Transport Policy 178 (2026) 103971

Contents lists available at ScienceDirect

Transport Policy

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

Comparing fare evasion severity by econometric and artificial intelligence models: An Italian case study

Benedetto Barabino, Roberto Ventura *

Department of Civil Engineering, Environment, Land and Architecture Engineering and Mathematics, University of Brescia, Via Branze 43, 25123, Brescia, Italy

ARTICLE INFO

Keywords:
Fare evasion
Severity
Econometric models
Machine learning models
Confusion matrices
Performance metrics

ABSTRACT

Fare evasion is a pressing issue in public transport networks, impacting the financial sustainability of Transit Agencies (TAs) and Public Transport Companies (PTCs). While prior studies have largely focused on the probability of fare evasion (or frequency), research on severity— e.g., the financial and operational impact of detected fare evasion cases—remains limited. This study addresses this gap by specifying, calibrating, and validating two prediction models for fare evasion severity using real-world survey data on passengers from a mid-sized Italian PTC. Two approaches are employed: an Econometric Approach (EA) that uses Logistic Regression Models (LRMs) and a Machine Learning Approach (MLA) leveraging an Artificial Neural Network Model (ANNM). Model performance is evaluated and compared using Confusion Matrices, Metrics robust to class imbalance (e.g., Area Under the Precision-Recall Curve, Balanced Accuracy), and Probability Calibration tools (e.g., reliability curves, Brier score). Probability thresholds (cut-offs) are enhanced to improve predictive performance under imbalanced conditions. Finally, each predictor effect is assessed for both models. Results indicate that the ANNM slightly outperforms the LRM in this case study, demonstrating higher predictive accuracy and a stronger ability to detect high-severity fare evasion cases. However, this gain entails a minor rise in false positives, reflecting the trade-off between predictive accuracy and calibration stability. The LRM remains valuable for policy analysis, offering consistent and interpretable probability estimates to help TAs/PTCs understand key factors influencing fare evasion severity. These findings provide critical insights for enhancing fare inspection policies and enforcement resource allocation.

1. Background

Fare evasion—when a passenger lacks a ticket or holds an incorrect one—significantly threatens the financial viability of Transit Authorities and Public Transport Companies (TAs/PTCs) (Delbosc and Currie, 2019; Barabino et al., 2020). In Proof-of-Payment Transit Systems (POP-TS), passengers must purchase tickets prior to boarding. However, fare payment is not systematically checked, enabling opportunistic behaviours[1] (among others) of passengers. To address this issue, TAs/PTCs can implement various actions, where the passenger inspections and fine imposition are the most practised (Multisystems Inc et al., 2002; Sasaki, 2014; Guarda et al., 2016; Dai et al., 2017; Barabino and Salis, 2019; Wolfgram et al., 2022; Alhassan et al., 2022; Barabino et al., 2024a; Escalona et al., 2024, 2025). These actions could be guided by assessing the severity of fare evasion, particularly in terms of its economic and operational impact on transit systems.

Understanding how fare evasion affects revenue, service delivery, and system equity is essential for maintaining financial sustainability and passenger trust. Severe cases of fare evasion can lead to substantial financial losses, strain enforcement resources, and damage the perceived fairness of the system, potentially discouraging fare-paying passengers. By identifying where the most economically harmful evasion occurs, TAs/PTCs can allocate resources more effectively and design policies that address the root causes without disproportionately affecting 'vulnerable' passengers. Indeed, while assessing the severity of fare evasion is essential, its management must also consider equity implications in enforcement strategies. Unequal targeting of vulnerable groups—such as low-income individuals, students, or the elderly—-through profiling raises serious ethical concerns and undermines social trust (Barabino et al., 2024a). Equitable inspection should avoid social bias and instead ensure fairness across all passenger segments. For instance, evidence from Chile showed that perceptions of social injustice

* Corresponding author.
E-mail addresses: benedetto.barabino@unibs.it (B. Barabino), roberto.ventura@unibs.it (R. Ventura).

[1] In this paper, opportunistic behaviour refers to individuals taking advantage of the perceived low likelihood of inspection to evade payment.

https://doi.org/10.1016/j.tranpol.2025.103971
Received 24 July 2025; Received in revised form 19 December 2025; Accepted 20 December 2025
Available online 22 December 2025

or opposition to privatised transport systems can fuel fare evasion ideologically (e.g., Busco et al., 2021, 2024). Therefore, effective fare policies must strike a balance between operational severity and perceived fairness to ensure both efficiency and public support. Addressing the severity of fare evasion also enhances transparency and public confidence, supporting long-term planning and policy.

This research is inspired by the preceding context and insights from the following literature review. It highlights existing research gaps and outlines the specific objectives addressed in this study.

1.1. State of the art

The probability (or frequency) and severity of fare evasion are key dimensions to describe this crucial threat for TAs/PTCs.

On the one hand, the probability of fare evasion reflects the chance that a passenger avoids payment—intentionally or not—during a trip and is influenced by various predictors (attributes, variables, or factors). This probability has been explored through Descriptive Models (DMs), Econometric Models (EMs), and more recently, Machine Learning Models (MLMs). DMs identify trends, as in Geographic Information System (GIS)-based studies in San Francisco and New York linking evasion to socio-economic disparities (Lee, 2011; Reddy et al., 2011). EMs, especially probit and logistic regressions, have modelled fare evasion (or the intent to evade the fare) using socio-demographics and behavioural traits according to one size-fits-all model[2] (Bucciol et al., 2013; Barabino et al., 2015; Dai et al., 2017; Cools et al., 2018). Moreover, fare evasion likelihood has been examined through *a priori* and *a posteriori* segmentation (Barabino et al., 2024a). The former classifies passengers by traits such as age, employment, and prior fines, identifying high-risk groups like young students and unemployed individuals with violations (Barabino and Salis, 2020; Barabino et al., 2022). Traits like honesty and ticketing familiarity also influence behaviour (Currie and Delbosc, 2017). The latter reveals consistent profiles in developed contexts, such as deliberate, unintentional, and non-evaders (e.g., Delbosc and Currie, 2016; Barabino and Salis, 2023). In contrast, studies in developing countries highlight sociocultural influences—social norms, service perception, and enforcement fear—as key drivers (e.g., González et al., 2019; Busco et al., 2021). Hybrid models further link age, prior fines, and satisfaction to evasion behaviour, revealing the influence of socioeconomic context (Guzman et al., 2021). Few studies have addressed evasion frequency: some have linked it to economic and operational factors (Troncoso and de Grange, 2017; Guarda et al., 2016; Cantillo et al., 2022), others found it associated with occupancy, service exposure, and inspections (e.g., Barabino et al., 2023; Freiria and Sousa, 2025). MLMs are still rarely used in fare evasion, despite their success elsewhere (Schmidhuber, 2015; Cao, 2017; Fatima and Pasha, et al., 2017; Boukerche and Wang, 2020; Sarker et al., 2021; Jhaveri et al., 2022). Early applications of MLMs in fare evasion identified key predictors like inspection type and passenger volume (Barabino et al., 2024b; Barabino and Ventura, 2025). Specifically, in these last studies "inspection type" refers to the organisational configuration of the inspection strategy, distinguishing targeted inspections (performed along pre-defined routes and often assisted by checkers) from random inspections (conducted by inspectors only). MLMs found that fare evasion frequency tends to increase under targeted inspections—due to their focus on high-demand segments—and with higher passenger volumes, which are positively associated with the likelihood of fare evasion events.

On the other hand, as far as the authors know, only few studies have explored fare evasion severity events. It could be referred to as the most likely consequence of a possible evader, comprising things like lost revenues and/or violence on board. In the specific domain of economic studies, a handful of studies tried to quantify the severity according to the revenue losses by DMs and considered variables such as the type of evasion (e.g., no ticket *vs* unvalidated ticket), fare levels, and passenger flows (e.g., Cosby, 1985; Israel and Strathman, 2002; Prokosch and Gartsman, 2017). More precisely, Cosby (1985) proposed a model targeting both passenger and employee misconduct. He uses crew duty blocks as sampling units and simulation of fraud-free conditions by comparing actual revenues with those recorded during inspector-accompanied trips. It was applied to Greater Manchester's bus services and estimated relevant losses for different evasion types. Israel and Strathman (2002) proposed a model to provide a disaggregated analysis of fare evasion by type and time period within a free fare zone boundary, integrating both enforcement data and financial loss estimates. The model used scheduled trips as sampling units using both surveyors and inspectors to collect data. Specifically, fare inspectors accompanied surveyors onboard buses and trains to verify passengers' fares. Inspectors checked tickets and identified the type of violation, while surveyors recorded the data. Surveyors did not interact with passengers; their role was purely observational. Consequently, the type of violation derives from direct inspection data based on observed forms of non-compliance. They classified evasion into six types reflecting procedural irregularities (i.e., (i) no proof of payment, (ii) unvalidated ticket, (iii) zone violation, (iv) expired time, (v) use of a youth fare by passengers over 18, and (vi) use of an honoured citizen fare without the required permission), calculated fare evasion rates and estimated revenue losses using observed frequencies and Tri-Met (Portland) financial data. Revenue losses were estimated at around 20 % on light rail and 8 % for buses, suggesting increasing enforcement. Prokosch and Gartsman (2017) provided the financial and operational consequences of implementing all-door boarding on Boston's MBTA Green Line. It quantified fare revenue losses under a hybrid fare collection model using a detailed trip-along observational methodology. Inspectors and surveyors monitored 110 trips to measure fare evasion at unmonitored rear doors, combining quantitative data with contextual observations. The findings indicate an annual revenue loss that is about 2 % of the Green Line's total fare revenue: it is attributed primarily to crowding-induced fare evasion: more significant losses occurring during peak periods due to increased boarding pressure and confusion.

Other studies have emphasized that not all fare evasion events carry the same weight or implications for policy. Beyond the intentionality of the act, the occurrence with which individuals evade fares plays a crucial role in defining the severity of the behavior. In this regard, Delbosc and Currie (2016, 2019) identified a subgroup of recidivist fare evaders—individuals who repeatedly evade fares and are responsible for a disproportionate share of total revenue loss. This perspective suggests that fare evasion severity should be conceptualized not only in terms of behavioral intent (accidental, unintentional, or deliberate), but also as a function of recurrence. Recidivist fare evasion thus represents a chronic and more severe form of non-compliance, comparable to persistent offending patterns observed in other forms of deviant behavior.

Only recently, Barabino et al. (2023) used EMs encompassing logistic regression to assess fare evasion severity as a proxy for revenue loss, based on passenger survey data instead of trip data. By distinguishing between cases with severe evasion and cases with no severe evasion, results showed severity is influenced by socio-demographics, behaviour, and context: males, youth under 26, employed individuals, and captive riders were associated with higher severity. In contrast, higher education, commuting for work or study, and satisfaction with service reduced severity. Prior violations were the strongest predictor of severe reoffending. Barabino et al. (2024b) extended this analysis using Artificial Neural Networks (ANNs), confirming honesty, prior fines, education, and trip purpose as key drivers. Honesty significantly lowered severity, while education effects were mixed. Inspection perception had minimal impact, suggesting limited deterrence value.

[2] A "one size-fits-all model" refers to a solution, approach, or framework that is designed to be universally applicable across different situations, users, or contexts—without any customization or adaptation.

1.2. Literature gaps and objective

All these studies have contributed to understanding fare evasion by estimating its frequency and the severity through various modelling techniques, but several research gaps remain.

First, while many studies on the probability of fare evasion are available, the severity dimension has remained largely unexplored. This limits the understanding of how passenger and contextual characteristics contribute not just to the likelihood of evasion, but to the financial significance of such behaviour.

Second, as far as the authors' knowledge, while Barabino and Ventura (2025) compared the predictive performance of EMs and MLMs in forecasting the frequency of fare evasion, no existing research has systematically compared these modelling approaches in the prediction of the severity of fare evasion measured by the (potential) economic impact rather than mere occurrence. Such a comparison is critical, as these models offer complementary analytical strengths: EMs (such as LRMs) are well-suited for interpretability, hypothesis testing, and causal inference, making them particularly valuable in policy evaluation. Conversely, MLMs (such as ANNMs) excel in uncovering intricate, often non-linear relationships, and often outperform traditional EMs in predictive accuracy—especially in high-dimensional settings.

Third, while prior literature acknowledges the predictive superiority of ANNMs and their robustness to multicollinearity, they are often criticised for their "black box" nature, which limits transparency in identifying the relative importance of predictors (e.g., Chang, 2005; Zhang, 2010; Hegde and Rokseth, 2020). Although permutation feature importance can offer insights, no study has systematically quantified the effect of each factor on fare evasion severity within ANNMs and compared these results with those obtained from LRMs, as opposed to the related frequency (Barabino et al., 2024b; Barabino and Ventura, 2025).

Fourth, when developing a classification model, it is essential to select an appropriate probability threshold (cut-off) for assigning class labels, report metrics that are robust to class imbalance, and evaluate probability calibration (e.g., Lipton et al., 2014; Van Calster et al., 2019; Raccagni et al., 2025). These considerations are closely linked to predictive performance, particularly when the outcome variable is highly imbalanced—as in fare evasion severity—where severe cases occur far less frequently than non-severe ones. To the best of our knowledge, these methodological aspects have not been explicitly addressed in studies focused on predicting fare evasion severity.

