Construction and Building Materials 440 (2024) 137397

Contents lists available at ScienceDirect

Construction and Building Materials

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

Machine learning models for predicting physical properties in asphalt road construction: A systematic review

Joerg Leukel [a,*,1], Luca Scheurer [a,2], Vijayan Sugumaran [b,3]

[a] Department of Information Systems and Digital Technologies, University of Hohenheim, Stuttgart, Germany
[b] Department of Decision and Information Sciences, Oakland University, Rochester, MI, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Feature selection
Intelligent compaction
Neural networks
Pavement
Prediction

ABSTRACT

Prediction models using machine learning assume an important role in supporting decisions in asphalt road construction, such as the scheduling of tasks and the control of compaction operations. The development of prediction models for physical properties can be informed by insights from the adoption of specific machine learning techniques. However, the available evidence has not yet been synthesized. To address this deficit, we systematically selected and analyzed 30 eligible studies published in peer-reviewed journals between 2011 and 2023 for data collection and preprocessing as well as training and evaluation of prediction models. The results establish a comprehensive picture of the adoption of machine learning techniques for predicting physical properties in asphalt road construction. Specifically, the review revealed the following findings: (1) a large range of input variables and sensors used; (2) pre-specified models using few input variables that made feature selection unnecessary; (3) an emphasis on Artificial Neural Networks although empirical evidence for their higher performance is yet ambiguous; (4) low adoption rates of unitless performance metrics, which are necessary for the integration of evidence from different studies; and (5) the need for greater completeness and clarity in the reporting of training and test data used.

1. Introduction

Decision-making in asphalt road construction calls for accurate information about the current and future physical properties of materials, road layers, and infrastructure used [1,2]. The direct measurement of many of these properties is only possible with complex technical procedures, which require specially trained personnel, cost-intensive equipment, and working time [3]. For instance, this requirement applies to asphalt tests carried out in the laboratory or using devices, such as a falling weight deflectometer (FWD) [4] and non-nuclear density gauge (NNDG) [5]. An alternative approach lies in prediction models that are learned from potentially relevant input variables, even if the exact influence of these variables is not known in advance. Improvements in machine learning (ML) in recent years have decisively promoted this approach, as shown by the large number and variety of applications in asphalt road construction.

The field of ML prediction models for physical properties is rapidly emerging and concerned with different construction activities, such as compaction and quality inspection before handover, and different road layers. Prediction models have been proposed to replace laborious and time-consuming tests that determine, for instance, the California Bearing Ratio (CBR), subgrade modulus, and asphalt layer density. Moreover, an increasing number of studies develop models to assist machine operators during compaction [6,7]. Collectively, evidence of the usefulness of prediction models for physical properties in asphalt road construction is growing.

However, the development of effective prediction models is made difficult by the variety of ML techniques that are available for the preprocessing of input variables, training of models from past data, and assessment of prediction performance. To better assist in the development of models, the available evidence from published studies should first be identified and mapped onto a conceptual framework of ML techniques. This mapping is essential for understanding the state of the field and gaining insights into the significance of ML techniques. In this

* Corresponding author.
E-mail addresses: joerg.leukel@uni-hohenheim.de (J. Leukel), luca.scheurer@uni-hohenheim.de (L. Scheurer), sugumara@oakland.edu (V. Sugumaran).
[1] ORCID: 0000-0002-4146-9812
[2] ORCID: 0009-0003-5658-2425
[3] ORCID: 0000-0003-2557-3182

https://doi.org/10.1016/j.conbuildmat.2024.137397
Received 29 February 2024; Received in revised form 22 May 2024; Accepted 7 July 2024
Available online 13 July 2024

respect, previous literature reviews are limited: they have either focused on pavement condition assessment during road usage but not construction [8–10], or they have concentrated on studies employing Artificial Neural Networks (ANNs) but not other algorithms [11]. Against this backdrop, further assessment of prediction models for physical properties is warranted. Our research addresses this need by conducting a systematic review. The specific objectives of our research are to: (1) determine the adoption of ML techniques in studies on the prediction of continuous physical properties in asphalt road construction, and (2) synthesize the results to propose future research directions.

This systematic review critically evaluates the methodological approaches in developing prediction models for physical properties in asphalt road construction. Our investigation provides a comprehensive overview of current practices and identifies key areas for improvement to enhance the generalizability and rigor of these models. Specifically, we propose expanded research into alternative ML algorithms, advocate for standardized performance metrics, and recommend controlled experiments to determine the impact of training data volume on model performance. These contributions aim to guide future research and refine model development in this rapidly advancing area.

2. Method

We adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement to guide our search and data extraction from studies [12]. Fig. 1 shows an overview of the research design, which includes the following processes: a) initial identification of studies, b) selection of relevant studies, and c) data collection from these studies. Details on each of these processes are provided in this section.

2.1. Eligibility criteria

Studies were included if they developed and evaluated a ML model that predicted a continuous physical property in asphalt road construction. This application domain was defined as all activities of execution of construction works, starting with the site set-up and ending with the quality control before the road is used. Additional criteria used for study selection included: peer-reviewed article in a journal, written in English, and published from 2011 onwards. Exclusion criteria consisted of the following: (1) no investigation of construction but planning (e.g., prediction of project costs), design of materials (e.g., asphalt mixtures), maintenance (e.g., remaining useful life), or other types of buildings (e.g., concrete roads, tunnels); (2) prediction of a categorical or ordinal variable; (3) no evaluation of a prediction model (e.g., literature review); (4) no adoption of ML (e.g., mathematical modeling); and (5) no collection of real-world data (e.g., simulated data or secondary data from databases only).

2.2. Information sources and search strategy

We employed Scopus and Web of Science as our primary sources for literature retrieval. The decision to use these two electronic databases was guided by their extensive coverage of articles in peer-reviewed journals, ensuring a broad spectrum of relevant research captured across multiple fields and publishers [13,14]. We first searched Scopus on 2023–10–22 and Web of Science on 2023–09–04 to identify relevant articles published from 2011 onwards. The limitation to this period facilitated a more homogeneous analysis by incorporating studies that employ comparable ML techniques and at the same time ensured a comprehensive coverage of developments in the field (by spanning twelve and three-quarter years). The automated search query used the following fields: article title, abstract, and keywords. Variations of the following search terms were used: "road" AND "construction" AND "predict" AND "machine learning". We omitted physical properties from the search query to prevent oversight of articles that examined physical properties not specified in the list of search terms. We were aware that the less specific search query would lead to more irrelevant records, however, it reduces the risk of overlooking relevant articles. The Appendix provides the full search terms and limiters used for each database.

To enhance the inclusivity of our review, we employed three other methods in addition to the automated search: supplementary search, author searching, and citation searching. In supplementary search, we identified articles on intelligent compaction to ensure a targeted exploration of this subfield. Author searching involved the retrieval of articles authored by the same individuals who had contributed to the already selected studies (after full-text assessment). Citation searching involved identifying articles that cited the selected studies.

2.3. Selection process

After removing duplicates from the database search, two review authors independently screened the titles, keywords, and abstracts of the records against the eligibility criteria using a standardized form. Disagreements were resolved through discussion. We retrieved the full texts of all records deemed potentially eligible, and the same review authors independently evaluated their eligibility. Again, inconsistent evaluations were resolved through discussion. The raw agreement between review authors was 95% in the screening stage and 86% in the eligibility stage.

2.4. Data collection process

Two review authors used a standardized form to extract data independently from each eligible study. The form asked to record data points for data items, which are defined in the following section. Any discrepancies in the data points were resolved through discussion.

