Methodology for the Characterisation of Linear Rail Transport Infrastructures with the Machine Learning Technique and Their Application in a Hyperloop Network

José Ángel Fernández Gago¹ · Federico Collado Pérez-Seoane¹

Received: 29 December 2020 / Revised: 17 February 2021 / Accepted: 16 April 2021 / Published online: 20 May 2021
© The Author(s) 2021

Abstract The layout solution for linear rail transport infrastructure will always alternate ‘surface’ sections with ‘tunnel’ and ‘viaduct’ sections. The capital expenditure (CapEx) linked at the planning stage to this type of public asset is strongly connected to the quantity of tunnels and viaducts planned. In this context, for similar lengths, a railway line using 15% tunnels and 7% viaducts to link two cities should not have the same financial viability as one using 8% tunnels and 3% viaducts to link the same cities. The process of planning, design and construction of linear works is heavily scrutinised by public administrations in all countries, and in many cases similar standards of work are shared. Firstly, this research paper highlights the existence of hidden geometric patterns in all linear transport infrastructures worldwide. Secondly, it proposes to exploit the existence of such patterns for the benefit of planners through the computational power available today in machine learning-as-a-service (MLaaS) platforms. This article demonstrates how geometric features extracted from any succession of rectangular trapeziums in linear infrastructures can predict the quantity of kilometres in ‘surface’, ‘tunnel’ and ‘viaduct’ sections in future linear rail transport infrastructures that have not yet been built. The practical application of the proposed working methodology has made it possible to intuit

the characteristics of a future Hyperloop transport network in Europe of more than 12,000 km in length.

Keywords Infrastructures planning · Machine learning · Rail transport infrastructure · Hyperloop · Project finance · Capital expenditure

1 Introduction

The planning of transport infrastructures is the first step that any administration must take to make them a reality. In most cases, more than 10 years may elapse from the bidding of the first planning studies until the commissioning of these public goods.

A technically sound understanding of the most relevant characteristics of an infrastructure in the planning phase can be vital for its capital expenditure and future sustainability.

In the case of roads and railway lines, their layout solution will always alternate ‘surface’ sections with ‘tunnel’ and ‘viaduct’ sections. It seems reasonable to think that, among several layout alternatives of the same typology, which offer citizens similar levels of service when they come into operation, the one which minimises the construction of singular works (tunnels and viaducts) will be the one which also presents the highest financial viability.

The planning, design and construction of linear infrastructures throughout the world follows very similar procedures. In each of the above-mentioned milestones, the topography of the terrain strongly conditions the type of layout solution adopted for each section of road or railway.

In this way, a linear infrastructure in operation can always be discretised into a succession of rectangular trapeziums. The geographical coordinates of two points of a route separated by a few metres, together with their

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40864-021-00149-4>.

✉ José Ángel Fernández Gago
jose.gago@ineco.com

¹ Ingeniería y Economía del Transporte S.M.E. M.P. S.A. (Ineco), Avenida Partenón, 4, 28042 Madrid, Comunidad Autónoma de Madrid, Spain

Communicated by Zijia Wang.

elevations, allow the configuration of a trapezium to which a qualification label can be added: ‘surface’, ‘tunnel’ or ‘viaduct’. It is at this point when two possibilities open up:

- The construction of databases with real information, targeted if desired by geographical areas and/or type of infrastructure.
- The training of algorithms based on supervised learning techniques through classification methods.

The computational power currently offered by various machine learning-as-a-service (MLaaS)-type platforms makes it possible to solve everyday problems in a very efficient way that were unmanageable a decade ago.

In 2013, Elon Musk, through Hyperloop Alpha [1], modernised a concept that had been in man’s head for more than 200 years. The idea of transporting people or freights on board pressurized capsules, which move inside tubes in a quasi-vacuum, is a natural evolution of the railway, trying to provide the most advanced trains with magnetic levitation, in a very low-pressure atmospheric environment. Since then, a global force has been generated to drive this innovative project that links the best trained human capital, the most capable financial capital and the boldest governments and institutions.

Such an infrastructure could present up to four different layout solutions (see Fig. 1).

In this context, it is worth studying the predictive potential that certain geometrical patterns could have in the development of new infrastructures. Unmasking their existence could be used to the benefit of countries infrastructure planning services, in order to provide taxpayers with greater efficiency in the management of public funds linked to this type of investment projects.

Concretely, this research addresses the following questions: Is it possible for a machine learning algorithm to determine patterns that define the layout characteristics of an existing transport network? Is it possible to infer the layout characteristics that similar lines or transport networks not yet built anywhere in the world would have? At a time when public administrations have not yet endorsed a

massive production of Hyperloop network studies, this work aims to provide developers of this type of project with innovative planning methodologies.

This article firstly offers a review of the scientific literature about the layout of linear infrastructures and machine learning algorithms. Secondly, the article deals with the description of the proposed methodology. Subsequently, the structure and value of the different datasets is declared, linked to an existing high-speed rail network and a plausible European Hyperloop network of more than 12,000 km in length. With the set of declared data, the described methodology is applied, its soundness is checked, and the results obtained are offered in a tabulated way. The article concludes with a discussion of these results and the formulation of several conclusions.

2 Literature Review

This section reviews the scientific literature linked to the two key concepts on which this research article is based, namely, the layout of linear infrastructures and the technique of machine learning.

2.1 About Layout of Linear Infrastructures in Civil Engineering

There is an ancient legend that tells that the best route for a road between two points in a mountainous area is that defined by a loose donkey with freedom of movement heading to its desired destination.

This ‘liberal’ vision of the geometric design of a road transited to more interventionist positions in the eighteenth century on the part of the public authorities, which materialised through the approval of technical regulations of various kinds. In this context, it can be pointed out that in Spain in 1761 a Royal Decree was published with the aim of creating ‘straight and solid roads’.

This regulation, emanating from the public sector, was based on the previous scientific production of military

Fig. 1 Layout typologies in an Hyperloop infrastructure

engineers of that time such as the French Henri Gaultier and Bernard Forest de Belidor, the Spanish Miguel Sánchez Taramas, or the English John Müller.

In line with European tradition, the American Association of State Highway and Transportation Officials (AASHTO) emerged in the United States in 1914, with the aim of shaping the legislation and coordinating the various state policies on transport infrastructure in general. However, it was not until 1937 when a Special Committee on Administrative Design Policies within AASHTO [2] began to issue publications on geometric road design, as a common basis for the improvement of existing roads and the building of new ones.

