THE BALTIC JOURNAL
OF ROAD
AND BRIDGE
ENGINEERING
2019/14(3)

ISSN 1822-427X/eISSN 1822-4288
2019 Volume 14 Issue 3: 360–383
https://doi.org/10.7250/bjrbe.2019-14.448

ROAD FUNCTIONAL CLASSIFICATION USING PATTERN RECOGNITION TECHNIQUES

GAETANO BOSURGI[1], ORAZIO PELLEGRINO[2], GIUSEPPE SOLLAZZO[3*]

[1,2]*Dept of Engineering, University of Messina, Messina, Italy*
[3]*Dept of Engineering, University of Palermo, Palermo, Italy*

Received 19 December 2018; accepted 15 May 2019

Abstract. The existing international standards suggest a methodology to assign a specific functional class to a road, by the values of some features, both geometrical and use-related. Sometimes, these characteristics are in contrast with each other and direct the analyst towards conflicting classes for a road or, worse, one or more of these features vary heterogeneously along the road. In these conditions, the analyst assigns the class that, by his capability and experience, he retains the most appropriate, in a very subjective way. On the contrary, the definition of an automatic procedure assuring an objective identification of the most appropriate functional class for each road would be desirable. Such a solution would be useful, especially when the road belongs to the existing infrastructure network or when it was not realised by out of date standards. The proposed procedure regards the definition of a classification model based on Pattern Recognition techniques, considering 13 input variables that, depending on their assumed value, direct the analyst towards one of the four functional classes defined by the Italian standards. In this way, it is possible to classify a road even when its characteristics are heterogeneous and conflicting. Moreover, the authors analysed the model limitations, in terms of errors and dataset size, considering observation and variable numbers. This

* Corresponding author. E-mail: giuseppe.sollazzo@unipa.it

Gaetano BOSURGI (ORCID ID 0000-0002-0782-5510)
Orazio PELLEGRINO (ORCID ID 0000-0002-7990-3581)
Giuseppe SOLLAZZO (ORCID ID 0000-0002-7116-7946)

approach, representing a beneficial decision support tool for the decision-maker, is exploitable for both planned and existing roads and becomes particularly advantageous for road agencies aiming to optimally allocate their limited funds for specific interventions assuring the achievement of a fixed functional class.

Keywords: functional classification, pattern recognition, road classification, road network.

Introduction

As known, the vehicular flow mobility has numerous analogies with the blood circulation: the more the transport system is similar to the blood vessel system in the human body, the higher its efficiency level. In the human body, a network of thick vessels start from the hearth and, moving away from it, they become thinner and thinner, to thoroughly and efficiently perfuse all the human organs. Similarly, the same gradation and harmony have to characterise the road networks where the user, through the information provided by the external context, adapt his speed to reach safely and quickly his destination (Charlton & Starkey, 2017; Goto & Nakamura, 2016).

On the contrary, perceiving inhomogeneity, mainly when locally concentrated, causes hazards and issues; then, the road agency has to identify and solve these critical conditions through proper interventions (Anderson & Hernandez, 2017; Friedrich, 2017; Wang, You, & Wang, 2017).

Although road classification standards are slightly different among various countries (American Association of State..., 2004; Bosurgi, D'Andrea, & Pellegrino, 2011; Federal Highways Administration, 2004; Friedrich, 2017; Lamm, Psarianos, & Mailaender, 1999; Ministero delle Infrastrutture..., 2001; The British Standards Institution, 2002), it is possible to assess that, generally, it depends on a series of geometrical features and the kind of service offered to users. The offered service is a function of the road movement type (Connection, Collection, Penetration, Local Movement), the movement entity, its function in the crossed context and the admitted traffic components.

The problem gets slightly complicated when the various net functional levels (primary – connectors, main – manifolds, secondary, local), to which the examined road is linked, have to correspond to the constructive typologies defined by the standards (Chen, Namdeo, & Bell, 2008; Giummarra, 2003; Jaarsma, 1997; Kaptein & Claessens, 1998; Liu, Yan, & Wang, 2017).

In Italy, these typologies are the following six: motorways, main highways, highways, urban highways, urban roads, local roads. Other countries follow similar classifications: in USA and Canada, roads are

divided into Freeways, Arterials, Collector e Local roads; in the United Kingdom they are classified in motorway, primary *A*-road, non-primary *A*-road, *B* road, *C* road, unclassified; in France, autoroutes, route nationale, routes départmentales, routes communales.

In any case, the standard methodologies create interpretation difficulties, especially for existing roads or when the external context has been strongly modified from an urbanistic point of view, making the road inappropriate for its novel function. At this regard, some authors focused on re-thinking the existing classification procedures. For example, Liu, Yan, & Wang (2017) evidenced that urban roads classification rules must be extended, avoiding considering only car traffic, as made so far. There is a need, instead, to deeply consider the role of internal city spaces and means of transport, also the alternative ones, such as bikes or public means, generally promoted by the local institutions for their limited environmental impact.

In short traits, it is also possible that the values of variables considered by the standards are lower than the reference threshold value for that specific class. Then, in similar scenarios, the analyst has only two options: assign the road to a lower level class or, even, unclassified the road. These choices identify transitional states, preparatory for proper improvement interventions aimed to re-direct the road functionality to the values chosen by the agency (Dickerson, Peirson, & Vickerman, 2000; Forkenbrock & Foster, 1997; Gomes, 2013; Karlaftis & Golias, 2002; Kashani & Mohaymany, 2011; Kockelman, 2001; Nowakowska, 2010; Rodrigues, Ribeiro, & da Silva Nogueira, 2015).

Where the functional class assignment task is more complicated than usual, extension and significance of the geometrical, functional, or constructive features conflicting with standards must be evaluated. It is strongly suggested to complete such an analysis through objective procedures, avoiding relying on a subjective decision, even when performed by an expert.

Then, the methods for inspections and the consequent subjective estimation must be replaced by instrumental investigations able to organise a sufficiently automatized analytical model (Rodrigues, Ribeiro, & da Silva Nogueira, 2015; Wang, You, & Wang, 2017). For instance, surveys performed with high-speed and high-performance techniques (Wang, Yang, Zhang, Wang, Cao, & Eklund, 2016) have very high sampling frequency, with sufficiently reduced computational costs, but they usually have the drawback of providing large quantities of data, complex to handle using traditional analyses. Therefore, it is strategical not only to recognize redundant or useless data for the fixed scope, to choose an acceptable sampling frequency for the final information quality level – avoiding useless and critical increase in the dataset

size - but also, above all, to identify the most appropriate techniques to extract the desired outputs (D'Andrea & Pellegrino, 2012; D'Andrea, Cappadona, La Rosa, & Pellegrino, 2014; Kuehnle & Burghout, 1998; Mena, 2003; Paclık, Novovičová, Pudil, & Somol, 2000; Pellegrino, 2011; Praticò & Giunta, 2011; Sohn & Lee, 2003).