To fill these gaps, this study develops, calibrates, and validates both a LRM and an ANNM using real-world fare inspection data from a mid-sized PTC in Italy. This dual-model comparison helps support a more comprehensive understanding of fare evasion behaviour and offers practical implications for TAs/PTCs. By leveraging the strengths of both approaches—causal interpretability from EMs and pattern recognition from MLMs—this study contributes to advancing data-driven and context-sensitive fare enforcement strategies.

1.3. Paper structure

The structure of the paper is as follows: Section 2 describes the fare evasion dataset, and the modelling adopted to analyse it. This includes the identification of predictors, the development of LRMs and ANNMs, as well as the evaluation of their predictive performance and the influence of each factor on fare evasion severity. Section 3 presents the results and provides a detailed discussion of the findings. Finally, Section 4 offers concluding remarks and proposes new avenues for future research.

2. Data and methods

2.1. Research context, data and modelling fare evasion severity event

This study was conducted in the Cagliari metropolitan area (Italy), with about 400k residents. Public transit is mainly operated by CTM, with 270+ vehicles, covering 12+M*veh*km annually across 32 routes (Tilocca et al., 2017; CTM, 2023). In Cagliari, various ticket types are available (e.g., annual, seasonal, monthly, single-ride, etc.) across all routes, and fare evasion can occur with any of them—although it is generally considered more prevalent among single-ride tickets. Fare evasion severity data[3] were collected through a survey conducted with passengers onboard[4] by questionnaires, during a four-month campaign between 2013 and 2015,[5] involving 12,000+ raw records.

Specifically, an onboard intercept survey was preferred over mail or web surveys for its higher response rates and direct interaction with actual users (Transit Cooperative Research Program, 2005). Given the questionnaire's complexity, a paper-and-pencil interview was adopted to improve accuracy and clarity, particularly among elderly or less literate passengers (e.g., Barabino et al., 2011, 2013). The questionnaire was administered by trained survey teams under uniform procedures and supervision to ensure consistency across all survey locations. Surveyors were instructed to apply the same questionnaire format and response options in every case, minimizing interviewer bias and ensuring data comparability. The onboard survey took place through face-to-face interviews with passengers during regular service hours over a six-week period (07:00 a.m.–07:00 p.m.). Weekdays were selected, as fare evasion occurs more frequently on these days (e.g., CTM, 2015). Interviewers approached passengers across various routes, recording responses directly, with an average of 69.35 % as for response rate and 8.98 % as for sampling rate with respect to the total passenger population of a day.

A five-part questionnaire from (a) to (e) was developed by drawing on the literature available in 2013 and was used as a survey tool. The questionnaire was structured to collect data on: (i) final fare evasion factors (events of fare evasion) and (ii) intermediate factors (potential predictors of fare evasion).

According to (i), respondents were asked to self-report their fare evasion behaviour with the question (a): "*How many times have you travelled without a valid ticket in the past four months?*" A four-month timeframe was considered appropriate for this purpose (Barabino and Salis, 2023). Responses were categorized into five levels, ranging from "never" to "always", where "never" was associated to an honest behaviour of the passenger. Someone could argue on the accuracy of self-reported data in measuring fare evasion severity; however, several studies suggest that individuals often experience psychological discomfort that discourages dishonest reporting—due to factors such as intrinsic costs of lying, personal honesty, or conditional cooperation (e.g., Abeler et al., 2014; Traxler, 2010).

Fare evasion severity may be interpreted along several dimensions—economic, operational, legal, behavioral, and social. In this study, the authors mainly refer to its economic aspects, such as lost revenues, which is among the most direct and measurable consequences for PTCs. Specifically, it was evaluated as a proxy of revenue loss related to the fare evasion behaviour previously stated. Naturally, the greater the severity of the evasion, the higher the associated revenue loss. More precisely, the severity of fare evasion could be represented using a

[3] According to CTM, an event of fare evasion occurs when a passenger intentionally fails to purchase, validate, or obtain a ticket and/or pass for travel.

[4] The Independent Ethical Board of the University of Cagliari was consulted on the entirety of this research project before it was submitted, which attested that this experiment did not require official ethical approval (Ref. PG/2018/13345).

[5] We cannot use more recent data owing to the secrecy policy of CTM.

response-ordered discrete variable organised e.g., as follows: complete absence of a ticket, tickets that are unvalidated or incorrectly validated[6] (even if purchased), and minor infractions (e.g., just expired ticket, forgotten passes[7]). In contrast, passengers who properly validate their tickets are not considered to have committed fare evasion. However, since many TAs/PTCs may lack the detailed and structured data necessary for such granular analysis, especially in the case of data collected by a survey, this study adopts a simplified approach. Specifically, fare evasion severity is modelled as a binary variable: individuals who placed themselves in the top three previous levels of fare evasion behaviour were classified as committing high (serious) fare evasion (i.e., Always, Often, Fair), while the others were considered as committing low (not serious) fare evasion (i.e., Rarely or Never). This binary classification would be particularly appropriate when relying on self-reported survey data, given the potential uncertainties associated with such responses. In fact, the use of this indirect approach to estimate the severity of fare evasion was necessary because the data were collected through a survey, without the involvement of inspectors to directly observe or record the various forms of fare evasion. Moreover, this choice was also motivated by the fact that most fare evasion is committed by recidivist offenders, who represent a market segment that deserves particular attention. For instance, Currie and Delbosc (2021) showed that habitual fare evaders are responsible for most revenue losses in Melbourne—an 8 % share of chronic evaders accounted for 68 % of total revenue loss. Furthermore, unlike Israel and Strathman (2002), grouping two broader severity levels focuses on the magnitude of potential revenue loss rather than the operational nature of the infraction.

To be more formal, let:

- I be the set of fare evasion events.
- S_i be the severity of a given event $i \in I$.

Severity is modelled as follows:

$$S_i = \begin{cases} 1; \text{ if } i \in I \text{ is a}'' \text{high severity event}'' \text{ (i.e., the answer provided from the passenger is } \textit{Always}, \textit{Often} \text{ or } \textit{Fair}) \\ 0; \text{ if } i \in I \text{ is a low severityevent (i.e., the answer provided from the passenger is } \textit{Rarely} \text{ or } \textit{Never}) \end{cases} ; \forall i \in I \quad (1)$$

According to (ii), respondents were asked regarding: (b) Socio-demographic details, including gender, age, education level, employment status, car ownership, and reasons for choosing bus transport; (c) Travel behaviour patterns, such as trip purpose, time of the trip, in-bus time, use of other public transport modes, travel frequency, and overall satisfaction with the service; (d) Contextual or situational aspects, including general awareness of fare evasion, perceived frequency of ticket inspections within a specific timeframe, awareness of the fine amount, history of fare violations (i.e., whether the passenger had previously been fined), and self-reported fare evasion frequency. Finally, (e) the honesty was assessed by the following question: "*If there were no inspections, would you still buy a ticket?*" This question aims to capture passengers' attitudes towards fare evasion, regardless of the final choice to evade the fare (e.g., Barabino et al., 2015; Barabino and Salis, 2020).

Notably, while including a broader set of attributes—such as perceived ticket cost, risk of detection, and income—could have yielded more comprehensive insights, the survey design prioritised practical feasibility. It was intended for administration to all passengers on board, necessitating a concise format. Consequently, the selection was limited the previous attributes to ensure the questionnaire remained manageable. The target completion time was 5–10 min, aligned with the CTM's estimate of a 15-min average minimum travel time within Cagliari.

Once data were collected, a dataset was built. Table 1 provides a summary of all predictors incorporated into the modelling process, including a brief description and descriptive statistics for each of them. The predictors are either categorized or transformed into binary format for modelling purposes. For clarity, dummy variables are presented in italics to facilitate comparison. Table 1 is self-explicative.

2.2. Econometric method (EM) – logistic regression model (LRM)

Once the experimental dataset was prepared, the severity model was first generated using a traditional EM. Specifically, an EM applies a LRM (Greene, 1993; Ortúzar and Willumsen, 2024). Since the severity is modelled as a binary variable, a binomial logistic regression model is applied. The results are expressed in terms of Odds Ratios (ORs), which represent the likelihood of a severe evasion occurring relative to a non-severe evasion. Each OR is obtained by exponentiating the corresponding estimated coefficient.

More formally, let:

- $LRM(\cdot)$ be the function representing the LRM.
- $TR \subset I$ and $TE \subset I$ be the training and testing subsets, respectively.
- $\boldsymbol{X}$ be the vector of severity predictors, and x be a generic predictor.
- $\boldsymbol{X_i}$ and x_i be the values taken by $\boldsymbol{X}$ and x for a given event $i \in I$, respectively.
- α and $\boldsymbol{\beta}$ be the intercept and the vector of the coefficients of the logistic regression, respectively.
- dr be the deviance ratio of the model.

Hence, the probability of a high-severity fare evasion event is modelled as follows:

$$P(S_i = 1) = LRM(\boldsymbol{X}_i) = \frac{e^{\alpha + \boldsymbol{\beta}\boldsymbol{X}_i}}{1 + e^{\alpha + \boldsymbol{\beta}\boldsymbol{X}_i}}; \ \forall i \in I; \quad (2)$$

The logarithm of the odds – defined as *logit* (S_i) – is the linear combination of predictors x_i; S_i is in the interval [0–1] for each value taken by the predictors:

$$logit(S_i) = log\left[\frac{P(S_i)}{1 - P(S_i)}\right] = log(odds) = \alpha + \boldsymbol{\beta}\boldsymbol{X}_i; \ \forall i \in I; \quad (3)$$

This model is assessed based on the deviance ratio (dr), which is computed as the ratio between the average deviance explained by the model to the degrees of freedom of the model, and its statistical relevance using the chi-squared test. Additionally, the direction and significance of the regression coefficients, along with the ORs, are examined.

To ensure an unbiased model validation, the experimental dataset is randomly partitioned into training (TR) and testing (TE) subsets before fitting. The TR is used to calibrate the model's coefficients, whereas the

[6] The "incorrectly validated ticket" might refer to cases where the passenger possesses a useable ticket but fails to validate it correctly or the ticket validator is out of order (e.g., validating it in the wrong direction or, in paper-based tickets, date and time are not printed).

[7] Many TAs/PTCs treat forgetting a pass as equivalent to travelling without a ticket; however, to our knowledge, if passengers show (*a posteriori*) that passes are really provided, the fine is sometime deleted.

Table 1
Dataset for evaluating the severity of fare evasion: List of variable and descriptive statistics.

Variables	Abbreviation	Description		% Value
(a) Severity of fare evasion	Severe	Severe	User engages in severe fare evasion	8.6 %[a]
	Not severe	*Not severe*	*User engages in minimal or no fare evasion*	91.4 %
(b) Socio demographic characteristics	Gender	Gen_M	Male	61.5 %
		Gen_F	*Female*	38.5 %
	Age[+]	Above 65	Above 65 years old	5.1 %
		51–65	Between 51 and 65 years old	10.7 %
		36–50	Between 35 and 50 years old	16.0 %
		26–35	Between 26 and 35 years old	16.1 %
		18–25	Between 18 and 25 years old	37.4 %
		Under_18	*Under 18 years old*	14.7 %
	Educational qualification	Upper_sc	Upper school graduate	16.0 %
		Middle_sc	Middle school graduate	45.5 %
		Middle_sc_n	Middle school not graduate	35.6 %
		Primary sc	Primary_school	2.7 %
		No_Sc	*No_school*	0.2 %
	Employment	Work_y	Worker	30.4 %
		Work_n	*Not worker*	69.6 %
		Stud_y	Student	49.3 %
		Stud_n	*Not student*	50.7 %
		Unempl_y	Unemployed	20.2 %
		Unempl_n	*Employed*	79.8 %
	Car availability	Car_y	Has a car	31.2 %
		Car_n	*Does not have a car*	68.8 %
	Reason to use the bus	Other_use_bus	Use of the bus for other reasons (Cheap; More ecological; It is difficult to find parking; Lower stress; Safer than other means)	38.7 %
		No_alter_use_bus	*Use of the bus because there are no alternatives*	61.3 %
(c) Travel behaviour characteristics	Trip purpose	Work_trips	Trips for work	25.7 %
		Study_trips	Trips for study	45.0 %
		Other_trips	*Other trips (shopping, sport, leisure, personal services, etc.)*	29.3 %
	Time of the day	Rush_hour_y	Rush hours trips (7.30–9.30 and 12.30–13.30 and 17.30–19.30)	57.6 %
		Rush_hour_n	*Not rush hour trips*	42.4 %
	In-vehicle time	In_vehicle_time_more15	Travel time more than 15 min	73.2 %
		In_vehicle_time_less15	*Travel time less than 15 min*	26.8 %
	Other transit systems use	Other_transit_y	Use of other transit systems	38.7 %
		Other_transit_n	*No use of other transit systems*	61.3 %
	Bus use frequency	Freq_traveler_y	The user travels more than 3 days a week	87.2 %
		Freq_traveler_n	*The user travels less than 3 days a week*	12.8 %
	Quality rating	Satisf_y	Satisfied user (grade on the overall service more than sufficient, *i.e., 6–10 in a 10 points scale*)	96.2 %
		Satisf_n	*Not satisfied user (grade on the overall service less than sufficient, i.e., 1–5 in a 10 points scale)*	3.8 %
Situational factors	Perceived inspection frequency	Insp_freq_more10	User has seen the inspectors boarding the vehicle to perform control activities more than 10 times in 4 months.	3.7 %
		Insp_freq_more6	User has seen the inspectors boarding the vehicle to perform control activities more than 6 times in 4 months.	13.1 %
		Insp_freq_1_5	User has seen the inspectors boarding the vehicle to perform control activities several times between 1 and 5 in 4 months.	55.0 %
		Insp_freq_null	*User has never seen the inspectors boarding the vehicle to perform control activities in*4 months.	25.2 %
	Amount of the fine	Know_fine_y	User knows the amount of the fine	68.0 %
		Know_fine_n	*User does not know the amount of the fine*	32.0 %
	Fines in the past	Fine_past_y	User has already been fined in the past	28.8 %
		Fine past_n	*User has not been fined in the past*	71.2 %
Personality var.	Value/Principles - Honesty	Honesty_y	User buys the ticket also if no checks are performed	77.7 %
		Honesty_n	*User does not buy the ticket if no checks are performed*	22.3 %

[+] Age groups were adapted for the purpose of this study to ensure adequate sample size and behavioural relevance within each category.

