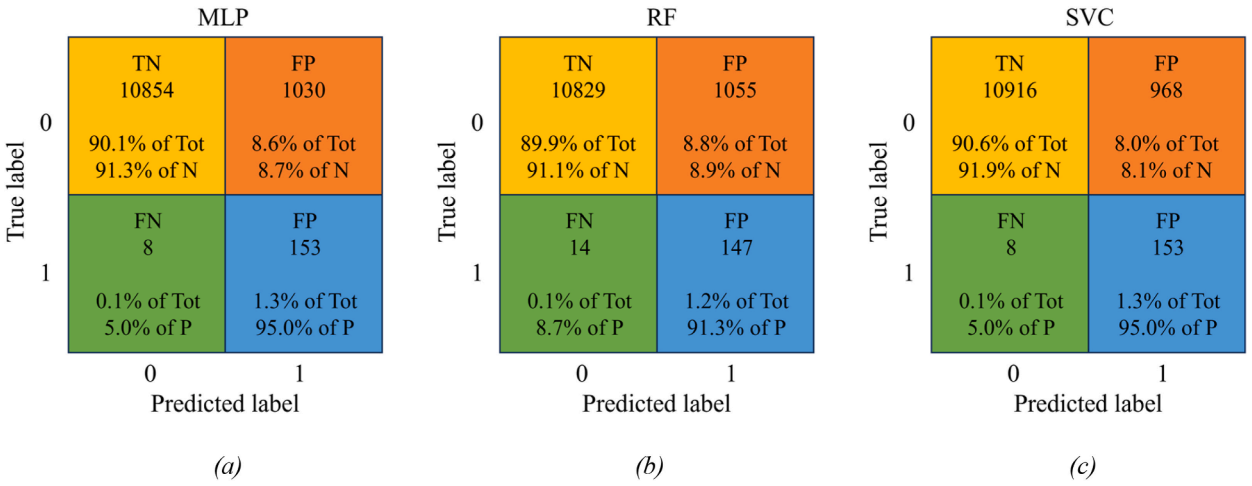


Fig. 10. Confusion matrix for a threshold of 0.5 for MLP (a), RF (b), SVC (c) algorithms. In each confusion matrix the top left square (true label = 0, predicted label = 0) contains the number of True Negatives (TN). The top right square (true label = 0, predicted label = 1) contains the number of False Positives (FP). The bottom left square (true label = 1, predicted label = 0) contains the number of False Negatives (FN). Finally, the bottom right square (true label = 1, predicted label = 1) contains the number of True Positives (TP). In each square it is reported the total number of samples that fell in the corresponding category, the percentage of samples over the total number of samples (Tot), and the percentage of samples over the total samples with the same true label (i.e., N for negative samples, and P for positive samples, respectively).

Table 4
Test set metrics for the implemented ML algorithms – i.e., Multi-Layer Perceptron (MLP), Random Forest (RF) and Support Vector Classifier (SVC).

	Not weighted metrics				Weighted metrics			
	Precision	Recall	F1 score	Accuracy	Weighted Precision	Weighted Recall	Weighted F1 score	Weighted accuracy
MLP	0.13	0.95	0.23	0.91	0.94	0.95	0.94	0.93
RF	0.12	0.91	0.22	0.91	0.93	0.91	0.92	0.91
SVC	0.14	0.95	0.24	0.92	0.94	0.95	0.95	0.94

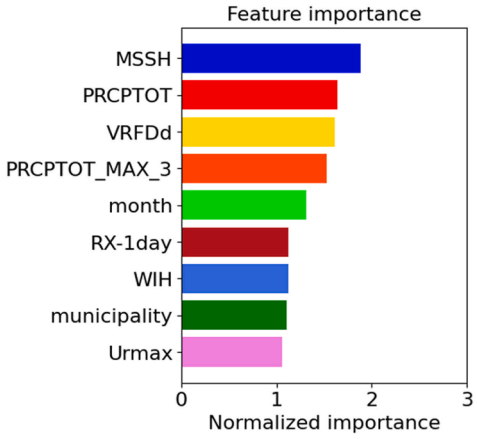
In *Group 1*, composed of the coastal municipalities of San Michele al Tagliamento, Caorle and Eraclea, the main driver of impact is the total precipitation, which dominates the importance ranking, as shown in Fig. 11-*b*. This behaviour is corroborated by Ref. [74] where several extreme events occurred in the area have been analysed. The study highlights how summer and autumn seasons are characterized by marked convective activities, with precipitation systems also affecting the lowland area that faces the Adriatic coast. As described in Fig. 6, the low permeability that characterizes the soil of these municipalities could exacerbate the effects of the precipitations, hence contributing to the occurrence of impacts.

Group 2 (i.e., Jesolo, Cavallino-Treporti and Chioggia) and Venice (*Group 3*) present the highest number of impacts, possibly because of the size (Venice) and of the high percentage of *anthropic land use* in both groups (Fig. 6). The impacts are mainly governed by extreme sea-level related events (Fig. 11-*c* and Fig. 11-*d*). These phenomena are exacerbated by the combined effect of sea-level rise and subsidence [53], that makes these areas more impacted (as detailed in Section 4.1). Besides that, although Group 2 is influenced by the intense precipitation coming from North (as seen for *Group 1*), Venice is famously affected by winds (i.e., Bora and Scirocco) which represents the major cause of *Acqua Alta* (i.e., high tide).