This brief literature review evidences some issues that this research aims to solve:

- high dependence of the road functional classification procedures proposed by the standards on the subjective decision of the analyst;
- difficult evaluation for the existing roads, especially where the external context evolved from an urbanistic perspective;
- the relative complexity of the proposed theoretical models aiming to automatize the classification procedures, which become too complicated to be handled by the technical personnel in the public offices.

The proposal of the authors, thus, regards the definition of an analytical procedure to solve these issues. The paper describes the organisation of a model for automatic identification of road segment functional class, based on several features, particularly significant for the road standards, related to the road context.

1. Methodology

The idea is to define a model able to judge any possible scenario, and it requires an appropriate training phase, in which the analyst "teaches" the classification behaviour suggested by the standards. The goal of the methodology is then an automatic identification and classification of various "objects" or observations, related to specific road segments, in some classes (already known), by some features or variables. For the mathematical model definition, some initial hypotheses – like the a priori knowledge of the output classes suggested by the standards and the assignment of novel observations in one of the previously defined classes (classification problem) – led the authors to adopt Pattern Recognition (PR) techniques, able to identify common behaviours in highly chaotic dataset.

In general, the available analytical procedures are divided into four main technique groups:

- Template Matching;
- Statistical Classification;
- Syntactic or Structural Matching;
- Artificial Intelligence.

In this study, the authors considered the second group (statistical classification), in which the model is built on a data set, including all the considered features (input variables). The methodology tends to identify objects (or observations) belonging to different classes since it is possible to consider them mathematically separated in a multidimensional space. An appropriate training and learning phase teaches the model how to group the available observations, in a supervised approach. Then, the model attempts to generalise the classification learned and assigns novel observations to the available classes.

Cluster analysis and soft computing techniques (neural networks, fuzzy logic) are generally considered as further alternatives for solving this kind of problem (Hastie, Tibshirani, & Friedman, 2009). However, the cluster analysis, both in the hard and fuzzy versions, is an unsupervised technique and, thus, it is ineffective to build a predictive model. The artificial intelligence methods, on the contrary, have this capability, but they need huge datasets (more and more than the number considered in this study) and, also, they determine higher computational costs for the elaborations. Moreover, their operative philosophy is similar to black-box models, in which input-output relationships remain hidden and unexplained, although it is possible to validate the final quality (Bosurgi, Pellegrino, & Sollazzo, 2019; Sollazzo, Fwa, & Bosurgi, 2017).

By these motivations, in this paper, the authors considered Linear Discriminant Analyses (LDA and Fisher) and Quadratic Discriminant Analysis (QDA). In the following, for length reasons, some technical notices regarding only the linear analyses techniques will be provided, since they will assure the best performance (as described in the Results section). Moreover, the QDA methodology is very similar and relies on common analogous hypotheses, except for the shape of the discriminant function that, in this case, becomes of a superior order.

Regarding the training dataset, it has been built considering ideal observations concerning a significant part of possible scenarios identifiable in a road environment. The Italian standards, similarly to other international regulations, requires the knowledge of particular variables and indicates specific ranges for assigning the road to different functional classes. Then, the proposed model was trained using theoretical values extracted from the standards, for defining numerous possible ideal scenarios characterised by homogeneous data from different categories which identify a specific class. Then, the training dataset includes 160 observations with 13 input variables, with values limited by the thresholds fixed by the Italian road standards (Ministero delle Infrastrutture..., 2001), varying with the related functional class and covering almost the entire problem analytical domain.

In this way, the model is trained to identify the examined road class in the ideal conditions defined by the standards. Such a model, defined on these examples, will be able to analyse novel real observations in which the functional class is unknown and allocate them to the most appropriate group (Duin & Pękalska, 2012; Hastie, Tibshirani, & Friedman, 2009; Huang & Guan, 2015; Jain, Duin, & Mao, 2000; Ravi, Reddy, & Zimmermann, 2000; Shanahan, Thomas, Mirmehdi, Martin, Campbell, & Baldwin, 2000).

This study focuses only on rural roads and, thus, the considered functional classes are the following four:

A – primary roads;

B – connectors, main roads;

C – manifolds, secondary roads;

D – local roads.

Since the values of the input variables are included in intervals of various sizes, the authors applied a standardisation by dividing them by their standard deviations. In this way, all the variables will have similar weight in the study, avoiding single variable prevalence due to its higher absolute value.

2. A brief note about the Linear Discriminant Analysis (LDA)

In the case of the Linear Discriminant Analysis (LDA), the separation of the detected observations (or objects) is obtained with straight lines in a two-dimensional space, planes in a 3D space or, moreover, hyperplanes in an nD space (where n is the number of input variables).

In general, it is possible to say that all the surveyed observations, reported in a data set X of $[m \times n]$ dimension (where m is the number of features and n is the number of objects), will be assigned to the C classes representing, in this case, different functional classes. Therefore, N_i objects $\{x_1, x_2, \ldots, x^{N_i}\}$ belonging to the class ω_i are obtained.

Then, the purpose is to project the objects belonging in X on a C–1 dimension hyperplane called Y, where it is more convenient to perceive the separation among the classes.

For example, in the simplest case where $C = 2$, the m-dimension data set will have a number of N_1 samples belonging to ω_1 and N_2 belonging to ω_2. Thus, the goal is to get an y scaling for projecting the x observations on an opportune straight line (Eq. (1)):

$$y = w^T x, \tag{1}$$

where

$$x = \begin{bmatrix} x_1 \\ \dots \\ x_m \end{bmatrix} \tag{2}$$

and where

$$w = \begin{bmatrix} w_1 \\ \dots \\ w_m \end{bmatrix}. \tag{3}$$

The following procedure will be applied to obtain the axis which guarantees the best separability among the classes.

