

Article

Stiffness Moduli Modelling and Prediction in Four-Point Bending of Asphalt Mixtures: A Machine Learning-Based Framework

Nicola Baldo ^{1,*}, Fabio Rondinella ¹, Fabiola Daneluz ¹, Pavla Vacková ², Jan Valentin ²,
Marcin D. Gajewski ³ and Jan B. Król ³

- ¹ Polytechnic Department of Engineering and Architecture (DPIA), University of Udine, Via del Cotonificio 114, 33100 Udine, Italy; fabio.rondinella@uniud.it (F.R.); fabiola.daneluz@uniud.it (F.D.)
² Faculty of Civil Engineering, Czech Technical University, Thákurova 7, 166 29 Prague, Czech Republic; pavla.vackova@fsv.cvut.cz (P.V.); jan.valentin@fsv.cvut.cz (J.V.)
³ Faculty of Civil Engineering, Warsaw University of Technology, 00-637 Warsaw, Poland; marcin.gajewski@pw.edu.pl (M.D.G.); jan.krol@pw.edu.pl (J.B.K.)
* Correspondence: nicola.baldo@uniud.it; Tel.: +39-0432-558-745

Abstract: Stiffness modulus represents one of the most important parameters for the mechanical characterization of asphalt mixtures (AMs). At the same time, it is a crucial input parameter in the process of designing flexible pavements. In the present study, two selected mixtures were thoroughly investigated in an experimental trial carried out by means of a four-point bending test (4PBT) apparatus. The mixtures were prepared using spilite aggregate, a conventional 50/70 penetration grade bitumen, and limestone filler. Their stiffness moduli (SM) were determined while samples were exposed to 11 loading frequencies (from 0.1 to 50 Hz) and 4 testing temperatures (from 0 to 30 °C). The SM values ranged from 1222 to 24,133 MPa. Observations were recorded and used to develop a machine learning (ML) model. The main scope was the prediction of the stiffness moduli based on the volumetric properties and testing conditions of the corresponding mixtures, which would provide the advantage of reducing the laboratory efforts required to determine them. Two of the main soft computing techniques were investigated to accomplish this task, namely decision trees with the Categorical Boosting algorithm and artificial neural networks. The outcomes suggest that both ML methodologies achieved very good results, with Categorical Boosting showing better performance (MAPE = 3.41% and $R^2 = 0.9968$) and resulting in more accurate and reliable predictions in terms of the six goodness-of-fit metrics that were implemented.

Keywords: stiffness modulus; asphalt mixture; machine learning; categorical boosting; artificial neural network



Citation: Baldo, N.; Rondinella, F.; Daneluz, F.; Vacková, P.; Valentin, J.; Gajewski, M.D.; Król, J.B. Stiffness Moduli Modelling and Prediction in Four-Point Bending of Asphalt Mixtures: A Machine Learning-Based Framework. *CivilEng* **2023**, *4*, 1083–1097. <https://doi.org/10.3390/civileng4040059>

Academic Editors: Angelo Luongo and Francesco D'Annibale

Received: 30 August 2023
Revised: 5 October 2023
Accepted: 8 October 2023
Published: 16 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the civil engineering domain, transportation infrastructure represents one of the major application fields, and flexible pavements prepared with asphalt concretes are still the most used technological solutions in every road network all over the world [1]. To evaluate whether an asphalt mixture (AM) is suitable for civil engineering applications, road agencies charged with pavement maintenance and construction often refer to performance indicators related to the mechanical behavior of the mixture under consideration. A failure to reach the standard thresholds related to the parameters considered could result in unsuitable mechanical behavior, causing the formation of typical fatigue or low-temperature cracks that would inevitably reduce the service life of the pavement. Serviceability would then be compromised, thus causing serious safety-related issues for every road user. For this reason, it is crucial to properly characterize the mechanical behavior of asphalt mixtures from a performance-based perspective [2–4].

The conventionally adopted approach is experimental: it allows the mechanical performances of the investigated AMs to be accurately evaluated, but it is time-consuming, and very expensive laboratory tests are often required. Furthermore, this approach has other drawbacks, such as (i) the need to repeat the experimental campaign when any compositional variable of the investigated mixture changes, with resulting impacts on time and costs; and (ii) the presence of skilled laboratory technicians that must be familiar with all the test protocols [5–10].

To overcome these drawbacks, the efforts of the scientific community have been focused on the development of mathematical models that allow each parameter involved in the mechanical characterization of a mixture to be individually handled. This achievement involved identifying complex constitutive equations that, embedded within specialized materials mechanics software, allow the performance parameters of a mixture to be accurately predicted [11–18]. These advanced constitutive models provide a detailed understanding of the mechanical responses of bituminous mixtures, coupled with useful methodologies for pavement monitoring [19].

However, in recent years, alternative non-physically-based models which use soft computing techniques have gained popularity within the academic community. Unlike constitutive equations, these machine learning (ML) models do not depend on the nature of a problem, and they possess the outstanding capability of being able to model even very different phenomena [20]. They do not require a priori knowledge of the relationships between inputs and their corresponding outputs, but they still allow accurate, fast, and reliable predictions to be produced [21–24].

Some of the most commonly used soft computing techniques are based on artificial neural networks (ANNs) or decision trees (DTs). The functioning of the former is intended to mimic that of the human biological nervous system by using simple elementary units named artificial neurons. Being highly interconnected and organized in successive layers, artificial neurons allow neural networks to model even highly nonlinear phenomena by producing both fast and accurate predictions in terms of several output variables, namely permeability, interface shear stiffness modulus, compressive strength, stress at failure, and so on [25–34]. However, the difficulty of interpreting neural models coupled with the challenges associated with the optimization of the multiple hyperparameters involved [35] make them somewhat difficult to handle.

In contrast, simple decision rules allow DT-based algorithms to produce equally accurate predictions [36,37], making them competitive with those produced by ANN models, but without the interpretability issues. This factor makes them preferable for solving many regression or classification problems [38]. Therefore, two different soft computing techniques were investigated and implemented in the present study: the former involved the development of an ANN-based model, whereas the latter was based on a DT algorithm known in the scientific literature as CatBoost. By way of example, in recent years, CatBoost was successfully implemented to predict the Pavement Condition Index values of asphalt concrete overlays [39].

The main purposes of the present study are (i) to mechanically characterize two different asphalt mixtures for pavement construction by investigating a fundamental behavioral parameter, such as stiffness modulus (SM); and (ii) to model and predict the performance of each mixture using ANN- and CatBoost-based ML algorithms. The two investigated mixtures were designed for binder and base layers and were prepared using spilite aggregate, a conventional 50/70 penetration grade bitumen, and limestone filler. An extensive four-point bending test (4PBT) experimental campaign was carried out to determine the stiffness modulus values of the mixtures under 11 loading frequencies and 4 testing temperatures, ranging from 0.1 to 50 Hz and from 0 to 30 °C, respectively. Three to five specimens were tested for each condition, and the results were averaged to obtain the dataset subsequently used to train and test the two different ML models.