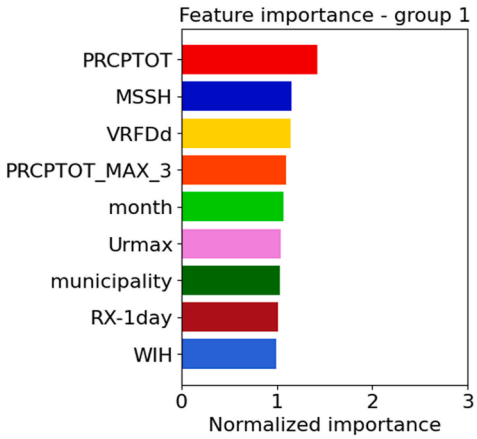
Finally, impacts recorded in *Group 4* (i.e., Rosolina, Porto Viro, Porto Tolle and Ariano nel Polesine) are equally triggered by winds and sea-level related events (Fig. 11-*e*). These, together with the low elevation of the area (Fig. 6), have historically made the river Po Delta area a famous hotspot for saltwater intrusion, especially during dry seasons, exposing the vast cultivated fields [75]. However, the importance of sea-level and precipitation is lower compared to the previous groups of municipalities due to the limited availability of impact data (as discussed in Section 2.2.3), as the impact dataset is constrained by the *'Stato di crisi'* records reported by the Veneto region. This outcome may be biased by the low rate of urbanisation and population density, hence by the low vulnerability, of the municipalities included in *Group 4* (described in Section 4.1): the low vulnerability of these municipalities leads to only the most serious hazards being recorded to the exclusion of less serious ones. As it is very likely that most severe hazards are characterised by the compounding presence of precipitation, storm surge and wind, it is reasonable to expect that the feature importance of these indicators are similar.

5. Discussion and conclusions

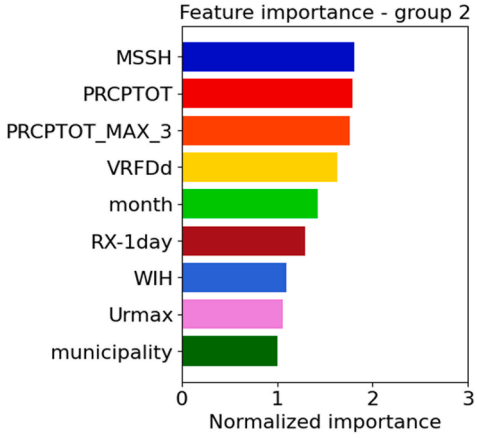
The study was developed to create an algorithm able to estimate, given a set of environmental indicators, a daily risk score for each coastal municipality of the Veneto region, and to understand the extreme events that caused the occurrence of impacts, through the analysis of feature importance in the reference timeframe (2009–2019).


Feature importance
MSSH
PRCPTOT
VRFDd
PRCPTOT_MAX_3
month
RX-1day
WIH
municipality
Urmax
0
1
2
3
Normalized importance

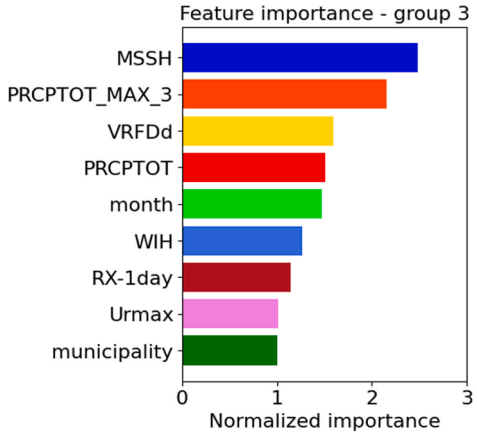
(a)


Feature importance - group 1
PRCPTOT
MSSH
VRFDd
PRCPTOT_MAX_3
month
Urmax
municipality
RX-1day
WIH
0
1
2
3
Normalized importance

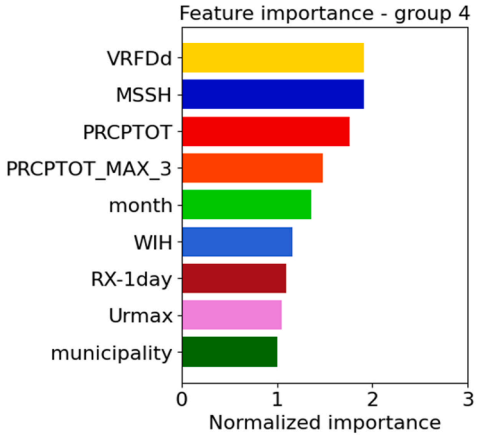
(b)


Feature importance - group 2
MSSH
PRCPTOT
PRCPTOT_MAX_3
VRFDd
month
RX-1day
WIH
Urmax
municipality
0
1
2
3
Normalized importance

(c)


Feature importance - group 3
MSSH
PRCPTOT_MAX_3
VRFDd
PRCPTOT
month
WIH
RX-1day
Urmax
municipality
0
1
2
3
Normalized importance

(d)


Feature importance - group 4
VRFDd
MSSH
PRCPTOT
PRCPTOT_MAX_3
month
WIH
RX-1day
Urmax
municipality
0
1
2
3
Normalized importance

(e)

Fig. 11. Feature importance obtained for the whole Veneto coastal area (a) and for each group of municipalities: (b) is related to Group 1, which includes San Michele al Tagliamento, Caorle and Eraclea; (c) Group 2 covers the lagoonal municipalities of Jesolo, Cavallino-Treporti and Chioggia; (d) Group 3 is entirely devoted to Venice; and Group 4 represented in (e) contain the Rovigo municipalities of Rosolina, Porto Vito, Porto Tolle and Ariano nel Polesine.

