

## Article

# Application of Data Mining Techniques to Predict Luminance of Pavement Aggregate

Grzegorz Mazurek \*  and Paulina Bąk-Patyna

Department of Transportation Engineering, Faculty of Civil Engineering and Architecture, Kielce University of Technology, Al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland; pbak@tu.kielce.pl  
\* Correspondence: gmazurek@tu.kielce.pl

**Abstract:** The primary purpose of the analysis presented here is to assess the feasibility of effectively predicting the aggregate luminance coefficient. Current road lighting standards and recommendations are based on assessing the level and distribution of luminance on the road surface. The brightness of a road surface depends on the amount of light falling on it, as well as the reflective properties of the road surface, which in turn depend on its physical condition, type and mineralogical composition. The complexity of the factors on which the value of the luminance coefficient depends it makes that data mining techniques the most appropriate tools for evaluation luminance coefficient phenomenon. This article uses five types of techniques: C&RT, boosted trees, random forest, neural network, and support vector machines. After a preliminary analysis, it was determined that the most effective technique was the boosted tree method. The results of the analysis indicated that the actual value of the luminance coefficient has multiple modal values within a single aggregate stockpile, depending on the mineralogical composition and grain size, and cannot be determined by a single central measure. The present model allowed us to determine the value of the luminance coefficient  $Q_d$  with a mean error of  $4.3 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ . In addition, it was found that the best aggregate for pavement brightening that allows high visibility during the day  $Q_d$  and at night RL is a limestone aggregate. In the group of those that have the ability to potentially brighten the pavement were quartzite and granite aggregates.



**Citation:** Mazurek, G.; Bąk-Patyna, P. Application of Data Mining Techniques to Predict Luminance of Pavement Aggregate. *Appl. Sci.* **2023**, *13*, 4116. <https://doi.org/10.3390/app13074116>

Academic Editors: Filipe Moura and Manuel Marques

Received: 23 February 2023  
Revised: 13 March 2023  
Accepted: 22 March 2023  
Published: 23 March 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/>).

**Keywords:** luminance; data mining techniques; aggregate; model validation; road lighting

## 1. Introduction

According to the nomenclature used in European countries, light-colored pavements are those that use light-colored aggregates and synthetic binders that are either colorless or light-colored using the corresponding pigment. Lightened pavements, on the other hand, are those that use traditional asphalt as a binder, while the aggregates are of the light-colored group. Light-colored pavements are defined as pavements whose wearing course visually gives the impression of a light-colored pavement. Invariably in these type of solutions, the important element is the aggregate itself [1].

In 1982, the Commission Internationale de l'Eclairage (CIE) [2] introduced the concept of luminance to evaluate pavement brightness. Following this, a number of methods have emerged for evaluating this parameter. Nevertheless, the evaluation of this criterion is complex and still requires an improvement of measurement methods [3]. CIE introduced the R classification system according to which each type of pavement is classified into one of four classes, from R1 to R4. Each of the four classes has a defined reflectance table (r-table) containing the reduced luminance coefficient [4]. Unfortunately, the R classification is based on measurements from 1960 to 1970, and it should be taken into account that there have been changes in asphalt pavement technology (mma) and requirements for macro-texture, in addition to the introduction of new grain sizes of mineral mixtures and types of asphalt binders [5,6].

The purpose of road brightening, assessed through luminance coefficient  $q$ , is to cause structures and traffic participants on them to be visible, minimizing driver discomfort. According to the authors of [7], pavement brightening results in an approximately 30% reduction in road accidents. It should be kept in mind that another very important part of road brightening is the reduction in the amount of energy required for adequate visibility at night. In this aspect, approx. 1.3% of the energy consumed in the EU is used by road lighting [8]. However, there are still questions of what to focus on when brightening the pavement.

Reflection from the road surface is a function of the physical condition of the road and its nature, as well as the direction of illumination and viewing conditions [5]. Different pavements can have different reflection characteristics, which depend on the texture of the pavement, its age, the paving materials, the bonding materials, and the paving methods. The reflection characteristics change with weather conditions and the physical condition (for example, when bumps or cracks appear). As the reflective properties of road pavements change, so does their luminance [9]. Luminance is closely related to the type of aggregate. Light aggregates such as: granite, gabbro, and quartzite, in contrast to limestone, are extracted from acidic rocks with high  $\text{SiO}_2$  content. Therefore, these are aggregates with a low asphalt affinity. The macro-texture of most light rocks reflects light in a diffuse rather than a directional manner, minimizing driver glare exposure. In addition, light-colored rocks such as quartzite, granite, and gabbro, in contrast to sedimentary rocks, have a high roughness coefficient, enabling an increase in the roughness of the pavement [1]. Therefore, there are some contradictions between the luminance of the aggregate and its applicability. Nowadays, it is much more common for pavement materials to be dark-colored and the standard r-table does not reflect their reflective properties [10]. Consequently, more energy is required to illuminate the road surface. Despite the fact that the type of material in the pavement affects its luminance, the phenomenon of light reflection is not directly taken into account during the design [11]. In the European Union, the design of road lighting is subject to regulation with some derogations to the member states. For the region of Poland, there is a document [12] that dictates the manner of lighting design, taking into account the luminance coefficient  $Q_0$ . In addition, there are also additional recommendations for expressways according to which the value of the luminance coefficient for pavement  $Q_d$  should be higher than  $70 \text{ mcd} \cdot \text{m}^{-2} \text{ lx}^{-1}$ , while in tunnels,  $Q_d > 90 \text{ mcd} \cdot \text{m}^{-2} \text{ lx}^{-1}$ .