At the end of the twentieth century, emulating AASHTO, some countries began to issue their own technical regulations for the design of their land communication routes. Thus, in 1972 the National Association of Australian State Road Authorities published the ‘Guide Policy for Geometric Design of Freeways and Expressways’ [3, 4]. In 1973, the Indian Roads Congress published the ‘Recommendations about the Alignment Survey and Geometric Design of Hill Roads’ [3, 5]. As a final example, in 1976, the Mexican Ministry of Public Works published the first edition of its ‘Manual de Proyecto Geométrico de Carreteras’ [3, 6].

Road geometric layout regulations have historically been followed by the production of technical regulations in the layout of linear railway works. In this way, the existence of a technical heritage that has been successfully transferred to different software for the layout of linear works has been consolidated globally. Different versions of them have been appearing on the market since the beginning of the 1980s (ISTRAM[®], CLIP[®], CIVIL 3D[®], ROADENG[®] among others).

It was in 2014 when the Spanish company Actisa Ltd. began marketing the software TADIL[®] [7], introducing it to the market as ‘the first software for infrastructure design through artificial intelligence’. This software, available in the field of road and rail, allows a process of self-design of linear works from boundary conditions such as the start and end point of the infrastructure, applicable geometric regulations, environmental, climatic, geotechnical restrictions, etc.

Even though each of the above-mentioned software offers results rich in detail, obtaining prior information for its use is not exempt from the consumption of technical and economic resources.

2.2 Examples of the Application of Machine Learning in Civil Engineering

Machine learning is an increasingly common tool for solving problems related to civil engineering. Over the last decade, experts from all over the world related to this field of knowledge have relied on machine learning techniques

to try to provide answers to many different questions. Different types of regression models have been used to identify significant impact factors associated with station ridership at different periods of the day [8], to the early prediction of rail contact fatigue [9] and to investigate the significant risk factors related to car traffic environment, driver characteristics and vehicle types [10, 11]. Neural networks (also deep convolutional neural networks) have been used for the prediction of the diameter of columns as a preliminary step in the design of jet grouting applications [12], for building urban accident prediction models [10] or for the detection of rail surface defects [13, 14]. Random forest models have been applied when trying to predict highway crash likelihood with traffic data collected by discrete loop detectors and web-crawl weather data [15]. Different techniques associated with deep learning have been used to establish an accurate prediction model for renewable energy resources [10, 16] and to inspect the rail surface with 3D laser cameras [13, 17].

In addition to the above references, which are more focused on technical aspects, machine learning has also been used for the study and prediction of construction phase costs of civil engineering projects [18]. Examples include the development of a cost estimation model for residential building [18, 19] and the prediction of costs associated with construction projects through back-propagation neural networks [18, 20].

2.3 Contributions with Respect to the State of the Art

As has been shown in the previous section, machine learning techniques are widely used to respond to different problems associated with civil engineering, but there is currently no similar approach to respond to the objective set out in this article: to characterise new linear infrastructures layouts planned over a territory, first discovering and then using the patterns that emerge after the analysis of massive geometric data linked to other similar infrastructures, already in service.

Over time, the increasing availability of geographic information system (GIS) and building information modeling (BIM) data linked to linear infrastructures will make it possible to generate datasets with more instances than those used in this research. Given this changing situation, another novelty with respect to previous approaches will be to include in the methodological process the use of MLaaS platforms, which provide elastic and easily adaptable solutions.

As will be exemplified below, MLaaS tools will enable the fast identification of changes in patterns, and the simple re-training of predictive models that will readjust to new realities. The low cost in economic and temporal terms, and

the scalability and computational power offered by these platforms, will make it possible in a very short time, to test a wide variety of algorithms of different types that provide a proven response to classification problems. In this way, changes in the available data may or may not be associated with changes in the patterns and, in turn, changes in patterns may or may not imply changes in the typology and characteristics of the machine learning model to be used (see Fig. 2). The methodology proposed will facilitate the identification of these changes when they occur, and the design of algorithms that adapt to the new datasets, without affecting the quality of the predictions already offered.

3 Research Methodology

The type of layout of the linear works is not unrelated to the amount of the capital consumed in their execution. For the same connection on a Hyperloop line, a layout with a length on the surface of 95%, is not the same as one with a length on the surface of 85%. The cost of building elevated, underground, or underwater sections determines the final cost of this type of work.

To undertake any work using machine learning techniques, it will first be necessary to have a dataset with historical information already observed. This data is the training basis for a machine learning algorithm, which at a later stage makes it possible to predict the layout linked to a specific section of a hypothetical line.

In this research, the original dataset contains quantitative and qualitative information linked to 14-line sections of the Spanish high-speed rail network. The 1835.84 km analysed, divided into segments of 33.33 meters on average, generates a database of 55,080 records that constitute the observable reality. The choice of this type of railway infrastructure as a working element is justified by its

similarity to a not yet built Hyperloop network. From the layout of the new transport network, which can be located anywhere in the world, its quantitative information can always be extracted. After a training process, it will be the best qualified algorithm, the one that allows one to obtain the qualitative characteristics of the new network (length in 'surface', 'underground', 'elevated' or 'underwater' mode).

The practical application of this research has consisted in the evaluation of a hypothetical Hyperloop transport network located within the European Union. Our network is 12,067.29 km long (363,385 records or segments) and consists of 28 lines connecting all European conurbations with more than two million inhabitants.

The proposed methodology revolves around the geometric figure of the rectangular trapezium and its proven predictive potential. The geographical coordinates of the points of origin and end of each section of the route of a linear transport infrastructure, together with its elevation, provide the minimum information necessary for the elevation representation of its topographic profile. The figure below (see Fig. 3) shows an approach to this idea.

In this context, each trapezium in the chain deduced from a line can be attributed various geometric magnitudes. Some of these magnitudes are related to the trapezium itself, such as the unevenness and the slope. Other magnitudes take into consideration the trapeziums before and after it, besides the trapezium itself. This link with the adjacent sections is made because of the rigidity of the civil engineering linear works layouts, where the boundary conditions imposed to the radii of curvature in plant, or to the slopes in elevation, determine the final solution of the layout. Thus, in a given segment, the unevenness of the ground is as relevant as the section in front and behind it. With this methodology, a set of values linked to each section is born, which takes into consideration the average unevenness of the terrain in the vicinity of the trapezium

Fig. 2 Graphical exemplification of the re-training and re-testing of predictive models.

