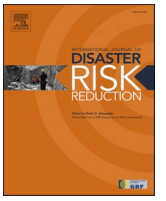
International Journal of Disaster Risk Reduction 117 (2025) 105139

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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



Predicting traffic volumes on road infrastructures in the context of multi-risk assessment frameworks

Paolo Intini *, Gianni Blasi, Francesco Fracella, Antonio Francone, Roberto Vergallo, Daniele Perrone

Department of Innovation Engineering, University of Salento, sp 6 Lecce-Monteroni, 73047, Lecce, Italy

ARTICLE INFO

Keywords:
Traffic volume
Multi-risk assessment
Heavy vehicles
Prediction
Machine learning

ABSTRACT

In multi-risk assessment frameworks involving road infrastructures, measures of exposure to natural hazards include traffic volumes. However, traffic counts are usually collected through traffic counter/radar stations which only cover a small part of the road network. In this study, country-wide Annual Average Daily Traffic (AADT) prediction models based on Italian data were developed to provide direct risk exposure measures both in terms of traffic volumes (continuous variable) and traffic volume discrete classes, using province-/municipality-related geographic, socio-economic and road-related variables as predictors. To ease transferability and applicability of the models, only publicly available predictors were selected. Traditional statistical techniques (generalized linear models for predicting traffic values and ordered logistic models for traffic classes) and Machine Learning (ML) approaches (XGBoost for both regression and classification problems) were used. Both the direct estimation of traffic volumes and the classification into traffic ranges provided satisfactory results in terms of goodness-of-fit and predictive accuracy metrics. Results show that population, occupation, tourism, density, number of lanes, urban environment, complex intersections and ring roads were generally related to an increase in traffic volumes. Distance from large cities and accessibility metrics are inversely related to traffic instead. The application of the XGBoost ML approach proved to be more accurate than traditional approaches only for heavy vehicles. It was discussed how the obtained models can be used as input modules for overall multi-risk assessment frameworks involving road infrastructures.

1. Introduction

Transport infrastructures are critical systems in modern society and their loss of functionality may lead to major consequences on the local economy and general population wellness, in terms of fatalities or injuries, economic loss and impact on public confidence (as stated by e.g., Ref. [1]). The high vulnerability of transport infrastructures to natural hazards was evidenced in recent studies [2]. Among possible causes of their high vulnerability can be included the broad extensiveness, the high density and the fast expansion of such networks requiring the realization of specific assets (tunnels, bridges, etc.) that might significantly affect the risk related to natural hazards if not adequately designed. Simplified risk assessment procedures mainly consider the convolution of three main features, namely hazard, exposure and vulnerability [3]. The definition and computation of such features varies depending on the

* Corresponding author.
E-mail addresses: paolo.intini@unisalento.it (P. Intini), gianni.blasi@unisalento.it (G. Blasi), francesco.fracella@unisalento.it (F. Fracella), antonio.francone@unisalento.it (A. Francone), roberto.vergallo@unisalento.it (R. Vergallo), daniele.perrone@unisalento.it (D. Perrone).

https://doi.org/10.1016/j.ijdrr.2024.105139
Received 28 August 2024; Received in revised form 20 December 2024; Accepted 21 December 2024
Available online 27 December 2024

specific risk considered, although they are generally assumed as the probability of occurrence of a hazardous event, the value of the asset affected and its resilience to the event [4–6].

The management of a great number of infrastructures in their multiple operational and environmental conditions requires a multi-hazard approach [7], to analyze not only individual risks but also their interactions and concomitance, as also emphasized by international risk reduction frameworks [8]. In fact, considering tropical cyclones, earthquakes and flooding, Koks et al. [9] estimated that almost one third of worldwide transport infrastructure assets are exposed to at least one hazard. An overview of the state-of-the art on multi-risk assessment frameworks related to road infrastructures is presented as follows. After that, the attention is particularly focused on how exposure to risks can be defined and measured at different scales in such frameworks, discussing the relevance of estimating traffic volumes on road infrastructure networks and the related scientific background specific to the problem. In fact, it is here anticipated that the main motivation of this study is to fill a research gap in terms of modelling road traffic volumes in the specific context of multi-risk assessment frameworks. The specific contributions of this study will be further specified at the end of this section.

1.1. Background on multi-risk assessment related to road infrastructures

Multi-risk methodologies accounting together for different hazards, either account for the mutual influence of all hazards in a considered area or consider simultaneously each single hazard and its cascading events [10]. Hence, the "chain of events" can be defined as a direct or indirect result of an initial event [11]. The triggering event considered may change depending on the typical hazard of the region examined, such as cyclones in tropical zones [12–14], or avalanches in colder countries [15,16]. One of the most common cascading events are earthquake-induced landslides. The peculiarity of these events is their dependence on both the triggering event and geological, geomorphological, hydrological, environmental conditions of the affected region [17].

The adoption and the study of multi-risk approaches were increasingly encouraged during the last two decades by critical infrastructures stakeholders. This is due to several reasons, such as the increasing interconnection of society or the higher frequency of extreme events/disasters related to climate change [18]. In particular, among infrastructures, particular attention was paid to road networks. In fact, the most relevant road infrastructures in an area (i.e., at the regional, state or continental level) are critical assets, especially when particularly exposed to natural hazards [19]. Some multi-risk assessments related to road infrastructures were described in previous literature (e.g., Ref. [20–22]). In Clarke and O'Brien [23], earthquakes, landslides and floods, alongside their interaction and cascading events, were considered to perform stress tests on the European infrastructure network. A spatio-temporal database was also developed including steady geographic information and the event time-variable data.

A typical interaction that should be taken into account in multi-risk assessment frameworks is the one between road networks and structures which are directly part of them (such as bridges) or closely located (like buildings on the roadside). For example, Argyroudis et al. [24] implemented the European Commission's SYNER-G methodology to englobe the hazard uncertainty, fragility models of road system assets, alongside interdependencies such the collapsed building debris disposition into a single probabilistic systemic risk analysis framework. The methodology was applied to several case studies, showing high frequency of road service loss due to debris blockage in historic city centers, where buildings have higher seismic vulnerability. One of the main causes of service interruption is bridge structural damage or collapse [25]. The behavior of bridges under a single risk is usually studied using fragility curves, which express the probability of damage as event intensity increases. The fragility functions, defined for different damage states and several hazards, are combined to propose multi-risk fragility surfaces [26–28]. Other multi-risk assessment frameworks for bridges have also been proposed in previous research (e.g., Ref. [29]). The bridge damage or collapse events cause direct economic losses, and indirect losses related to road closure [30]. This latter effect is important on main roads, mainly because of the travel delay. According to Abarca et al. [31], considering both direct and indirect costs in Average Annual Losses allows a more effective prioritization of retrofit interventions.

Another peculiar interaction pertains to road infrastructures in coastal areas. This interaction is crucial given that sea level rise driven by climate change poses a significant threat to coastal regions around the world (e.g., Ref. [32]), causing increased flooding, erosion, and habitat loss. As sea levels continue to rise, the risk for road infrastructures in coastal areas becomes more pronounced, being exposed to frequent flooding and structural damage (e.g., Ref. [33]). Adaptation measures to deal with sea level rise are crucial to the resilience of coastal infrastructures. These measures include (i) planned retreat of structures and infrastructures, (ii) accommodating natural system effects by adjusting land use of the coastal zone, crop types, and flood resilience measures, (iii) protecting the zone with soft or hard barriers such as nourished beaches, dunes, or seawalls, and (iv) attacking by building seaward and upwards to claim new land for economic development [34]. In the context of planned retreat, relocation of road infrastructures to less exposed areas plays a crucial role. Strategic road setback can mitigate the risk of traffic disruptions due to flooding, erosion damage and ground failure due to liquefaction (see e.g., Ref. [24]) by ensuring the proper functioning of the road system in the coastal environment. In addition, relocating roads that are currently in high-risk coastal areas and have significant traffic volume can reduce maintenance costs and extend the service life of the infrastructure.

1.2. Exposure to risks and road infrastructures

Regardless of the particular natural hazard considered or the assessment framework, the estimation of exposure to risk is crucial. Exposure is generally defined [35] as "*the presence of people, livelihoods, species or ecosystems, environmental functions, services, and resources, infrastructure, or economic, social, or cultural assets in places and settings that could be adversely affected.*" Hence, in case of risk assessments, especially in case of natural disasters which may potentially endanger large areas, different measures of exposure to risks can be considered, also depending on the scale of the problem. Nevertheless, when assessments are conducted at a vast scale, exposure

can be measured through general variables. For example, Satta et al. [36] evaluated the exposure to coastal risk of the whole Mediterranean coastal area based only on population density and land cover type, thus only indirectly capturing the presence of infrastructures.

However, as already anticipated, there is also a consistent body of research on risk analysis procedures specifically dedicated to transport infrastructures in a given area, considering single or multiple hazards (see e.g., Ref. [9,19,37]). In case of multi-risk assessment frameworks involving road infrastructures, it is crucial to consider, among other variables, traffic volumes (see e.g., Ref. [15,31,38]). In this context, traffic volumes can be used as a direct measure of exposure to calculate the fatality risk on a given road section (see e.g. Ref. [15], in case of avalanches) or to estimate loss metrics in the overall risk assessment, i.e. economic losses derived from increased travel times for changes in the travel path due to a link with damaged structures [31,38].

Elaborating more on the cited studies, traffic volumes are intended as an indicator of the importance of the considered road segments within the infrastructural system. The disruption of a segment with high traffic volume related to natural hazard-induced damage leads to major consequences on several social assets, such as freight distribution, connection to and between strategic facilities and so on. Additionally, since these consequences do not only involve individuals living or interacting with disaster-affected area whatsoever (which is a likely scenario in case of main roads with heavy traffic), also non-direct economic losses are obtained from a disruptive event. The disruption of a main road carrying high traffic volumes causes a re-distribution of traffic among undamaged roads, increasing their congestion and slowing down services which may be fundamental, particularly in case of emergency. Concluding, a combination of the natural hazard, the vulnerability of the road network (e.g. due to the presence of a high number of old bridges) and the possible direct and non-direct economic losses, represent a reliable mean for disaster risk assessment of a specific area. In this scenario, the traffic volume alone may not be a comprehensive indicator of the exposure; however, it surely represents a feature expressing the amount of goods and services under threat in case of a disruptive event (e.g. [39]). Consequently, amplification factors depending on traffic volumes on specific roads may be applied to single-hazard derived risk indexes in multi-risk frameworks.

Apart from being used in risk assessment frameworks including different types of hazards, traffic volumes are also related to intrinsic road-related risks. In this sense, traffic crashes are a major contributing factor to deaths worldwide, especially among young people [40]. While they may depend on a wide range of possible causes and being related to several concurring factors, crashes can be generally modelled as a function of traffic volume, a typical measure of the users' exposure to road-related risk [41]. However, natural disasters may alter the normal operating conditions of road infrastructures, especially in case of evacuating population [42–45]. Drivers are directly affected by damage or functionality disruption due to hazardous events. For example, in case of wildfires, drivers' speeds can be influenced by the produced smoke [46]. These effects, combined with the variations in the traffic volumes in the network, may increase the probability of incidents [47–49]. It is evident that, in these scenarios, road users are exposed to other risks, in addition to the intrinsic road-related risks.

1.3. Background on traffic volume predictions

Traffic data are usually collected through monitoring stations and are used to compute synthetic indicators such as the Annual Average Daily Traffic (AADT), hourly peak volumes or the percentage of heavy vehicles. Traffic prediction models (i.e., typically AADT prediction models) can be obtained through linear regressions (see e.g., Ref. [50–55]), spatial models (see e.g., Ref. [56–60]) and/or Machine Learning (ML) techniques (see e.g., Ref. [61–64]).

Traditional regression-based models (e.g., the ordinary least-square linear regression – OLS –), associate multiple independent variables, such as road classes or number of lanes, to the AADT (e.g., [60,64]). Besides these models which are widely used, other authors employed spatial models such as Geographically Weighted Regression or kriging, in which the relation among variables take different weights according to their distance from the studied case ([56,58]; [60]; [65]). Several types of ML techniques have also been experimented for predicting traffic volumes (e.g., Random Forest and the Support Vector Regressions by Ref. [64]; or decision rules in Ref. [63]). In some cases, ML approaches for data extraction are also coupled with the availability of aerial images to estimate AADT [62,66]. Regardless of the specific modelling approach, most of the previous studies used geographic and socio-economic factors such as distance from cities/urban areas, land use, accessibility, population, density, occupation and income. Road-related variables are usually limited to road type, speed limits and the number of lanes, while the intersection types or the presence of alternative routes are not explicitly modelled. Moreover, in none of the reviewed studies, land use includes information about being in the coastal area or not.

It is important to note that, in all the reviewed studies except for Song et al. [59] and Sfyridis and Agnolucci [67], the composition of the traffic flow (i.e., the percentage of heavy vehicles) is disregarded in the predictions. However, apart from the usual importance of this information for road pavement management, the heavy vehicle volume can be important for fatigue life predictions of bridges [68,69], which is not negligible in the context of multi-risk assessments. Structural vulnerability assessment of bridges is indeed a key aspect for the evaluation of network resilience [30,70]. Several approaches were adopted in last years for structural health monitoring of bridges, based on satellite data [71] or operational modal analysis [72,73]. In most cases, the accuracy in predicting traffic loads is fundamental when performing structural vulnerability assessment, particularly for road bridges subjected to degradation, as evidenced in the literature (e.g. Ref. [74,75]). Moreover, the distinction between light and heavy vehicle volumes can be important as well to determine indirect economic losses within risk assessments, such as in the framework proposed by Ishibashi et al. [38].

Moreover, several studies in previous research have focused on low-volume/minor roads in the network [51,53,54,56,57,60,63], given that traffic counts are usually sparser on them than on major roads. However, given the aims of this study, a comprehensive prediction including major road classes is sought here. This prediction is intended to be integrated into overall multi-risk assessment

frameworks, which would be applicable universally and encompass all relevant major arterials. For the same reason, this study also investigates the possibility of directly predicting traffic classes (e.g., from low to high; such approach was only found in Ref. [53], based on logistic models), to be integrated with index-based risk models.

