Machine learning techniques applied to construction: A hybrid bibliometric analysis of advances and future directions

Jose Garcia[a,*], Gabriel Villavicencio[a], Francisco Altimiras[b], Broderick Crawford[c], Ricardo Soto[c], Vinicius Minatogawa[a], Matheus Franco[d], David Martínez-Muñoz[e], Víctor Yepes[e]

[a] *Escuela de Ingeniería de Construcción y Transporte, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362804, Chile. jose.garcia@pucv.cl; gabriel.villavicencio@pucv.cl; vinicius.minatogawa@pucv.cl*

[b] *Facultad de Ingeniería y Negocios, Universidad de las Américas, Santiago 7500975, Chile.*

[c] *Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile. broderick.crawford@pucv.cl; ricardo.soto@pucv.cl*

[d] *Geosciences Institute, Department of Science and Technology Policy, University of Campinas, Campinas 13083-855, SP, Brazil. matheusfranco400@gmail.com*

[e] *Institute of Concrete Science and Technology (ICITECH), Universitat Politècnica de València, 46022 València, Spain. vyepesp@cst.upv.es; damarmu1@cam.upv.es*

Abstract

Complex industrial problems coupled with the availability of a more robust computing infrastructure present many challenges and opportunities for machine learning (ML) in the construction industry. This paper reviews the ML techniques applied to the construction industry, mainly to identify areas of application and future projection in this industry. Studies from 2015 to 2022 were analyzed to assess the latest applications of ML techniques

*Corresponding author: Jose García, e-mail: jose.garcia@pucv.cl

in construction. A methodology was proposed that automatically identifies topics through the analysis of abstracts using the Bidirectional Encoder Representations from Transformers technique to select main topics manually subsequently. Relevant categories of machine learning applications in construction were identified and analyzed, including applications in concrete technology, retaining wall design, pavement engineering, tunneling, and construction management. Multiple techniques were discussed, including various supervised, deep, and evolutionary ML algorithms. This review study provides future guidelines to researchers regarding ML applications in construction.
Keywords: Machine Learning, BERT, Construction, Concretes, Retaining Walls, Tunnels, Pavements, Construction Management.

1. Introduction

Nowadays, machine learning (ML) techniques are widely applied to multiple tasks and challenges. Herewith, the availability of a more powerful computing infrastructure provides the necessary tools for implementing advanced ML techniques to solve complex industrial problems. In this way, we can improve decision-making in industries, increasing their sustainability and productivity. The fourth industrial revolution (Industry 4.0) is changing all the industries in different aspects [1]. One of the industries that is expected to benefit significantly from ML implementation is the construction industry. Multiple articles raise the need to automate construction to improve the way this industry works, including the need to improve the construction supply chains,[1, 2, 3]. In this work, a review of ML applications for smart construction was developed. Articles published in recent years that consider the concepts of ML and construction were analyzed. The initial database obtained was more than 5000 articles, so it was decided to use a methodology based on topic modeling, Section 2, to make an initial grouping of the most interesting topics to later delve into each of these.

The objective of topic modeling is to group documents and words that have similar meanings. It is widely used in a variety of domains, including natural language processing (NLP) and information retrieval (IR). It uses unsupervised ML algorithms to extract topics from document collections. There are several topic modeling approaches available, for example, Probabilistic Latent Semantic Analysis (PLSA), [4], Latent Dirichlet Assignment (LDA), [5]. Another interesting method, nonnegative matrix factorization (NMF), is an unsupervised technique for reducing the dimension of nonnegative matrices, [6], which has been widely utilized to deduce underlying links between texts and to find latent themes [7]. Although these approaches do not require labels to operate, they require specifying the number of categories to perform the grouping. However, a growing number of topic modeling systems are based on LDA and NMF, although they require considerable work in hyperparameter tuning to generate meaningful topics.

In general, the methods outlined above have some drawbacks. One of these limitations is that they ignore semantic relationships between words when using bag-of-words representations. These representations do not consider the context of words in a sentence, which may make it difficult for them to display documents correctly. This article uses a semi-automatic method to carry out a bibliographic analysis. In the first stage, a search is carried out on the Scopus database, and a set of abstracts related to the search is obtained. These abstracts are modeled across topics using BERTopic, [8]. This method has been used to model topics and provides a better contextual perspective than previous methods.

Based on the latter, this article uses a semi-automatic method to carry out bibliographic searches. In the first stage, a search is carried out on the Scopus database, and a set of abstracts related to the search is obtained. These abstracts are modeled across topics using Bidirectional Encoder Representations from Transformers topis (BERTopic), [8]. Subsequently, the main topics are validated for consistency by an expert to select the relevant topics. Using the relevant terms of each of these topics, new Scopus queries are generated to finally carry out a traditional bibliographic analysis with the result of said queries and a clustering analysis based on bigrams.

This study aims to determine the latest applications of ML tools in the construction industry through a semi-automated method that integrates ML techniques and expert knowledge. The main objective is to determine in what areas and what ML techniques have been developed and implemented to solve problems in the construction industry. This state-of-the-art review includes articles from the last seven years, where the search focused on applications of machine learning in construction areas.

A brief summary of the structure of the content of the following sections: Throughout the Section 2, the procedure used to carry out the bibliographic analysis is explained. In Sections 3 and 4, the bibliographical analysis of the selected articles is detailed. First, The BERT topics are selected, and a general scientometric analysis is carried out in 3. Later, for each selected topic, a bigram analysis is carried out in Section 4, plus the traditional bibliographic analysis. In Section 5, future directions are developed and finally in Section 6, develop the conclusions and the next steps.

2. Methodology

This section describes the proposed methodology. First, an overview of the method is given to later describe each of the stages. In Figure 1, the detail of the methodology used to carry out the review is shown. In the first stage, a search on Scopus is carried out using the concepts of "Machine Learning" and "Construction." Later these are filtered for articles in English retrieved in the last seven years. These results are analyzed using the methodology developed in section 2.1. Each of the topics obtained is validated by experts in the area who determine validity, evaluating the coherence between the main terms obtained. For the topics that pass the expert criteria for each of them, a search is performed again based on the attributes obtained in the topic. With this new search, the selection of articles is carried out according to expert criteria again, and for this selected set, a bigram analysis is carried out on the one hand, which is detailed in the section 2.2, in addition to a traditional review that implies reading of the article and extraction of the main characteristics is realized.

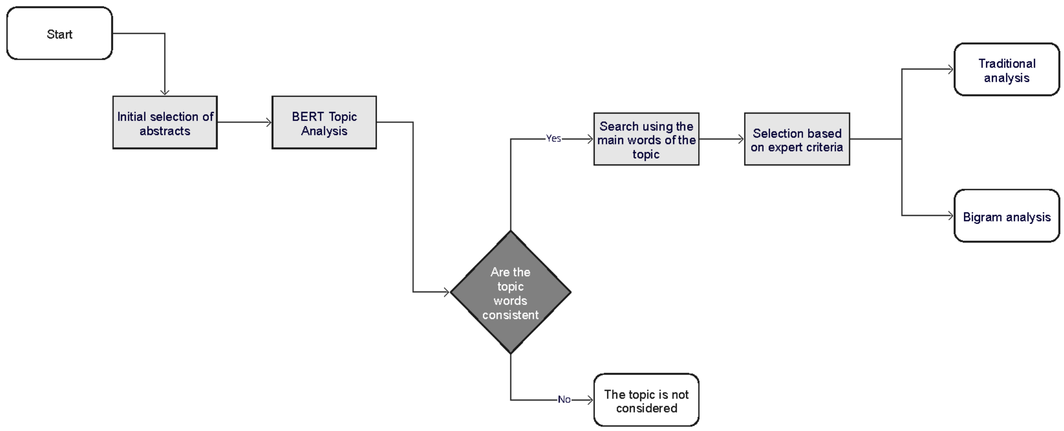

Figure 1: Flowchart of the semi-automated literature review methodology.

2.1. Topic analysis

The selection of topics is made by analyzing the abstracts of all the retrieved documents. In order to make the selection of these, the process consists of three stages. In the first stage, a numerical and contextual representation of each of the terms is generated. To perform this representation, a pre-trained model of a neural network, Bidirectional Encoder Representations from Transformers (BERT), [9], was used. This embedding is very powerful for language comprehension as it captures the semantic relationships between words.

Once the words are embedded in a vector, in order to analyze and group the concepts in a meaningful way, a dimensionality reduction process must be carried out. Several techniques allow the reduction process to be carried out. In this case, as the reduction process requires preserving global and local components of the data space, the uniform manifold approximation and projection for dimension reduction (UMAP) technique, [10], is used. This algorithm uses the concept of simplex obtained from algebraic topology in addition to manifold theory to be able to develop dimensionality reduction. Once the dimensionality reduction has been carried out, it is necessary to perform the groupings in order to find the similarities that allow us to obtain the topics. Following on from the work done in [8], at this stage, (HDBSCAN), [11] is used to generate the topics.

2.2. Bigram analysis

A bigram is a sequence of two adjacent elements of a chain of tokens; in our specific case, they correspond to words. The objective is to carry out a statistical analysis of the frequency distribution of these bigrams in the different analyzed abstracts. To perform the analysis of each of the topics identified by BERT, the R-bibliometrix [12] package was used. Specifically, four visualizations were used. The first corresponds to the Treemap. This aims to identify the frequency of the main bigrams in each of the topics. Subsequently, the thematic map is used; this graph uses the concept of density (internal associations) together with that of centrality (external associations), [13, 14].

This visualization is divided into four quadrants; quadrant 1 identifies high density and high centrality. And the main topics that appear in the articles are considered. The second quadrant corresponds to high centrality and low density, which are basic and transversal topics. Quadrant 3 corresponds to high density and low centrality topics and is related to the niche or specialized topics. Finally, the fourth quadrant corresponds to emerging or poorly developed topics.

Finally, the last two visualizations correspond to conceptual maps and dendrograms. The conceptual structure visualization creates a conceptual structure map of each of the topics obtained by BERT. Specifically, multidimensional scaling (MDS) is performed on terms extracted from the abstracts of the documents. In addition to analyzing the relationship between the terms in a hierarchical way, the conceptual structure is also displayed through a dendrogram.

3. BERT Topics and General bibliometrics

This section details the results obtained from the analysis of topics, and later with the selected articles of each topic, a general analysis of the journals, authors, and the thematic evolution of the main concepts is carried out. According to the methodology detailed in the section 2; The analysis begins with generating topics using BERT to later select the most important topics according to expert criteria. Figure 2 shows the selection made for the topics. In particular, five themes are selected. Concrete, retaining walls, pavements, tunnels, and construction management. With the keywords obtained in each topic, a manual selection of the articles to be analyzed was made. Figure 3 shows the main journals analyzed. Automation in construction, construction and building materials, and engineering with computers were the main sources of articles. Figure 4 shows an analysis of the contribution by country as well as an analysis of author networks. In the case of countries, in the upper right diagram of Figure, the country with the greatest contribution corresponds to the USA with a frequency of 91 author appearances, followed by China with 57 and further down Iran with 30, South Korea with 20 and Canada with 17. Additionally, the visualization represents a collaboration between countries, in which if the frequency of authors between countries with articles in common exceeds the value 5, a connection is drawn between them. At this point, the collaboration between the USA and China, the USA and Iran, and Spain and Chile stands out.

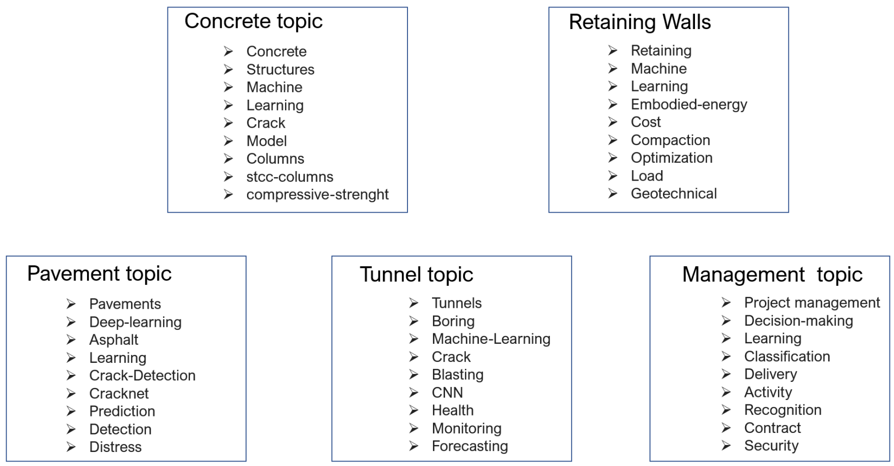

Figure 2: BERT topics selection results.

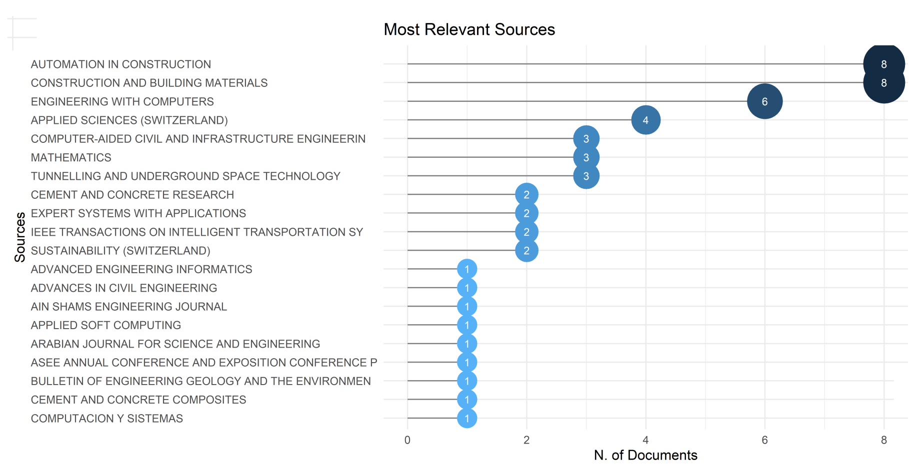

Figure 3: Most relevant sources.

The upper left diagram of Figure 4 shows a network analysis of the authors. There are seven main groups in the diagram. Where the most significant collaborative group is highlighted in red, the author's network, Zhang, A; from the USA; Fei, Y, from the USA; Chen, C, from the USA; Liu, Y; from the USA; and, Li, B, from China. The lower diagram highlights the publications with important impact factors in the red group between 2017 and 2020. Their publication area is related to the detection of cracks in the asphalt pavement area through the use of deep learning techniques. Another collaborative network of authors is the one led by Koopialipoor, M; of Iran, which considers collaborations with the USA and Vietnam. In the lower diagram, they have had a significant number of publications in 2019 and 2020, in addition to a significant number of citations. The publication line is related to applying ML techniques such as deep learning to tunnels. The inspection and detection of cracks in tunnels have been addressed by Doulamis A; Protopadakis E ; Doulamis, N, and other collaborators. They stand out with publications and important impact factors in 2015 and 2017.

Figure 5 depicts a diagram for assessing the topic evolution of the articles under consideration. Combining performance analysis and scientific mapping, this method identifies and visualizes conceptual subdomains, [15]. Co-word analysis is utilized in a longitudinal context to identify the many study subjects covered during a specific time period. The Figure shows that machine learning and deep learning topics appear strongly in the first window of time. The above is quite natural since the review is focused on ML techniques. It is also observed that these concepts are maintained in the different time windows. Another interesting point in the first time window is crack detection. We see that already at this time, this concept was already addressed significantly through ML techniques. When we move to the second time window, we see that deep learning techniques are strongly related to Crack Detection applications, the construction industry and management, and health monitoring. Finally, two additional concepts appear in the last window of time; ML and deep learning techniques have been focused on and strongly converged into prediction models. On the other hand, a new area of application related to pavement conditions appears.