Fig. 3 Floor layout of a linear infrastructure and its elevation and topographic profile of terrain

under analysis. As this characteristic has to be quantified for each proposed segment, two moving averages with predictive interest will be calculated. One that takes into account the segment in question and the 99 segments ahead of it, and another that takes into consideration the same reference segment and the 99 adjacent segments behind it. At this point it is interesting to note that the technique that uses moving averages to forecast the evolution of a certain stock market value, is also valid for determining the forecast of the evolution of the layout of a linear work. By analogy between stock market curves and topographical profiles of the terrain, financial techniques converge here with civil engineering techniques.

Once the dataset has been constructed from real data of the Spanish network, together with the proposed fields suitably calculated, a machine learning algorithm is trained using two tools that are specific to these techniques: BigML[®] [21] and KNIME[®] [22]. In this context, it should be noted that BigML[®] offers a higher training power than that offered by KNIME[®], which is why the algorithm generated by BigML[®] will be the reference, while the algorithm generated by KNIME[®] will be the verification.

At a later stage after the training and testing process, once the BigML[®] algorithm that offers the best evaluation metrics has been chosen, a first contrast test will be performed. This test consists of extracting the geometric and layout data linked to two project alternatives for a high-speed rail corridor between Santiago de Chile and Valparaiso. Since 2018, a Chinese-Chilean consortium is promoting the development of such a connection in the southern country. In this context, a Spanish engineering company (Actisa, Ltd.) has been offering an open document [23] on its website since 2020, specifying the layout details of the two corridors (North and South) between both cities.

The last phase of the research methodology consists of applying the algorithm selected as optimal through BigML[®] and KNIME[®] to a Hyperloop network in Europe.

On the lines of the network that cross the Alps or run almost entirely through the Netherlands, it is possible to intuit a priori the relative amount of each type of layout that they will present as a solution.

Figure 4 shows a summary of the steps taken to implement this research methodology.

4 Data

This section refers specifically to the process of obtaining the data necessary to train algorithms and apply them in specific situations.

4.1 Historical Dataset

The National Centre for Geographic Information in Spain [24] provides public access to the basic topographic data of its rail transport network. On 31 December 2018, the aforementioned railway network was 25,468 km long, divided into various types (funicular railway, rack railway, underground, tram, light rail and conventional train).

If it looks at the part of the network classified as conventional trains (24,863 km), it finds that only 4049 km have international gauge or UIC standard, and of these, 2827 km are in use on the date indicated. In spite of the above, the process of analysing the open information has not allowed all the kilometres of network evaluated to be made eligible, mainly due to formal defects in the data. Therefore, the final length is considered to be 1836 km (see Fig. 5). With this length as a basis, the preparation of a dataset with historical information is possible.

Dividing the above-referenced 1836 km into segments of approximately 33.33 meters, a dataset of 55,080 instances is generated, each one with 14 features. In this sense, the first feature will have an instrumental purpose and will identify each trapezium with a correlative number

Fig. 4 Steps of the methodology

Fig. 5 Baseline data for the construction of the historical dataset and picture of the 14 lines used

within the same line it is part of. The following four features, considered primary, correspond to the longitude and latitude of the initial and final points of each segment, and would be obtained through GIS software¹. The following two features, which are a derivation of the previous four primary ones, represent the value of the ground elevation in meters². The following six features, which are the result of previous calculations, refer to magnitudes in meters of horizontal length, differences in vertical elevation and percentages of slope. The last of the features, qualified as the target, refers to the engineering solution with which the

layout of the line was resolved in the historical past ('surface', 'underground' or 'elevated').

Figure 6 shows an example of the topology linked to a line belonging to a transport network and notation of defining elements of each rectangular trapezium considered in each instance.

The six calculated features, linked to each instance of the datasets, are obtained according to the mathematical equalities, equations and boundary conditions set out below.

Equation (1) describes the geographical equality that exists between point A of a trapezium and point B of the preceding trapezium.

$$(Long_{a_N}, Lat_{a_N}, Alt_{a_N}) \equiv (Long_{b_{N-1}}, Lat_{b_{N-1}}, Alt_{b_{N-1}}) \quad (1)$$

In the same way, Eq. (2) describes the geographical equality that exists between point B of a trapezium and point A of the posterior trapezium.

¹ The software used by the authors in the research process is QGIS® [25] in its Noosa version.

² The elevation in meters of each georeferenced point has been obtained using a specific software of free use called GPS Visualizer® [26]. When this software is fed online with a list of geographical coordinates belonging to a route, the web tool returns a file in CSV format with the elevation of each point.

Fig. 6 Notation of defining elements of each rectangular trapezium considered in each instance

$$(Long_{b_N}, Lat_{b_N}, Alt_{b_N}) \equiv (Long_{a_{N+1}}, Lat_{a_{N+1}}, Alt_{a_{N+1}}) \quad (2)$$

In both equalities, Long is the geographical Longitude, Lat is the geographical Latitude and Alt the Altitude. Equation (3) expresses as a result the difference in level in meters, in absolute terms, between points A and B of the same trapezium.

$$USS_N = |Alt_{b_N} - Alt_{a_N}| \quad (3)$$

where USS means ‘Unevenness Single Segment’.

Equation (4) expresses as a result the difference in level in meters, in absolute terms, between point A of the anterior trapezium and point B of the posterior trapezium. This result is incorporated into the dataset as a specific feature of the central trapezium (N).

$$TSU_N = |Alt_{b_{N+1}} - Alt_{a_{N-1}}| \quad (4)$$

where TSU means ‘Three Segment Unevenness’.

Equation (5) expresses as a result the distance in meters obtained through the Haversine formula, which exists between the projection in ground of points A and B of the same trapezium.

$$HD_N = 6,371 \times \cos^{-1}[\cos(\text{rad}(90^\circ - Lat_{a_N})) \times \cos(\text{rad}(90^\circ - Lat_{b_N})) + \sin(\text{rad}(90^\circ - Lat_{a_N})) \times \sin(\text{rad}(90^\circ - Lat_{b_N})) \times \cos(\text{rad}(Long_{a_N} - Long_{b_N}))] \quad (5)$$

where HD means ‘Haversine Distance’.

Equation (6) expresses as a result the percentage of the slope in absolute value of the line joining points A and B of the same trapezium.

$$S_N = USS_N / HD_N \quad (6)$$

where S means ‘Slope’.