First, it is necessary to identify some indices measuring the class separation, as the mean vectors in *x* and *y* (Eq. (4) and Eq. (5)):

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x, \tag{4}$$

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in \omega_i} y = \frac{1}{N_i} \sum_{x \in \omega_i} w^T x = w^T \frac{1}{N_i} \sum_{x \in \omega_i} x = w^T \mu_i. \tag{5}$$

The distance between projected averages (like being the distance between centroids), calculated through Eq. (6), represents an acceptable criterion for the final decision:

$$J(w) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T \mu_1 - w^T \mu_2\right| = \left|w^T (\mu_1 - \mu_2)\right|. \tag{6}$$

However, in this way, there will be no knowledge about dispersion within the classes. For this reason, Fisher (1936) introduced into the above-mentioned objective function *J*(w) normalization on a measure representative of this dispersion, called scatter. $\tilde{S}_i^2$ represents in such a way the variability within the class ω_i after projecting it along the new axis *y,* as in Eq. (7):

$$\tilde{S}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2. \tag{7}$$

The sum $\tilde{S}_1^2 + \tilde{S}_2^2$ corresponds to the variability into the two classes after the projection on the new *y*-axis, and the analyst aims to find the $w^T x$ linear function that maximises the function *J*(*w*), reported in Eq. (8):

$$J(w) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|^2}{\tilde{S}_1^2 + \tilde{S}_2^2}. \tag{8}$$

In conclusion, in the ideal representation on the new axis, the observations belonging to the same class are very close to each other and, at the same time, the averages of the different classes are as far as possible. It is necessary to express $J(w)$ as an explicit function of w to find the function maximum. Then, it is possible to define a scatter in the multivariate space x, like Eq. (9) and Eq. (10):

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T, \tag{9}$$

$$S_W = S_1 + S_2, \tag{10}$$

where S_i is the covariance matrix of the ω_i class and S_w is the within-class scatter matrix.

The scatter of the projection on y, expressed as a function of the scatter matrix in the x space, is expressed through Eq. (11)and Eq. (12):

$$\tilde{S}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2 = \sum_{x \in \omega_i} (w^T x - w^T \mu_i)^2 = \sum_{x \in \omega_i} w^T (x - \mu_i)(x - \mu_i)^T w = w^T S_i w, \tag{11}$$

$$\tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_1 w + w^T S_2 w = w^T (S_1 + S_2) w = w^T S_w w = \tilde{S}_w, \tag{12}$$

where $\tilde{S}_w$ is the scatter matrix referred to the class projected on the y-axis.

Similarly, it is possible to derive the differences among averages projected in the y-axis in terms of averages in the original x space, as reported in Eq. (13):

$$\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2 = (w^T \mu_1 - w^T \mu_2)^2 = w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = w^T S_B w = \tilde{S}_B. \tag{13}$$

The S_B matrix represents the scatter between the class of the original observations, while $\tilde{S}_B$ is that reported on the y-axis.

So, Eq. (8) becomes Eq. (14):

$$J(w) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} = \frac{w^T S_B w}{w^T S_w w}, \tag{14}$$

where $J(w)$ is a measure of the difference among the considered classes means, normalised by the value of the within-class scatter matrix.

The derivative equal to zero, as is known, gives the maximum of the function. The final expression (bypassing the other steps) is represented by Eq. (15):

$$S_w^{-1} S_B w - J(w) w = 0. \tag{15}$$

By solving the eigenvalue problem, it derives Eq. (16):

$$S_w^{-1} S_B w = \lambda w, \tag{16}$$

where $\lambda = J(w)$ is a scalar.

The so-called linear discriminant gives the optimal solution, reported in Eq. (17):

$$w^* = \arg\max_w J(w) = \arg\max_w \left(\frac{w^T S_B w}{w^T S_w w} \right) = S_w^{-1} \left(\mu_1 - \mu_2 \right). \tag{17}$$

Whether the classes are more than 2 (for example C), there will be C–1 projection vectors w_i (instead of the only y) but the procedure will be the same.

As usual for an analytical procedure, some limitations and restrictive hypotheses exist: the discriminant analysis provides results strictly related to the input data, then the higher the variable discriminant power, the greater its influence on the final result, despite the unremarkable conceptual importance. Furthermore, modifications in the number and type of variables (including or removing some variables) will substantially change the decisional space. However, these drawbacks do not cause relevant errors in the model response, and this will be widely proved in the Result section.

3. Results

Although it is possible to adapt this procedure to every international road standard easily, the author performed a numerical examination by the Italian Standard requirements. Shortly, the road functional class depends on some features (or variables) listed in Table 1 with the reference value for the four considered functional classes (*A*, *B*, *C*, and *F*).

Referring to Table 1, the authors defined a reference dataset, including 160 records, equally distributed for the four functional classes. In detail, the first 40 records provide values of the 13 variables included in the admittible ranges of the *A* functional class. The second 40 records identify the functional class *B*, and so on for classes *C* and *F*. In Table 2, a short extract of this dataset (only 15 records), used in the numerical tests for both training and testing the model, is presented. Some lexical

Table 1. Variables or features considered by the Italian road standards* for road classification

Features	Variable	Class *A*	Class *B*	Class *C*	Class *F*
1	Net function	Statewide	Regional	Provincial	Municipal
2	Movement type	Connector	Collector	Penetration	Local
3	Capacity, vph	3600–7200	1800–3600	1200–1800	400–1200
4	Design Speed, km/h	90–140	70–120	60–100	40–100
5	V_{85} - Operating Speed, km/h	115–135	95–115	80–95	60–80
6	Radius – Minimum, m	339	178	118	45
7	Lane Minimum Width, m	3.75	3.75	3.75	3.50
8	Traffic Island Minimum Width, m	2.60	2.50	0.00	0.00
9	Left Shoulder – Minimum Width, m	0.70	0.50	0.00	0.00
10	Right Shoulder – Minimum Width, m	2.50	1.75	1.50	1.00
11	Light Vehicle Flow, vph	1100	1000	600	450
12	Traffic Components	Restrict	Restrict	All eligible	All eligible
13	Access	Not allowable	Not allowable	Allowable	Allowable

Note: *Ministero delle Infrastrutture... (2001)

Table 2. Example of the ideal training dataset

1	2	3	4	5	6	7	8	9	10	11	12	13	Class
4	4	7200	140	135	2200	4.5	3.0	1.6	3.4	1190	1	1	*A*
4	4	4367	119	120	1101	4.2	2.9	1.2	2.6	1141	1	1	*A*
4	4	5005	95	124	1416	4.2	2.9	1.3	2.8	1134	1	1	*A*
4	4	6331	126	134	1595	4.3	2.7	1.0	2.8	1117	1	1	*A*
4	4	5433	121	128	725	4.2	2.9	0.8	3.4	1159	1	1	*A*
4	4	5462	135	116	1081	3.9	2.8	0.8	2.7	1113	1	1	*A*
4	4	6508	99	127	1208	4.4	2.7	1.1	2.8	1179	1	1	*A*
4	4	4678	118	129	629	4.1	2.8	0.8	2.9	1153	1	1	*A*
4	4	4642	92	126	1612	4.5	2.9	1.1	3.0	1128	1	1	*A*
4	4	6944	114	134	1251	4.2	2.9	0.8	2.8	1147	1	1	*A*
4	4	3944	130	128	821	3.8	2.7	1.0	3.0	1123	1	1	*A*
4	4	5551	124	132	1681	4.1	2.9	1.4	3.1	1103	1	1	*A*
4	4	5458	103	134	2137	4.4	2.6	1.3	2.9	1180	1	1	*A*
4	4	5907	125	133	365	4.3	2.7	1.4	2.7	1116	1	1	*A*
4	4	6470	101	121	1749	4.2	3.0	1.0	2.9	1166	1	1	*A*
...	...	...	...	...	...	...	...	...	...	...	...	...	...

variables have been transformed into numbers to enter only values in a numerical format. For example, for Movement Type, Local = 1, Penetration = 2, Collector = 3, Connector = 4, and other.

