Prediction of the severity of exceeding design traffic loads on highway bridges

Roberto Ventura [*], Benedetto Barabino, Giulio Maternini

Department of Civil, Environmental, Architectural Engineering and Mathematics (DICATAM), University of Brescia, Via Branze, 43, 25123, Brescia, Italy

A B S T R A C T

Being a driver of failure consequences, forecasting the severity of events where design traffic load limits on bridges have been exceeded (DLEEs) is fundamental for road safety. Previous research has focused on estimating failure consequences by direct and indirect cost metrics. Only recently has research assessed severity unconventionally, in which the type of DLEEs was predicted by applying econometric models through Binomial Logistic Regression (BLR). Since machine learning models using Artificial Neural Networks (ANN) have not yet been explored, this study will enhance the literature as follows. First, two different 'severity' models were set up as a function of bridge-side, temporal-context, and traffic load hazard variables. Whilst the former relied on a BLR, the latter used an ANN. Second, the performance of these models was assessed using confusion matrixes, some performance indicators, and a cross-entropy parameter. Raw Weigh-In-Motion data on 7.4 M+ individual vehicle transits on a bridge along a primary roadway in Brescia (Italy) were processed. Although a similarly strong performance was achieved for BLR and ANN, the results indicated that ANN was able to predict severity records with a higher level of confidence than BLR on the case study dataset, with the cross-entropy of the ANN less than one third of that of the BLR. These analyses can support road authority traffic management to safeguard bridges from traffic load hazards. Finally, this study recommends future developments, such as considering the structural effects of traffic loads in the modelling, prioritizing traffic management actions among bridges at network level, and exploring the impact of ANN models in risk assessment.

1. Introduction

Vehicular traffic is one of the main hazards that undermine bridge safety. Indeed, according to Ref. [1], vehicular traffic actions (i.e., overloads and collisions) are the most frequent sources of bridge failures after hydraulic actions (e.g., flood and scour). Focusing on the overload risk, exceptionally loaded vehicles pose a serious risk to the soundness of bridges [2]. In fact, truck loads along road infrastructures have been found to frequently exceed legal load restrictions as traffic volumes have increased, and occasionally this has led to bridge failures [3]. Data on extremely high traffic load conditions on bridges can be collected by emerging Intelligent Transportation Systems (ITSs) based architectures. On the one hand, these data can be gathered, saved, and examined for supervising traffic load, which would help Road Authorities (RAs) with bridge maintenance and operational planning. On the other hand, the data can be

---

Nomenclature

ANN	Artificial Neural Network
BLR	Binomial Logistic Regression
CE	Cross Entropy
CM	Confusion Matrix
CPU	Central Processing Unit
DLEE	Design Load Exceedance Event
EM	Econometric Model
FP	Filtering Procedure
GLR	Generalized Liner Regression
GVM	Gross Vehicle Mass
ISLS	Irreversible Serviceability Limit State
ITS	Intelligent Transportation System
LS1	Load Scheme 1
MLM	Machine Learning Model
OIML	Organisation Internationale de Métrologie Légale
PI	Performance Indicator
QCA	Quality Control Algorithm
RAM	Random Access Memory
RA	Road Authority
RSLS	Reversible Serviceability Limit State
SVM	Support Vector Machine
TC	Traffic Code
ULS	Ultimate Limit State
WIM	Weigh-In-Motion

a useful tool for setting up models that forecast bridge failure events induced by traffic load hazards, and thus for aiding data-driven traffic management actions. Amongst ITS-based architectures, Weigh-In-Motion (WIM) systems have been shown to be a valuable way to gather data about traffic loads on bridges [4].

WIM systems provide a dynamic measurement of vehicle mass, allowing for enforcement measures against overloaded trucks and the prevention of roadway and bridge damage [5]. Additionally, these systems can collect information on the total amount of traffic that passes on any given roadway, the date, and time that it passes, as well as the speed and the size of the vehicles, the number of axles on those vehicles, the mass acting on each axle, the type of axles, and the interaxle distances [6]. The main advantage of WIM systems is their ability to capture vehicle parameters in real-time for all passing traffic without the need for a human checker to estimate these data by selecting random vehicles and performing manual weighing operations [7]. On the one hand, WIM systems are common in certain countries (e.g., China, Canada, and United States), where they are exploited for several purposes, involving, for example, freight movement analysis, traffic flow simulation, weight enforcement, bridge design and management, and road pavement design and management (e.g. Refs. [8–13], and [14]). On the other hand, less research has been done in Europe, where WIM are often considered as experimental devices (e.g. Refs. [5,7], and [15]).

Concentrating on the design and management of bridges, much research has utilized WIM data records to achieve several objectives. These goals include enhancing bridge safety by developing customized traffic load models grounded in WIM data, taking into account, e.g., uncertainties stemming from the specific traffic conditions at bridge sites, evolving environmental factors, and the growing number of multi-lane and long span cable-stayed bridges (e.g. Refs. [16–20], and [21]).

A prerequisite for the management of bridge safety is processing WIM data so that the frequency and the severity of failure events induced by traffic load hazards can be promptly predicted. Indeed, without the ability to foresee such events beforehand, it is difficult to take traffic management action to lessen the threat posed by extremely overloaded vehicles. Additionally, the frequency and severity of bridge failure events could be viewed as two key components of risk assessment models. In fact, frequency, and severity are often understood as drivers of the probability and consequences of failure events caused by traffic load hazards, respectively [22].

Particularly, as for the severity, which is the focus of this study, prior research has assessed the consequences of bridge failure events in terms of cost metrics that are direct (e.g., rebuilding and repair costs) and indirect (e.g., increase in vehicular operating costs, increase in users' travel time, injuries, and fatalities), (e.g. Refs. [23–25], and [26]).

However, the procedures for estimating these metrics are time and resource consuming because they require: i) detailed cost analyses to assess rebuilding and repairing costs (e.g. Ref. [26]); ii) refined traffic models that account for the road network topology to determine increases in vehicular operating costs and user travel times (e.g. Refs. [24,27]); iii) complex impact speed/injury probability models to forecast injury and fatality occurrences and, thus, to derive the related costs (e.g. Ref. [26]). Hence, implementing dynamic bridge safety management strategies driven by direct and indirect cost metrics might be problematic. Moreover, financial restrictions often hinder RAs from calibrating such refined cost, traffic, and impact speed/injury probability models.

Conversely, indirectly measuring the consequences of bridge failure events induced by traffic load hazards could become a

straightforward task. Additionally, it could provide a contribution to the cost metrics estimated by previous literature. This process involves detecting different types of events where the design traffic load limits of the bridge have been exceeded (referred to as Design Load Exceedance Events - DLEEs). Particularly, after identifying various levels of design traffic loads according to existing Structural Design Codes, a severity metric can be established. This metric would indirectly measure the consequences of possible bridge failure events[1] by classifying the types of DLEEs according to the different magnitude levels. This severity metric could then be estimated by some models that link it to several traffic-related predictors, to understand both their effects and the significance of the severity. These models could be Econometric Models (EMs) and Machine Learning Models (MLMs).

By applying EMs, [22] specified, calibrated, and validated a Generalized Linear Regression (GLR) model to forecast the severity of DLEEs. This model was a component of a wider framework set up to assess the risk related to traffic load hazards on road bridges. According to ISO 39001 [28], intermediate safety factors have been recognized as predictors of severity. Results indicate that a GLR model can be an interesting tool for predicting the severity component. Notably, parameters associated with adherence to the mass limits stipulated by the Traffic Code (TC) are what exerted the most substantial impact on these predictions.

As for the MLMs, they have been widely recognized computational tools extensively utilized in the field of engineering (e.g. Refs. [29–32], and [33]). Focusing on bridge engineering, MLMs have achieved encouraging levels of performance for several purposes. For example, they can: (a) establish data-driven bridge condition deterioration models at a regional level by relying on inspection reports; (b) develop fragility functions for bridges subjected to pier scour under vehicular loads; (c) quantify bridge damage by processing accelerometric signals ([34,35], and [36]).

Hegde & Rokseth ([37]) categorized MLMs based on their prediction accuracy with varying data availability. When data are limited, boosting techniques and Support Vector Machine (SVM) methods tend to excel, although their performance declines with larger datasets due to extended training times ([38,39]). Conversely, Artificial Neural Network (ANN)-based models are commonly used for forecasting risk components due to their strong predictive capabilities, particularly when ample data, even if noisy, are available ([38]). ANNs can automatically learn essential latent features, making them well-suited for situations like analysing extensive datasets from WIM systems.

Comparing ANNs to GLRs reveals certain strengths and weaknesses that both approaches have shown. On the one hand, ANNs typically offer superior predictive performance and excel in modelling non-linear phenomena ([40]). Conversely, ANNs are considered "black box" models, which means that the parameters of their functional relationships lack physical significance. This poses a challenge for the comprehension of the precise influence of each predictor on the response variable. Nevertheless, methods employing feature importance indicators can efficiently rank predictors based on their significance ([41]).

Ventura et al. [41] had applied an ANN model to forecast the frequency of the occurrence of DLEEs caused by traffic load hazards on bridges. They had obtained promising results; however, no study has yet adopted an MLM to predict the severity component. Hence, there appears to be the need to explore the suitability of ANN models for predicting the severity of DLEEs due to traffic load hazards on bridges. This represents an intriguing research gap that ought to be addressed, including a comparative analysis of their performance in relation to traditional GLR models.