1.4. Objectives of the study

Road traffic volumes are a key input for multi-risk assessment frameworks that include road infrastructures. However, the number of traffic count stations is not comparable to the number of roads in a network. To address the problem of unavailability of traffic volumes for all the roads in a network, models to predict both light and heavy annual average daily traffic volumes, which could be integrated into multi-risk assessment frameworks, are developed in this study. Models are based on the Italian publicly available traffic volume dataset related to State roads. Both a traditional regression model and a ML approach were used to predict traffic volumes, by using geographic, socio-economic and detailed road-related factors as independent variables.

The work discussed herein is part of a multi-disciplinary research project aimed at developing a multi-risk assessment framework combining traffic-related risks to seismic and coastal flood hazards for road infrastructures. While the interaction between roads and the seismic/coastal hazards could be relevant worldwide, this research project addresses the Italian territory, where a very dense road network (168.129 km -[76]-) is placed in context with high seismic hazard and an extensive coastal development (length >8,000 km), causing high road network exposure to several natural hazard-related risks (e.g., Ref. [77]).

Given the considered context, the main contributions to the state of the art of this article are reported as follows:

- enlarging the body of knowledge on the relationships between traffic volumes and socio-economic/road-based predictors, investigating their geographic variability;
- specifically exploring the above-reported relationships for heavy vehicle traffic volumes, which were mostly disregarded in previous studies;
- developing predictive models representing a trade-off between rigorousness and flexibility, i.e.,
  - o by including detailed geographic, socio-economic and road-related variables (trying to enlarge the set of potential predictors with respect to previous literature), which however can be easily collected by other researchers and practitioners and potentially transferred to other contexts;
  - o by using both ML techniques (the eXtreme Gradient Boosting technique, never used in previous research for this particular scope) and traditional regression approaches which can be easily implemented in overall multi-risk assessment frameworks;
  - o by modelling the dependent variables both as quantitative measures (AADT traffic volumes) and as traffic classes (i.e., low to high), which again can be easily integrated in multi-risk assessments.

Hence, this study wants to enlarge the pool of available traffic prediction models, developing a tool calibrated for the European context (almost all the previous studies were based on other areas and the geographic variability can be particularly relevant for the problem at hand) and, in particular, for Italy. However, this study is specifically dedicated to developing traffic prediction models which can be integrated into general multi-risk assessment frameworks by: separately predicting light and heavy vehicle volumes, introducing class-based dependent variables, referring to a country-wide major road network and accurately selecting easily collectable and potentially transferrable dependent variables, including some potential predictors relevant for multi-risk assessments usually not considered in previous research (e.g., coastal zone, presence of alternatives, accessibility and tourism indexes).

The methods used in this research are described in Section 2. Results from prediction models are then presented in Section 3 and discussed in Section 4. Finally, Section 5 draws the conclusions.

2. Methods

In this section, the set of independent variables which will be used to predict traffic volumes is presented, together with the relative data collection methods and characteristics. Traditional statistical methods and Machine Learning -ML- approaches used for prediction purposes are then described.

2.1. Traffic volume dataset

The dataset of traffic volumes used in this study is the publicly available Average Annual Daily Traffic (AADT) dataset published online by the Italian National Road Agency (ANAS),[1] which manages the network of nationally relevant roads ("State roads") and some motorways in Italy.

Published data regard 1268 traffic counter stations spread across all Italian regions, excluding only Trentino-Alto Adige, in a 10-year period from 2013 to 2022. Annual average volumes are separately reported for heavy vehicles (vehicles having size corresponding to payloads of 3500 kg, trucks and buses) and light vehicles (all other vehicles). The geographic localization of these stations is presented in Fig. 1.

[1] https://www.stradeanas.it/it/le-strade/osservatorio-del-traffico/dati-traffico-medio-giornaliero-annuale. Lastly accessed on December 15th, 2023.

2.1.1. Data treatment

As declared in the explanatory note to the dataset, the AADT of both light and heavy vehicles is computed as follows:

$$AADT_k = \frac{\sum_{j=1}^{d}\sum_{i=1}^{p}(V_5)_{i,j,k}}{d} \quad [1]$$

where:

$AADT_k$ = bi-directional Annual Average Daily Traffic related to the k-component of traffic, that is light vehicles ($k = \mathrm{l} \rightarrow AADT_l$) or heavy vehicles ($k = \mathrm{h} \rightarrow AADT_h$);

$(V_5)_{i,j,k}$ = recorded traffic volume of the k-component in the valid i-th 5-min period of the valid j-th day of the year;

p = number of valid 5-min period in each day of the year ("valid" means that data were correctly recorded within the time period), if all 5-min periods are valid: $p = 288$;

d = number of valid days during the year ("valid" indicates days for which p is at least 282, that means at least 98 % valid 5-min periods); if $\mathrm{d} < 365/2 \sim 183$, $AADT_k$ is considered a missing data for that year.

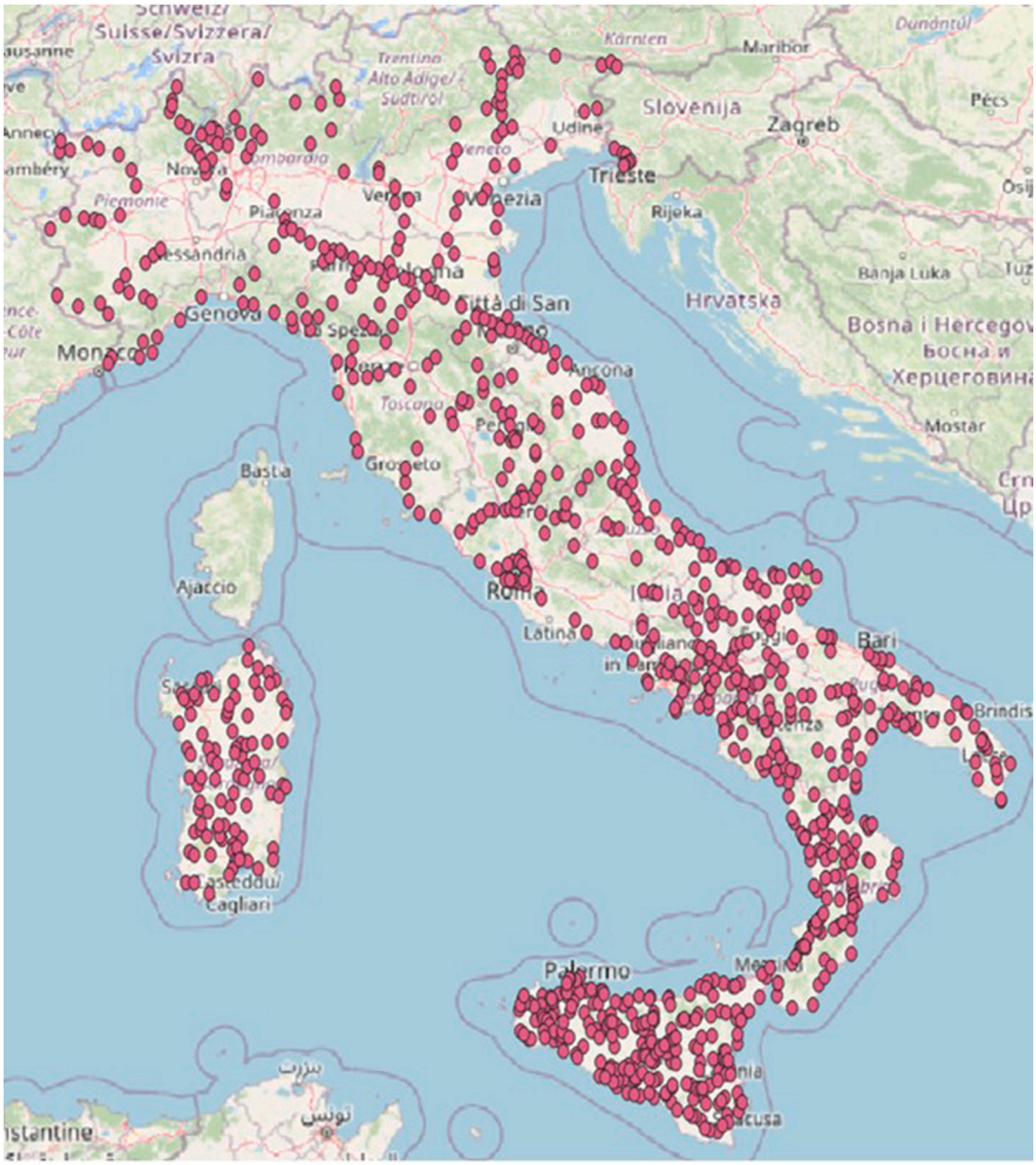


Fig. 1. Position of all Italian ANAS traffic counter stations in the dataset on OpenStreetMap base map.

The mean $AADT_k$ over the entire dataset of available traffic stations was computed for each year from 2013 to 2022. The variation of the mean $AADT_k$ over the years is reported in Fig. 2.

For what concerns light vehicles, different phases can be noted: 1) a first steep 2013–2014 increase, 2) a soft and almost uniform increase in the 2014–2017 period, 3) a sort of plateau in the 2017–2019 period, 4) an evident drop in 2020 due to the Covid-19 pandemics (see also [78]) and 5) a subsequent increase up to 2022, to values comparable with 2019. On average, heavy vehicles were less variable in the 2013–2019 period (even if a 2013–2014 increase and a 2018–2019 decrease can be noted). Differently than light vehicles, the heavy vehicles drop due to the pandemics continued up to 2022.

Given the high time variability of traffic volumes, which may highly depend on the socio-economic conditions [79], it was decided to exclude the 2020–2021 period from the analysis. Given the relatively stable tendency during years before 2020 and in order to include the most extended and recent available period, a 5-year period was thus selected for further analyses: 2015–2019. Hence, the dataset was filtered to include traffic stations for which at least one yearly traffic count was available over the time span 2015–2019. Out of the initial 1268 stations, 281 were removed (because they only included counts since 2020), to form the final sample of 987 traffic stations. Only 37 % of stations have valid data for each year in the period and thus, for the remaining stations, discontinued data is available (among the total dataset of 987 stations x 5 years = 4935 counts, about 27 % yearly traffic counts are missing).

Given that macro-level issues due to the temporal variability were already addressed by removing the 2020–2021 period and that the remaining 2015–2019 period is relatively stable, the average traffic volume over the considered period (2015–2019) was calculated. This was deemed as acceptable given that: a) the aim of this study is not directed towards catching the yearly variability but in providing reliable estimates to be used in overall risk assessment frameworks, b) post-pandemic trends are still to be established to extrapolate reliable future yearly variation tendencies (after 2022).

However, to compute the average yearly volume, the micro-level yearly variability in the considered period should be considered as well. In fact, if missing data had been retained, the average volume could have been underestimated/overestimated depending on the specific year in which data were missing. For example, by looking at Fig. 2, if only 2015 and 2016 data were available for a given station, it is likely that the average traffic count would have been underestimated because traffic volumes increased, on average, in the next three years. Thus, average yearly growth rates were computed for each 1-year time interval in the period (i.e., 2015–2016, 2016–2017, etc.), based on the available data. Those rates were used to replace yearly missing counts starting from the available data. For example, if the 2016 count was missing, it was estimated by applying the calculated 2015–2016 increasing rate to the 2015 available count.

The above-explained process of data treatment, repeated for both light and heavy vehicles, led to estimate average light and heavy yearly volumes in the 2015–2019 period for each of the 987 traffic count stations in the dataset. Boxplots of light and heavy $AADT_k$ volumes (namely, $AADT_l$: mean 11871, st. dev.: 16578 vehicles/day; and $AADT_h$: mean 749, st. dev.: 1049 vehicles/day) are reported in Fig. 3. It is possible to note how traffic data are not normally distributed: they are skewed towards zero (especially heavy vehicle volumes). This can be explained by the presence of a small portion of high-volume roads in the sample, with respect to most road sections (by looking at boxplots, 75 % of $AADT_l$ are widely below 20000 vehicles/day, while 75 % of $AADT_h$ are below 1000 vehicles/day).

2.1.2. Additional pre-processing stage: traffic volume classes

Depending on the specific multi-risk assessment strategy, two alternative measures of risk exposure may be needed, i.e.:

- a specific traffic volume measure or, alternatively,
- a traffic exposure class, i.e., a range of traffic volumes.

Hence, given that this study wants to provide tools which could be generally applied within different possible risk assessment frameworks, both options were considered here. While traffic volume measures obtained from traffic counts (sub-section 2.1.1) can be directly used, a definition of traffic volume classes is needed.

Since the dataset includes traffic volume counts for all the major Italian highway classes (two-way two-lane road arterials and multi-lane divided or undivided highways, including freeways, spanning from 4 to 6 lanes), the sample of traffic counts can be regarded as a representative sample of the major road network. For this reason, instead of using a-priori traffic classification into ranges, a one-dimensional clustering algorithm was used to identify traffic classes, namely, the k-means algorithm.

The k-means algorithm is an unsupervised ML technique which shards a dataset composed of n observations into m clusters, by minimizing the variance (with respect to the cluster mean) within each cluster (see e.g., Ref. [80]). While it is not ensured to reach the global optimum, the algorithm easily finds convergence to a local optimum by using different possible heuristics. The procedure proposed by Hartigan and Wong [81] is used here. The *cluster* library, based on Kaufman and Rousseeuw [82], was used to run the k-means algorithm in *R* environment. The k-means algorithm can be repeated for different possible numbers of clusters. In this study, three methods are screened to find the optimal number of clusters: the elbow method, the silhouette and the gap statistic. For both cases of light and heavy vehicles, after having compared the results obtained from the three above-reported methods, the selected optimal number of clusters is three. Results from cluster analysis are reported in Fig. 4, where boxplots of the three clusters of both light and heavy traffic volumes are represented; and Table 1, where the main descriptive statistics are reported for each cluster.