State-of-the-art procedures were implemented in both the developed methodologies, including k-fold cross-validation, overfitting detection, and extensive grid searches to

find the best hyperparameter sets for each algorithm. Six different performance metrics were investigated to determine the accuracy of the developed models and to evaluate their generalization capabilities, namely mean absolute error, mean absolute percentage error, mean squared error, root mean squared error, Pearson correlation coefficient, and determination coefficient.

The results obtained emphasize that the developed ML model can provide accurate and reliable stiffness modulus predictions, thus allowing these values to then be exploited within the well-established design procedures. This represents the main contribution provided by the present study to the existing scientific literature. In addition, the sensitivity analysis that was carried out confirms that ML models are able to understand the functional relationships between the variables that were investigated, thus inspiring future scientific applications.

The remainder of the paper is organized as follows: Section 2 describes the volumetric characterization of the prepared mixtures and the experimental campaign that was carried out. Section 3 provides an overall explanation of the ML framework, explaining all the pre-processing and resampling techniques that were implemented in the developed models. Section 4 describes the obtained predictive results and compares the performances of the CatBoost and ANN models. Finally, Section 5 outlines the main conclusions and points out future developments.

2. Materials and Methods

Two asphalt mixtures were investigated in the present study. They were designed for binder and base course layers, respectively. The former involved the utilization of spilitite aggregate with a nominal maximum aggregate size of 16 mm (AML16) obtained from the Bělčice quarry (Benešov, Czech Republic). The latter was also prepared using spilitite aggregate, but the nominal maximum aggregate size was 22 mm (AMP22). The bitumen used to prepare both the mixtures was a conventional one, with a penetration grade (PG) at 25 °C ranging between 50 and 70 mm/10, meeting the technical specifications set by the European standard EN 12591 [40] for paving grade binders (see Table 1). The binder was obtained from the Litvinov refinery (Litvínov, Czech Republic). Finally, limestone filler from the Velke Hydčice quarry was used.

Table 1. Characteristics of the use bituminous binder.

Basic Bitumen Characteristic	Unit	Value
Penetration at 25 °C	mm/10	59
Softening point (ring and ball)	°C	50.6
Breaking point after Fraas	°C	−11
Remaining penetration after short-term ageing (RTFOT)	%	41
Softening point after short-term ageing (RTFOT)	°C	54.2

The mix design followed the requirements set by Czech technical standard CSN 73 6121 [41]. The grading curves of both mixtures is given in Table 2. A volumetric characterization of both the mixtures was carried out, and the results are summarized in Table 3. The following additional requirements were also tested: (i) moisture susceptibility according to EN 12697-12 [42] determined at 15 °C; and (ii) stiffness tested on 6 cylindrical test specimens via repeated indirect tensile stress tests (IT-CY) according to EN 12697-26 [43] at 15 °C.

Table 2. Grading curves of tested asphalt mixtures and requirements according to the national standard (CSN 73 6121 [41]).

Control Sieve (mm)	Grading Curve—Passing (%)			
	AML16	Mix Requirements	AMP22	Mix Requirements
32	100		100	
22	100		96	100–90
16	97	100–90	80	85–60
11	77		68	
8	64	80–52	55	65–40
5.6	55		48	
4	44	61–31	38	
2	28	45–20	24	38–22
1	18		17	
0.5	14		14	
0.25	11		12	
0.125	10	16–4	10	13–5
0.063	8.5	10–3	9.0	9–4

Table 3. Volumetric characterizations of AML16 and AMP22 mixtures.

	AML16 Mix	AMP22 Mix
Binder content	4.5%	4.2%
Air voids content	5.2%	5.3%
Bulk density	2.417 g/cm ³	2.421 g/cm ³
Moisture susceptibility	84%	81%
IT-CY stiffness at 15 °C	7537 MPa	8257 MPa

An experimental trial using a four-point bending test (4PBT) apparatus was performed under several testing conditions. The test method principles and general conditions related to the testing apparatus and the test specimens that were used are given in Annex B. The resulting stiffness values were investigated by exposing the prismatic specimens to different temperatures and frequencies. According to EN 12697-26 [43], the dimensions of the test specimens should be 405 × 50 × 50 mm. To determine stiffness, eleven different loading frequencies (from 0.1 to 50 Hz) and four different testing temperatures (from 0 to 30 °C) were selected. As for the AMP22 mixture, five testing temperatures were investigated since the stiffness moduli were also tested at 15 °C. The technical standard does not specifically set a fixed number of frequencies to be tested, and, similarly, particular temperatures are not prescribed. For national products, the standards may specify additional requirements, e.g., one particular temperature, which is then linked to the mechanistic modelling of the performance of the pavement structure. According to existing good practice in asphalt mix testing, several temperatures are usually used to either simulate the behavior response of the mixture to the conditions occurring in the pavement—e.g., 0 °C for winter and 30 °C for summer—or at least three different temperatures are selected to plot the so called master curve, in which the resulting stiffness values are shifted by applying the time–temperature superposition principle to one selected temperature. This curve is then used to interpret the behavior of the asphalt mix under the different conditions which can occur on a road, such as low or high temperatures and different traffic loading intensities. The latter effect is simulated in the test by the used frequencies, i.e., very low or very high frequencies simulate either low or fast-moving traffic, and the pavement is put through cyclic loading and resting periods. The frequency is changed according to the speed and intensity that is being simulated. To dynamically simulate this traffic loading effect, test methods such as 4PBT are able to determine a large variability in the responses of the specimen depending on the loading frequency and temperature. The test setup and instrumentation for 4PBT is shown in Figure 1. For each temperature, and for the whole range of frequencies between 0.1 Hz and 50 Hz, three to five specimens were tested, and the average results are reported

in Table 4. These outcomes constituted the dataset that was later used to train and test the developed ML algorithms.

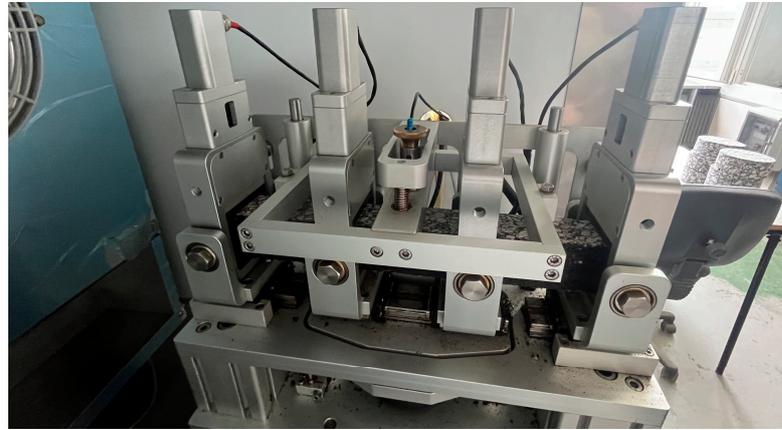


Figure 1. Test setup and instrumentation for 4PBT.