[a] Overall, the consistency between our 8.6 % survey-based estimate and the 7–7.3 % fine-based rates recorded by CTM over 2013–2015 suggests that the actual fare evasion level, including escaping from the vehicle and validating moments before boarding inspectors, likely ranges between 8 % and 9 % (CTM, 2013, 2014 and 2015).

The same dataset was adapted from Barabino et al. (2023).

TE serves exclusively to provide an objective evaluation of the model's predictive performance. Next, an automated stepwise procedure (combining forward selection and backward elimination) identifies the best model based on the highest dr and chi-squared significance ($p_{\chi 2}$). Predictor relevance is assessed by examining coefficient significance (p_{x_i}), direction, magnitude, and corresponding ORs.

2.3. Machine learning method (MLM) – Artificial Neural Network Model (ANNM)

There are many approaches to set up an architecture for MLM (Flach, 2012). In this paper, the architecture uses an ANNM for binary classification, based on a two-layer feed-forward architecture: an input layer (variables), a hidden layer (neurons with activation functions), and an output layer (binary outcome), chosen for its simplicity and effectiveness (e.g., Schmidhuber, 2015; Ventura et al., 2024). Specifically, this configuration features a unidirectional flow of information, progressing from input nodes to hidden nodes and finally to the output node. A hyperbolic tangent sigmoid function is used in the hidden layer, chosen for its antisymmetric properties, as validated by previous studies (e.g., Sramka et al., 2019). The output layer employs a logistic activation function, appropriate for predicting fare evasion severity as a proportion of outcome.

More formally, let:

- $ANN(\cdot)$ be the function representing the ANNM.
- $TR \subset I$, $VA \subset I$ and $TE \subset I$ be the training, validation and testing subsets, respectively.
- $\boldsymbol{W}$ and $\boldsymbol{b}$ be the matrix of the weights and the vector of the biases for the hidden layer, respectively.
- $f(\cdot)$ be the (hyperbolical tangent) activation function for the hidden layer.
- $\overline{\boldsymbol{W}}$ and $\overline{b}$ be the vector of the weights and the (scalar) bias for the output layer, respectively.
- $g(\cdot)$ be the (sigmoidal) activation function of the output layer.
- $loss(\cdot)$ be the (cross entropy) loss function of the model.

The probability of a high-severity fare evasion event is modelled as:

$$P(S_i = 1) = ANN(\boldsymbol{X_i}) = g(\overline{b} + \overline{\boldsymbol{W}} f(\boldsymbol{W}\boldsymbol{X_i} + \boldsymbol{b})); \ \forall i \in I; \tag{4}$$

Before fitting, the dataset (I) is split into training (TR), validation (VA), and testing (TE) subsets. The TR optimises weights and biases, the VA monitors generalisation and prevents overfitting, and TE provides an unbiased evaluation of the model's performance. Note that, while TR and TE performed the same roles as defined in the econometric model, VA had a distinct role—evaluating the model's ability to generalise across the network and, thus, mitigating the risk of overfitting. Training was halted once generalisation ceased to improve. During the training phase, the parameters $(\boldsymbol{W}, \boldsymbol{b}, \overline{\boldsymbol{W}}, \overline{b})$ were fine-tuned based on the TR data until a specific loss function $loss(\cdot)$ computed on VA was minimised. Specifically, the cross entropy was chosen, which is commonly employed for classification problems. Additionally, a backpropagation algorithm was utilised, as it is recognised as one of the most effective in training feedforward neural networks (e.g., Goodfellow et al., 2016). Thus, the training proceeded as in eqn. (5):

$$(\boldsymbol{W_0}, \boldsymbol{b_0}, \overline{\boldsymbol{W}}_0, \overline{b}_0) = \underset{(\boldsymbol{W}, \boldsymbol{b}, \overline{\boldsymbol{W}}, \overline{b})}{\arg\min} loss(\boldsymbol{W}, \boldsymbol{b}, \overline{\boldsymbol{W}}, \overline{b})_{VA} \tag{5}$$

Multiple models with varying neuron count in the hidden layer are trained to find the optimal configuration, selected by the lowest cross-entropy.

To assess predictor impact on the estimated severity, the Permutation Feature Importance (PFI) method is employed. PFI_x quantifies the importance of each predictor x by measuring performance degradation when x_i values are randomly permuted (e.g., Breiman, 2001; Strobl et al., 2007; Barabino et al., 2024b). PFI_x are computed on TE to ensure an unbiased evaluation. Moreover, for ease of interpretation, PFI_x are normalised so their sum equals one. Specifically, let:

- Ω be the set of random permutation instances.
- $loss(\boldsymbol{W_0}, \boldsymbol{b_0}, \overline{\boldsymbol{W}}_0, \overline{b}_0)_{TE}$ be the loss function of the trained ANNM, evaluated on TE considering the original predictors.
- $loss(\boldsymbol{W_0}, \boldsymbol{b_0}, \overline{\boldsymbol{W}}_0, \overline{b}_0)_{TE,x,\omega}$ be the loss function of the trained ANNM at the permutation $\omega \in \Omega$, evaluated on TE considering a random rearrangement of the x_i values.
- $\lambda = -loss(\boldsymbol{W_0}, \boldsymbol{b_0}, \overline{\boldsymbol{W}}_0, \overline{b}_0)_{TE}$ be the reference score of the trained ANNM on the original predictors.
- $\lambda_{x,\omega} = -loss(\boldsymbol{W_0}, \boldsymbol{b_0}, \overline{\boldsymbol{W}}_0, \overline{b}_0)_{TE,x,\omega}$ be the score of the trained ANNM considering a random rearrangement of the x_i values.

Hence, several permutation instances ($|\Omega|$) are issued for each x and the relative PFI_x is obtained according to eqn. (6).

$$PFI_x = \lambda - \frac{1}{|\lambda|} \sum_{\omega \in \Omega} \lambda_{x,\omega}; \tag{6}$$

This technique breaks the association between the predictor and the response variable, and the resulting decline in model performance reflects the degree to which the model depends on that specific predictor. Thus, the higher PFI_x, the higher the importance of x.

Although PFI_x is a valuable metric for assessing the performance of MLMs, it does not provide direct insights into the individual effect of each predictor to the severity of fare evasion. Therefore, for a meaningful comparison between EMs and MLMs, it is necessary to evaluate the influence of each variable. Among others, one viable method involves calculating the partial derivatives for each observation in the dataset and examining their relative frequency distributions. This approach is particularly useful for ANNMs, where nonlinear relationships are expected to occur. In such models, the marginal effect of a given predictor can vary across observations due to interactions with other inputs. Conversely, in linear models, the partial derivatives remain constant, reflecting the additive structure of the model. As a result, the distribution of partial derivatives may serve as an indicator of the degree of nonlinearity: higher dispersion suggests a stronger nonlinear behaviour, as the predictor's effect shifts depending on the context; lower dispersion, approaching zero in the linear case, indicates a more uniform influence across all records.

Specifically, let:

- $\frac{\partial ANN}{\partial x}(\boldsymbol{X_i})$ be the partial derivative of the function $ANN(\cdot)$, relative to the predictor x, for a given event $i \in I$.
- $\mu\left(\frac{\partial ANN}{\partial x}\right)$, $\sigma\left(\frac{\partial ANN}{\partial x}\right)$, and $CV\left(\frac{\partial ANN}{\partial x}\right)$ be the mean, the standard deviation and the coefficient of variation of the $\frac{\partial ANN}{\partial x}(\boldsymbol{X_i})$ values computed across the dataset, respectively.
- $P\left(\frac{\partial ANN}{\partial x}\right)^+$ be the probability that $\frac{\partial ANN}{\partial x}(\boldsymbol{X_i}) \geq 0$ within the dataset.
- $\Delta x \approx 0$ be a marginal increment.
- δ_x be a vector with the same dimension as $\boldsymbol{X}$, with all components equal to 0 except the one corresponding to predictor x, which is set to 1.

Thus, $\frac{\partial ANN}{\partial x}(\boldsymbol{X_i})$, $\mu\left(\frac{\partial ANN}{\partial x}\right)$, $\sigma\left(\frac{\partial ANN}{\partial x}\right)$, $CV\left(\frac{\partial ANN}{\partial x}\right)$ and $P\left(\frac{\partial ANN}{\partial x}\right)^+$ are numerically estimated as follows:

$$\frac{\partial g}{\partial x}(\boldsymbol{X_i}) \cong \frac{ANN(\boldsymbol{X_i} + \Delta x \cdot \boldsymbol{\delta}_x) - ANN(\boldsymbol{X_i})}{\Delta x} \tag{7}$$

$$\mu\left(\frac{\partial ANN}{\partial x}\right)=\frac{\sum_{i\in I}\frac{\partial ANN}{\partial x}(\boldsymbol{X}_i)}{|I|};\tag{8}$$

$$\sigma\left(\frac{\partial ANN}{\partial x}\right)=\sqrt{\frac{\sum_{i\in I}\left(\frac{\partial ANN}{\partial x}(\boldsymbol{X}_i)-\mu\left(\frac{\partial ANN}{\partial x}\right)\right)^2}{|I|}};\tag{9}$$

$$CV\left(\frac{\partial ANN}{\partial x}\right)=\frac{\sigma\left(\frac{\partial ANN}{\partial x}\right)}{\left|\mu\left(\frac{\partial ANN}{\partial x}\right)\right|};\tag{10}$$

$$P\left(\frac{\partial ANN}{\partial x}\right)^+=\frac{\left|\left\{\frac{\partial ANN}{\partial x}\geq 0:\forall i\in I\right\}\right|}{|I|};\tag{11}$$

A positive value of $\mu\left(\frac{\partial ANN}{\partial x}\right)$ indicates that the severity forecast averagely increase when x increases as well, and *vice versa*. Similarly, a $P\left(\frac{\partial ANN}{\partial x}\right)^+ > 0.5$ implies that positive effects are more numerous than negative effects, and *vice versa*. In other words, if the $P\left(\frac{\partial ANN}{\partial x}\right)^+$ are larger than 50 % of cases, the positive effects are more numerous than negative ones.

These metrics offer distinct and complementary perspectives. While $\frac{\partial ANN}{\partial x}$ mirrors the average magnitude of the predictor's influence, $P\left(\frac{\partial ANN}{\partial x}\right)^+$ measures the relative frequency with which the effect is positive, both of which are essential for a comprehensive assessment of the predictor's role. Additionally, a higher $CV\left(\frac{\partial ANN}{\partial x}\right)$ suggests increased nonlinearity in the ANNM with respect to x, indicating that the predictor's influence varies substantially across different observations. Therefore, although neural networks are often "black box" models, these metrics help enhance interpretability (Barabino and Ventura, 2025).

2.4. Performance comparison

After estimating LRM and ANNM, a comparative analysis is performed to assess which model offers a better fit and predictive accuracy for fare evasion severity in this specific study. The evaluation uses three complementary strategies.

I. Confusion Matrices (CMs).[8] CMs are computed for *TR*, *VA* and *TE* subsets. Each matrix shows the actual (target) severity class (columns) versus the predicted class (rows), in a 2 × 2 format. True Negatives ($\mathcal{TN}$) and True Positives ($\mathcal{TP}$) are on the main diagonal, while False Positives ($\mathcal{FP}$) and False Negatives ($\mathcal{FN}$) are on the antidiagonal. A lower $\mathcal{FP}$ and $\mathcal{FN}$ indicate better model's performance, with zero indicating perfect classification. Notably, the models output the probability of a severe fare evasion event. To convert these probabilities into discrete classes (severe vs. non-severe), it is necessary to define a cut-off threshold. However, the conventional threshold value of 0.5 may not be appropriate in this context due to the pronounced class imbalance (i.e., non-severe cases are far more frequent than severe ones). Using this default threshold often leads to an underestimation of severe cases, as the model tends to favour the majority class. To enhance the model's ability to accurately identify severe events, the threshold should be adjusted (e.g., Raccagni et al., 2025). This calibration process typically involves optimizing the threshold to achieve a suitable trade-off between $\mathcal{FP}$ and $\mathcal{FN}$. Further details on the threshold selection are provided in Appendix.