In addition to this dataset, the analysis focused on an existing road. In particular, the authors considered a rural road segment, in the city area of Messina (Italy), 11 km long. Along this road, the values of the 13 reference variables have been measured in 24 sections, representing homogeneous segments along the road (in terms of geometrical and traffic features, by the involved variables). Table 3 provides these values: it is interesting to notice how the output column is absent in Table 3 (it was, on the contrary,

Table 3. Dataset measured for the analysed existing road

Feature												
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1	2	1200	73	67	200	4.35	0.50	0.30	1.25	11	0	1
1	2	1200	100	80	700	3.07	0.36	0.00	1.00	11	0	1
1	2	1200	76	73	220	3.07	0.36	0.00	1.00	11	0	1
1	2	1200	51	53	80	3.25	0.36	0.00	1.00	11	0	1
1	2	1200	100	80	500	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	100	86	800	4.19	0.50	0.30	1.25	11	0	1
1	2	1200	92	72	350	3.26	0.36	0.00	1.00	11	0	1
1	2	1200	45	56	60	3.07	0.36	0.00	1.00	11	0	1
1	2	1200	46	52	62	3.07	0.36	0.00	1.00	11	0	1
1	2	1200	58	72	110	3.59	0.50	0.30	1.25	11	0	1
1	2	1200	40	49	45	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	51	58	80	3.79	0.36	0.00	1.00	11	0	1
1	2	1200	45	57	60	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	37	48	37	4.50	0.36	0.00	1.00	11	0	1
1	2	1200	45	57	60	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	37	50	38	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	44	48	55	3.58	0.36	0.00	1.00	11	0	1
1	2	1200	40	53	45	3.76	0.36	0.00	1.00	11	0	1
1	2	1200	56	57	100	3.25	0.50	0.30	1.25	11	0	1
1	2	1200	62	68	130	4.76	0.50	0.30	1.25	11	0	1
1	2	1200	66	73	150	3.77	0.36	0.00	1.00	11	0	1
1	2	1200	73	71	200	3.25	0.50	0.30	1.25	8	0	1
1	2	1200	100	94	1000	5.30	0.50	0.30	1.25	8	0	1
1	2	1200	73	68	200	3.82	0.50	0.30	1.25	8	0	1

present in the previous table), since the model has to identify the most appropriate functional class and assign each section to it.

A graphical representation of the dataset in a multidimensional space is impractical. However, it is useful to plot data in two-dimensional charts alternatively, to appreciate their disposition in the analysis space. For example, Figure 1 provides the reference dataset represented in a 2D chart in terms of Desing Speed (V_d, vpd) and Capacity.

Although for length motivation, only LDA has been described in the Methodology section, other methods have been applied. Indeed, QDA and Fisher's analysis have also been considered in numerical applications to identify the most effective approach. Error evaluation is used for model validation and for comparing different approaches. The errors are related to the number of erroneous classification in the testing dataset (part of the 160 record dataset), the correct classification of which was already known. It is possible to compute two different errors: the apparent and the real error. The apparent error is calculated on the training record set, while the real error is computed on the testing one. Generally, the apparent error is more optimistic than the real. However, their comparison generally represents an indirect measure of overtraining, consisting of a model adaptation to the training data only and an inability to novel external records.

The entire dataset division in the training and testing parts has been performed with the common cross-validation technique. In practice, the authors divided the starting dataset (160 records) into ten groups (16 records for each group), repeating the classifier modelling procedure

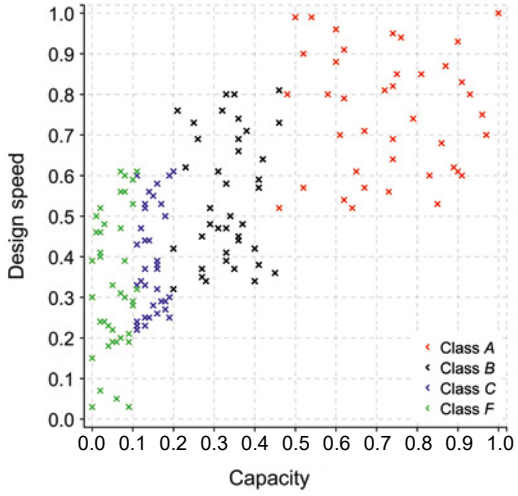


Figure 1. Training dataset in a 2D space (Design Speed vs Capacity)

ten times. For each repetition, 9 of the ten groups have been adopted for training, leaving the other group for testing. After the completion of the ten tests, the average errors have been calculated. Results have shown how LDA provided better behaviour in terms of both classification errors (Table 4 and Figure 3) and computational costs.

In Figure 2a, the linear discriminant segments are shown in the same chart previously provided in Figure 1. It is also interesting to consider other variable couples to observe data from different perspectives and understand the global classification quality reached by the method (Figures 2b–2d), proved by the error number.

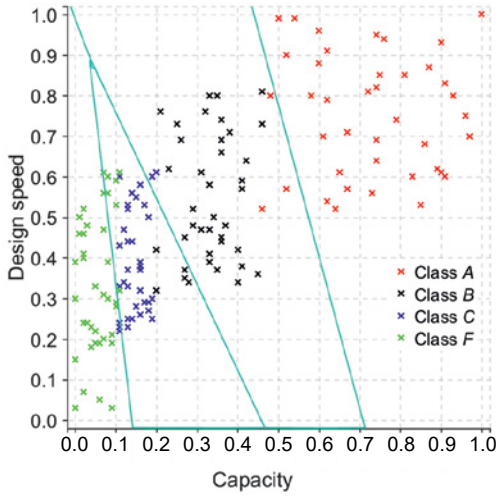
A very relevant element in such a kind of analysis is the dataset size. The authors evaluated the classification error as a function of the training dataset size to verify the procedure effectiveness. Figures 3a and 3b provide the trend of real and apparent error, depending on the number of observations, respectively. It is interesting to notice that the errors tend to an asymptote (near to 0). Fisher performance is not reported in Figure 3b, as it almost matches the LDA trend, except for the final values not tending to an asymptote.