Table 4. Outcomes of the 4PBT experimental trial carried out on AML16 and AMP22 mixtures.

Mix	Testing Temperature (°C)	Stiffness Modulus (SM) (MPa) under Loading Frequency (Hz)										
		0.1	1	2	3	5	8	10	15	20	30	50
AML16	0	14,606	18,939	20,203	20,984	21,874	22,763	23,067	23,785	24,104	23,862	23,751
	10	7346	11,335	12,659	13,497	14,510	15,453	15,855	16,639	17,108	17,733	17,844
	20	3001	5325	6276	6912	7769	8574	9009	9722	10,317	11,238	11,618
	30	1265	2140	2506	2831	3285	3754	4005	4496	4891	5777	6387
AMP22	0	13,719	18,268	19,697	20,435	21,364	22,196	22,654	23,245	23,672	24,133	23,691
	10	6633	10,741	12,142	13,017	14,097	15,108	15,613	16,485	16,959	17,847	17,912
	15	5887	8803	9845	10,503	11,320	12,123	12,534	13,172	13,606	14,396	14,664
	20	1975	4521	5467	6089	6857	7757	8210	8897	9414	10,414	11,155
	30	1222	1681	2073	2472	2962	3433	3660	4061	4377	5192	5896

With the aim of obtaining a detailed understanding of the collected dataset, a preliminary exploratory analysis was carried out by diagramming the stiffness modulus (SM) as a function of the testing temperature and loading frequency. This practice is useful in identifying potential existing relationships and correlations between different pairs of features [44]. The marked trends between the considered variables can be observed in Figure 2. Since the data came from the analysis of two different mixtures, cyan and red triangles were used to represent the data collected during the tests of the AML16 and AMP22 mixture specimens, respectively.

To quantify the correlations between SM and the influencing variables under consideration, Pearson correlation factors were computed, thus creating a so-called Pearson correlation matrix (Figure 3). This tool provides a measure of the strength of an estimated linear correlation between two given variables [45]. This strength can vary within a range from -1 to $+1$. The minus and plus signs represent inverse and direct proportionality, respectively. The closer the absolute value of the Pearson correlation factor is to unity, the stronger the correlation between the variables considered.

It can be observed that the testing temperature showed the strongest negative correlation with the stiffness modulus ($R = -0.92$), while the loading frequency displayed a small positive correlation ($R = +0.28$). A categorical variable was assigned to distinguish between the two mixtures, and it was subsequently encoded according to alphabetical order to determine its Pearson correlation factor with the SM values. As such, 0 and +1 were assigned to identify the AML16 and AMP22 mixtures, respectively. However, the Pearson correlation between the encoded categorical variable and the stiffness modulus

was comparatively low ($R = -0.04$). No correlation could be found between the loading frequency, the testing temperature, and the categorical variable, thus allowing these three variables to be considered independently and therefore making them usable as inputs in subsequent predictive modeling steps [46].

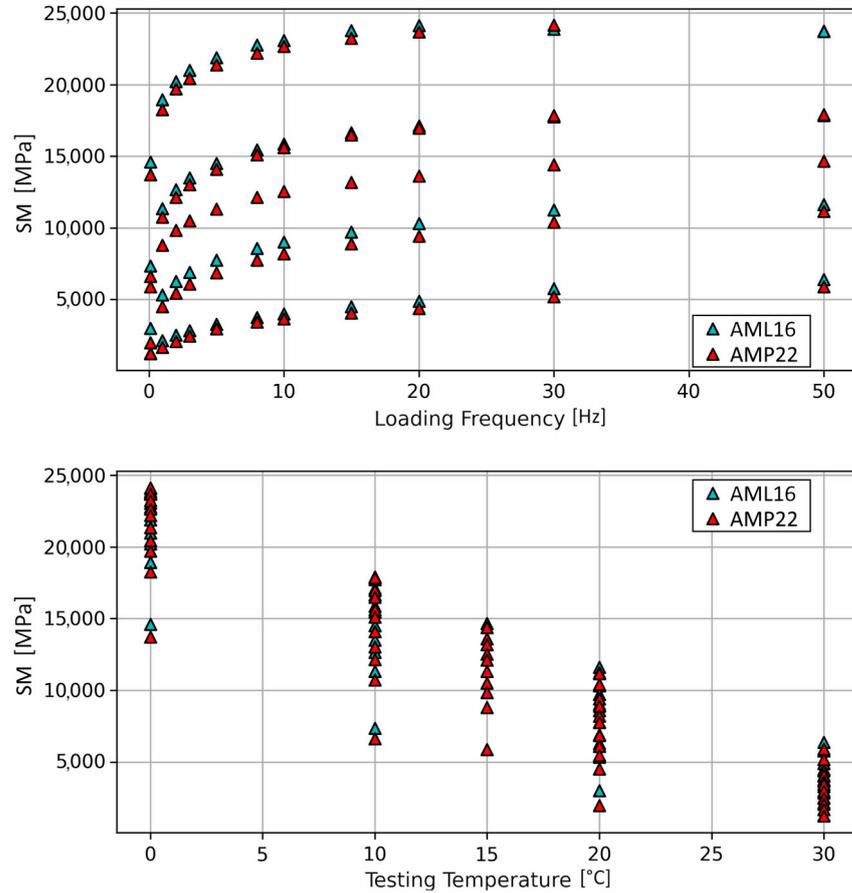


Figure 2. Graphical trends of SM vs. loading frequency (up) and testing temperature (down).

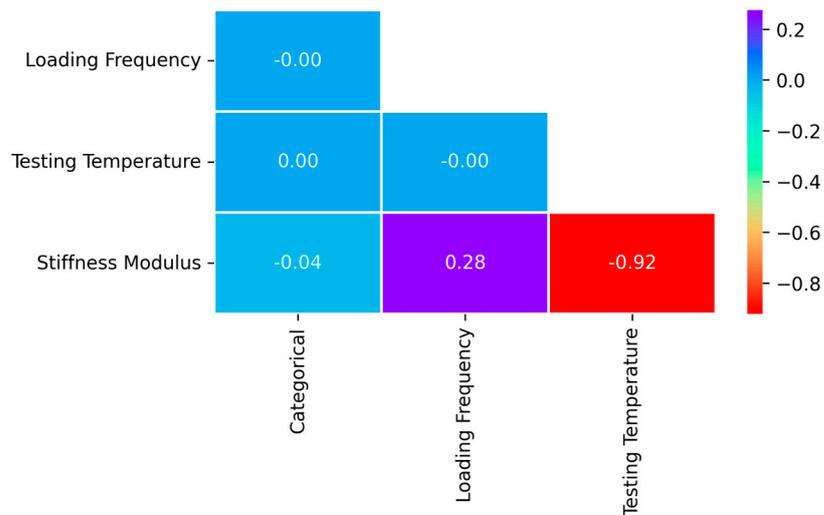


Figure 3. Pearson correlation factors between SM and its influencing variables.