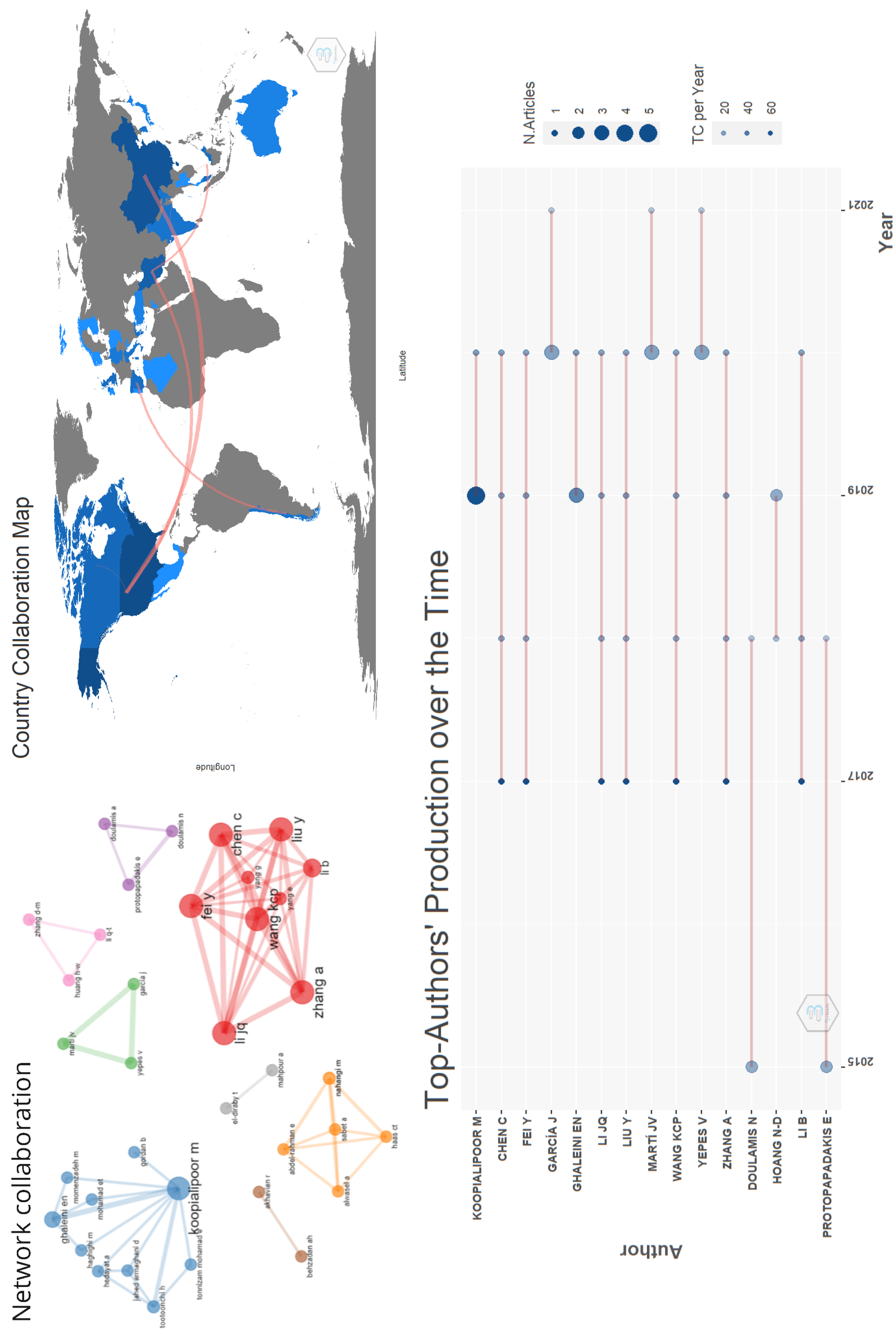

Figure 4: Country and author's collaboration map.

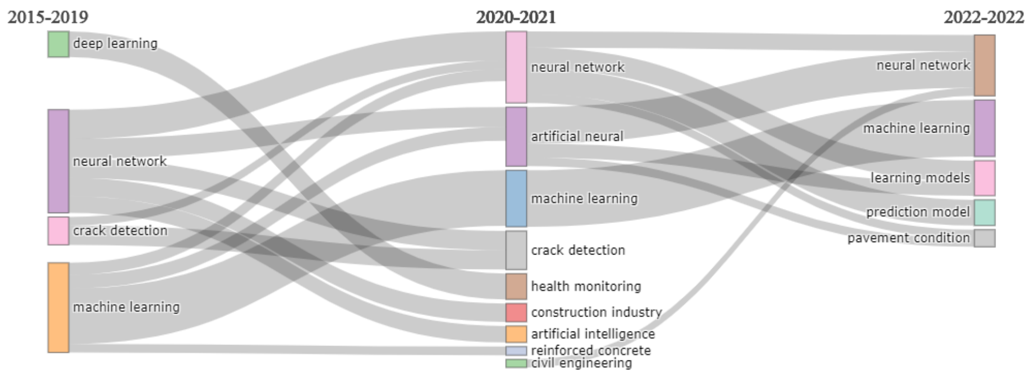

Figure 5: Thematic evolution map.

4. Bigram and traditional results

This section details the analysis for each of the five topics obtained in the previous section. The analysis, according to the methodology proposed in the section 2 consists of two parts. First, an analysis of bigrams is carried out, from which groups of related words are extracted to obtain an overview of the topic. Then a traditional analysis of the selected articles on each topic is developed.

4.1. Concrete Structures

Concrete is the most widely used artificial material in buildings, pavements, and retaining walls. Concrete technology deals with the study of the properties of concrete and its practical applications. Concrete is used to construct foundations, columns, beams, slabs, and other load-bearing elements in building construction. The production of concrete requires large quantities of coarse and fine aggregates. To preserve natural resources, it is of the utmost importance to pay close attention to the use of waste materials and by-products in concrete mixes. For this purpose, predictive models based on ML have been used to determine the properties of concrete in order to save time, cost, and energy.

4.1.1. Bigram document analysis

When performing the bigram analysis and structuring the most relevant concepts, we see in the upper left graph in Figure 6 that the main concepts related to the artificial intelligence techniques appear: artificial neural networks (ANN), and support vector machines. When observing the concepts related to concrete techniques, reinforced concrete, concrete mix, retaining walls, and compressive strength, appear as the main concepts.

When the co-words analysis is applied, the concepts are later grouped.

The result can be seen in the lower right Figure. In this Figure, it is observed that the concrete and reinforced concrete structures are related to prediction models. On the other hand, the study of compressive strength is in conjunction with neural networks. The part of crack detection and the concrete surface appears strongly related to convolutional neural networks. When the bigrams are grouped further, three clusters mainly stand out. These results are shown in the two figures below. In the lower-left Figure, we see that there is a cluster that is related to the structural, sustainable design and its optimization. On the other hand, there is a whole group related to crack detection, structure health monitoring, and convolutional neural networks. Finally, a large group relates a significant number of machine learning techniques to concrete design and production variables such as compressive strength of reinforced concrete, mixture proportions, and compressive strength.

Figure 6: Tree, Thematic, conceptual and dendrogram maps applied a concrete data set.

4.1.2. Traditional analysis

In Table 1, a summary of the different articles selected for Concrete structures is shown. The table highlights the use of ANN, RF, and SVM techniques. On the other hand, applications for monitoring structures, crack and prediction of concrete properties appear more frequently. Following the groups found in the bigram analysis, the main group related to the design and production of concrete was found. Concrete is the most widely used artificial material in buildings, pavements, and dams. Concrete production requires large amounts of coarse and fine aggregates. To preserve natural resources, much attention has been paid to the use of waste materials and by-products in concrete mixes. The fresh and hardened properties of concrete mixes containing waste foundry sand (WFS) residues as a partial or total replacement for fine aggregate have been the focus of several recent studies. To manufacture molds and cores, the ferrous (iron and steel) and nonferrous (copper, aluminum, and brass) metal-casting industries discard WFS. Using predictive models for concrete properties can save time and energy and provide information on scheduling activities such as frame removal. In [16], the M5P (decision tree) algorithm was used to model the strength, modulus of elasticity, strength, and tensile strength at the break of these

Reference	Application	Techniques	Results	Data
[17]	The effect of adding rubber to concrete on steel corrosion and concrete deterioration is studied.	1. Bayesian Ridge 2. K-nearest neighbors 3. RF	1) R2 = 0.87; MAE=2.84; MSE = 12.73 2) R2 = 0.97; MAE=1.26; MSE = 2.88 3) R2 = 0.97; MAE=1.22; MSE = 2.82	Data derived from ultrasonic tests applied to samples subjected to accelerated corrosion methods.
[18]	Estimating the axial strength of steel tube confined concrete	1. SVR-GWO 2. ANN 3. SVR 4. RF	1) R2 = 0.99, MAPE= 7.0 , RMSE = 209.1 2) R2 = 0.98, MAPE= 17.3 , RMSE = 337.8 3) R2 = 0.88, MAPE = 19.9, RMSE = 857.1 4) R2 = 0.98, MAPE = 10.3, RMSE = 337.5	136 samples of STCC columns infilled with various strength concrete were collected to develop and evaluate the proposed model.
[19]	Monitoring and tracking of changes in the structural integrity and durability of concrete structures	1. Single and Dual Convolutional neural network	1) AUC: Only the charts are reported 2) Accuracy: Dual model trained with the SNR of -20 dB exhibited the best accuracy. Over 80% in almost all experiments.	1) Acoustic Emission (AE) signals emitted from compressive failure of concrete specimens; 2) noise signals by man-made activities; and 3) AE signals acquired during the physical model experiments
[20]	Structural Health Monitoring to build reliable automatic damage-assessment procedures.	1) Extreme ML (ELM)-stress 2) Extreme ML (ELM)-vibration	1) RMSE= 0.07; R2 = 0.94; MaxErr = 0.16 2) RMSE= 0.42; R2 = -0.33; MaxErr = 0.81	Stress and vibration dataset
[21]	durability design and service life prediction of concrete structures in civil engineering projects.	1) ANN 2) MARS-L 3) Gaussian Process Regression 4) Hybrid ensemble model	1) R2 = 0.87, MAPE= 18.4 , RMSE = 0.075 2) R2 = 0.88, MAPE= 17.1 , RMSE = 0.072 3) R2 = 0.88, MAPE= 17.1, RMSE = 0.071 4) R2 = 0.89, MAPE= 16.3, RMSE = 0.070	1030 records have been compiled from the machine-learning repository of the University of California, Irvine.
[22]	Prediction of the surface chloride the concentration of concrete, for durability design and prediction of the service life of concrete structures in the marine environment.	1) Ensemble ML 2) SVM 3) ANN 4) RF	1) R2 = 0.83, MAPE= 39.1 , RMSE = 0.16 2) R2 = 0.47, MAPE= 68.7 , RMSE = 0.27 3) R2 = 0.76, MAPE= 37.3 , RMSE = 0.16 4) R2 = 0.81, MAPE= 37.2 , RMSE = 0.16	642 records of field exposure data of surface chloride concentration in marine concrete are collected
[23]	Quantification of digitally crack patterns on reinforced concrete shell elements.	1) Bagged Trees 2) Subspace KNN 3) RUSBoosted Trees	1) Accuracy: 89.3 2) Accuracy: 80.1 3) Accuracy: 71.4	A dataset with 119 images from crack patterns of reinforced concrete shells
[24]	The debonding behavior between High-performance fiber reinforced concrete and Normal Concrete subjected to direct shear loading is analyzed.	1) Linear-GP techniques	1) R2= 0.97, MAE= 0.24, RMSE= 0.3	HPFRC-NC data is manufactured using two bonding strategies, ie mechanical surface treatments with and without chemical agent.
[16]	Prediction of concrete properties applied to the programming of framework removal activities.	1) MP5	1) R2 = 0.91, MAPE= 0.09 , RMSE = 1.97 The averages of the different predicted variables are reported.	A dataset containing information on the mixture proportions and the values of the mechanical properties at different ages was collected.
[25]	A data-driven approach to classifying the in-plane failure modes of infill frames.	1) RF 2) Adaptative Boosting 3) SVM	1) Accuracy: 81.1 2) Accuracy: 78.5 3) Accuracy: 77.2	A database consisting of 114 infill frame specimens are constructed.
[26]	Prediction of the coefficient of thermal expansion of concrete.	1) RF 2) Linear Regression	1) R2= 0.76, RMSE= 0.22 2) R2= -0.04, RMSE= 0.46	Wisconsin database of concrete mixes
[27]	Automation, safety, cost, and time savings through observation and estimation of crack propagation	1) Voronoi processes digital images	Thickness of the crack geometry.	Real cases of concrete cracks
[28]	Prediction of the compressive strength of concrete to improve the safety and durability of this	1) ANN	The comparison is reported through bar charts.	The database is generated, from numerous sources, including literature, companies, institutions and laboratories.
[29]	The reliable prediction of the carbonation depth of concrete structures applied to the maintenance of structures.	1) Artificila Neural Network 2) Decision Tree 3) Bagged decision Tree 4) Boosted decision Tree	1) MSE= 0.24, MAE= 0.29, RMSE=0.49 2) MSE= 0.42, MAE= 0.32, RMSE=0.64 3) MSE= 0.38, MAE= 0.34, RMSE=0.61 4) MSE= 0.26, MAE= 0.31, RMSE=0.51	The data used for the development of the prediction model was prepared in the Finnish DuraInt-project.
[30]	Percussion-based method to identify the moisture level of concrete	1) SVM	1) Accuracy >98%	The four cubic specimens, with dimensions 150 mm Œ 150 mm Œ 150 mm and cubic compressive strength is 50 MPa.
[31]	Optimization of embodied energy and carbon dioxide emissions of a reinforced concrete column.	1) Cost 2) CO2 3) Embodied Energy	when a 10% cost increase is assumed, embodied energy and emissions are reduced by up to 22% and 63%, respectively.	A short RC column with a square section, subjected to both axial force and moment, is modeled.

Table 1: Summary of applications and techniques in concrete structures.

concretes. A complete containing information on mixed proportions and mechanical property values at different ages was compiled using internationally published documents. Various performance metrics were used to evaluate the performance of the developed models, including the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the coefficient of determination (R2), and the correlation coefficient (R). The results indicated that the proposed models could provide reliable predictions of the target mechanical properties.

The coefficient of thermal expansion (CTE) significantly influences the performance of the concrete. However, CTE measurements are expensive; therefore, CTE is often predicted from empirical equations based on historical data and concrete composition. In [26], the authors were focused on applying linear and random forest (RF) regression methods to predict CTE and other properties from a Wisconsin concrete mix database. The results of this article show that the accuracy of the RF model is significantly better than the prediction methods recommended by the American Association of Highway and Transportation Officials (AASHTO) for CTE. Additionally, RF significantly outperformed the linear regression technique, where the value of R2 was much lower. The latter shows that the behavior of CTE does not have a linear dependence on the independent variables.

The compressive strength of concrete is a fundamental parameter in the design of durability and the prediction of the useful life of concrete structures in civil engineering projects. Therefore, being able to predict this resistance has a significant practical utility. In [21] the authors proposed a hybrid ensemble surrogate ML technique for predicting the compressive strength of concrete. The proposed model is robust in handling overfitting problems and is therefore suitable for predicting the compressive strength of concrete.

Predicting the carbonation depth of concrete structures is essential for optimizing their design and maintenance. In [29], a way to improve the prediction of carbonation is proposed using a model based on ML. The model in question considers the parameters that influence the carbonation process. In the study, an example is carried out that allows us to see the model's applicability, which allows predicting the depth of carbonation with high precision. Underwater and hydraulic concrete structures require periodic inspection due to the constant water loads. Determining the humidity in the structures is very important since it guarantees the correct functioning of the structures. In [30], the authors proposed a method for determining humidity based on percussion. The method includes the Mel Frequency Cepstral Coefficients

(MFCC) used as a characteristic of the sound included by impact. A microphone was also used with which the impact-induced sound signals were obtained. The use of ML techniques, particularly a support vector machine (SVM), is proposed to predict moisture in the concrete. Finally, the authors report that the proposed system has a precision greater than 98%.

Estimating the axial strength of concrete columns confined with steel tubes is essential when making structural designs. However, this estimation is challenging because it depends non-linearly on a series of parameters such as the compressive strength of the concrete, the elastic limit of the steel, the diameter of the column, the thickness of the steel tube, the length of the column. In [18], an optimized hybrid ML model was proposed with the aim of predicting the axial force in columns. To address this challenge, a hybrid method was used that integrates the support vector regression method with the Gray wolf optimization metaheuristic. To verify the quality of the results, they were compared with models that use neural networks, random forest, and linear regression. With the hybrid method, an R2 coefficient was obtained with respect to the real values of 0.992 and an average error percentage of 7%.

Concrete mixing is a complex process that contains several stages. In

[28], ML techniques are used to improve the design of concrete mixes. By building and analyzing an extensive database of concrete recipes and their respective laboratory validations. One of the main results of this study is the translation of the architecture of the proposed ANN to a mathematical equation that can be used in practical applications in the real world.

One of the most common uses of machine learning is to generate prediction models. In [22], the use of ML models to predict chloride concentration in marine concrete surfaces is addressed. The study uses a ML ensemble model to predict the concentration of surface chloride (Cs) in concrete. In the first place, a database is established that is then used to train five ML models, which are: linear regression (LR), Gaussian process regression (GPR), support vector machine (SVM), artificial neural network multilayer perceptron (MLP-ANN) and RF. In addition, the metaheuristic combination of predictions of RF, MLP-ANN, and SVM achieves greater precision when predicting compared to each model independently.

The use of machine learning methods also applies to sustainable concrete design. Specifically, in [31] the embodied energy and carbon dioxide emissions of a reinforced concrete column are optimized. Conventionally, the design of reinforced concrete structures focuses on minimizing construction costs while satisfying the structural design code. However, the aspect of sustainability is a relevant dimension in structural design. According to the experiments, it is concluded that when a cost increase of 10% is assumed, the embodied energy and the CO_2 emissions can suffer an overall reduction of up to 22% and 63%, respectively.

A second group identified in the bigram analysis corresponded to crack detection and concrete monitoring. Checking the damage status of a structure is essential when checking concrete structures. In the article [32], it is proposed to design a framework for the automated probabilistic classification of cracks in cementitious components based on acoustic emission (AE) signals. Waveform parameters, including RA and average frequency (AF) values, are grouped by an unsupervised grouping algorithm dictated by density. Using the Support Vector Machine (SVM) algorithm, clusters that intersect in the data are separated through a hyperplane. Finally, it is possible to establish that the expectations based on the compound theory are correct; this is achieved through the cracking modes that are obtained from the proposed machine learning approach.