The Eq. (7) offers as a solution the difference in meters between the average altitude of a section and the moving average corresponding to the altitudes of 100 consecutive trapeziums in front of the reference trapezium in the same line. As expressed, this equation for the construction of the dataset is applicable by lines. In this context, the last 100

trapeziums of the line must take into consideration the expressed boundary conditions.

$$MAF_N = \left(\frac{Alt_{a_N} + Alt_{b_N}}{2} \right) - \frac{1}{100 - K_N} \sum_{i=N}^{N+99-K_N} \left(\frac{Alt_{a_i} + Alt_{b_i}}{2} \right) \quad (7)$$

$$\text{where} \begin{cases} 1 \leq N \leq N_{MAX} \\ K_N = 0, \text{ provided that } 1 \leq N \leq N_{MAX} - 99 \\ K_N = 99 - (N_{MAX} - N) \end{cases}$$

and MAF means ‘Moving Average at the Front’.

Equation (8) offers as a solution the difference in meters between the average altitude of a section and the moving average corresponding to the altitudes of 100 consecutive trapeziums behind the reference trapezium in the same line (Moving Average at the Back, MAB). As expressed, this equation is applicable by lines. In this context, for the first 100 trapeziums of the line, the established contour conditions must be taken into consideration.

$$MAB_N = \left(\frac{Alt_{a_N} + Alt_{b_N}}{2} \right) - \frac{1}{\min(K_N, 100)} \sum_{i=1-N+K_N}^N \left(\frac{Alt_{a_i} + Alt_{b_i}}{2} \right) \quad (8)$$

$$\text{where} \begin{cases} 1 \leq N \leq N_{MAX} \\ K_N = N, \text{ provided that } 1 \leq N \leq 100 \\ K_N = 2 \times N - 100, \text{ provided that } 100 \leq N \leq N_{MAX} \end{cases},$$

and MAB means ‘Moving Average at the Back’.

Both USS, HD and S refer to intrinsic properties of a single trapezium. On the other hand, the TSU calculation features to the central trapezium a characteristic derived from the existence of three consecutive trapeziums, linking in some way the layout solution of the central section with the layout solutions adopted for the anterior and posterior sections. This creates a trend in the transition from one type of layout to another. MAF and MAB operate in the same way as TSU, trying to link to each section under analysis

the layout trend that will occur in the 99 sections before or after that section.

The data mining executed with the initial source of geographic information and the calculation of features with the help of the exposed mathematical formulation allow the creation of a historical dataset (Online Resource 1) in which the target feature ('surface', 'underground' or 'elevated') is known (see Fig. 7).

The relevant numerical information from this dataset is shown in Table 1.

4.2 Algorithm Training

Once the historical dataset has been obtained, it is time to train the machine learning algorithm. The work process is proposed in two phases. In Phase I, algorithms of different typologies will be trained and evaluated. At the end of Phase I, the optimal algorithm for this research will be selected. Phase II will consist of the validation of the optimal algorithm through a contrast test, and the application of the algorithm to a plausible European Hyperloop network.

4.2.1 Phase I: Training and Selection of the Optimal Algorithm

For the phase I: Training and Selection of the Optimal Algorithm, the flow chart is shown in Fig. 8 (a). The

process of training and selection are mainly carried out in MLaaS tools which offer powerful and interesting computational and evaluation capabilities. This research will use two different MLaaS tools. BigML[®], and in particular its OptiML[®] functionality, for the training and evaluation of the different algorithms, and KNIME[®] to test the conclusions obtained in BigML[®], as exemplified in Fig. 8 (b).

To start the training process, OptiML[®] asks the user for an error metric with which to perform its evaluations. In this case, the chosen metric is the Receiver Operating Characteristic (Area Under the Curve)-ROC (AUC).

From that moment on, OptiML[®] works in two steps. In the first step, the 'Bayesian parameter search', the algorithm performs a series of initial random partitions of the historical dataset. Eighty percent of the data from each partition (training set) will be used to train models of very different typologies (decision trees, random forest, deep neural networks, and logistic regressions), applying Bayesian parameter optimisation techniques. The evaluation of each algorithm or model is performed with its corresponding test set, e.g. if an algorithm has been generated with a training set A, it is evaluated with its test set A. The result of the first step is a set of 389 models of different typologies with promising ROC (AUC) results.

In the second step, OptiML[®] performs 'Monte Carlo cross-validation'. From N new partitions, OptiML[®] re-trains and re-evaluates the 389 referenced models, associating to each of them several ROC (AUC) results. With

Fig. 7 Aspect of the features and instances of the historical dataset

Table 1 Statistical metrics for historical dataset

Dataset	Segments Qty.	Statistical metrics	Features							
			Alt A	Alt B	HD	USS	TSU	S	MAF	MAB
HISTORICAL	55,080	Minimum value	3.70	3.70	32.00	0.00	0.00	0.00	−783.39	−338.24
		First quartile	223.48	223.38	33.00	0.40	0.90	0.01	−10.46	−12.47
		Median	660.00	659.95	33.00	0.90	2.10	0.03	1.09	−1.38
		Third quartile	809.00	809.00	33.00	1.80	4.20	0.05	11.79	9.67
		Maximum value	2163.40	2163.40	35.00	25.90	793.00	0.78	340.62	776.69
		Average value	572.82	572.79	33.16	1.53	3.92	0.05	0.57	−0.60
		Standard deviation	350.13	350.14	0.37	2.02	8.48	0.06	47.18	47.18
		Standard error	1.49	1.49	0.00	0.01	0.04	0.00	0.20	0.20

the average of these results, OptiML[®] provides the user with a ranking of the 101 best algorithms.

Up to this moment, the process has been fully automatic and has only required a few hours to be completed. After completion, OptiML[®] provides the user with an interactive tool to analyse and compare the results of the 101 models with the most commonly used evaluation metrics, not only ROC (AUC). Using this tool, the optimal model will be selected for the case presented here.

The 10 best models trained offer very similar ROC (AUC) values (between 0.980 and 0.978). For this reason, it will draw upon on another evaluation metric to determine which is the optimal algorithm: the average precision. The historical dataset is strongly biased towards ‘surface’ results, so the model must be precise in discerning whether a segment should be classified as ‘underground’ or ‘elevated’. Among the 10 best models according to ROC (AUC), the one with the highest average precision value (91.7%) is number 9, a random forest algorithm composed of 95 decision trees. This algorithm is the one chosen as optimal for this research.

The following table, shows a comparison between the algorithm chosen as optimal for this research and the best model of each trained typology, based on the average precision (see Table 2):

To finalize Phase I, the optimal algorithm is replicated in another MLaaS platform: KNIME[®]. A random forest of 95 decision trees is built, modulated with the Bayesian parameters obtained by OptiML[®]. The similarity between the resulting evaluation metrics and those obtained in BigML[®], validate the correct choice of the algorithm.

Details of the comparison between the optimal model and its replication in KNIME[®] are provided in section 5 of this article.