2.2. Collected variables

The set of independent variables which will be used to predict traffic volumes is presented, together with the relative data collection

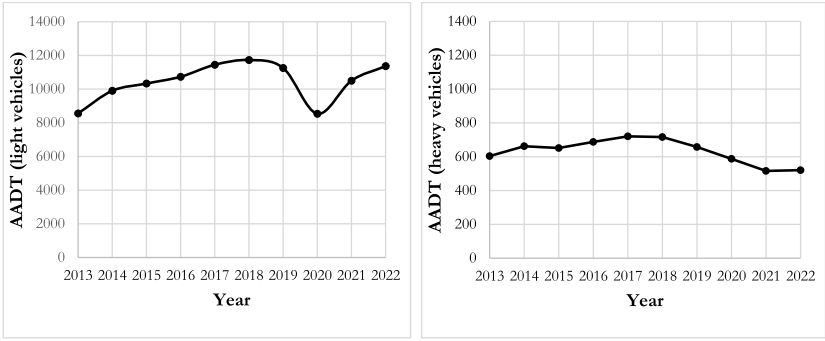


Fig. 2. Mean of the average daily traffic volumes over the 10-year period for both light vehicles (left) and heavy vehicles (right) across all available traffic counter stations in the dataset (y-axis graphical scale of the $AADT_h$ is multiplied by a factor of 10 in order to appreciate yearly variations of $AADT_h$).

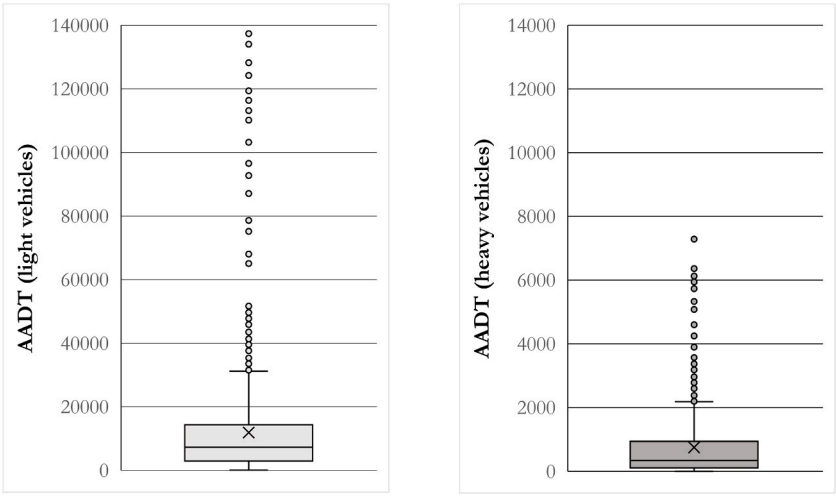


Fig. 3. Boxplots of average daily traffic volumes over the 2015–2019 period for both light vehicles (left) and heavy vehicles (right) across all available traffic counter stations in the dataset (y-axis graphical scale of the $AADT_h$ is multiplied by a factor of 10 in order to appreciate the distribution of $AADT_h$).

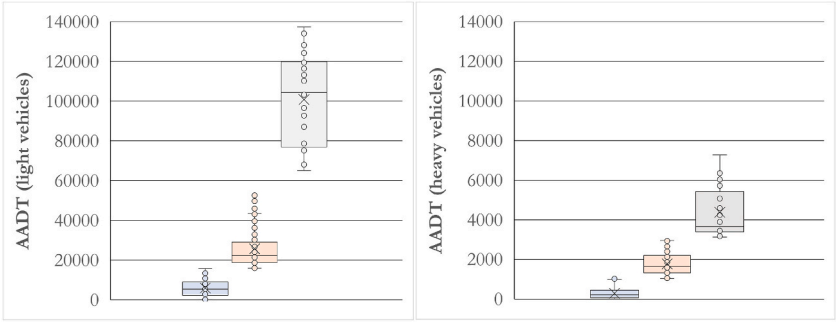


Fig. 4. Boxplots of the three clusters representing traffic volume ranges for both light (left) and heavy (right) AADT volumes (y-axis graphical scale of the $AADT_h$ is multiplied by a factor of 10 in order to appreciate the distribution of $AADT_h$ boxplots).

methods and features. Three groups of variables were considered: province-, municipality- and road-related, with respect to the traffic count station location. The main descriptive statistics regarding these variables are reported in Table 2, alongside those related to light and heavy vehicle volumes (dependent variables).

For province- and municipality- related variables, the relevant information was retrieved from the Italian National Institute of

Table 1
Descriptive statistics of the three clusters representing traffic volume ranges for both light and heavy AADT volumes.

Class	$AADT_l$ (vehicles/day)				$AADT_h$ (vehicles/day)			
	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Low (0)	5924	4190	72	15718	288	256	1	1026
Medium (1)	25607	9106	15899	53260	1775	553	1034	2962
High (2)	100804	23236	65053	137375	4387	1216	3128	7284

Statistics (ISTAT) online data[2] for the investigated period, or the most recent applicable period when not available. Population density data were obtained from the baseline population and area width data. The accessibility index measures the average travel time (in minutes) required to reach urban or logistic nodes (in particular, to reach the three closest infrastructures among ports, airports, railway stations and freeway interchanges from municipality centroids in free-flow conditions). The air traffic index measures the number of landed and boarded passengers per 100 inhabitants. The occupation index indicates the percentage of working people in the age 15–64. The tourism index indicates the number of days of residential tourism per inhabitant per year. The municipality elevation zone is based on the average elevation of the municipality area, initially divided by ISTAT into 8 classes: 0–299 m, 300–599 m, 600–899 m, 900–1199 m, 1200–1499 m, 1500–1999 m, 2000–2499 m, 2500+ m. A more parsimonious classification is used here, namely the "0 – low" zone (0–299 m), "1 - medium" zone (300–899 m), "2 - high" zone (900+ m). The alongshore and the coastal zones identified with the "1" code indicate, namely, municipalities which are directly on the seaside or at <10 km from the seaside. The urbanization index developed by ISTAT is an overall indication, based on both population and area, of the degree of urbanization of the municipality (0 – "high", 1 – "medium", 2 – "low"). Finally, the "distance from large city" variable indicates the distance (in km) between the municipality and the closest large city (having 100,000+ inhabitants) and it was calculated in a GIS environment.

Road-related variables include detailed information about the road section on which the count station is located. All information were retrieved by visually inspecting road sections by means of online tools. The road-related variables are: the number of lanes ("0": two lanes, "1": four lanes, "2": six lanes); the road environment ("0": rural, "2": urban, "1": sub-urban, that is a mostly rural environment with significant urban settlements); the particular function with respect to large urban settlements ("1": ring road of a region/province capital city, "0" if otherwise); presence of physical medians to separate carriageways ("1": yes, "0": no); intersection types along the road section ("0": at-grade intersections, "1": mix of at-grade and grade-separated intersections, "2": grade-separated intersections/junctions, "3": freeway junctions); presence of possible higher-level alternatives to the road section ("0": no comparable alternatives, "1": presence of neighboring/parallel alternative roads having similar road functions, "2": presence of alternative roads having similar road functions).

2.3. Data analysis

Traditional regression models (generalized linear models) and machine learning -ML- (eXtreme Gradient Boosting -XGBoost-technique) were used to predict traffic volumes and traffic volume ranges, as explained in the following. The choice of both approaches was based on previous research in which they were generally used alternatively to predict traffic volumes. In particular, ML approaches are increasingly used for traffic prediction purposes [83] and their outputs are comparable with other traditionally used tools, in the context of simulations [84]. Given this background and, to increase the flexibility of using different outputs from this study for multi-risk assessments depending on the particular application, traditional statistics and ML were both used, and their outputs compared.

The overall data analysis procedure is summarized in next Fig. 5.

2.3.1. Traditional regression models

Generalized linear models were first used to predict both traffic volume values and classes (ranges). In particular:

- linear regression to predict traffic volumes;
- ordered logistic regression to predict traffic volume ranges.

2.3.1.1. Theoretical background. The linear regression model is expressed as follows [85]:

$$(AADT_k)_i = \beta_0 + \sum_{j=1}^{n} \beta_j X_{i,j} + \varepsilon_i \quad [2]$$

where:

$(AADT_k)_i$ is the i-th traffic volume observation in the dataset;

β_j is the j-th model coefficient to be estimated (β_0 is the intercept, set to 0 to avoid considering negative volumes);

[2] http://dati.istat.it/. Lastly accessed on January 15th, 2024.

Table 2
Descriptive statistics of the collected variables.

Variable		Mean	St. dev.	Max.	Min.	Count	Percent (%)
Traffic volume	$AADT_l$ (vehicles/day)	11871	16578	137375	72	–	–
	$AADT_h$ (vehicles/day)	749	1049	7284	1	–	–
Province-related	Population (inhabitants)[a]	657775	692787	4258886	84441	–	–
	Area (km^2)[a]	3716	1902	7692	213	–	–
	Density (inh./km^2)	222	321	2589	37	–	–
	Accessibility index (minutes)	56	14	101	27	–	–
	Air traffic index (pax/100 inh.)	162	291	1231	0	–	–
	Occupation index (−)	52	11	73	37	–	–
	Tourism index (days/inh.)	6	6	47	0	–	–
Municipality-related	Population (inhabitants)	73342	347992	2749031	78	–	–
	Area (km^2)	135	188	1287	2	–	–
	Density (inh./km^2)	338	546	3957	1	–	–
	Elevation zone – 0 (low)	–	–	–	–	195	20
	Elevation zone – 1 (med.)	–	–	–	–	330	33
	Elevation zone – 2 (high)	–	–	–	–	462	47
	Elevation (m)	322	286	1684	0	–	–
	Alongshore zone – 0 (no)	–	–	–	–	661	67
	Alongshore zone – 1 (yes)	–	–	–	–	326	33
	Coastal zone – 0 (no)	–	–	–	–	611	62
	Coastal zone – 1 (yes)	–	–	–	–	376	38
	Urbanization index – 0 (high)	–	–	–	–	141	14
	Urbanization index – 1 (med.)	–	–	–	–	460	47
	Urbanization index – 2 (low)	–	–	–	–	386	39
	Distance from large city (km)	55	35	164	2	–	–
Road section-related	Number of lanes – 0 (2 lanes)	–	–	–	–	744	75
	Number of lanes – 1 (4 lanes)	–	–	–	–	224	23
	Number of lanes – 2 (6 lanes)	–	–	–	–	19	2
	Environment – 0 (rural)	–	–	–	–	777	79
	Environment – 1 (suburban)	–	–	–	–	59	6
	Environment – 2 (urban)	–	–	–	–	151	15
	Ring road – 0 (no)	–	–	–	–	876	89
	Ring road – 1 (yes)	–	–	–	–	111	11
	Median – 0 (no)	–	–	–	–	756	77
	Median – 1 (yes)	–	–	–	–	231	23
	Intersection type – 0 (at-grade)	–	–	–	–	555	56
	Intersection type – 1 (mix)	–	–	–	–	131	13
	Intersection type – 2 (grade-sep.)	–	–	–	–	225	23
	Intersection type – 3 (freeway)	–	–	–	–	76	8
	Higher-level alternative – 0 (no)	–	–	–	–	754	76
	Higher-level alternative – 1 (yes, parallel)	–	–	–	–	162	16
	Higher-level alternative – 2 (yes)	–	–	–	–	71	7

[a] Million inhabitants and thousands km^2 are considered for data analysis.

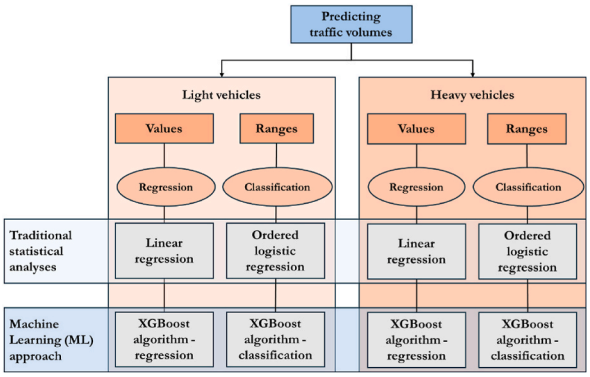


Fig. 5. Framework of the proposed data analysis procedure.

$X_{i,j}$ is the i-th observation of the j-th predictor (up to n predictors included in the model);
ε_i is the i-th value of normally distributed errors;
k subscript indicates the type of modelled traffic volumes (two separate models are estimated for light vehicles, $k = l$, and for heavy vehicles, $k = h$).

The ordered logistic regression model, in the proportional odds variant, is expressed as follows [86]:

$$logit\{P[(AADT_k)_C \le c]\} = log\frac{P[(AADT_k)_C \le c]}{P[(AADT_k)_C > c]} = \beta_{0,c} - \sum_{j=1}^{n} \eta_j X_{i,j} \quad [3]$$

where:

$(AADT_k)_C$ is the m-th traffic volume class c, ordered from the lowest to the highest volume class;

$\beta_{0,c}$ is the model intercept referred to the m-th class c;

η_j is the j-th model coefficient estimate;

$X_{i,j}$ is the i-th observation of the j-th predictor (up to n predictors included in the model);

k subscript indicates the type of modelled traffic volumes (two separate models are estimated for light vehicles, $k = l$, and for heavy vehicles, $k = h$).

The estimated coefficients in Eq. (3) (η_j) can be exponentiated to obtain odds ratios. In this case, they can be interpreted as the odds of being in a class greater than c versus lower or equal classes for a one-unit change in the continuous predictor (or a shift from 0 to 1 in the binary categorical predictor), with all other conditions being equal. Given the proportional odds model assumption, the odds ratios are the same across categories. Note that an ordered logistic model was preferred over a standard (unordered) multinomial regression, given that traffic volume ranges can be ordered from the lowest volume to the highest volume class.

2.3.1.2. Model training and evaluation. Before fitting both categories of models, a preliminary screening was run on the dataset to identify multi-collinearity among potential predictors. After this stage, some of the variables included in Table 1 were omitted (province population density, municipality population, elevation, alongshore zone, urbanization index and road median). For the same reason, after having dummy-coded all the categorical variables, the "0" category was excluded from further analyses. The final dataset on which models are built is composed of 987 items (rows) and 24 variables (columns): 4 independent variables (light and heavy AADT, both in form of numerical values and traffic classes), 9 continuous predictor variables and 11 dummy-coded categorical predictor variables.