Cracks in concrete structures are certainly an indicator that something is wrong, and over the years, the process of detecting these indicators has been carried out manually; that is, there must be a person in charge of the process that generates the precision of the measurements is not entirely correct. In [33], the way to perform this inspection automatically using ML techniques is proposed. In principle, there is a training stage where images are binarized, used to extract possible regions of cracks, then classification models with a convolutional neural network. Finally, the proposed method is evaluated with other concrete images that contain and do not contain cracks. The same is raised in [34], where they proposed automatically detecting cracks through images using a convolutional neural network.

In [27], the Voronoi Diagram algorithm was used to estimate crack patterns and spread on a random concrete surface. A random photo of a concrete crack located on the surface of a fountain is taken, and the dimensions and directions of the crack are measured. After that, the crack was divided into 12 parts to assess the algorithm's ability to estimate the crack pattern, including its direction. As a result of the study, it is identified that this method is precise, fast, economical, and useful for monitoring and estimating the propagation of cracks in concrete surfaces.

High-Performance Fiber Reinforced Concrete (HPFRC) is a standard concrete (NC) structure repair material. In [24], a prediction model based on HPFRC and ML to address repair problems in concrete structures is addressed. This is achieved in the first instance by conducting a study on the disunity behavior between HPFRC and NC subjected to a direct shear load. A finite element (FE) model is then developed to predict the direct debarking response. Finally, a ML model is developed that makes it possible to formulate the shear strength of HPFRC-NC.

In concrete crack analysis, acoustic emission monitoring has taken an important role since it allows for monitoring changes in structural integrity and durability. However, it is necessary to distinguish crack signals from ambient noise. In [19] a convolutional network model is explored, allowing us to distinguish environmental noise signals from the crack's own signals. In particular, a two-dimensional convolutional model was proposed, able to distinguish and separate both sets successfully.

In [35] the authors address the problem of automatic detection of cracks in concrete structures from images. The article indicates that a more practical and precise method is necessary, for which they propose a method based on image processing using the light gradient magnification machine (LightGBM). It is possible to obtain a precision of the proposed method of 99.7%, a sensitivity of 75.71%, a specificity of 99.9%, a precision of 68.2% and an F measure of 0.6952. With these results, it is possible to demonstrate that the proposed method manages to detect cracks with great precision in concrete structures.

In [25], a classification of in-plane failure modes are established for concrete frames using ML. In the first instance, an experimental database is built, then six ML algorithms are implemented and evaluated for the failure mode classification. In this article, it was obtained a result that the highest precision (85.7%) was achieved with the Adaptive Boosting and Support Vector Machine algorithms.

In [23], a study is presented proposing an automated approach to quantifying digitally documented crack patterns in reinforced concrete shell elements subjected to reverse cyclical shear loads. A set of artificial cracks is analyzed using multifractal analysis. With the results of the parametric study, a multiclass classification model is trained and used to estimate the level of damage for cracked concrete elements. Finally, the multifractal characteristics manage to translate the shape of the crack patterns into meaningful information with an accuracy of 89.3%.

4.2. Retaining Walls design

Retaining walls are rigid concrete walls used to laterally support the soil so they can be retained at different levels on the two sides. Optimizing cost and CO2 emissions in retaining walls is a relevant issue for the competitiveness of construction companies and the environmental impact of the construction of these structures. Within ML applications in the efficient design of retaining walls, hybrid models have been used to estimate safety factors. The particle swarm optimization (PSO) algorithm has been used to calculate the optimal construction cost of reinforced concrete retaining walls. Models that combine ANN with the artificial bee colony algorithm (ABC) have also been used to estimate and optimize the safety factors of retaining walls.

4.2.1. Bigram document analysis

This section details the bigram analysis performed for the concepts of machine learning and retaining walls. The results are shown in Figure 7. When analyzing the treemap in the upper left corner, retaining wall concepts such as geotechnical engineering, carbon emissions, bearing capacity, and loads, all of them typical of the retaining wall subject. However, ML concepts such as forecasting, classification, neural networks, mean square error, and convolutional neural networks are also mentioned. Additionally, a third group is observed that is related to optimization, with concepts such as optimization algorithms and artificial bee colonies appearing. When co-words are analyzed, and subsequent grouping occurs, the lower right figure illustrates groups associated with retaining walls, wall height, friction angles, and artificial intelligence algorithms or prediction models. Additionally, there is a subgroup for optimization, specifically of reinforced concrete walls, and metaheuristic algorithms such as harmony search or hybrid algorithms. When creating a conceptual structure map, we notice that the major groups correspond to two (lower left Figure): on the one hand, concepts related to retaining walls and ML algorithms such as neural networks appear predominantly in red. On the other hand, another group appears in blue, which is concerned with optimizing the design of walls and metaheuristic algorithms. The dendrogram illustrates the relationship between the various concepts mentioned previously (Figure top right).

4.2.2. Traditional analysis

In Table 2, a summary of the different articles selected for retaining wall structures is shown. There is an important group of applications related to metaheuristics, machine learning, and optimization of costs, emissions, and embodied energy. On the other hand, there are also ML applications in re-

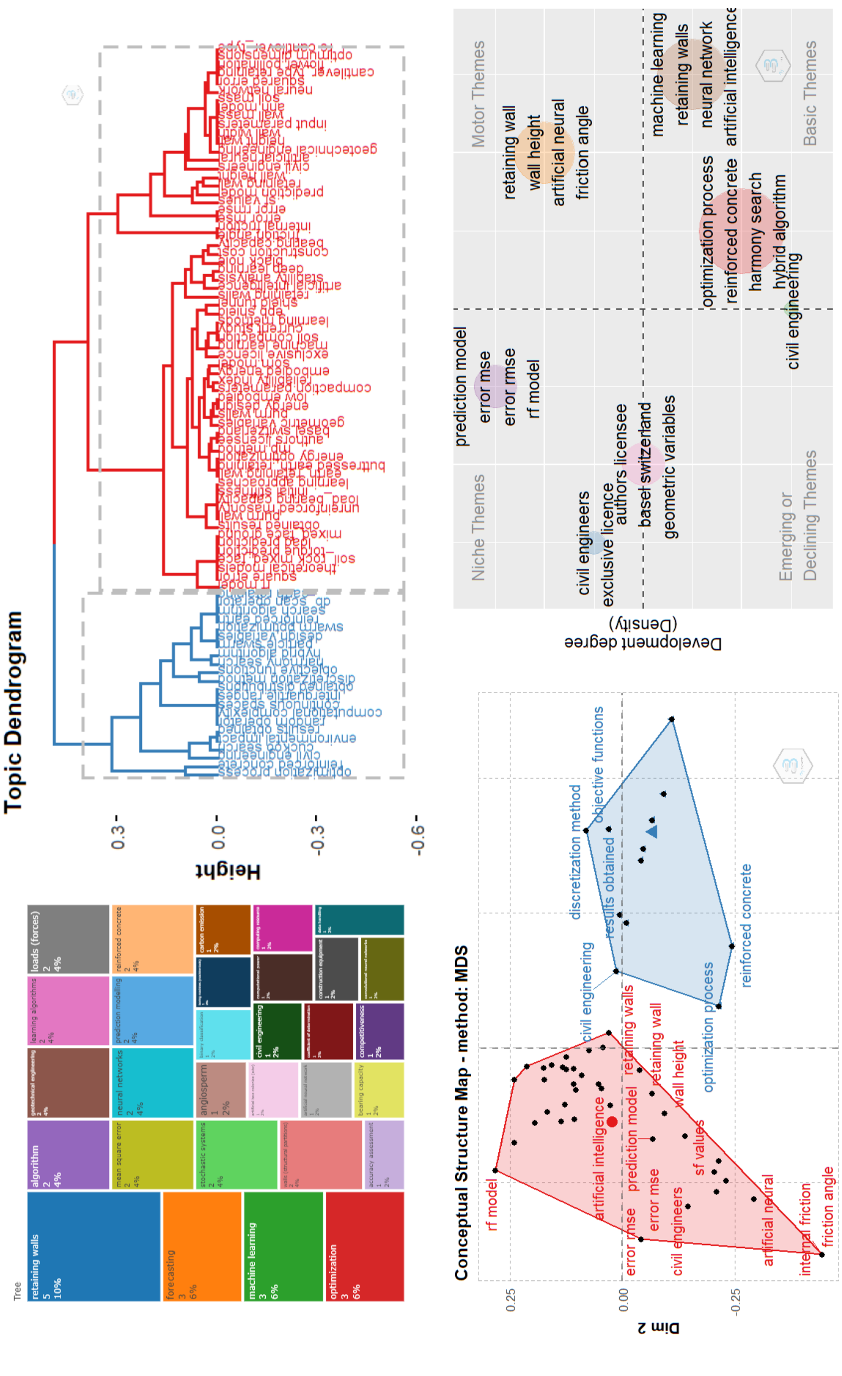

Figure 7: Tree, Thematic, conceptual and dendrogram maps applied a retaining wall data set.

Reference	Application	Techniques	Results	Data
[36]	sustainable design of counterfort retaining walls	Black Hole metaheuristic optimization	Reduction in Cost and CO_2 emissions	A retaining wall defined at different heights and applying restrictions of the construction regulations
[37]	Predicting safety factor of retaining walls in geotechnics	1) ANN 2) ABC- ANN	1) RMSE= 0.038 , R^2= 0.916 2) RMSE= 0.018 , R^2= 0.983 Average over models.	Safety factors of 2800 retaining walls were modeled.
[38]	The buttressed walls problem optimization	PSO	Reduction in Cost and CO_2 emissions	A buttressed wall defined at different heights and applying restrictions of the construction regulations
[39]	Estimating soil compaction parameters for seeking safe and economic of retaining walls	1) RF		Database of 147 samples collected from different studies.
[40]	The retaining walls problem optimization	Cuckoo search algorithm	Reduction in Cost and CO_2 emissions	A retaining wall defined at different heights and applying restrictions of the construction regulations
[41]	Design of safety factors of retaining wall under both static and dynamic conditions	1) ABC- ANN	1) RMSE= 0.005 , R^2= 0.998 Average over models.	Safety factors of 2800 retaining walls were modeled.
[42]	Optimization of factors for the selection of retaining wall techniques during the decision-making stage.	Self-organizing maps	1) Accuracy: 79.8 %	Data from 129 excavation project cases without missing values were collected from building construction companies in large South Korean cities.
[43]	automated process to obtain low embodied energy buttressed earth-retaining wall optimum designs	1) Simulated annealing with mutation operator	Reduction in Cost and embodied-energy	A retaining wall defined at different heights and applying restrictions of the construction regulations
[44]	Image-data-driven deep learning in the stability analysis of geosystems	1) Convolutional Neural Network	1) Accuracy: 97.94%	2D images for retaining walls, organized as data sets of sizes 500 to 200,000, labeled using a traditional mechanical method.
[45]	Evaluation of the failure probability of retaining walls	1) MARS 2) EmNN 3) SOS-LSSVM	1) RMSE=0.0017,R^2=0.9999 2) RMSE=0.0183,R^2=0.9860 3) RMSE=0.0002,R^2=1.0000	Datasets were generated based on consideration of uncertainties in soil parameters
[46]	Calculation of the load capacity through different methods of optimization of reinforced concrete retaining walls	1) PSO	Cost Optimization	A retaining wall defined at different heights and applying restrictions of the construction regulations

Table 2: Summary of applications and techniques in retaining walls.

taining walls related to safety factors. When going into detail in the articles regarding the group related to optimization, metaheuristic or hybrid techniques are mainly explored to solve the optimization of costs, emissions, or energy consumption. It was found that optimizing cost and CO_2 emissions in earth retaining walls is critical for a construction company's competitiveness and that optimizing emissions is critical for the environmental impact of construction. In [36], the optimization based on the black hole algorithm was used, along with a discretization mechanism based on min-max normalization. The results obtained were compared with another algorithm that solves the problem (Harmony Search algorithm). Solutions that minimize CO_2 emissions prefer the use of concrete rather than those that optimize cost. When compared to another algorithm, the results show good performance in optimization using the black hole algorithm. In [38], the buttressed walls problem was determined using an application of a hybrid clustering PSO algorithm. In this study, the focus was the optimization in the design of reinforced earth retaining walls, particularly minimizing the amount of CO_2 emissions generated in its construction and the economic cost. This problem has high computational complexity since it involves 32 design variables. The authors propose a hybrid algorithm in which the PSO method is integrated that solves optimization problems in continuous spaces with the db-scan clustering technique. The db-scan operator significantly improves the solutions' quality, showing good results compared to the harmony search algorithm.

In [40], a hybrid k-means cuckoo search algorithm was applied to the counterfort retaining walls problem. In [46] a PSO algorithm is employed to calculate the optimum construction cost of reinforced concrete retaining walls. Geotechnical and structural limitations are considered constraints for the optimization problem. The critical role of building in natural resource use is driving structural design professionals to develop more efficient structural designs that reduce emissions and energy consumption. In [43], an automated approach to generating optimal buttressed earth retaining wall designs with minimal embodied energy is described. In this research, two objective functions were used to compare the cost optimization and embodied energy optimization strategies. This study employed a hybrid simulated optimization algorithm to determine the geometry, concrete resistances, and concrete and material quantities required to create the optimal buttressed earth-retaining wall with the lowest embodied energy. A relationship was discovered between the two optimization criteria, implying that cost and energy optimization are inextricably related. This permits the statement that a 1 cost reduction results in a 4.54 kWh reduction in energy consumption.

The other interesting group obtained from the bigrams analysis was the application of ML techniques to prediction and classification. Particularly in [37], the authors present intelligent models to solve problems related to retaining walls. For this, the safety factors of 2800 retaining walls were modeled and recorded, considering different effective parameters of retaining walls. This includes the following parameters: wall height, wall thickness, friction angle, soil density, and rock density. A combination of the artificial bee colony (ABC) and ANN algorithm was used to approximate the safety factors of the retaining wall (compared to a previously developed ANN without ABC). The performances of the generated models were evaluated using coefficients of determination (R^2) and performance indices of the error (RMSE). The new hybrid model (ANN + ABC) can significantly increase the performance capacity of the network (compared to ANN without ABC). R^2 values of 0.982 and 0.985 for training and testing of the ABC + ANN model, respectively, compared to values of 0.920 and 0.924 for the ANN model (without ABC). In conclusion, the results showed that the new hybrid model could be introduced as a sufficiently capable technique in the field of this study to estimate the safety factors of RW. In [41], a combination of ANN and artificial bee colony (ABC) is employed for predicting and optimizing safety factors of retaining walls. A comprehensive database of 2880 datasets was used; the input parameters included wall height, wall width, wall mass, soil mass, and internal angle. A critical point in the study of retaining walls is the structure's failure probability. In [45], a reliability study of the structure is conducted using ML techniques, incorporating geotechnical variables. They are predicted using Neural Networks, Multivariate Adaptive Regression Splines, and vector machine support techniques. The application of these techniques yielded results that deviated by less than two % of the real values, simplifying the process of calculating these safety factors.

Making design decisions is a subjective process that considers multiple dimensions such as economic, social, and environmental. In [42], self-organizing maps (SOM) were used to simulate decision-making in order to determine the most appropriate retaining wall technique. N-fold cross-validation was used to validate the model. This study demonstrates that self-organized maps are beneficial for decision-making when selecting a retaining wall method. The SOM had a maximum accuracy of 81.5 percent and a mean accuracy of 79.8 percent. Through the use of classification convolutional neural networks, in

[44], models were built that were trained using previously classified retaining wall images. These images indicated whether the constructed wall was safe or not. In the training process of the convolutional network, image sets that had between 500 and 200,000 images were used to verify the results against 20,000 images later in the testing stage. The result of the models achieved an accuracy of 97.94 % in the safety classification of a wall. In [39], an estimation of compaction parameters is performed. Estimating these parameters is an essential point in the design of retaining walls. The Proctor Test is usually used to make this estimate. However, this test is expensive and time-consuming. The study developed a new model for predicting compaction parameters based on eleven new progressive ML methods to overcome these limitations. The modeling phase was performed using a database of 147 samples collected from different studies. Model performance was evaluated across six metrics in addition to incorporating K-fold cross-validation. The comparative study demonstrated the effectiveness of the RF technique, which showed the highest performance in predicting soil compaction parameters.

4.3. Pavement Engineering

Pavement engineering is a discipline that uses engineering techniques to optimize the design and maintenance of flexible asphalt and rigid concrete pavements. Determining the shear strength of soil is an essential task in the design phase of a pavement construction project. For this purpose, models integrating the support vector machine (SVM) algorithm and cuckoo search optimization (CS) have been used. Some architectures based on convolutional neural networks (CNN) have also been used for the detection of pavement cracks on asphalt surfaces. With this same purpose, deep convolutional neural networks with transfer learning have been used to detect and classify pavement faults based on computational vision automatically.