Compared with other available decision support systems for climate change adaptation in coastal areas (e.g., DSS-DESYCO and THESUS; [76,77]), the novelty of the Machine Learning (ML) model developed within this application is that, to the authors' knowledge, for the first time Artificial Intelligence (AI) methodologies are used to estimate the risks as a result of extreme weather events, along the coastal municipalities of the Veneto region. Furthermore, although the algorithms developed for this application are simple, the study offers insight into how ML models can be employed for environmental and multi-risk assessment under climate change. Therefore, the developed model represents an early prototype decision support tool underpinning climate change risk assessment and the definition of adaptation strategies at the regional scale.

After a preliminary analysis of the input dataset, several ML strategies for the estimation of the daily risk score have been considered: the SVC and the MLP algorithms provided similar results, which were satisfactory in terms of performances and excluded the risk of overfitting, meaning that the developed tools are reliable and ready to be used.

The MLP algorithm was used to calculate the relative importance that each type of hazard plays in the creation of impacts (i.e., feature importance). This analysis was performed also per group of municipalities, highlighting how the vulnerability and exposure characteristics of each group can amplify or mitigate the occurrence of impacts. This analysis shows that southern coastal areas (i.e., province of Rovigo), characterized by a larger natural area, experienced less impacts, whereas more urbanized and more densely populated municipalities (e.g., Venice, Jesolo, Chioggia) were the most affected.

One of the challenges of the implementation of the ML algorithm was the skewness of the input dataset: most input samples are associated with a negative label, while approximately 1 % of the samples are associated to a positive label. The skewness of the data is a very common issue in several science fields [78], and it is almost always addressed using weights or oversampling in the training, validation, and testing phase, and in the definition of the loss function and the metric. In this work the same strategy has been applied. Drawbacks of this choice are the high number of false positives in the results; the advantage, on the other hand, is the ability to keep the minority class in the right account, which is of course the main priority of the work. Several binary classification metrics were used to analyse the results: non-weighted metrics such as precision and F1 score are strongly influenced by the high number of false positives, while weighted metrics are not. This result is not ideal, but this problem is intrinsic in all binary classification problems on skewed dataset. It is important to understand the meaning of these results: even if the risk score of a sample assumes a high value, there is the possibility that no impact would occur on that day, i.e. that the sample is a false positive. Stakeholders and local administrations that would like to use the results or the algorithm for adaptation and mitigation planning should be aware of this limitation.

Even if the first results obtained applying the ML model for the coastal area of the Veneto region are promising, some improvements could be achieved with the use of a more detailed impact dataset: since the preliminary data analysis, it was clear that the main limitation of this study is the scarcity and low accuracy of the impact dataset: this issue is critical, as the results of any machine learning tool are at best as good as the input data. The impact list retrieved from the '*Stato di crisi*' reports from Civil Protection, and cross-checked and integrated with analysis of local newspapers still presents some main problematics: namely incompleteness and inaccuracy (it is possible that some impacts are not present), lack of the number of events per municipality on a given date, lack of economical information on the impact, lack of typology of the impact, lack of the precise location, and hence vulnerability and exposure, of the target.

In order to partially overcome these issues, a robust procedure that includes the use of simple algorithms, feature analysis, a manual analysis of the outliers, an error analysis on the train set, and the use of regularization coefficients to reduce the risk of overfitting was implemented. However, it is important to remember that no processing techniques can make up for the low data quality.

Much more meaningful results could have been achieved with very similar algorithms with more impact information available. If the entity of the impacts (like the economic loss), or at least the number of impacts on a given day in each municipality were available, it would be possible to normalize the impacts over the municipality size or the population. The actual impossibility of normalizing the impacts has a major consequence on the analysis, as any measure of the contribution of the municipality, with its exposure and vulnerability characteristics, in the determination of impacts, concludes that bigger and more populated municipalities are more affected. If detailed information on the impact were available (e.g., location, entity, type of impact, magnitude, information on the structure/activity affected by the impact), local exposure and vulnerability factors could be considered in the analysis: geographic, hydrographic and geological information, on local or municipality level, together with infrastructural networks maps, will also be included in the analysis as indicators. Similarly, socioeconomic dynamics (e.g., urbanization, population growth, migrations, tourism, economical activities) could be integrated to consider the effect of their interaction with climatic drivers in exacerbating the risk and vulnerability towards disasters. Moreover, if the characteristics of the impacts would be known, the plethora of assessment endpoints could be extended, including different types of damages (e.g., structural damage, flood damage, water quality alteration, infrastructure disruption).