4.2.2 Phase II: Validation and Application of the Optimal Algorithm

Once the optimal algorithm has been determined and validated, the work process moves into Phase II: Validation and Application of the Optimal Algorithm. The optimal algorithm will be verified by using it in a contrast test and then applying it to a plausible European Hyperloop network. The work process of Phase II is shown in Fig. 9. Details of the results of this phase are provided in section 5 of this article.

4.3 Evaluable Datasets

The evaluable datasets should have the same structure as that expressed for the historical dataset (see Fig. 10).

In the case of the preliminary contrast test linked to the high-speed railway line between Santiago de Chile and Valparaíso, the BigML[®] algorithm evaluates a dataset of 2778 segments in the North Corridor and 3281 segments in the South Corridor. Table 3 shows relevant numerical information from this dataset.

With respect to the Hyperloop network of the 28 lines proposed for Europe, the algorithms trained in BigML[®] and KNIME[®] obtain the layout characteristics of a dataset of 363,385 instances or segments (Online Resource 2). Table 4 shows relevant numerical information from this dataset.

5 Results

5.1 Metrics and Qualification of the Trained Algorithms

Table 5 shows the results of the main error metrics for the predictive models, obtained from the work done in

Fig. 8 Work process—Phase I

BigML® and KNIME®. The ROC (AUC) values allow to advance a high technical solvency for the algorithms in the prediction process.

In the Table 5, the acronym ‘SUR’ refers to ‘surface’, the acronym ‘UDG’ refers to ‘underground’, and the acronym ‘ELE’ refers to ‘elevated’.

5.2 Contrast Test

As already mentioned, there is a self-drawing software in the market called TADIL®, whose owner company has published in 2020 an example of application [23] through its website. Specifically, the website offers a PDF file that documents very clearly the layout solutions linked to two high-speed rail corridor alternatives between Santiago de Chile and Valparaiso.

This example allows one, in the field of the proposed research, to verify the reliability in the application of the optimal algorithm trained with geographical data from Spain (Northern Hemisphere), when applied in very distant routes (Southern Hemisphere) (see Table 6).

Because of the absence of a high number of reference points on the PDF document linked to TADIL®, the replicas of the lines drawn up with GIS software and evaluated later with BigML® are not 100% coincidental (Online Resource 3). In spite of the above, the results offered by the algorithm in the rectified line³ are quite well adapted to the results offered by TADIL®. In both cases,

the layout of the new infrastructure presents similar orders of magnitude, confirming that the machine learning algorithm developed with the proposed methodology can be an alternative to other types of developments.

5.3 Characterisation of the European Hyperloop Network

Both the algorithm derived from BigML® and the one derived from KNIME® have allowed the characterisation of a 12,000-km proposal of this new transport network (see Table 5). As explained in section 4.2 of this research, the model trained in BigML® has been endorsed using the ‘Monte Carlo cross-validation’ method, that is, it has given optimal and consistent results for a significant number of random partitions of the data. The model trained with KNIME® has only been tested for a single partition (80–20%), so their results, while useful for validating the model building algorithms by the OptiML® functionality, cannot be considered with the same level of strength of the BigML® results.

Having understood this point, the constructive solution has been chosen as the one offered by the optimal algorithm, trained in BigML®. The solution predicted by the

³ A rectified line is the one that results from multiplying the tentative route by a factor that equals the total length of this route with the total length of the original TADIL® line.

Fig. 8 continued

Table 2 Comparison of metrics between algorithms

Algorithm	Average precision (%)	Average ROC (AUC)	Average recall (%)	Accuracy (%)	Average PHI	Average F measure
Random forest (95 decision trees) -OPTIMAL-	91.70	0.9786	75.34	94.43	0.749	0.8123
Best deep neural network	82.63	0.9370	70.78	92.38	0.656	0.7520
Best decision tree	63.91	0.9204	83.24	84.66	0.586	0.6971
Best logistic regression	54.35	0.8103	42.83	86.50	0.258	0.4503

Fig. 9 Work process—Phase II **Fig. 10** Structure of the evaluable dataset linked to the Chilean corridors and European Hyperloop network**Table 3** Statistical metrics for contrast dataset

Dataset	Segments Qty.	Statistical metrics	features							
			Alt A	Alt B	HD	USS	TSU	S	MAF	MAB
Contrast	6059	Minimum value	97.90	89.80	32.48	0.00	0.00	0.00	−388.69	−517.99
		First quartile	197.40	197.25	33.30	0.20	0.90	0.01	−18.37	−16.43
		Median	346.40	346.20	34.40	1.00	2.10	0.03	0.57	−1.13
		Third quartile	490.00	490.00	36.31	3.60	10.00	0.10	14.30	15.34
		Maximum value	1763.00	1763.00	38.95	24.00	105.40	0.71	670.10	497.80
		Average value	395.85	395.74	34.90	2.66	7.56	0.08	4.69	−4.42
		Standard deviation	257.10	257.14	1.87	3.91	11.44	0.11	100.99	101.01
		Standard error	3.30	3.30	0.02	0.05	0.15	0.00	1.30	1.30

Table 4 Statistical metrics for European Hyperloop Network dataset

Dataset	Segments Qty.	Statistical metrics	Features							
			Alt A	Alt B	HD	USS	TSU	S	MAF	MAB
European Hyperloop network	363,385	Minimum value	−9.50	−9.50	23.69	0.00	0.00	0.00	−855.22	−769.69
		First quartile	60.40	60.40	30.80	0.30	0.80	0.01	−7.20	−6.99
		Median	143.60	143.60	32.85	0.80	1.80	0.02	−0.17	−0.01
		Third quartile	309.40	309.40	35.95	1.60	4.00	0.05	6.84	7.08
		Maximum value	3358.70	3358.70	46.59	97.10	197.70	3.13	855.00	817.39
		Average value	242.24	242.24	33.21	1.49	3.83	0.04	0.01	−0.01
		Standard deviation	290.52	290.52	4.02	2.43	6.72	0.07	47.03	47.03
		Standard error	0.48	0.48	0.01	0.00	0.01	0.00	0.08	0.08