Therefore, to summarize, the main innovations introduced by this study are:

- The introduction of a severity metric that indirectly measures the consequences of possible bridge failure events by classifying the type of Design Load Exceedance Events (DLEEs).
- The specification, calibration, validation, and testing of two severity prediction models using Binomial Logistic Regression (BLR) and Artificial Neural Networks (ANN).
- The provision of an extensive quantitative comparison on the fit and predictive performance of both models.

Raw WIM data from 7.4 M+ individual vehicle transits collected near a major bridge along the heavily trafficked ring road in Brescia (Italy) during a 15-month observation period were processed to fit and compare these models.

The remainder of the paper has been organized as follows: Section 2 describes materials and method used to acquire and process WIM data, detect severity predictors, and to build the BLR and ANN models whilst comparing their performances. Section 3 presents and examines the results of these analyses. Lastly, Section 4 draws some conclusions and offers new perspectives.

2. Materials and method

The overall method was organized into four blocks, according to the scheme outlined in Fig. 1. Each block meets a specific function described below. The first block leveraged WIM systems to acquire data related to the traffic load flow on the monitored bridge. These data were then processed to evaluate the severity predictors and measure the response variables for that severity. Next, the second block established a model for predicting severity through an EM, which involved fitting a BLR and analysing the significance and effect of each predictor integrated into the model. Similarly, the third block built a severity prediction model through an MLM, by using an ANN and scrutinizing the importance of each predictor within the model. Lastly, the fourth block conducted a comparative evaluation of the performance of the two severity models using a set of quantitative metrics.

More details regarding the four blocks are provided in the following subsections.

---

[1] In this study, the term "failure event" does not exclusively signify a collapse. Rather, it encompasses any circumstance that could result in the inability of an existing bridge or its components to meet the specified design and construction requirements.
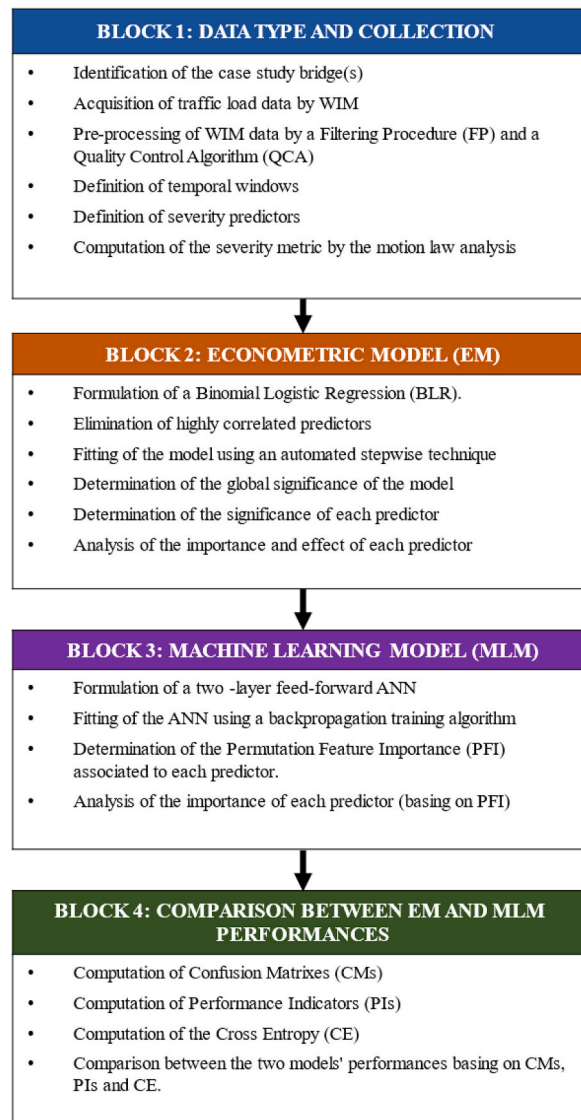
Fig. 1. Flowchart of the proposed methodology.

2.1. BLOCK 1: data type and collection

The data for this study were collected from a bridge on the South Ring Road in Brescia, Italy. Brescia is the second-most populous city in the Lombardy Region and one of Italy's major industrial and economic hubs ([42,43]). The South Ring Road was selected in agreement with the local RA (i.e., the Province of Brescia) because it is a portion of the primary road network and one of the arterial roads with the largest traffic volume and proportion of heavy vehicles in the Province [44]. The case-study bridge (23.5 m span length) is a simply supported overpass structure, with two lanes in each direction of each carriageway, over a secondary urban road (Fig. 2).

An experimental WIM system was installed on the north roadway transition embankment to gather data about the traffic loads passing over the bridge during a monitoring period (denoted by *T*). Only the right lane was instrumented, both because heavy vehicle transit is not allowed in the left lane and because of financial constraints (Fig. 3). The WIM device consisted of two stainless steel plates that were positioned on the road pavement, outfitted with fibre optic sensors, and connected to a data logger. According to the Organisation Internationale de Métrologie Légale (OIML) criteria for WIM devices it was classified as a "10 accuracy class" instrument [45].

Because the WIM data were collected automatically, they may have included errors that could compromise the reliability of the resulting severity analysis. Therefore, the datasets were pre-processed to remove anomalies and outliers. In this study, anomalies were records from which the WIM system was unable to accurately determine some or all of the vehicular parameters, such as tire passage, out plate borders, or passing speed out of the measurement range. Anomalies were automatically detected by the WIM system through
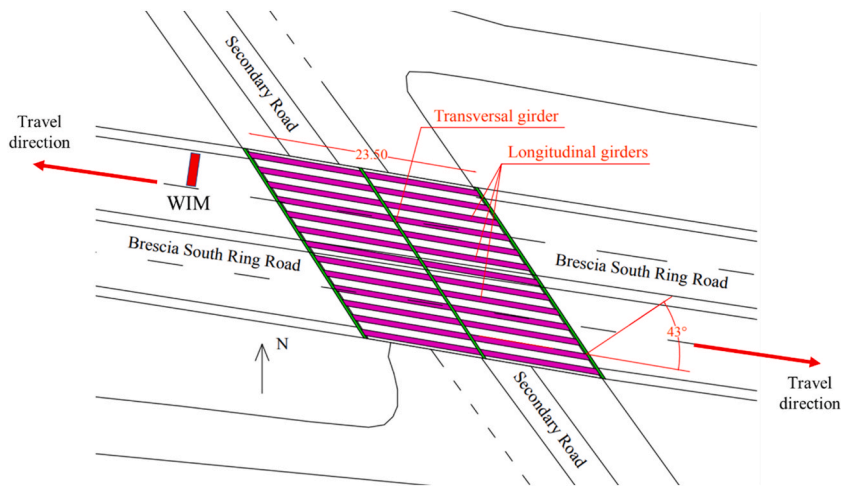
Fig. 2. The case study bridge.

Fig. 3. Plan of the case study bridge.

a proprietary algorithm and eliminated by a straightforward Filtering Procedure (FP). Conversely, outliers were records having inconsistent values for one or more of the vehicular parameters, e.g., if the sum of the mass on each axle was considerably dissimilar from the whole gross vehicular mass. Outliers were removed by a set of filtering criteria implemented through an appropriate Quality Control Algorithm (QCA) developed by the authors by tailoring existing procedures proposed in the literature ([46,47], and [48]). For more details about the QCA, please refer to Ref. [22]. Next, since the arrival of incoming traffic is a stochastic process, the monitoring period was partitioned into equally sized and disjoined temporal windows, to account for the temporal variability of the vehicle loads. This division also made it possible to analyse the variations over time in the characteristics of vehicles on the bridge and in the severity of DLEEs. More formally, let:

- T be the monitoring period during which the severity was modelled.
- S be the set of temporal windows and let $T(s)$ be the subset of T in the temporal window $s \in S$.
- F be the set of severity predictors, let $f \in F$ be the generic predictor, and let f_s be the value of $f \in F$ measured in $T(s)$.
- V_s be the severity of DLEEs measured in $T(s)$, intended as a driver of the associated consequences.

According to Ref. [22], the set of severity predictors were referred to as intermediate safety factors. These predictors included bridge side factors (describing the bridge's geometrical properties), temporal context factors (identifying the day and the time during which each temporal window occurred), and traffic load hazard factors (describing the traffic load hazards on the monitored bridge). Traffic load hazard factors were further allocated into five subgroups: traffic flow characteristics, vehicular characteristics, interaction between vehicular and bridge characteristics, compliance with TC requirements and induced actions on the structure.

Through processing the WIM dataset, the total vehicular load exerted on the bridge lane at each instant $t \in T$ was calculated using axle mass records. Notably, the preference for disaggregated WIM records of axle masses over aggregated gross vehicle masses arose from the possibility of non-integer fractions of vehicles that were simultaneously operating on the bridge lane. Specifically, a motion law analysis was conducted for each vehicular axle to determine the application point of each load during periods other than when it was transiting over the WIM system. This step was essential because the WIM system recorded axle data at a single point, even though the bridge deck spanned a certain length of space. Subsequently, the total vehicle load for each instant $t \in T$ was computed by summing the masses of the axles present on the bridge lane at that instant and performing the necessary conversions from mass to load units.