For these considerations, it is strongly recommended that local authorities and the scientific community would cooperate for the creation of a standardised impact dataset. The existence of such a dataset would improve massively the quality of multi-risk studies and would help the scientific communities to share their know-how.

The same algorithm can be applied to other geographic areas, with the same or different indicators and assessment endpoints, provided that the algorithm is retrained or fine-tuned on local data.

This tool is also meant to be the first step of a methodology for the estimation of the number of impacts per year over the Veneto coastal area, which will be applied to modelled future projections of marine and atmospheric indicators to estimate the impact of climate change in future timeframes.

Future development of the methodology may include the implementation of ML tools such as Recurrent Neural Networks or Graph neural networks, which would be able to take into account spatial and temporal dependencies in the analysis. The development of the algorithm followed the guidelines for robustness and explainability of artificial intelligence developed by JRC [79]: the developed algorithm is robust, as it has been tested on new data to avoid the risk of overfitting; results are reproducible, as it has been built in an ad hoc environment and any random seed is hardcoded, and it is transparent, as mainly python libraries are used, and all the parameters are summarized in this paper. The algorithm is also explainable, as a function that calculates the feature importance has been implemented. No sensitive data has been used to train and test the algorithm. The algorithm is not developed to take autonomous decisions that would affect people's lives, but it has been designed to be used by researchers and local authorities as an instrument able to increase the understanding of the phenomena underlying the occurrence of impacts in different area of the Veneto coast, and hence helping the mitigation planning process. It can also be used as an early warning system when applied to short term weather forecasts. It is recommended to retrain or fine tune the algorithm regularly with new data, to consider changes in the dynamics, or new mitigation structures in place.

CRediT authorship contribution statement

Maria Katherina Dal Barco: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Margherita Maraschini:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Davide Ferrario, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Ngoc Diep Nguyen:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Silvia Torresan:** Writing – review & editing, Validation, Supervision, Conceptualization. **Sebastiano Vascon:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Andrea Critto:** Validation, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

All of the reported work in the attached manuscript is original and the manuscript has not been previous published in whole or in part. I also have read and abided by the statement of ethical standards for manuscripts submitted to this journal. All authors have seen the manuscript and approved its submission to the *International Journal of Disaster Risk Reduction*.

Moreover, all the authors declare that they have no known competing financial interests or personal relationships that could had influenced the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The research leading to these results has been funded by the AdriaClim (Climate change information, monitoring and management tools for adaptation strategies in Adriatic coastal areas, https://programming14-20.italy-croatia.eu/web/adriaclim) project, funded by the Interreg ITA-CRO Programme 2014–2020 (Project ID: IT-HR 10252001). The authors gratefully acknowledge Eng. Alessandro De Sabbata and the Veneto region's Post-Emergency Disaster Events Management Office (*Direzione Gestione Post Emergenze Connesse ad Eventi Calamitosi e altre attività commissariali*), as well as Dr. Elena Allegri for their significant support in retrieving information on the extreme weather events that have affected the Veneto region in the last decades. A special thanks to our colleagues Olinda Rufo and Heloisa Labella Fonseca for their contribution in the final phase of the data pre-processing.

References

[1] Fox-Kemper, B., H.T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J.-B. Sallée, A.B.A. Slangen, and Y. Yu, 2021: Ocean, Cryosphere and Sea Level Change. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1211–1362, doi: 10.1017/9781009157896.011.

[2] Intergovernmental Panel on Climate Change, H. Lee, K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W.W.L. Cheung, S.L. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, Z. Zommers, Synthesis report of the IPCC Sixth Assessment Report (AR6) - Longer report (2023). https://report.ipcc.ch/ar6syr/pdf/IPCC_AR6_SYR_LongerReport.pdf.

[3] U. Schickhoff, M. Bobrowski, J. Böhner, B. Bürzle, R.P. Chaudhary, L. Gerlitz, J. Lange, M. Müller, T. Scholten, N. Schwab, Climate change and Treeline dynamics in the Himalaya, Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya (2016), https://doi.org/10.1007/978-3-319-28977-9_15.

[4] S.R. Weiskopf, M.A. Rubenstein, L.G. Crozier, S. Gaichas, R. Griffis, J.E. Halofsky, K.J.W. Hyde, T.L. Morelli, J.T. Morisette, R.C. Muñoz, A.J. Pershing, D.L. Peterson, R. Poudel, M.D. Staudinger, A.E. Sutton-Grier, L. Thompson, J. Vose, J.F. Weltzin, K.P. Whyte, Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States, Sci. Total Environ. xxxx (2020), https://doi.org/10.1016/j.scitotenv.2020.137782.

[5] A. AghaKouchak, F. Chiang, L.S. Huning, C.A. Love, I. Mallakpour, O. Mazdiyasni, H. Moftakhari, S.M. Papalexiou, E. Ragno, M. Sadegh, Climate extremes and compound hazards in a Warming world, Annu. Rev. Earth Planet Sci. 48 (2020) 519–548, https://doi.org/10.1146/annurev-earth-071719-055228.