After, the initial dataset was randomly split into a training (75 % of the initial dataset) and a test dataset (remaining 25 %), in which all outcome classes were adequately represented. Models were fitted to the training dataset. To compare potential candidate alternative models (for both linear and ordered logistic models), likelihood ratio tests (LRTs) were used to evaluate the improvements provided by additional predictors at the 5 % significance level.

The same LRTs were used to compare the final fitted models with the corresponding null models. Moreover, for linear models, the R^2 is computed to assess the goodness-of-fit. To assess in-sample predictions for ordered logistic models, accuracy and Cohen's K are computed. They are defined as follows:

$$Accuracy\ (ACC) = \frac{Correct\ predictions}{All\ predictions}(\%) \quad [4]$$

$$Cohen's\ K = \frac{ACC - p_e}{1 - p_e} \quad [5]$$

where p_e is the expected probability of agreement by chance (by using the confusion matrix for the random classification). Cohen's K is included between −1 and 1 (where 1 indicates perfect agreement). The accuracy is also tested against the No-Information Rate (NIR, largest proportion of observed classes) to assess its significance.

For linear models, the evaluation of out-of-sample predictions was based on the metrics defined as follows: root mean square error (RMSE) and related coefficient of variation (CV-RMSE), computed over the test dataset.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[(AADT_k)_{predicted} - (AADT_k)_{observed,test\ dataset}\right]_i^2}{n}} \quad [6]$$

$$CV - RMSE = \frac{RMSE}{\frac{\sum_{i=1}^{n}[(AADT_k)_{observed,test\ dataset}]_i}{n}} \quad [7]$$

For the evaluation of ordered logistic models, several metrics are obtained starting from the confusion matrix. First, the overall accuracy is computed, for the test dataset (Eq. (4)). Moreover, the following additional predictive accuracy metrics are computed for each traffic class: balanced accuracy, sensitivity, specificity, precision, recall and F1 score. They are defined as follows.

$$Precision_c = \frac{TP_c}{TP_c + FP_c}(\%) \quad [8]$$

$$Recall_c = \frac{TP_c}{TP_c + FN_c}(\%) \quad [9]$$

$$Specificity_c = \frac{TN_c}{TN_c + FP_c}(\%) \quad [10]$$

$$Balanced\ accuracy_c = \frac{Recall_c + Specificity_c}{2}(\%) \quad [11]$$

$$F1_c = 2\frac{Precision_c * Recall_c}{Precision_c + Recall_c}(\%) \quad [12]$$

where:

TP_c, FP_c, TN_c, FN_c are, namely, the number of true positive, false positive, true negative and false negative predictions related to the m-th class c.

Models were estimated in *R* environment as well as the related metrics. The *MASS* library [87] was used to perform ordered logistic regressions.

2.3.2. Machine learning approach

A parallel ML approach was used to predict traffic volumes and traffic volume ranges. In particular, the XGBoost (eXtreme Gradient Boosting) algorithm, a scalable ML framework belonging to the family of gradient tree boosting algorithms. It was selected for this study because it can be applied for both regression and classification problems and it was previously successfully used in other traffic engineering problems, with particular regard to traffic safety (see e.g., Refs. [88–90]). However, its specific application to traffic volume predictions is explored in this study for the first time, to the best of the authors' knowledge.

2.3.2.1. Theoretical background. A brief description of the XGBoost technique is reported as follows (see Ref. [91], for more details). As a gradient tree boosting algorithm, a tree ensemble model is used to predict the output variable by means of additive functions in the space of Classification and regression trees (CARTs), each function f characterized by a number of leaves N and a weight score w on each leaf l. The goal is to minimize the particular objective function O used in the XGBoost algorithm, reported as follows:

$$O = \sum\nolimits_i L(\widehat{y}_i, y_i) + \sum\nolimits_k \Omega(f_k) \quad [13]$$

where $L(\widehat{y}_i, y_i)$ is a loss function depending on the difference between the prediction $\widehat{y}_i$ and the target variable y_i, while the function Ω is a regularization term which helps in penalizing complex models, defined as follows:

$$\Omega(f) = \gamma N + \frac{1}{2}\lambda \sum_{l=1}^{N} w_l^2 \quad [14]$$

The model is additively trained by iteratively adding the function which provides the best improvement according to the previous objective function. By operating some simplifications useful to quicken the optimization process, including a second-order approximation, the following score can be computed for a tree structure s obtained in each iteration t:

$$O^t(s) = -\frac{1}{2}\sum_{l=1}^{N} \frac{\left(\sum_{i \in I_l} g_i\right)^2}{\sum_{i \in I_l} h_i + \lambda} + \gamma N \quad [15]$$

where g_i and h_i are, respectively, the first and second order gradient statistics on the loss function L and I_l is the set of instances for the leaf l.

Given that it is usually impossible to generate all the possible tree structures, the algorithm works starting by a single leaf and iteratively adding branches to the tree by selecting the best split (i.e., dividing the set of instances I into a left I_L and right I_R set) according to the following variant of the objective function:

$$O_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma \quad [16]$$

The choice of the loss function is fundamental to characterize the problem. Given the statistical methods used and described in the previous section, two different loss functions were used for both light and heavy vehicles:

- the squared error for regression, in case of traffic volumes, and
- the softmax function for multi-classification, in case of traffic volume ranges.

Model performance is influenced by the set of hyperparameters used, which should be tuned by trading-off between improving

predictive accuracy and preventing overfitting. The hyperparameters are: learning rate (η, a shrinkage scale factor of weights, used to reduce the influence of each individual tree, with discrete values chosen in a range between 0.01 and 0.30), minimum split loss (γ, minimum loss reduction required to furthers shards a leaf node, with discrete values chosen in a range between 0 and 2), maximum tree depth (maximum number of allowed splits, with discrete values chosen in a range between 1 and 15), subsample column ratio (defines the subsample of columns selected for each tree, with discrete values chosen in a range between 0.10 and 1.00), minimum child weight (minimum sum of instance weights required to further partitioning the tree, with discrete values chosen in a range between 1 and 10), subsample ratio (defines the subsample instances at each iteration, with discrete values chosen in a range between 0.10 and 1.00), maximum iterations (maximum number of trees fitted in the model, with discrete values chosen in a range between 100 and 500).

2.3.2.2. Model training and evaluation. To ensure comparability between model outputs obtained from traditional regression and ML, the same datasets were used to fit XGBoost models (both in terms of observations, variables and training/test split). Model hyperparameters were tuned by means of a grid search algorithm and a 5-fold cross-validation. Once tuned, the XGBoost model is retrained on the training dataset (75 %) and tested on the remaining 25 %.

The same applicable evaluation metrics used for traditional regression models are also computed for the obtained XGBoost models. Moreover, to support model interpretation and comparison with the previously fitted regression model coefficients, the following two complementary approaches are used:

- rank the variables based on the average information gain (or simply "*gain*") that the variable obtains in all the trees (see e.g., Ref. [89]): the top 10 variables showing the highest gain are reported.
- The SHAP (SHapley Additive exPlanations) values [92] were computed (see e.g., the application made by Ref. [88]), used to define a linear model g able to explain the original ML model, defined as follows:

$$g(z') = \phi_0 + \sum_{i=1}^{N} \phi_i z_i' \quad [17]$$

where z' is the i-th simplified input binary variable (up to N) and ϕ_i is the effect attributed to each variable, set equal to Shapley values [93]: weighted average of all possible differences between predictions obtained from models $f(x)$ trained on a subset of variables S including the i-th variable and models trained without it:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f(x_{S \bigcup \{i\}}) - f(x_S) \right] \quad [18]$$

Table 3
Results from linear regression models.

Predictor	Outcome variable: $AADT_l$				Outcome variable: $AADT_h$			
	Estimate	Std. Error	t value	p-value	Estimate	Std. Error	t value	p-value
Province-related								
Population	3869.180	530.707	7.291	<0.001	–	–	–	–
Area	−706.343	148.053	−4.771	<0.001	–	–	–	–
Accessibility index	–	–	–	–	−5.694	1.545	−3.687	<0.001
Occupation index	86.980	14.871	5.849	<0.001	12.082	1.378	8.768	<0.001
Tourism index	91.833	47.938	1.916	0.056	–	–	–	–
Municipality-related								
Area	8.986	1.946	4.736	<0.001	–	–	–	–
Density	6.145	0.682	9.005	<0.001	0.267	0.048	5.519	<0.001
Distance from large city	−28.084	7.851	−3.577	<0.001	−1.823	0.789	−2.311	0.021
Road-related								
Lanes – 4 (1)	9472.225	982.430	9.642	<0.001	832.071	84.910	9.799	<0.001
Lanes – 6 (2)	56368.724	2998.387	18.800	<0.001	3112.434	222.010	14.147	<0.001
Environment - Suburban (1)	3594.056	1185.792	3.031	0.003	–	–	–	–
Environment - Urban (2)	1597.232	806.002	1.982	0.048	–	–	–	–
Ring road – Yes (1)	4487.250	1030.272	4.355	<0.001	–	–	–	–
Intersections - Mixed (1)	1485.147	878.534	1.690	0.091	266.683	71.756	3.717	<0.001
Intersections - Grade-separated (2)	3289.687	906.831	3.628	<0.001	487.849	77.811	6.270	<0.001
Intersections – Freeway (3)	3587.573	1498.922	2.393	0.017	744.889	125.634	5.929	<0.001
Goodness-of-fit measures								
Likelihood Ratio Test (vs. null model)	χ^2 (15) = 1492.7, p-value <0.001				χ^2 (9) = 1095.9, p-value <0.001			
R^2	0.797				0.653			
Predictive accuracy metrics								
Root mean square error (RMSE)	6383				566			
CV-RMSE	0.51				0.84			

XGBoost models were estimated in *R* environment as well as the related metrics. In particular, the *xgboost* and *shapviz* library were used.

3. Results

Results from data analysis with both traditional regression analyses and ML are reported as follows. First, the results regarding traffic volume values are presented, then the results regarding traffic classes are reported.

3.1. Results: traffic volume predictions

Results from linear regression and XGBoost models for traffic volumes are reported, namely, in Tables 3 and 4.

Province population, occupation and tourism indexes, municipality area and density, the increase in the number of lanes, the urban/sub-urban area (in particular sub-urban), ring roads, intersections different than at-grade intersections on the road are related to an increase in light traffic volumes. Province area, distance from large city are related to a decrease in light traffic volumes instead.

Most of the above reported effects can be easily retrieved also in SHAP plots obtained after the XGBoost models (Fig. 6). Some of these factors are also the ones contributing the most to the model, such as roads with 4 or 6 lanes, municipality area and density, province population and occupation index, distance from large city.

Province occupation index, municipality density, the increasing number of lanes, intersections different than at-grade intersections on the road increase the heavy traffic volumes. Province accessibility index and distance from large city lead to a decrease in heavy traffic volumes. For heavy vehicle volumes as well, the same effects are visible from the SHAP plots and from the calculated importance gain of individual predictors.

Goodness-of-fit of the XGBoost models based on the R^2 calculation are considerably higher than the corresponding linear models, both for light and heavy vehicles (0.928 vs. 0.797 and 0.816 vs. 0.653, respectively). However, when considering metrics obtained from generalizing the models for other sample datasets (RMSE and CV-RMSE), the performance of linear and XGBoost models is

Table 4
Results from the XGBoost regression models.

Outcome variable: $AADT_l$				Outcome variable: $AADT_h$		
Feature		Gain		Feature		Gain
Category	Predictor			Category	Predictor	
Road	Lanes – 6 (2)	0.243		Road	Lanes – 4 (1)	0.304
Municipality	Area	0.191		Road	Lanes – 6 (2)	0.255
Municipality	Density	0.187		Municipality	Distance from large city	0.175
Road	Lanes – 4 (1)	0.100		Province	Occupation index	0.051
Province	Population	0.095		Municipality	Density	0.048
Municipality	Distance from large city	0.054		Road	Intersections – Freeway (3)	0.033
Province	Occupation index	0.023		Province	Accessibility index	0.032
Province	Accessibility index	0.019		Road	Intersections - Grade-separated (2)	0.025
Province	Area	0.018		Municipality	Area	0.024
Road	Intersections - Grade-separated (2)	0.014		Province	Population	0.014
Hyperparameters						
	300		Maximum iterations		100	
	6		Maximum tree depth		3	
	0.01		Learning rate *(η)*		0.01	
	0		Minimum split loss (γ)		2	
	0.75		Subsample column ratio		0.5	
	1		Minimum child weight		10	
	0.5		Subsample ratio		0.75	
Goodness-of-fit measures						
	0.928		R^2		0.816	
Predictive accuracy metrics						
	6801		RMSE		525	
	0.54		CV-RMSE		0.78	

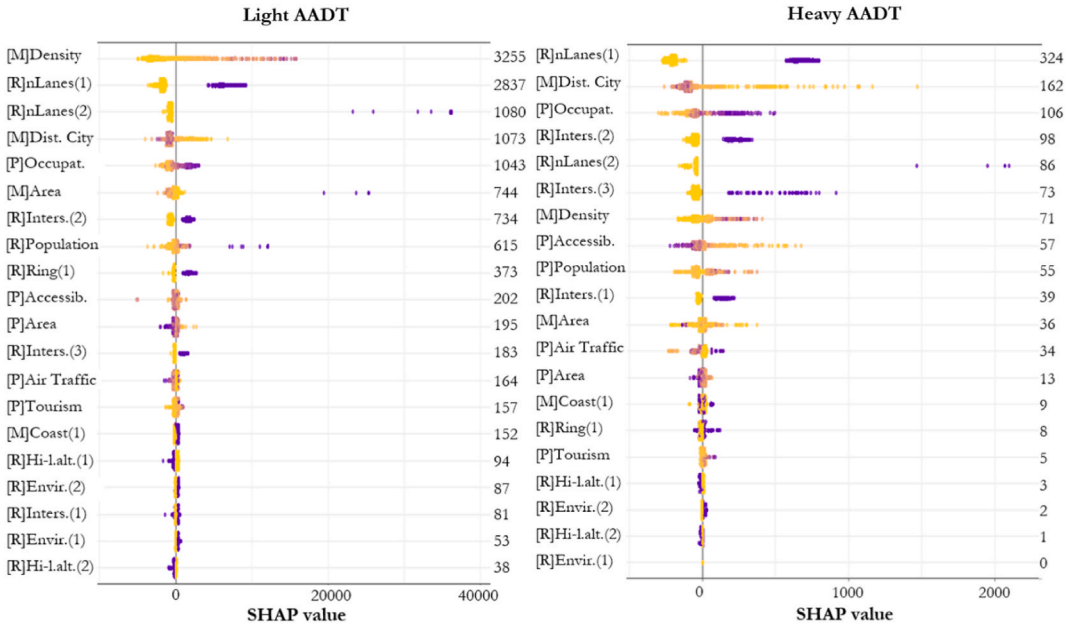


Fig. 6. SHAP values for light and heavy AADT volumes (the color scale goes from yellow to violet, following the low to high contribution provided by each individual predictor, for which the mean SHAP value is also reported; note for variables: [P] = Province-related, [M] = Municipality-related, [R] = Road-related).