4.3.1. Bigram document analysis

This section details the bigram analysis performed for ML and pavement concepts. The results are shown in Figure 8. When analyzing the treemap, concepts related to crack detection, monitoring, and conditions and the prediction of coefficients or variables related to the pavement are highlighted. This can be seen in the upper left Figure by complementing the analysis with an analysis of co-words and clustering, which is shown in the lower image on the right. We see that there is a group related to pavement maintenance policies. Another group is associated with cracks, and a third group is related to pavement condition prediction. On the other hand, techniques such as deep learning and RF stand out. Finally, when performing a conceptual map clustering, two groups stand out. The first group in blue is mainly distinguished maintenance and policies related to pavement maintenance. On the other hand, the red cluster is a little more diffuse, highlighting the application of ML techniques related to cracks in the pavement analysis of parameters such as vibration, shear strength, and pavement surface. This is complemented by the dendrogram shown in the upper right image, which indicates the closeness between the different concepts.

4.3.2. Traditional analysis

In Table 3, a summary of the different articles selected for Pavements structures is shown. Among the main techniques used, different architectures of convolutional networks stand out, in addition to ANN multilayer perceptron, RF and SVM. Regarding the applications, the detection of Crack, and prediction of indicators related to its monitoring and deterioration stand out. When performing traditional analysis driven by the topics found in bigram analysis. An interesting group that appears is related to pavement maintenance policies. The effects of climate change in particular, which are related to temperature changes, directly impact the pavement. Having a guide to guarantee the adequate maintenance of the pavements allows efficiencies to be made when maintaining them. In [55] the authors address this problem

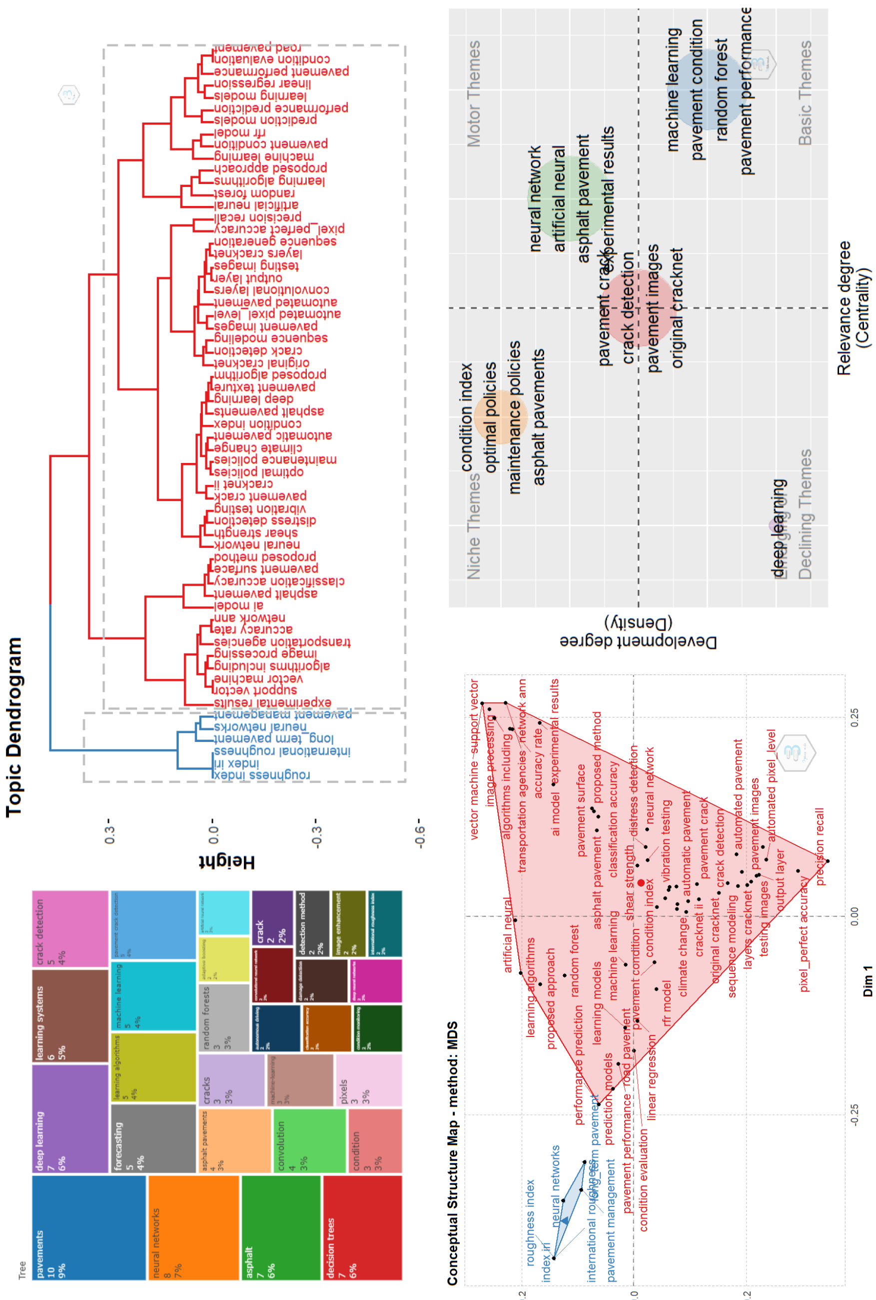

Figure 8: Tree, Thematic, conceptual and dendrogram maps applied a pavement data set.

Reference	Application	Techniques	Results	Data
[47]	Pavement surface deterioration detection and classification system	1) ANN 2) SVM	1) Accuracy=97.5 2) Accuracy=95.0 Averages	Images of 600 pavements of road surfaces in Thailand
[48]	Detect and analyze road macrotexture on pavement types	1) Augm-RF 2) GAN-DenseNet	1) Accuracy=58 2) Accuracy=82	Small dataset of images of pavement
[49]	automated pixel-level crack detection in 3D asphalt pavement images	1) CrackNet-V	1)Pr:84.3, Re: 90.1, F-1:87.1	500 test images
[50]	Automatic pavement crack segmentation	1) U-net-A 2) U-net-B 3) U-net-C	1)Pr:96.9, Re: 93.5, F-1:95.0 2)Pr:97.3, Re: 94.3, F-1:95.8 3)Pr:95.8, Re: 82.4, F-1:87.7	CFD and AigleRN
[51]	Estimate the international roughness index of flexible pavements using deterioration, traffic, weather, maintenance, and structural data.	1) RF	1) MSE=0.974, R2=0.006	Data with over 12.300 samples of distress, 28.700 of rutting data, and 19.900 of IRI data for asphalt pavement (LTPP)
[52]	Automated pavement distress detection and classification	1) VGG-16 2) VGG-16+RF 3) VGG-16+LR	1)Pr:90.0, Re: 90.0, F-1:90.0 2)Pr:86.0, Re: 86.0, F-1:85.0 3)Pr:88.0, Re: 88.0, F-1:87.0	1056 pavement images (HMA-surfaced and PCC-surfaced) from the FHWA/LTPP
[53]	Detecting potholes on asphalt pavement surface	1) LS-SVM 2) ANN	1) CAR=88.75, AUC=0.96 2)CAR=85.25, AUC=0.92	A data set consisting of 200 image samples has been collected
[54]	detecting and classifying asphalt pavement crack.	1) SVM 2) ANN 3) RF	1) CAR=87.5 2) CAR=84.3 3) CAR=70.0	The data set consists of 200 samples
[55]	Sustainability of the pavements against the increase in temperature	1) GBR 2) RF 3) ANN	1)Accuracy=90.71 2)Accuracy=86.92 3)Accuracy=77.65	The data used was provided by a local company, 537 records of asphalt pavement segments.
[56]	Pavement crack prediction	1) ANN-Model1	1)R2=0.89, RMSE=0.525 Averages	The FDOTs PMS data of about 9109 pavement segments were monitored over a period of 40 years.
[57]	find optimal maintenance policies in a road network	1) GBR 2) RF 3) ANN	1) Accuracy=91.2 2) Accuracy=87.9 3) Accuracy=78.7 Averages	The data is from asset management companies in the entire network of Iran.
[58]	pavement performance prediction models in pav-ement management systems	1) RF	1) R2=0.955, MSE=0.279 Averages for IR indicator	Different datasets were created for the 5 and 10-years predictions.
[59]	Pavement Surface Technical Condition Index Deterioration Prediction Model	1) LightGBM	1) R2=0.754, MSE=2.651	Highway Pavement Data.
[60]	Pavement condition monitoring and maintenance	ANN, SVM , Neuro Fuzzy, Linear Regression	R2, MSE, MAE. It is a summary of prediction models of the asphalt pavement degradation condition	Long-term pavement performance
[61]	Prediction of soil shear strength for road construction	1) LSSVM	1) RMSE=0.082, MAPE=14.841, R2=0.885	A dataset of 332 soil samples collected from the Trung Luong National Expressway Project in Viet Nam
[62]	Automated pixel-level crack detection on three-dimensional asphalt pavement surfaces	1) CrackNet-R	1) Pr=88.9, Re=95.0,F-1=91.8	3000 diverse 3D images for training and 500 for testing
[63]	Automated pixel-level crack detection on three-dimensional asphalt pavement surfaces	1) CrackNet-II	1) Pr=90.2, Re=89.1, F1=89.6	2500 diverse 3D images for training and 200 for testing
[64]	Automated pixel-level crack detection on three-dimensional asphalt pavement surfaces	1) CrackNet	1) Pr=90.1, Re=87.6, F-1=88.9	1800 diverse 3D images for training and 200 for testing

Table 3: Summary of applications and techniques in pavements.

in the case of Iran. Particularly in certain areas, climate change has changed from a cold semi-desert to a relatively hot semi-desert. In the article, ML algorithms are used to develop a methodology that allows evaluating the necessary maintenance differences and thus developing a maintenance policy. This policy allows an adequate evaluation of the costs involved in the maintenance process due to the effects of climate change. In [57], a framework was proposed using ML to find optimal maintenance policies in a road network. The stages included grouping the network based on relevant factors, identifying criteria that impact optimal policies, and determining policies and application periods. Additionally, regression algorithms such as gradient boost regression, lasso, ridge, RF regression, and neural network, among others, were used to quantify and predict the cost of policies.

A second line found in the bigram analysis is related to the detection of cracks and distress in the pavement. In [64], an architecture based on Convolutional Neural Networks (CNN) called CrackNet, is developed and employed for pavement crack detection on threedimensional (3D) asphalt surfaces. This same group of authors from the School of Civil and Environmental Engineering at Oklahoma State University (USA) published three new versions, in [63], of the CNN-based pavement crack detection architecture, CrackNet II, CrackNet-V, and CrackNet-R . In [62], a CrackNet version using recurrent neural networks was developed (called CrackNet-R), four times faster and with better accuracy than the original CrackNet version. This version proposes a gated recurrent multilayer perceptron (GRMLP) to update the internal memory recursively. GRMLP is intended for deeper input and hidden state abstractions by conducting multilayer nonlinear transforms at gating units. The training of CrackNetR is completed using 3,000 diverse 3D images. The analysis using 500 testing pavement images shows a precision of 88.9%, a recall of 95.0%, and a Fmeasure of 91.84%. In [49], CrackNet-V was developed as a more efficient version of the CNN-based architecture. This version has a deeper architecture but fewer parameters, with improved accuracy and efficient feature extraction.

In [52], Deep Convolutional Neural Networks (DCNN) with transfer learning were applied for computer vision-based automated pavement distress detection and classification. The FHWA/LTPP database with multiple Pavement images datasets was used. The truncated DCNN was used to build deep features for road imaging. Various ML classifiers were trained using semantic image vectors. A neural network classifier trained in deep transfer learning vectors gave the best results.

In [47], a novel method based on a hybrid ABC-ANN model for pavement surface distress detection and classification was used. In this study, the ANN was used to classify a hazard area as a specific hazard type, including transverse cracks, longitudinal cracks, and potholes. The study results demonstrate that the hybrid ABC-ANN approach works well for pavement distress detection and can classify types of distress on pavement images with reasonable precision. The precision obtained by the proposed ABC-ANN method achieves an increase of 20% compared to the existing algorithms.

In [54], the performance of different ML algorithms was analyzed for asphalt pavement crack classification, including support vector machine (SVM), ANN, and the RF. The feature set consisting of the properties derived from the projective integral and the properties of crack objects can offer the most desirable result. Experimental results show that SVM has achieved the highest classification accuracy rate (87.50%), followed by ANN (84.25%) and RF (70%). The proposed approach may be useful in assisting transportation agencies and inspectors in the task of assessing the condition of the pavements.

A relevant issue in public safety is related to cracks in the pavement, despite advances in imaging techniques and segmentation. Segmenting or recognizing pavement cracks is a non-trivial problem. This is because there is no regularity in the pavement cracks, so there is no clear pattern. At In, [50], a variation of the U-net topology was developed to perform automatic pavement crack detection. To validate the proposal, benchmark data such as CFD and AigleRN were used.

The primary non-destructive pavement evaluation methods are image recognition models, ML algorithms, and visual inspections. While the previous methodologies are efficient, they include uncertainty, noise, and overfitting. By and large, the cracks do not follow a predictable pattern. The use of ANN to predict the qualification of cracks in pavements is addressed in [56] to strengthen the results of the learning models already used in predicting cracks in pavements. An interesting facet of the work is the data used. The model formulation incorporates variables such as average daily traffic and truck factor, road functional class, asphalt thickness, and pavement condition time series data. By and large, the work concludes that ANNs are considered suitable ML models for crack classification.

In [53] also uses ML techniques to detect potholes on the asphalt pavement surface. In this case, Gaussian filters, steerable filters, and integral projection are used to extract features from digital images. Once the feature set was generated, the robustness of the LS-SVM and ANN methods was evaluated. The evaluation was performed using 200 images as a training and validation set. Both methods had values in the precision indicator above 85% and a ROC-AUC of 0.96. Particularly LS-SVM was the one that obtained the best results.

An application thinking of autonomous cars corresponds to detecting the texture of the road since it directly affects the operation of the tires and braking. In [48], deep learning is used to perform pavement texture recognition. As a first step, the captured images were pore-processed and subsequently augmented using the Generative adversarial networks (GANs). Finally, the RF technique and the Densenet network were used for the texture identification process. The latter obtained better precision than RF. Particularly when using the data augmented with GANs, a better quality database is obtained, and therefore when training with this new set of images, it is observed that the accuracy improves from 59% to 82%. To train the adversary network, 250,000 iterations were used. These methods were also found to work better than manual methods.

Regarding the third group related to using ML in order to predict pavement properties. The shear strength property of the soil is critical. Determining the shear strength of the soil is an important task in the design phase of the construction project. In [61], the authors present a hybrid AI model that integrates the Least squares support vector machine (LSSVM) algorithm and the cuckoo search optimization (CSO). A data set of 332 soil samples collected from the Luong National Highway Project in Vietnam was used to construct and validate the model. The input variables used in this study were: the depth of the sample, the percentage of sand, the percentage of clay, the percentage of clay, the moisture content, the wet density of the soil, the specific gravity, the liquid limit, the plastic limit, the plastic index, and the liquid index. LSSVM is used to generalize functional mapping that estimates shear strength from the information provided by the input variables. The LSSVM model requires proper configuration of the regularization and parameters of the kernel function; instead, the CSO algorithm is used to determine these parameters automatically. The experimental results show that the prediction precision of the LSSVM and CSO hybrid method (RMSE = 0.082, MAPE = 14.841, and $R^2 = 0.885$) is better than that of the reference approaches that include the standard LSSVM, the ANN, and the tree regression. Therefore, the proposed method is a promising alternative to assist construction engineers in estimating the shear strength of the soil.

Another interesting indicator to consider in flexible pavements is the international roughness index (IRI). The RF technique is used in [51] to perform automatic prediction on this indicator. Eleven thousand samples were used to create the data set. Eighty percent of the data was used in the training process, with the remaining twenty percent reserved for validation. Sampling was conducted at random. The results outperformed regularized linear regression models, with indicators exceeding 95%. When the importance of variables is analyzed, it is discovered that the primary influencing variables are the initial value of IRI, as well as the average rainfall, fatigue cracking, and transverse cracking. In [58] a general ML technique be used to construct models for pavement performance prediction in pavement management systems (PMS). The proposed models were developed using a RF algorithm and datasets that included past IRI observations as well as structural, meteorological, and traffic data. The proposed approach is compatible with a variety of machine learning algorithms and emphasizes generalization performance. A case study is presented for the prediction of the IRI over the next five and ten years utilizing the Long-Term Pavement Performance.

Pavement condition prediction is a powerful and critical tool for determining the most effective maintenance approaches and treatment processes.

Similar to previous works, in [60], use ML methods to forecast the IRI and pavement condition indices (PCI). These performance indices are frequently used in pavement monitoring to correctly determine the state of a pavement's health. Additionally, the paper discusses the most critical variables that pavement condition prediction models include. In [65], the prediction of the PCI indicator is addressed through the use of cascade models. The goal is to be able to replace visual inspections, and in order to calibrate the models, they chose the six most frequent defects: patches, alligator cracks, transverse and longitudinal cracks, shoving, and potholes. The cascade architecture uses traditional learning models integrated with a neural network. After applying the statistical cross-validation techniques, the results show that the model can predict the index with an adequate degree of precision. Finally, the pavement maintenance quality index (PQI) prediction is covered in [59]. The study proposes a prediction model for the deterioration of the technical condition index of the pavement surface based on the Light Gradient Boost Machine. To properly fit the model, the grid-search technique was used. The prediction result is compared with the prediction result using a RF. The comparison indicates that the boost method has a good prediction; this is observed when analyzing the R_2 indicator, which obtained a value of

0.754 and the MAE that reaches 2.651.