In the 1990s, an alternative to  $Q_0$  was introduced in the form of the luminance coefficient under diffuse illumination,  $Q_d$ . The luminance coefficient used in the data analysis,  $Q_d$ , is intended to characterize the reflectivity of road markings in daylight or under road lighting. Portable devices, available on the market, are used in large numbers to determine the brightness of horizontal road markings and can also be used to surface roads [13]. The alternative "brightness" parameter  $Q_d$ , i.e., the luminance coefficient under diffuse illumination, has been adopted for road surface markings, but the  $Q_0$  parameter is still used in the pavement classification system adopted for road lighting. It is important to be able to accurately reduce  $Q_0$  values from r-tables to  $Q_d$ , which is included to some extent in the CIE document [14]. Despite the presence of the  $Q_d$  parameter, the parameter that is still widely accepted for assessing brightness is the mean (equivalent) luminance coefficient  $Q_0$ , which determines the weighted mean of the solid angle of the luminance coefficients in the r-table. The average luminance coefficient,  $Q_0$ , can be calculated from the r-table using the weighting coefficient procedure [15]. It was shown that the mean luminance coefficient  $Q_0$  is strongly correlated with the mean luminance produced on the road surface [5].

Unfortunately, there are still no mathematical models for predicting the luminance coefficient of an aggregate/asphalt mixture under diffused illumination  $Q_d$ . The existence of such a model or method of determination would allow the photometric properties of aggregates to be taken into account in the design of the mineral-asphalt mixture. Data mining can be proposed as an effective method providing simple regression or classification rules. However, it requires a large learning dataset. In addition, it also requires full validation of the adopted model. Solutions to complex problems, for which no a priori

mathematical model is known, in road engineering using data mining techniques are quite rare and only recently began to be applied. Data mining techniques are so far mainly used to correctly predict the mechanical or physical properties of road materials. The undeniable advantage of data mining (DM) methods is that they can take into account numerous variables, both qualitative and quantitative. In addition, they are quite resistant to the occurrence of abnormal values, which many times can carry some peculiar and interesting information about the phenomenon under study. In their paper, Rebelo et al. [16] used a number of DM techniques to effectively predict the water resistance of mineral-asphalt mixtures. Some authors have used DM to improve pavement rutting resistance [17]. In their work, Guo and Hao [18] used a random forest algorithm to assess pavement durability using information on emerging damage. The estimation of the stiffness modulus was successfully determined using a falling weight deflectometer (FWD), taking into account a number of factors related to the measurement, using artificial neural network (ANN) or support vector machines (SVMs) [19,20]. Additionally, SVM methods were used to predict the ITS parameter [21]. DM techniques have also excelled in predicting IRI [22] or skid resistance [23].

Far fewer applications of DM techniques have been reported for predicting the luminance coefficient of asphalt pavements and of the aggregate itself. It should be noted that the evaluation of these parameters is complex. It depends on the aggregate system, its origin, the macro-texture of the pavement, etc. Therefore, the degree of complexity in predicting the luminance coefficient indicates that traditional regression methods basing on the method of least squares will be neither effective nor efficient. There are very few publications that attempt to apply DM techniques to luminance coefficient evaluation. In their paper, A. Del Corte-Valiente et al., used DM techniques to manage street lighting [24]. In addition, Kazanasmaz et al., applied the ANN algorithm to predict the daylight illumination of office buildings [25]. In contrast, in the paper [26], the authors attempted to ensure adequate luminance of an indoor sports court by determining it using DM techniques. It should be noted that it is possible to use traditional techniques, such as multiple regression, but these make it impossible to predict nonlinear effects. In the article by Qin [27], a multiple regression model was used to calculate the perceived luminance in tunnel interior lighting conditions. It was pointed out that this type of regression technique is applicable only to the object under study and its generalization requires the consideration of more factors. If the number of factors is increased, the effectiveness of the multiple regression model will decrease significantly.