Table 5
Results from the ordinal logistic regression models (for both light and heavy vehicle classes, 0 = low, 1 = medium, 2 = high).

Predictor	Outcome variable: $AADT_l$ class				Outcome variable: $AADT_h$ class			
	Estimate	**St.Err.**	**t value**	**p-value**	**Estimate**	**St.Err.**	**t value**	**p-value**
Coefficient (low|medium)	7.690	1.023	7.520	<0.001	5.709	0.994	5.741	<0.001
Coefficient (medium|high)	14.290	1.429	10.001	<0.001	9.379	1.063	8.825	<0.001
Province-related								
Population	0.552	0.199	2.779	0.005	0.314	0.180	1.744	0.081
Area	−0.168	0.081	−2.063	0.039	–	–	–	–
Accessibility index	–	–	–	–	−0.027	0.010	−2.819	0.005
Occupation index	0.070	0.014	5.134	<0.001	0.057	0.012	4.984	<0.001
Tourism index	0.042	0.021	2.053	0.040	–	–	–	–
Municipality-related								
Density	0.001	0.000	4.878	<0.001	0.001	0.000	2.438	0.015
Distance from large city	−0.012	0.004	−2.670	0.008	–	–	–	–
Road-related								
Lanes – 4 (1)	2.271	0.394	5.758	<0.001	2.632	0.348	7.564	<0.001
Lanes – 6 (2)	4.266	1.151	3.708	<0.001	3.990	1.018	3.918	<0.001
Environment - Suburban (1)	2.177	0.541	4.025	<0.001	–	–	–	–
Environment - Urban (2)	1.817	0.446	4.071	<0.001	1.143	0.401	2.853	0.004
Ring road – Yes (1)	0.955	0.373	2.562	0.010	–	–	–	–
Intersections - Mixed (1)	1.056	0.486	2.174	0.030	1.338	0.408	3.279	0.001
Intersections - Grade-separated (2)	2.145	0.485	4.423	<0.001	2.269	0.419	5.411	<0.001
Intersections – Freeway (3)	2.641	0.624	4.229	<0.001	2.752	0.541	5.086	<0.001
Goodness-of-fit measures/In-sample accuracy metrics								
Likelihood Ratio Test (vs. null model)	χ^2 (14) = 466.23, p-value <0.001				χ^2 (10) = 466.16, p-value <0.001			
Accuracy (%) (vs NIR %)	88.4 (88.4 > NIR = 77.2, p-value <0.001)				86.6 (86.6 > NIR = 75.5, p-value <0.001)			
Cohen's Kappa	0.66				0.64			
Predictive accuracy metrics								
Overall accuracy (%)	89.9				85.4			
Precision (by class) (%)	94 (0), 69 (1), 100 (2)				91 (0), 58 (1), 100 (2)			
Recall (by class) (%)	93 (0), 71 (1), 100 (2)				93 (0), 56 (1), 60 (2)			
Specificity (by class) (%)	76 (0), 94 (1), 100 (2)				63 (0), 92 (1), 100 (2)			
Balanced accuracy (by class) (%)	85 (0), 82 (1), 100 (2)				78 (0), 74 (1), 80 (2)			
F1 (by class) (%)	94 (0), 70 (1), 100 (2)				92 (0), 57 (1), 75 (2)			

comparable. There is a slight improvement in the XGBoost model for heavy vehicles and a slight worsening for light vehicles compared to the corresponding linear models. Hence, in this case, while a trade-off between the 20 % of unexplained variance ($R^2 = 0.797$) and the higher generalization (both lower RMSE and CV-RMSE values) can be acceptable for the simple linear model for light vehicles, the XGBoost model should be clearly preferred for heavy vehicle predictions.

3.2. Results: traffic class predictions

Results from logistic regression and XGBoost models for traffic classes are reported, namely, in Tables 5 and 6.

Most of the predictors included in the ordered logistic regression models, as well as the sign of the related coefficients estimates, are similar to those in the corresponding linear models. For what concerns light vehicles, the only difference is in the missing traffic increasing effect (here in terms of higher traffic classes) provided by an increasing municipality area. For heavy vehicles, a higher likelihood of having high volume classes can be related to an increase in population province and urban roads, while the distance from large city predictor was not included in the model.

Also in this case, most of the above reported effects can be easily retrieved in the SHAP plots referred to the three traffic classes obtained after training the XGBoost models (Figs. 7 and 8). Analyzing the factors contributing the most to the model, SHAP values confirm that higher values of municipality density correspond to a shift towards higher traffic classes (especially for light vehicles). Roads with 4 lanes are related to an increase in the traffic classes for both light and heavy vehicles, evidently associated to the Medium class. High values of the distance from large cities are particularly associated with the low traffic class (both for light and heavy vehicles, even if this variable was not included in the ordinal logistic model for heavy vehicles).

The accuracy of the XGBoost models (based on both the accuracy and Cohen's Kappa metrics) is considerably higher than the ordered logistic models, both for light and heavy vehicles. However, also in this case, while considering predictive accuracy metrics obtained from generalizing the models for other sample datasets (overall accuracy and metrics disaggregated by traffic classes), performances of ordered logistic and XGBoost models are comparable, noting a slight improvement for the XGBoost model for heavy vehicles and a slight worsening for the XGBoost model for light vehicles, compared to the corresponding ordered logistic models. In detail, in the light volume class prediction, the overall accuracies are almost equal between the ordered logistic and the XGBoost models (namely, 89.9 % vs. 89.1 %). The only minor difference seems to be provided by the slightly higher capability of the ordered logistic regression model to identify the High class (even if based on a small number of observations in the test dataset).

Table 6
Results from the XGBoost classification models (for both light and heavy vehicle classes, 0 = low, 1 = medium, 2 = high).

Outcome variable: $AADT_l$ class			Outcome variable: $AADT_h$ class		
Feature		Gain	Feature		Gain
Category	Predictor		Category	Predictor	
Municipality	Density	0.273	Road	Lanes – 4 (1)	0.308
Road	Lanes – 4 (1)	0.196	Municipality	Distance from large city	0.111
Municipality	Distance from large city	0.104	Municipality	Density	0.092
Province	Occupation index	0.083	Municipality	Area	0.077
Municipality	Area	0.070	Province	Occupation index	0.063
Road	Intersections - Grade-separated (2)	0.040	Province	Population	0.056
Road	Lanes – 6 (2)	0.038	Province	Area	0.050
Road	Ring road – Yes (1)	0.034	Province	Accessibility index	0.047
Province	Population	0.028	Province	Tourism index	0.033
Road	Intersections – Freeway (3)	0.026	Road	Intersections - Grade-separated (2)	0.031
Hyperparameters					
200		Maximum iterations	200		
3		Maximum tree depth	6		
0.025		Learning rate (η)	0.01		
0		Minimum split loss (γ)	0		
0.50		Subsample column ratio	0.75		
1		Minimum child weight	1		
1		Subsample ratio	0.5		
Goodness-of-fit/In-sample accuracy metrics					
93.4 (93.4 > NIR = 77.2, p-value <0.001)		Accuracy (%) (vs NIR%)	93.1 (93.1 > NIR = 75.5, p-value <0.001)		
0.81		Cohen's Kappa	0.82		
Predictive accuracy metrics					
89.1		Overall accuracy (%)	86.2		
93 (0), 68 (1), 100 (2)		Precision (by class) (%)	90 (0), 61 (1), 100 (2)		
94 (0), 66 (1), 88 (2)		Recall (by class) (%)	95 (0), 54 (1), 50 (2)		
71 (0), 94 (1), 100 (2)		Specificity (by class) (%)	61 (0), 93 (1), 100 (2)		
83 (0), 80 (1), 94 (2)		Bal. accuracy (by class) (%)	78 (0), 73 (1), 75 (2)		
93 (0), 67 (1), 93 (2)		F1 (by class) (%)	93 (0), 57 (1), 67 (2)		

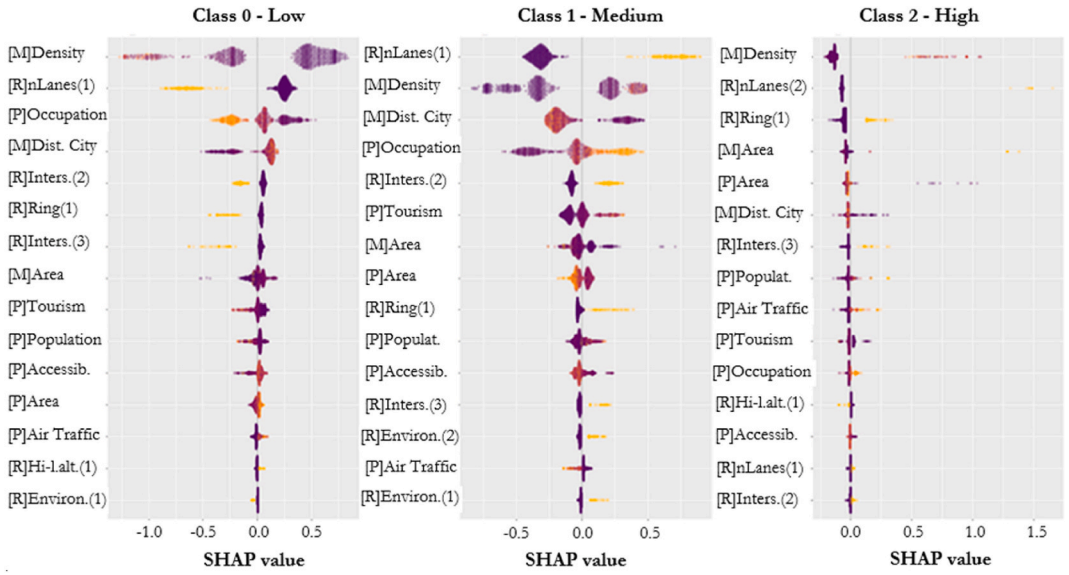


Fig. 7. SHAP values for light vehicle volume classes (the color scale goes from yellow to violet, following the low to high contribution provided by each individual predictor; note for variables: [P] = Province-related, [M] = Municipality-related, [R] = Road-related).

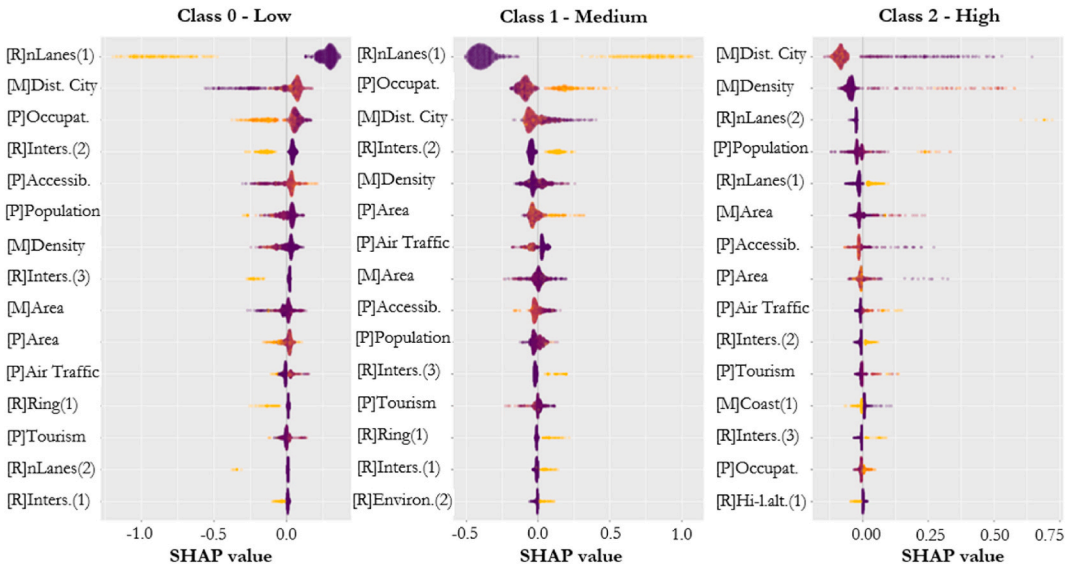


Fig. 8. SHAP values for heavy vehicle volume classes (the color scale goes from yellow to violet, following the low to high contribution provided by each individual predictor; note for variables: [P] = Province-related, [M] = Municipality-related, [R] = Road-related).

4. Discussion

Results are discussed as follows, based on the peculiar objectives of this study. First, the factors influencing light and heavy traffic volumes are discussed in light of previous research. Then, the feasibility of integrating such traffic prediction models in overall multi-risk frameworks is argued and demonstrated.

4.1. Factors influencing traffic volumes

All the relationships that emerged between predictors and traffic volumes (both light and heavy) seem logical and in accordance with previous research. Note however that, as previously stated, most previous studies predicted the overall traffic volume without making differences between light and heavy vehicles. Given that, usually, the traffic volume is dominated by light vehicles, the comparison of results for light vehicles from this study are compared with previous findings for overall volumes.