2.5. Data items

Fig. 2 shows the conceptual framework by defining the ML process and indicating the major data items. This process begins with the prediction problem and subsequent phases for data collection, data preprocessing, model training, and model evaluation. We adopted the framework from previous systematic reviews of predictive modeling in industrial maintenance [15], crop production [16], and pasture management [17] and adjusted the definition of a few data items to the context of asphalt road construction. Prediction was defined as the process of estimating the current or future value of a physical property. The relevant physical properties can refer to objects such as materials processed in asphalt mixture plants, road layers from the soil

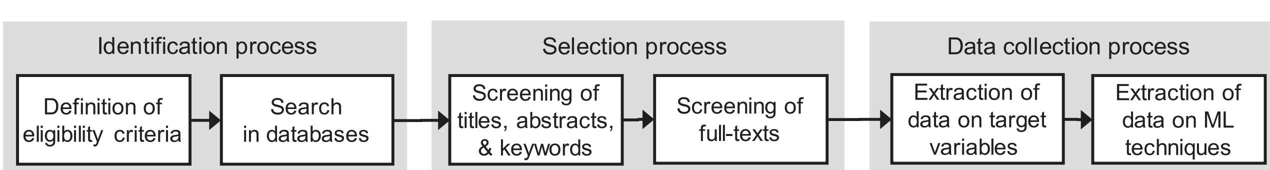

Fig. 1. Processes and activities for identification and selection of studies and data collection.

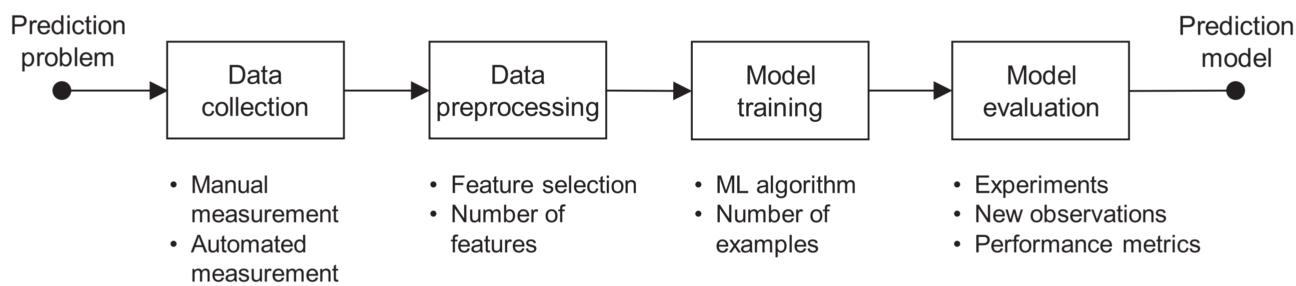

Fig. 2. Conceptual framework of the review.

underground up to the surface layer, and equipment used. Physical properties are usually continuous variables with a standard unit of measurement, such as resilient modulus expressed in megapascals (MPa).

Data collection is concerned with acquiring data that are representative for the variability of the application domain in which the prediction model should be used. This activity requires collecting true values of the predicted variable as well as of input variables that are deemed relevant. The collection of input variables can be carried out either manually, automatically, or in both ways. *Manual measurement* includes foremost laboratory tests (e.g., sieve analysis), whereas *automated measurement* relies upon sensors, such as accelerometers and ground penetrating radar.

The aim of data preprocessing is to transform raw data into a feature-based representation so that a prediction model can be trained and evaluated in subsequent phases. Features can either correspond directly to the input variables or be redefined on the basis of one or more input variables. Furthermore, input variables can be deliberately ignored, i.e., they do not become a feature. To determine the relevant features, *feature selection* techniques can be used [18]. For instance, pairs of strongly correlated features can be identified using correlation analysis in order to leave only one feature in the feature set (assuming that strong correlations indicate redundancy among features). Another technique is the iterative removal or inclusion of features in trained models to evaluate their relevance. Moreover, some ML algorithms already provide built-in feature selection. Regardless of the approach to feature definition, a specific set and *number of features* must be defined.

During model training, a function is learned that describes the relationship between features and the target variable. Learning takes place based on input-output pairs, also known as examples, i.e., pairs of observed values of the features and true values of the target variable. The function is determined using a supervised learning *algorithm*, such as ANN [19], Multiple Linear Regression (MLR), and Random Forests (RF) [20]. Although no exact rules are known for the minimum *number of examples* required, the performance of all algorithms depends largely on the number of examples used to train the model [21,22].

In the final phase, the prediction performance of the trained model must be evaluated. Model evaluation is often carried out not just for one particular model, but requires experimentation with different models. Thus, experimental manipulation of factors is required in order to compare alternative models and better understand the effects of decisions made during data collection, data preprocessing, and model training. Evaluation is defined as the application of the trained model to *new observations* and the measurement of the model's performance [23]. An observation is called new if it has not already been used for model training (also called unseen data). Three basic techniques are available that also can be used in combination: (1) Cross-validation (CV) splits the original dataset into multiple parts before the training begins, trains a model iteratively on different combinations of these parts, and evaluates each model on the left-out part to calculate the average performance. (2) Split-sample validation divides the original dataset into two parts, trains the model on the first part (often called 'training set'), and evaluates it on the second part (often called 'test set'). The split is usually given as a ratio, such as 70:30 or 80:20. A variant is to cut out a third part in order to first carry out a preliminary evaluation on a so-called validation set and then the final evaluation on a test set (e.g., using a ratio of 70:15:15). (3) *External validation* evaluates the model on data that is independent from the data used for training in order to demonstrate the generalizability of the model. Therefore, this new data must not be taken from the original dataset but intentionally be collected, for example, at a different construction site or time. Because cross-validation and split-sample validation use no data other than the original dataset, these techniques are also referred to as *internal validation*, i.e., performance is evaluated with the dataset at hand [24].

Regardless of the evaluation technique used, the performance is calculated by comparing predicted and observed values of the target variable, for which numerous *performance metrics* specific to continuous variables are available [25]. Metrics either have a unit, such as the root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE) [26,27], or are unitless, such as the coefficient of determination (R^2).

3. Results

3.1. Study selection

Fig. 3 shows the study selection in a PRISMA 2020 flow diagram. Following the deletion of duplicates, the search in Scopus and Web of Science yielded 1086 unique records. Based on the screening of titles, abstracts, and keywords, 77 articles were forwarded to the full-text assessment complemented with 21 additional articles identified via other methods. Of these 98 articles, 42 were excluded as they were not examining asphalt road construction, 6 were for predicting categorical variables, and 8 were due to techniques other than machine learning. Further, 12 articles were excluded because no data from construction sites were collected. In the end, 30 articles met the inclusion criteria and each reports an original study.

An overview of the studies according to the prediction object and predicted variable is shown in Table 1. Twenty-one studies targeted characteristics of the subgrade layer, including CBR ($n = 5$), compactness ($n = 4$), and other measures of load bearing capacity. The subbase layer was the subject of three studies, and properties of the asphalt layer were targeted in five studies. Most of the roads were highways ($n = 16$). One study each predicted the gas consumption of the asphalt mixing plant, the fuel consumption of trucks for earthworks, and the asphalt temperature during transportation from the plant to the construction site, respectively. Regarding the year of publication, more than two-thirds of the studies fall within the last three years of the period under review (2021: 4; 2022: 5; 2023: 13).