It should be noted that a high luminance of the aggregate does not necessarily mean that a highly bright pavement will be obtained. This will definitely depend on the texture of the pavement, the amount of binder, and the pavement's surface condition. Fotios et al. [28] indicated a significant impact of the mixture type, its structure, and the time of exposure to climatic conditions on the value of the obtained reflectance coefficient. Therefore, in order to determine the final value of the luminance coefficient, the composition of the bituminous mixture should also be taken into account. Nevertheless, the aggregate is responsible for a substantial part of the luminance coefficient value. This was indicated in the work of [11], as well as in experimental tests performed on samples of bituminous mixtures [29]. Observing the limited number of publications, it can be concluded that regarding the luminance coefficient problem, very little attention was paid. However, the minimum requirements for the aggregate luminance coefficient are frequently defined as a precondition for designing, for example, pavements in tunnels. Thus, it should be emphasized that aggregate luminance is crucial for pavement brightening [11] and a reliable prediction model of the aggregate luminance coefficient is important; its results will be presented in this article.

## 2. Materials and Methods

### 2.1. Aggregate

This study used aggregates that meet the requirements of Polish standards WT-1/2014 [28] and PN-EN 13043 [29] for the construction of wearing courses of asphalt pavements. Aggregates were taken randomly from the collection commonly available and used for pavement construction layers in the southern part of Poland. In order to increase the effectiveness of modeling, attention was paid to selecting aggregates obtained from sedimentary, metamorphic and igneous rocks. A set of aggregates was used for the analysis, and individual aggregate types were grouped by petrographic description and grain size (Table 1).

**Table 1.** Performance of aggregate set used to create a DM model.

No.	Type of Aggregate	Grain Size	LA EN 1097-2 [30]	PSV EN 1097-8 [31]	$\rho_a$ EN 1097-6 [32]	Methylene Blue Test EN 933-9 [33]	Grading of Filler Aggregates EN 933-1 [34]	Flakiness Index EN 933-3 [35]
1.	Amphibolite	0/2, 5/8, 8/11	max. 25	min. 56	2.84		max. 16 */max. 2 **	max. 20 **
2.	Basalt	2/5, 5/8, 8/11, 16/22	max. 10	min. 50	2.96		max. 1 **	max. 18 **
3.	Gabbro	2/5, 5/8, 8/11	max. 15	min. 50	2.63 ÷ 3.0		max. 1 **	max. 15 **
4.	Granite	2/8, 8/16, 16/22	max. 15	min. 50	2.65	5 ÷ 7 *	max. 1 **	max. 15 **
5.	Quartzite	0/2, 2/5, 5/8, 8/11, 8/16,	max. 25	min. 56	2.64		max. 14 */max. 1 **	max. 14 **
6.	Melaphyre	0/2, 2/5, 5/8, 8/11, 8/16,	max. 15	min. 56	2.75		max. 14 */max. 1 **	max. 18 **
7.	Limestone	0/2, 2/8, 4/11, 5/11, 5/8, 8/11, 8/16, 16/22	25 ÷ 30	44 ÷ 56	2.69 ÷ 2.72	-	max. 16 */max. 2 **	max. 17 **

\*—fine-crushed aggregate ( $D \leq 2$  mm), \*\*—coarse-crushed aggregate ( $D > 2$  mm).

Table 1 also provides information on selected physical parameters of the aggregate, which is intended for wearing course. Evaluation of the coarse aggregate crushing strength (LA) and coarse aggregate polishing resistance (PSV) features was performed only for coarse fractions. The results given are from the declarations of the manufacturers of the supplied aggregates and are the minimum value or range of the results. Limestone aggregates have by far the least favorable physical characteristics from the standpoint of making a mineral mixture composed entirely of aggregate. Therefore, the best option would be one in which the aggregate achieves a low LA and a maximum high PSV [36]. In contrast, such aggregates usually have a low asphalt affinity (quartzite and granite) or are dark in color (melaphyre).

### 2.2. Luminance Coefficient

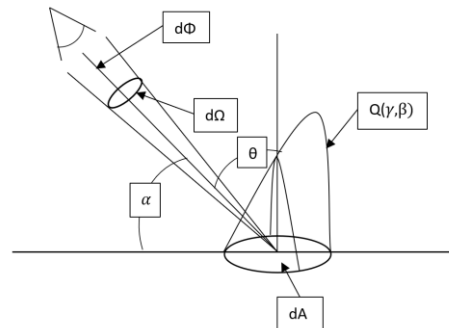
In Europe, road brightness and visibility under artificial light are related to the distribution of luminance and illumination on the road surface. The illuminance on a road surface refers only to the amount of light reaching the surface, which does not indicate how bright the surface is. Illuminance (E) is the amount of incident light (luminous flux  $\Phi$ ) per unit of area (A). The SI unit of illuminance is lux (lx). According to the SI system, illumination is expressed by Formula (1):

$$E = \frac{d\Phi}{dA} \quad (1)$$

Luminance, on the other hand, is defined as the luminous flux per unit of projected area and is a function of the illuminance on the road and the reflectance characteristics of the road surface [37] according to Formula (2):

$$L = \frac{d\Phi}{d\Omega \cdot dA \cdot \cos \theta} \quad (2)$$

where:  $\Phi$ —luminous flux,  $A$ —surface area,  $\Omega$ —solid noun,  $\theta$ —angle between the direction of the solid noun  $\Omega$  and the normal of the emitter or reflecting surface  $A$ . The geometry required to determine the luminance coefficient is shown in Figure 1.