To measure the severity of bridge overload events, some design traffic lane load thresholds of different magnitude levels were defined. These thresholds needed to be compared with the actual traffic loads on the monitored bridge lane during each temporal window. Eurocode1[2] was chosen as a reference among the existing Structural Calculation Codes. Hence, according to the load combinations recommended by Eurocode 1 for Ultimate Limit State (ULS), Irreversible Serviceability Limit State (ISLS), and Reversible Serviceability Limit State (RSLS), three different design traffic lane load thresholds were established. Specifically, Load Scheme 1 (LS1) was preferred because LS1 is the model set for global structural verifications and, therefore, it was the one that best suited the definition of design traffic load thresholds relating to the whole lane [49].

Typically, severity is described mathematically using an ordered discrete response variable. However, due to the infrequent nature of ULS-associated design load exceedance – an event that may not occur at all during the monitoring period – the severity of such design load exceedance events was represented through a binary variable in the modelling process. Formally, for each $s \in S$, a binary variable (denoted by z_s), which assumed the value of 0 if the event resulted in merely exceeding the design load associated with RSLS and the value of 1 if the event resulted in exceeding the design load associated with ISLS and/or ULS. These values were employed to model the severity (V_s). As a result, V_s was defined as the conditional chance that z_s would equal 1 in $s \in S$ when at least a DLEE was detected during $T(s)$, as formalized in eqn. (1).

$$V_s \overset{\text{def}}{=} P(z_s = 1\,); \; \forall s \in S : \text{"At least one DLEE was detected during } T(s)\text{"}; \tag{1}$$

Further information regarding the severity measurement process were offered in Ref. [22].

2.2. BLOCK 2: econometric model

After computing V_s, the severity forecasting model was initially set in a more traditional manner by employing an EM. Specifically, considering the binary nature of the response variable, a BLR was chosen, as has been applied in other fields (e.g. Refs. [50,51], and [52]). This decision was made to make the results easier to interpret and to suggest possible severity mitigation actions. In fact, they were able to be evaluated using the Odds Ratio (OR), which is easily calculated by taking the exponent of the parameter estimate, which yields the number of "successes" (a severe event) against each of the "non-successes" (a non-severe event). More formally, let:

- $TR \subset S$ and $TE \subset S$ be the training and test subsets, respectively.
- $\widetilde{V}_s$ be the predicted value for observed V_s.
- α and β_f be the coefficients of the severity model.
- p_{GL} be the p-value linked to the model, indicating the overall statistical significance of the model based on the chi-square test.
- p_f be the p-value linked to the severity predictor f, indicating the significance of each individual predictor based on the t-test.
- dr be the deviance ratio, defined as the proportion between the regression deviance and the degree of freedom.

Thus, the prediction of severity was determined based on the functional form indicated in eqn. (2):

$$\widetilde{V}_s = \frac{e^{\alpha + \sum_{f \in F} \beta_f f_s}}{1 + e^{\alpha + \sum_{f \in F} \beta_f f_s}}; \; \forall \, s \in S \tag{2}$$

The dataset was randomly divided into training (TR) and test (TE) subsets before the fitting phase to enable the implementation of a model validation technique that depended on objective out-of-sample evaluations. Thus, the fitting process began with the detection of the model predictors (f). A straightforward selection technique was then used to retrieve the predictors from F. To begin with, an initial filtering process was performed to avoid multicollinearity problems by deleting strongly correlated predictors, as outlined below:

- Calculate the correlation matrix between pairs of predictors $f \in F$, next between each predictor $f \in F$ and the response variable V_s.
- Identify pairs of strongly correlated predictors (i.e., correlation index above 0.8, according to Ref. [53]).
- For each pair of strongly correlated predictors, remove the one that has the lower correlation index with the response variable V_s.

Next, an automated stepwise method incorporating both forward selection and backward elimination was employed to ascertain the optimal set of predictors from the list generated during the initial filtering phase. A forward or backward stepwise approach was favoured based on the estimated model's highest $d.r.$ and p_{GL} values. Once the model was fitted, the significance of each predictor (p_f), the sign of each coefficient and the associated ORs were evaluated to determine the impact of each variable on the forecasted severity.

2.3. BLOCK 3: machine learning model

The severity prediction model was constructed in a more innovative manner, by MLMs. ANNs were preferred over other Machine Learning algorithms, according to Ref. [41]. Let:

---

[2] It is the standard that provides guidelines for the actions to be taken into consideration in the design of bridges within the European Union [72].

- *TR⊂S, VA⊂S* and *TE⊂S* be the training, validation, and test subsets, respectively.
- *IN* ∈ $\mathbb{R}^{|S|,|F|}$ be the input matrix for the ANN fitting procedure, i.e., the matrix of the severity predictors (i.e., *F*) throughout each $T(s) \in S$.
- *TG* ∈ $\mathbb{R}^{|S|}$ be the target vector for the ANN fitting procedure, i.e., the vector of the severity of DLEEs (i.e., V_s) throughout each $T(s) \in S$.
- $\widetilde{TG}$ ∈ $\mathbb{R}^{|S|}$ be the forecasted value for *TG*.
- ω be a function that establishes the relationship between the input matrix *IN* and the target vector *TG*.
- $\widetilde{\omega}$ be an approximation of the function ω.
- *P* be the set of ANN hidden layer perceptrons and let $p \in P$ represent the generic perceptron. In this context, a perceptron serves as the fundamental unit of the network, emulating a biological neuron.
- $\theta \in \mathbb{R}^{|F|\bullet|P|+2|P|+1}$ be the generic vector including the ANN parameters and let $\theta_0 \in \mathbb{R}^{|F|\bullet|P|+2|P|+1}$ be the vector found in the learning phase.
- PFI_f be the permutation feature importance of the predictor $f \in F$.
- $\varepsilon \in \mathbb{R}^{|S|}$ be the vector of residual values, indicating the difference between the target vector *TG* and the predicted one $\widetilde{TG}$.
- $CE(\theta)_{VA}$ be the cross entropy calculated over the validation subset (*VA*), depending on the parameter vector (θ).

Therefore, let us suppose that there is a function ω that links the input matrix *IN* and the target vector *TG*, described as $TG = \omega(IN)$. This can be thought of as a calculation model that connects the causes (*IN*) with their seen consequences (*TG*). The ANN establishes a mapping $TG = \widetilde{\omega}(IN, \theta) + \varepsilon$ and discovers optimal value for the parameter vector that yields the most accurate emulation of ω [54]. The two-layer feed-forward ANN was chosen to carry out this mapping [55].

In this framework, the data progressed forward from input nodes, traversed the hidden nodes and moved towards the output nodes. The network was composed of two layers of perceptrons: a hidden layer with |P| perceptrons and an output layer with a single perceptron. A hyperbolic tangent sigmoid activation function was chosen for the hidden layer due to its antisymmetric form: a quality that facilitates the learning of the network ([56,57]). A sigmoid was selected as the activation function for the output layer because of the binary nature of the response variable.

Prior to beginning the training phase, the entire observation set (i.e., *S*) was randomly separated into three partitions: training, validation, and test. During training, the network was given the training set (*TR*), and the network parameters were tuned to match the training data. The validation set (*VA*) was employed to assess the network generalization and to stop training when generalization ceased to advance, to prevent the overfitting phenomenon. The test set (*TE*) exerted no influence on the training process, thereby becoming a reliable indicator of network performance both throughout the training phase and afterwards. Over the course of the training phase, θ was calibrated to lessen a specific cost function calculated over the validation subset [58]. Specifically, the $CE(\theta)_{VA}$ was chosen as the cost function because it was the most appropriate one for classification problems when outputs are interpreted as probabilities of membership in an indicated class [59]. Therefore, the formalization of the training procedure was as in eqn. (3):

$$\theta_0 = \arg\min_{\theta} CE(\theta)_{VA} \tag{3}$$

A backpropagation algorithm was chosen, because it is widely recognized as a leading method for training feedforward neural networks [54]. Upon completion of the training phase, the optimal parameter vector was identified, resulting in the formulation of the functional form of the ANN severity prediction model (eqn. (4)):

$$\widetilde{V}_s = \widetilde{\omega}(\{f_s \in F\}, \theta_0); \forall s \in S; \tag{4}$$

Notably, training produced different results because of the different initial conditions and the random partitioning. Hence, many instances were generated, and the best model was chosen based on the lowest CE. The relevance of each predictor was then assessed to understand how it affected the severity forecast and to suggest possible mitigation actions. The permutation feature importance (PFI_f) was chosen to achieve this goal since it has been acknowledged as a reliable indicator, particularly for non-linear or opaque estimators [60]. The PFI_f was stated as a reduction in a model score while a specific factor was randomly rearranged [61]. This strategy disrupted the connection between the factor and the target, causing a drop in the model's score. The magnitude of this drop indicated the extent to which the model was reliant on the specific factor. The benefit of this method is that it can be computed over numerous instances with different rearrangements of the factor and that it is model-independent. In this study, the indicator was determined for the test dataset instead of for the training dataset to stress the importance of features in out-of-sample prediction. More precisely, let:

- *N* be the set of random rearrangement instances and let $n \in N$ be the generic instance.
- $IN_{nf} \in \mathbb{R}^{|S|,|F|}$ be the damaged form of the input matrix at the rearrangement $n \in N$, which is generated by randomly reordering the column related to $f \in F$ of the undamaged input matrix *IN*.
- $CE(\theta_0)_{TE}$ be the *CE* of the trained ANN model, computed over the test dataset *TE*, using the undamaged input matrix *IN*.
- $CE(\theta_0)_{TEnf}$ be the *CE* of the trained ANN model, computed over on the test dataset *TE*, using the damaged input matrix IN_{nf}.
- $s = -CE(\theta_0)_{TE}$ be the reference score of the trained ANN model on the undamaged input matrix *IN*.
- $s_{nf} = -CE(\theta_0)_{TEnf}$ be the score of the trained ANN model on the damaged input matrix IN_{nf}.