[6] L. Sorokin, The experience of disaster risk reduction and economic losses reduction in Malaysia during the water crisis 1998 in the Context of the Next El Nino strongest on record maximum 2015, in: C. MalS, R. Singh, Huggel (Eds.), Climate Change, Extreme Events and Disaster Risk Reduction, 2018, https://doi.org/10.1007/978-3-319-56469-2_16.

[7] B.N. Goswami, V. Venugopal, D. Sengupta, M.S. Madhusoodanan, P.K. Xavier, Increasing Trend of extreme Rain events over India in a warming environment, Science 314 (5804) (2006) 1442–1445, https://doi.org/10.1126/science.1132027.

[8] A. Vecere, M. Martina, R. Monteiro, C. Galasso, Satellite precipitation–based extreme event detection for flood index insurance, Int. J. Disaster Risk Reduc. 55 (2021) 102108, https://doi.org/10.1016/j.ijdrr.2021.102108.

[9] J. Zscheischler, O. Martius, S. Westra, E. Bevacqua, C. Raymond, R.M. Horton, B. van den Hurk, A. AghaKouchak, A. Jézéquel, M.D. Mahecha, D. Maraun, A.M. Ramos, N.N. Ridder, W. Thiery, E. Vignotto, A typology of compound weather and climate events, Nat. Rev. Earth Environ. 1 (2020) 333–347, https://doi.org/10.1038/s43017-020-0060-z.

[10] F. Zennaro, E. Furlan, C. Simeoni, S. Torresan, S. Aslan, A. Critto, A. Marcomini, Exploring machine learning potential for climate change risk assessment, Earth Sci. Rev. 220 (2021) 103752, https://doi.org/10.1016/J.EARSCIREV.2021.103752.

[11] A. Bolle, L. das Neves, S. Smets, J. Mollaert, S. Buitrago, An impact-oriented early warning and bayesian-based decision support system for flood risks in zeebrugge harbour, Coast Eng. 134 (October 2017) (2018) 191–202, https://doi.org/10.1016/j.coastaleng.2017.10.006.

[12] M.K. Dal Barco, E. Furlan, H.V. Pham, S. Torresan, K. Zachopoulos, N. Kokkos, G. Sylaios, A. Critto, Multi-scenario analysis in the Apulia shoreline: a multi-tiers analytical framework for the combined evaluation and management of coastal erosion and water quality risks, Environ. Sci. Pol. 153 (2024) 103665, https://doi.org/10.1016/j.envsci.2023.103665.

[13] Ó. Ferreira, T.A. Plomaritis, S. Costas, Effectiveness assessment of risk reduction measures at coastal areas using a decision support system: findings from Emma storm, Sci. Total Environ. 657 (2019) 124–135, https://doi.org/10.1016/j.scitotenv.2018.11.478.

[14] W.S. Jäger, E.K. Christie, A.M. Hanea, C. den Heijer, T. Spencer, A Bayesian network approach for coastal risk analysis and decision making, Coast Eng. 134 (January) (2018) 48–61, https://doi.org/10.1016/j.coastaleng.2017.05.004.

[15] H.V. Pham, M.K. Dal Barco, M.P. Shahvar, E. Furlan, A. Critto, S. Torresan, Bayesian network analysis for shoreline dynamics, coastal water quality, and their related risks in the Venice littoral zone, Italy, J. Mar. Sci. Eng. 12 (1) (2024) 139, https://doi.org/10.3390/jmse12010139.

[16] T.A. Plomaritis, S. Costas, Ó. Ferreira, Use of a Bayesian Network for coastal hazards, impact and disaster risk reduction assessment at a coastal barrier (Ria Formosa, Portugal), Coast Eng. 134 (September) (2018) 134–147, https://doi.org/10.1016/j.coastaleng.2017.07.003.

[17] M. Sanuy, J.A. Jiménez, Probabilistic characterisation of coastal storm-induced risks using Bayesian networks, Nat. Hazards Earth Syst. Sci. 21 (1) (2021) 219–238, https://doi.org/10.5194/nhess-21-219-2021.

[18] S.J. Park, D.K. Lee, Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms, Environ. Res. Lett. 15 (9) (2020) 94052, https://doi.org/10.1088/1748-9326/aba5b3.

[19] S. Janizadeh, S.C. Pal, A. Saha, I. Chowdhuri, K. Ahmadi, S. Mirzaei, A.H. Mosavi, J.P. Tiefenbacher, Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future, J. Environ. Manag. 298 (2021) 113551, https://doi.org/10.1016/j.jenvman.2021.113551.

[20] M.S. Rana, C. Mahanta, Spatial prediction of flash flood susceptible areas using novel ensemble of bivariate statistics and machine learning techniques for ungauged region, Nat. Hazards 115 (2023) 947–969, https://doi.org/10.1007/s11069-022-05580-9.

[21] J. Sampurno, V. Vallaeys, R. Ardianto, E. Hanert, Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta, Nonlinear Process Geophys. 29 (2022) 301–315, https://doi.org/10.5194/npg-29-301-2022.

[22] R.C. Chen, C. Dewi, S.W. Huang, R.E. Caraka, Selecting critical features for data classification based on machine learning methods, Journal of Big Data 7 (1) (2020), https://doi.org/10.1186/s40537-020-00327-4.