Finally, the acceptability of the adopted input variable number has been tested. In this particular case, the feature number is fixed because the standards explicitly indicate them. However, it is possible to evaluate if this number is too large for this approach, also as a function of the dataset size, quite limited in this case. This processing determines the errors by the number of features, continuously increasing the number of involved variables using a predefined order. When the dataset disposition is changed (for example, by exchanging two or more columns between them), the result also changes slightly. To make this analysis useful, the authors considered the configuration shown in Table 1 and other possible random configurations of the same variables. In detail, three different configurations (by varying the feature position) of the dataset have been tested. However, results show similar behaviour of the model in the different configurations (Figure 4).

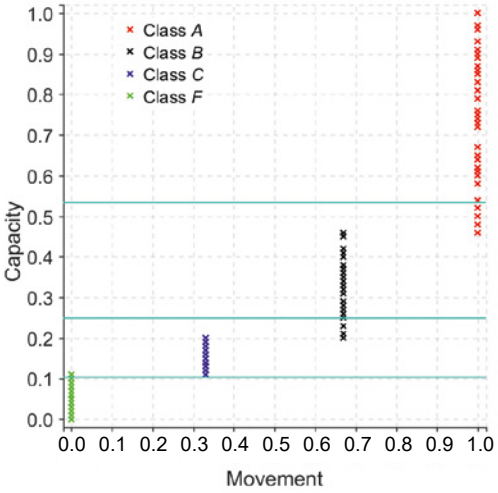
Finally, the classifier results for the 24 observations related to the analysed existing road are reported in Table 5 concerning the LDA method.

Table 4. Percentage of unclassified records for the three techniques

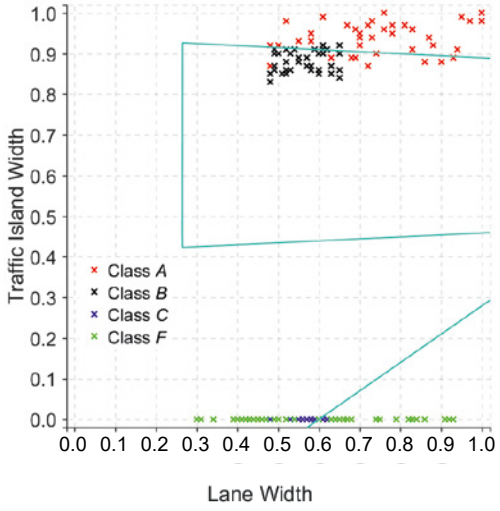
Linear Discriminant Analysis	Quadratic Discriminant Analysis	Fisher
1.25%	6.25%	1.25%



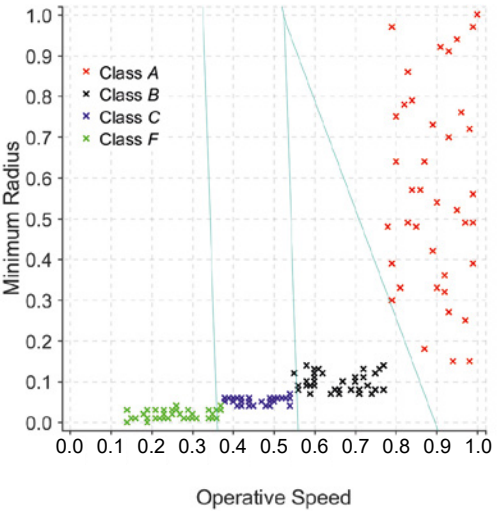
a) Design Speed vs Capacity



b) Capacity vs Movement Type

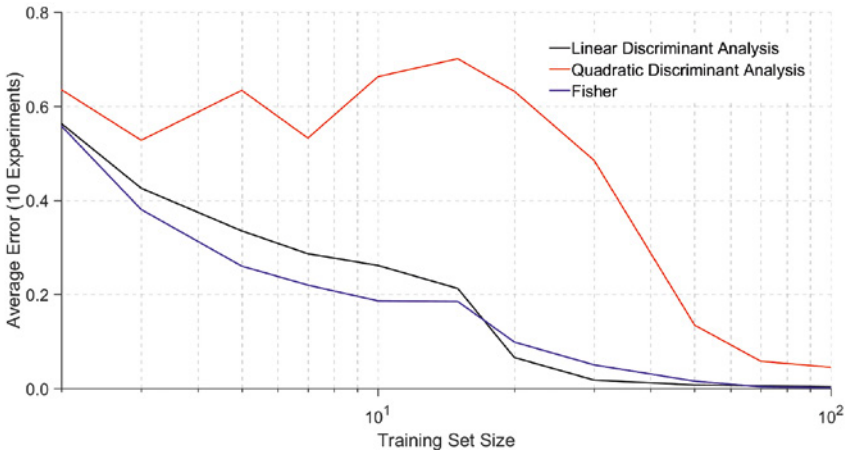


c) Traffic Island Width vs Lane Width

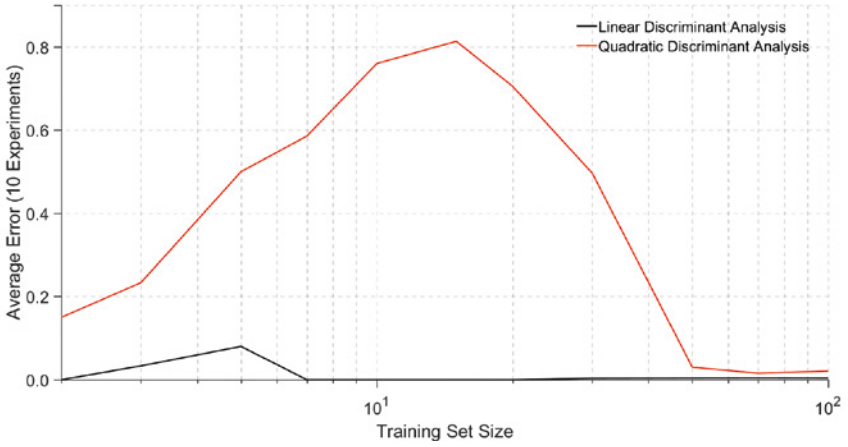


d) Minimum Radius vs Operating Speed

Figure 2. The training dataset is shown in several 2D spaces with the Linear Discriminant Analysis segments



a) real errors



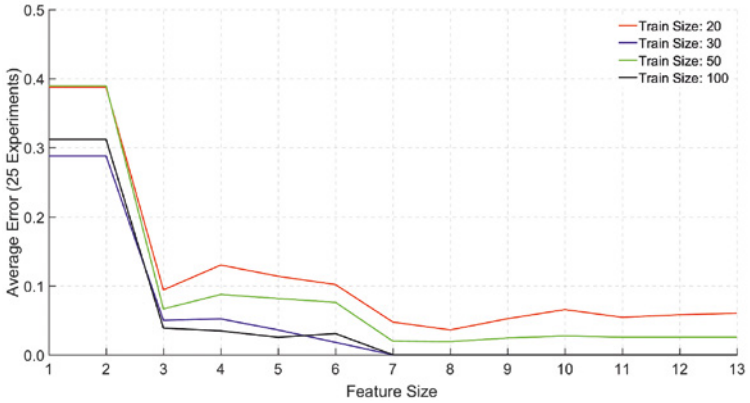
b) apparent errors

Figure 3. Learning curves, representing the performance of the classifiers for different dataset sizes