II. Performance Metrics. The True Positive Rate (*TPR*, also referred to as sensitivity or recall), True Negative Rate (*TNR*, or specificity), Positive Predictive Value (*PPV*, or precision), Accuracy (*ACC*), Balanced Accuracy (*BACC*), Area Under the Precision-Recall Curve (*AUPRC*), and F-score (*F*1), are calculated for *TR*, *VA* and *TE* subsets following Eqns. from 12 to 18. Higher values of these metrics reflect better performance.

$$TPR=\frac{\mathcal{TP}}{\mathcal{TP}+\mathcal{FN}};\tag{12}$$

$$TNR=\frac{\mathcal{TN}}{\mathcal{TN}+\mathcal{FP}};\tag{13}$$

$$PPV=\frac{\mathcal{TP}}{\mathcal{TP}+\mathcal{FP}};\tag{14}$$

$$ACC=\frac{\mathcal{TP}+\mathcal{TN}}{\mathcal{TP}+\mathcal{FN}+\mathcal{TN}+\mathcal{FP}};\tag{15}$$

$$BACC=\frac{TPR+TNR}{2};\tag{16}$$

$$AUPRC=\int_0^1 PPV(TPR)\cdot dTPR;\tag{17}$$

$$F1=\frac{2\cdot TPR\cdot PPV}{TPR+PPV;};\tag{18}$$

III. Probability Calibration. Reliability Curves (RCs) and the Brier Score (*BS*) are assessed for *TR*, *VA* and *TE*. RCs are diagnostic tools that graphically illustrate how the probability predictions produced by a model correspond to the observed outcomes. An RC is constructed by plotting the predicted probability from the model (e.g., 0.1, 0.2, …, 0.9) on the horizontal axis and the observed frequency of the positive class (1) within each probability interval on the vertical axis. If the model is well-calibrated, the points of the curve lie close to the diagonal line, indicating good agreement between the predicted and observed probabilities. Conversely, when the model is poorly calibrated, the points deviate from the diagonal line: points located above the diagonal indicate that the model tends to underestimate the true probabilities, whereas points below the diagonal suggest that the model tends to overestimate them. *BS* is a score that quantifies how accurately a binary classification model predicts probabilities. Lower *BS* values indicate better calibration and overall predictive accuracy. The score ranges from 0, representing perfect predictions, to 1, corresponding to completely incorrect predictions. For the training set (*TR*), the *BS* is computed as shown in Eq. (19). For the validation (*VA*) and test (*TE*) sets, the *BS* is computed in the same way, by replacing *TR* with *VA* or *TE*, respectively.

[8] A confusion matrix is a contingency table that summarises the performance of a classification model by comparing predicted and observed outcomes. It shows the number of correctly and incorrectly classified cases for each category, allowing the assessment of model accuracy and misclassification patterns. In this study, the confusion matrix is used to evaluate how well the model distinguishes between severe and non-severe fare evasion cases. A true positive event represents an observed case of severe fare evasion correctly predicted by the model, whereas a true negative corresponds to an observed non-severe or compliant behavior accurately classified as such. A false positive (false alarm) occurs when a non-severe case is incorrectly predicted as severe, while a false negative (missing) event refers to a severe case of fare evasion wrongly predicted as non-severe.

$$BS = \frac{1}{|TR|} \cdot \sum_{i \in TR} (P(S_i = 1) - S_i)^2; \tag{19}$$

IV. Importance and Effect of the Factors (IEF). For each $x_i \in X_i$, importance was assessed by the individual p-value p_{x_i} for the LRM, and by the permutation feature importance PFI_{xi} for the ANNM. Thus, predictors were then ranked in ascending order of p_{x_i} for the LRM, and in descending order of PFI_{xi} for the ANNMs, resulting in a prioritised list of the most important ones for each model. The LRM list reflects decreasing statistical significance (higher p_{x_i} values), while the ANNM list reflects decreasing feature importance (lower ρ_x values). Moreover, the direction of the effect was evaluated by the sign of coefficient estimate for LRM, and by $P\left(\frac{\partial ANN}{\partial x}\right)^+$ for the ANNM. If both models run in the same direction, the effects were assumed equal.

3. Results and discussion

Fare evasion severity (S_i) was measured as in Eqn. (1). The LRM was specified according to the EA (Eqn. (2)). The dataset was randomly split into TR and TE subsets using a 70 %/30 % split (e.g., Ferlitsch, 2021; Ventura et al., 2024). Next, the severity model was estimated using the MLA based on an ANNM (Eqn. (4)). As in the previous case, the dataset was split into TR, VA and TE using a 70 %-15 %-15 % ratio (e.g., Flach, 2012; Ventura et al., 2024).

Table 2 shows the results. In this table, columns 3 to 5 report the results (at the best fit) of the LRM, whereas columns 5 to 9 report those of the ANNM.

On the one hand, the LRM includes estimated coefficients (Estimates) associated with the significance levels (i.e., the corresponding p-value, p_{x_i}) and the ORs. For ease of reading, ORs<1 indicate reduced odds of severe fare evasion per unit increase in the predictor; ORs>1 indicate higher odds. The LRM shows a strong fit $\left(p_{\chi 2} < 0.001\right)$, and key predictors mostly align with expected patterns. On the other hand, after multiple iterations, the best-performing ANN architecture featured 40 hidden neurons and was trained using the Levenberg-Marquardt algorithm.

Table 2 includes the normalised PFIs, the mean partial derivative (i.e., $\mu(\partial ANN/\partial x_i)$), and the Std. Dev of the partial derivative (i.e., $\sigma(\partial ANN/\partial x)$), the coefficient of variation of the partial derivative (i.e., $CV(\partial ANN/\partial x_i)$) and the probability that the partial derivative is positive

Table 2
Results of the best-fit LRM and ANNM. The contrast variable is indicated in italic font in brackets for dummy variables.

		LRM ($d.r = 66.77$, $p_{\chi 2} < .001$)			ANNM (40 neurons, $loss = 0.17$)				
Levels	**Predictors (X_i)**	**Estimates (α, β)**	**p-val (p_{x_i})**	**ORs**	**Normalised PFIs (PFI_x)**	**Mean Part. Der. ($\mu(\partial ANN/\partial x)$)**	**Std. Dev Part. Dev. ($\sigma(\partial ANN/\partial x)$)**	**Coeff. Of Var. Part. Der. ($CV(\partial ANN/\partial x)$)**	**Prob. Part. Dev. > 0 ($P(\partial ANN/\partial x)^+$)**
	Constant	0.939	0.036	2.557					
Sociodemographic characteristics	Gender (*Women*)								
	Male	0.384	<.001	1.468	0.025	0.027	0.059	2.138	92.61 %
	Age (*Under 18*)								
	Above 65	-2.026	<.001	0.132	0.004	-0.032	0.070	2.191	18.36 %
	51-65	-2.329	<.001	0.097	0.025	-0.054	0.096	1.778	1.88 %
	36-50	-1.09	<.001	0.336	0.004	-0.032	0.056	1.760	5.24 %
	26-35	-0.639	0.007	0.528	<0.001	-0.029	0.062	2.140	23.23 %
	18-25	0.242	0.125	1.274	0.029	0.024	0.058	2.429	77.46 %
	Educational qualification (*None*)								
	Upper school graduate	-2.382	<.001	0.092	0.084	-0.065	0.111	1.704	0.66 %
	Middle school graduate	-1.79	<.001	0.167	0.088	-0.034	0.072	2.092	20.73 %
	Middle school not graduate	-1.132	<.001	0.322	0.021	-0.008	0.055	7.124	45.59 %
	Employment (*Not worker*)								
	Worker	0.456	0.076	1.578	<0.001	0.037	0.062	1.689	94.62 %
	Reason to use the bus (*No alternatives*)								
	Other	-0.18	0.207	0.835	0.013	-0.021	0.044	2.055	12.85 %
Travel behaviour characteristics	Trip purpose (*Other*)								
	Work trips	-0.7	0.011	0.497	0.008	-0.009	0.050	5.777	29.64 %
	Study trips	-0.971	<.001	0.379	0.084	-0.061	0.109	1.790	26.83 %
	In-vehicle time (*Less than 15 min*)								
	More than 15 min	-0.455	<.001	0.634	0.021	-0.033	0.057	1.734	6.81 %
	Quality rating (*Not satisfied*)								
	Satisfied	-0.504	0.02	0.604	0.008	-0.016	0.054	3.504	35.96 %
Situational factors	Perceived inspection frequency (*Null*)								
	Between 6 and 10 times in 4 months	-0.303	0.089	0.739	<0.001	-0.021	0.048	2.237	12.24 %
	Between 1 and 5 times in 4 months	-0.131	0.296	0.877	<0.001	-0.011	0.053	4.633	51.27 %
	Amount of the fine knowledge (*No*)								
	Yes	0.418	0.015	1.519	0.029	0.032	0.068	2.146	91.28 %
	Fines in the past (*No*)								
	Yes	1.212	<.001	3.360	0.134	0.079	0.110	1.384	100.00 %
Personality variables	Value/Principles – Honesty (*No*)								
	Yes	-2.262	<.001	0.104	0.420	-0.143	0.150	1.043	0.00 %

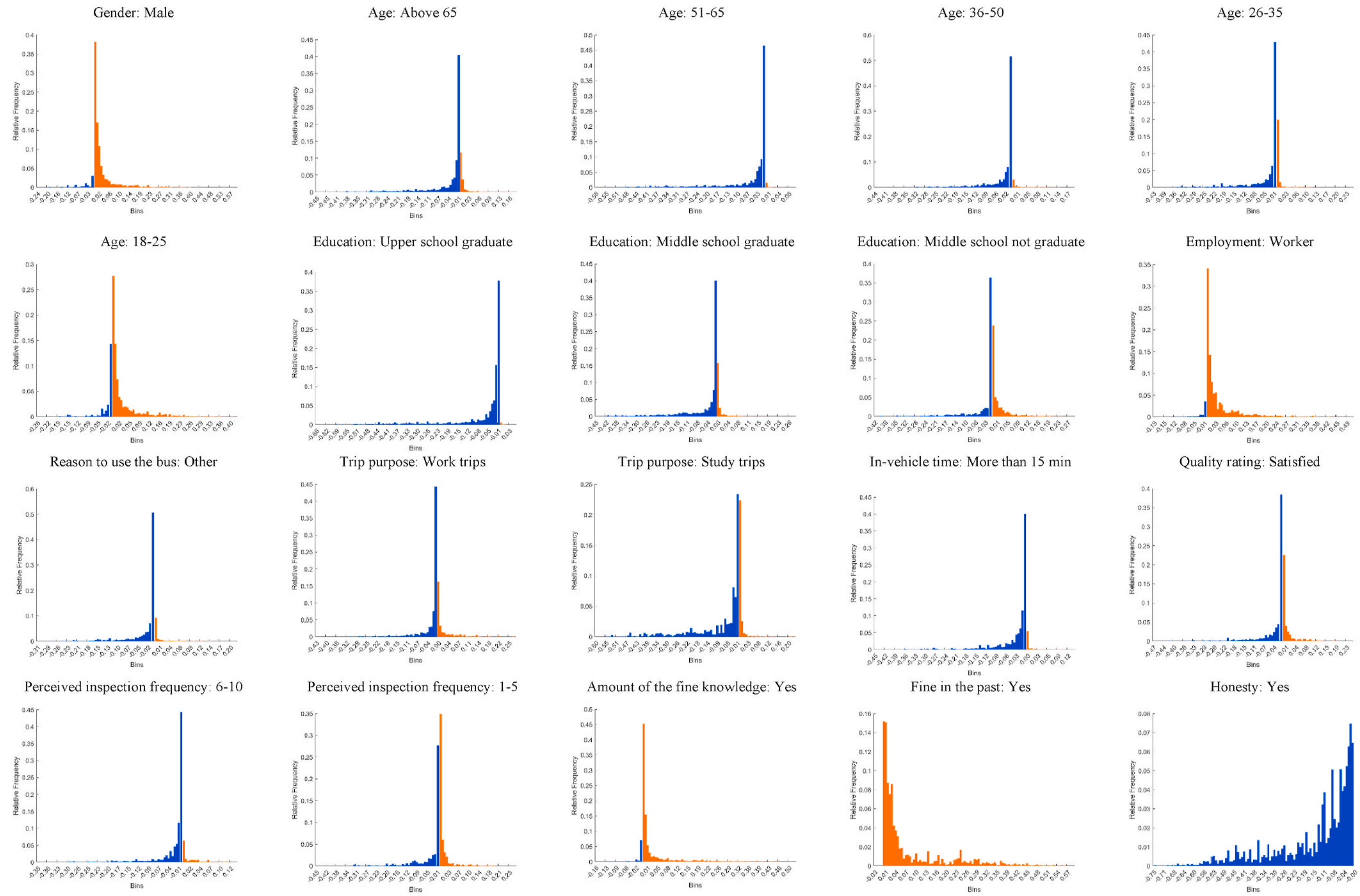

Fig. 1. Relative frequency distribution of the partial derivatives for each factor in the ANNM. Orange indicates a positive sign (i.e., the variable increases fare evasion severity); blue indicates a negative contribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(i.e., $P(\partial ANN/\partial x)^{+}$).

Furthermore, Fig. 1 (which is self-explanatory) depicts the relative frequency distribution of the partial derivative for each factor in the ANNM. In this figure, the x-axis labelled *Bins* represents the discrete intervals used to group the partial derivative values for the histogram visualization. Each bin corresponds to a specific range of the of the partial derivative values, and the bar height indicates the relative frequency of observations falling within that range.

Gender presents a modest but consistent effect on the severity of fare evasion. The LRM finds that males are 1.48 times more likely to commit severe fare evasion than females (OR = 1.468). This result is in line with past research, even if it was carried out considering the likelihood of fare evasion (Currie and Delbosc, 2017) and the intention to evade the fare (e.g., Barabino et al., 2015; Barabino et al., 2022), respectively. The ANN confirms this positive association ($PFI_x = 0.025$, $\mu(\partial ANN/\partial x_i) = 0.027$), with a high probability of a positive gradient ($P(\partial ANN/\partial x)^{+} = 92.61$ %), indicating a robust and consistent relationship despite the relatively low importance score.