[23] S. Terzi, S. Torresan, S. Schneiderbauer, A. Critto, M. Zebisch, A. Marcomini, Multi-risk assessment in mountain regions: a review of modelling approaches for climate change adaptation, J. Environ. Manag. 232 (February) (2019) 759–771, https://doi.org/10.1016/j.jenvman.2018.11.100.

[24] C. Huntingford, E.S. Jeffers, M.B. Bonsall, H.M. Christensen, T. Lees, H. Yang, Machine learning and artificial intelligence to aid climate change research and preparedness, Environ. Res. Lett. 14 (2019) 124007, https://doi.org/10.1088/1748-9326/ab4e55.

[25] A. Tilloy, B.D. Malamud, H. Winter, A. Joly-Laugel, A review of quantification methodologies for multi-hazard interrelationships, Earth Sci. Rev. (September) (2019) 102881, https://doi.org/10.1016/j.earscirev.2019.102881.

[26] N. Malik, U. Ozturk, Rare events in complex systems: understanding and prediction, Chaos 30 (2020) 90401, https://doi.org/10.1063/5.0024145.

[27] M. de Coste, Z. Li, Y. Dibike, Machine-learning approach for predicting the occurrence and timing of mid-winter ice breakups on canadian rivers, Environ. Model. Software 152 (2022), https://doi.org/10.1016/j.envsoft.2022.105402.

[28] V. Rao, R. Maulik, E. Constantinescu, M. Anitescu, A machine-learning-based importance sampling method to compute rare event probabilities, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12142 LNCS (2020) 169–182, https://doi.org/10.1007/978-3-030-50433-5_14/FIGURES/.

[29] S. Hegelich, Decision Trees and random forests: machine learning techniques to classify rare events, European Policy Analysis 2 (1) (2016) 98–120, https://doi.org/10.18278/EPA.2.1.7.

[30] A. Dogan, D. Birant, Machine learning and data mining in manufacturing, Expert Syst. Appl. 166 (2021) 114060, https://doi.org/10.1016/J.ESWA.2020.114060.

[31] S. Luca, P. Karsmakers, K. Cuppens, T. Croonenborghs, A. van de Vel, B. Ceulemans, L. Lagae, S. van Huffel, B. Vanrumste, Detecting rare events using extreme value statistics applied to epileptic convulsions in children, Artif. Intell. Med. 60 (2) (2014) 89–96, https://doi.org/10.1016/J.ARTMED.2013.11.007.

[32] N. Yousefi, M. Alaghband, I. Garibay, A Comprehensive Survey on Machine Learning Techniques and User Authentication Approaches for Credit Card Fraud Detection (2019), https://doi.org/10.48550/arXiv.1912.02629.

[33] J. Coffinet, J.N. Kien, Detection of rare events: a machine learning toolkit with an application to banking crises, The Journal of Finance and Data Science 5 (4) (2019) 183–207, https://doi.org/10.1016/J.JFDS.2020.04.001.

[34] D.S. Hain, R. Jurowetzki, Introduction to rare-event predictive modeling for inferential statisticians-A hands-on application in the prediction of breakthrough patents. *Financial econometrics: bayesian analysis, quantum uncertainty, and related topics, 427*(ECONVN 2022, Studies in Systems, Decision and Control) (2020), https://doi.org/10.1007/978-3-030-98689-6_5.

[35] P. Ruol, L. Martinelli, C. Favaretto, T. Pinato, F. Galiazzo, S. Patti, U. Anti, R. Piazza, P. Simonin, G. Selvi, Gestione integrata della Zona costiera studio e monitoraggio per la definizione degli interventi di difesa dei litorali dall'erosione nella Regione Veneto-linee guida, 2016.

[36] A. Barbi, A. Cagnati, G. Cola, F. Checchetto, A. Chiaudani, A. Crepaz, I. Delillo, M. L, M. G, M. P, P. Sg, Atlante climatico del Veneto. Precipitazioni-Basi informative per l'analisi delle correlazioni tra cambiamenti climatici e dinamiche forestali nel Veneto, 2013.

[37] A. Barbi, M. Monai, R. Racca, A.M. Rossa, Recurring features of extreme autumnall rainfall events on the Veneto coastal area, Nat. Hazards Earth Syst. Sci. 12 (8) (2012) 2463–2477, https://doi.org/10.5194/nhess-12-2463-2012.

[38] A. Bezzi, S. Pillon, D. Martinucci, G. Fontolan, Inventory and conservation assessment for the management of coastal dunes, Veneto coasts, Italy, J. Coast Conserv. 22 (2018) 503–518, https://doi.org/10.1007/s11852-017-0580-y.

[39] Regione Veneto, Analysis of ICZM practice in Italy, Veneto Region (2012). https://sistemavenezia.regione.veneto.it/content/pianificazione-spaziale-marittima.

[40] P. Ruol, L. Martinelli, C. Favaretto, Vulnerability analysis of the Venetian littoral and adopted mitigation strategy, Water 10 (8) (2018) 984.

[41] S. Torresan, A. Critto, M. Dalla Valle, N. Harvey, A. Marcomini, Assessing coastal vulnerability to climate change: comparing segmentation at global and regional scales, Sustain. Sci. 3 (1) (2008) 45–65, https://doi.org/10.1007/s11625-008-0045-1.