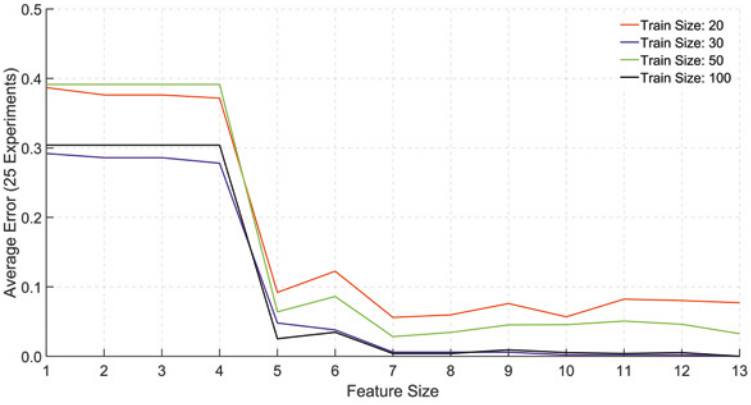
4. Discussion

To define an objective road classifier, the authors applied PR techniques, owing to their characteristics:

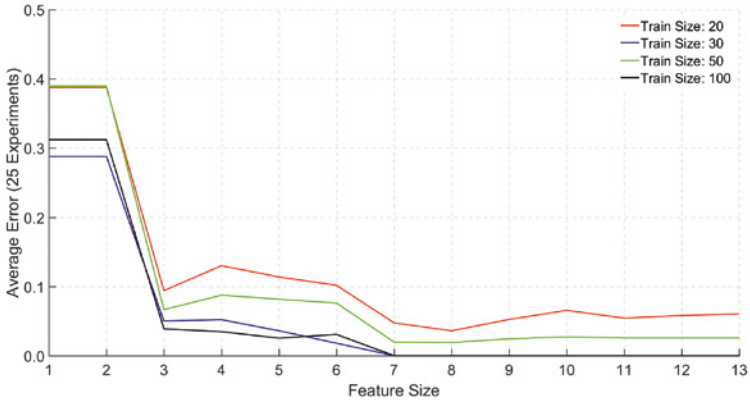
- reliable results also considering reduced databases;
- possibility to define a model including several input variables with very different existence ranges;



a) configuration No. 1



b) configuration No. 2



c) configuration No. 3

Figure 4. Feature curves, representing the trend of the errors when the feature number increases

Table 5. Prediction of the functional class for the 24 observations by the Linear Discriminant Analysis methodology

Obss. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	Predicted Class
OBS1	1	**2**	1200	73	67	**200**	4.35	0.50	0.30	1.25	11	0	**1**	*F*
OBS2	1	**2**	1200	100	80	**700**	3.07	0.36	0.00	1.00	11	0	**1**	*F*
OBS3	1	**2**	1200	76	73	**220**	3.07	0.36	0.00	1.00	11	0	**1**	*F*
OBS4	1	**2**	1200	51	53	80	3.25	0.36	0.00	1.00	11	0	**1**	*F*
OBS5	1	**2**	1200	100	80	**500**	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS6	1	**2**	1200	100	**85.7**	**800**	4.19	0.50	0.30	1.25	11	0	**1**	*F*
OBS7	1	**2**	1200	92	72	**350**	3.26	0.36	0.00	1.00	11	0	**1**	*F*
OBS8	1	**2**	1200	45	56	60	3.07	0.36	0.00	1.00	11	0	**1**	*F*
OBS9	1	**2**	1200	46	52	62	3.07	0.36	0.00	1.00	11	0	**1**	*F*
OBS10	1	**2**	1200	58	72	110	3.59	0.50	0.30	1.25	11	0	**1**	*F*
OBS11	1	**2**	1200	40	49	45	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS12	1	**2**	1200	51	58	80	3.79	0.36	0.00	1.00	11	0	**1**	*F*
OBS13	1	**2**	1200	45	57	60	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS14	1	**2**	1200	37	48	37	4.50	0.36	0.00	1.00	11	0	**1**	*F*
OBS15	1	**2**	1200	45	57	60	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS16	1	**2**	1200	37	50	38	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS17	1	**2**	1200	44	48	55	3.58	0.36	0.00	1.00	11	0	**1**	*F*
OBS18	1	**2**	1200	40	53	45	3.76	0.36	0.00	1.00	11	0	**1**	*F*
OBS19	1	**2**	1200	56	57	100	3.25	0.50	0.30	1.25	11	0	**1**	*F*
OBS20	1	**2**	1200	62	68	**130**	**4.76**	0.50	0.30	1.25	11	0	**1**	*F*
OBS21	1	**2**	1200	66	73	**150**	3.77	0.36	0.00	1.00	11	0	**1**	*F*
OBS22	1	**2**	1200	73	71	**200**	3.25	0.50	0.30	1.25	8	0	**1**	*F*
OBS23	1	**2**	1200	100	**94**	**1000**	**5.30**	0.50	0.30	1.25	8	0	**1**	*F*
OBS24	1	**2**	1200	73	68	**200**	3.82	0.50	0.30	1.25	8	0	**1**	*F*

- the output provided in classes;
- supervised methodologies, also useful for forecasting analyses;
- error estimation as a function of dataset size in terms of observations and features numbers.

In this case, the authors selected the discriminant analysis, because of its application simplicity, aiming to reduce complications for the technical personnel that manage roads, generally lacking in a significant scientific background.

In the numerical applications, three similar discriminant analyses have been tested for identifying the most appropriate one, since, despite their similarities, the methods provide slightly different performance. The numerical outcomes reported in Table 4 (classification errors) and, mainly, in Figure 3 prove that LDA represents the best methodology in this application, but the total error percentage is limited for all the tested methods. This behaviour is likely related to the specific characteristics of the training database, appropriately defined using clear and correct ideal observations univocally related to one of the classes. In truth, the classification accuracy increases whether all the possible feature combinations are defined, but this solution is expensive from a computational point of view. Moreover, also the adoption of a more advanced discriminant method than the linear one in some conditions determines improvements in the output precision. However, the authors aimed to define a simple analytical approach with a reduced record number for the training dataset and the proposed architecture assured remarkable reliability.