With regards to geographic and socio-economic variables, population was related to an increase in light traffic volumes, coherently with findings from previous studies [50,53,55,58]. Similarly, the increase in AADT with the studied area and related population density was also found by, namely, Caceres et al. [55]; Xia et al. [51] and Das & Tsapakis [63]. Note however that, in this study, the effect of the area of the municipality in which the traffic count is included is related to an increase in the light vehicle volume (except for the logistic regression model), while the province area is related to its decrease. The traffic increasing effect induced by an increase in the occupation index is similar to other effects found for other variables related to economic activities and occupation ([54,55,58,63]; effect specifically related to income by Ref. [57]). To the authors' knowledge, the specific effect of tourism on traffic was first revealed in this study. However, this effect may be someway included in the general effects of economic activities, previously discussed. The increase in distance from large cities was related to a decrease in light volumes, similarly to Caceres et al. [55], who used the distance from mean center of population as specific variable. Land use is here studied through the effect of urban/sub-urban environments, which are related to increasing traffic with respect to the rural environment. This tendency was also revealed by Eom et al. [57]; Pulugurtha & Kusam [94], Apronti et al. [53]. Duddu & Pulugurtha [61] found more complex relationships, depending on the detailed land use classification considered in their study.

Significant effects of road-related variables were mostly noted for the number of lanes and the intersection types. The logical tendency for light traffic volume to increase with the number of lanes was also found in several previous studies [51,52,54,55,57,94]. The presence of intersections other than at-grade intersections was also related to an increase in light traffic volumes. The importance of this detailed variable was first revealed in this study, to the authors' knowledge. However, it is strictly related to other road characteristics such as the road functional classification and the speed limit, discussed in other studies. In fact, it is most likely to find grade-separated intersections on high-level/high-speed roads. The increase in road importance was related to an increase in traffic volumes in other studies [51,52,54,57,58], as well as the increase in speeds [57,58,94]. However, the use of the number of lanes and the intersection type can be more easily transferable than the road functional classification (depending on specific country/regional standards) or the speed limit (an information which is not always easy to collect). Finally, the ring road is associated with the traffic increasing, as also noted by Sfyridis and Agnolucci [64].

The effects found for heavy vehicles are similar to those already discussed for light vehicles (related to occupation, density, number of lanes, and intersection types), with some notable exceptions. Specifically, the accessibility index, which relates to the presence of infrastructures other than roads, is clearly inversely related to heavy vehicle traffic volume, whereas it was not relevant to light vehicle volumes. Hence, this study suggests that the presence of other infrastructures in the studied area should be considered when predicting heavy vehicle traffic volumes. Moreover, the distance from large cities was not included in the logistic regression model for heavy vehicles. This may be because high heavy-volume roads are mostly major roads in the network, where heavy vehicle transit is not particularly influenced by local conditions and often serves long intercity travel distances.

It is important to note that this study used all traffic counts available for the national (Italian) road network to build traffic prediction models. While tendencies revealed in this study were found to be generally coherent with previous research, most of previous studies used data from low-volume/local roads ([53,54,60,63]), non-state roads [51], non-freeway/expressway facilities [56,57]. Moreover, except for the study by Caceres et al. [55], which used Spanish traffic counts and by Fu et al. [62], which used Irish data (even if based on a spatial approach different than the one used in this study), AADT prediction studies were mostly conducted outside of Europe, typically in the United States. Hence, on one hand, it appears that general relationships between traffic volumes and the main predictors are independent of the geographic context; on the other hand, the definition of variables typically depends on local factors (e.g., administrative units such as municipalities, counties, provinces etc., or the different road functional classification).

4.2. Integration with multi-risk assessment frameworks

4.2.1. Discussion of results in the context of multi-risk assessment

Factors which were revealed as predictors are easily collectable by other researchers and practitioners such as population, density and area. Other indexes such as the occupation (percentage of working people in the age 15–64) and the tourism (number of days of residential tourism per inhabitant per year) can be retrieved from local statistics. The accessibility index, which mostly influences the heavy vehicle volumes, even if not immediately available, can be determined as previously indicated in Section 2.2, starting from travel time estimations on the network. Additionally, all the road-related information included in the models, such as the number of lanes, intersection types, road environment, ring road, can be easily found from online data sources. Hence, while models developed in this study are based on Italian data, the modeling approach could be ideally transferred anywhere.

The integration of traffic prediction models in multi-risk assessment frameworks depends on the particular framework used. In fact, multi-risk methodologies are mainly based on analytical/quantitative and on classification/qualitative approaches, which have different details of data needed. Such approaches can be applied to all the components of the overall risk estimation: hazard, exposure and vulnerability. It is not rare that both approaches are used simultaneously to describe different factors into the same methodology: in this case, quantitative estimates are used for variables for which more detailed data are available [5].

In the case of the numerical approach, variables are directly transformed from continuous data (possibly normalized) to establish various indices and define the risk parameters calculation or to develop data regression models [28,95,96]. The outputs can be economic losses, fragility curves, and resilience indices (see e.g., Ref. [26,38]). Using such approach, traffic volumes can be directly used (eventually normalized or weighted) as an input variable in calculating the risk exposure and quantifying damages to the road network after critical events (see e.g., Ref. [15,97]). Both the models developed here for light and heavy vehicles could provide traffic volume predictions to integrated risk assessments frameworks. Specifically, the generalized linear models can be more easily integrated in other frameworks, especially those related to light vehicles, which proved to be more reliable than those for heavy vehicles,

when compared to the corresponding XGBoost models.

Nevertheless, risk assessment frameworks using qualitative approaches are usually based on categorical variables (sometimes discretized from continuous values) and their interaction through matrices and simple logical operations (typically product or weighted mean). In this case, outputs are often warning indices or simplified risk maps, which may easily support decision-makers. Classifications may involve hazard parameters such as hazard zone classification, vulnerability parameters such as soil or shoreline type classes and exposure factors (see e.g., Ref. [5,14,31]). The models developed here for predicting light and heavy vehicle traffic volume classes can be easily integrated in such frameworks. In this case, depending on the overall framework structure, the traffic input "module" which only provides a class number (in this case from one to three) can be integrated regardless of the particular approach used (ordered logistic model or ML approach).

The integration of traffic predictions into multi-risk assessment frameworks will be demonstrated through the following two practical examples.

4.2.2. Example application #1: Vulnerability of coastal roads to erosion and flooding

This example application is based on the study by Drejza et al. [33], in which the vulnerability of coastal roads to erosion and flooding is measured through a proposed index (CREFVI -Coastal Road Erosion and Flooding Vulnerability Index-), obtained as the average between two sub-indexes related to erosion and flooding:

$$CREFVI = \frac{EV + FV}{2} = \frac{\sqrt{\sqrt[10]{\prod_{i=1}^{10} e_i} + \sqrt{\sqrt[10]{\prod_{j=1}^{10} f_i}}}}{2} \quad [19]$$

where: EV = Erosion Vulnerability, FV = Flooding Vulnerability, e_i = each of the 10 parameters related to erosion, variable on a 5-points scale, f_i = each of the 10 parameters related to flooding, variable on a 5-points scale (most of them, including traffic, are the same e_i parameters). CREFVI = 0 implies no intervention, 0<CREFVI <10 indicates low vulnerability rank (monitoring and long-term intervention planning), 10≤CREFVI<25 indicates medium vulnerability rank (medium-term or case-by-case intervention planning), 25≤CREFVI<50 indicates high vulnerability rank (necessary intervention), CREFVI≥50 indicates critical vulnerability (immediate intervention).

Considering that the 5-points scale traffic index based on AADT ranges is included in the list of both e_i and f_i parameters, the previous equation can be rewritten, for the particular scope of this example, as follows:

$$CREFVI = \sqrt{AADT\ index}^{*} \frac{\sqrt{\sqrt[10]{\prod_{i=1}^{9} e_i} + \sqrt{\sqrt[10]{\prod_{j=1}^{9} f_i}}}}{2} \quad [20]$$

where the index based on the AADT was isolated from the calculation of the two sub-indexes (the list of e_i and f_i is then limited to 9 out of 10 parameters) to determine the contribution of traffic volume to the overall vulnerability index. Based on the previous equation, it is possible to quantify the impact on the overall index deriving from considering different traffic classes (other conditions being equal), based on the following ratio:

$$\frac{CREFVI_{AADT\ index=i}}{CREFVI_{AADT\ index=j}} = \frac{\sqrt{AADT\ index = i}}{\sqrt{AADT\ index = j}} \quad [21]$$

Results from this calculation are shown in next Table 7 for all the possible combinations of traffic indexes.

It is immediately visible how the traffic volume may significantly influence the vulnerability index. As a direct consequence, a wrong specification of the traffic class may lead to errors in the rank classification and, consequently, in the definition of intervention priorities. This is particularly relevant in cases in which coastal vulnerability should be assessed for roads which are not provided with traffic data. For example, if, in case of missing data, a traffic class equal to 1 (AADT <1000) is assigned, while the real traffic volume should be instead assigned to the 4th class (4000 ≤ AADT ≤5999); this would result in underestimating the CREFVI by 50 %. Clearly, if the opposite occurs, overestimations are possible as well. Such errors may result in not assigning the correct vulnerability rank to the coastal road for mitigation interventions against erosion and flooding.

4.2.3. Example application #2: Economic assessments of post-earthquake scenarios

The second example application is based on the study by Ishibashi et al. [38], in which indirect losses due to changes in the travel path are computed based on functionality loss of bridges in the area of interest due to post-earthquake damage. This scenario may cause unavailability of links and, consequently, detours. Hence, both the increased travel times and distances generate indirect costs which should be summed up to direct recovery costs.

The two basic indirect costs (cost due to increased distance C_d and cost due to increased travel time C_t) which, summed up, generate the total indirect cost $C(I)$, are reported as follows from the cited study:

Table 7
Ratios between the calculated CREFVI index considering a given traffic class and the same index considering another traffic class (among the 5 classes defined in the study by Ref. [33]).

AADT index	1 (AADT <1000)	2 (1000 ≤ AADT ≤1999)	3 (2000 ≤ AADT ≤3999)	4 (4000 ≤ AADT ≤5999)	5 (AADT ≥6000)
1 (AADT <1000)	1.00	1.41	1.73	2.00	2.24
2 (1000 ≤ AADT ≤1999)	0.71	1.00	1.22	1.41	1.58
3 (2000 ≤ AADT ≤3999)	0.58	0.82	1.00	1.15	1.29
4 (4000 ≤ AADT ≤5999)	0.50	0.71	0.87	1.00	1.12
5 (AADT ≥6000)	0.45	0.63	0.77	0.89	1.00

$$C_d = \int_{t_o}^{t_o+\Delta} \left\{ \left[c_{d,lv} * (1 - \%hv) + c_{d,hv} * (\%hv) \right] * (L_{det} - L) * q'_{det}(t) \right\} dt \quad [22]$$

$$C_t = \int_{t_o}^{t_o+\Delta} \left\{ \left[c_{t,lv} * (1 - \%hv) + c_{t,hv} * (\%hv) \right] * \left[q'_{det}(t) * \left(T'_{det}(t) - T \right) + q'(t) * (T'(t) - T) \right] \right\} dt \quad [23]$$

where c_d and c_t are the unitary costs per, respectively, travelled km and minute spent for the travel, both differentiated for light (lv) and heavy vehicles (hv); $\%hv$ is the average percentage of heavy vehicles in the flow; L and L_{det} are, respectively, the length of the link and the detour link; q' and q'_{det} are, respectively, the post-disaster link flow and portion of original link flow q which uses the detour, T' and T'_{det} are, respectively, the post-disaster link and detour travel times; T is the original travel time on the link in normal conditions; t_o is the disaster time instant and Δ is a generic time interval (to get the overall cost it depends on the post-disaster recovery time). Readers are referred to Ishibashi et al. [38] for details about the theoretical background and the calculation of each of the reported variables.

Assuming, for the sake of the example application, a simple scenario formed by a 10-km long road link with limited functionality after a given disaster (which reduced free flow speed and capacity by 25 %) and a detour which is 50 % longer, it is possible to compute the total indirect costs $C(I)$ due to increased distance and travel times. The computed indirect costs for a 1 h time interval are reported in the following diagrams as a function of the original AADT on the link in normal conditions (assuming $\%hv = 0.05$), in the two following scenarios:

- detour from a two-way two-lane road link to another similar link (Fig. 9a);
- detour from a four-lane divided road link to a two-way two-lane road link (Fig. 9b).

Previous diagrams evidently show how indirect costs are more than linearly dependent on the AADT, especially when the detour involves a road with lower features than the damaged road link. The simple numerical example shown demonstrates how correctly specifying the AADT may help estimating post-disaster indirect costs on a road network. Moreover, it sheds some light on how those indirect costs may be significantly variable depending on the AADT, especially for high-level roads, again demonstrating the importance of this source of information (including the percentage of heavy vehicles) for multi-risk assessments.

5. Conclusions

Measures of traffic volumes are often based on sparse traffic counter stations which are mainly located on rural arterials belonging to the main road network. Defining traffic volumes for each road section in a road network is almost always impossible regardless of the specific country or region. Hence, traffic volume prediction models can be useful for addressing this gap, regardless of the specific application for which they are required. This study was focused on developing prediction models for both traffic volume values and classes, to be used in the context of a multi-disciplinary research project focused on the multi-risk assessment of critical infrastructures.

Province- and municipality-related geographic and socio-economic variables (such as population, occupation, tourism, population density, urban environment, distance from large cities, accessibility index) coupled with road-related variables (such as number of

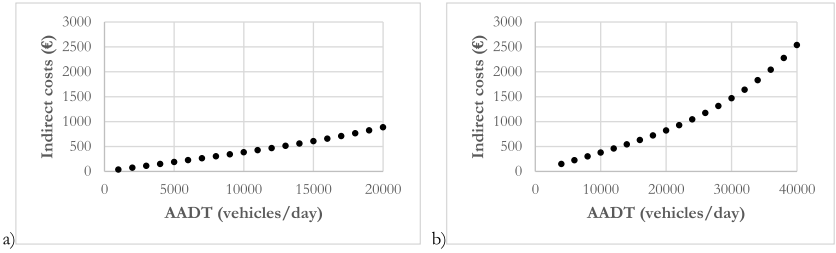


Fig. 9. Computed indirect costs in the example application of 50 % longer detour from a damaged 10-km long link in a 1 h post-disaster time interval (a, on the left: detour from a two-way two-lane road to a similar road; b, on the right: detour from a four-lane divided road to a two-way two-lane road).

lanes, intersection types, ring roads) were used to satisfactorily predict both traffic volume values and classes (volume ranges from low to high). Two classes of models were developed for both light and heavy vehicle traffic volumes, using traditional statistical techniques (generalized linear model for predicting traffic values and ordered logistic models for predicting traffic classes) and ML approach (XGBoost technique for both the regression and classification problems). The latter proved to be more effective than traditional approaches for the estimation of heavy vehicles, while it provided comparable results for light vehicles.