Table 5 Results of the training process

Concept			BigML®			KNIME®		
			[i] SUR	[j] UND	[k] ELE	[i] SUR	[j] UND	[k] ELE
Algorithm type			RADOM FOREST 95			RADOM FOREST 95		
Number of instances	Dataset	[A] = [B+C]	55,080			55,080		
	Training SET	[B]	44,064			44,064		
	TEST SET	[C]	11,016			11,016		
Results	True positives	[D]	9282	846	274	9290	803	276
	True negatives	[E]	1135	9893	10,390	1090	9904	10,391
	False positives	[F]	505	72	37	534	88	25
	False negatives	[G]	94	205	315	102	221	324
Metrics	Recall	[H] = [D]/[D+G]	99.00%	80.49%	46.25%	98.91%	78.42%	46.00%
	Recall (average)	[I] = [H _i +H _j +H _k]/3	75.34%			74.44%		
	Specificity	[J] = [E]/[E+F]	69.21%	99.28%	99.65%	67.12%	99.12%	99.76%
	Specificity (average)	[K] = [J _i +J _j +J _k]/3	89.38%			88.67%		
	Precision	[L] = [D]/[D+F]	94.84%	92.16%	88.10%	94.56%	90.12%	91.69%
	Precision (average)	[M] = [L _i +L _j +L _k]/3	91.70%			92.13%		
	Accuracy	[N] = [D _i +D _j +D _k]/[C]	94.43%			94.13%		
	ROC (AUC)		0.9779	0.9901	0.9679	0.9690	0.9874	0.9613

Qualification of algorithms according to the AUC value of the roc curve

RANDOM	AUC = 0.50							
BAD	0.50 < AUC < 0.60							
REGULAR	0.60 ≤ AUC < 0.75							
GOOD	0.75 ≤ AUC < 0.90							
VERY GOOD	0.90 ≤ AUC < 0.97				X	X		X
EXCELLENT	0.97 ≤ AUC ≤ 1.00	X	X				X	

KNIME® algorithm is considered in this section to evaluate the consistency of the optimal algorithm predictions.

Under these circumstances, a clear adaptation can be observed between the layout solutions calculated by both algorithms (see Fig. 11):

The numerical details obtained by each algorithm for the quantity of layout in ‘surface’, ‘underground’ and ‘elevated’ for the entire network and for each of its 28 lines can be found in Table 7.

Table 6 Contrast test results

Concept		TADIL [®]		BigML [®] algorithm			Difference % [2] – [1]
		Meters	% [1]	Tentative layout	Rectified layout		
				Meters	Meters	% [2]	
North corridor	Total length	94,803	100.00	95,791	94,803	100.00	0.00
	Surface	68,328	72.07	68,438	67,733	71.45	−0.63
	Underground	20,525	21.65	22,498	22,266	23.49	1.84
	Elevated	5950	6.28	4855	4805	5.07	−1.21
South corridor	Total length	113,171	100.00	115,662	113,171	100.00	0.00
	Surface	82,846	73.20	90,985	89,026	78.66	5.46
	Underground	25,225	22.29	21,726	21,258	18.78	−3.51
	Elevated	5100	4.51	2950	2887	2.55	−1.96

5.4 Aspect of the Network and Final Results for Optimal Algorithm

The attached figure (see Fig. 12) shows a potential Hyperloop transport network in Europe.

This network would directly serve all European population centres of more than 2 million people through 12,067.29 km. This length of network, with 28 lines and 28 stations, would have the direct capacity to connect one-third of Europe's GDP and one-fourth of its workforce. In addition, almost one-fifth of the goods moved annually in the Union could benefit from the existence of this new mode of transport.

The algorithm designed in BigML[®] and selected as optimal could not be studied during their training process segments with 'underwater' layout ('UDW'), because this type of layout does not exist in the analysed railway network. Therefore, the sections of lines '07', '19' and '20' that run over obviously maritime areas are classified as 'UDW' (see Table 8).

6 Discussion

This research proposes the application of dynamic machine learning techniques. Current MLaaS platforms allow a kind of machine learning within machine learning to be applied, whereby the best algorithm for a dataset is one in particular, and it can be changed to another if the dataset also changes in a relevant way (e.g. more observations).

It is highly reliable that the algorithm selected as optimal—a random forest of 95 decision trees—offers a

99.37% surface layout like solution for the Katowice–Warsaw line, situated in one of the most extensive plains in Europe.

It is highly reliable that the optimal algorithm offers a solution with a 98.73% surface layout for the Geraadsbergen–Amsterdam line.

It is highly reliable that the optimal algorithm offers as a solution for the 124.90 km stretch from the vicinity of Bussoleno (in Italy) to the vicinity of Le Pont de Beauvoisin (in France) a 67.89% 'underground' layout and an 8.01% 'elevated' layout. It should be noted that this particular section represents the passage of Hyperloop through the Alps, the most mountainous area in Europe, and is integrated into line 28 which would link Vercelli to Paris.

The training of algorithms through databases based on high-speed rail networks allows to obtain the characterisation of high-speed rail lines, or of Hyperloop lines (by similarity) not yet built.

Everything seems to indicate that the same methodology applied in this research, on databases linked to linear high-speed railway infrastructures, could be applied to databases built on other subjects: regional or suburban railway networks, motorways, conventional roads, oil pipelines, or high-voltage electricity networks, for example.

7 Conclusion

It is shown that the rectangular trapezium as a unit of characterisation of a linear infrastructure, enriched by only a few topographical features, holds enormous predictive potential. Its use, combined with machine learning