3. Machine Learning Framework

3.1. Categorical Boosting

This section focuses on the development of a predictive model called Categorical Boosting (or CatBoost) which is capable of predicting the mechanical behavior of AMs in terms of SM. The methodology is based on a decision tree ensemble approach in which each generated decision tree sequentially learns from previous ones to fine-tune its learning and improve its predictive performance [47].

In 2017, Yandex (Moscow, Russia) engineers proposed CatBoost as an advanced machine learning algorithm. Unlike standard gradient-boosting techniques, CatBoost employs innovative ordered boosting to address target leakage and prediction shift issues [48]. Furthermore, one of the most significant advantages of this algorithm lies in its ability to deal with various data formats and sizes, unlike other conventional ML techniques. CatBoost can automatically operate with categorical variables, encoding them without showing any conversion issues.

Overall, the main advantages of CatBoost can be summarized as follows [49,50]: (a) effectiveness in dealing with small-scale datasets and with categorical features; (b) high stability and efficiency; (c) less parameter tuning; and (d) oblivious tree building reduces overfitting, improving accuracy and generalizability.

3.2. Artificial Neural Network

Unlike decision tree-based models, a neural model is based on an artificial neural network whose structure is similar to that of the human nervous system. Such mathematical models are characterized by multiple and interconnected artificial neurons which are typically organized in sequences of three different layers called input, hidden, and output [51]. The first layer is composed of as many neurons as there are pieces of input information, namely the variables influencing the SM. The second layer serves to process the input information, and the number of neurons or which it is composed determines the computational power of the neural network. Finally, the last layer provides the predicted value of the target variable, here referred to as the SM. Both the hidden and the output layers are equipped with activation functions. To deal with nonlinearities, the former usually employs a nonlinear activation function, while the latter usually employs a simple linear transformation.

Each connection is weighted, and the strength of this methodology lies in the variability of the connection weights, which are iteratively determined based on the loss function results achieved during the training phase. In general, for each iteration of the training algorithm, the output parameter estimation provided by a neural model (Equation (1)) can be described as follows:

$$f(X) = W_2[f_A(W_1X)] \quad (1)$$

where X denotes the input information vector, f_A denotes the activation function of the hidden layer, and W_1 and W_2 denote the matrices of the weights related to the connections between the input and hidden layers and between the hidden and output layers, respectively. The mathematical formulations by which the weight matrices are accordingly adjusted are known as training algorithms, and the main ones can be deeply explored via the relevant scientific literature [52–54].

Over the years, machine learning models based on artificial neural networks have proven able to successfully approximate even highly nonlinear functions, returning outstanding performance in several pavement engineering applications [55–57].

3.3. Grid Search and k-Fold Cross-Validation

To enhance both the speed of convergence and the accuracy of the model, each feature underwent min–max normalization. This pre-processing procedure ensured that the observation values were scaled to a range between 0 and +1, according to Equation (2):

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where, for each variable, x_{norm} is the normalized value, x is the observed value, and x_{min} and x_{max} represent its minimum and maximum values, respectively. Subsequently, roughly the 75% of the dataset was used to train the models (75 observations out of 99). A five-fold cross-validation technique was employed to fairly evaluate the resulting training and validation performances of both models (CatBoost and ANN), and to subsequently optimize their parameters according to a grid search [58]. Detailed descriptions of the hyperparameters and search ranges of each model are provided in Table 5.

Table 5. Grid search for the best hyperparameters for each model.

ML Model	Hyperparameter	Search Range	Optimal Value
CatBoost	Max depth	3–6	3
	Learning rate	0.005, 0.01, 0.05	0.01
	Max iterations	500, 1000, 5000	5000
ANN	Hidden layer size	1–50	38
	Activation function	Identity, Logistic, TanH, ReLU	ReLU
	Solver	SGD [52], Adam [59]	Adam
	Max iterations	500, 1000, 5000	1000

Once the best hyperparameter combination was identified, the best-calibrated model was tested on an independent testing set made from the remaining 24 observations (roughly 25% of the dataset). This allowed several goodness-of-fit measures to be determined, including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), Pearson correlation coefficient (R), and determination coefficient (R^2) [47].

Their mathematical formulations (Equations (3)–(8)) are presented below:

$$MAE = \frac{1}{n} \sum_{i=1}^n |SM_{T_i} - SM_{P_i}| \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{SM_{T_i} - SM_{P_i}}{SM_{T_i}} \right| \times 100 \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (SM_{T_i} - SM_{P_i})^2 \quad (5)$$

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (SM_{T_i} - SM_{P_i})^2} \quad (6)$$

$$R = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{SM_{T_i} - \mu_{SM_{T_i}}}{\sigma_{SM_{T_i}}} \right) \left(\frac{SM_{P_i} - \mu_{SM_{P_i}}}{\sigma_{SM_{P_i}}} \right) \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (SM_{T_i} - SM_{P_i})^2}{\sum_{i=1}^n (SM_{T_i} - \mu_{SM_{T_i}})^2} \quad (8)$$

where SM_{T_i} is the actual i -th value of the stiffness modulus, SM_{P_i} is the model-predicted stiffness modulus value for the i -th observation, i is the observation index, n is the total

number of observations, and μ and σ are the mean and the standard deviation values, respectively. An overfitting detection algorithm was also implemented to avoid any potential overfitting phenomena. During the training phase, before proceeding with the subsequent iteration, both ML models determined their corresponding best loss value coupled with the number of iterations since this optimal value was achieved. If the latter overtook a predetermined upper threshold (here set as 20), the algorithm automatically stopped the training phase.

4. Results and Discussion

To justify the choice of the CatBoost approach and highlight its better performance compared with that of the ANN model, an in-depth comparative analysis was carried out in the present study. It is important to mention that both models were calibrated and subsequently tested using the same observations to ensure a fair performance comparison in terms of the same goodness-of-fit measures.

A one-to-one visual comparison between the experimental target vector, the outputs predicted by CatBoost, and those predicted by the ANN can be observed in Figure 4. The histogram plot displays black bars representing the test set values, grey bars representing the CatBoost predictions, and cyan bars representing the ANN predictions. The horizontal axis shows the IDs of the 24 observations from which the test set was composed. Both the ANN and CatBoost predictions closely match the observed SM experimental data, with CatBoost performing slightly better than the ANN. This finding can be considered significant since it highlights the accuracy and the reliability of both developed models.