4.4. Tunnels

Tunnels are underground infrastructure that seeks to connect two external points by crossing flat surfaces, mountainous accidents, and even seas. One of the main challenges in tunnel engineering is the inspection, evaluation, maintenance, and safe operation of the infrastructure. In order to study structural damage in tunnels, computer vision techniques have been used, including combinations of convolutional neural networks (CNN) and fuzzy spectral clustering (Fuzzy spectral clustering). On the other hand, predicting machinery performance is critical for accurate cost estimation in tunnel construction projects. For this purpose, deep neural network models have been used to predict the penetration rate of tunnel boring machinery. These systems offer high detection accuracy compared to existing methods.

4.4.1. Bigram document analysis

The bigram analysis is shown in Figure 9. In the upper left Figure, tunnel inspection and crack detection are obtained as major issues being developed in tunnels. This is confirmed in the graphs below. In the thematic map, shown in the lower right Figure, we see that three groups appear. One group is related to tunnel inspection, another group is related to crack detection, and a third group does not have a precise meaning. When analyzing the clusters generated by the conceptual map, shown in the Figure lower left. It is noted that two clusters appear; the blue one is related to the concept of monitoring and structural health with image segmentation, ML, deep learning, and convolutional networks. In a second cluster in red, the concept of crack detection appears related to penetration rates, excavations, and geotechnics and in conjunction with metaheuristic optimization techniques, deep learning, and ML. When analyzing the dendrogram in the upper right Figure, we see that tunnel inspection is very close to convolutional networks and image segmentation concepts. On the other hand, in the red group, crack detection concepts are related to metaheuristic techniques such as artificial bee colony and ML regression and classification techniques.

4.4.2. Traditional analysis

In Table 4, a summary of the techniques, applications, and results obtained in the different works analyzed is shown. Regarding the applications, the inspections and monitoring of tunnels stand out, in addition to the prediction of penetration rates and performances. Among the techniques, the use of SVM, convolutional networks and Multilayer perceptron stands out.

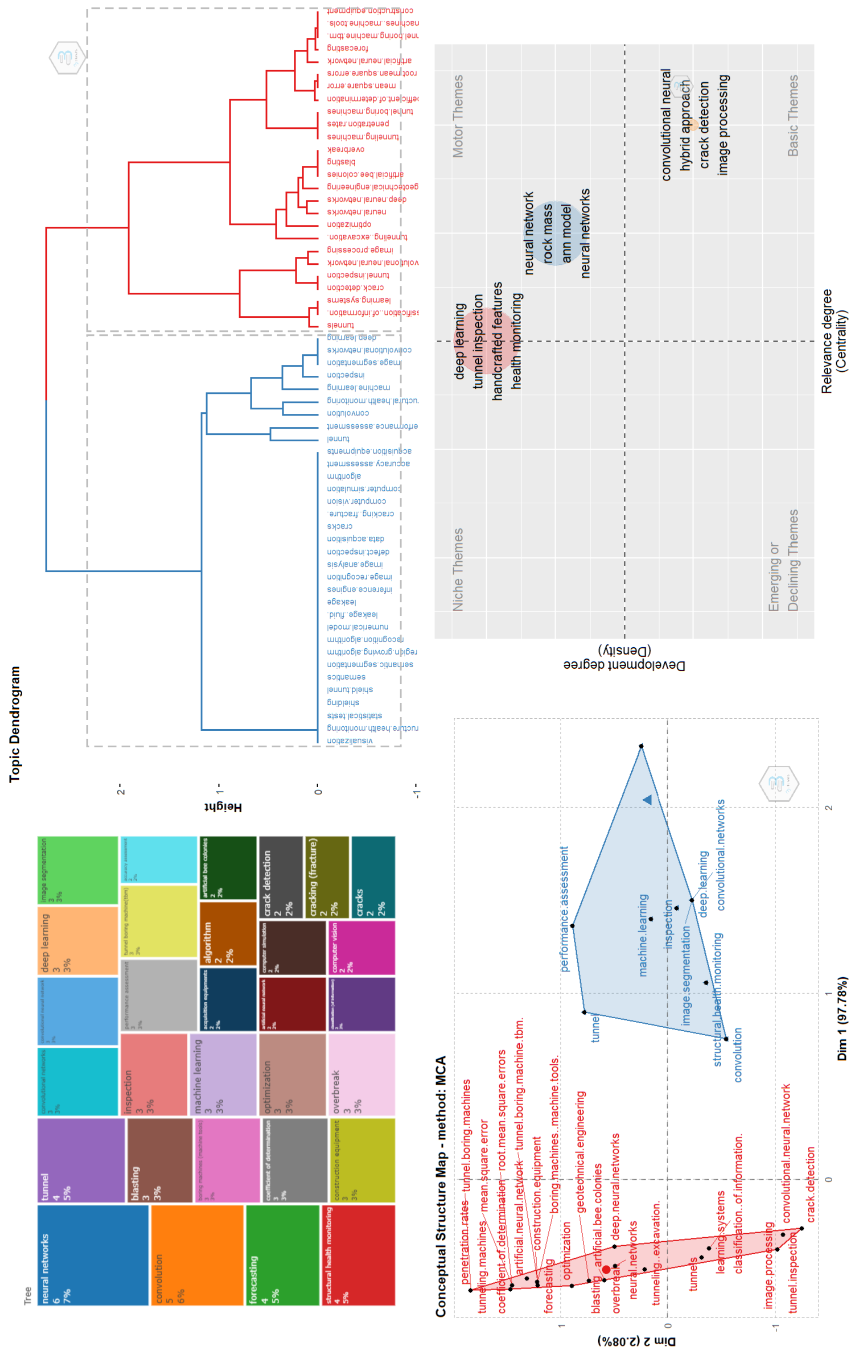

Figure 9: Tree, Thematic, conceptual and dendrogram maps applied a Tunnel data set.

Reference	Application	Techniques	Results	Data
[66]	Tunnel inspection for structural monitoring	1) CNN-FuzzySpectral 2) SVM (Rbf) 3) SVM (Linear)	1) Acc=0.64, F1=0.49 2) Acc=0.58, F1=0.33 3) Acc=0.54, F1=0.31	Images are from the Metsovo motorway tunnel in Greece.
[67]	Automatic visual inspection of cracks in tunnels	A proprietary framework that includes Crack enhancement, threshold segmentation, and filtering.	1) Acc=0.85, Re=0.93	100 cracks as experimental data. In the experiment, a positive detection is considered if more than 50% of the crack is identified
[68]	Predict overtopping induced by blasting operations in tunnels.	1) ABC-ANN	1) R2=0.923, RMSE=0.428	Dataset of Gardaneh Rokh tunnel, Iran.
[69]	Predict overtopping induced by blasting operations in tunnels.	1) ABC-ANN 2) ANN	1) R2=0.904, RMSE=0.090 2) R2=0.947, RMSE=0.065 Best values	330 data sets
[70]	Reliability for the evaluation of the stability of tunnel structures	1) UD-SVM	1) R2=0.997, MSE=0.023	20 training samples and 10 testing samples
[71]	Periodic inspections of electrical energy, railway, structural and signaling infrastructures.	1) SVM (Linear)	1) Pr=0.98, Re=0.91, F1=0.94 Averages without consider others	The datasets used come from LYNX Mobile Mapper from Optech Inc.
[72]	Inspection, evaluation, maintenance and safety of tunnels	1) CNN 2) AnchorGraph 3) SVM (Rbf)	1) Acc=0.886, F1=0.886 2) Acc=0.757, F1=0.822 3) Acc=0.719, F1=0.795	Over 100,000 samples acquired with a single monocular camera
[73]	prediction of the penetration rate of tunnel boring machines.	1) ANN 3 Layers (DNN)	1) R2=0.934, RMSE=0.032	A database comprising 1286 datasets of five parameters was considered.
[74]	Performance prediction in mechanized tunnel projects	1) KNN 2) ANN 3) SVM	1) R2=0.907, RMSE=0.204 2) R2=0.924, RMSE=0.180 3) R2=0.914, RMSE=0.183	Data 209 records generated in 13 km of the PSRWT tunnel.
[75]	The convergence rates of two tunnels from the Namaklan Twin Tunnel were predicted	1) ANN-MLP 2) ANN-RBF	1) R2=0.93,RMSE=0.17,MAE=0.12 2) R2=0.81,RMSE=0.27,MAE=0.22	The data set was collected through field investigations and laboratory experiments.
[76]	Prediction of lining response for twin tunnels	1) MARS 2) Decision Tree	1) R2=0.968 2) R2=0.994	A total of 682 cases were modeled considering five key parameters on twin-tunnel structural forces
[77]	Detection of concrete defects in tunnels	1) CNN 2) SVM (Poly) 3) SVM (Rbf)	1) Acc=0.886, F1=0.886 2) Acc=0.877, F1=0.719 3) Acc=0.864, F1=0.795	Detections are captured and validated by an expert
[80]	Automatic detection and segmentation of concrete cracks in tunnels	1) U-Net 2) CrackSegNet-Dilated 3) CrackSegNet-FocalLoss	1)Pr=63.85,Re=47.46,F1=54.45 2)Pr=74.84,Re=70.46,F1=72.58 1)Pr=66.07,Re=85.54,F1=74.55	A total of 409 images, 4032Œ3016, were obtained in a tunnel in Huzhou

Table 4: Summary of applications and techniques in Tunnels.

On the other hand, the first line identified in the bigram analysis is related to tunnel crack detection. In [66], convolutional neural networks and fuzzy spectral clustering were used for real-time crack detection in tunnels. This article proposes a computational vision model for tunnel crack detection, a challenging process due to low visibility, curvature, and crack structures that, although very narrow, are very deep. The proposed system integrates a robot that examines tunnels in real-time as it moves through the infrastructure. Initially, a convolutional neural network is used to detect cracks. Then, a combined fuzzy spectral clustering is introduced to refine the detected crack regions. The model was tested in tunnels on the Egnatia Highway. Due to the low visibility and geometry of the system, the accuracy and F1-score values are not that high; however, the system offers a considerable improvement in detection compared to existing methods. Additionally, the ability of the robot to touch the crack allows for on-site measurements with accuracy.

In [67], an image acquisition system is designed, which uses multi-line scanning cameras. The objective is to capture images of the tunnel surface to generate a model for automatic crack detection. For the training of the model, three stages were developed. The first is an improvement of the data set through a frequency-domain improvement algorithm. A filter is then generated to remove noise generated by water stains and existing devices on the tunnel's surface. Finally, a segmentation algorithm is used to segment the cracks. The algorithm was tested on Line 1 of the Beijing subway, surpassing state-of-the-art algorithms.

Predicting cracks or overflows in the face of critical conditions is vital in monitoring and maintaining essential infrastructure. In [68, 69], a neural network was built, which was used to predict the overbreak induced by the blasting operations of the Gardaneh Rock tunnel. R2 values of 0.923 were obtained in the validation set. With this model and considering that overbreak is one of the main difficulties in tunnel excavations, the excavation operation is improved. Specifically, extra drilling of 47% was achieved.

In [70] stability evaluation using reliability was applied; the main difficulty of the above is the nature of the limit state function. The article developed a hybrid approach, integrating the uniform design with a regression model using the support vector machine technique, was developed. The hybrid proposal was evaluated in three tunnels with different characteristicsa first simplified case and later two real cases. The results concluded that the hybrid method could train adequately with less data than traditional methods, maintaining the quality of the predictions.

The second line of research obtained from bigram analysis is related to tunnel inspection and analysis of operational conditions. One way to detect the health status of structures in tunnels is by laser scanning. This form is proposed in the article by [71], where they focus directly on railway tunnels because they represent one of the tunnels whose accidents can be more catastrophic. However, it is mentioned that the human component in these types of constructions continues to be predominant, which is why it is worrying and generates a need to advance through automation. The study determined that laser scanning in conjunction with custom processing tools can provide data for additional structural operations. A methodology is used divided into the preprocessing of the point cloud, then the division of the cloud into terrestrial and non-terrestrial points, and finally, the detection of the elements present and each of the clouds.

In [72], Deep convolutional neural networks were used for efficient vision-based tunnel inspection. One of the main challenges facing engineers today is the safe inspection, evaluation, maintenance, and operation of civil infrastructure. For this process, manual processes are used, which are slow and produce subjective results, or automated approaches, which depend on complex handmade characteristics, where it is seldom known in advance which characteristics are important for the problem in question. This article proposes a fully automated tunnel evaluation approach. Complex features were hierarchically constructed with a monocular camera using a deep learning model. The obtained features were used to train a defect detector using a convolutional neural network to build high-level features and, as a detector, a multilayer perceptron was used due to its global function approximation properties. Very rapid predictions were obtained with the proposed system due to the advancing nature of convolutional neural networks and multilayer perceptrons.

In [73], an application of deep neural networks was employed to predict the penetration rate of tunnel boring machines(TBM). Performance prediction is critical to accurate and reliable cost estimation using a TBM in mechanized tunnel construction projects. A wide variety of artificial intelligence methods have been used in predicting the penetration rate of TBM. This focuses on developing a deep neural network (DNN) based model, an advanced version of an ANN, for predicting the penetration rate of TBM based on data obtained from the transfer tunnel of raw water Pahang-Selangor in Malaysia. Based on the results obtained from the coefficient of determination and the root mean square error (RMSE), a significant increase in the prediction of the performance of the penetration rate is achieved through developing a predictive DNN model. The DNN model demonstrated better performance for estimating the penetration rate than the ANN model.

In [74], a supervised machine learning technique was used to predict tunnel boring machine penetration rate. Prediction of the penetration rate is a complex and challenging task due to the interaction between the tunnel boring machine (TBM) and the rock mass. This article discusses the use of supervised ML techniques, including k-nearest neighbor (KNN), chi-squared automatic interaction detection (CHAID), SVM, classification and regression trees (CART), and ANN to predict the penetration rate (PR) of a TBM. To achieve this goal, an experimental database based on field observations and laboratory tests was created for a tunnel project in Malaysia. In the database, uniaxial compressive strength, Brazilian tensile strength, rock quality designation, weathering zone, push force, and revolution per minute was used as inputs to predict the TBM PR. Then KNN, CHAID, SVM, CART, and NN predictive models were developed to select the best. In this article, the KNN model has the best performance to predict the PR of TBM. The KNN model identified uniaxial compressive strength (0.2) as the most important and revolution per minute (0.14) as the least important factor in predicting the TBM penetration rate.

In [75], the topic of tunnel convergence prediction using ML methods is addressed. The study focuses on the construction of a tunnel in Namaklan where ANN, multivariate linear regression (MLR), multivariate nonlinear regression (MNR), support vector regression (SVR), Gaussian process regression (GPR), regression trees (RT), to predict the convergence rate (CR). Six predictive parameters were selected, which are: cohesion, internal friction angle, uniaxial compressive strength of the rock mass, rock mass classification, overburden height, and the number of rock bolts installed. Using the coefficient of determination (R^2) it was possible to determine that the MLP-ANN model is the most optimal, with $R^2 = 0.93$. In contrast, the MLR model has a prediction with the lowest $R^2 = 0.61$, and the RT and GPR models are the least indicated for predicting these indicators.

In [76], it is mentioned how to predict the linear response for tunnels built in anisotropic clay. This is important when building a tunnel because it considerably impacts the duration and safety it will have over time. Five parameters were taken into account to measure: Burial depth, the center-to-center distance of the tunnel, soil resistance, stiffness ratio, and degree of anisotropy. These are known as finite elements (FE). Then, through the application of multivariate adaptive regression splines and decision tree regression methods, the prediction of the bending moment within the linings of the first tunnel is evaluated based on the cases of FE constructed. This allows engineers to estimate the structural response of tunnels with greater reliability.

In [77], the use of an automated robotic inspector that can assess the condition of a tunnel is proposed. This inspector has mobile autonomy, has a crane arm, and is directed by the crack detector based on computer vision. In addition, the robotic inspector has ultrasound sensors, stereo cameras, and a laser scanner. The inspector's method is initially crack detection through a deep learning approach, using a visual inspection based on convolutional neural networks. Then this generates a detailed 3D model of the cracked area using photogrammetric methods. In [80], the idea of detecting cracks in tunnels and their segmentation is raised. They do this using a convolutional deep neural network technique called "CrackSegNet," and a dense segmentation of cracks is carried out in the form of pixels. The network consists of a backbone, dilated convolution, spatial pyramid cluster, and jump connection modules. The proposed network achieves significantly higher precision and generalizability than the compared methods, thus achieving greater efficiency at a low cost.

The manual inspection procedure for cracks and leaks in metro shield tunnels is slow. One of the main causes of the slowness is the difficulty, which is an interference defect that occurs in the tunnels. In [78], the manual procedure was replaced with an automatic procedure based on deep learning. In particular, a semantic segmentation algorithm is proposed to identify cracks and leaks. The proposed method was compared against state-of-the-art methods, finding that the semantic segmentation algorithm is superior to the other methods analyzed. This superiority was not only in the quality of the recognition but also in the processing times to obtain the result. Robotics is a fundamental actor in the automation of tunnel inspection. In [79], a robotic inspector is used for tunnel evaluation. Among the important features, the robotic inspector is able to navigate autonomously in the structure. In addition, it captures images and finally analyzes them to identify defects in the structure. The cracks are detected through deep learning techniques, and later the robot can create a 3D model with the detail of the cracked area. The autonomous system was evaluated in railway and road tunnels.