Thus, numerous permutation instances (|N|) were operated for each explanatory factor $f \in F$ and the related PFI_f was computed by
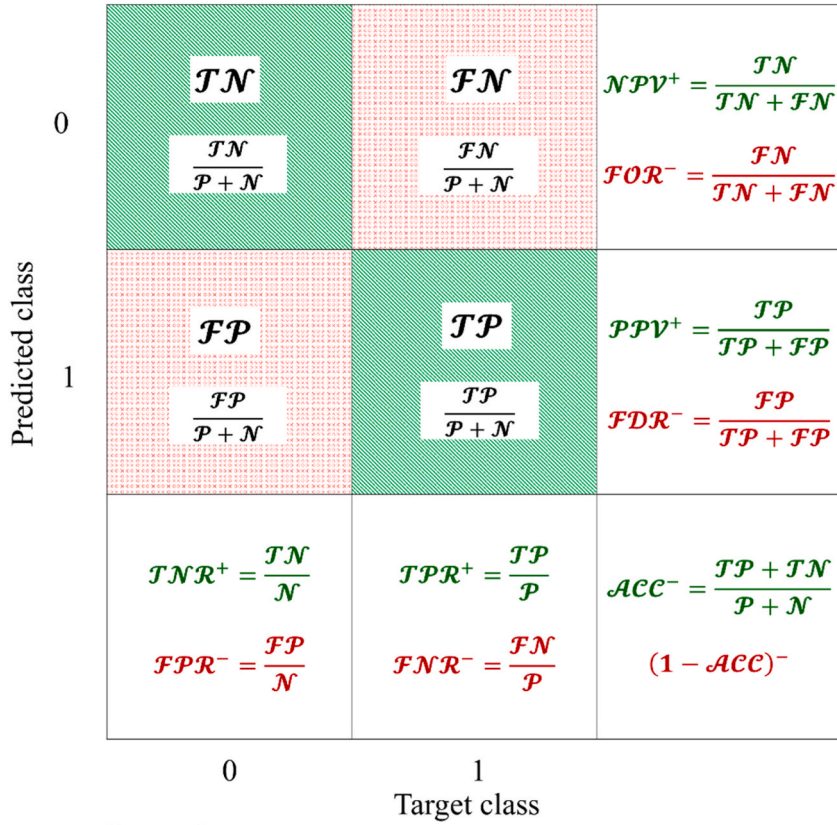
Legend

Filling	Symbol	Meaning
	PI^+	Positively oriented PI (a unit value indicates a perfect classification)
	PI^-	Negatively oriented PI (a null value indicates a perfect classification)

Fig. 4. Confusion matrix and definitions of Performance Indicators (PIs).

averaging the scores across various damaged datasets (eqn. (5)):

$$PFI_f = s - \frac{1}{|N|}\sum_{n \in N} s_{n,f}; \forall f \in F;$$

(5)

2.4. BLOCK 4: comparison of performance between econometric and machine learning

Once the BLR and ANN models were set, they were compared to determine which would be best for fitting and prediction. This comparative analysis involved three approaches based on:

I. Confusion Matrixes (CMs).
II. Performance Indicators (PIs).
III. Cross Entropy (CE).

As for I, the CMs related to the two models were preliminarily graphed for the *TR*, *VA* and *TE* subsets and then compared to one another. Precisely, the CM is a two-dimensional matrix with in one dimension the true class of an object and in the other the class that the classifier assigns. This matrix was obtained by comparing the observed and the predicted values together. Since the output of the BLR and ANN severity models is the predicted conditional probability that the binary variable z_s will be equal to 1 in $s \in S$ (i.e., $\tilde{V}_s$), a cut-off threshold of 0.5 was set to determine the related prediction for z_s [62]. More formally, let:

- $\mathscr{P}$ be the number of real positive cases in the sample, i.e., the total amount of observed z_s unit values.
- $\mathscr{N}$ be the number of real negative cases in the sample, i.e., the total amount of observed z_s null values.
- $\mathscr{TP}$ be the number of true positive cases in the sample, i.e., the total amount of observed z_s unit values that are correctly predicted by the model.
- $\mathscr{TN}$ be the number of true negative cases in the sample, i.e., the total amount of observed z_s null values that are correctly predicted by the model.
- $\mathscr{FP}$ be the number of false positive cases in the sample, i.e., the total amount of observed z_s null values that are wrongly predicted as unit values by the model.
- $\mathscr{FN}$ be the number of false negative cases in the sample, i.e., the total amount of observed z_s unit values that are wrongly predicted as null values by the model.

Therefore, the CM was defined as shown in Fig. 4. Since a binary variable was concerned, the CM turned out to be in a 2*x*2 matrix with the $\mathscr{TN}$ and $\mathscr{TP}$ values along the main diagonal, and the $\mathscr{FP}$ and $\mathscr{FN}$ values along the antidiagonal ([63,64]). Next, the CMs were compared with the understanding that the closer to 0 that the values were along the antidiagonal, the greater the model performance would be since null values for $\mathscr{FP}$ and $\mathscr{FN}$ indicate a perfect prediction.

As for II, some PIs were preliminarily computed on the *TR*, *VA* and *TE* datasets. These PIs included several performance metrics commonly employed when dealing with classifier models. Let:

- $\mathscr{TPR}$ be the true positive rate (or sensitivity) of the severity model.
- $\mathscr{TNR}$ be the true negative rate (or specificity) of the severity model.
- $\mathscr{PPV}$ be the positive predictive value (or precision) of the severity model.
- $\mathscr{NPV}$ be the negative predictive value of the severity model.
- $\mathscr{FNR}$ be the false negative rate (or miss rate) of the severity model.
- $\mathscr{FPR}$ be the false positive rate (or fall-out) of the severity model.
- $\mathscr{FDR}$ be the false discovery rate of the severity model.
- $\mathscr{FOR}$ be the false omission rate of the severity model.
- $\mathscr{ACC}$ be the accuracy of the severity model.

These PIs were defined as stated in the third row and third column of the matrix in Fig. 4.

From these definitions, all the PIs will always fall within the interval $[0,1]$. Next, these indicators were divided into two groups according to their positive or negative relationship with model performance. The former, i.e., (a), where the higher the values were, the higher the prediction performance (i.e., $\mathscr{TPR}$, $\mathscr{TNR}$, $\mathscr{PPV}$, $\mathscr{NPV}$ and $\mathscr{ACC}$) would be. Whilst the latter, i.e., (b), where the higher the values were, the lower the prediction performance (i.e., $\mathscr{FNR}$, $\mathscr{FPR}$, $\mathscr{FDR}$, $\mathscr{FOR}$ and $1 - \mathscr{ACC}$) would be. For ease of visualization, in Fig. 4 these indicators were coloured green and red, respectively. Afterwards, the PIs were compared, considering that a perfect prediction would be designated by unit values for (a) and null values for (b).

As for III, the *CE* parameter was first computed on the *TR*, *VA* and *TE* datasets. Formally, for the training subset (*TR*), this quantity was computed using eqn. (6):

$$CE = -\frac{1}{|TR|} \sum_{s \in Va} z_s \, log(\widetilde{V}_s) + (1 - z_s) \, log(1 - \widetilde{V}_s); \tag{6}$$

As for the *VA* and *TE* subsets, the *CE* was computed using eqn. (6) by replacing *TR* with *VA* and *TE*, respectively. Next, *CE* was compared considering that: a) *CE* heavily penalizes outputs that are extremely inaccurate (i.e., $\widetilde{V}_s$ near $1 - z_s$), with very little penalty for approximately correct classifications (i.e., $\widetilde{V}_s$ near z_s); b) minimizing *CE* generally leads to good classifiers [65].

3. Results and discussion

3.1. BLOCK 1: data type and collection

Over the course of a 15-month observation period (*T*), raw WIM data on 7.4 M+ individual vehicle transits were gathered. Because of its nested structure and substantial size (approximately 9.5 GB), managing the dataset with spreadsheets was not feasible. As a result, all the processing outlined in Section 2 was carried out using a MATLAB© script installed on a mobile workstation (Intel(R) Core

Table 1
Specifics regarding the Weigh-In-Motion (WIM)data collected throughout the monitoring period.

Beginning date	January 1, 2022
Ending date	March 31, 2023
Quantity of raw vehicular records	7,459,312
Quantity of validated vehicular records (after performing FP and QCA)	3,650,538
Length of each temporal window [s]	3600
Quantity of temporal windows	10,824

Table 2
Vehicle classes identified by the Weigh-In-Motion (WIM) device. The proportion of vehicles within each class pertains to the validated traffic database.