[42] Legambiente, Salviamo le coste italiane (2013). http://issuu.com/legambienteonlus/docs/doss_consumo_di_suolo_costiero.It is a report issued by Legambiente, a NGO that, in more than 35 years of activity, has organised many environmental monitoring activities on air quality, sea pollution and marine litter (https://www.legambiente.it/).

[43] MATTM, Piano Nazionale di Adattamento ai Cambiamenti Climatici PNACC, 2017.

[44] D. Camuffo, Four centuries of documentary sources concerning the sea level rise in Venice, Climatic Change 167 (3–4) (2021) 1–16, https://doi.org/10.1007/s10584-021-03196-9.

[45] C. Cavalieri, Extreme-city-territories. Coastal geographies in the Veneto region, Journal of Urbanism: International Research on Placemaking and Urban Sustainability (2020) 1–19.
[46] S. Torresan, A. Critto, J. Rizzi, A. Marcomini, F.J. Mendez, S. Leschka, P. Fraile-Jurado, Assessment of coastal vulnerability to climate change hazards at the regional scale: the case study of the North Adriatic Sea, Nat. Hazards Earth Syst. Sci. 12 (7) (2012).
[47] J. Rizzi, V. Gallina, S. Torresan, A. Critto, S. Gana, A. Marcomini, Regional Risk Assessment addressing the impacts of climate change in the coastal area of the Gulf of Gabes (Tunisia), Sustain. Sci. 11 (3) (2016) 455–476, https://doi.org/10.1007/s11625-015-0344-2.
[48] C. Ferrarin, A. Roland, M. Bajo, G. Umgiesser, A. Cucco, S. Davolio, A. Buzzi, P. Malguzzi, O. Drofa, Tide-surge-wave modelling and forecasting in the Mediterranean Sea with focus on the Italian coast, Ocean Model. 61 (2013) 38–48, https://doi.org/10.1016/j.ocemod.2012.10.003.
[49] C. Ferrarin, P. Lionello, M. Orlić, F. Raicich, G. Salvadori, Venice as a paradigm of coastal flooding under multiple compound drivers, Sci. Rep. 12 (2022) 5754, https://doi.org/10.1038/s41598-022-09652-5.
[50] G. Umgiesser, M. Bajo, C. Ferrarin, A. Cucco, P. Lionello, D. Zanchettin, A. Papa, A. Tosoni, M. Ferla, E. Coraci, S. Morucci, F. Crosato, A. Bonometto, A. Valentini, M. Orlić, I.D. Haigh, J.W. Nielsen, X. Bertin, A.B. Fortunato, R.J. Nicholls, The prediction of floods in Venice: methods, models and uncertainty (review article), Nat. Hazards Earth Syst. Sci. 21 (8) (2021) 2679–2704, https://doi.org/10.5194/nhess-21-2679-2021.
[51] L. Cavaleri, M. Bajo, F. Barbariol, M. Bastianini, A. Benetazzo, L. Bertotti, J. Chiggiato, C. Ferrarin, F. Trincardi, G. Umgiesser, The 2019 flooding of Venice and its implications for future predictions, Oceanography 33 (1) (2020) 42–49.
[52] A. Crespi, S. Terzi, S. Cocuccioni, M. Zebisch, B. Julie, H.-M. Füssel, Climate-related hazard indices for Europe, European Topic Centre on Climate Change Impacts, Vulnerability and Adaptation (ETC/CCA) (2020), https://doi.org/10.25424/cmcc/climate_related_hazard_indices_europe_2020.
[53] P. Lionello, D. Barriopedro, C. Ferrarin, R.J. Nicholls, M. Orlic, M. Reale, G. Umgiesser, M. Vousdoukas, D. Zanchettin, Extremes floods of Venice: characteristics, dynamics, past and future evolution, Nat. Hazards Earth Syst. Sci. (2020) 1–34 https://doi.org/10.5194/nhess-2020-359, November.
[54] D. Zanchettin, S. Bruni, F. Raicich, P. Lionello, F. Adloff, A. Androsov, F. Antonioli, V. Artale, E. Carminati, C. Ferrarin, V. Fofonova, R. Nicholls, S. Rubinetti, A. Rubino, G. Sannino, G. Spada, R. Thiéblemont, M. Tsimplis, G. Umgiesser, S. Zerbini, Review article: sea-level rise in Venice: historic and future trends, Natural Hazards and Earth System Sciences Discussions (November) (2020) 1–56, https://doi.org/10.5194/nhess-2020-351.
[55] S. Tarquini, I. Isola, M. Favalli, F. Mazzarini, M. Bisson, M.T. Pareschi, E. Boschi, TINITALY/01: a new triangular irregular network of Italy, Ann. Geophys. 50 (2007) 407–425. http://www.annalsofgeophysics.eu/index.php/annals/article/view/4424.
[56] S. Tarquini, L. Nannipieri, The 10 m-resolution TINITALY DEM as a trans-disciplinary basis for the analysis of the Italian territory: current trends and new perspectives, Geomorphology 281 (2017) 108–115, https://doi.org/10.1016/j.geomorph.2016.12.022.
[57] P. Lionello, M.B. Galati, E. Elvini, Extreme storm surge and wind wave climate scenario simulations at the Venetian littoral, Phys. Chem. Earth (2012) 40–41 https://doi.org/10.1016/j.pce.2010.04.001 (Parts A/B/C), 86–92.