Fig. 11 Comparative results of the algorithms

Table 7 Characterisation results

European Hyperloop network				Solution proposed by BigML [®]			Validation by KNIME [®]			
Line number, description and code				Length [km]	SUR [%]	UDG [%]	ELE [%]	SUR [%]	UDG [%]	ELE [%]
01	LISBON–MADRID	LISMAD	537.66	92.42	2.10	5.48	96.22	1.98	1.80	
02	VALENCIA–MADRID	VALMAD	360.02	94.55	2.89	2.56	95.21	2.97	1.82	
03	MADRID–BARCELONA	MADBAR	544.65	82.95	11.04	6.01	84.24	11.77	3.99	
04	MADRID–PARIS	MADPAR	1108.71	81.39	14.50	4.11	87.66	9.96	2.38	
05	MARSEILLE–PARIS	MARPAR	787.92	89.12	7.56	3.33	92.57	4.70	2.73	
06	PARIS–GERAARDSBERGEN	PARGER	247.53	81.57	15.32	3.11	91.83	5.60	2.57	
07	GERAARDSBERGEN–LONDON	GERLON	317.56	92.45	5.78	1.77	94.83	4.04	1.13	
08	LONDON–STOKE-ON-TRENT	LONSOT	238.33	90.80	6.99	2.21	94.35	3.37	2.28	
09	GERAADSBERGEN–AMSTERDAM	GERAMS	198.93	98.73	1.23	0.03	98.60	1.40	0.00	
10	HAMBURG–AMSTERDAM	HAMAMS	391.34	97.79	0.62	1.59	99.21	0.50	0.29	
11	BERLIN–HAMBURG	BERHAM	262.40	95.68	4.02	0.30	97.34	2.59	0.07	
12	PRAGUE–BERLIN	PRABER	306.67	91.94	5.80	2.26	91.84	4.74	3.42	
13	VIENNA–PRAGUE	VIEPRA	280.89	92.51	4.91	2.58	93.67	4.52	1.81	
14	VIENNA–MUNICH	VIEMUN	364.96	93.32	5.20	1.48	93.44	5.39	1.16	
15	MUNICH–STUTTGART	MUNSTU	236.62	85.36	3.09	11.55	94.50	3.49	2.01	
16	STUTTGART–FRANKFURT	STUFRA	169.32	87.65	9.78	2.56	93.76	4.43	1.81	
17	FRANKFURT–DORTMUND	FRADOR	240.14	81.00	9.47	9.53	85.14	8.38	6.47	
18	DORTMUND–GERAARDSBERGEN	DORGER	274.62	95.74	3.78	0.47	97.20	2.26	0.54	
19	HAMBURG–COPENHAGEN	HAMCOP	464.36	92.61	7.34	0.04	95.04	4.87	0.10	
20	COPENHAGEN–ESTOCKHOLM	COPEST	595.13	91.10	5.82	3.08	95.10	2.17	2.73	
21	VIENNA–KATOWICE	VIEKAT	304.38	98.61	0.29	1.10	98.44	0.23	1.33	
22	KATOWICE–WARSAW	KATWAR	260.25	99.37	0.04	0.58	98.50	0.04	1.46	
23	BUDAPEST–VIENNA	BUDVIE	229.06	95.62	3.73	0.66	97.47	2.18	0.35	
24	BUCHAREST–BUDAPEST	BUCBUD	746.49	91.21	6.62	2.17	92.79	5.92	1.30	
25	ATHENS–BUCHAREST	ATHBUC	1079.47	76.19	21.24	2.57	76.38	21.01	2.61	
26	NAPLES–ROME	NAPROM	198.12	81.34	14.13	4.53	84.48	13.05	2.47	
27	ROME–VERCELLI	ROMVER	561.18	70.97	22.60	6.43	74.39	21.42	4.20	
28	VERCELLI–PARIS	VERPAR	760.58	83.12	14.14	2.75	81.14	16.46	2.41	
Total			12,067.29	87.60	9.26	3.14	89.98	7.89	2.13

Fig. 12 Hyperloop network proposed for Europe (28 lines)

methodologies, based on supervised learning, yields brilliant results in predicting the layout solution in unknown segment observations. In the case presented, a dataset elaborated from public and open data, provided with a few calculated features, has been able to train two predictive algorithms with very good and excellent evaluation metrics.

The current offer of MLaaS platforms and open source software provides the average user with enormous storage and calculation capacities at a low or zero cost, in addition to numerous functionalities and evaluation and validation tools. This allows the methodology outlined here to be easily accessed and quickly implemented. The algorithm trained in BigML® and validated with KNIME®, has been able to provide in a few seconds of characteristics to a plausible Hyperloop network of 12,067.29 km.

The robustness of the algorithms obtained, the low cost involved in obtaining them, and the rapid application on new routes yet to be constructed (easily outlined and characterized with GIS tools) make this methodology a more than interesting alternative when it comes to preparing and enriching planning studies, preliminary projects or audits of construction projects. The testing and discrimination of many different layout alternatives, in a reliable, quick and inexpensive way, will make it possible

to focus resources only on those alternatives that are really suitable for the intended objectives.

Although, as demonstrated, the main topographical and geometrical characteristics of a section have been more than sufficient to build solid and consistent predictive models, it should not be forgotten that supervised learning is a living method that can always be improved. In this sense, is likely that, as far as this methodology is concerned, improvement will occur as a result of the enrichment of the datasets with other features which, a priori, are already perceived as having predictive potential. These could include the geological characteristics of the terrain, where the first obstacle to overcome will be to obtain information that is not currently available as open data. Other features to be explored will be ‘subjective’, exclusive to each type of linear work.

Being able to incorporate the specific conditioning factors for each type of linear work that lead to one constructive solution or another will undoubtedly be enriched when it comes to improving the algorithms and providing them with greater precision. Thus, the use of the algorithms in real projects and the collaboration with professionals in the sector will be the best way to identify these subjective features and relate them to the cornerstone of this methodology: the rectangular trapezium.

Table 8 Final results

European Hyperloop network			Solution proposed							
Line number and description		Length [km]	SUR		UDG		ELE		UDW	
			[km]	[%]	[km]	[%]	[km]	[%]	[km]	[%]
01	LISBON–MADRID	537.66	496.91	92.42	11.28	2.10	29.47	5.48	0.00	0.00
02	VALENCIA–MADRID	360.02	340.39	94.55	10.40	2.89	9.23	2.56	0.00	0.00
03	MADRID–BARCELONA	544.65	451.77	82.95	60.14	11.04	32.74	6.01	0.00	0.00
04	MADRID–PARIS	1108.71	902.38	81.39	160.76	14.50	45.57	4.11	0.00	0.00
05	MARSEILLE–PARIS	787.92	702.16	89.12	59.55	7.56	26.21	3.33	0.00	0.00
06	PARIS–GERAARDSBERGEN	247.53	201.91	81.57	37.92	15.32	7.70	3.11	0.00	0.00
07	GERAARDSBERGEN–LONDON	317.56	258.98	81.55	18.36	5.78	5.55	1.75	34.67	10.92
08	LONDON–STOKE-ON-TRENT	238.33	216.39	90.80	16.67	6.99	5.26	2.21	0.00	0.00
09	GERAADSBERGEN–AMSTERDAM	198.93	196.41	98.73	2.45	1.23	0.07	0.03	0.00	0.00
10	HAMBURG–AMSTERDAM	391.34	382.71	97.79	2.42	0.62	6.22	1.59	0.00	0.00
11	BERLIN–HAMBURG	262.40	251.07	95.68	10.54	4.02	0.80	0.30	0.00	0.00
12	PRAGUE–BERLIN	306.67	281.95	91.94	17.79	5.80	6.93	2.26	0.00	0.00
13	VIENNA–PRAGUE	280.89	259.84	92.51	13.79	4.91	7.26	2.58	0.00	0.00
14	VIENNA–MUNICH	364.96	340.59	93.32	18.97	5.20	5.40	1.48	0.00	0.00
15	MUNICH–STUTTGART	236.62	201.97	85.36	7.31	3.09	27.34	11.55	0.00	0.00
16	STUTTGART–FRANKFURT	169.32	148.41	87.65	16.57	9.78	4.34	2.56	0.00	0.00
17	FRANKFURT–DORTMUND	240.14	194.50	81.00	22.75	9.47	22.89	9.53	0.00	0.00
18	DORTMUND–GERAARDSBERGEN	274.62	262.93	95.74	10.39	3.78	1.30	0.47	0.00	0.00
19	HAMBURG–COPENHAGEN	464.36	407.61	87.78	34.10	7.34	0.21	0.04	22.45	4.83
20	COPENHAGEN–ESTOCKHOLM	595.13	537.68	90.35	34.64	5.82	18.33	3.08	4.48	0.75
21	VIENNA–KATOWICE	304.38	300.16	98.61	0.87	0.29	3.36	1.10	0.00	0.00
22	KATOWICE–WARSAW	260.25	258.62	99.37	0.11	0.04	1.52	0.58	0.00	0.00
23	BUDAPEST–VIENNA	229.06	219.02	95.62	8.53	3.73	1.50	0.66	0.00	0.00
24	BUCHAREST–BUDAPEST	746.49	680.90	91.21	49.41	6.62	16.18	2.17	0.00	0.00
25	ATHENS–BUCHAREST	1079.47	822.49	76.19	229.28	21.24	27.70	2.57	0.00	0.00
26	NAPLES–ROME	198.12	161.15	81.34	27.99	14.13	8.98	4.53	0.00	0.00
27	ROME–VERCELLI	561.18	398.24	70.97	126.85	22.60	36.08	6.43	0.00	0.00
28	VERCELLI–PARIS	760.58	632.16	83.12	107.53	14.14	20.90	2.75	0.00	0.00
Total		12,067.29	10,509.31	87.09	1117.35	9.26	379.03	3.14	61.60	0.51