A graphical examination further proves the remarkable quality level reached through the LDA (Figure 2). Although the various charts of Figure 2 represent only a partial point of view in terms of the model efficiency – because only a few cases are reported and results are shown in 2D charts – they are helpful to evidence the substantial correctness of the linear classifier and how the various classes are separated in the different dimensions. For instance, by comparing Figures 1 and 2a (referred to the same dimensional space), it is possible to notice that, also from a qualitative perspective, the result is satisfactory and remarkable, despite some residual incorrectly classified observations (for instance, 3 for class *A*, 5 for class *B*).

The proposed procedure aims to solve a problem which is strongly constrained from an analytical point of view. For a rural road, the possible classes are only 4, and the input variables identified by the Italian standards (and considered by most of the international ones) are fixed in number and typology. Consequently, any possible calibration and refinement of the model have to avoid the determination of different formal configurations, obtained by reducing the input number or the output groups. However, the learning and feature curves have been examined to verify the acceptability of the model architecture and, thus, the approach applicability. In particular, it has been investigated whether 13 input variables are too many for a reduced dataset including

160 records only, determining solutions less accurate than those provided by models based on considerably fewer features. The numerical result, in this case too, was positive. Figures 3a and 3b show that the error level is sharply reduced for datasets including more than 50 observations, at least for LDA. Beyond this threshold, the error tends to an asymptote near to zero. Therefore, the 160 observations are widely sufficient to maintain errors below a significant value. The same Figures show that QDA needs a larger record dataset to provide equally acceptable results, confirming the decision to rely on more straightforward methodology, as LDA.

Moreover, considering the adopted database, QDA is affected by a remarkable difference between the apparent and the real errors. This condition evidences a relevant overfitting risk: the model properly fits the training data but, when adopted to classify novel observations, it assures unsatisfactory performance.

Figures 4a–4c provide the feature curves, i.e. the error value for different feature numbers. The numerical outcomes, in this case, depend on the number of features, but also on their order. Then, the authors tested three different configurations – in terms of the order – for higher reliability. For a fixed feature sequence, different model architectures are defined, by considering the fixed features one by one, passing from a single-variable model to a 13-features model. Then, for each architecture, the classification error is computed. Configuration 1 represents features as listed in Table 1, while their order was randomly changed in configurations 2 and 3. The similarities shown in these Figures in terms of error trend allow result generalisation: more than four values are enough for the considered dataset sizes and determine limited errors; on the contrary, fewer features cause a significant reduction in the model accuracy. When 13 features are considered, the error level is almost equal to zero, for at least 100 observations and all the analysed configurations.

Finally, the examination of the real road test is relevant to further prove the approach effectiveness. The previously defined LDA model classified the 24 observations reported in Table 5 in the functional class *F*. As previously said, the proposed approach seems to be particularly helpful to the analyst when the final classification is relatively complex, due to contrasting feature values and this test further confirmed this aspect. In Table 5, the values conflicting with the class *F* ranges suggested by the standards are evidenced in bold on grey background. In detail, all the observations have out-of-range values for “Movement Type” (“penetration” instead of “local”) and “Access” (no access surveyed). In general, as only two features drive the classification towards a different class (in this case superior), the LDA classifier neglected this deviation and assigned the road sections to the

class *F*. It is worthy of underlining that these values are ameliorative for class *F* and do not require the definition of corrective activities on the field or non-classification decisions. Concerning other particular cases, it is possible to evidence:

- feature 5 (V_{85}) slightly overpasses twice the related limit values;
- for feature 6 (minimum radius), 11 of 24 observations show higher values than thresholds;
- feature 7 (lane – minimum width) values are beyond standard limits in two cases.

Again, in this case, all the features assume ameliorative values for class *F*, but their number and entity are so reduced to have minimal influence on the final decision to justify the transition to the upper class. The most critical scenario regards observations 23, in which 5 of the 13 features are out of bounds: the authors believe the LDA classifier "decided" correctly in this questionable case also.

Conclusions

In this paper, the authors proposed the application of an analytical methodology for functional road classification. The classification approach shows remarkable advantages:

- it is independent of the subjective judgment of the analyst;
- it applies to existing roads also;
- the methodology is based on simple analytical tools, that also technical personnel with a limited scientific background is able to manage.

It is known that international standards base this task on the knowledge of some reference variables, conditioning the outcome. Mainly for existing roads, due to constructive and utilisation inhomogeneities, these variables sometimes drive the classification towards different functional classes, making the assignment complicated. Moreover, these variables often vary suddenly along the road in particular segments. Then, there is a need to identify a methodology measuring these inhomogeneities and to support the analyst in classifying roads beyond ideal standard conditions.

Since this method assigns a road (or its segments) to the most appropriate class by the variable deviations from the ideal conditions, the analyst is supported in the identification of the possible modifications to require to the road agency for bringing it back to the expected class. Naturally, this procedure has to be only considered as a straightforward and quick data-based support tool, representing a significant help for the decision-maker, but not his replacement.

In the present version, this approach simplifies the decisional phase by mimic traditional considerations of the analysts and extend the derived applications to similar scenarios. In future developments, it would be interesting to evolve the artificial intelligence core of such a methodology, increasing the methodology analysis possibility for dealing with some critical aspects, like reducing subjective influence in the selection of some variable values or introducing external and more complicated factors (for instance, consider evaluation of "net function" feature). Among the future developments of this research, it will also be necessary to investigate other road typologies.

Acknowledgements

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

American Association of State Highway and Transportation Officials (AASHTO) (2004). *Green Book: A Policy on Geometric Design of Highways and Streets*, 5th edition, American Association of State Highway and Transportation, Washington, DC.

Anderson, J., & Hernandez, S. (2017). Roadway classifications and the accident injury severities of heavy-vehicle drivers. *Analytic Methods in Accident Research*, *15*, 17-28. https://doi.org/10.1016/j.amar.2017.04.002

Bosurgi, G., D'Andrea, A., & Pellegrino, O. (2011). Context sensitive solutions using interval analysis. *Transport*, *26*(2), 171-177. https://doi.org/10.3846/16484142.2011.589425

Bosurgi, G., Pellegrino, O., & Sollazzo, G. (2019). Optimizing Artificial Neural Networks For The Evaluation Of Asphalt Pavement Structural Performance. *The Baltic Journal of Road and Bridge Engineering*, *14*(1), 58-79. https://doi.org/10.7250/bjrbe.2019-14.433

Charlton, S. G., & Starkey, N. J. (2017). Driving on urban roads: How we come to expect the 'correct'speed. *Accident Analysis & Prevention*, *108*, 251-260. https://doi.org/10.1016/j.aap.2017.09.010

Chen, H., Namdeo, A., & Bell, M. (2008). Classification of road traffic and roadside pollution concentrations for assessment of personal exposure. *Environmental Modelling & Software*, *23*(3), 282-287. https://doi.org/10.1016/j.envsoft.2007.04.006

D'Andrea, A., & Pellegrino, O. (2012). Application of Fuzzy Techniques for Determining the Operating Speed Based on Road Geometry. *Promet-Traffic&Transportation*, *24*(3), 203-214. https://doi.org/10.7307/ptt.v24i3.313

D'Andrea, A., Cappadona, C., La Rosa, G., & Pellegrino, O. (2014). A functional road classification with data mining techniques. *Transport*, *29*(4), 419-430. https://doi.org/10.3846/16484142.2014.984329

Dickerson, A., Peirson, J., & Vickerman, R. (2000). Road accidents and traffic flows: an econometric investigation. *Economica*, *67*(265), 101-121. https://doi.org/10.1111/1468-0335.00198

Duin, R. P., & Pękalska, E. (2012). The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, *33*(7), 826-832. https://doi.org/10.1016/j.patrec.2011.04.019

Federal Highways Administration (FHWA) (2004). *Flexibility in Highway Design*, Federal Highways Adminstration, US Department of Transportation, Washington, DC.