3.2. Data collection

Information on the different types of measurements for the input variables is presented in Table 2. The following groups of studies can be differentiated: In eleven studies, only manual measurements were carried out, in each case for several soil characteristics, such as distribution

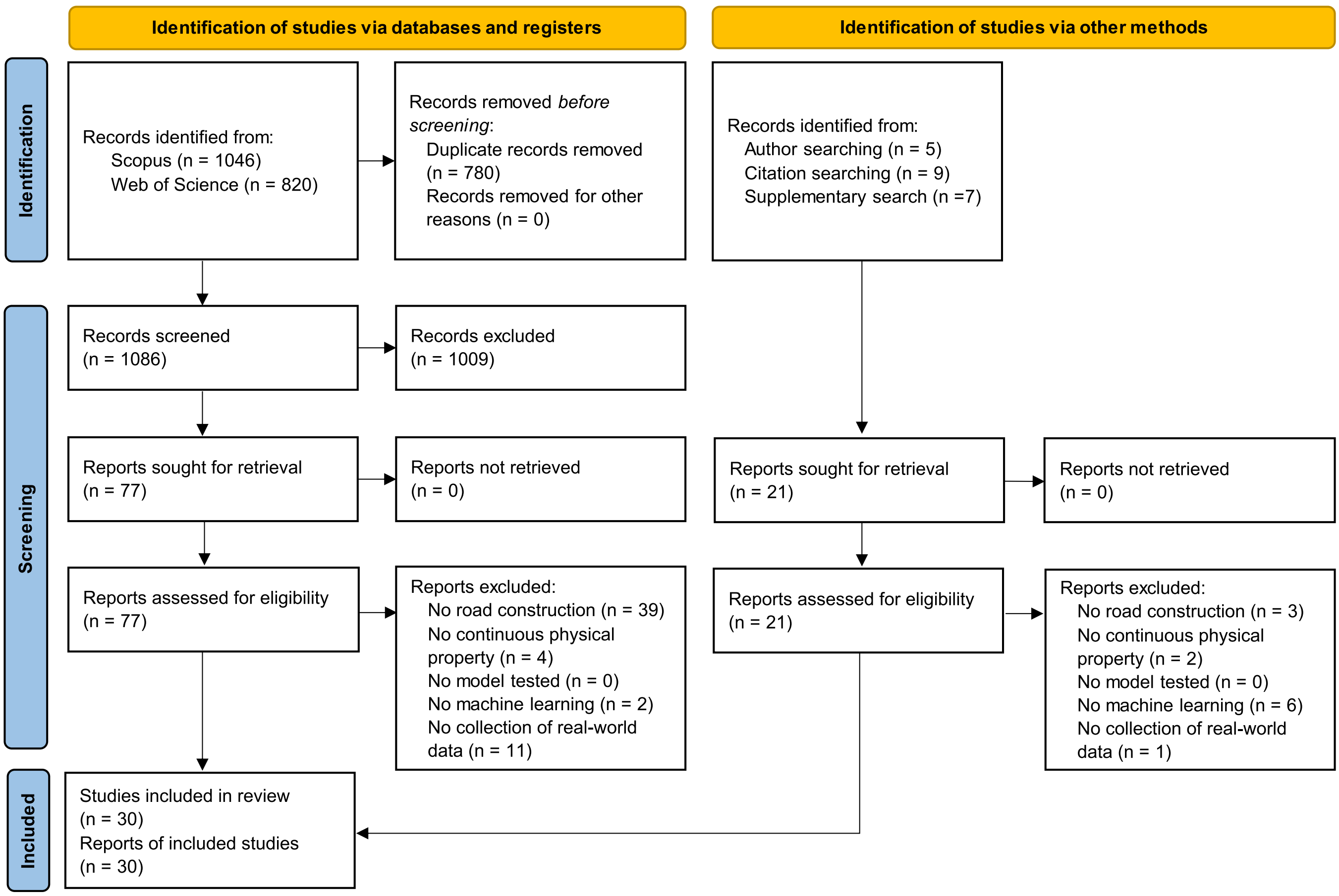

Fig. 3. PRISMA 2020 flow diagram of the study selection.

Table 1
Prediction objects and predicted variables in studies ($n = 30$).

Prediction object	Predicted variable (unit of measurement)	Studies
Subgrade layer	California Bearing Ratio	[28–32]
	Compactness (%)	[33–36]
	Density (%)	[37,38]
	Dynamic modulus (MPa)	[39]
	Foundation settlement (cm)	[40]
	Maximum dry density (g/cc)	[41,42,36]
	Optimum moisture content (%)	[41,42,36]
	Resilient modulus (MPa)	[43–45]
	Sheer strength (kPa)	[46]
	Subgrade modulus (MPa)	[33,47]
Subbase layer	Base modulus (MPa)	[47]
	California Bearing Ratio	[48,49]
Asphalt layer	Air voids (%)	[50]
	Asphalt temperature (K)	[51,52]
	Core density (%Gmm)	[53]
	Density (%)	[52]
	Density (g/cm^3)	[54]
	Roller amplitude (mm)	[52]
Asphalt plant	Gas consumption (m^3/t)	[55]
Truck	Fuel consumption (L)	[56]
Transportation	Asphalt temperature (°C)	[57]

of aggregate particles by size. Sixteen other studies only conducted automated measurements with sensors for acceleration ($n = 14$), position ($n = 13$), temperature ($n = 5$), subsurface properties ($n = 2$), etc. Three studies combined manual and automated measurements.

3.3. Data preprocessing

Information on the features derived from the collected data was available in 27 studies of which 14 provided feature definitions in table format. The largest prediction model included 13 features [29], whereas seven models had only three features ($M = 6.2$, $SD = 2.6$). Given the relatively few input variables, only every fifth study applied feature selection to reduce the number of features (Table 3). Based on correlation analysis, the reductions were from 12 to 5 [42], 6–3 [41], and 8–6 features [31], respectively; two other studies reported no details. One study adopted backward elimination for the MLR model to remove 5 of 13 initial features, whereas the ANN model retained all features [29].

3.4. Model training

Table 4 summarizes the adoption of ML algorithms by showing 11 different groups of algorithms. By far the most common group was ANN ($n = 21$), followed by MLR ($n = 7$) and SVR ($n = 6$). The number of specific algorithms per study was either one ($n = 18$), two ($n = 7$), three ($n = 2$), four ($n = 1$), six ($n = 1$), or seven ($n = 1$). Regarding the mapping of algorithms on predicted variables, no clear relationship emerged. For instance, ANN was applied to a broad range of variables, including modulus ($n = 5$), density ($n = 5$), compactness ($n = 3$), CBR ($n = 3$), and also air voids, fuel consumption, gas consumption, and optimum moisture content (each $n = 1$). Similarly, MLR and SVR were used for very different variables.

Seven studies indicated software tools that implemented the algorithms, such as *MATLAB* [40], and specific packages for programming languages, such as *Python* [32,33] and *R* [56]. The smallest training sets

Table 2
Data collection for input variables in studies ($n = 30$).