4.5. Construction Management

Due to the complex and dynamic nature of many construction and infrastructure projects, the ability to detect and classify key on-site activities by various teams and human personnel can improve the quality and management of construction projects. One of the approaches in this matter is using sensors integrated with smartphones as data collection and transmission nodes to detect activities in construction equipment. These systems of recognition and classification of the activity of construction workers are combined with data collected from sensors and ML models. In this way, it is possible to assess the condition, behavior, and surrounding context of construction workers to effectively manage and control projects. Another example is related to safety in construction management. Safety Leading Indicators are a way of flagging sites that are most at risk. Some works propose using machine learning to develop safety indicators that classify sites according to their safety risk in construction projects.

4.5.1. Bigram analysis

Figure 10 shows the bigram analysis performed for the management concept. In the upper left figure, the treemap indicates that Construction projects, Contract delivery, price index, and activity recognition correspond to the most frequent bigram. Regarding ML techniques, we see that the support vector machine is the only technique that appears in the treemap. When analyzing the thematic map, lower right figure, we see an important group related to project management and delivery and other groups related to the activity recognition. In the conceptual structure map, two groups are distinguished in light blue a group related to management and delivery and a more diffuse red group. In the red group, the concepts of productivity monitoring and construction productivity appear again, but there are also the concepts of activity recognition and construction safety.

4.5.2. Traditional analysis

In Table 5, a summary of the articles analyzed in the management area is shown. Among the applications that stand out is the detection of critical activities in relation to safety on the construction site. On the other hand, there are also works related to the prediction of cost indicators or the progress of the project. From the point of view of techniques, KNN and ANN are the main techniques used. By complementing this information with the bigram analysis, we observe a first group related to security and recognition of activities. Activity recognition is an emerging general area with great potential in the Construction Engineering Management (CEM) domain. Due to the com-

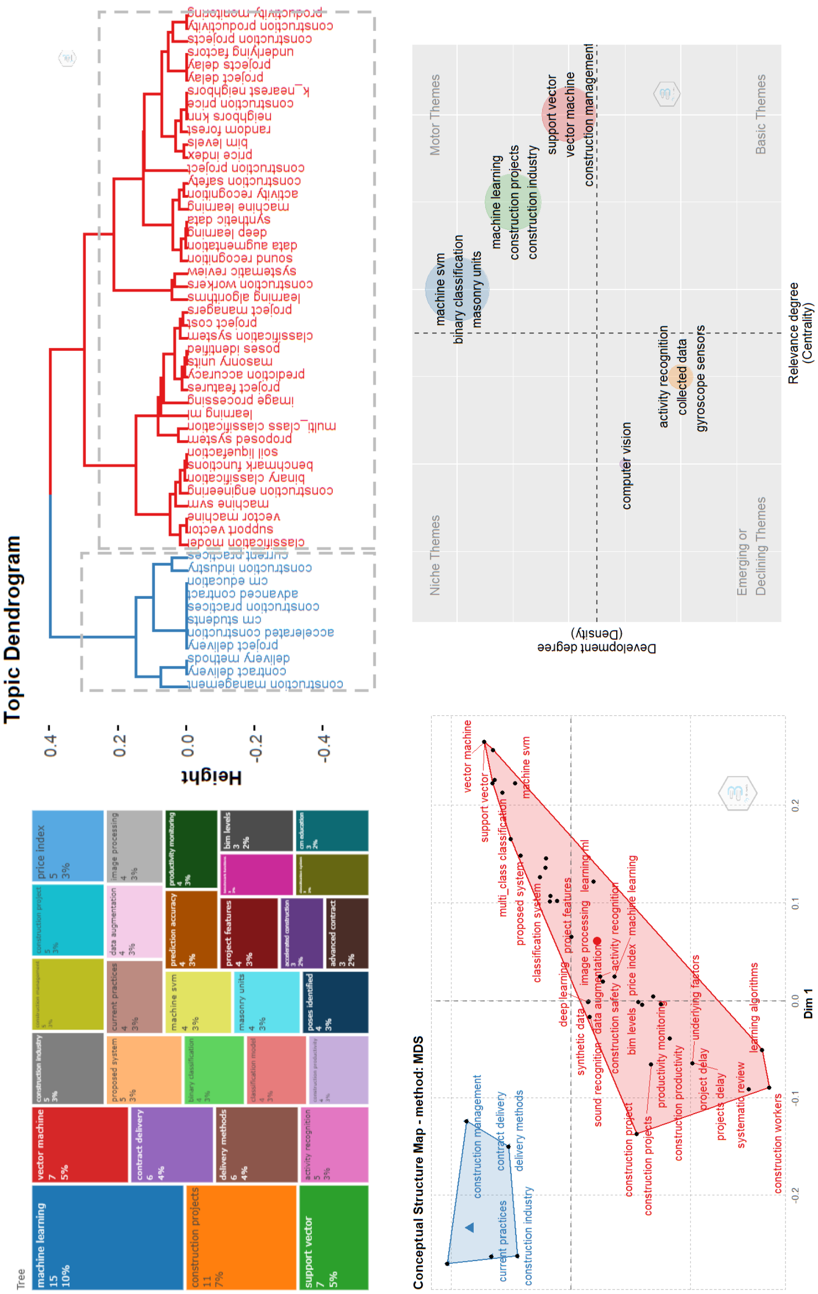

Figure 10: Tree, Thematic, conceptual and dendrogram maps applied a Management data set.

Reference	Application	Techniques	Results	Data
[81]	Detect and classify key activities carried out in the field by various teams and human personnel	1) ANN 2) Decision Tree 3) KNN	1) Acc = 88.7 2) Acc = 84.1 3) Acc = 87.5; Averages	Embedded smartphone sensors as data collection and transmission nodes
[82]	Detect and classify key activities carried out in the field by various teams and human personnel	1) ANN 2) Logistic regression 3) KNN	1) Acc = 90.7 2) Acc = 88.2 3) Acc = 90.5; Averages	Embedded smartphone sensors as data collection and transmission nodes
[83]	base decisions related to construction safety under the uncertainty of knowledge extracted from objective empirical data	1) Random Forest 2) SGTB	Rank Probability Skill Score 1) RPSS = 0.1148 2) RPSS = 0.0865; Averages	Using NLP, a dataset of 4400 attributes and safety outcomes was built.
[84]	Development of indicators that classify sites according to their safety risk in construction projects.	1) Decision Tree 2) RF 3) KNN	1) Acc=0.71 2) Acc=0.78 3) Acc=0.73	Data were obtained from a large contractor in Singapore and the data were accumulated from the year 2010 to 2016.
[85]	Monitoring workers activities	1) SVM	Confusion matrices are reported for activity recognition. Productivity analysis, the time in seconds are reported	Smartphone Sensors
[86]	Automatically detect various types of building materials in building images	1) ANN 2) RBF 3) SVM	1) Pr = 65.3, Re = 60.0 2) Pr = 91.1, Re = 70.4 3) Pr = 88.1, Re = 68.0 Averages were considered	The dataset contains 750 images taken of various constructions, the job site was collected
[87]	Monitor the implementation of each individual part of the buildings and reflect them in the BIM models	A framework that includes different machine learning techniques such as CNN and SVM	No metric defined	Data from the original BIM models and the as-built images
[88]	Analysis of construction sound data to monitor project procedures.	1) Hidden Markov Model	1) Acc = 94.3	Mel-frequency cepstral coefficients are extracted as the features of the six types of sound data.
[89]	Enhance Construction Cost Index forecasting	1) PERT 2) KNN 3) ARIMA	1) MAPE = 0.83,MSE=7415, MAE=77 2) MAPE = 0.78,MSE=9138, MAE=70 3) MAPE = 3.97,MSE=161996,MAE=368 Log term prediction	Short-, mid-, and long-term. Data from January 1985 to December 2014 is collected from ENR and the bureau of labor statistics

Table 5: Summary of applications and techniques in Construction Management.

plex and dynamic nature of many construction and infrastructure projects, the ability to detect and classify key activities carried out in the field by diverse teams and human personnel can improve project decision-making and control quality and reliability.

In [81], embedded smartphone sensors are proposed as ubiquitous multimodal data collection and transmission nodes to detect detailed activities of construction teams. Accelerometer and gyroscope sensors are used to train supervised learning classifiers. To evaluate the models, the selection of discriminatory characteristics was used to extract, the sensitivity analysis of the size of the data segmentation window, and the choice of the classifier to train. Choosing the level of detail (LoD) in describing team actions (classes) is an important factor with a major impact on ranking performance. Computational efficiency and end-use of the classification process may well influence the decision for selecting an optimal LoD to describe team activities (classes).

In [82], a smartphone-based construction workers' activity recognition and classification system is proposed. Assessing the condition, behavior, and surrounding context of construction workers is essential for effective project management and control. The embedded sensors of ubiquitous mobile phones offer a great opportunity to automate the recognition of worker activity. This study proposes the use of smartphones to capture body movements by collecting data using integrated gyro and accelerometer sensors. The collected data is used to train five different types of ML algorithms. Activity recognition precision analysis has been performed for all different ML activity categories and classifiers in user-dependent and independent ways. The results indicate that neural networks outperform other classifiers by offering accuracy ranging from 87% to 97% for user-dependent categories and from 62% to 96% for user-independent categories.

Construction safety is one of this industry's most relevant and concerning issues. Although ML has been considered by construction research for more than two decades, it has not yet been applied to safety concerns. In [83], RF and Stochastic Gradient Tree Boosting (SGTB) models are proposed to a set of categorical safety attributes data extracted from a large set of textual reports of construction injuries. The integration of a natural language processing tool (NLP) developed by the same researchers in previous works is proposed. Both models can predict the type of injury, the type of energy, and the part of the body with great performance ($0.236 < \text{RPSS} < 0.436$), surpassing the parametric models found in the literature. This work opens the door to a new field of research, where construction safety is considered an empirically founded quantitative science.

The construction industry is one of the most dangerous in many countries. Safety leading indicators are a way to mark sites that are most at risk. ML is not widely used in the construction industry, especially in the development of safety-leading indicators. In [84], an ML approach to developing safety leading indicators that rank sites according to their safety risk on construction projects is proposed. In this study, five ML algorithms were compared for predicting the occurrence and severity of accidents. The data includes safety inspection records, accident cases, and project-related data. These data were obtained from a large contractor in Singapore, and the data was accumulated from 2010 to 2016. From thirty-three input variables, 13 input variables were selected using a combination of Boruta technical feature selection and decision tree. Of the 13 input variables selected, six of them are related to the project, and seven of them are elements in the Contractor safety inspection checklists. During validation, the RF model provided the best prediction performance with an accuracy of 0.78 and has achieved substantial strength according to the Weighted-Kappa statistics of 0.70.

Constant monitoring of work progress and identifying deviations from plans are critical to designing a more efficient and safe workplace. Sustained physical work will result in work-related musculoskeletal disorders (WMSD) that can adversely affect the health of workers and the project's budget, schedule, and productivity. To prevent WMSD, health and safety organizations have established rules and regulations limiting labor-intensive activities' duration and frequency. In [85], a wearable sensor data and ML system was used for activity recognition, productivity analysis, and ergonomic risk assessment. The model implements embedded smartphone sensors and a multi-class Support vector machine (SVM) to recognize worker activities in the field and extract duration and frequency information, which will ultimately be used to assess productivity and ergonomic risks associated with each activity.

Project management, control, and delivery were other important groups identified in the bigram analysis. In [86], Digital images and video clips collected at construction job sites are commonly used for extracting useful information. Exploring new applications for image processing techniques within construction engineering and management is a steadily growing field of research. One of the initial steps for various image processing applications is automatically detecting various construction materials on construction images. In this paper, the authors conducted a comparison study to evaluate the performance of different ML techniques for detecting three common building materials: Concrete, red brick, and OSB boards. The employed classifiers in this research are: Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machine (SVM). To achieve this goal, the feature vectors extracted from image blocks are classified to compare the efficiency of these methods for building material detection. The results indicate that for all three types of materials, SVM outperformed the other two techniques in accurately detecting the material textures in images. The results also reveal that the common material detection algorithms perform very well in cases of detecting materials with distinct colors and appearance (e.g., red brick). In contrast, their performance for detecting materials with color and texture variance (e.g., concrete) and materials containing similar color and appearance properties with other elements of the scene (e.g., ORB boards) might be less accurate. For example, OSB surfaces and flooring can have similar color and texture values, making the detection process more challenging. In these cases, an interesting line to explore is strengthening the database with more images. These images can be real or artificially generated through GANs, for example.

In [87], while unavoidable, inspections, progress monitoring, and comparing as-planned with as-built conditions in construction projects do not readily add tangible intrinsic value to the end-users. In large-scale construction projects, the process of monitoring the implementation of every single part of buildings and reflecting them on the BIM models can become highly labor-intensive and error-prone due to the vast amount of data produced in the form of schedules, reports and photo logs. In order to address the mentioned methodological and technical gap, this paper presents a framework and a proof of concept prototype for on-demand automated simulation of construction projects, integrating some cutting-edge IT solutions, namely image processing, ML, BIM, and Virtual Reality. This study utilized the Unity game engine to integrate data from the original BIM models and the as-built images, which were processed via various computer vision techniques. These methods include object recognition and semantic segmentation for identifying different structural elements through supervised training in order to superimpose the real-world images on the as-planned model. The proposed framework leads to an automated update of the 3D virtual environment with the states of the construction site. This framework empowers project managers and stockholders with an advanced decision-making tool, highlighting the inconsistencies in an effective manner. This paper contributes to body knowledge by providing a technical exemplar for the integration of ML and image processing approaches with immersive and interactive BIM interfaces, the algorithms and program codes which can help replicability of these approaches by other scholars.

In [88], the sound recognition technology, which has been adopted in diverse disciplines, has not received much attention in the construction industry. Since each working and operation activity on a construction site generates its distinct sound, its identification provides imperative information regarding work processes, task performance, and safety-relevant issues. Thus, accurate sound data analysis is vital for project participants to monitor project procedures, make data-driven decisions, and evaluate task productivities. To accomplish this objective, this paper investigates the sound recognition technology for construction activity identification and task performance analyses. Mel-frequency cepstral coefficients are extracted for sound identification as the features of the six types of sound data. In addition, a supervised ML algorithm called Hidden Markov Model is used to perform sound classification. The research findings show that the maximum classification accuracy is 94.3% achieved by a 3-state HMM. This accuracy of the adopted technique is expected to reliably execute the construction sound recognition, which significantly leverages construction monitoring, performance evaluation, and safety surveillance approaches.

In [89], the Construction Cost Index (CCI) is calculated monthly and published by Engineering News-Record (ENR). CCI is utilized for capital project budgeting and construction cost estimation, especially when mid- and long-term forecasts are needed. Accurate prediction of CCI helps avoid underestimating and overestimating project costs. However, the current prevailing time series prediction models do not show promising results, especially in mid-and long-term forecasting. The capability of two machine-learning algorithms, k nearest neighbor (KNN) and perfect random tree ensembles (PERT), are utilized to enhance CCI forecasting, especially in the mid-and long-term. The proposed machine-learning algorithms can significantly enhance forecasting CCI's predictability in all the short-, mid-, and long-term scenarios. Data from January 1985 to December 2014 is collected from ENR and the bureau of labor statistics to conduct empirical studies and quantitatively measure the performance of the proposed methods. As the outcomes show, the prediction accuracies of both proposed methods are better than those of current prevailing time series models under all the tested scenarios. It is anticipated that cost estimators can benefit from CCI forecasting by incorporating predicted price variations in their estimates, preparing more-precise bids for contractors, and developing more accurate budgets for owners.

5. Future directions

Figure 11 shows a summary diagram of the five main topics obtained along with the lines that are being developed in each of the topics. In addition, Table 6 has been introduced, which proposes four groups related to challenges and future lines. The first group in the Table, is related to the prediction of variables. The second group is concerned with safety applications, the third group with images and convolutional networks, and the fourth group with the optimization of structural designs. For the first group, which corresponds to the prediction or classification of variables, in the topic of concrete, we find the prediction of its mechanical properties or, in the case of retaining walls, the prediction of geotechnical variables. When analyzing the metrics of the ML models, it is observed that, in general, the ML models are capable of predicting the variables with outstanding results. So the challenge is to move to the second level of ML application. With this, we mean: that the previous studies have been carried out with historical datasets compiled by the authors. How can the model now be put into a production environment?

The first stage is to be able to generate a data lake with information holistic and related to the production processes. The creation of the data lake implies the capture of the variables of interest to subsequently carry out all the engineering and data governance for the proper development of this. On the other hand, how does the result of this prediction fit into decision-making? A model that has good predictions but that is not useful for making decisions does not generate value within an industrial process. These same challenges related to the prediction of variables appear in tunnels, for example, for certain variables such as penetration or overtopping rates or the prediction of costs related to project management.