Forkenbrock, D. J., & Foster, N. S. (1997). Accident cost saving and highway attributes. *Transportation*, *24*(1), 79-100.

Friedrich, M. (2017). Functional Structuring of Road Networks. *Transportation Research Procedia* 25, pp. 568–581. https://doi.org/10.1016/j.trpro.2017.05.439

Giummarra, G. J. (2003). Establishment of a road classification system and geometric design and maintenance standards for low-volume roads. *Transportation research record*, *1819*(1), 132-140. https://doi.org/10.3141%2F1819a-20

Gomes, S. V. (2013). The influence of the infrastructure characteristics in urban road accidents occurrence. *Accident Analysis & Prevention*, *60*, 289-297. https://doi.org/10.1016/j.aap.2013.02.042

Goto, A., & Nakamura, H. (2016). Functionally hierarchical road classification considering the area characteristics for the performance-oriented road planning. *Transportation research procedia*, *15*, 732-748. https://doi.org/10.1016/j.trpro.2016.06.061

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: prediction, inference and data mining. *Springer-Verlag, New York*.

Huang, Y., & Guan, Y. (2015). On the linear discriminant analysis for large number of classes. *Engineering Applications of Artificial Intelligence*, *43*, 15-26. https://doi.org/10.1016/j.engappai.2015.03.006

Ministero delle Infrastrutture e dei Trasporti (2001). *Norme Funzionali e Geometriche per la Costruzione Delle Strade. Suppl. Ord. G.U. 04.01.2002 n°3.* (in Italian)

Jaarsma, C. F. (1997). Approaches for the planning of rural road networks according to sustainable land use planning. *Landscape and urban planning*, *39*(1), 47-54. https://doi.org/10.1016/S0169-2046(97)00067-4

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(1), 4-37. https://doi.org/10.1109/34.824819

Kaptein, N. A., & Claessens, F. M. M. (1998). Effects of cognitive road classification on driving behaviour: a driving simulator study. *Soesterberg: TNO Human Factors Research Institute*.

Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science*, *49*(10), 1314-1320. https://doi.org/10.1016/j.ssci.2011.04.019

Kockelman, K. M. (2001). Modeling traffic's flow-density relation: Accommodation of multiple flow regimes and traveler types. *Transportation*, *28*(4), 363-374.

Kuehnle, A., & Burghout, W. (1998). Winter road condition recognition using video image classification. *Transportation Research Record*, *1627*(1), 29-33. https://doi.org/10.3141%2F1627-05

Lamm, R., Psarianos, B., & Mailaender, T. (1999). *Highway design and traffic safety engineering handbook*.

Liu, B., Yan, L., & Wang, Z. (2017). Reclassification of urban road system: integrating three dimensions of mobility, activity and mode priority. *Transportation research procedia*, *25*, 627-638. https://doi.org/10.1016/j.trpro.2017.05.447

Mena, J. B. (2003). State of the art on automatic road extraction for GIS update: a novel classification. *Pattern recognition letters*, *24*(16), 3037-3058. https://doi.org/10.1016/S0167-8655(03)00164-8

Nowakowska, M. (2010). Logistic models in crash severity classification based on road characteristics. *Transportation Research Record*, *2148*(1), 16-26. https://doi.org/10.3141%2F2148-03

Paclık, P., Novovičová, J., Pudil, P., & Somol, P. (2000). Road sign classification using Laplace kernel classifier. *Pattern Recognition Letters*, *21*(13-14), 1165-1173. https://doi.org/10.1016/S0167-8655(00)00078-7

Pellegrino, O. (2011). Road context evaluated by means of fuzzy interval. *Cognition, Technology & Work*, *13*(1), 67-79. https://doi.org/10.1007/s10111-010-0155-2

Praticò, F. G., & Giunta, M. (2011). Speed distribution on low-volume roads in Italy: from inferences to rehabilitation design criteria. *Transportation research record*, *2203*(1), 79-84. https://doi.org/10.3141%2F2203-10

Ravi, V., Reddy, P. J., & Zimmermann, H. J. (2000). Pattern classification with principal component analysis and fuzzy rule bases. *European Journal of Operational Research*, *126*(3), 526-533. https://doi.org/10.1016/S0377-2217(99)00307-0

Rodrigues, D. S., Ribeiro, P. J. G., & da Silva Nogueira, I. C. (2015). Safety classification using GIS in decision-making process to define priority road interventions. *Journal of transport geography*, *43*, 101-110. https://doi.org/10.1016/j.jtrangeo.2015.01.007

Shanahan, J., Thomas, B., Mirmehdi, M., Martin, T., Campbell, N., & Baldwin, J. (2000). A soft computing approach to road classification. *Journal of Intelligent and Robotic Systems*, *29*(4), 349-387. https://doi.org/10.1023/A:1008158907779

Sohn, S. Y., & Lee, S. H. (2003). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. *Safety Science*, *41*(1), 1-14. https://doi.org/10.1016/S0925-7535(01)00032-7

Sollazzo, G., Fwa, T. F., & Bosurgi, G. (2017). An ANN model to correlate roughness and structural performance in asphalt pavements. *Construction and Building Materials*, *134*, 684-693. https://doi.org/10.1016/j.conbuildmat.2016.12.186

The British Standards Institution (2002). *DMRB Volume 6 Section 1 Part 1* (TD 9/93) *Road Geometry. Links. Highway link design* (includes Amendment No.1 dated February 2002)

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., & Eklund, P. (2016). A review of road extraction from remote sensing images. *Journal of traffic and transportation engineering (english edition)*, *3*(3), 271-282. https://doi.org/10.1016/j.jtte.2016.05.005

Wang, X., You, S., & Wang, L. (2017). Classifying road network patterns using multinomial logit model. *Journal of Transport Geography*, *58*, 104-112. https://doi.org/10.1016/j.jtrangeo.2016.11.013