Results obtained from this study can stimulate further research on this topic. In fact, the presented flexible framework, considering both regression and classification problems, and the use of both traditional statistical approaches and ML techniques can pave the way for exploring the integration of a traffic prediction module into other risk models focused on natural hazards (see e.g., the coastal vulnerability index proposed by Ref. [98]), which usually neglect users' exposure on road infrastructures. Conversely, the importance of accurately predicting traffic volumes for risk assessments related to different types of disasters was shown in the previous section, through two application examples based on previous research.

Findings of this study can be useful in different applications, ranging from basic traffic engineering purposes to the broad context of multi-risk assessment of critical infrastructures, in which this study was conceived. In particular:

- traffic volume prediction models can be used by traffic engineers (i.e., for safety assessments, pavement management and maintenance, design applications), in case of unavailable traffic counts;
- the estimated traffic volume values can be used as input for analytical risk assessment approaches, while traffic classes can be directly used for simplified assessments in which the drivers' level of exposure to risk on the road network is of interest. Specifically, the estimated traffic volume values could be subsequently rearranged as well into classes, depending on the application and traffic range definition;
- heavy vehicle traffic volume predictions can be specifically used for multi-risk assessments of critical infrastructures such as bridges;
- the use of integrated frameworks including traffic volume predictions can be useful for public administrations and decision-makers to define risk management and intervention strategies.

For what concerns the mentioned applications, it is worth noting that road agencies which manage different levels of networks may consider implementing new traffic counter stations in locations particularly exposed to external hazards, in the optic of a multi-risk assessment (see also [37]). In this way, prediction errors could be minimized at those locations where the exposure levels should be estimated with the greatest level of accuracy. In cases where this implementation is not feasible, models such as those presented in this study can be used to supply traffic volume data and update them in real-time. In this case, road agencies should store and update the necessary input data in order to feed prediction models relevant to the particular road network under investigation.

This study is not without limitations. In particular, some assumptions were made while defining the methodological framework. First, a *k*-means algorithm was used to define both the number and the range of variability of traffic classes, while generally those labels are set a-priori, depending on expert judgement and/or context-based evaluation. However, this was done to make the approach more scalable and applicable to various contexts. To limit this threat to the validity of results, we set up the *k*-means algorithm to find only the main 3 classes (i.e., from low to high), hence minimizing the probability of error, while checking on a sample basis the output of the k-means algorithm. Moreover, some predictors that are not easily retrievable from national/local statistics and online sources may have improved the model prediction (e.g., other road-related specific variables). Further research could be dedicated to specifically improving prediction models, with particular regard to heavy vehicles. However, the selection of variables was guided by the search for a trade-off between model accuracy and flexibility, by prioritizing easily collectable and generalizable variables, for the sake of broader applicability and transferability. In fact, while both the approaches and variables used can be transferred anywhere, the presented models are based on Italian data, used as testbed given its proneness to the combined seismic and coastal hazard, alongside the high road network density. Hence, caution should be taken while using these models for predictions outside of the study context, without preliminary calibrations. Finally, the temporal variability of traffic volumes should be monitored, and modeling approaches should consider this aspect, if relevant.

CRediT authorship contribution statement

Paolo Intini: Writing – review & editing, Writing – original draft, Visualization, Methodology, Funding acquisition, Data curation, Conceptualization. **Gianni Blasi:** Writing – review & editing, Funding acquisition, Conceptualization. **Francesco Fracella:** Writing – original draft, Visualization. **Antonio Francone:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **Roberto Vergallo:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Daniele Perrone:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by the Department of Innovation Engineering of the University of Salento through the research grant "MASCOT – Multi-risk ASsessment of Critical Outdated infrasTructures".

Data availability

Data will be made available on request.

References

[1] European Council, Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection, Off. J. Eur. Union 23 (2008) 27–31.
[2] S. Balakrishnan, B. Cassottana, InfraRisk: an open-source simulation platform for resilience analysis in interconnected power–water–transport networks, Sustain. Cities Soc. 83 (2022) 103963, https://doi.org/10.1016/j.scs.2022.103963.
[3] W. Marzocchi, A. Garcia-Aristizabal, P. Gasparini, M.L. Mastellone, A. Di Ruocco, Basic principles of multi-risk assessment: a case study in Italy, Nat. Hazards 62 (2012) 551–573, https://doi.org/10.1007/s11069-012-0092-x.
[4] I. Iervolino, Dinamica delle strutture e ingegneria sismica: Principi e applicazioni, Hoepli Editore, 2021.
[5] E.E. Koks, B. Jongman, T.G. Husby, W.J. Botzen, Combining hazard, exposure and social vulnerability to provide lessons for flood risk management, Environ. Sci. Pol. 47 (2015) 42–52, https://doi.org/10.1016/j.envsci.2014.10.013.
[6] International Standard Organization, ISO 31000 Risk Management — Principles and Guidelines, 2009, p. 13, 61010-1 © Iec:2001 2009.
[7] A. Urlainis, D. Ornai, R. Levy, O. Vilnay, I.M. Shohet, Loss and damage assessment in critical infrastructures due to extreme events, Saf. Sci. 147 (2022) 105587, https://doi.org/10.1016/j.ssci.2021.105587.
[8] UNODRR -United Nations Office for Disaster Risk Reduction-, Sendai Framework for Disaster Risk Reduction 2015-2030, 2015, p. 2015.
[9] E.E. Koks, J. Rozenberg, C. Zorn, M. Tariverdi, M. Vousdoukas, S.A. Fraser, S. Hallegatte, A global multi-hazard risk analysis of road and railway infrastructure assets, Nat. Commun. 10 (1) (2019) 2677.
[10] M.S. Kappes, M. Keiler, K. von Elverfeldt, T. Glade, Challenges of analyzing multi-hazard risk: a review, Nat. Hazards 64 (2012) 1925–1958, https://doi.org/10.1007/s11069-012-0294-2.
[11] G. Zuccaro, D. De Gregorio, M.F. Leone, Theoretical model for cascading effects analyses, Int. J. Disaster Risk Reduc. 30 (2018) 199–215, https://doi.org/10.1016/j.ijdrr.2018.04.019.
[12] K.N. Scheitlin, J.B. Elsner, S.W. Lewers, J.C. Malmstadt, T.H. Jagger, Risk assessment of hurricane winds for Eglin air force base in northwestern Florida, USA, Theor. Appl. Climatol. 105 (2011) 287–296, https://doi.org/10.1007/s00704-010-0386-4.
[13] V. Poompavai, M. Ramalingam, Geospatial analysis for coastal risk assessment to cyclones, J. Indian Soc. Rem. Sens. 41 (2013) 157–176, https://doi.org/10.1007/s12524-011-0198-8.
[14] C. Do, Y. Kuleshov, Multi-hazard tropical cyclone risk assessment for Australia, Rem. Sens. 15 (3) (2023) 795, https://doi.org/10.3390/rs15030795.
[15] A. Zischg, S. Fuchs, M. Keiler, G. Meißl, A. Zischg, S. Fuchs, M. Keiler, G. Meißl, Modelling the system behaviour of wet snow avalanches using an expert system approach for risk management on high alpine traffic roads, Nat. Hazards Earth Syst. Sci. 5 (6) (2005) 821–832, https://doi.org/10.5194/nhess-5-821-2005.
[16] A. Sinickas, B. Jamieson, M.A. Maes, Snow avalanches in western Canada: investigating change in occurrence rates and implications for risk assessment and mitigation, Struct. Infrastruct. Eng. 12 (4) (2016) 490–498, https://doi.org/10.1080/15732479.2015.1020495.
[17] X. Shao, C. Xu, Earthquake-induced landslides susceptibility assessment: a review of the state-of-the-art, Nat. Hazards Res. 2 (3) (2022) 172–182, https://doi.org/10.1016/j.nhres.2022.03.002.
[18] V. Gallina, S. Torresan, A. Critto, A. Sperotto, T. Glade, A. Marcomini, A review of multi-risk methodologies for natural hazards: consequences and challenges for a climate change impact assessment, J. Environ. Manag. 168 (2016) 123–132, https://doi.org/10.1016/j.jenvman.2015.11.011.
[19] S.A. Argyroudis, S.A. Mitoulis, M.G. Winter, A.M. Kaynia, Fragility of transport assets exposed to multiple hazards: state-of-the-art review toward infrastructural resilience, Reliab. Eng. Syst. Saf. 191 (2019) 106567.
[20] D. Murgese, G. Giraudo, D. Testa, G. Airoldi, R. Cagna, M. Bugnano, S. Castagna, Multi-risk assessment of Cuneo province road network, in: Engineering Geology for Society and Territory-Volume 6: Applied Geology for Major Engineering Projects, Springer International Publishing, 2015, pp. 657–661.
[21] A. Karatzetzou, S. Stefanidis, S. Stefanidou, G. Tsinidis, D. Pitilakis, Unified hazard models for risk assessment of transportation networks in a multi-hazard environment, Int. J. Disaster Risk Reduc. 75 (2022) 102960.
[22] A. Solheim, K. Sverdrup-Thygeson, B. Kalsnes, Hazard and risk assessment for early phase road planning in Norway, Nat. Hazards 119 (2) (2023) 943–963.
[23] J. Clarke, E. O'Brien, Multi-hazard risk assessment methodology, stress test framework and decision support tool for resilient critical infrastructure, Transport. Res. Procedia 14 (2016) 1355–1363, https://doi.org/10.1016/j.trpro.2016.05.208.
[24] S. Argyroudis, J. Selva, P. Gehl, K. Pitilakis, Systemic seismic risk assessment of road networks considering interactions with the built environment, Comput. Aided Civ. Infrastruct. Eng. 30 (7) (2015) 524–540, https://doi.org/10.1111/mice.12136.
[25] K. Pitilakis, S. Argyroudis, K. Kakderi, J. Selva, Systemic vulnerability and risk assessment of transportation systems under natural hazards towards more resilient and robust infrastructures, Transport. Res. Procedia 14 (2016) 1335–1344, https://doi.org/10.1016/j.trpro.2016.05.206.
[26] X. Guo, Y. Wu, Y. Guo, Time-dependent seismic fragility analysis of bridge systems under scour hazard and earthquake loads, Eng. Struct. 121 (2016) 52–56, https://doi.org/10.1016/j.engstruct.2016.04.038.
[27] G. Ganesh Prasad, S. Banerjee, The impact of flood-induced scour on seismic fragility characteristics of bridges, J. Earthq. Eng. 17 (6) (2013) 803–828, https://doi.org/10.1080/13632469.2013.771593.
[28] S.A. Argyroudis, S.A. Mitoulis, Vulnerability of bridges to individual and multiple hazards-floods and earthquakes, Reliab. Eng. Syst. Saf. 210 (2021) 107564, https://doi.org/10.1016/j.ress.2021.107564.
[29] J.M. Andrić, D.G. Lu, Risk assessment of bridges under multiple hazards in operation period, Saf. Sci. 83 (2016) 80–92, https://doi.org/10.1016/j.ssci.2015.11.001.
[30] A. Abarca, R. Monteiro, G.J. O'Reilly, Simplified methodology for indirect loss–based prioritization in roadway bridge network risk assessment, Int. J. Disaster Risk Reduc. 74 (2022) 102948, https://doi.org/10.1016/j.ijdrr.2022.102948.
[31] A. Abarca, R. Monteiro, G.J. O'Reilly, Seismic risk prioritisation schemes for reinforced concrete bridge portfolios, Struct. Infrastruct. Eng. (2023) 1–21, https://doi.org/10.1080/15732479.2023.2187424.
[32] M. VanKoningsveld, J.P. Mulder, M.J. Stive, L. VanDerValk, A.W. VanDerWeck, Living with sea-level rise and climate change: a case study of The Netherlands, J. Coast Res. 24 (2) (2008) 367–379, https://doi.org/10.2112/07A-0010.1.
[33] S. Drejza, P. Bernatchez, G. Marie, S. Friesinger, Quantifying road vulnerability to coastal hazards: development of a synthetic index, Ocean Coast Manag. 181 (2019) 104894, https://doi.org/10.1016/j.ocecoaman.2019.104894.
[34] R.J. Nicholls, Adapting to Sea-Level Rise. Resilience, 2018, pp. 13–29, https://doi.org/10.1016/B978-0-12-811891-7.00002-5.
[35] IPCC (Intergovernmental Panel on Climate Change), in: H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (Eds.), Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to