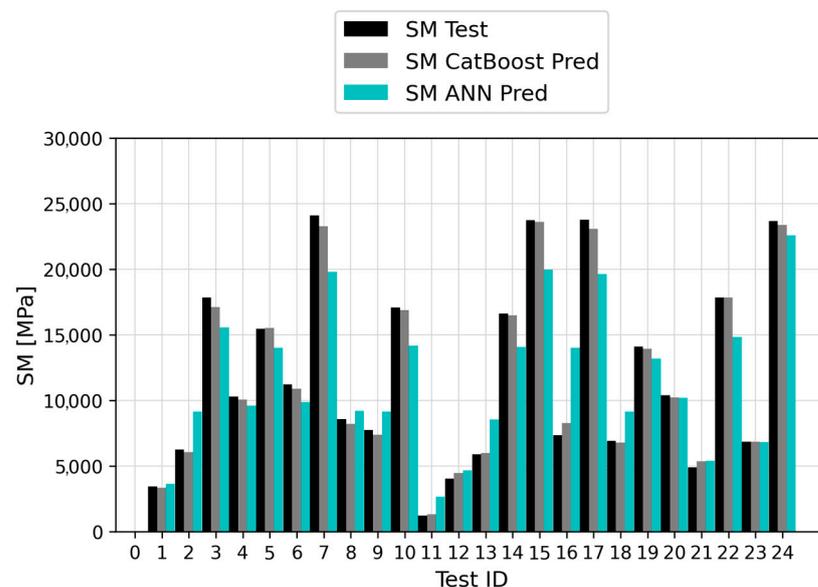


Figure 4. Observed SM values and those predicted by CatBoost and the ANN.

To gain a deeper comprehension of the capabilities of each model, regression diagrams for the two outlined models can be observed in Figure 5. The observed experimental values and the corresponding model predictions are plotted on x-axis and the y-axis, respectively. The 45-degree black line represents perfect accuracy, i.e., 100% accurate predictions. Training, validation, and testing observations are denoted by red triangles, green crosses, and blue circles, respectively. A testing Pearson correlation coefficient for each model is displayed at the top of the corresponding regression diagram. It can be observed that the predictions of both models are closely aligned to the perfect accuracy line, which further confirms the excellent results that we obtained. Again, the CatBoost predictions showed a slightly higher accuracy, highlighted by a higher Pearson correlation coefficient (0.9990).

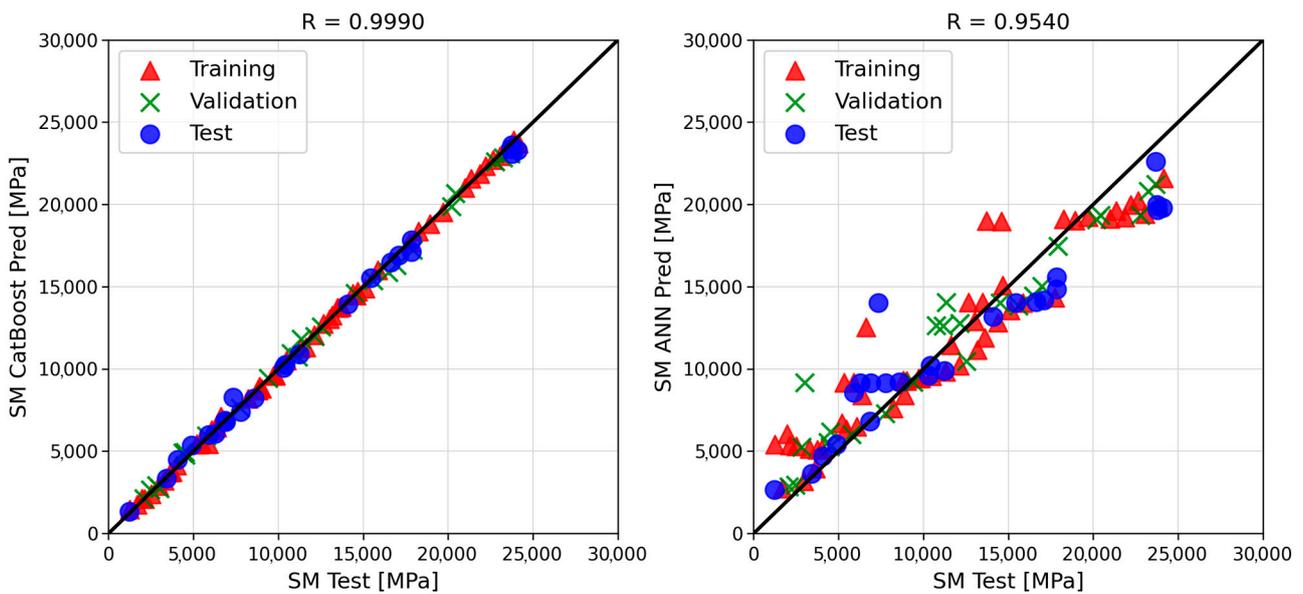


Figure 5. Regression plots for both the CatBoost (left) and the ANN (right) models.

This condition is further highlighted in Figure 6, in which only the results obtained during the testing phase are shown. It can be observed that, based on the same testing vector, the predictions made by the CatBoost model (grey circles) are closer to the perfect accuracy line with than those made by the ANN model (cyan circles).

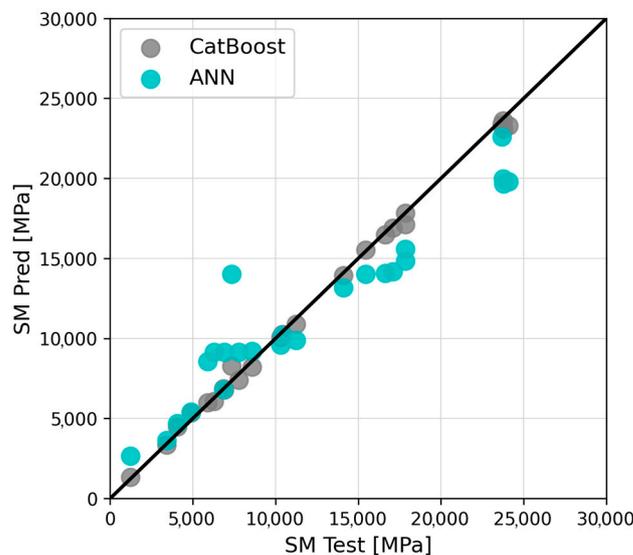


Figure 6. Direct comparison between the CatBoost and ANN model predictions.

A detailed comparison of the results achieved by the two outlined models can be found in Table 6, described in terms of the six goodness-of-fit measures. As was previously mentioned, to ensure a fair comparison, both ML models were calibrated using the same observations, and both employed the same pre-processing and resampling algorithms.

Based on the results obtained during the testing phase, it can be observed that the difference between the two models in terms of prediction accuracy is roughly an order of magnitude for each goodness-of-fit metric. The CatBoost model displayed the lowest error metric values (MAE, MAPE, MSE, and RMSE) and the highest correlation metric values (R, and R^2), as is shown in Table 6.

Although the performance of each model was remarkable, it can be concluded that the most reliable and successful one was CatBoost. For this reason, we decided to perform a

sensitivity analysis of the results obtained in terms of SM with respect to the input variables of the model, namely a categorical variable identifying the mixture, loading frequency, and testing temperature.

Table 6. Training and testing error metrics of the investigated ML models.