Class ID	Class name	Pictorial representation	Percentage of vehicles within the class [%]
1	Cars and vans		93.237%
2	Single unit trucks and buses		0.710%
3	Articulated trucks up to 6 axles		5.047%
4	Road trains up to 6 axles		0.725%
5	Vehicles with more than 6 axles		0.274%
6	Isolated trailers		0.003%
7	Unidentified vehicles	?	0.005%

(TM) i7-10750H CPU @ 2.60 GHz with 2.59 GHz processor, 32.0 GB RAM, Windows 11 Pro 64 bit). Following the pre-processing step, the validated traffic dataset still contained 3.6 M+ vehicle records, or roughly 49 % of the overall traffic flow (Table 1).

Despite the removal of a significant portion of data, this primarily pertains to cars and vans, which are light and quick vehicles for which certain metrics were challenging to measure using WIM. The reason behind this finding was that the WIM data were gathered to assess the weight of the heavier vehicles rather than the lighter ones. Regardless, since the lighter vehicles contributed only a small fraction of the estimated loads acting on the bridge, the accuracy of the subsequent severity analyses was not compromised.

The WIM system identified seven distinct classes of vehicles (Table 2). Whilst cars and vans were the most prevalent vehicle types, a relatively notable percentage (approximately 6.8 %) of commercial vehicles (i.e., classes 2 to 6) were also identified. This observation confirms the substantial presence of these commercial vehicles on the South Ring Road.

Since the severity modelling was done on a yearly basis, 1 h was chosen as the temporal window duration. This led to the creation of a set of 10,824 hourly temporal windows (S) (Table 1). Afterwards, the set of severity predictors (F) was calculated for each of the 6184 temporal windows during which at least one design traffic load exceedance event was detected. Table A1 in the Appendix provides the list of these predictors along with a brief description of each one as well as some descriptive statistics. The severity of DLEEs, which occurred during each $T(s)$ (i.e., the response variable V_s) was then calculated by comparing the total amount of traffic load on the monitored bridge lane at each slot with the design load thresholds imposed by Eurocode 1.

3.2. BLOCK 2: econometric model

Once the WIM data were processed, the severity prediction model was initially fitted by the EM, as shown in eqn. (2). The training (TR) and test (TE) subsets were determined using a splitting ratio of 70 % and 30 %, [66]. Table 3 shows the coefficient estimates and their significance for each factor of the best fit model.

This model provided a good data fit. In fact, the statistical χ^2 test on $d.r.$ yielded a low p_{GL}-value. As a result, the null hypothesis that all the regression coefficients were zero could be refused. The findings demonstrate that many predictors were highly significant (i.e., $p_f \leq 0.05$), because they exhibited a powerful regression effect. When examining each highly significant predictor individually, traffic flow characteristics and compliance with TC prescriptions were the subgroups that had the strongest influence on severity predictions.

As for traffic flow characteristic factors, two predictors proved to be highly significant: Class 6 vehicle fractions and mean speed. Regarding the former, a unit increase in the fraction of isolated trailers strongly increased the probability of recording severe events, as indicated by the very high positive coefficient (i.e., $\beta_1 > 0$) (or, equivalently, by the OR tending toward infinity). This was a novel result, which could be explained by assuming that some carriers may have added a trailer to an already overloaded vehicle to carry a

Table 3
Econometric model. Results of the best fit BLR severity prediction model.

Item	Description	Est. [a]	OR	Item	Description	Est. [a]	OR
α	Regression constant	−1132***	–	β_4	Axle mass – Maximum	0.000693	1.001
β_1	Class 6 (*Isolated trailers*) fraction	5828***	∞	β_5	Interaxle - Mean	16.23*	1.118E+07
β_2	Class 7 (*Unidentified vehicles*) fraction	−3053*	0	β_6	GVM limit ratio – Maximum – All classes	4.09**	59.740
β_3	Speed - Mean	−0.213**	0.808	β_7	GVM limit ratio – Maximum – Class 5 (*Vehicles > 6 axles*)	429.9***	5.049E+186

Source		Degree of freedom	Deviance	Mean deviance	Parameter	Value
Regression		7	1382.43	197.490	d.r.	197.49
Residual		4321	19.36	0.0045	χ^2	.001
Total		4328	1401.79	0.324		

[a] * Denotes that the factor is significant at 0.10 level or lower. ** Denotes that the factor is significant at 0.05 level or lower. *** Denotes that the factor is significant at 0.001 level or lower.

larger payload and, thus, to optimize operating costs. Clearly, this behaviour would have a negative impact on the safety of the bridge because it would lead to an increase in the loads acting on the lane and, thus, to the probability of exceeding ISLS and ULS design traffic load thresholds.

Regarding the latter, the negative coefficient ($\beta_3 < 0$) implies that a 1 km/h increase in the mean speed would decrease the odds of severe events by about 24 % (i.e., OR< 1). This outcome might be explained as follows. First, faster vehicles commonly have a lower GVM and are more likely to fall into lighter categories [7]. Furthermore, as speed increases, vehicles spend less time on the deck, reducing the probability of multiple vehicles being present on the bridge simultaneously. This suggests that implementing excessively low speed limits on bridge decks might have an adverse impact on safety. Given that RAs sometimes set extremely low speed limits for heavy vehicles travelling on damaged or steep bridge decks to reduce dynamic effects, this is a noteworthy finding that merits further attention [2].

As for the compliance with TC prescription factors, two predictors emerged as highly significant: the maximum GVM limit ratio for all vehicular classes and the maximum GVM limit ratio for Class 5 vehicles. Regarding the former predictor, the severity odds increased by about 60 times when the maximum ratio (computed for all vehicle classes) among GVM and the mass limit imposed by the TC took a unit increase ($\beta_6 > 0$, or, equivalently, OR> 1). This result suggests that vehicles surpassing the mass limits stipulated by the TC would pose a greater risk to the bridge than those that adhere to the specified limits. Moreover, and not surprisingly, these findings were in line with those discussed in the literature. Indeed, it has been shown that the likelihood of bridge failure increases with the percentage of overweight trucks [67], especially when they cross the bridge simultaneously (e.g. Refs. [24,25], and [67]). Consequently, compensation fees for overweight trucks have been recommended to balance the economic costs associated with reduced bridge life [68].

Regarding the latter predictor, the extremely high OR (or, equivalently, $\beta_7 \gg 0$) associated with the maximum ratio (computed only for Class 5 vehicles) among the GVM and the mass limits imposed by the TC denotes that, on the one hand, overloaded vehicles with more than 6 axles are those that contribute most towards increasing the likelihood of severe DLEEs on the case study bridge. This result can be explained considering that Class 5 vehicles are the biggest on the roads. Consequently, the load limits imposed by the TC were higher. Thus, the more severe the events will be where design traffic load limits have been exceeded. It is noteworthy that this conclusion should be subject to interpretation, especially when one considers that it is only based on the weight of the vehicles and because it refers to the total vehicle load acting on the monitored bridge lane rather than to the internal actions induced on the individual bridge elements. Indeed, in contrast, it has been proven – e.g., the Federal Bridge Formula (FBF) in the United States – that heavier trucks with more axles produce smaller stresses on bridge elements compared to semi-trailer trucks with 5 axles [69]. Hence, further investigation is required.

3.3. BLOCK 3: machine learning model

The severity model was then fitted using an MLM built on an ANN method according to eqns. (3) and (4). To designate the training (*TR*), validation (*VA*) and test (*TE*) subsets, respectively, a splitting ratio of 70 %, 15 %, and 15 % was chosen [70]. By tuning the hidden layer's number of neurons and the training procedure, numerous attempts were made seeking improvement. Consequently, a network with 35 perceptrons in the hidden layer was selected, as it provided the greatest data fitting. This network was trained using a fast supervised learning algorithm called the Scaled Conjugate Gradient algorithm. As shown in Fig. 5, the training operation was stopped at epoch 88, which corresponded to the minimum value for the cost function calculated on the validation subset ($CE(\theta)_{VA}$).

At the best fit, the permutation feature importance of each severity predictor (PFI_f) was computed according to eqn. (5). The results are shown in the Appendix (Table A2). According to a general perspective, the interaction between vehicular and bridge characteristics and compliance with TC prescriptions were the subgroups of factors with the highest impact on the severity of design load exceedance events. Examining the top 10 most influential predictors individually, the following observations emerge. When considering traffic flow characteristics factors, only one predictor proved highly significant, i.e., the fraction of Class 3 vehicles, which ranked tenth. The

greater importance of this vehicular class over the others can be explained by bearing in mind the fact that there were more Class 3 vehicles among all the commercial vehicles surveyed. When considering vehicular characteristics factors, two predictors turned out to be highly significant (GVM – Mean and GVM – Std. deviation), which were both related to the simple vehicular mass and were ranked second and seventh, respectively. This was a consistent result, because the greater the GVM of the vehicles passing on the bridge was, the greater was the expected severity of design load exceedance events.

Regarding the factors describing the interaction between vehicular and bridge characteristics, two predictors turned out to be highly significant (Normalized GVM – Mean and GVM-length ratio – Std. deviation). The Normalized GVM – Mean was the factor that had the strongest impact on predicting the severity of design load exceedance events, since it ranked first. Hence, it outperformed the other factors that were related to the simple GVM of each passing vehicle. This was a noteworthy result that reflected the findings of [41] for the frequency of design load exceedance events. Indeed, it proved that the load exerted on the bridge deck was more closely associated to the interplay between the GVM and the bridge-span-length/vehicle-length ratio than to the simple GVM alone. One possible explanation could be that if the bridge span was longer than the vehicle, there would be at least one instant during which the whole GVM would be borne by the deck. Otherwise, only a portion of the GVM would be borne by the deck at any given moment. This outcome can be seen as a reminder to RAs and carriers that they should thoroughly assess the ratio of bridge span length to vehicle length on a case-by-case basis before allowing the passage of exceptionally heavily loaded vehicles over a bridge.