Severity of fare evasion is most pronounced among younger passengers, particularly those aged 18–25. Hence, it decreases with age. According to the LRM, passengers over 65 are 7.58 times less likely to commit severe fare evasion than those under 18 (1/OR = 1/0.132), a finding that aligns with previous studies on the probability of fare evasion (Bucciol et al., 2013; Barabino et al., 2015, 2022; Delbosc and Currie, 2016; Cools et al., 2018; Barabino and Salis, 2020; Cantillo et al., 2022), although some studies diverge (e.g., Eddy, 2010). The ANN confirms this trend: older age groups show consistently negative partial derivatives, and the 51–65 group exhibits a notably negative gradient ($\mu(\partial ANN/\partial x_i) = -0.054$), with a substantially low $P(\partial ANN/\partial x)^{+}$ (1.88 %), reinforcing the inverse relationship between age and severity. The moderate $CV(\partial ANN/\partial x_i)$ (1.778–2.191) across age brackets reflects modest dispersion around the average effects, yet directional consistency is preserved. Interestingly, the 18–25 group exhibits a slightly increased severity compared to those under 18, suggesting a transitional behavioural pattern.

Education also shows a strong mitigating effect. LRM results indicate that upper school graduates are 10.87 times less likely to offend severely than the least educated (1/OR = 1/0.092), in agreement with findings from Barabino et al. (2015), though contrasting results have been observed (e.g., Bucciol et al., 2013; Delbosc and Currie, 2016). ANN results support this conclusion: education variables such as "Middle school graduate" and "Upper school graduate" rank among the most influential predictors ($PFI_x = 0.088$ and 0.084) and are associated with strong negative gradients ($\mu(\partial ANN/\partial x_i) = -0.034$ and -0.034, respectively). The latter also shows a near zero $P(\partial ANN/\partial x)^{+}$ (0.66 %), confirming directional certainty. However, the $CV(\partial ANN/\partial x_i)$ of the derivative for "Middle school not graduate" is high (7.124), reflecting notable variability and suggesting heterogeneity in this group's behaviour.

Employment status yields a more nuanced picture. LRM results suggest that workers are more prone to severe fare evasion than non-workers (OR = 1.578), a finding that is partially consistent with Delbosc and Currie (2016) but diverges from others (e.g., Bucciol et al., 2013; Barabino et al., 2015, 2022). The ANN supports a small positive effect ($\mu(\partial ANN/\partial x_i) = 0.037$), with a high $P(\partial ANN/\partial x)^{+}$ (94.62 %), indicating a consistent directional impact despite its negligible importance ($PFI_x < 0.001$).

Trip purpose strongly affects severity. The LRM shows that study-related trips reduce the odds of severe evasion by 2.63 times (1/OR = 1/0.379), and work-related trips by 2.01 times (1/OR = 1/0.497). This suggests fare evasion is less severe when passengers travel for systematic trips, since they are likely to hold season passes (e.g., Barabino et al., 2023). The ANN confirms the importance of trip purpose, especially for study trips ($PFI_x = 0.084$, $\mu(\partial ANN/\partial x_i) = -0.061$), which register the second most negative gradient in the model. However, the $CV(\partial ANN/\partial x_i)$ for work trips is relatively high (5.777), pointing to more heterogeneous effects for this group.

Trip duration also contributes to lower severity. Passengers travelling more than 15 min are 1.58 times less likely to offend severely (OR = 1/0.634) according to the LRM, possibly due to greater exposure to inspections as shown in past research evaluation the intention to evade the fare (Barabino and Salis, 2020). The ANN mirrors this finding with a negative derivative ($\mu(\partial ANN/\partial x_i) = -0.033$) and a low probability of positive effect ($(P(\partial ANN/\partial x)^+ = 6.81$ %), though its overall impact remains moderate ($PFI_x = 0.021$).

Service satisfaction appears to reduce severity as well. The LRM suggests that satisfied passengers are 1.66 times less likely to commit severe fare evasion (OR = 1/0.604), in line with Barabino et al. (2015), Delbosc and Currie (2016), and Barabino et al. (2022), although other studies show mixed results depending on passenger type (e.g., Bucciol et al., 2013; Currie and Delbosc, 2017). The ANN identifies a similar direction ($\mu(\partial ANN/\partial x_i) = -0.016$), but with lower overall influence ($PFI_x = 0.008$). The relatively high $CV(\partial ANN/\partial x_i)$ (3.504) suggests that, although the average effect is negative, the individual-level variation may be substantial. Thus, more studies are suggested to better evaluate these results.

Among situational factors, the perceived frequency of inspections has a deterrent effect. The LRM shows that passengers inspected 6–10 times in four months are 1.35 times less likely to offend (OR = 1/0.739), supporting previous research (Barabino et al., 2015; Cools et al., 2018; Porath and Galilea, 2020). The ANN concurs with a negative gradient ($\mu(\partial ANN/\partial x_i) = -0.021$) and a low probability of positive derivatives ($(P(\partial ANN/\partial x)^+ = 12.24$ %). This reinforces the preventive nature of frequent inspections, though with low PFI_x (<0.001).

Conversely, knowledge of fine amounts is associated with higher severity. The LRM finds an OR of 1.519 for passengers aware of the fine, in line with Barabino et al. (2015), Bucciol et al. (2013), Barabino and Salis (2020, 2023) and Barabino et al. (2022). This suggests that awareness may foster normalisation or risk tolerance, as well as perceived inefficacy in enforcement, with some passengers believing penalties are too low or won't be collected after a certain period. The ANN reflects this with a moderate importance ($PFI_x = 0.029$), a positive derivative ($\mu(\partial ANN/\partial x_i) = 0.032$), and high probability of a positive effect ($(P(\partial ANN/\partial x)^+ = 91.28$ %), indicating both consistency and certainty in this association.

One of the strongest predictors in both models is a history of past fines. LRM shows that previously fined individuals are 3.36 times more likely to reoffend severely (OR = 3.360), consistent with Barabino et al. (2015) even if shown in the case of the probability of evade the fare. Moreover, it should be noted that Dai et al. (2017) investigated a very specific sample of recently fined passengers (less than forty individuals) through a lab-in-the-field experiment. Their finding that these passengers tended to misbehave more often in the experimental setting likely reflects a short-term rebound or compensatory effect rather than a long-term behavioural pattern, unlike this study. The ANN reinforces this finding, assigning high importance ($PFI_x = 0.134$) and the highest positive gradient ($\mu(\partial ANN/\partial x_i) = 0.079$). The probability of a positive derivative is 100 %, and the $CV(\partial ANN/\partial x_i)$ (1.384) is relatively low, underscoring the stability and strength of this effect. These results might suggest a potential "recovery" behaviour among some type of fare evaders or confirm that these passengers seem unresponsive to more stringent enforcement measures (Barabino et al., 2024a). As for the last claim, the research highlights a segment of "chronic fare evaders" who habitually avoid paying and are unresponsive to stricter enforcement often do not actively hide their evasion, perhaps offering false identification or ignoring fines (e.g., Hauber, 1980; Currie and Delbosc, 2021; Delbosc and Currie, 2016; González et al., 2019). Separately, there are other habitual "artful evaders" who exploit their in-depth knowledge of transit system weaknesses to evade fares with minimal risk (Bijleveld, 2007).

Lastly, personality traits have substantial effects. According to the LRM, honest passengers are 9.62 times less likely to commit severe fare evasion (OR = 1/0.104), in line with *a posteriori* segmentation studies (Delbosc and Currie, 2016; Currie and Delbosc, 2017; Barabino and Salis, 2023). The ANN further highlights this variable's dominance, with the highest PFI_x (0.420), a strong negative gradient ($\mu(\partial ANN/\partial x_i) = -0.143$), and a 0 % probability of a positive derivative, confirming the certainty and strength of this deterrent effect. The $CV(\partial ANN/\partial x_i)$ of 1.043 suggests a stable behavioural pattern linked to ethical principles.

After estimating the LRM and ANNM, their fit and predictive accuracy were assessed to determine the most effective modelling approach, in this case study.

Focusing on the I comparison strategy, the resulting CMs are shown in Table 3.

Globally, the ANNM successfully detects 490 high-severity cases ($\mathscr{TP}$), compared to 437 in the LRM, thereby reducing the number of false negatives ($\mathscr{FN}$) from 319 to 266. This enhancement shows that the neural model is more effective in recognizing severe fare evasion events, in this case study. However, while both models achieve a high number of true negatives ($\mathscr{TN}$), the LRM produces slightly fewer false positives ($\mathscr{FP}$)—449 vs 491—meaning it misclassifies low-severity cases as high-severity somewhat less often than the ANNM. In practical terms, the ANNM detects more severe cases but also generates a slightly higher number of false alarms, reflecting the typical balance between detection ability and misclassification in datasets with uneven class distributions. Notably, the conversion of predicted probabilities into discrete classes (severe vs. non-severe) was carried out using the optimal cut-off values identified for each model: 0.30 for the LRM and 0.26 for the ANNM. Further details are provided in Appendix.

Moving on to the II comparison strategy, the results of the key performance metrics (computed according to Eqns. from 12 to 18) are provided in Table 4.

Both models achieve high predictive performance across the training, validation, and test sets. The ANNM consistently yields higher values for the *TPR*, improving from 0.587 to 0.636 in training and from 0.555 to 0.677 in testing, which indicates a stronger capability in detecting high-severity fare evasion cases. This enhanced detection ability makes the ANNM particularly valuable for deterrence and enforcement applications, where minimizing false negatives is crucial. Nevertheless, these results depend on the specific binary classification of fare evasion severity adopted in this study and should not be generalised without external validation.

While the ANNM detects more severe cases, it does so with a slight reduction in the *TNR*—approximately 0.938 compared to 0.943 for the LRM—showing that the improvement in detection comes at the cost of a small increase in misclassification of low-severity cases. Still, both models maintain high specificity, ensuring reliable identification of non-severe events and reducing undue penalties for minor infractions.

The *PPV* shows comparable results between the two models. However, the ANNM achieves a clear advantage in the test phase (0.557 vs. 0.484), indicating greater precision in identifying actual high-severity cases. The overall *ACC* remains high and nearly identical for both models across all subsets (≈0.91–0.92), confirming their strong generalization ability and consistent predictive reliability.

The *BACC*, which accounts for class imbalance, further highlights the ANNM's more even performance across both classes. Similarly, *AUPRC* increases from 0.522 in the LRM to 0.558 in the ANNM on the test dataset, confirming the superior ability of the neural model to distinguish between severe and non-severe cases under skewed class

Table 3
I comparison strategy: Confusion Matrices for LRM and ANNM.

		Training		Validation		Test		All	
LRM	Predicted \ Target	0	1	0	1	0	1	0	1
	0	5′220	222	n.a.	n.a.	2′257	97	7′477	319
	1	320	316	n.a.	n.a.	129	121	449	437
ANNM	**Predicted \ Target**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
	0	5′223	185	1′110	39	1′102	42	7′435	266
	1	347	323	74	79	70	88	491	490

n.a. = not applicable.

Table 4
II comparison strategy: Performance metrics for LRM and ANNM.

		LRM	ANNM			LRM	ANNM
Training	***TPR***	0.587	0.636	**Test**	***TPR***	0.555	0.677
	TNR	0.942	0.938		***TNR***	0.946	0.940
	PPV	0.497	0.482		***PPV***	0.484	0.557
	ACC	0.911	0.912		***ACC***	0.913	0.914
	BACC	0.765	0.787		***BACC***	0.751	0.809
	AUPRC	0.525	0.529		***AUPRC***	0.522	0.558
	F1	0.539	0.550		***F1***	0.517	0.611
Validation	***TPR***	n.a.	0.669	**All**	***TPR***	0.578	0.648
	TNR	n.a.	0.938		***TNR***	0.943	0.938
	PPV	n.a.	0.516		***PPV***	0.493	0.500
	ACC	n.a.	0.913		***ACC***	0.912	0.913
	BACC	n.a.	0.804		***BACC***	0.761	0.793
	AUPRC	n.a.	0.564		***AUPRC***	0.525	0.541
	F1	n.a.	0.583		***F1***	0.532	0.564

distributions.

Finally, the F1-score—representing the harmonic balance between detection capability and precision—improves from 0.517 in the LRM to 0.611 in the ANNM, demonstrating a substantial gain in predictive consistency and generalization in classifying high-severity fare evasion events in the test phase.

These findings are consistent with previous research, which has shown the superior flexibility and predictive capability of artificial neural models over logistic regression when capturing nonlinear relationships (e.g., Mussone et al., 2017; Ventura et al., 2024; Wilson et al., 2024). Nevertheless, applying these methods to the domain of fare-evasion severity represents a novel contribution, and the generalizability of the results should be further validated using external datasets.

Focusing on the III comparison strategy, RCs and the associated *BS* values are presented in Fig. 2.