Goodness-of-Fit Measure	ML Model	
	CatBoost	ANN
MAE (MPa)	300.49	2000.98
MAPE (%)	3.41	22.68
MSE (MPa ²)	1.55×10^5	6.49×10^6
RMSE (MPa)	393.33	2547.83
R	0.9990	0.9540
R ²	0.9968	0.8674

Determining the influence of a specific feature on the predictions of an ML model can be difficult. To address this challenge, an algorithm for the feature importance calculation was implemented. This algorithm helps determine how much the CatBoost predictions change on average when the values of a particular feature are modified. A higher importance value determines a greater prediction variation when that particular feature is modified. The importance values were also normalized to return a total importance value equal to 100%. As can be observed in Figure 7, the testing temperature had the highest importance value (79.82%), followed by the loading frequency (18.93%) and the categorical variable (1.25%).

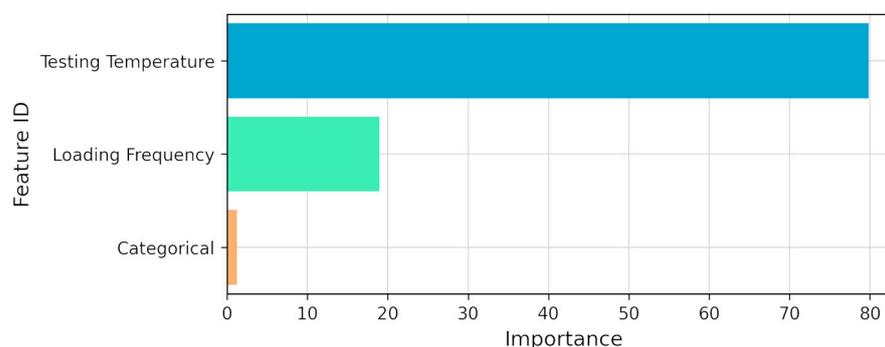


Figure 7. CatBoost feature importance.

The results that were obtained show how the SM values are mainly dependent on the testing temperature (as was expected from the Pearson correlation matrix results), and secondarily dependent on the loading frequency. The categorical variable introduced to distinguish the two mixture types is of minor importance, probably because of the quite similar mechanical behavior exhibited by the two investigated mixtures.

In this respect, it should be mentioned that all the results described in the present study refer to the experimental trial under consideration. Therefore, any other application that differs from the proposed one will mandatorily require new calibration efforts and new research to identify the best model hyperparameters.

5. Conclusions

To produce high-performance pavement for road infrastructure, the stiffness modulus (SM) of the asphalt mixture needs to be determined during the mix design procedures since it represents a crucial mechanical parameter. Expensive and time-consuming experimental trials are usually carried out to this end, but machine learning-based methodologies could provide useful tools to reduce laboratory efforts. If properly developed and implemented,

reliable SM predictions of asphalt mixture performance could be provided by these soft computing techniques, thus helping pavement engineers during data analysis procedures.

The current paper aimed to discuss two different approaches to modeling and predicting the stiffness modulus in four-point bending tests (4PBTs). The first approach is based on decision trees and is named Categorical Boosting (CatBoost), and the second relies on artificial neural networks (ANNs). Both methodologies were developed using Python 3.9.12 [60]. The experimental campaign involved two selected mixtures: one developed for binder layers and the other for base layers. Both were prepared using standard 50/70 PG bitumen, spilite aggregate, and limestone filler. Three to five specimens were tested for each of the eleven loading frequencies (from 0.1 to 50 Hz) and four testing temperatures (from 0 to 30 °C). The obtained results were averaged and used to train and subsequently test the developed ML models. The following conclusions can be drawn:

- Based on mixture composition and testing conditions, both the models were able to reliably predict the resulting stiffness modulus of each mixture, properly balancing accuracy and generalizability. This was ensured by the careful optimization of the hyperparameters of both models using three different algorithms, namely an extensive grid search, a five-fold cross-validation, and an overfitting detection.
- The optimal CatBoost model was characterized by a maximum tree depth of 3, a learning rate of 0.01, and a maximum number of training iterations of 5000. Conversely, the optimal ANN model involved the Adam solver, and its architecture was characterized by 38 hidden neurons, a ReLU activation function, and a maximum number of training iterations of 1000.
- Based on six goodness-of-fit metrics, CatBoost proved to be the most suitable algorithm to model the phenomena under investigation, outperforming the ANN. Its predictions were characterized by outstanding accuracy, expressed by MAE, MAPE, and R^2 values equal to 300.49 MPa, 3.41%, and 0.9968, respectively. The corresponding ANN error metrics were roughly an order of magnitude higher, resulting in a comparatively lower prediction accuracy.
- A sensitivity analysis carried out on the CatBoost model revealed that the testing temperature had the strongest influence on the SM predictions (79.82% of total importance), followed by the loading frequency (18.93%) and the categorical variable (1.25%).

The outlined methodology represents an encouraging starting point, and it could help pavement engineering professionals enhance mix design procedures.

As for future developments, the present research could be expanded in many ways: (i) the generalization capabilities of the CatBoost model could be improved by collecting more observations from several alternative mixtures; (ii) different optimization algorithms could be further investigated to find new potential optimal solutions for the best hyperparameter combinations; and (iii) other alternative modelling variables (such as fatigue or permanent deformation resistance) could be implemented as output, thus allowing a comprehensive mechanical characterization to be properly carried out.

Author Contributions: Conceptualization, N.B.; methodology, N.B., F.R., J.V., M.D.G. and J.B.K.; software, F.R.; validation, N.B., F.R., F.D., P.V., J.V., M.D.G. and J.B.K.; formal analysis, N.B., F.R., F.D., P.V., J.V., M.D.G. and J.B.K.; investigation, P.V., J.V., M.D.G. and J.B.K.; resources, N.B., J.V. and J.B.K.; data curation, N.B., F.R. and J.V.; writing—original draft preparation, N.B., F.R., F.D., P.V., J.V., M.D.G. and J.B.K.; writing—review and editing, N.B., F.R., F.D., P.V., J.V., M.D.G. and J.B.K.; visualization, N.B., F.R., F.D., P.V., J.V., M.D.G. and J.B.K.; supervision, N.B., J.V. and J.B.K.; project administration, N.B., J.V. and J.B.K.; funding acquisition, N.B., J.V. and J.B.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was conceptualized and developed as part of activities related to project GA22-04047K, funded by The Czech Scientific Foundation (GACR), and project No. 2021/03/Y/ST8/00079, funded by the Polish National Science Centre (NCN) under the Weave-UNISONO 2021.