Considering the overtopping case and safety factor prediction applications such as in the management topic, related to safety and activity recognition or in the case of safety factor prediction in retaining walls. In addition to the two previous challenges, there is a challenge that these predictions must be carried out in times close to real. This generates challenges of having to integrate these safety models with big-data techniques in order to execute decision-making in real-time. The above can also be complemented with all the technologies developed by cloud providers. Another group of interesting applications is related to detecting cracks in concrete, pavements, retaining walls, tunnels, or the case of activity recognition. Usually, the techniques used are related to convolutional neural networks. Convolutional networks, in general, are quite intensive in computation, especially in the training part and if they have a significant number of layers, also when making predictions. Again thinking about the productive case, it is interesting for networks with many layers to be able to generate simpler architectures, with fewer layers, capable of operating on simple hardware, for example, cell phones. This allows, for example, in the case of security applications to be able to carry out close detection in real-time directly in the hardware. On the other hand, in the case of having to train neural networks, it is interesting to explore the capabilities of cloud providers to generate better training in less time. Here we also emphasize the importance of generating a data lake for future experiments and development.

Finally, there is a group of applications related to the optimization of structures. Usually, what is found here are cost optimizations, CO2, or embodied energy. We believe that a fundamental point that would make it easier to integrate into decision-making is to consider different sustainability criteria: economic, environmental, social, and constructability, which naturally implies multi-objective optimization with multi-criteria decisions.

When defining the objective function that guides this optimization, the complete life cycle analysis must be considered: Manufacturing, Construction, Use, Maintenance, and End of Life. Furthermore, all structural designs involve variability and uncertainty. The initial parameters, the structure's dimensions, the materials' mechanical characteristics, and the loads may differ from the design values. Therefore, the optimization should naturally consider this uncertainty to obtain a robust design.

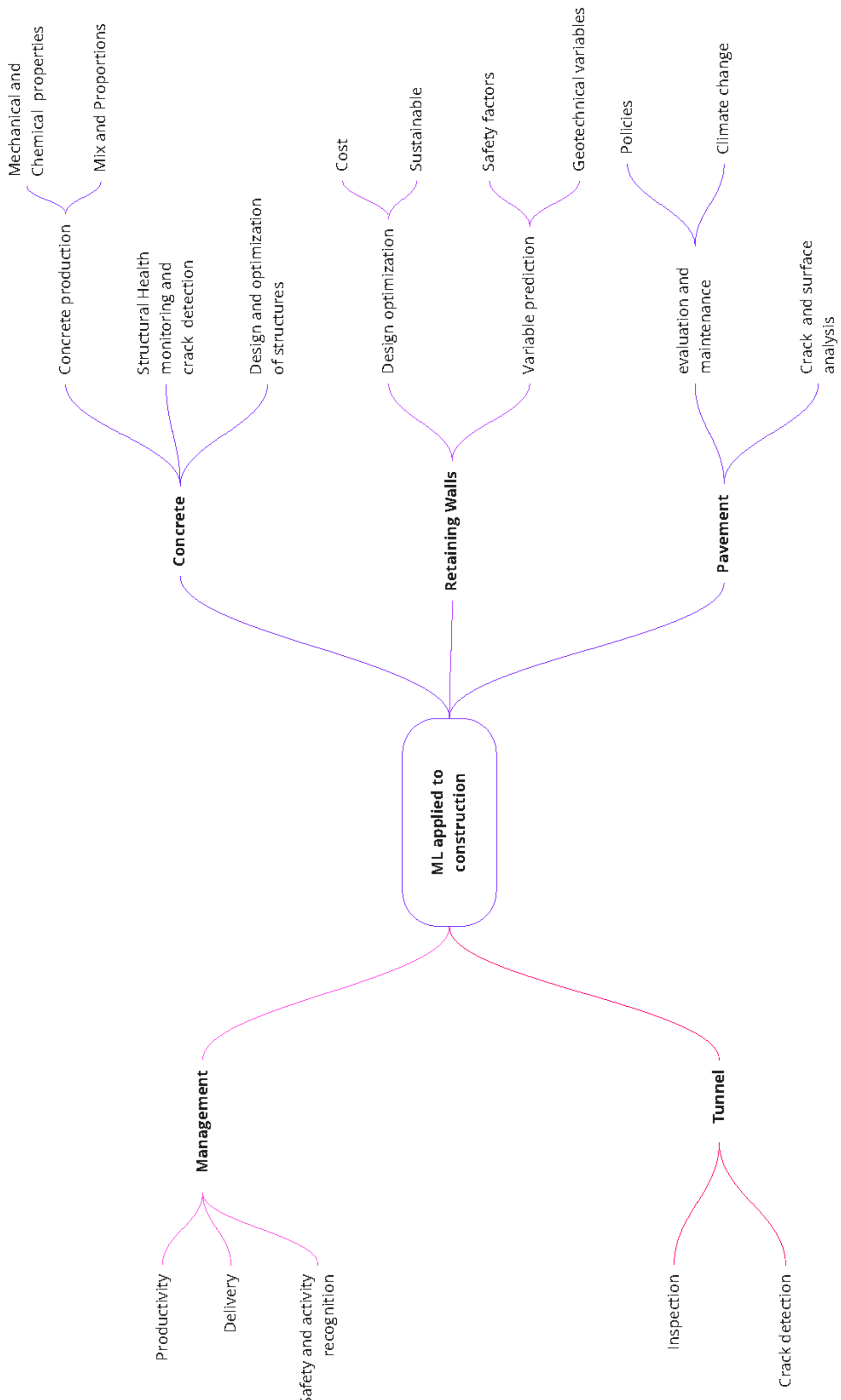

Figure 11: Summary of the main topics identified and lines developed in each of them.

6. Conclusions

In this work, we propose a hybrid methodology. As a first instance, we used the bidirectional encoder representation for the transforms technique to find topics in the abstracts of articles obtained from Scopus. Later we used the expert knowledge to select the relevant topics. This methodology found five topics of ML applications to construction: concrete structures, retaining walls, pavement, tunnels, and management. The leading journals in this area of research are Automation in Construction, Construction and Building Materials, and Computer Engineering.

On the topic of concrete, we distinguish two main research lines; the first is strongly related to automatic crack detection and monitoring of structures, and the second cluster is associated with the prediction or automatic identification of parameters for an efficient and sustainable design of concrete. Regarding retaining walls, the main lines of research have to do with optimizing the design of walls where hybrid techniques between ML and metaheuristics have obtained good performance. On the other hand, the prediction of design parameters of the structure through ML techniques has been studied. Regarding the pavement topic, an essential line of research is related to pavement maintenance policies and how events such as climate

Group	Area	Reference	Actual state	ML-Challenges
Prediction or classification Variables	Concrete Retaining wall Pavement Tunnel Management	[17], [18], [22], [24], [16], [25], [26],[28], [30] [39] [55], [56], [57], [58], [59], [60], [61] [71], [72], [73], [75], [76] [88], [89]	Traditional Machine Learning Models	1) Automatic data acquisition and structuring of the data lake. 2) Models in productive environments, integrated into the decision making 3) Evaluation of the impact of the model on the production process and feedback of the models for future recalibrations
Safety applications	Concrete Retaining wall Pavement Tunnel Management	[20], [21], [27] [41],[45] [50] [68], [69] [81], [82], [83], [84], [85]	Traditional Machine Learning Models	The same challenges as the previous group. Additionally, the incorporation of analytics and big data techniques for real-time
Crack and failure detection	Concrete Retaining Wall Pavement Tunnel Management	[19], [23] [44] [47], [49], [48], [50] , [52], [53], [54], [62], [63], [64] [66], [67], [77],[80] [86], [87]	Usually, models that incorporate deep learning techniques such as deep convolutional networks	In addition to the challenges analyzed in the previous groups, there are challenges in: 1) More efficient network models 2) Use of cloud components to perform more efficient training, process more images, and with better metrics
Structure optimization	Concrete Retaining wall	[31] [36], [38], [40],[42],[43], [46]	Optimization models that integrate machine learning techniques and metaheuristics algorithms. CO_2 , costs and energy optimization criteria are considered	Incorporate cradle-to-grave analysis, robust multi-objective optimization that incorporates multi-criteria decisions in environmental, social, economic, and constructability dimensions.

Table 6: Summary of machine learning challenges.

change affect them. A second line is related to monitoring and detecting cracks and distress in the pavement. In the case of tunnels, structure monitoring appears again as a main line of research in addition to identifying, predicting, and optimizing operational variables such as penetration rates, excavations, and geotechnical variables. Finally, in the case of construction management, incorporating ML in the control, management, costs, and delivery of projects is a line of interest. Still considering project management and administration, another line is related to the safety of workers and the identification of activities within the work.

There is an opportunity to strengthen the proposed hybrid review technique regarding the next steps. We would particularly like to carry out the analysis of other construction themes and consider other areas. Considering the research lines found, we observe that most investigations focus on obtaining the model. However, the model must be inserted into the decision-making process to generate value. At this point, we see an opportunity to extend much of the research. In the case of lines that incorporate optimizations, a large number of fixed parameters are usually considered; an extension would be to consider a robust and multi-objective optimization, considering not only the cost of the optimization but also variables such as environmental or social.

The study is particularly useful for supporting decision-making processes and optimizing the effectiveness and sustainability of construction processes. The results have their roots in the BERT methodology, which leverages ML to investigate prominent and relevant topics. Thus, identifying critical research lines that have the most significant influence in practice provides clear guidance for management to identify, select, and analyze which ML method makes sense to improve their companies performance and sustainability.

This is particularly relevant since the practical application of ML demands a high-skilled workforce and capabilities, which companies do not easily reach. First, information technology resources are highly disputed and often scarce. Second, construction demands compliance because of strict rules and norms, which adds further resources. Hence, having a study setting out the base and the state-of-the-art regarding ML for construction is vital for accelerating and reducing costs for achieving a more pervasive effect on the market.

Another significant implication is the results of the herein applied methodology. We uncovered critical areas in the construction sector by combining BERT methodology with experts knowledge. Expanding such technic to include patents and other scientific and technological knowledge sources may be valuable for recognizing innovation opportunities. Considering that the construction sector is not broadly recognized for high innovativeness and given its relevance for the worlds economy and sustainability, this might have a path for attracting entrepreneurs and companies to pursue innovations, primarily business model innovations combined with product innovations.

7. Acknowledgment

The authors gratefully acknowledge the funding received from the following research projects:

- José García was supported by the Grant CONICYT/FONDECYT/INICIACION/11180056.
- José García and Vinicius Minatogawa was supported by PROYECTO DI INVESTIGACIÓN INNOVADORA INTERDISCIPLINARIA: 039.414/2021.
- Víctor Yepes was supported by Grant PID2020-117056RB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by ERDF A way of making Europe.
- Francisco Altimiras was supported by the INF-PUCV Scholarship.
- Broderick Crawford is supported by Grant CONICYT/FONDECYT/REGULAR/1210810.

References

[1] Dallasega, P., Rauch, E. & Linder, C. Industry 4.0 as an enabler of proximity for construction supply chains: A systematic literature review. *Computers In Industry*. **99** pp. 205-225 (2018), https://doi.org/10.1016/j.compind.2018.03.039.

[2] Osunsanmi, T., Aigbavboa, C. & Oke, A. Construction 4.0: the future of the construction industry in South Africa. *International Journal Of Civil And Environmental Engineering*. **12**, pp. 206-212 (2018), https://doi.org/10.5281/zenodo.1315923.

[3] Rauch, E., Linder, C. & Dallasega, P. Anthropocentric perspective of production before and within Industry 4.0. *Computers & Industrial Engineering*. **139** pp. 105644 (2020), https://doi.org/10.1016/j.cie.2019.01.018.

[4] Hofmann, T. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*. **42**, pp. 177-196 (2001), https://doi.org/10.1023/A:1007617005950.

[5] Blei, D., Ng, A. & Jordan, M. Latent dirichlet allocation. *Journal Of Machine Learning Research*. **3**, pp. 993-1022 (2003), https://doi.org/10.5555/944919.944937.

[6] Lee, D. & Seung, H. Learning the parts of objects by non-negative matrix factorization. *Nature*. **401**, pp. 788-791 (1999), https://doi.org/10.1038/44565.

[7] Arora, S., Ge, R. & Moitra, A. Learning topic modelsgoing beyond SVD. *2012 IEEE 53rd Annual Symposium On Foundations Of Computer Science*. pp. 1-10 (2012), https://doi.org/10.1109/FOCS.2012.49.

[8] Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv Preprint ArXiv:2203.05794*. (2022)

[9] Devlin, J., Chang, M., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*. (2018).

[10] McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv Preprint ArXiv:1802.03426*. (2018), https://arxiv.org/abs/1802.03426.

[11] Campello, R., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. *Pacific-Asia Conference On Knowledge Discovery And Data Mining*. pp. 160-172 (2013), https://doi.org/10.1007/978-3-642-37456-2_14.

[12] Aria, M. & Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal Of Informetrics*. **11**, pp. 959-975 (2017), https://doi.org/10.1016/j.joi.2017.08.007.

[13] Grivel, L., Mutschke, P. & Polanco, X. Thematic mapping on bibliographic databases by cluster analysis: a description of the sdoc environment with solis. *Journal Of Knowledge Organization*. **22**, pp. 70-77 (1995), https://doi.org/10.5771/0943-7444-1995-2-70.

[14] López-Fernández, M., Serrano-Bedia, A. & Pérez-Pérez, M. Entrepreneurship and family firm research: A bibliometric analysis of an emerging field. *Journal Of Small Business Management*. **54**, pp. 622-639 (2016), https://doi.org/10.1111/jsbm.12161.

[15] Cobo, M., López-Herrera, A., Herrera-Viedma, E. & Herrera, F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal Of Informetrics*. **5**, pp. 146-166 (2011), https://doi.org/10.1016/j.joi.2010.10.002.

[16] Behnood, A. & Golafshani, E. Machine learning study of the mechanical properties of concretes containing waste foundry sand. *Construction And Building Materials*. **243** (2020), cited By 38, https://doi.org/10.1016/j.conbuildmat.2020.118152.

[17] Zhang, J., Zhang, M., Dong, B. & Ma, H. Quantitative evaluation of steel corrosion induced deterioration in rubber concrete by integrating ultrasonic testing, machine learning and mesoscale simulation. *Cement And Concrete Composites*. **128** (2022), cited By 0, https://doi.org/10.1016/j.cemconcomp.2022.104426.

[18] Ngo, N., Pham, T., Le, H., Nguyen, Q. & Nguyen, T. Axial strength prediction of steel tube confined concrete columns using a hybrid machine learning model. *Structures*. **36** pp. 765-780 (2022), cited By 0, https://doi.org/10.1016/j.istruc.2021.12.054.

[19] Han, G., Kim, Y., Kim, H., Oh, T., Song, K., Kim, A., Kim, Y., Cho, Y. & Kwon, T. Auto-detection of acoustic emission signals from cracking of concrete structures using convolutional neural networks: Upscaling from specimen. *Expert Systems With Applications*. **186** (2021), cited By 0, https://doi.org/10.1016/j.eswa.2021.115863.

[20] Mariniello, G., Pastore, T., Asprone, D. & Cosenza, E. Layout-aware Extreme Learning Machine to Detect Tendon Malfunctions in Prestressed Concrete Bridges using Stress Data. *Automation In Construction*. **132** pp. 103976 (2021), https://doi.org/10.1016/j.autcon.2021.103976.

[21] Asteris, P., Skentou, A., Bardhan, A., Samui, P. & Pilakoutas, K. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cement And Concrete Research*. **145** (2021), cited By 35, https://doi.org/10.1016/j.cemconres.2021.106449.

[22] Cai, R., Han, T., Liao, W., Huang, J., Li, D., Kumar, A. & Ma, H. Prediction of surface chloride concentration of marine concrete using ensemble machine learning. *Cement And Concrete Research*. **136** (2020), cited By 30, https://doi.org/10.1016/j.cemconres.2020.106164.

[23] Athanasiou, A., Ebrahimkhanlou, A., Zaborac, J., Hrynyk, T. & Salamone, S. A machine learning approach based on multifractal features for crack assessment of reinforced concrete shells. *Computer-Aided Civil And Infrastructure Engineering*. **35**, 565-578 (2020), cited By 12, https://doi.org/10.1111/mice.12509.

[24] Jiao, P., Roy, M., Barri, K., Zhu, R., Ray, I. & Alavi, A. High-performance fiber reinforced concrete as a repairing material to normal concrete structures: Experiments, numerical simulations and a machine learning-based prediction model. *Construction And Building Materials*. **223** pp. 1167-1181 (2019), cited By 13, https://doi.org/10.1016/j.conbuildmat.2019.07.312.

[25] Huang, H. & Burton, H. Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. *Journal Of Building Engineering*. **25** (2019), cited By 36, https://doi.org/10.1016/j.jobe.2019.100767.

[26] Nilsen, V., Pham, L., Hibbard, M., Klager, A., Cramer, S. & Morgan, D. Prediction of concrete coefficient of thermal expansion and other properties using machine learning. *Construction And Building Materials*. **220** pp. 587-595 (2019), cited By 23, https://doi.org/10.1016/j.conbuildmat.2019.05.006.