the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022, p. 3056, https://doi.org/10.1017/9781009325844.
[36] A. Satta, M. Puddu, S. Venturini, C. Giupponi, Assessment of coastal risks to climate change related impacts at the regional scale: The case of the Mediterranean region, Int. J. Disaster Risk Reduct. 24 (2017) 284–296.
[37] M. Gazzea, A. Miraki, O. Alisan, M.M. Kuglitsch, I. Pelivan, E.E. Ozguven, R. Arghandeh, Traffic monitoring system design considering multi-hazard disaster risks, Sci. Rep. 13 (1) (2023) 4883, https://doi.org/10.1038/s41598-023-32086-6.
[38] H. Ishibashi, M. Akiyama, D.M. Frangopol, S. Koshimura, T. Kojima, K. Nanami, Framework for estimating the risk and resilience of road networks with bridges and embankments under both seismic and tsunami hazards, Struct. Infrastruct. Eng. 17 (4) (2020) 494–514, https://doi.org/10.1080/15732479.2020.1843503.
[39] Ministero delle Infrastrutture e dei Trasporti, Consiglio Superiore dei Lavori Pubblici (in English: Italian Ministry of Infrastructures and Transport, High Council of Public Works). Linee Guida per la Classificazione e Gestione del Rischio, la Valutazione della Sicurezza ed il Monitoraggio dei Ponti Esistenti (in English: Guidelines for Classifying and Managing Risk, Assessing Safety and Monitoring of Existing Bridges), 2020.
[40] G.C. Patton, C. Coffey, S.M. Sawyer, R.M. Viner, D.M. Haller, K. Bose, C.D. Mathers, Global patterns of mortality in young people: a systematic analysis of population health data, Lancet 374 (9693) (2009) 881–892, https://doi.org/10.1016/S0140-6736(09)60741-8.
[41] R. Elvik, T. Bjørnskau, Safety-in-numbers: a systematic review and meta-analysis of evidence, Saf. Sci. 92 (2017) 274–282, https://doi.org/10.1016/j.ssci.2015.07.017.
[42] P. Murray-Tuite, B. Wolshon, Evacuation transportation modeling: an overview of research, development, and practice, Transport. Res. C Emerg. Technol. 27 (2013) 25–45, https://doi.org/10.1016/j.trc.2012.11.005.
[43] K. Kim, P. Pant, E. Yamashita, Integrating travel demand modeling and flood hazard risk analysis for evacuation and sheltering, Int. J. Disaster Risk Reduc. 31 (2018) 1177–1186, https://doi.org/10.1016/j.ijdrr.2017.10.025.
[44] E. Ronchi, S.M. Gwynne, G. Rein, P. Intini, R. Wadhwani, An open multi-physics framework for modelling wildland-urban interface fire evacuations, Saf. Sci. 118 (2019) 868–880, https://doi.org/10.1016/j.ssci.2019.06.009.
[45] P. Intini, E. Ronchi, S. Gwynne, A. Pel, Traffic modeling for wildland–urban interface fire evacuation, J. Transport. Eng., Part A: Systems 145 (3) (2019) 04019002, https://doi.org/10.1061/JTEPBS.0000221.
[46] P. Intini, J. Wahlqvist, N. Wetterberg, E. Ronchi, Modelling the impact of wildfire smoke on driving speed, Int. J. Disaster Risk Reduc. 80 (2022) 103211, https://doi.org/10.1016/j.ijdrr.2022.103211.
[47] A. Khattak, X. Wang, H. Zhang, Are incident durations and secondary incidents interdependent? Transport. Res. Rec. 2099 (1) (2009) 39–49, https://doi.org/10.3141/2099-05.
[48] E.I. Vlahogianni, M.G. Karlaftis, F.P. Orfanou, Modeling the effects of weather and traffic on the risk of secondary incidents, J. Intell. Transport. Syst. 16 (3) (2012) 109–117, https://doi.org/10.1080/15472450.2012.688384.
[49] R.M. Robinson, A.J. Collins, C.A. Jordan, P. Foytik, A.J. Khattak, Modeling the impact of traffic incidents during hurricane evacuations using a large scale microsimulation, Int. J. Disaster Risk Reduc. 31 (2018) 1159–1165, https://doi.org/10.1016/j.ijdrr.2017.09.013.
[50] D. Mohamad, K.C. Sinha, T. Kuczek, C.F. Scholer, Annual average daily traffic prediction model for county roads, Transport. Res. Rec. 1617 (1) (1998) 69–77, https://doi.org/10.3141/1617-10.
[51] Q. Xia, F. Zhao, Z. Chen, L.D. Shen, D. Ospina, Estimation of annual average daily traffic for nonstate roads in a Florida county, Transport. Res. Rec. 1660 (1) (1999) 32–40, https://doi.org/10.3141/1660-05.
[52] F. Zhao, S. Chung, Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models, Transport. Res. Rec. 1769 (1) (2001) 113–122, https://doi.org/10.3141/1769-14.
[53] D. Apronti, K. Ksaibati, K. Gerow, J.J. Hepner, Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods, J. Traffic Transport. Eng. 3 (6) (2016) 493–506.
[54] M. Shojaeshafiei, M. Doustmohammadi, S. Subedi, M. Anderson, Comparison of estimation methodologies for daily traffic count prediction in small and medium sized communities, Int. J. Traffic Transport. Eng. 6 (4) (2017) 71–75, https://doi.org/10.5923/j.ijtte.20170604.01.
[55] N. Caceres, L.M. Romero, F.J. Morales, A. Reyes, F.G. Benitez, Estimating traffic volumes on intercity road locations using roadway attributes, socioeconomic features and other work-related activity characteristics, Transportation 45 (2018) 1449–1473, https://doi.org/10.1007/s11116-017-9771-5.
[56] F. Zhao, N. Park, Using geographically weighted regression models to estimate annual average daily traffic, Transport. Res. Rec. 1879 (1) (2004) 99–107, https://doi.org/10.3141/1879-12.
[57] J.K. Eom, M.S. Park, T.Y. Heo, L.F. Huntsinger, Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method, Transport. Res. Rec. 1968 (1) (2006) 20–29, https://doi.org/10.1177/0361198106196800103.
[58] B. Selby, K.M. Kockelman, Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression, J. Transport Geogr. 29 (2013) 24–32, https://doi.org/10.1016/j.jtrangeo.2012.12.009.
[59] Y. Song, X. Wang, G. Wright, D. Thatcher, P. Wu, P. Felix, Traffic volume prediction with segment-based regression kriging and its implementation in assessing the impact of heavy vehicles, IEEE Trans. Intell. Transport. Syst. 20 (1) (2018) 232–243, https://doi.org/10.1109/TITS.2018.2805817.
[60] S.S. Pulugurtha, S. Mathew, Modeling AADT on local functionally classified roads using land use, road density, and nearest nonlocal road data, J. Transport Geogr. 93 (2021) 103071, https://doi.org/10.1016/j.jtrangeo.2021.103071.
[61] V.R. Duddu, S.S. Pulugurtha, Principle of demographic gravitation to estimate annual average daily traffic: comparison of statistical and neural network models, J. Transport. Eng. 139 (6) (2013) 585–595, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000537.
[62] M. Fu, J.A. Kelly, J.P. Clinch, Estimating annual average daily traffic and transport emissions for a national road network: a bottom-up methodology for both nationally-aggregated and spatially-disaggregated results, J. Transport Geogr. 58 (2017) 186–195, https://doi.org/10.1016/j.jtrangeo.2016.12.002.
[63] S. Das, I. Tsapakis, Interpretable machine learning approach in estimating traffic volume on low-volume roadways, Int. J. Transport. Sci. Technol. 9 (1) (2020) 76–88, https://doi.org/10.1016/j.ijtst.2019.09.004.
[64] A. Sfyridis, P. Agnolucci, Annual average daily traffic estimation in England and Wales: an application of clustering and regression modelling, J. Transport Geogr. 83 (2020) 102658, https://doi.org/10.1016/j.jtrangeo.2020.102658.
[65] B. Shamo, E. Asa, J. Membah, Linear spatial interpolation and analysis of annual average daily traffic data, J. Comput. Civ. Eng. 29 (1) (2015) 04014022, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000281.
[66] A. Ganji, M. Zhang, M. Hatzopoulou, Traffic volume prediction using aerial imagery and sparse data from road counts, Transport. Res. C Emerg. Technol. 141 (2022) 103739, https://doi.org/10.1016/j.trc.2022.103739.
[67] A. Sfyridis, P. Agnolucci, Factors affecting road traffic: identifying drivers of annual average daily traffic using least absolute shrinkage and selection operator regression, Transp. Res. Rec. 2677 (5) (2023) 1178–1192.
[68] T.L. Wang, C. Liu, D. Huang, M. Shahawy, Truck loading and fatigue damage analysis for girder bridges based on weigh-in-motion data, J. Bridge Eng. 10 (1) (2005) 12–20, https://doi.org/10.1061/(ASCE)1084-0702(2005)10:1(12).
[69] P. Chotickai, M.D. Bowman, Truck models for improved fatigue life predictions of steel bridges, J. Bridge Eng. 11 (1) (2006) 71–80, https://doi.org/10.1061/(ASCE)1084-0702(2006)11:1(71).
[70] A. Miano, M. Civera, F. Aloschi, A. Mele, V. De Biagi, B. Chiaia, A. Prota, A framework for the assessment of road network resilience: application to a densely populated urban context, Procedia Struct. Integr. 64 (2024) 311–318, https://doi.org/10.1016/j.prostr.2024.09.253.
[71] M. Calò, S. Ruggieri, A. Nettis, G. Uva, A MTInSAR-based early warning system to appraise deformations in simply supported concrete girder bridges, Struct. Control Health Monit. 2024 (1) (2024) 8978782, https://doi.org/10.1155/2024/8978782.
[72] C. Rainieri, G. Fabbrocino, Automated output-only dynamic identification of civil engineering structures, Mech. Syst. Signal Process. 24 (2010) 678–695, https://doi.org/10.1016/j.ymssp.2009.10.003.
[73] M. Civera, V. Mugnaini, L. Zanotti Fragonara, Machine learning-based automatic operational modal analysis: a structural health monitoring application to masonry arch bridges, Struct. Control Health Monit. 29 (10) (2022) e3028.

[74] M.A. Mendoza-Lugo, M. Nogal, O. Morales-Nápoles, Estimating bridge criticality due to extreme traffic loads in highway networks. Engineering Structures, Eng. Struct. 300 (2024) 117172, https://doi.org/10.1016/j.engstruct.2023.117172.
[75] A. Nettis, A. Nettis, S. Ruggieri, G. Uva, Corrosion-induced fragility of existing prestressed concrete girder bridges under traffic loads, Eng. Struct. 314 (2024) 118302, https://doi.org/10.1016/j.engstruct.2024.118302.
[76] Ministero delle Infrastrutture e dei Trasporti,, Conto Nazionale delle Infrastrutture e dei Trasporti - Anni 2021-2022 (in English: National Count of Infrastructures and Transport – Years 2021-2022), 2023 (in English: Italian Ministry of Infrastructures and Transport, Office of Statistics).
[77] M. Fiorucci, S. Martino, M. Della Seta, L. Lenti, A. Mancini, Seismic response of landslides to natural and man-induced ground vibrations: evidence from the Petacciato coastal slope (central Italy), Eng. Geol. 309 (2022) 106826, https://doi.org/10.1016/j.enggeo.2022.106826.
[78] P. Colonna, P. Intini, Compensation effect between deaths from Covid-19 and crashes: the Italian case, Transp. Res. Interdiscip. Perspect. 6 (2020) 100170, https://doi.org/10.1016/j.trip.2020.100170.
[79] A. Moyano, M. Stępniak, B. Moya-Gómez, J.C. García-Palomares, Traffic congestion and economic context: changes of spatiotemporal patterns of traffic travel times during crisis and post-crisis periods, Transportation 48 (6) (2021) 3301–3324, https://doi.org/10.1007/s11116-021-10170-y.
[80] D. Lord, X. Qin, S.R. Geedipally, Highway Safety Analytics and Modeling, Elsevier, 2021.
[81] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a k-means clustering algorithm, J. Roy. Stat. Soc. Ser. C (Appl. Stat.) 28 (1) (1979) 100–108, https://doi.org/10.2307/2346830.
[82] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: an Introduction to Cluster Analysis, John Wiley & Sons, 2009.
[83] A. Boukerche, J. Wang, Machine learning-based traffic prediction models for intelligent transportation systems, Comput. Network. 181 (2020) 107530, https://doi.org/10.1016/j.comnet.2020.107530.
[84] A. Sroczyński, A. Czyżewski, Road traffic can be predicted by machine learning equally effectively as by complex microscopic model, Sci. Rep. 13 (1) (2023) 14523, https://doi.org/10.1038/s41598-023-41902-y.
[85] S. Chatterjee, J.S. Simonoff, Handbook of Regression Analysis, John Wiley & Sons, 2013.
[86] C.R. Bilder, T.M. Loughin, Analysis of Categorical Data with R, Chapman and Hall/CRC, 2014.
[87] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S-PLUS, Springer Science & Business Media, 2013.
[88] A.B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, A.K. Mohammadian, Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis, Accid. Anal. Prev. 136 (2020) 105405, https://doi.org/10.1016/j.aap.2019.105405.
[89] F. Jiang, J. Ma, A comprehensive study of macro factors related to traffic fatality rates by XGBoost-based model and GIS techniques, Accid. Anal. Prev. 163 (2021) 106431, https://doi.org/10.1016/j.aap.2021.106431.
[90] C. Yang, M. Chen, Q. Yuan, The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: an exploratory analysis, Accid. Anal. Prev. 158 (2021) 106153, https://doi.org/10.1016/j.aap.2021.106153.
[91] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.
[92] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. (2017) 4765–4774.
[93] L.S. Shapley, A value for n-person games, Contrib. Theory Games 2 (28) (1953) 307–317, https://doi.org/10.1515/9781400881970-018.
[94] S.S. Pulugurtha, P.R. Kusam, Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths, Transport. Res. Rec. 2291 (1) (2012) 53–60, https://doi.org/10.3141/2291-07.
[95] R. Bubbico, S. Di Cave, B. Mazzarotta, Risk analysis for road and rail transport of hazardous materials: a GIS approach, J. Loss Prev. Process. Ind. 17 (6) (2004) 483–488, https://doi.org/10.1016/j.jlp.2004.08.011.
[96] P. Peduzzi, H. Dao, C. Herold, F. Mouton, Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index, Nat. Hazards Earth Syst. Sci. 9 (4) (2009) 1149–1159, https://doi.org/10.5194/nhess-9-1149-2009.
[97] R. Rahman, S. Hasan, A deep learning approach for network-wide dynamic traffic prediction during hurricane evacuation, Transport. Res. C Emerg. Technol. 152 (2023) 104126, https://doi.org/10.1016/j.trc.2023.104126.
[98] D. Pantusa, F. D'Alessandro, F. Frega, A. Francone, G.R. Tomasicchio, Improvement of a coastal vulnerability index and its application along the Calabria Coastline, Italy, Sci. Rep. 12 (1) (2022) 21959, https://doi.org/10.1038/s41598-022-26374-w.