Data Availability Statement: Some or all of the data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dash, S.S.; Chandrappa, A.K.; Sahoo, U.C. Design and performance of cold mix asphalt—A review. *Constr. Build. Mater.* **2022**, *315*, 125687. [[CrossRef](#)]
2. Zhou, F.; Scullion, T.; Sun, L. Verification and modeling of three-stage permanent deformation behavior of asphalt mixes. *J. Transp. Eng.* **2004**, *130*, 486–494. [[CrossRef](#)]
3. Gandomi, A.H.; Alavi, A.H.; Mirzahosseini, M.R.; Nejad, F.M. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *J. Mater. Civ. Eng.* **2011**, *23*, 248–263. [[CrossRef](#)]
4. Alavi, A.H.; Ameri, M.; Gandomi, A.H.; Mirzahosseini, M.R. Formulation of flow number of asphalt mixes using a hybrid computational method. *Constr. Build. Mater.* **2011**, *25*, 1338–1355. [[CrossRef](#)]
5. Dias, J.L.F.; Picado-Santos, L.; Capitão, S. Mechanical performance of dry process fine crumb rubber asphalt mixtures placed on the Portuguese road network. *Constr. Build. Mater.* **2014**, *73*, 247–254. [[CrossRef](#)]
6. Liu, Q.T.; Wu, S.P. Effects of steel wool distribution on properties of porous asphalt concrete. *Key Eng. Mater.* **2014**, *599*, 150–154. [[CrossRef](#)]
7. Garcia, A.; Norambuena-Contreras, J.; Bueno, M.; Partl, M.N. Influence of steel wool fibers on the mechanical, thermal, and healing properties of dense asphalt concrete. *J. Test. Eval.* **2014**, *42*, 1107–1118. [[CrossRef](#)]
8. Pasandín, A.; Pérez, I. Overview of bituminous mixtures made with recycled concrete aggregates. *Constr. Build. Mater.* **2015**, *74*, 151–161. [[CrossRef](#)]
9. Zaumanis, M.; Mallick, R.B.; Frank, R. 100% hot mix asphalt recycling: Challenges and benefits. *Transp. Res. Procedia* **2016**, *14*, 3493–3502. [[CrossRef](#)]
10. Wang, L.; Gong, H.; Hou, Y.; Shu, X.; Huang, B. Advances in pavement materials, design, characterisation, and simulation. *Road Mater. Pavement Des.* **2017**, *18*, 1–11. [[CrossRef](#)]
11. Masad, E.; Tashman, L.; Little, D.; Zbib, H. Viscoplastic modeling of asphalt mixes with the effects of anisotropy, damage and aggregate characteristics. *Mech. Mater.* **2005**, *37*, 1242–1256. [[CrossRef](#)]
12. Giunta, M.; Pisano, A.A. One dimensional viscoelastoplastic constitutive model for asphalt concrete. *Multidiscip. Model. Mater. Struct.* **2006**, *2*, 247–264. [[CrossRef](#)]
13. Erkens, S.M.J.G.; Liu, X.; Scarpas, A. 3D finite element model for asphalt concrete response simulation. *Int. J. Geomech.* **2002**, *2*, 305–330. [[CrossRef](#)]
14. Costanzi, M.; Cebon, D. Generalized phenomenological model for the viscoelasticity of idealized asphalts. *J. Mater. Civ. Eng.* **2014**, *26*, 399–410. [[CrossRef](#)]
15. Pasetto, M.; Baldo, N. Computational analysis of the creep behaviour of bituminous mixtures. *Constr. Build. Mater.* **2015**, *94*, 784–790. [[CrossRef](#)]
16. Collop, A.C.; McDowell, G.R.; Lee, Y. Use of the distinct element method to model the deformation behavior of an idealized asphalt mixture. *Int. J. Pavement Eng.* **2004**, *5*, 1–7. [[CrossRef](#)]
17. Abbas, A.; Masad, E.; Papagiannakis, T.; Harman, T. Micromechanical modelling of the viscoelastic behavior of asphalt mixtures using the discrete-element method. *Int. J. Geomech.* **2007**, *7*, 131–139. [[CrossRef](#)]
18. Dondi, G.; Simone, A.; Vignali, V.; Manganelli, G. Numerical and experimental study of granular mixes for asphalts. *Powder Technol.* **2012**, *232*, 31–40. [[CrossRef](#)]
19. Liu, Z.; Gu, X.; Wu, C.; Ren, H.; Zhou, Z.; Tang, S. Studies on the validity of strain sensors for pavement monitoring: A case study for a fiber Bragg grating sensor and resistive sensor. *Constr. Build. Mater.* **2022**, *321*, 126085. [[CrossRef](#)]
20. Asteris, P.G.; Kokoris, S.; Gavriilaki, E.; Tsoukalas, M.Z.; Houpas, P.; Paneta, M.; Koutzas, A.; Argyropoulos, T.; Alkayem, N.F.; Armaghani, D.J.; et al. Early prediction of COVID-19 outcome using artificial intelligence techniques and only five laboratory indices. *Clin. Immunol.* **2023**, *246*, 109218. [[CrossRef](#)]
21. Li, J.; Sha, A.; Jiang, W.; Ren, X. Investigation on internal mechanical response of asphalt mixture during creep test utilizing the Smart Aggregate. *Constr. Build. Mater.* **2022**, *349*, 128800. [[CrossRef](#)]
22. Majidifard, H.; Jahangiri, B.; Buttler, W.G.; Alavi, A.H. New machine learning-based prediction models for fracture energy of asphalt mixtures. *Measurement* **2019**, *135*, 438–451. [[CrossRef](#)]
23. Ghafari, S.; Ehsani, M.; Nejad, F.M. Prediction of low-temperature fracture resistance curves of unmodified and crumb rubber modified hot mix asphalt mixtures using a machine learning approach. *Constr. Build. Mater.* **2022**, *314*, 125332. [[CrossRef](#)]
24. Rahman, A.A.; Mendez Larrain, M.M.; Tarefder, R.A. Development of a nonlinear rutting model for asphalt concrete based on Weibull parameters. *Int. J. Pavement Eng.* **2019**, *20*, 1055–1064. [[CrossRef](#)]
25. Baldo, N.; Miani, M.; Rondinella, F.; Valentin, J.; Vackcová, P.; Manthos, E. Stiffness Data of High-Modulus Asphalt Concretes for Road Pavements: Predictive Modeling by Machine-Learning. *Coatings* **2022**, *12*, 54. [[CrossRef](#)]
26. Baldo, N.; Miani, M.; Rondinella, F.; Manthos, E.; Valentin, J. Road Pavement Asphalt Concretes for Thin Wearing Layers: A Machine Learning Approach towards Stiffness Modulus and Volumetric Properties Prediction. *Period. Polytech. Civ. Eng.* **2022**, *66*, 1087–1097. [[CrossRef](#)]
27. Tarefder, R.A.; White, L.; Zaman, M. Neural network model for asphalt concrete permeability. *J. Mater. Civ. Eng.* **2005**, *17*, 19–27. [[CrossRef](#)]