Acknowledgements The authors would like to extend their thanks to Victoria Laura Barrio Cagigal, from the Department of Chemical and Environmental Engineering of the University of the Basque Country (Spain), for her support and guidance in the doctoral thesis of which this research is part; to Cristina Delgado Quirarte, Carolina Sanz Pecharromán, Carlos Sanchez Martín, and Miriam Yepes Pérez from the Area of Economics and Transport Policy of Ineco (Spain), for the critical review of the article; to Félix Guijarro de Vicente, from the Area of Transport Infrastructures of Deloitte Spain, for the help in the literature review; and to Kieran Alec O'Regan and Stephanie McDermott Flores for the proofreading of the article in its English translation.

Funding The authors declare that no external funding has been necessary for the materialisation of this article.

Availability of data and materials The authors declare that they are in possession of all data and materials mentioned in this article. This

data has been deposited as Supplementary Information in the repositories enabled for this purpose (online resources).

Declarations

Conflict of interest The authors declare that they have no conflict of interest with any organisation mentioned in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted

use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Musk, E (2013) Hyperloop Alpha. SpaceX
2. AASHTO (2001) A policy on geometric desing of highways and streets. pp. ix (Preface)
3. LópezArchila MC, EcheverryArciniegas CA (1999) Manual de DiseñoGeométrico para Carreteras. Ciencia e IngenieríaNeogranadina 8(145–156):155
4. National Association of Australian State Road Authorities (1972) Guide Policy for Geometric Design o Freeways and Expressways
5. The Indian Road Congress (1973) Recommendations about the Alignment Survey and Geometric Design of Hill Roads
6. Secretaria de Comunicaciones y Transportes, Dirección General de Servicios Técnicos de los Estados Unidos Mexicanos (1976) Manual de Proyecto Geométrico de Carreteras
7. Actisa Development Team (2014) Tadil software. Spain
8. Zhu Y, Chen F, Wang Z (2019) Spatio-temporal analysis of rail station ridership determinants in the built environment. *Transportation* 46:2269–2289
9. Sysyn M, Gerber U, Nabochenko O (2019) Prediction of rail contact fatigue on crossings using image processing and machine learning methods. *Urban Rail Transit* 5:123–132
10. Ghasedi M, Sarfjoo M, Bargegol I (2021) Prediction and analysis of the severity and number of suburban accidents using logit model, factor analysis and machine learning: a case study in a developing country. *SN Appl. Sci.* 3:13
11. Yan X, Radwan E, Abdel-Aty M (2005) Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accid Anal Prev* 37(6):983–995
12. Ochmanski M, Modoni G, Bzowka J (2015) Prediction of the diameter of jet grouting columns with artificial neural networks. *Soils Found* 55(2):425–436
13. Hadj-Mabrouk H (2019) Contribution of artificial intelligence to risk assessment of railway accidents. *Urban Rail Transit* 5:104–122
14. Faghih-Roohi S, Hajizadeh S, Núñez A, Babuska R, De Schutter B (2016) Deep convolutional neural networks for detection of rail surface defects. In: *International joint conference on neural networks (IJCNN)*, Vancouver, BC, 2016, pp 2584–2589
15. You J, Wang J, Guo J (2017) Real-time crash prediction on freeways using data mining and emerging techniques. *J Mod Transport* 25:116–123
16. Shamshirband S, Rabczuk T, Chau K (2019) A survey of deep learning techniques: application in wind and solar energy resources
17. Santur Y, Karaköse M, Akin E (2017) A new rail inspection method based on deep learning using laser cameras. In: *International artificial intelligence and data processing symposium (IDAP)*, Malatya, pp 1–6 (2017)
18. TayefehHashemi S, Ebadati OM, Kaur H (2020) Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Appl Sci* 2:1703
19. Amin M (2017) Development of cost estimation model for residential building. *Int J CivStructEng Res* 5(1):1–4
20. Jiang Q (2019) Estimation of construction project building cost by back-propagation neural network. *J Eng Des Technol* 18(3):601–609
21. BigML (2011) BigML, Inc. Corvallis, Oregon, USA
22. Berthold M (2006) KNIME (Konstanz Information Miner). Universität Konstanz
23. Actisa SL (2020) High speed railway line from Santiago de Chile to Valparaiso and Viña del Mar. Northern Corridor or Southern Corridor?
24. Centro Nacional de Información Geográfica, file: RT-ESPANA-PORMODOS.ZIP. 15 November 2018
25. QGIS Development Team (2002) QGIS Noosa. Windows—GNU/Linux—Mac OS X—BSD—Android. USA: Open Source Geospatial Foundation
26. Schneider A (2002) GPS visualizer. USA: GPS Visualizer