Study	Manual measurement	Automated measurement
[48]	Soil characteristics	–
[55]	Moisture content	Thermometer
[43]	–	Accelerometer, global positioning system
[44]	–	Accelerometer, global positioning system
[45]	–	Accelerometer, global positioning system
[51]	–	Ground penetrating radar
[37]	–	Accelerometer, global positioning system, infrared thermometer
[47]	–	Accelerometer, global positioning system
[28]	Soil characteristics	–
[54]	–	Ground penetrating radar
[49]	Soil characteristics	–
[29]	Soil characteristics	–
[33]	–	Accelerometer, global positioning system
[56]	–	Inertial measurement unit, global positioning system
[39]	–	Accelerometer, global positioning system
[30]	Soil characteristics	–
[31]	Soil characteristics	–
[41]	Soil characteristics	–
[42]	Soil characteristics	–
[32]	Soil characteristics	–
[46]	–	Accelerometer, global positioning system
[34]	Particle size	Accelerometer, global positioning system
[35]	–	Accelerometer, global positioning system
[36]	Soil characteristics	–
[57]	–	Infrared thermometer
[40]	Soil characteristics	–
[53]	–	Accelerometer, gyroscope, magnetometer, stress gauge, thermometer
[38]	–	Accelerometer, global positioning system
[52]	–	Accelerometer, global positioning system
[50]	Asphalt mixture characteristics	Infrared thermometer

Table 3
Feature selection methods in studies ($n = 30$).

Type	No. of studies	Studies
Correlation analysis	5	[31,32,41,42,46]
Backward elimination	1	[29]
None	24	All other

Table 4
Machine learning algorithms in studies ($n = 30$).

Algorithm (group)	No. of studies	Studies
Artificial Neural Networks	21	[28-30,33-37,39,41-47,50, 53-56]
Multiple Linear Regression	7	[28,29,33,44,45,48,56]
Support Vector Regression	6	[38,40,50,52,56,57]
Random Forests	4	[33,50,51,56]
Gradient Boosting	3	[33,51,57]
Gaussian Process Regression	2	[32,50]
k-Nearest Neighbor	2	[32,33]
Adaptive Neuro-Fuzzy Inference System	1	[30]
Genetic Programming	1	[31]
Kernel Ridge Regression	1	[32]
Regression Trees	1	[50]

included 19 [32], 25 [54], and 28 [45] examples, respectively, compared to seven training sets that had at least 250 examples [34,36, 39,46,47,50,57]. Overall, specific information on the training set was only available in 19 studies.

3.5. Model evaluation

This section reports on: (1) which factors were manipulated in the experiments, (2) how models were assessed using cross-validation, split-sample validation, and external validation, and (3) which unitless metrics and metrics with units were adopted in the evaluation.

3.5.1. Experimental manipulation

As shown in Table 5, ten studies compared the performance of at least two different ML algorithms. In six studies that tested ANN models, the performance of ANN models was higher compared to other models in three studies [28,29,50] and lower in two studies [30,56]; in one study, ANN models outperformed three other algorithms for predicting the subgrade modulus but their performance was lower for predicting the compaction degree [33]. In summary, evidence for higher performance of ANN models is ambiguous. The second most frequently manipulated factor was the feature set (eight studies). All other factors played a marginal role and this applied both to ML factors, such as structure of ANN and size of training sets, and study conditions, such as soil type and asphalt mixture. Taken together, the experimental designs were rather simple, by either examining one ($n = 10$), two ($n = 4$), or three ($n = 4$) factors, while 12 studies were limited to one condition.

3.5.2. Model assessment

Concerning the techniques for assessing prediction models, two studies exclusively adopted cross-validation [33,56] (Table 6). Twenty-three studies applied split-sample validation, with the test sets representing between 5 and 61 percent of the original dataset. Regarding the number of examples in the test set, exact figures were only reported in nine studies. For 11 studies, we were able to determine this number either based on percentage values or dots shown in scatter plots. Seven test sets had more than 100 examples, whereas five test sets included less than 15 examples. External validation was reported in ten studies of which seven studies collected data from a different (independent) road project, and three studies used data from a different section of the same road on which the model was trained. The number of examples was greater than 100 in two studies, but less than 15 in three studies.

3.5.3. Performance metrics

Table 7 summarizes which performance metrics were reported for the final prediction models and indicates their R^2 (if reported). The unitless metric R^2 and the relative metric MAPE were only present in 20 and 7 studies, respectively. Regarding prediction models for the subgrade modulus, R^2 ranged between 0.63 and 0.98 ($n = 5$), while the range for the subgrade CBR was 0.56–0.991 ($n = 5$). RMSE was the most frequently used metric with unit ($n = 16$), followed by MAE ($n = 10$) and MSE ($n = 7$). The number of metrics per study was either one ($n = 8$), two ($n = 10$), three ($n = 6$), four ($n = 4$), six ($n = 1$) or nine ($n = 1$).

Table 5
Experimental manipulation in studies ($n = 30$).

Factor	No. of studies	Studies
ML algorithm	10	[28-30,32,33,49-51,56,57]
Feature set	8	[28-31,34,42,47,55]
Soil type	3	[41,44,47]
ANN structure	3	[28,29,42]
Activation function	2	[29,49]
Additive type	1	[44]
ANN training algorithm	1	[28]
Asphalt mixture	1	[55]
Compacted layer thickness	1	[44]
Roller passes	1	[38]
Training set	1	[49]
No manipulation	12	[35-37,39,40,43,45,46,48,52-54]

Note. ANN = Artificial Neural Networks.

Table 6
Model assessment in studies (n = 30).

Study	Cross-validation Folds	Split-sample validation		External validation	
		Ratio	Examples	Type	Examples
[48]	–	39:61[(*)]	30	–	–
[55]	–	70:30	10–43	–	–
[43]	–	80:20	NR	–	–
[44]	–	80:20	63	–	–
[45]	–	–	–	Different project	10; 9
[51]	5	–	–	Different project	18
[37]	–	95:5	14	–	–
[47]	–	80:20	NR (>100)	Different project	NR (>100)
[28]	–	70:30	47[(*)]	–	–
[54]	–	NR	7	Different project	21
[49]	–	90:10; 80:20; 70:30	45; 30; 15[(*)]	–	–
[29]	–	60:20:20	15; 15	–	–
[33]	5	–	–	–	–
[56]	3 (10 iterations)	–	–	–	–
[39]	–	–	–	Different road section	4
[30]	–	95:5[(*)]	14	–	–
[31]	–	80:20	202	Different road section	69
[41]	–	80:20	50[(*)]	–	–
[42]	–	80:20	106[(*)]	Different road section	32
[32]	5	80:20	202[(*)]	–	–
[46]	–	89:11	186[(*)]	–	–
[34]	–	70:10:20	46[(*)]	–	–
[35]	–	80:20	35	–	–
[36]	–	70:15:15	1) 337; 2) 47; 3) 47[(*)]	–	–
[57]	–	80:20	NR (>100)	Different project	NR (>100)
[40]	NR	NR	9[(*)]	–	–
[53]	–	–	–	Different project	11; 2
[38]	–	80:20	9[(*)]	–	–
[52]	–	NR	16–30[(*)]	–	–
[50]	–	–	–	Different project	8–10

Note. NR = not reported. [(*)] = no exact figures reported.

Table 7
Performance metrics reported in studies (n = 30).