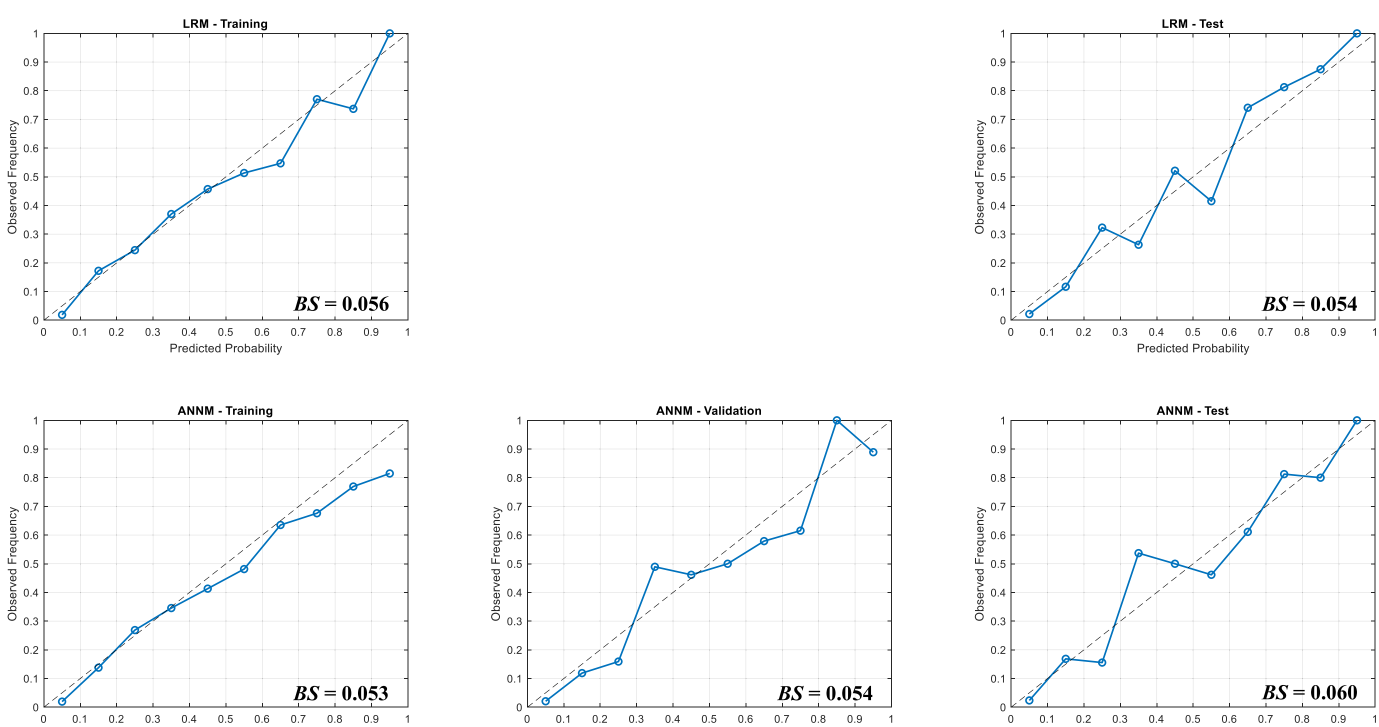

Fig. 2. III comparison strategy: Reliability curves for LRM and ANNM.

Table 5
Comparison between the ranking and effects returned by LRM and ANNM.

Levels	Predictors (X_i)	LRM - P-val Ranking	ANNM - PFI Ranking	Effect Direction
	Constant	–	–	
Intermediate	Gender (*Women*)			
Sociodemographic characteristics				
	Male	10	8	▲
	Age *(Under 18)*			
	Above 65	9	15	▲
	51–65	6	9	▲
	36–50	7	16	▲
	26–35	12	19	▲
	18–25	19	7	▲
	Educational qualification (*None*)			
	Upper school graduate	4	4	▲
	Middle school graduate	5	3	▲
	Middle school not graduate	11	10	▲
	Employment (*Not worker*)			
	Worker	17	17	▲
	Reason to use the bus (*No alternatives*)			
Intermediate	Other	20	12	▲
Intermediate	Trip purpose (*Other*)			
Travel behaviour characteristics				
	Work trips	13	13	▲
	Study trips	3	5	▲
	In-vehicle time (*Less than*15 min)			
	More than 15 min	8	11	▲
	Quality rating (*Not satisfied*)			
	Satisfied	15	14	▲
Intermediate	Perceived inspection frequency (*Null*)			
Situational factors				
	Between 6 and 10 times in 4 months	18	20	▲
	Between 1 and 5 times in 4 months	21	18	▼
	Amount of the fine knowledge (*No*)			
	Yes	14	6	▲
	Fines in the past (*No*)			
	Yes	2	2	▲
Personality variables	Value/Principles – Honesty (*No*)			
	Yes	1	1	▲

▲ = The models return the same effect; ▼ The models return an opposite effect.

LRM exhibits a generally smooth and well-aligned calibration across both training and test sets, with predicted probabilities closely matching the observed frequencies. Deviations are limited and mostly occur in the mid-probability range (around 0.4–0.6), suggesting that LRM provides stable and interpretable probability estimates. This impression is supported by the *BS* values, which remain low and consistent across subsets (≈0.056 in training and 0.054 in testing), confirming the robustness of LRM's probability calibration under unseen data.

ANNM also follows the diagonal trend reasonably well in training and validation sets, confirming good calibration during model development. However, the test curve exhibits greater variability compared to LRM, particularly at intermediate probabilities, where deviations from the ideal line are more pronounced. This reduced stability is reflected in the *BS*, which increases to 0.060 in the test phase, indicating less reliable probability estimates despite acceptable calibration overall. Interestingly, this observation contrasts with the performance metrics discussed in the II comparison strategy, where ANNM demonstrated stronger predictive capability. This highlights a common trade-off between calibration stability and classification performance: while ANNM excels in detecting severe cases, LRM provides more consistent and well-calibrated probability estimates (Alba et al., 2017; Van Calster et al., 2019). The choice between these models should therefore consider the operational context—whether priority is given to accurate probability calibration for resource allocation or to maximizing detection of high-severity fare evasion.

When the attention moves towards the IV comparison strategy, the results in terms of ranking and effect direction are shown in Table 5. This detailed overview allows for insights into both the alignment and divergences in variable importance (i.e., ranking) and impact.

When examining the ranking of predictors, both models demonstrate strong alignment on several critical variables. Notably, the Honesty consistently emerges as the most influential variable (ranked 1st) by both LRM (P-val Ranking) and ANNM (PFI Ranking). This highlights the consistent and paramount role of personal integrity in deterring fare evasion. Similarly, "Fines in the past" is ranked 2nd by both models, underscoring the significant impact of prior enforcement encounters. Predictors "Middle school graduate" and "Upper school graduate" also show relatively close rankings between the two models (LRM 5th, ANNM 3rd; and LRM 4th, ANNM 4th respectively), indicating shared importance.

Beyond these strong areas of consensus, rankings begin to diverge. For instance, "Age (18–25)" is ranked 19th by LRM but jumps significantly to 7th by ANNM. Conversely, "Age (Above 65)" is 9th in LRM but drops to 15th in ANNM. These discrepancies might stem from ANNM's inherent ability to implicitly account for complex, non-linear relationships and interactions between variables, down-weighting less distinctive factors and emphasising those with unique or interaction-driven contributions, which LRM treats as independent. Another notable divergence is seen in "Amount of the fine knowledge" ranked 14th by LRM but significantly higher at 6th by ANNM, suggesting ANNM captures a more pronounced importance for awareness of fine amounts.

As for the effect directions for the predictors, the two models exhibit a strong overall convergence, reinforcing the robustness of the findings across both generalised linear (LRM) and non-linear (ANNM) modelling paradigms. As indicated by "▲"in Table 5, almost all factors demonstrate the same influence. This consistency spans across sociodemographic characteristics (e.g., gender, age groups, educational qualifications, employment status), travel behaviour characteristics (e.g., trip purpose, in-vehicle time, quality rating), and situational factors like "Fines in the past" and "Amount of the fine knowledge," which consistently show the same effect direction. The consistent top-ranking of "Honesty" with a stable effect direction further solidifies its critical role.

However, a single, notable divergence between LRM and ANNM concerns the predictor "Perceived inspection frequency (Between 1 and 5 times in 4 months)," marked with a "▾" in Table 5. While the specific nature of the opposing effect is better detailed in Table 4, this contrast likely stems from the ANNM's superior capability to capture non-linearities and interactions among variables. In this case, the effect of perceived inspection frequency may not be uniformly positive or negative but instead context-dependent—shaped by factors not explicitly modelled in the linear framework of the LRM. The ANNM, by contrast, can dynamically adjust this effect based on surrounding conditions, revealing latent heterogeneity that the LRM may overlook. Notably, this divergence likely reflects a skewed distribution with a long negative tail, where most partial effects are mildly positive ($P(\partial ANN/\partial x)^{+} > 50\%$), but a smaller subset exerts stronger negative impacts (as indicated by $(\partial ANN/\partial x_i) < 0$)—a nuance only captured through the ANNM's flexibility.

From a policy perspective, these findings illustrate that while distinct modelling approaches may prioritise certain factors differently, they generally point towards complementary strategic directions for TAs/PTCs. The consistent high ranking and effect direction for "Honesty" and "Fines in the past" across both models underscore their profound influence on the severity of fare evasion. The divergences, particularly in the ranking of age groups or knowledge of fines, and the differing effect direction for specific perceived inspection frequencies, provide a more nuanced basis for decision-making. In practice, this suggests that policymakers could draw on insights from both models to design a comprehensive deterrence strategy. This would involve not only leveraging the impact of past enforcement (fines) but also considering the complex, non-linear ways in which factors like perceived inspection frequency and demographic profiles interact to influence fare evasion. Such a multi-faceted approach, informed by both linear and non-linear insights, could lead to more effective and targeted interventions.

Given that "Honesty" and "Fines in the past" are consistently highly ranked and influential factors in the prediction of fare evasion severity, a practical strategy should focus on a two-pronged approach: (1) fostering a culture of integrity and consistently (2) enforcing consequences for repeated offenders while implementing welfare programs for low-income riders.

As for (1), TAs/PTCs should leverage voluntary compliance, that has been shown to be more effective than punitive measures (e.g., Feldman, 2025). Specifically, to foster a stronger culture of honesty and civic duty—and thereby reduce the severity of fare evasion—they could launch targeted public awareness campaigns that vividly communicate the collective benefits of fare compliance. These campaigns should highlight how fare revenue directly funds essential service improvements, enhances accessibility, and strengthens the overall reliability of public transport for all users.

Messaging should be carefully crafted to appeal to passengers' sense of civic pride and personal responsibility, using clear and impactful slogans as: "*Your fare fuels better journeys for our city*." This approach is backed by evidence showing that perceptions of fare evasion's occurrence (e.g., social norms) are key drivers of compliance (Ayal et al., 2021; Celse and Grolleau, 2023). Notably, Ayal et al. (2021) observed a reduced likelihood of fare evasion among passengers exposed to messages communicating low evasion rates—particularly when these were presented alongside control posters featuring eye cues. In contrast, threatening messages focused solely on inspection operations or the severity of fines did not appear to significantly influence behaviour (Celse, under review).

Aligned with this, TAs/PTCs should adopt a strategy of "positive" enforcement, shifting from merely deterring fare evaders to highlighting most passengers who regularly pay their fares. This could include sharing inspiring stories or compelling statistics about honest behaviour via digital displays, social media, and visually engaging posters—thereby reinforcing the social norm of fare compliance (e.g., Mendoza and Wielhouwer, 2015; Ariel et al., 2019).

Moreover, community engagement programs could be developed in partnership with local schools and organisations to educate younger generations on the value of supporting public services and embracing the concept of the common good. Finally, while not directly tied to honesty, simplifying fare payment options plays a subtle yet essential role: making the process as easy and frictionless as possible removes common excuses and gently encourages compliant behaviour.

As for (2), the finding that "Fines in the past" strongly influence (with an increasing effect) fare evasion severity suggests that, for some passengers, the current penalty fare system fails to deter future offenses. These individuals appear largely unaffected by the risk of being fined—likely because they have learned to exploit weaknesses in the fare enforcement and collection process. As noted by Bijleveld (2007), they show limited behavioural response to fines. Simply increasing enforcement presence does not guarantee these offenders will face consequences. Moreover, raising fine amounts alone may not effectively reduce their evasion (e.g., Buehler et al., 2017).

Thus, a more effective strategy could involve combining higher fines for repeat offenses with a streamlined, fast-track judicial process—provided the legal framework allows such. Efficient fine collection is essential, ensuring rigorous follow-up for unpaid fines. The deterrent effect of past fines is weakened if violations go unpunished, underlining the need for inter-agency cooperation and legal reforms (e.g., Barabino et al., 2022).

Nevertheless, is worth noting that repeated fare evasion is not solely an issue of compliance—it also intersects with economic hardship and social exclusion. Low-income riders may evade fares simply because they cannot afford them. In such cases, punitive measures are inappropriate and may conflict with equity principles. Instead, supportive actions—such as providing free fare cards via welfare programs—could offer a fair and socially inclusive solution (Schwerdtfeger, 2016; Perrotta, 2017; Barabino et al., 2020).

By combining appeals to intrinsic values (honesty) with clear, consistent extrinsic motivators (enforcement, welfare), these strategies aim to create an environment where paying the fare is both the expected social norm and the unavoidable smart choice.

4. Conclusions

Fare evasion remains a tenacious challenge for Transit Authorities and Public Transport Companies (TAs/PTCs) worldwide, consistently requiring innovative solutions. This study offers a novel contribution to deepening our understanding of fare evasion severity, which can guide the formulation of diverse strategic responses. Specifically, to the best of the authors' knowledge, this research for the first time:

- Implements and compares Logistic Regression Models (LRMs) and Artificial Neural Network Models (ANNMs) in predicting fare evasion severity, rather than simple occurrence, thus offering a novel lens for economic impact analysis.
- Compares model performance through Confusion Matrices, Performance Metrics robust to class imbalance and probability calibration tools.
- Fine-tunes probability thresholds (cut-offs) to enhance predictive performance on imbalanced datasets.
- Investigates the relative importance of key variables, aiming to inform equitable and efficient fare enforcement strategies.