Then, the GVM-length ratio – Std. deviation predictor was ranked fourth, implying that the mass linear density of the vehicles strongly influences severity, which has been a consistent finding in our research. Indeed, (for a fixed GVM) shorter vehicles are expected to induce a greater overall load on the bridge rather than longer ones since more vehicles can be on the deck at the same time. Hence, the MLM, the greater the number of vehicles on the span the greater the severity.

Four predictors proved to be highly significant when considering compliance with TC prescription factors. Three of these predictors were related to the ratio among the actual GVM and the corresponding limit prescribed by the TC (i.e., GVM limit ratio – Maximum – Class 5, GVM limit ratio – Mean – Class 2 and GVM limit ratio – Mean – All classes), which were ranked third, fifth and eighth, respectively. These outcomes were consistent with the EM, indicating that the more vehicles that exceeded the mass limits stipulated by the TC the higher would be the likelihood of severe load exceedance events. Moreover, these results proved to be in line with findings in the literature (e.g. Refs. [24,25,67], and [68]). Hence, there is little doubt as to how crucial it is that RAs exercise greater vigilance before granting permits for overloaded vehicles. Additionally, it is strongly advised that more inspections should be conducted to identify illegally overloaded vehicles being driven on bridges without proper authorization. Finally, the last of these four predictors, which concerns the fraction of overloaded Class 5 vehicles, and which was ranked sixth, confirmed the evidence from the EM that overloaded vehicles with more than 6 axles, being the heaviest in the dataset, pose a significant threat to bridge safety.

3.4. BLOCK 4: comparison of performance between econometric and machine learning

To determine the most effective modelling strategy, the fitting, and prediction capabilities of the two severity models were compared. Regarding comparison approach I, Fig. 6 (a) and Fig. 6 (b) present the Confusion Matrixes (CMs) associated with the GLM and ANN models, respectively, for all *TR*, *VA* and *TE* subsets. On the one hand, the CMs suggested similar performance for BLR and ANN models. Indeed, both models exhibited a strong overall fitting and predictive power since they produced only three incorrect predictions out of 6150+ records: one on the training dataset (*TR*), and two on the test dataset (*TE*), respectively. On the other hand, the main distinction among BLR and ANN models was related to the different types of wrong predictions. Whilst two false positives ($\mathscr{FP}$) were predicted by the BLR model, one false positive ($\mathscr{FP}$) and one false negative ($\mathscr{FN}$) were forecasted by the ANN model.
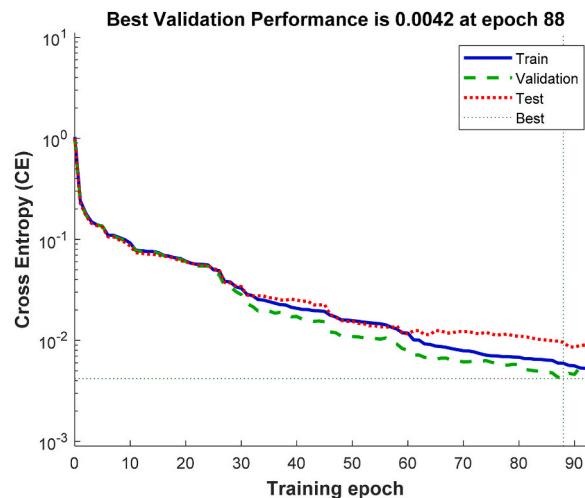

Fig. 5. MLM. The cost function (i.e., cross-entropy) for the ANN severity model plotted against the training epoch. The best validation performance was achieved at epoch 88.

This result might depend on the different modelling structures. Additionally, from a practical perspective, this finding suggested that the BLR was slightly more conservative than the ANN, since no true severe events were missed. Yet, the higher number of the false positives ($\mathscr{FP}$) predicted by the BLR than by the ANN could lead to greater costs due to unnecessary traffic management actions.

Regarding comparison approach II, the values of the Performance Indicators (PIs) that were computed for the two severity models on all *TR*, *VA* and *TE* subsets are summarized in Table 4. Moreover, for ease of visualization, these results are also shown in Fig. 6 (a) and Fig. 6 (b) for the BLR and the ANN severity models, respectively. In these figures, green and red coloured text denotes positively and negatively oriented PIs, respectively.

From one standpoint, identical PI values were obtained on the training dataset (*TR*) for both the BLR and the ANN models. The close-to-unity values for the positively oriented PIs and the close-to-zero values for the negatively oriented PIs indicate that both severity models can strongly fit training data. On the other hand, a slightly greater true positive rate, i.e., $\mathscr{TPR}$, (or, equivalently, a slightly lower false negative rate, i.e., $\mathscr{FNR}$) was found in the test dataset (*TE*) for the BLR model. This was because no truly severe events were missed by the BLR model, whereas one false negative ($\mathscr{FN}$) was predicted by the ANN model.

Regarding comparison approach III, the Cross Entropy (CE) values for both models are summarized in Table 4. These results clearly show that the ANN model significantly outperformed the BLR model both on fitting and forecasting tasks when the CE parameter was considered. Indeed, the CE value for the ANN model was about one order of magnitude lower than for the BLR model, for both the *TR* and *TE* subsets, respectively. This finding suggests that the ANN model predicted severity records with a higher level of confidence than the BLR model. In other words, the predictions of the ANN model were on average farther from the cut-off threshold than those of the BLR model. Indeed, the CE parameter is clearly a measure of the discordance between the correct class of each response variable record

Fig. 6. a) Confusion Matrixes for the best fit BLR severity model; b) Confusion Matrixes for the best fit ANN severity model.

Table 4
Comparison of the fitting and prediction performance of BLR and ANN models.

Parameter	Typology	BLR model		ANN model		
		Training	Test	Training	Validation	Test
II comparison strategy						
True positive rate ($\mathscr{TPR}$)	+	99.4 %	100 %	99.4 %	100 %	97.0 %
True negative rate ($\mathscr{TNR}$)	+	100 %	99.9 %	100 %	100 %	99.9 %
Positive predictive value ($\mathscr{PPV}$)	+	100 %	97.0 %	100 %	100 %	97.0 %
Negative predictive value ($\mathscr{NPV}$)	+	100 %	100 %	100 %	100 %	99.9 %
False negative rate ($\mathscr{FNR}$)	−	0.6 %	0.0 %	0.6 %	0.0 %	3.0 %
False positive rate ($\mathscr{FPR}$)	−	0.0 %	0.1 %	0.0 %	0.0 %	0.1 %
False discovery rate ($\mathscr{FDR}$)	−	0.0 %	3.0 %	0.0 %	0.0 %	3.0 %
False omission rate ($\mathscr{FOR}$)	−	0.0 %	0.0 %	0.0 %	0.0 %	0.1 %
Accuracy ($\mathscr{ACC}$)	+	100 %	99.9 %	100 %	100 %	99.8 %
III comparison strategy						
Cross Entropy (CE)	−	0.0202	0.0435	0.0059	0.0042	0.0096

+ : Positively oriented score (more is better).
− : Negatively oriented score (less is better).

and of the probability that the same response variable record belongs to the correct class according to the model's forecast.

The reason of these findings could be attributed to the greater capacity of ANNs to model the intricate (non-linear) relationships between the severity of bridge overload events and the factors that influence it. Moreover, the outcome appears to confirm what has already been observed in other fields of engineering system safety literature (e.g. Refs. [40,71]). Indeed, for example [71], proved that ANNs outperformed traditional statistical models in forecasting the severity of traffic crash injuries. Consequently, this result also seems to suggest that RAs and practitioners should give preference to ANNs over GLRs when modelling the severity of bridge overload events, particularly when the priority is predictive accuracy rather than delving into the effects of individual explanatory factors.

4. Conclusions

4.1. Contributions

Truck weights often exceed legally defined load limits, occasionally leading to bridge failures. Therefore, an essential prerequisite to improving bridge safety would be the provision of models to predict the severity of events in which bridge design traffic loads are exceeded (DLEEs).

This research is meant to contribute to the literature by setting up two severity prediction models. The first is an Econometric Model that applies a Binomial Logistic Regression (BLR) and the second is a Machine Learning model that applies an Artificial Neural Network (ANN). A comparison of the performance of the two models has been accomplished through associating Confusion Matrixes (CMs), some Performance Indicators (PIs) and a Cross Entropy (CE) parameter. The large dataset gathered by a Weigh-In-Motion (WIM) system near a bridge along a main road in Brescia (Italy) was set up to fit the two severity models. On the one hand, findings indicated a similar (and strong) fit and predictive power for the BLR and ANN models when the Confusion Matrixes (CMs) and the Performance Indicators (PIs) were considered. On the other hand, the ANN model showed a CE value one order of magnitude lower than the BLR model, implying that the former predicted severity records with higher level of confidence than the latter.

As far as the authors know, this research provides the first empirical contribution into the potentialities of MLMs in predicting the severity of DLEEs brought on by traffic load hazards on a road bridge.