Study	R^2	MAPE	RMSE	MAE	MSE	Other
[48]	–	–	–	–	Yes	–
[55]	0.7638	–	–	–	–	–
[43]	0.89	–	–	–	–	–
[44]	0.63	–	Yes	–	–	–
[45]	0.63	–	–	–	–	–
[51]	–	Yes	–	–	–	–
[37]	0.791	–	–	–	–	–
[47]	1) 0.80; 2) 0.98	–	–	–	–	Standard error of estimate
[28]	–	–	Yes	–	Yes	–
[54]	–	–	–	Yes	–	Mean relative error
[49]	0.996	–	Yes	–	–	–
[29]	0.945	–	Yes	Yes	–	Maximum absolute error
[33]	1) 0.71; 2) 0.87	–	Yes	–	–	–
[56]	–	–	Yes	–	–	–
[39]	–	–	–	–	–	Relative error
[30]	0.991	–	Yes	Yes	–	–
[31]	0.56	–	Yes	Yes	–	Various
[41]	–	–	–	Yes	–	Maximum relative error
[42]	1) 0.84; 2) 0.78	–	Yes	Yes	–	–
[32]	0.746	Yes	Yes	Yes	Yes	Various
[46]	0.9108	Yes	Yes	–	–	–
[34]	0.902	Yes	Yes	–	–	–
[35]	–	–	–	–	–	Relative error
[36]	1) 0.61; 2) 0.73; 3) 0.72	Yes	Yes	–	–	–
[57]	0.9941	Yes	Yes	–	–	–
[40]	–	–	–	Yes	Yes	Residual sum of squares
[53]	–	Yes	Yes	–	–	–
[38]	0.76	–	–	Yes	Yes	Normalized RMSE by range
[52]	1) 0.7593; 2) 0.2201; 3) 0.732	–	–	–	Yes	–
[50]	0.8531	–	Yes	Yes	Yes	–

Note. MAE = mean absolute error. MAPE = mean absolute percentage error. MSE = mean square error. R^2 = coefficient of determination. RMSE = root mean square error.

4. Discussion

This systematic review determined the adoption of ML techniques in studies on the prediction of physical properties in asphalt road construction. The principal findings and implications are discussed in this section, followed by a discussion of the limitations of the review.

4.1. Key findings and implications for future research

4.1.1. Data collection

The thirty studies included in this review reflect the wide scope of construction activities supported by prediction models and the high level of performance of these models. Most studies focused on predictions to assist in the compaction of subgrade, subbase, and asphalt layers. This focus is consistent with the overarching goals of digitalization in asphalt road construction, such as the automation of laborious and time-critical tasks, the support of information-intensive decision-making, and the tracking and reporting of process states [58]. However, the construction process includes not only the activities carried out directly on the construction site but also the production of asphalt in mixture plants and their time-critical transportation to the site. Surprisingly, only a handful of studies dealt with predictions related to the production and transportation of asphalt. We suggest that the focus be broadened to include prediction problems that arise upstream of the construction site. For this purpose, new prediction models with a significantly longer prediction horizon need to be developed, for example, the time between departure from the mixing plant and filling of the asphalt paver.

It is noteworthy that more than one half of the studies collected all input data using automated measurements, thus no manual data collection by personnel was required. These studies employed sensors for acceleration and position of rollers and trucks, temperature of asphalt, or subsurface properties; many of these sensors were commercial products or have been developed in previous research in intelligent compaction technology [59–61]. In this case, the predictions made can be used in real time to inform machine operators or control the machine [62]. Although the increasing use of automated measurement techniques simplifies data collection for ML models and thus benefits the researchers involved, we did not find that such data is more frequently made available to third parties. No study provided a publicly accessible dataset and only every third study stated that the dataset would be available on request. Overall, data availability is very low compared to other disciplines [63,64]. Enhancing open data practices in road construction could greatly facilitate cumulative research, allowing for the reuse and refinement of existing datasets to accelerate development of ML prediction models and improve their accuracy and robustness.

4.1.2. Data preprocessing

Our review reveals that the number of input variables and features was rather small (the models in half of the studies included between three and six features. An often-emphasized characteristic of ML is the ability to train prediction models for high-dimensional input data [65, 66]. However, this requirement does not seem to arise in predicting physical properties in asphalt road construction. The reason can be seen in the fact that most prediction models were specified in advance, based on established relationships between input variables and the target variable. Given the low number of input variables, determining a minimum set of relevant features from a larger set of candidate features assumes minor relevance. At the same time, our results highlight the need for every study, regardless of the number of features, to clearly report each feature by name and unit of measurement, and specify its distribution parameters. A good example is the study by Kurnaz et al. [28], in which a table shows seven features and the outcome variable together with their descriptive statistics (mean, standard deviation, minimum and maximum value). This form of extended reporting would make it easier for follow-up studies to understand and reproduce the feature sets and compare them with feature sets developed in these works.

4.1.3. Model training

An important finding of our review is the focus of the studies on ANN. This group of algorithms was used in 21 studies and even listed in the title of 9 articles. All other algorithms played a marginal role, including MLR, SVR, and RF, which are much more frequently used in other fields. Given the ambiguous evidence provided by the few studies that compared ANN with alternative algorithms, the focus on ANN was surprising to us. In fact, only 3 out of 6 comparison studies showed higher performance for ANN. We suggest that future research rigorously evaluates the performance of ANN compared to alternative algorithms. These evaluations are required for prediction problems in asphalt road construction because every algorithm's performance is highly contingent on the characteristics of the data used and the specific task it is applied to. In other words, no ML algorithm is a priori superior to another [67].

In view of the small number of features in the models examined, it is understandable that newer algorithms such as Convolutional Neural Networks (CNNs) were not used, as CNNs are designed for high-dimensional data [68]. However, our review highlights a promising trajectory for future research: the adoption of Physics-Informed Neural Networks (PINNs), which allow to incorporate the underlying physical laws into the learning process [69]. PINNs can be particularly useful for complex engineering tasks such as modeling in road construction and ensure that the prediction models are physically plausible. Future research should explore the integration of PINNs in prediction models and experimentally assess these models under varied environmental conditions to substantiate their applicability.

Detailed information on the data used for model training can help readers comprehend the steps taken, assess generalizability beyond the study setting, and make decisions for their own models. An important issue is the number of examples required to train an effective model. It is unfortunate that one-thirds of the studies did not report the number of examples used in the training, and only one study examined how additional training examples affect prediction performance [49]. This finding implies that future research is needed to investigate the impact of the number of training examples on performance, and then to estimate what minimum number is required to reach a predefined level of accuracy and generalizability.

4.1.4. Model evaluation

The performance of a prediction model must be examined in a rigorous evaluation by applying the model to data that allow conclusions to be drawn about its generalizability. For instance, a model for predicting the density of asphalt layers trained from data for one road using specific asphalt mixtures might perform worse, equally well, or better when applied to a different road, even if the asphalt mixtures are similar. The results of our review show that previous studies have taken very different paths by adopting a broad spectrum of evaluation techniques and reporting different sets of metrics. It should be highlighted that every third study applied the model to data from a different project or different road section, thus deliberately assessed the generalizability of the model beyond the original dataset. This type of evaluation is also the most complex as it requires additional data collection in the field.

Regarding the number of examples to which the models were applied, the studies differed greatly (from less than 15 to more than 100 examples). Given the high importance of the number of examples for the interpretation of evaluation results, we recommend that studies report this number clearly in the method and results sections. This information is as important as the number of training examples and can effectively be reported in a single sentence.