An empirical experiment was executed using authentic data derived from a mid-sized PTC in Cagliari, Italy. The findings revealed the honesty, history of fines, and education levels had a pronounced impact on fare evasion severity. Within this dataset, the ANNM slightly outperformed the LRM in terms of detecting high-severity fare-evasion cases, achieving higher recall and F1-scores, thereby reducing the number of severe instances missed. However, this improvement comes with a minor increase in misclassified low-severity cases, reflecting the typical trade-off between classification performance and calibration stability observed in the reliability analysis; overall, these results suggest that while neural models are more effective at identifying complex patterns associated with severe events, logistic regression provides more consistent and interpretable probability estimates.

Furthermore, while both models often produce similar effects, each may underscore distinct facets of the factors influencing evasion severity due to their respective abilities to discern nonlinear relationships. Crucially, these differences should not be viewed as limitations but rather as reflections of complementary modelling philosophies: the LRM emphasises interpretability and transparency, while the ANNM prioritises predictive accuracy.

Both models underwent internal cross-validation—meaning the testing and validation subsets were drawn from the same data distribution—though no external validation was conducted (i.e., testing on data from other cities, transit systems, or time periods). Consequently, while these findings might be transferable to similar environments—such as those with comparable ticketing systems, inspection strategies, passenger volumes, TA/PTC enforcement policies and definition and classification of fare evasion severity—they cannot be automatically generalised across all transit systems without further rigorous validation. Thus, a crucial next step is to apply the proposed modelling framework to datasets from diverse urban transit systems. Such endeavours would be instrumental in assessing the transferability of this approach and could help to either confirm or refine the conclusions presented herein.

These findings offer actionable insights for TAs/PTCs to enhance fare enforcement:

- Leverage key behavioural drivers. The strong influence of "Honesty" and "Fines in the past" on fare evasion severity predictions suggests that TAs/PTCs should actively promote civic responsibility and highlight the benefits of fare payment, reinforcing honesty as a social norm. At the same time, fines must not only be issued to repeat offenders but also effectively collected through rigorous follow-up—while ensuring that welfare programs support disadvantaged groups. This dual approach—nurturing integrity while guaranteeing tangible consequences for non-compliance—is essential to effectively curb fare evasion.
- Align model choice with strategic goals. Model selection should reflect operational priorities. LRMs are well suited for transparent policy evaluation, offering interpretable insights into factor effects (e.g., the impact of passenger age on evasion severity). Conversely, ANNMs provide higher predictive accuracy and are better suited for real-time detection and targeting (e.g., identifying zones with high evasion severity), making them ideal for adaptive planning and enforcement deployment, but keeping equity in inspection.

This study paves the way for novel research. Firstly, incorporating novel intermediate factors could deepen the understanding of this multifaceted phenomenon. Second, conducting sensitivity analyses or scenario-based simulations might assist in evaluating the effectiveness of specific policy interventions, for instance, by forecasting how fare evasion severity shifts under varying conditions using the trained models. Third, future research should investigate cost-sensitive thresholding and class-weighted learning approaches to better align model decisions with enforcement priorities and operational costs. Although such analyses would require explicit cost information, their implementation could substantially enhance the practical applicability and policy relevance of predictive models for fare evasion severity. Fourth, exploring alternative statistical (e.g., zero-inflated models) and machine learning (e.g., boosting techniques; Support Vector Machine) approaches may further improve predictive accuracy, particularly in imbalanced or data-constrained settings. In addition, hybrid methods combining the interpretability of LRM with the predictive power of ANNM warrant investigation, as they may offer an optimal balance between transparency and accuracy. Fifth, passengers using discounted fares without eligibility were not explicitly considered in this study, as these specific data were not collected. Future research could include this type of passengers to refine the measure of the severity. Sixth, analysing a wider range of severity levels directly related to the revenue loss (e.g., unvalidated ticket, expired tickets, forgotten passes) could enable a more granular assessment of the financial impacts associated with fare evasion. Effectively, TAs/PTCs can classify different level of severity and not adopt neither a proxy of loss revenues nor a binary classification provided in our study. Therefore, novel research in this issue is strongly suggested. Finally, the honesty construct was measured using a single self-reported item, which may be affected by social desirability and common method bias. Future studies should incorporate validated multi-item scales, such HEXACO personality trait scale, to better capture individual ethical orientation toward fare payment.

CRediT authorship contribution statement

Benedetto Barabino: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization, Project Management. **Roberto Ventura:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI) to improve English language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflict of interest

declare no conflict of interest, financial or otherwise.

Acknowledgement

The initial phase of this paper, including the completion of the first draft, occurred while the first author was employed at CTM SpA. We extend our gratitude to the General Management of CTM SpA for their support of this research and for allowing access to their real data for experimentation. This research received partial funding from the Department of Civil, Environment, Land and Architecture Engineering and Mathematics (DICATAM) at the University of Brescia through the research grant titled "Valuation of the Risk of Fare Evasion in an Urban Public Transport Network," CUP: D73C22000770002. The opinions, findings, conclusions, or recommendations expressed in this material are solely those of the authors and do not necessarily represent the views of CTM. Any remaining errors are the sole responsibility of the authors.

Appendix

This Appendix describes the procedure used to calibrate the probability threshold (cut-off) in the severity classification model. In binary classification, the choice of the threshold is crucial, as it determines how predicted probabilities are translated into categorical outcomes. In a balanced dataset—where positive (1) and negative (0) cases occur in roughly equal proportions—the conventional threshold of 0.5 is typically appropriate. Under this rule, an event is classified as positive when its predicted probability is at least 0.5 and as negative otherwise.

In this study, however, a major challenge lies in the pronounced class imbalance: severe fare evasion events account for only 8.6 % of the total observations. Using the default 0.5 threshold in this context biases the model toward predicting the majority class (non-severe events), resulting in a systematic underestimation of severe cases and, consequently, a high rate of false negatives. Lowering the threshold allows the model to become more sensitive to severe events, thereby reducing false negatives (higher *TPR*) at the expense of a moderate increase in false positives (lower *PPV*) (Fig. 3). This trade-off is acceptable—and even desirable—when the main goal is to improve the detection of severe events, which are of greater concern from a policy and operational perspective (Branco et al., 2016).

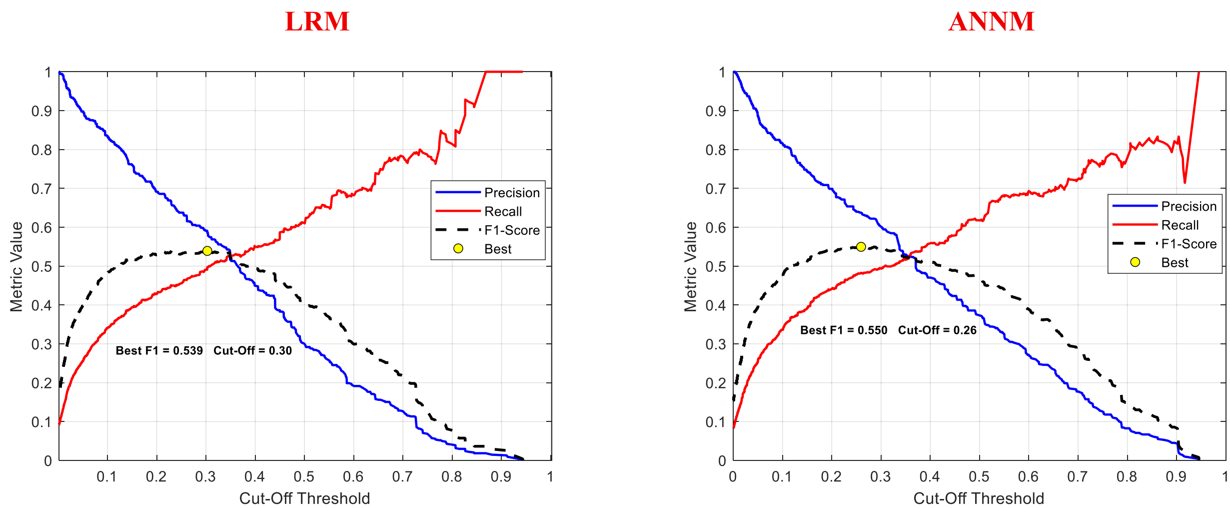

Fig. 3. Trend of precision, recall, and F1-score as a function of the cut-off probability threshold.

Threshold calibration can be effectively guided by the Precision-Recall Curve (PRC), a widely used tool for assessing model performance across different probability thresholds on imbalanced datasets. The PRC plots precision (*PPV*) against recall (*TPR*), providing a more informative assessment than the ROC curve in cases of strong class imbalance (Saito and Rehmsmeier, 2015). By examining how precision and recall vary with the threshold, one can identify the point that achieves the best compromise between the two metrics.

A common approach is to select the threshold that maximizes the F1 score, a measure that captures the harmonic balance between precision and recall (Lipton et al., 2014). This criterion identifies the probability cut-off that optimally balances the correct identification of severe events with the control of false alarms. Adopting such a calibrated threshold enhances the model's sensitivity to severe fare evasion cases while maintaining a reasonable level of specificity—an outcome consistent with the analytical objectives of this study.

Following this procedure, the optimal cut-off values were identified as 0.30 for the LRM and 0.26 for the ANNM (Figs. 3 and 4). To avoid overfitting, the training dataset was used for this procedure, preserving the test dataset for generalization assessment.

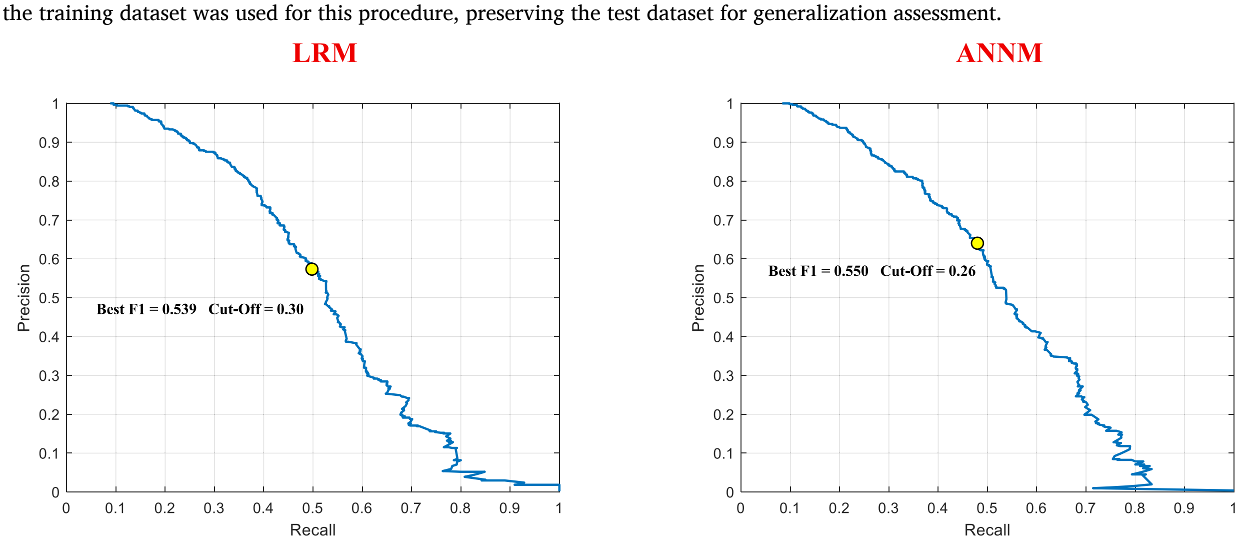

Fig. 4. Precision-Recall Curve (PRC) and best cut-off probability threshold.

The slightly lower threshold for the ANNM is consistent with its probability estimates being less sharply separated: to maximize *F1*, the model requires a smaller predicted probability to classify an event as severe, which effectively increases recall. In contrast, the LRM produces more calibrated and stable probabilities, so a slightly higher threshold is sufficient.

This difference aligns with the model calibration characteristics, as discussed in III comparison strategy: while ANNM achieves higher detection rates for severe events, LRM provides more consistent and interpretable probability estimates. Therefore, the choice between these models should consider the operational objective—whether the priority is maximizing detection of severe fare evasion or ensuring stable probability calibration for a more nuanced decision making.

Data availability

The data that has been used is confidential.

References

Abeler, J., Becker, A., Falk, A., 2014. Representative evidence on lying costs. J. Publ. Econ. 113, 96–104.

Alba, A.C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Guyatt, G., 2017. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA 318 (14), 1377–1384. https://doi.org/10.1001/jama.2017.12126.

Alhassan, I., Matthews, B., Toner, J., Susilo, Y., 2022. Seamless public transport ticket inspection: exploring users' reaction to next-generation ticket inspection. Journal of Public Transportation 24, 100004.

Ariel, B., Englefield, A., Denley, J., 2019. I heard it through the grapevine: a randomized controlled trial on the direct and vicarious effects of preventative specific deterrence initiatives in criminal networks. J. Crim. Law Criminol. 109, 819.

Ayal, S., Celse, J., Hochman, G., 2021. Crafting messages to fight dishonesty: a field investigation of the effects of social norms and watching eye cues on fare evasion. Organ. Behav. Hum. Decis. Process. 166, 9–19.

Barabino, B., Carra, M., Currie, G., 2024a. Fare inspection in proof-of-payment transit networks: a review. Journal of Public Transportation 26, 100101.