4.2. Implications for research and practice

The results of this research have important theoretical and practical implications. From the theoretical perspective, the availability of straightforward but effective models to forecast severity of design load exceedance events can be a meaningful contribution to the existing cost metrics of bridge failure consequences. From a practical viewpoint, grasping the effects and significance of traffic-related predictors on severity can serve as a valuable regulatory guide for Road Authorities (RAs). This level of understanding could very well empower RAs to deploy efficient traffic management strategies that would enhance the safety of bridges against the risks posed by the traffic load hazards. For example, industrial applications derived from Intelligent Transportation System (ITS) architectures, incorporating elements such as traffic lights, variable message signals, and cloud computing platforms, could be designed to redirect groups of vehicles that might present a significant risk of overloading a specific bridge.

4.3. Limitations and future developments

Lastly, this study also recommends the application of new research developments to overcome certain limitations. First, this research relied principally on total lane load values as a basic indicator to gauge the severity of overloading events and their

subsequent consequences. However, bridge design codes have established limit-state thresholds concerning the effects of traffic loads, such as internal forces or structural deformations, which can arise from various configurations of axle loads on the bridge deck. Therefore, a more accurate assessment of a bridge's structural response is recommended. This assessment should also consider the dynamic effects of vehicles passing over the bridge. A thorough investigation in this regard would enhance the identification of limit state exceedance events caused by traffic overloading. Clearly, these findings will be applicable to future research.

Moreover, obviously, this study was able to analyse traffic load data from just one WIM station on just one bridge. Consequently, it was not capable of establishing any universal patterns for all the bridges within a network. To address this deficiency, the methodology ought to be expanded to a network level, which should also involve prioritizing traffic management actions among bridges by considering the impact of traffic disruptions on users' travel times. Finally, it would be of vital importance to investigate the impact of ANN modelling approaches in assessing the risks of traffic loads on bridges (e.g. Ref. [22]).

Data availability statement

The data processed in this study have not been deposited into a public repository due to their confidential nature.

CRediT authorship contribution statement

Roberto Ventura: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Benedetto Barabino:** Writing – review & editing, Visualization, Supervision, Methodology, Funding acquisition, Conceptualization. **Giulio Maternini:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Benedetto Barabino reports financial support was provided by Ministero dell'Istruzione, dell'Università e della Ricerca. Roberto Ventura reports equipment, drugs, or supplies was provided by Provincia di Brescia. Benedetto Barabino serves as an Associate Editor for the Heliyon journal. If there are other authors, they 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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23374.

References

[1] D. Proske, Bridge Collapse Frequencies versus Failure Probabilities, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-73833-8.

[2] R. Ventura, B. Barabino, D. Vetturi, G. Maternini, Bridge safety analysis based on the function of exceptional vehicle transit speed, Open Transport. J. 14 (2020) 222–236, https://doi.org/10.2174/1874447802014010222.

[3] G. Zhang, Y. Liu, J. Liu, S. Lan, J. Yang, Causes and statistical characteristics of bridge failures: a review, J. Traffic Transport. Eng. 9 (2022) 388–406, https://doi.org/10.1016/j.jtte.2021.12.003.

[4] M. Sujon, F. Dai, Application of weigh-in-motion technologies for pavement and bridge response monitoring: state-of-the-art review, Autom. ConStruct. 130 (2021), 103844, https://doi.org/10.1016/j.autcon.2021.103844.

[5] R. Ventura, B. Barabino, D. Vetturi, G. Maternini, Monitoring vehicles with permits and that are illegally overweight on bridges using Weigh-In-Motion (WIM) devices: a case study from Brescia, Case Stud. Transp. Pol. (2023), 101023, https://doi.org/10.1016/j.cstp.2023.101023.

[6] ASTM, ASTM E1318-09, Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Methods, 2017.

[7] R. Ventura, B. Barabino, D. Vetturi, G. Maternini, Bridge's vehicular loads characterization through Weight-In-Motion (WIM) systems. The case study of Brescia, Eur. Transport/Trasport. Eur. 00 (2023) 1–12, https://doi.org/10.48295/ET.2023.90.6.

[8] P. Huang, J. Wang, X. Xu, G. Yang, S. Chen, Y. Yuan, W. Han, Improved multi-lane traffic flow simulation based on weigh-in-motion data, Measurement 188 (2022), 110408, https://doi.org/10.1016/j.measurement.2021.110408.

[9] J. Ren, R.G. Thompson, L. Zhang, Impact of payload spectra of heavy vehicles on pavement based on weigh-in-motion data, J. Transport. Eng., Part B: Pavements. 145 (2019), 04019005, https://doi.org/10.1061/jpeodx.0000099.

[10] S.V. Hernandez, A. Tok, S.G. Ritchie, Integration of Weigh-in-Motion (WIM) and inductive signature data for truck body classification, Transport. Res. C Emerg. Technol. 68 (2016) 1–21, https://doi.org/10.1016/j.trc.2016.03.003.

[11] H.-J. Roh, P.K. Sahu, S. Sharma, S. Datla, B. Mehran, Statistical investigations of snowfall and temperature interaction with passenger car and truck traffic on primary highways in Canada, J. Cold Reg. Eng. 30 (2016), https://doi.org/10.1061/(asce)cr.1943-5495.0000099.

[12] G. Fiorillo, M. Ghosn, Procedure for statistical categorization of overweight vehicles in a wim database, J. Transport. Eng. 140 (2014) 1–12, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000655.

[13] C.-F. Liao, Generating reliable freight performance measures with truck GPS data, transportation research record, J. Transport. Res. Board 2410 (2014) 21–30, https://doi.org/10.3141/2410-03.

[14] L. Guo, J. Wei, S. Cui, Study of load spectra of the jaozhou bay expressway based on MEPDG, Appl. Mech. Mater. 97–98 (2011) 402–407. https://dx.doi.org/10.4028/www.scientific.net/AMM.97-98.402.

[15] F. Schmidt, B. Jacob, F. Domprobst, Investigation of truck weights and dimensions using WIM data, Transport. Res. Procedia 14 (2016) 811–819, https://doi.org/10.1016/j.trpro.2016.05.029.

[16] J. Kim, J. Song, Bayesian updating methodology for probabilistic model of bridge traffic loads using in-service data of traffic environment, Struct. Infrastruct. Eng. 0 (2021) 1–16, https://doi.org/10.1080/15732479.2021.1924797.

[17] J. Zhou, Z. Chen, J. Yi, H. Ma, Investigation of multi-lane factor models for bridge traffic load effects using multiple lane traffic data, Structures 24 (2020) 444–455, https://doi.org/10.1016/j.istruc.2020.01.031.

[18] E.S. Hwang, D.Y. Kim, Live load model for long span steel cable bridges considering traffic congestion scenarios, Int. J. Steel Struct. 19 (2019) 1996–2009, https://doi.org/10.1007/s13296-019-00259-7.

[19] N. Lu, Y. Ma, Y. Liu, Evaluating probabilistic traffic load effects on large bridges using long-term traffic monitoring data, Sensors (2019) 19, https://doi.org/10.3390/s19225056.

[20] E.A. Micu, A. Malekjafarian, E.J. Obrien, M. Quilligan, R. McKinstray, E. Angus, M. Lydon, F.N. Catbas, Evaluation of the extreme traffic load effects on the Forth Road Bridge using image analysis of traffic data, Adv. Eng. Software 137 (2019), 102711, https://doi.org/10.1016/j.advengsoft.2019.102711.

[21] N. Lu, Y. Liu, M. Beer, Extrapolation of extreme traffic load effects on a cable-stayed bridge based on weigh-in-motion measurements, Int. J. Reliab. Saf. 12 (2018) 69–85, https://doi.org/10.1504/IJRS.2018.092504.

[22] R. Ventura, B. Barabino, G. Maternini, Traffic hazard on main road's bridges: real-time managing the risk of design load overcoming events - preprint, Soc. Sci. Res. Netw. (SSRN). (2023), https://doi.org/10.2139/ssrn.4396. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4396914 (Accessed 3 April 2023).

[23] 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.

[24] G. Fiorillo, M. Ghosn, Risk-based life-cycle analysis of highway bridge networks under budget constraints, Struct. Infrastruct. Eng. 0 (2022) 1–15, https://doi.org/10.1080/15732479.2022.2059525.

[25] G. Fiorillo, M. Ghosn, Risk-based importance factors for bridge networks under highway traffic loads, Struct. Infrastruct. Eng. 15 (2019) 113–126, https://doi.org/10.1080/15732479.2018.1496119.

[26] S.M. Wong, C.J. Onof, R.E. Hobbs, Models for evaluating the costs of bridge failure, Proc. Inst. Civ. Eng.: Bridge Eng. 158 (2005) 117–128, https://doi.org/10.1680/bren.2005.158.3.117.

[27] D.Y. Yang, D.M. Frangopol, Life-cycle management of deteriorating bridge networks with network-level risk bounds and system reliability analysis, Struct. Saf. 83 (2020), 101911, https://doi.org/10.1016/j.strusafe.2019.101911.

[28] ISO, ISO 39001: Road Traffic Safety (RTS) Management Systems: Requirements with Guidance for Use, 2012.

[29] R. Ventura, A. Ghirardi, D. Vetturi, G. Maternini, B. Barabino, Comparing the vibrational behaviour of e-kick scooters and e-bikes. Evidence from Italy, Int. J. Transport. Sci. Technol. (2023), https://doi.org/10.1016/j.ijtst.2023.10.010.