With respect to the reporting of performance metrics, the results of our review uncover an emphasis on metrics with units but a lack of unitless metrics. Metrics with units such as RMSE, MAE, and MAPE are particularly informative for practitioners as they retain the unit of the target variable (e.g., compaction degree in percentage). Unfortunately, the RMSE was only reported in about half of the studies and the MAE in less than one-thirds. In combination with scatter plots, these metrics provide information about the location and spread of the prediction error and thus help ascertain the practical usefulness of the model. However, they are also contingent upon the variability of the target variable, so that the metrics might be (very) different based on soils, asphalt mixtures, construction vehicles, processes, etc. Thus, metrics with units should be interpreted carefully when comparing different studies. Against this backdrop, unitless metrics become important as they allow comparison and integration of results from different studies. In particular, the R^2 represents the proportion of the variance in the target variable explained by the model; hence, its value is normalized by the variance of the target variable. Nevertheless, only two-thirds of the studies reported the R^2, and only every tenth abstract indicated the R^2 obtained in the final evaluation [30,44,49]. Another useful unitless metric, although not reported in any study, is the normalized RMSE (NRMSE), defined as RMSE divided by the standard deviation of the observed values. NRMSE gives the error measured in standard deviations, so that one can say, for example, the error is about one-third (NRMSE = 0.33) or one-fourth (NRMSE = 0.25) standard deviations.

Future research should view the two types of metrics discussed above as complements to each other, with each type being informative for a specific purpose [70]. A comparison of the evidence from different studies is not possible based on metrics with units, as long as the studies do not use the same data sets in the evaluation. In contrast, the unitless metrics R^2 and NRMSE (based on standard deviations) consider differences in the variability of the target variable and are thus more useful for the comparison of study results. Collectively, we suggest that studies report a wide range of metrics, including at least RMSE, MAE, MAPE, R^2 and NRMSE, and complement the metrics by scatter plots. Future studies can prevent misinterpretation of their evaluation results by defining each metric used explicitly (i.e., by formula). Because metrics with units such as RMSE and MAE cannot be appropriately interpreted if the variability of the target variable is not indicated at the same time, we suggest that studies always report the R^2 and the NRMSE (i.e., RMSE divided by SD) in the abstract, rather than limit the reporting to metrics with units.

4.2. Limitations

The findings of the review should be understood in light of the following limitations. Our data extraction was constrained by the level of detail reported in the studies. The studies indeed differed in that regard, so that incomplete reporting was found for many data items, which somehow limited the possibilities for qualitative synthesis. Additionally, the lack of established guidelines for assessing the risk of bias in ML studies made it difficult to consider the quality of studies in the overall conclusions drawn from the review. Furthermore, we cannot rule out that publication bias might be present in the literature being examined so that positive results would be more frequently published, e.g., in relation to certain algorithms. Finally, the diversity of the physical properties reduced the number of models whose performance could be directly compared based on shared metrics.

5. Conclusion

Development of ML prediction models for physical properties in asphalt road construction is rapid and evidence for the usefulness of such models has enhanced. This systematic review determined the methodological conduct for data collection, preprocessing, training, and evaluation, and thus provides a comprehensive picture of the state of the field. To enhance generalizability of prediction models and improve rigor, our results highlight three main areas. First, the evidence basis for algorithm selection for predicting physical properties in asphalt road construction should be expanded. Specifically, experimental studies are needed to assess the performance of alternate ML algorithms, which have been found to be viable alternatives to ANN in other fields. Second, although many studies reported remarkable levels of accuracies, integration of evidence is hindered by a lack of consensus on the performance metrics. Greater comprehensiveness and uniformity in the assessment of prediction performance is required, for which we submit specific recommendations. Third, a major challenge for the development of prediction models is to know how the number of training examples impacts accuracy and generalizability. Controlled ML experiments are recommended to examine how additional training examples affect performance and help understand the robustness of models.

Funding

This work was supported by the Ministry of Economic Affairs, Labor and Tourism Baden-Württemberg [grant: BW1_3049], Germany, as part of the collaborative research project “Artificial Intelligence for Digital and Sustainable Road Construction”.

CRediT authorship contribution statement

Vijayan Sugumaran: Writing – review & editing, Conceptualization. **Joerg Leukel:** Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization. **Luca Scheurer:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Appendix

Database search in Scopus (2023, October 22)	
Search terms	Records found
(("road" OR "asphalt*" OR "pavement" OR "highway*") AND ("construction" OR "project*" OR "test road*" OR "asphalt mixture*" OR "pavement engineer*") AND ("predict*" OR "estimat*" OR "forecast*" OR "determin*") AND ("machine learning" OR "random forest*" OR "support vector" OR "artificial intelligence" OR "neural network*")) Limiters: Year: From 2011 Document type: Article Source type: Journal Language: English	1046

Note. Database search within Article title, Abstract, and Keywords.

Database search in Web of Science (2023, September 4)	
Search terms	Records found
(("road" OR "asphalt*" OR "pavement" OR "highway*") AND ("construction" OR "project*" OR "test road*" OR "asphalt mixture*" OR "pavement engineer*") AND ("predict*" OR "estimat*" OR "forecast*" OR "determin*") AND ("machine learning" OR "random forest*" OR "support vector" OR "artificial intelligence" OR "neural network*")) Limiters: Publication years: From 2011–2023 Document types: Early Access or Article Languages: English	820

Note. Database search within Title, Abstract, Author Keywords, and Keywords Plus.

References

[1] C. Han, T. Han, T. Ma, Z. Tong, S. Wang, A BIM-based framework for road construction quality control and quality assurance, Int. J. Pavement Eng. 24 (2023) 2209903, https://doi.org/10.1080/10298436.2023.2209903.

[2] R. Kuenzel, J. Teizer, M. Mueller, A. Blickle, SmartSite: intelligent and autonomous environments, machinery, and processes to realize smart road construction projects, Autom. Constr. 71 (2016) 21–33, https://doi.org/10.1016/j.autcon.2016.03.012.

[3] S. Sivagnanasuntharam, A. Sounthararajah, J. Ghorbani, D. Bodin, J. Kodikara, A state-of-the-art review of compaction control test methods and intelligent compaction technology for asphalt pavements, Road. Mater. Pavement Des. 24 (2023) 1–30, https://doi.org/10.1080/14680629.2021.2015423.

[4] M. Livneh, Y. Goldberg, Quality assessment during road formation and foundation construction: use of falling-weight deflectometer and light drop weight, Transp. Res. Rec. 1755 (2001) 69–77, https://doi.org/10.3141/1755-08.

[5] W. van den Bergh, C. Vuye, P. Kara, K. Couscheir, J. Blom, P. van Bouwel, The use of a non-nuclear density gauge for monitoring the compaction process of asphalt pavement, IOP Conf. Ser.: Mater. Sci. Eng. 236 (2017) 12014, https://doi.org/10.1088/1757-899X/236/1/012014.

[6] D. Liu, Y. Wang, J. Chen, Y. Zhang, Intelligent compaction practice and development: a bibliometric analysis, Eng., Constr. Archit. Manag. 27 (2020) 1213–1232, https://doi.org/10.1108/ECAM-05-2019-0252.