[27] Bayar, G. & Bilir, T. A novel study for the estimation of crack propagation in concrete using machine learning algorithms. *Construction And Building Materials*. **215** pp. 670-685 (2019), cited By 28, https://doi.org/10.1016/j.conbuildmat.2019.04.227.

[28] Ziolkowski, P. & Niedostatkiewicz, M. Machine learning techniques in concrete mix design. *Materials*. **12** (2019), cited By 40, https://doi.org/10.3390/ma12081256.

[29] Taffese, W., Sistonen, E. & Puttonen, J. CaPrM: Carbonation prediction model for reinforced concrete using machine learning methods. *Construction And Building Materials*. **100** pp. 70-82 (2015), cited By 35, https://doi.org/10.1016/j.conbuildmat.2015.09.058.

[30] Zheng, L., Cheng, H., Huo, L. & Song, G. Monitor concrete moisture level using percussion and machine learning. *Construction And Building Materials*. **229** pp. 117077 (2019), https://doi.org/10.1016/j.conbuildmat.2019.117077.

[31] Yoon, Y., Kim, K., Lee, S. & Yeo, D. Sustainable design for reinforced concrete columns through embodied energy and CO2 emission optimization. *Energy And Buildings*. **174** pp. 44-53 (2018), cited By 17, https://doi.org/10.1016/j.enbuild.2018.06.013.

[32] Das, A., Suthar, D. & Leung, C. Machine learning based crack mode classification from unlabeled acoustic emission waveform features. *Cement And Concrete Research*. **121** pp. 42-57 (2019), https://doi.org/10.1016/j.cemconres.2019.03.001.

[33] Kim, H., Ahn, E., Shin, M. & Sim, S. Crack and Noncrack Classification from Concrete Surface Images Using Machine Learning. *Structural Health Monitoring*. **18**, 725-738 (2019), cited By 87, https://doi.org/10.1177/1475921718768747.

[34] Yokoyama, S. & Matsumoto, T. Development of an Automatic Detector of Cracks in Concrete Using Machine Learning. *Procedia Engineering*. **171** pp. 1250-1255 (2017), cited By 43, https://doi.org/10.1016/j.proeng.2017.01.418.

[35] Chun, P., Izumi, S. & Yamane, T. Automatic detection method of cracks from concrete surface imagery using two-step light gradient

boosting machine. *Computer-Aided Civil And Infrastructure Engineering*. **36**, 61-72 (2021), https://doi.org/10.1111/mice.12564.

[36] Yepes, V., Martí, J. & García, J. Black hole algorithm for sustainable design of counterfort retaining walls. *Sustainability (Switzerland)*. **12** (2020), https://doi.org/10.3390/su12072767.

[37] Ghaleini, E., Koopialipoor, M., Momenzadeh, M., Sarafraz, M., Mohamad, E. & Gordan, B. A combination of artificial bee colony and neural network for approximating the safety factor of retaining walls. *Engineering With Computers*. **35**, 647-658 (2019), cited By 56, https://doi.org/10.1007/s00366-018-0625-3.

[38] García, J., Martí, J. & Yepes, V. The buttressed walls problem: An application of a hybrid clustering particle swarm optimization algorithm. *Mathematics*. **8** (2020), https://doi.org/10.3390/math8060862.

[39] Benbouras, M. & Lefilef, L. Progressive Machine Learning Approaches for Predicting the Soil Compaction Parameters. *Transportation Infrastructure Geotechnology*. (2021), https://doi.org/10.1007/s40515-021-00212-4.

[40] García, J., Yepes, V. & Martí, J. A hybrid k-means cuckoo search algorithm applied to the counterfort retaining walls problem. *Mathematics*. **8** (2020), https://doi.org/10.3390/math8040555.

[41] Gordan, B., Koopialipoor, M., Clementking, A., Tootoonchi, H. & Tonnizam Mohamad, E. Estimating and optimizing safety factors of retaining wall through neural network and bee colony techniques. *Engineering With Computers*. **35**, 945-954 (2019), https://doi.org/10.1007/s00366-018-0642-2.

[42] Kim, Y., Park, U., Whang, S., Ahn, D. & Kim, S. Selection of optimized retaining wall technique using self-organizing maps. *Sustainability (Switzerland)*. **13**, 1-13 (2021), https://doi.org/10.3390/su13031328.

[43] Martínez-Muñoz, D., Martí, J., García, J. & Yepes, V. Embodied energy optimization of buttressed earth-retaining walls with hybrid simulated annealing. *Applied Sciences (Switzerland)*. **11**, 1-16 (2021), https://doi.org/10.3390/app11041800.

[44] Liu, Z., Hu, S., Sun, Y. & Azmoon, B. An Exploratory Investigation into Image-Data-Driven Deep Learning for Stability Analysis of Geosystems. *Geotechnical And Geological Engineering*. (2021), https://doi.org/10.1007/s10706-021-01921-w.

[45] Mishra, P., Samui, P. & Mahmoudi, E. Probabilistic design of retaining wall using machine learning methods. *Applied Sciences (Switzerland)*. **11** (2021), https://doi.org/10.3390/app11125411.

[46] Moayyeri, N., Gharehbaghi, S. & Plevris, V. Cost-based optimum design of reinforced concrete retaining walls considering different methods of bearing capacity computation. *Mathematics*. **7** (2019), https://doi.org/10.3390/math7121232.

[47] Banharnsakun, A. Hybrid ABC-ANN for pavement surface distress detection and classification. *International Journal Of Machine Learning And Cybernetics*. **8**, 699-710 (2017), https://doi.org/10.1007/s13042-015-0471-1.

[48] Chen, N., Xu, Z., Liu, Z., Chen, Y., Miao, Y., Li, Q., Hou, Y. & Wang, L. Data Augmentation and Intelligent Recognition in Pavement Texture Using a Deep Learning. *IEEE Transactions On Intelligent Transportation Systems*. (2022), https://doi.org/10.1109/TITS.2022.3140586.

[49] Fei, Y., Wang, K., Zhang, A., Chen, C., Li, J., Liu, Y., Yang, G. & Li, B. Pixel-Level Cracking Detection on 3D Asphalt Pavement Images through Deep-Learning- Based CrackNet-V. *IEEE Transactions On Intelligent Transportation Systems*. **21**, pp. 273-284 (2020), https://doi.org/10.1109/TITS.2019.2891167.

[50] Escalona, U., Arce, F., Zamora, E. & Sossa, H. Fully convolutional networks for automatic pavement crack segmentation. *Computacion Y Sistemas*. **23**, pp. 451-460 (2019), https://doi.org/10.13053/cys-23-2-3047.

[51] Gong, H., Sun, Y., Shu, X. & Huang, B. Use of random forests regression for predicting IRI of asphalt pavements. *Construction And Building Materials*. **189** pp. 890-897 (2018), https://doi.org/10.1016/j.conbuildmat.2018.09.017.

[52] Gopalakrishnan, K., Khaitan, S., Choudhary, A. & Agrawal, A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction And Building Materials*. **157** pp. 322-330 (2017), https://doi.org/10.1016/j.conbuildmat.2017.09.110.

[53] Hoang, N. An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction. *Advances In Civil Engineering*. **2018** (2018), https://doi.org/10.1155/2018/7419058.

[54] Hoang, N. & Nguyen, Q. A novel method for asphalt pavement crack classification based on image processing and machine learning. *Engineering With Computers*. **35**, pp. 487-498 (2019), https://doi.org/10.1007/s00366-018-0611-9.

[55] Mahpour, A. & El-Diraby, T. Incorporating Climate Change in Pavement Maintenance Policies: Application to Temperature Rise in the Isfahan County, Iran. *Sustainable Cities And Society*. **71** (2021), https://doi.org/10.1016/j.scs.2021.102960.

[56] Inkoom, S., Sobanjo, J., Barbu, A. & Niu, X. Prediction of the crack condition of highway pavements using machine learning models. *Structure And Infrastructure Engineering*. **15**, pp. 940-953 (2019), https://doi.org/10.1080/15732479.2019.1581230.

[57] Mahpour, A. & El-Diraby, T. Application of Machine-Learning in Network-Level Road Maintenance Policy-Making: The Case of Iran. *Expert Systems With Applications*. **191** (2022), https://doi.org/10.1016/j.eswa.2021.116283.

[58] Marcelino, P., Lurdes Antunes, M., Fortunato, E. & Gomes, M. Machine learning approach for pavement performance prediction. *International Journal Of Pavement Engineering*. **22**, pp. 341-354 (2021), https://doi.org/10.1080/10298436.2019.1609673.

[59] Pei, L., Yu, T., Xu, L., Li, W. & Han, Y. Prediction of Decay of Pavement Quality or Performance Index Based on Light Gradient Boost Machine. *Lecture Notes On Data Engineering And Communications*

Technologies. **80** pp. 1173-1179 (2022), https://doi.org/10.1007/978-3-030-81007-8_135.

[60] Shtayat, A., Moridpour, S., Best, B. & Rumi, S. An Overview of Pavement Degradation Prediction Models. *Journal Of Advanced Transportation*. **2022** (2022), https://doi.org/10.1155/2022/7783588.

[61] Tien Bui, D., Hoang, N. & Nhu, V. A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong National Expressway Project (Vietnam). *Engineering With Computers*. **35**, 955-965 (2019), https://doi.org/10.1007/s00366-018-0643-1.

[62] Zhang, A., Wang, K., Fei, Y., Liu, Y., Chen, C., Yang, G., Li, J., Yang, E. & Qiu, S. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces with a Recurrent Neural Network. *Computer-Aided Civil And Infrastructure Engineering*. **34**, pp. 213-229 (2019), https://doi.org/10.1111/mice.12409.

[63] Zhang, A., Wang, K., Fei, Y., Liu, Y., Tao, S., Chen, C., Li, J. & Li, B. Deep Learning-Based Fully Automated Pavement Crack Detection on 3D Asphalt Surfaces with an Improved CrackNet. *Journal Of Computing In Civil Engineering*. **32** (2018), https://doi.org/10.1061/(ASCE)CP.1943-5487.0000775.

[64] Zhang, A., Wang, K., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J. & Chen, C. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. *Computer-Aided Civil And Infrastructure Engineering*. **32**, 805-819 (2017), https://doi.org/10.1111/mice.12297.

[65] Issa, A., Sammaneh, H. & Abaza, K. Modeling Pavement Condition Index Using Cascade Architecture: Classical and Neural Network Methods. *Iranian Journal Of Science And Technology - Transactions Of Civil Engineering*. **46**, pp. 483-495 (2022), https://doi.org/10.1007/s40996-021-00678-9.

[66] Doulamis, A., Doulamis, N., Protopapadakis, E. & Voulodimos, A. Combined convolutional neural networks and fuzzy spectral clustering for real time crack detection in tunnels. *Proceedings - Interna-*

tional Conference On Image Processing, ICIP. pp. 4153-4157 (2018), https://doi.org/10.1109/ICIP.2018.8451758.

[67] Gong, Q., Zhu, L., Wang, Y. & Yu, Z. Automatic subway tunnel crack detection system based on line scan camera. *Structural Control And Health Monitoring*. **28**, e2776 (2021), https://doi.org/10.1002/stc.2776.

[68] Koopialipoor, M., Ghaleini, E., Haghighi, M., Kanagarajan, S., Maarefvand, P. & Mohamad, E. Overbreak prediction and optimization in tunnel using neural network and bee colony techniques. *Engineering With Computers*. **35**, pp. 1191-1202 (2019), https://doi.org/10.1007/s00366-018-0658-7.

[69] Koopialipoor, M., Ghaleini, E., Tootoonchi, H., Jahed Armaghani, D., Haghighi, M. & Hedayat, A. Developing a new intelligent technique to predict overbreak in tunnels using an artificial bee colony-based ANN. *Environmental Earth Sciences*. **78** (2019), https://doi.org/10.1007/s12665-019-8163-x.

[70] Li, X., Li, X. & Su, Y. A hybrid approach combining uniform design and support vector machine to probabilistic tunnel stability assessment. *Structural Safety*. **61** pp. 22-42 (2016), https://doi.org/10.1016/j.strusafe.2016.03.001.

[71] Sánchez-Rodríguez, A., Riveiro, B., Soilán, M. & González-deSantos, L. Automated detection and decomposition of railway tunnels from Mobile Laser Scanning Datasets. *Automation In Construction*. **96** pp. 171-179 (2018), https://doi.org/10.1016/j.autcon.2018.09.014.

[72] Makantasis, K., Protopapadakis, E., Doulamis, A., Doulamis, N. & Loupos, C. Deep Convolutional Neural Networks for efficient vision based tunnel inspection. *Proceedings - 2015 IEEE 11th International Conference On Intelligent Computer Communication And Processing, ICCP 2015*. pp. 335-342 (2015), https://doi.org/10.1109/ICCP.2015.7312681.

[73] Koopialipoor, M., Tootoonchi, H., Jahed Armaghani, D., Tonnizam Mohamad, E. & Hedayat, A. Application of deep neural networks in predicting the penetration rate of tunnel boring machines. *Bulletin Of*

Engineering Geology And The Environment. **78**, pp. 6347-6360 (2019), https://doi.org/10.1007/s10064-019-01538-7.

[74] Xu, H., Zhou, J., Asteris, P., Armaghani, D. & Tahir, M. Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. *Applied Sciences (Switzerland)*. **9** (2019), https://doi.org/10.3390/app9183715.

[75] Torabi-Kaveh, M. & Sarshari, B. Predicting Convergence Rate of Namaklan Twin Tunnels Using Machine Learning Methods. *Arabian Journal For Science And Engineering*. **45**, pp. 3761-3780 (2020), https://doi.org/10.1007/s13369-019-04239-1.

[76] Zhang, W., Li, Y., Wu, C., Li, H., Goh, A. & Liu, H. Prediction of lining response for twin tunnels constructed in anisotropic clay using machine learning techniques. *Underground Space (China)*. **7**, pp. 122-133 (2022), https://doi.org/10.1016/j.undsp.2020.02.007.

[77] Protopapadakis, E. & Doulamis, N. Image based approaches for tunnels defects recognition via robotic inspectors. *International Symposium On Visual Computing*. pp. 706-716 (2015), https://doi.org/978-3-319-27857-5_63.

[78] Huang, H., Li, Q. & Zhang, D. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunnelling And Underground Space Technology*. **77** pp. 166-176 (2018), https://doi.org/10.1016/j.tust.2018.04.002.

[79] Protopapadakis, E., Stentoumis, C., Doulamis, N., Doulamis, A., Loupos, K., Makantasis, K., Kopsiaftis, G. & Amditis, A. Autonomous Robotic Inspection in Tunnels. *ISPRS Annals Of Photogrammetry, Remote Sensing & Spatial Information Sciences*. **3** (2016), https://doi.org/10.5194/isprs-annals-III-5-167-2016.

[80] Ren, Y., Huang, J., Hong, Z., Lu, W., Yin, J., Zou, L. & Shen, X. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Construction And Building Materials*. **234** (2020), https://doi.org/10.1016/j.conbuildmat.2019.117367.

[81] Akhavian, R. & Behzadan, A. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine

learning classifiers. *Advanced Engineering Informatics*. **29**, pp. 867-877 (2015), https://doi.org/10.1016/j.aei.2015.03.001.

[82] Akhavian, R. & Behzadan, A. Smartphone-based construction workers activity recognition and classification. *Automation In Construction*. **71**, pp. 198-209 (2016), https://doi.org/10.1016/j.autcon.2016.08.015.

[83] Tixier, A., Hallowell, M., Rajagopalan, B. & Bowman, D. Application of machine learning to construction injury prediction. *Automation In Construction*. **69** pp. 102-114 (2016), https://doi.org/10.1016/j.autcon.2016.05.016.

[84] Poh, C., Ubeynarayana, C. & Goh, Y. Safety leading indicators for construction sites: A machine learning approach. *Automation In Construction*. **93** pp. 375-386 (2018), https://doi.org/10.1016/j.autcon.2018.03.022.

[85] Nath, N. & Behzadan, A. Construction productivity and ergonomic assessment using mobile sensors and machine learning. *Computing In Civil Engineering 2017*. pp. 434-441 (2017), https://ascelibrary.org/doi/10.1061/9780784480847.054.

[86] Rashidi, A., Sigari, M., Maghiar, M. & Citrin, D. An analogy between various machine-learning techniques for detecting construction materials in digital images. *KSCE Journal Of Civil Engineering*. **20**, pp. 1178-1188 (2016), https://doi.org/10.1007/s12205-015-0726-0.

[87] Pour Rahimian, F., Seyedzadeh, S., Oliver, S., Rodriguez, S. & Dawood, N. On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning. *Automation In Construction*. **110** pp. 103012 (2020), https://doi.org/10.1016/j.autcon.2019.103012.

[88] Zhang, T., Lee, Y., Scarpiniti, M. & Uncini, A. A supervised machine learning-based sound identification for construction activity monitoring and performance evaluation. *Construction Research Congress 2018: Construction Information Technology - Selected Papers From The Construction Research Congress 2018*. **2018-April** pp. 358-366 (2018), https://doi.org/10.1061/9780784481264.035

[89] Wang, J. & Ashuri, B. Predicting ENR Construction Cost Index Using Machine-Learning Algorithms. *International Journal Of Construction Education And Research*. **13**, pp. 47-63 (2017), https://doi.org/10.1080/15578771.2016.1235063.