28. Ozsahin, T.S.; Oruc, S. Neural network model for resilient modulus of emulsified asphalt mixtures. *Constr. Build. Mater.* **2008**, *22*, 1436–1445. [[CrossRef](#)]
29. Liu, J.; Liu, F.; Zheng, C.; Fanijo, E.O.; Wang, L. Improving asphalt mix design considering international roughness index of asphalt pavement predicted using autoencoders and machine learning. *Constr. Build. Mater.* **2022**, *360*, 129439. [[CrossRef](#)]
30. Xiao, F.; Amirkhanian, S.; Juang, C.H. Prediction of fatigue life of rubberized asphalt concrete mixtures containing reclaimed asphalt pavement using artificial neural networks. *J. Mater. Civ. Eng.* **2009**, *21*, 253–261. [[CrossRef](#)]
31. Ahmed, T.M.; Green, P.L.; Khalid, H.A. Predicting fatigue performance of hot mix asphalt using artificial neural networks. *Road Mater. Pavement Des.* **2017**, *18*, 141–154. [[CrossRef](#)]
32. Bui, Q.A.T.; Nguyen, D.D.; Iqbal, M.; Jalal, F.E.; Prakash, I.; Pham, B.T. Prediction of Interface Shear Stiffness Modulus of Asphalt Pavement using Bagging Ensemble-based Hybrid Machine Learning Model. *Arab. J. Sci. Eng.* **2023**, *48*, 13889–13900. [[CrossRef](#)]
33. Piro, N.S.; Mohammed, A.; Hamad, S.M.; Kurda, R. Artificial neural networks (ANN), MARS, and adaptive network-based fuzzy inference system (ANFIS) to predict the stress at the failure of concrete with waste steel slag coarse aggregate replacement. *Neural Comput. Appl.* **2023**, *35*, 13293–13319. [[CrossRef](#)]
34. Mahmood, W.; Mohammed, A. Performance of ANN and M5P-tree to forecast the compressive strength of hand-mix cement-grouted sands modified with polymer using ASTM and BS standards and evaluate the outcomes using SI with OBJ assessments. *Neural Comput. Appl.* **2022**, *34*, 15031–15051. [[CrossRef](#)]
35. Yang, Z.; Zhang, A.; Sudjianto, A. Enhancing explainability of neural networks through architecture constraints. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 2610–2621. [[CrossRef](#)]
36. Baduge, S.K.; Thilakathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 104440. [[CrossRef](#)]
37. Koopialipour, M.; Asteris, P.G.; Mohammed, A.S.; Alexakis, D.E.; Mamou, A.; Armaghani, D.J. Introducing stacking machine learning approaches for the prediction of rock deformation. *Transp. Geotech.* **2022**, *34*, 100756. [[CrossRef](#)]
38. Slack, D.; Friedler, S.A.; Scheidegger, C.; Roy, C.D. Assessing the local interpretability of machine learning models. *arXiv* **2019**, arXiv:1902.03501.
39. Mansour, E.; Mousa, M.R.; Dhasmana, H.; Hassan, M. Machine-Learning-Based Framework for Prediction of the Long-Term Field Performance of Asphalt Concrete Overlays in a Hot and Humid Climate. *Transp. Res. Rec.* **2023**, *2677*, 03611981231161353. [[CrossRef](#)]
40. *SIST EN 12591*; Bitumen and Bituminous Binders—Specifications for Paving Grade Bitumens. European Committee for Standardization: Brussels, Belgium, 2009.
41. *ČSN 73 6121 (736121)*; Stavba Vozovek—Hutněné Asfaltové Vrstvy—Provádění a Kontrola Shody. Česká Technická Norma: Prague, Czech Republic, 2019.
42. *SIST EN 12697: Part 12*; Bituminous Mixtures-Test Methods for Hot Mix Asphalt-Determination of the Water Sensitivity of Bituminous Specimens. European Committee for Standardization: Brussels, Belgium, 2019.
43. *SIST EN 12697: Part 26*; Bituminous Mixtures-Test Methods for Hot Mix Asphalt-Stiffness. European Committee for Standardization: Brussels, Belgium, 2019.
44. Subasi, A. *Practical Machine Learning for Data Analysis Using Python*; Elsevier: London, UK, 2020.
45. Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*, 7th ed.; Routledge: London, UK, 2020. [[CrossRef](#)]
46. Zhou, B.Q.; Yang, R.C.; Li, H.P.; Wang, Y.J.; Zhang, C.Y.; Xiao, Z.J.; He, Z.Q.; Pang, W.H. Numeric and nonnumeric information input to predict adsorption amount, capacity and kinetics of tetracyclines by biochar via machine learning. *Chem. Eng. J.* **2023**, *471*, 144636. [[CrossRef](#)]
47. Rondinella, F.; Daneluz, F.; Hofko, B.; Baldo, N. Improved predictions of asphalt concretes' dynamic modulus and phase angle using decision-tree based categorical boosting model. *Constr. Build. Mater.* **2023**, *400*, 132709. [[CrossRef](#)]
48. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. *arXiv* **2017**, arXiv:1706.09516.
49. Huang, G.M.; Wu, L.F.; Ma, X.; Zhang, W.Q.; Fan, J.L.; Yu, X.; Zeng, W.Z.; Zhou, H.M. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* **2019**, *574*, 1029–1041. [[CrossRef](#)]
50. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big data* **2020**, *7*, 94. [[CrossRef](#)]
51. Moayed, H.; Mosallanezhad, M.; Rashid, A.S.A.; Jusoh, W.A.W.; Muazu, M.A. A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: Theory and applications. *Neural Comput. Appl.* **2020**, *32*, 495–518. [[CrossRef](#)]
52. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
53. Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **1994**, *5*, 989–993. [[CrossRef](#)]
54. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; MIT Press: Cambridge, MA, USA, 1992.

55. Althoey, F.; Akhter, M.N.; Nagra, Z.S.; Awan, H.H.; Alanazi, F.; Khan, M.A.; Javed, M.F.; Eldin, S.M.; Özkılıç, Y.O. Prediction models for marshall mix parameters using bio-inspired genetic programming and deep machine learning approaches: A comparative study. *Case Stud. Constr. Mater.* **2023**, *18*, e01774. [[CrossRef](#)]
56. Svilar, M.; Peško, I.; Šešlija, M. Model for Estimating the Modulus of Elasticity of Asphalt Layers Using Machine Learning. *Appl. Sci.* **2022**, *12*, 10536. [[CrossRef](#)]
57. Baldo, N.; Miani, M.; Rondinella, F.; Celauro, C. A Machine Learning Approach to Determine Airport Asphalt Concrete Layer Moduli Using Heavy Weight Deflectometer Data. *Sustainability* **2021**, *13*, 8831. [[CrossRef](#)]
58. Rondinella, F.; Daneluz, F.; Vacková, P.; Valentin, J.; Baldo, N. Volumetric Properties and Stiffness Modulus of Asphalt Concrete Mixtures Made with Selected Quarry Fillers: Experimental Investigation and Machine Learning Prediction. *Materials* **2023**, *16*, 1017. [[CrossRef](#)] [[PubMed](#)]
59. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
60. Van Rossum, G.; Drake, F.L. *Introduction to Python 3: Python Documentation Manual Part 1*; CreateSpace: Charleston, SC, USA, 2009.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.