[7] Y. Zhan, Y. Zhang, Z. Nie, Z. Luo, S. Qiu, J. Wang, A.A. Zhang, C. Ai, X. Tang, C. Tan, Intelligent paving and compaction technologies for asphalt pavement, Autom. Constr. 156 (2023) 105081, https://doi.org/10.1016/j.autcon.2023.105081.
[8] S. Cano-Ortiz, P. Pascual-Muñoz, D. Castro-Fresno, Machine learning algorithms for monitoring pavement performance, Autom. Constr. 139 (2022) 104309, https://doi.org/10.1016/j.autcon.2022.104309.
[9] R. Justo-Silva, A. Ferreira, G. Flintsch, Review on machine learning techniques for developing pavement performance prediction models, Sustainability 13 (2021) 5248, https://doi.org/10.3390/su13095248.
[10] N. Sholevar, A. Golroo, S.R. Esfahani, Machine learning techniques for pavement condition evaluation, Autom. Constr. 136 (2022) 104190, https://doi.org/10.1016/j.autcon.2022.104190.
[11] X. Yang, J. Guan, L. Ding, Z. You, V.C. Lee, M.R. Mohd Hasan, X. Cheng, Research and applications of artificial neural network in pavement engineering: a state-of-the-art review, J. Traffic Transp. Eng. 8 (2021) 1000–1021, https://doi.org/10.1016/j.jtte.2021.03.005.
[12] M.J. Page, D. Moher, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, J.E. McKenzie, PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews, BMJ 372 (2021) n160, https://doi.org/10.1136/bmj.n160.
[13] P. Mongeon, A. Paul-Hus, The journal coverage of Web of Science and Scopus: a comparative analysis, Scientometrics 106 (2016) 213–228, https://doi.org/10.1007/s11192-015-1765-5.
[14] V.K. Singh, P. Singh, M. Karmakar, J. Leta, P. Mayr, The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis, Scientometrics 126 (2021) 5113–5142, https://doi.org/10.1007/s11192-021-03948-5.
[15] J. Leukel, J. González, M. Riekert, Adoption of machine learning technology for failure prediction in industrial maintenance: a systematic review, J. Manuf. Syst. 61 (2021) 87–96, https://doi.org/10.1016/j.jmsy.2021.08.012.
[16] J. Leukel, T. Zimpel, C. Stumpe, Machine learning technology for early prediction of grain yield at the field scale: a systematic review, Comput. Electron. Agric. 207 (2023) 107721, https://doi.org/10.1016/j.compag.2023.107721.
[17] C. Stumpe, J. Leukel, T. Zimpel, Prediction of pasture yield using machine learning-based optical sensing: a systematic review, Precis, Agric (2024) 430–459, https://doi.org/10.1007/s11119-023-10079-9.
[18] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. (2003) 1157–1182.
[19] C.M. Bishop, Pattern recognition and machine learning, Springer, New York, 2006.
[20] T.K. Ho, Random decision forests. in: Proceedings of 3rd International Conference on Document Analysis and Recognition (ICDAR), IEEE,, 1995, pp. 278–282.
[21] A. L'Heureux, K. Grolinger, H.F. Elyamany, M.A.M. Capretz, Machine learning with big data: challenges and approaches, IEEE Access 5 (2017) 7776–7797, https://doi.org/10.1109/ACCESS.2017.2696365.
[22] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine learning algorithm validation with a limited sample size, PLoS ONE 14 (2019) e0224365, https://doi.org/10.1371/journal.pone.0224365.
[23] G. Shmueli, To explain or to predict? Stat. Sci. 25 (2010) https://doi.org/10.1214/10-STS330.
[24] E.W. Steyerberg, F.E. Harrell, Prediction models need appropriate internal, internal-external, and external validation, J. Clin. Epidemiol. 69 (2016) 245–247, https://doi.org/10.1016/j.jclinepi.2015.04.005.
[25] L.B. Sheiner, S.L. Beal, Some suggestions for measuring predictive performance, J. Pharmacokinet. Biopharm. 9 (1981) 503–512, https://doi.org/10.1007/BF01060893.
[26] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature, Geosci. Model Dev. 7 (2014) 1247–1250, https://doi.org/10.5194/gmd-7-1247-2014.
[27] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82, https://doi.org/10.3354/cr030079.
[28] F.T. Kurnaz, Y. Kaya, Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network, Eur. Phys. J. 134 (2019), https://doi.org/10.1140/epjp/i2019-12692-0.
[29] K. Othman, H. Abdelwahab, The application of deep neural networks for the prediction of California Bearing Ratio of road subgrade soil, Ain Shams Eng. J. 14 (2023) 101988, https://doi.org/10.1016/j.asej.2022.101988.
[30] T. Varol, H.B. Ozel, M. Ertugrul, T. Emir, M. Tunay, M. Cetin, H. Sevik, Prediction of soil-bearing capacity on forest roads by statistical approaches, Environ. Monit. Assess. 193 (2021) 527, https://doi.org/10.1007/s10661-021-09335-0.
[31] G. Verma, B. Kumar, Application of multi-expression programming (MEP) in predicting the soaked California bearing ratio (CBR) value of fine-grained soil, Innov. Infrastruct. Solut. 7 (2022), https://doi.org/10.1007/s41062-022-00858-0.
[32] G. Verma, B. Kumar, C. Kumar, A. Ray, M. Khandelwal, Application of KRR, K-NN and GPR algorithms for predicting the soaked CBR of fine-grained plastic soils, Arab. J. Sci. Eng. (2023), https://doi.org/10.1007/s13369-023-07962-y.
[33] J. Pang, J. Yang, B. Zhu, J. Qian, Study of regression algorithms and influent factors between intelligent compaction measurement values and in-situ measurement values, Appl. Sci. 13 (2023) 5953, https://doi.org/10.3390/app13105953.
[34] X. Wang, C. Cheng, J. Li, J. Zhang, G. Ma, J. Jin, Automated monitoring and evaluation of highway subgrade compaction quality using artificial neural networks, Autom. Constr. 145 (2023) 104663, https://doi.org/10.1016/j.autcon.2022.104663.
[35] X. Wang, C. Cheng, J. Zhang, G. Ma, J. Li, J. Jin, Real-time monitoring and quality assessment of subgrade compaction: key factors and ANN model, Acta Geotech. 18 (2023) 3349–3366, https://doi.org/10.1007/s11440-022-01769-1.
[36] X. Wang, X. Dong, J. Li, Z. Zhang, J. Zhang, G. Ma, Developing an advanced ANN-based approach to estimate compaction characteristics of highway subgrade, Adv. Eng. Inform. 56 (2023) 102023, https://doi.org/10.1016/j.aei.2023.102023.
[37] S. Commuri, A.T. Mai, M. Zaman, Neural network-based intelligent compaction analyzer for estimating compaction quality of hot asphalt mixes, J. Constr. Eng. Manag. 137 (2011) 634–644, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000343.
[38] W. Zhang, A. Raza Khan, S. Yoon, J. Lee, R. Zhang, K. Zeng, Investigation of the correlations between the field pavement in-place density and the intelligent compaction measure value (ICMV) of asphalt layers, Constr. Build. Mater. 292 (2021) 123439, https://doi.org/10.1016/j.conbuildmat.2021.123439.
[39] D. Singh, F.N. Beainy, A.T.T. Mai, S. Commuri, M. Zaman, In-situ assessment of stiffness during the construction of HMA pavements, Int. J. Pavement Res. 4 (2011) 131–139.
[40] R. Yang, S. Yuan, Monitoring and prediction of highway foundation settlement based on particle swarm optimization and support vector machine, Math. Probl. Eng. 2022 (2022) 1–8, https://doi.org/10.1155/2022/2754965.
[41] G. Verma, B. Kumar, Multi-layer perceptron (MLP) neural network for predicting the modified compaction parameters of coarse-grained and fine-grained soils, Innov. Infrastruct. Solut. 7 (2022), https://doi.org/10.1007/s41062-021-00679-7.
[42] G. Verma, B. Kumar, Artificial neural network equations for predicting the modified Proctor compaction parameters of fine-grained soil, Transp. Infrastruct. Geotechnol. 10 (2023) 424–447, https://doi.org/10.1007/s40515-022-00228-4.
[43] S. Asif Imran, M. Barman, S. Commuri, M. Zaman, M. Nazari, Artificial neural network–based intelligent compaction analyzer for real-time estimation of subgrade quality, Int. J. Geomech. 18 (2018) 04018048, https://doi.org/10.1061/(ASCE)GM.1943-5622.0001089.
[44] M. Barman, M. Nazari, S.A. Imran, S. Commuri, M. Zaman, Quality improvement of subgrade through intelligent compaction, Transp. Res. Rec.: J. Transp. Res. Board 2579 (2016) 59–69, https://doi.org/10.3141/2579-07.
[45] M. Barman, M. Nazari, S.A. Imran, S. Commuri, M. Zaman, F. Beainy, D. Singh, Quality control of subgrade soil using intelligent compaction, Innov. Infrastruct. Solut. 1 (2016), https://doi.org/10.1007/s41062-016-0020-0.
[46] X. Wang, X. Dong, Z. Zhang, J. Zhang, G. Ma, X. Yang, Compaction quality evaluation of subgrade based on soil characteristics assessment using machine learning, Transp. Geotech. 32 (2022) 100703, https://doi.org/10.1016/j.trgeo.2021.100703.
[47] A. Fathi, C. Tirado, S. Rocha, M. Mazari, S. Nazarian, A machine-learning approach for extracting modulus of compacted unbound aggregate base and subgrade materials using intelligent compaction technology, Infrastructures 6 (2021) 142, https://doi.org/10.3390/infrastructures6100142.
[48] M. Alawi, M. Rajab, Prediction of California bearing ratio of subbase layer using multiple linear regression models, Road. Mater. Pavement Des. 14 (2013) 211–219, https://doi.org/10.1080/14680629.2012.757557.
[49] T.V. Nagaraju, A. Bahrami, C.D. Prasad, S. Mantena, M. Biswal, M.R. Islam, Predicting California Bearing Ratio of lateritic soils using hybrid machine learning technique, Buildings 13 (2023) 255, https://doi.org/10.3390/buildings13010255.
[50] Y. Zhao, K. Zhang, Y. Zhang, Y. Luo, S. Wang, Prediction of air voids of asphalt layers by intelligent algorithm, Constr. Build. Mater. 317 (2022) 125908, https://doi.org/10.1016/j.conbuildmat.2021.125908.
[51] Y. Chen, F. Li, S. Zhou, X. Zhang, S. Zhang, Q. Zhang, Y. Su, Bayesian optimization based random forest and extreme gradient boosting for the pavement density prediction in GPR detection, Constr. Build. Mater. 387 (2023) 131564, https://doi.org/10.1016/j.conbuildmat.2023.131564.
[52] W. Zhang, K.N. Ahmad, Z. Tong, Z. Hu, H. Wang, M. Wu, K. Zhao, S. Yang, H. Farooq, L.N. Mohammad, In-time density monitoring of in-place asphalt layer construction via intelligent compaction technology, J. Mater. Civ. Eng. 35 (2023) 04022386, https://doi.org/10.1061/(ASCE)MT.1943-5533.0004558.
[53] S. Yu, S. Shen, M. Lu, Data sensing and compaction condition modeling for asphalt pavements, Autom. Constr. 154 (2023) 105021, https://doi.org/10.1016/j.autcon.2023.105021.
[54] Y. Meng, S. Mo, H. Li, T. Yan, J. Lei, L. Fan, Non-destructive prediction techniques for asphalt mixture based on back propagation neural networks, Int. J. Pavement Eng. 24 (2023) 2253965, https://doi.org/10.1080/10298436.2023.2253965.
[55] I. Androjić, Z. Dolaček-Alduk, Artificial neural network model for forecasting energy consumption in hot mix asphalt (HMA) production, Constr. Build. Mater. 170 (2018) 424–432, https://doi.org/10.1016/j.conbuildmat.2018.03.086.
[56] G. Pereira, M. Parente, J. Moutinho, M. Sampaio, Fuel consumption prediction for construction trucks: a noninvasive approach using dedicated sensors and machine learning, Infrastructures 6 (2021) 157, https://doi.org/10.3390/infrastructures6110157.
[57] X. Wang, P. Pan, J. Li, Real-time measurement on dynamic temperature variation of asphalt pavement using machine learning, Measurement 207 (2023) 112413, https://doi.org/10.1016/j.measurement.2022.112413.
[58] L.G. Sategna, D. Meinero, M. Volontà, Digitalising the Construction Sector: Unlocking the potential of data with a value chain approach, Committee for European Construction Equipment, Brussels, 2019.
[59] S. Commuri, M. Zaman, A novel neural network-based asphalt compaction analyzer, Int. J. Pavement Eng. 9 (2008) 177–188, https://doi.org/10.1080/10298430701232018.