Barabino, B., Deiana, E., Mozzoni, S., 2013. The quality of public transport service: the 13816 standard and a methodological approach to an Italian case. Ing. Ferrov. 68 (5), 475–499.

Barabino, B., Deiana, E., Tilocca, P., 2011. Urban transport management and customer perceived quality: a case study in the metropolitan area of Cagliari, Italy. Theor. Empir. Res. Urban Manag. 6 (1), 19–32.

Barabino, B., Di Francesco, M., Ventura, R., 2023. Evaluating fare evasion risk in bus transit networks. Transp. Res. Interdiscip. Perspect. 20, 100854. https://doi.org/10.1016/j.trip.2023.100854.

Barabino, B., Di Francesco, M., Ventura, R., Zanda, S., 2024b. Toward real-time deterrence against fare evasion risk in public transport. Transp. Res. Interdiscip. Perspect. 27, 101238. https://doi.org/10.1016/j.trip.2024.101238.

Barabino, B., Lai, C., Olivo, A., 2020. Fare evasion in public transport systems: a review of the literature. Public Transport 12, 27–88.

Barabino, B., Salis, S., Useli, B., 2015. What are the determinants in making people free riders in proof-of-payment transit systems? Evidence from Italy. Transport. Res. Pol. Pract. 80, 184–196.

Barabino, B., Salis, S., 2020. Do students, workers, and unemployed passengers respond differently to the intention to evade fares? An empirical research. Transp. Res. Interdiscip. Perspect. 7, 100215.

Barabino, B., Salis, S., Useli, B., 2022. Assessing the intention to evade fares for demographic segments of passengers: empirical research in Italy for building smart (er) cities. J. Urban Plann. Dev. 148 (1), 05021066.

Barabino, B., Salis, S., 2019. Moving towards a more accurate level of inspection against fare evasion in proof-of-payment transit systems. Network. Spatial Econ. 19 (4), 1319–1346.

Barabino, B., Salis, S., 2023. Segmenting fare-evaders by tandem clustering and logistic regression models. Public Transport 15 (1), 61–96.

Barabino, B., Ventura, R., 2025. Comparing econometric and machine learning models to fare evasion prediction. Transp. Res. Interdiscip. Perspect. 34, 101636.

Branco, P., Torgo, L., Ribeiro, R.P., 2016. A survey of predictive modeling on imbalanced domains. ACM Comput. Surv. 49 (2), 1–50.

Boukerche, A., Wang, J., 2020. Machine learning-based traffic prediction models for intelligent transportation systems. Comput. Netw. 181.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bijleveld, C., 2007. Fare dodging and the strong arm of the law: an experimental evaluation of two different penalty schemes for fare evasion. J. Exp. Criminol. 3, 183–199.

Bucciol, A., Landini, F., Piovesan, M., 2013. Unethical behaviour in the field: Demographic characteristics and beliefs of the cheater. J. Econ. Behav. Organ. 93, 248–257.

Buehler, S., Halbheer, D., Lechner, M., 2017. Payment evasion. J. Ind. Econ. 65 (4), 804–832.

Busco, C., Gonzalez, F., Jaqueih, Y., Jimenez, F., Alonso, B., 2021. Understanding transantiago users' motivations for paying or evading payment of bus fares. Journal of Public Transportation 23 (2), 1–15. ù.

Busco, C., González, F., San Martín, S., 2024. Bus fare evasion of the Santiago de Chile transportation system from a stakeholders perspective. Travel Behaviour and Society 35, 100715.

Cao, L., 2017. Data science: a comprehensive overview. ACM Comput. Surv. 50 (3), 1–42.

Cantillo, A., Raveau, S., Muñoz, J.C., 2022. Fare evasion on public transport: who, when, where and how? Transport. Res. Pol. Pract. 156, 285–295.

Zhang, P., 2010. Industrial control system simulation routines. In: Advanced Industrial Control Technology. Elsevier, pp. 781–810. https://doi.org/10.1016/B978-1-4377-7807-6.10019-1.

Celse, J. (under review). Behavioural Insights into Fare Evasion: How Message Framing Shapes Risk Perception and Information Retention in Public Transport. Available at: SSRN 5335590.

Celse, J., Grolleau, G., 2023. Fare evasion and information provision: what information should be provided to reduce fare-evasion? Transp. Policy 138, 119–128.

Chang, L.Y., 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Safety Sci. 43 (8), 541–557. https://doi.org/10.1016/j.ssci.2005.04.004.

Cools, M., Fabbro, Y., Bellemans, T., 2018. Identification of the determinants of fare evasion. Case Studies on Transport Policy 6 (3), 348–352.

Cosby, S., 1985. A method for measuring the revenue loss due to fraud within a public transport undertaking. Traffic Eng. Control 26 (2), 59–61.

CTM, 2015. Diagnosi dell'evasione tariffaria nella rete di continuità urbana del CTM, pp. 1–65. Technical Report (in Italian).

CTM, 2014. Diagnosi dell'evasione tariffaria nella rete di continuità urbana del CTM, pp. 1–53. Technical Report (in Italian).

CTM, 2013. Diagnosi dell'evasione tariffaria nella rete di continuità urbana del CTM, pp. 1–48. Technical Report (in Italian).

Currie, G., Delbosc, A., 2017. An empirical model for the psychology of deliberate and unintentional fare evasion. Transp. Policy 54, 21–29.

Currie, G., Delbosc, A., 2021. The paradigm shift in revenue protection research and practice. In: Currie, G. (Ed.), Handbook of Public Transport Research. Edward Elgar Publishing, pp. 115–135.

CTM, 2023. Carta mobilità 2023–2024. Available at: https://ctmcagliari.portaletrasparenza.net/it/trasparenza/servizi-erogati/carta-dei-servizi-e-standarddiqualita.html. (Accessed 16 July 2024).

Dai, Z., Galeotti, F., Villeval, M.C., 2017. The efficiency of crackdowns: a lab-in-the-field experiment in public transportations. Theor. Decis. 82 (2), 249–271.

Delbosc, A., Currie, G., 2016. Cluster analysis of fare evasion behaviours in Melbourne, Australia. Transp. Policy 50, 29–36.

Delbosc, A., Currie, G., 2019. Why do people fare evade? A global shift in fare evasion research. Transp. Rev. 39 (3), 376–391. https://doi.org/10.1080/01441647.2018.1482382.

Eddy, D., 2010. Fare evasion – is it a youth issue? Transit Australia – Australia's Urban Passenger Transport Journal 65 (12), 1–7.

Escalona, P., Brotcorne, L., Fortz, B., Ramirez, M., 2024. Fare inspection patrolling under in-station selective inspection policy. Ann. Oper. Res. 332 (1), 191–212.

Escalona, P., Brotcorne, L., Fortz, B., Wolf, N., 2025. Spot-fare inspection in urban bus transportation systems: strategy and unpredictability under a stackelberg game approach. Public Transport 1–36.

Fatima, M., Pasha, M., et al., 2017. Survey of machine learning algorithms for disease diagnostic. J. Intell. Learn Syst. Appl. 9 (1), 1.

Feldman, Y., 2025. Can the Public be Trusted?: on the Promise and Perils of Voluntary Compliance. Cambridge University Press.

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

Ferlitsch, A., 2021. Deep Learning Patterns and Practices. Simon and Schuster.

Freiria, S., Sousa, N., 2025. Determinants of fare evasion in urban bus lines: case study of a large database considering spatial components. Urban Sci. 9 (6), 231.

González, F., Busco, C., Codocedo, K., 2019. Fare evasion in public transport: grouping transantiago users' behavior. Sustainability 11 (23), 6543.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press, Cambridge. Vol. 1, No. 2.

Greene, W.H., 1993. Econometric Analysis. Macmillan Publishing Company, New York.

Guarda, P., Galilea, P., Paget-Seekins, L., Ortúzar, J.D.D., 2016. What is behind fare evasion in urban bus systems? An econometric approach. Transport. Res. Pol. Pract. 84, 55–71.

Guzman, L.A., Arellana, J., Camargo, J.P., 2021. A hybrid discrete choice model to understand the effect of public policy on fare evasion discouragement in Bogotá's bus rapid transit. Transport. Res. Pol. Pract. 151, 140–153.

Hegde, J., Rokseth, B., 2020. Applications of machine learning methods for engineering risk assessment–A review. Saf. Sci. 122, 104492.

Hauber, A.R., 1980. Daily life and the law: discrepancies in behavior. Int. J. Law Psychiatr. 3 (2), 187–192.

Israel, S.I., Strathman, J.G., 2002. Analysis of Transit Fare Evasion in the Rose Quarter. Technical Report, Center for Urban Studies. College of Urban and Public Affairs, Portland State University.

Jhaveri, R.H., Revathi, A., Ramana, K., Raut, R., Dhanaraj, R.K., 2022. A review on machine learning strategies for real-world engineering applications. Mob. Inf. Syst. 2022 (1), 1833507.

Lee, J., 2011. Uncovering San Francisco, California, Muni's Proof-of-payment patterns to help reduce fare evasion. Transport. Res. Rec.: J. Transport. Res. Board 2216 (1), 75–84. https://doi.org/10.3141/2216-09.

Lipton, Z.C., Elkan, C., Naryanaswamy, B., 2014. Optimal thresholding of classifiers to maximize F1 measure. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 225–239.
Mendoza, J.P., Wielhouwer, J.L., 2015. Only the carrot, not the stick: incorporating trust into the enforcement of regulation. PLoS One 10 (2), e0117212.
Multisystems, Inc., Mundle & Associates, Inc., and Parsons Transportation Group, Inc., 2002. A Toolkit for Self-Service, Barrier-free Fare Collection. Transit Cooperative Research Program, Report 80. TRB, Washington D.C.
Mussone, L., Bassani, M., Masci, P., 2017. Back-propagation neural networks and generalized linear mixed models to investigate vehicular flow and weather data relationships with crash severity in urban road segments. In: Transport Infrastructure and Systems. Taylor & Francis Group, London, pp. 731–739.
Ortúzar, J.D.D., Willumsen, L.G., 2024. Modelling Transport. Fifth Edition. John Wiley & Sons.
Perrotta, A.F., 2017. Transit fare affordability findings from a qualitative study. Publ. Works Manag. Pol. 22 (3), 226–252.
Porath, K., Galilea, P., 2020. Temporal analysis of fare evasion in transantiago: a socio-political view. Res. Transport. Econ. 83, 100958.
Prokosch, A., Gartsman, A., 2017. All-door boarding without Proof-of-Payment: revenue impacts and operational implications. Transp. Res. Board 17-06500, 1–14.
Raccagni, S., Ventura, R., Barabino, B., 2025. Evaluating road crash risk for a mega-event: an application to 2026 milano-cortina winter olympic games. Transp. Res. Interdiscip. Perspect. 34, 101709.
Reddy, A.V., Kuhls, J., Lu, A., 2011. Measuring and controlling subway fare evasion. Transport. Res. Rec.: J. Transport. Res. Board 2216 (1), 85–99. https://doi.org/10.3141/2216-10.
Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10 (3), e0118432.
Sarker, I.H., Hoque, M.M., Uddin, M.K., Alsanoosy, T., 2021. Mobile data science and intelligent apps: concepts, AI-based modeling and research directions. Mobile Network. Appl. 26 (1), 285–303.
Sasaki, Y., 2014. Optimal choices of fare collection systems for public transportations: barrier versus barrier-free. Transp. Res. Part B 60, 107–114.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
Schwerdtfeger, S., 2016. Fare Evasion: the Last Resort of Social Exclusion or Simply a Way to save Money? RGS-IBG annual international conference 2016. https://doi.org/10.13140/RG.2.2.18674.50889.
Sramka, M., Slovak, M., Tuckova, J., Stodulka, P., 2019. Improving clinical refractive results of cataract surgery by machine learning. PeerJ 29 (7), 1–23. https://doi.org/10.7717/peerj.7202.
Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinf. 8, 1–21.
Tilocca, P., Farris, S., Angius, S., Argiolas, R., Obino, A., Secchi, S., Mozzoni, S., Barabino, B., 2017. Managing data and rethinking applications in an innovative mid-sized bus fleet. Transp. Res. Procedia 25, 1899–1919.
Transit Cooperative Research Program (TCRP), 2005. On Board and Intercept Transit Survey Techniques – a Synthesis of Transit Practice, Synthesis 63. Transportation Research Board, Washington, DC.
Traxler, C., 2010. Social norms and conditional cooperative taxpayers. Eur. J. Polit. Econ. 26 (1), 89–103.
Troncoso, R., de Grange, L., 2017. Fare evasion in public transport: A time series approach. Transport. Res. A: Pol. Pract. 100, 311–318.
Van Calster, B., McLernon, D.J., van Smeden, M., Wynants, L., Steyerberg, E.W., 2019. Calibration: the achilles heel of predictive analytics. BMC Med. 17 (1), 230. https://doi.org/10.1186/s12916-019-1466-7.
Ventura, R., Barabino, B., Maternini, G., 2024. Prediction of the severity of exceeding design traffic loads on highway bridges. Heliyon 10 (1), e23374.
Wilson, A.A., Nehme, A., Dhyani, A., Mahbub, K., 2024. A comparison of generalised linear modelling with machine learning approaches for predicting loss cost in motor insurance. Risks 12 (4), 62.
Wolfgram, L., Pollan, C., Hostetter, K., Martin, A., Spencer, T., Rodda, S., Amey, A., 2022. Measuring and managing fare evasion. Transp. Res. Board. https://doi.org/10.17226/26514.