[58] S. Raveh-Rubin, H. Wernli, Large-scale wind and precipitation extremes in the Mediterranean: aclimatological analysis for 1979–2012, Q. J. R. Meteorol. Soc. 141 (681) (2015) 2404–2417, https://doi.org/10.1002/qj.2531.
[59] I. Goodfellow, Y. Bengio, A. Courville, Deep Feedforward Networks, 2016. https://mnassar.github.io/deeplearninghandbook/slides/06_mlp.pdf.
[60] G. King, G.H.E. Langche Zeng, J. Alt, J. Freeman, K. Gleditsch, G. Imbens, C. Manski, P. Mccullagh, W. Mebane, J. Nagler, B. Russett, K. Scheve, P. Schrodt, M. Tanner, R. Tucker, S. Bennett, P. Huth, L. Zeng, Logistic regression in rare events data, Polit. Anal. 9 (2) (2001) 137–163, https://doi.org/10.1093/OXFORDJOURNALS.PAN.A004868.
[61] N. Cahyana, S. Khomsah, A.S. Aribowo, Improving imbalanced dataset classification using oversampling and gradient boosting, in: 2019 5th International Conference on Science in Information Technology: Embracing Industry 4.0: towards Innovation in Cyber Physical System, 2019, pp. 217–222 https://doi.org/10.1109/ICSITECH46713.2019.8987499, ICSITech 2019.
[62] S. Ertekin, J. Huang, C.L. Giles, Active learning for class imbalance problem, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07, 2007, pp. 823–824, https://doi.org/10.1145/1277741.1277927.
[63] J. He, M.X. Cheng, Weighting methods for rare event identification from imbalanced datasets, Frontiers in Big Data 4 (108) (2021), https://doi.org/10.3389/FDATA.2021.715320/XML/NLM.
[64] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, A.S.R. Pritzel, T. Ewalds, F. Alet, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, J. Stott, O. Vinyals, S. Mohamed, P. Battaglia, GraphCast: Learning skillful medium-range global weather forecasting (2022), https://doi.org/10.48550/arXiv.2212.12794.
[65] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830.
[66] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2 (5–6) (1991) 183–197, https://doi.org/10.1016/0925-2312(91)90023-5.
[67] L.E.O. Breiman, Random Forests (2001) 5–32.
[68] C.-C. Chang, C.-J. Lin, Training v-support vector classifiers: theory and algorithms, Neural Comput. 13 (9) (2001) 2119–2147, https://doi.org/10.1162/089976601750399335.
[69] X. Ying, An overview of overfitting and its solutions, J. Phys. Conf. 1168 (2) (2019), https://doi.org/10.1088/1742-6596/1168/2/022022.
[70] P. Baldi, Autoencoders, unsupervised learning, and Deep architectures, Proceedings of Machine Learning Research 27 (2012) 37–49. https://proceedings.mlr.press/v27/baldi12a.html.
[71] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation Forest, Eighth IEEE International Conference on Data Mining, 2008, pp. 413–422, https://doi.org/10.1109/ICDM.2008.17.
[72] A. Ng, Machine Learning, Cs229., Stanford University, Stanford.Edu, 2021. http://cs229.stanford.edu/.
[73] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[74] P. Stocchi, S. Davolio, Intense air-sea exchanges and heavy orographic precipitation over Italy: the role of Adriatic sea surface temperature uncertainty, Atmos. Res. 196 (2017) 62–82, https://doi.org/10.1016/j.atmosres.2017.06.004.
[75] P. Tarolli, J. Luo, E. Straffelini, Y.-A. Liou, K.-A. Nguyen, R. Laurenti, R. Masin, D'Agostino Vicenzo, Saltwater intrusion and climate change impact on coastal agriculture, PLOS Water 2 (4) (2023) E0000121, https://doi.org/10.1371/journal.pwat.0000121.
[76] S. Torresan, A. Critto, J. Rizzi, A. Zabeo, E. Furlan, A. Marcomini, DESYCO: a decision support system for the regional risk assessment of climate change impacts in coastal zones, Ocean Coast Manag. 120 (2016) 49–63, https://doi.org/10.1016/j.ocecoaman.2015.11.003.
[77] B. Zanuttigh, D. Simcic, S. Bagli, F. Bozzeda, L. Pietrantoni, F. Zagonari, S. Hoggart, R.J. Nicholls, THESEUS decision support system for coastal risk management, Coast Eng. 87 (2014) 218–239, https://doi.org/10.1016/j.coastaleng.2013.11.013.
[78] M. Maalouf, T.B. Trafalis, Robust weighted kernel logistic regression in imbalanced and rare events data, Comput. Stat. Data Anal. 55 (1) (2011) 168–183, https://doi.org/10.1016/J.CSDA.2010.06.014.
[79] IPCC, 2023: 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 [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, 184 pp., doi: 10.59327/IPCC/AR6-9789291691647.