[60] H.-C. Dan, D. Yang, X. Liu, A.-P. Peng, Z. Zhang, Experimental investigation on dynamic response of asphalt pavement using SmartRock sensor under vibrating compaction loading, Constr. Build. Mater. 247 (2020) 118592, https://doi.org/10.1016/j.conbuildmat.2020.118592.

[61] C. Plati, A. Loizos, K. Gkyrtis, Assessment of modern roadways using non-destructive geophysical surveying techniques, Surv. Geophys. 41 (2020) 395–430, https://doi.org/10.1007/s10712-019-09518-y.

[62] D. Liu, M. Lin, S. Li, Real-time quality monitoring and control of highway compaction, Autom. Constr. 62 (2016) 114–123, https://doi.org/10.1016/j.autcon.2015.11.007.

[63] C. Jiao, K. Li, Z. Fang, Data sharing practices across knowledge domains: A dynamic examination of data availability statements in PLOS ONE publications, 016555152211018, J. Inf. Sci. (2022), https://doi.org/10.1177/01655515221101830.

[64] L. Tedersoo, R. Küngas, E. Oras, K. Köster, H. Eenmaa, Ä. Leijen, M. Pedaste, M. Raju, A. Astapova, H. Lukner, K. Kogermann, T. Sepp, Data sharing practices and data availability upon request differ across scientific disciplines, Sci. Data 8 (2021) 192, https://doi.org/10.1038/s41597-021-00981-0.

[65] L. Gao, J. Song, X. Liu, J. Shao, J. Liu, J. Shao, Learning in high-dimensional multimedia data: the state of the art, Multimed. Syst. 23 (2017) 303–313, https://doi.org/10.1007/s00530-015-0494-1.

[66] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection, ACM Comput. Surv. 50 (2018) 1–45, https://doi.org/10.1145/3136625.

[67] D.H. Wolpert, The supervised learning no-free-lunch theorems, in: R. Roy, M. Köppen, S. Ovaska, T. Furuhashi, F. Hoffmann (Eds.), Soft Computing and Industry, Springer London,, 2002, pp. 25–42.

[68] S. Sony, K. Dunphy, A. Sadhu, M. Capretz, A systematic review of convolutional neural network-based structural condition assessment techniques, Eng. Struct. 226 (2021) 111347, https://doi.org/10.1016/j.engstruct.2020.111347.

[69] S. Cuomo, V.S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, F. Piccialli, Scientific machine learning through physics–informed neural networks: where we are and what's next, J. Sci. Comput. 92 (2022), https://doi.org/10.1007/s10915-022-01939-z.

[70] D.L.J. Alexander, A. Tropsha, D.A. Winkler, Beware of R(2): simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models, J. Chem. Inf. Model. 55 (2015) 1316–1322, https://doi.org/10.1021/acs.jcim.5b00206.