[30] V. Vishnu, M.P. Harikrishnan, A.S. Warrier, N.K. Mahanti, M. Basil, T. Venkatesh, R. Pandiselvam, A. Kothakota, Design consideration and optimization of process parameters in fiber extraction unit via modelling studies, J. Food Process. Eng. 46 (2023), https://doi.org/10.1111/jfpe.14298.

[31] R. Pandiselvam, V. Prithviraj, M.R. Manikantan, P.P.S. Beegum, S.V. Ramesh, S. Padmanabhan, A. Kothakota, A.C. Mathew, K.B. Hebbar, A. Mousavi Khaneghah, Central composite design, Pareto analysis, and artificial neural network for modeling of microwave processing parameters for tender coconut water, Measurement: Food 5 (2022), https://doi.org/10.1016/j.meafoo.2021.100015.

[32] V. Srikanth, G.K. Rajesh, A. Kothakota, R. Pandiselvam, N. Sagarika, M.R. Manikantan, K.P. Sudheer, Modeling and optimization of developed cocoa beans extractor parameters using box behnken design and artificial neural network, Comput. Electron. Agric. 177 (2020), https://doi.org/10.1016/j.compag.2020.105715.

[33] R. Pandiselvam, M.R. Manikantan, S. Sunoj, S. Sreejith, S. Beegum, Modeling of coconut milk residue incorporated rice-corn extrudates properties using multiple linear regression and artificial neural network, J. Food Process. Eng. 42 (2019), https://doi.org/10.1111/jfpe.12981.

[34] Y. Xia, X. Lei, P. Wang, L. Sun, Artificial intelligence based structural assessment for regional short-and medium-span concrete beam bridges with inspection information, Rem. Sens. 13 (2021), https://doi.org/10.3390/rs13183687.

[35] S. Kameshwar, J.E. Padgett, Parameterized fragility assessment of bridges subjected to pier scour and vehicular loads, J. Bridge Eng. 23 (2018) 1–11, https://doi.org/10.1061/(asce)be.1943-5592.0001240.

[36] P.J. Chun, H. Yamashita, S. Furukawa, Bridge damage severity quantification using multipoint acceleration measurement and artificial neural networks, Shock Vib. (2015) 2015, https://doi.org/10.1155/2015/789384.

[37] J. Hegde, B. Rokseth, Applications of machine learning methods for engineering risk assessment – a review, Saf. Sci. 122 (2020), https://doi.org/10.1016/j.ssci.2019.09.015.

[38] X. Wen, Y. Xie, L. Jiang, Z. Pu, T. Ge, Applications of machine learning methods in traffic crash severity modelling: current status and future directions, Transport Rev. 41 (2021) 855–879, https://doi.org/10.1080/01441647.2021.1954108.

[39] Susmita Ray, A quick review of machine learning algorithms, in: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, 2019.

[40] A. Alqatawna, A.M. Rivas Alvarez, S.S.C. Garcia-Moreno, Comparison of multivariate regression models and artificial neural networks for prediction highway traffic accidents in Spain: a case study, Transport. Res. Procedia 58 (2021) 277–284, https://doi.org/10.1016/j.trpro.2021.11.038.

[41] R. Ventura, B. Barabino, G. Maternini, Estimating the Frequency of Design Traffic Loads Overcoming on Road's Bridges - Preprint, Social Science Research Network (SSRN), 2023.

[42] M. De Aloe, R. Ventura, M. Bonera, B. Barabino, G. Maternini, Applying cost–benefit analysis to the economic evaluation of a tram-train system: evidence from Brescia (Italy), Res. Transport. Bus. Manag. 47 (2022), 100916, https://doi.org/10.1016/j.rtbm.2022.100916.

[43] V. Martinelli, R. Ventura, M. Bonera, B. Barabino, G. Maternini, Effects of urban road environment on vehicular speed. Evidence from Brescia (Italy), Transport. Res. Procedia 60 (2022) 592–599, https://doi.org/10.1016/j.trpro.2021.12.076.

[44] C. Faccin, L. Zavanella, E. Savoldi, A. Boroni, P. Rossini, Piano del traffico della viabilitá extraurbana, PTVE), 2011.

[45] OIML, OIML R 134-1: Automatic Instruments for Weighing Road Vehicles in Motion and Measuring Axle Loads Part 1: Metrological and Technical Requirements – Tests, 2006, 2006, pp. 1–81.

[46] A.R. Babu, O. Iatsko, A.S. Nowak, Comparison of bridge live loads in US and Europe, Struct. Eng. Int. 29 (2019) 84–93, https://doi.org/10.1080/10168664.2018.1541334.

[47] B. Enright, E.J. Obrien, Cleaning weigh-in-motion data: techniques and recommendations, Int. Soc. Weigh-In-Motion (2011) 2015–2019. https://www.academia.edu/download/43335567/Cleaning_Weigh-in-Motion_Data_Techniques20160303-2533-8efmds.pdf.

[48] B. Sivakumar, M. Ghosn, F. Moses, Protocols for Collecting and Using Traffic Data in Bridge Design Prepared for National Cooperative Highway Research Program, 2008.

[49] R. Ventura, Traffic Hazard on Main Road's Bridges: Real-Time Evaluation and Management of the Risk Related to Design Load Overcoming Events, University of Brescia, 2023.

[50] B. Barabino, M. Bonera, G. Maternini, A. Olivo, F. Porcu, Bus crash risk evaluation: an adjusted framework and its application in a real network, Accid. Anal. Prev. 159 (2021), 106258, https://doi.org/10.1016/j.aap.2021.106258.

[51] B. Barabino, M. Di Francesco, R. Ventura, Evaluating fare evasion risk in bus transit networks, Transp. Res. Interdiscip. Perspect. (2023), https://doi.org/10.1016/j.trip.2023.100854.

[52] S. Kameshwar, J.E. Padgett, Multi-hazard risk assessment of highway bridges subjected to earthquake and hurricane hazards, Eng. Struct. 78 (2014) 154–166, https://doi.org/10.1016/j.engstruct.2014.05.016.

[53] N. Shrestha, Detecting multicollinearity in regression analysis, Am. J. Appl. Math. Stat. 8 (2020) 39–42, https://doi.org/10.12691/ajams-8-2-1.

[54] G. Ian, B. Yoshua, A. Courville, Deep Learning, 2016. https://www.deeplearningbook.org.

[55] J. Schmidhuber, Deep Learning in neural networks: an overview, Neural Network. 61 (2015) 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.

[56] M. Sramka, M. Slovak, J. Tuckova, P. Stodulka, Improving clinical refractive results of cataract surgery by machine learning, 2019, PeerJ (2019) 1–23, https://doi.org/10.7717/peerj.7202.

[57] I.v. Romero Reyes, I.v. Fedyushkina, V.S. Skvortsov, D.A. Filimonov, Prediction of progesterone receptor inhibition by high-performance neural network algorithm, Int. J. Math. Models Methods Appl. Sci. 7 (2013) 303–310.

[58] J. Brownlee, Loss and Loss Functions for Training Deep Learning Neural Networks, 2019. https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/. (Accessed 18 June 2022).

[59] R. Reed, R.J. Marks II, Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks, MIT Press, 1999.

[60] C. Strobl, A.L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: illustrations, sources and a solution, BMC Bioinf. 8 (2007), https://doi.org/10.1186/1471-2105-8-25.

[61] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[62] E.E. Hassan, D. Zhang, The usage of logistic regression and artificial neural networks for evaluation and predicting property-liability insurers' solvency in Egypt, Data Sci. Finance Econ. 1 (2021) 215–234, https://doi.org/10.3934/DSFE.2021012.

[63] MathWorks, Create Confusion Matrix Chart for Classification Problem, 2022. https://it.mathworks.com/help/stats/confusionchart.html. (Accessed 3 August 2022).

[64] K.M. Ting, Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, https://doi.org/10.1007/978-0-387-30164-8.

[65] MathWorks, Crossentropy - Neural Network Performance, 2022. https://it.mathworks.com/help/deeplearning/ref/crossentropy.html. (Accessed 19 June 2022).

[66] A. Fertlitsch, Deep Learning Patterns and Practices, 2021.

[67] G. Fiorillo, M. Ghosn, Fragility analysis of bridges due to overweight traffic load, Struct. Infrastruct. Eng. 14 (2018) 619–633, https://doi.org/10.1080/15732479.2017.1380675.

[68] O.E. Gungor, I.L. Al-Qadi, J. Mann, Detect and charge: machine learning based fully data-driven framework for computing overweight vehicle fee for bridges, Autom. ConStruct. 96 (2018) 200–210, https://doi.org/10.1016/j.autcon.2018.09.007.

[69] FHWA - Federal Highway Administration, Bridge Formula Weights, 2019. https://ops.fhwa.dot.gov/freight/publications/brdg_frm_wghts/. (Accessed 21 August 2023).

[70] P. Flach, Machine Learning: the Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, USA, 2012.

[71] Q. Zeng, H.-L. Huang, P.-P. Xu, M. Ma, Developing an Optimized Artificial Neural Network to Predict Traffic Crash Injury Severity, 2014.

[72] The European Union, Eurocode 1: Actions on Structures - Part 2: Traffic Loads on Bridges, 2003, p. 1.