**Figure 1.** The geometry of the luminance coefficient:  $\alpha$  (viewing angle),  $\beta$  (angle between the plane of incidence of light and the viewing plane),  $\gamma$  (angle of incidence of light), and  $\delta$  (angle between the viewing plane and the axis of the road on which the luminance coefficient  $q$  of the road surface is located depending on the viewing point  $P$  [38]).

In summary, the reflection of a surface area can be described by the luminance coefficient  $q$ , which is defined as the luminance of the surface area  $L$ , produced by illumination and reflection, divided by the illuminance of the surface plane of the area  $E$  according to Formula (3):

$$q = \frac{L}{E} \quad (3)$$

Luminance coefficient  $q$  is measured in  $\text{mcd} \cdot \text{m}^{-2} \text{lx}^{-1}$ .

### 2.3. Luminance Coefficient under Diffused Illumination $Q_d$

EN 1436 defines  $Q_d$  as the ratio of the luminance of a surface area under diffuse illumination in proportion to the illuminance at the surface plane.  $Q_d$  is measured using the spectral distribution of illumination according to a standard illuminator representing daylight and an observation angle of  $2.29^\circ$  representing a geometry (viewing distance) of 30 m. It is worth mentioning that the maximum luminance that can be obtained for a given area, when the surroundings have a constant luminance  $L$ , is  $318 \text{ mcd} \cdot \text{m}^{-2} \text{lx}^{-1}$ . In contrast, a value of approximately  $Q_d = 52 \text{ mcd} \cdot \text{m}^{-2} \text{lx}^{-1}$  is attributed to “black” pavements (for example: mastic asphalt) [13]. This type of luminance coefficient determines daytime visibility.

Road lighting creates a contrast between the luminance of a person, vehicle or object and the luminance of the immediate background, which is usually the road surface or its edges. Luminance contrast is a measure of the difference between the luminance of an object and the luminance of the background. Objects on the road are visible if they contrast with the road surface. Threshold contrast is the minimum contrast at which an observer can see an object against its surroundings. Luminance contrast is positive if the object is brighter than its background and negative if the object is darker than its background [38].

Surface reflection is strong in situations corresponding to driving into the sun. The ordinary reflection is strong in situations corresponding to driving with the sun shining from the back. Adequate light for a standard road configuration when confronted with dry and wet surface conditions is evaluated by optimizing the lighting distribution function [39].

The average luminance expressed by the  $Q_0$  coefficient, although still used, is a parameter that is difficult to measure using the small space of a portable instrument. Therefore, since the 1990s,  $Q_d$  has aimed to characterize the reflectivity of road markings in daylight or under road lighting. Portable devices are available and are used widely for road marking, and can also be applied to road surfaces. However, no portable instrument is available for measuring the specular coefficient. In this study, the luminance coefficient



$Q_d$  takes into account the effect of specular reflection, but to a lesser extent than  $Q_o$ . Therefore, the  $Q_d$  parameter can be considered a reasonable approximation of the  $Q_o$  value for determining the illuminance required for road lighting design. The  $Q_d$  coefficient is also relevant under weak daylight or twilight lighting conditions. In these cases, the human sense of vision is very sensitive to the contrast that results from dark objects against the surface of the road [39].

The normalized luminance coefficient  $Q_d$  for a pavement associated with an  $r$ -table can be determined by laboratory methods from a cut pavement sample under field conditions using a reflectometer, or by modern digital image analysis techniques [40].

#### 2.4. Surface Reflectance RL

The surface reflectance RL ( $\text{mcd m}^{-2} \text{lx}^{-1}$ ) is the quotient of the luminance  $L$  of the marking surface in the direction of observation by the illumination value  $E$  in the plane perpendicular to the direction of incident light and the area of the reflective surface. The definition of RL is very similar to that of the luminance coefficient  $q$ , and uses the same unit. The only difference is that the illuminance is not measured on the surface plane, but perpendicular to the illumination direction. Typically, the surface reflectance coefficient determines the reflectivity of pavement or horizontal markings illuminated by vehicle lighting. It is also referred to as nighttime visibility. It is important in the absence of road lighting.

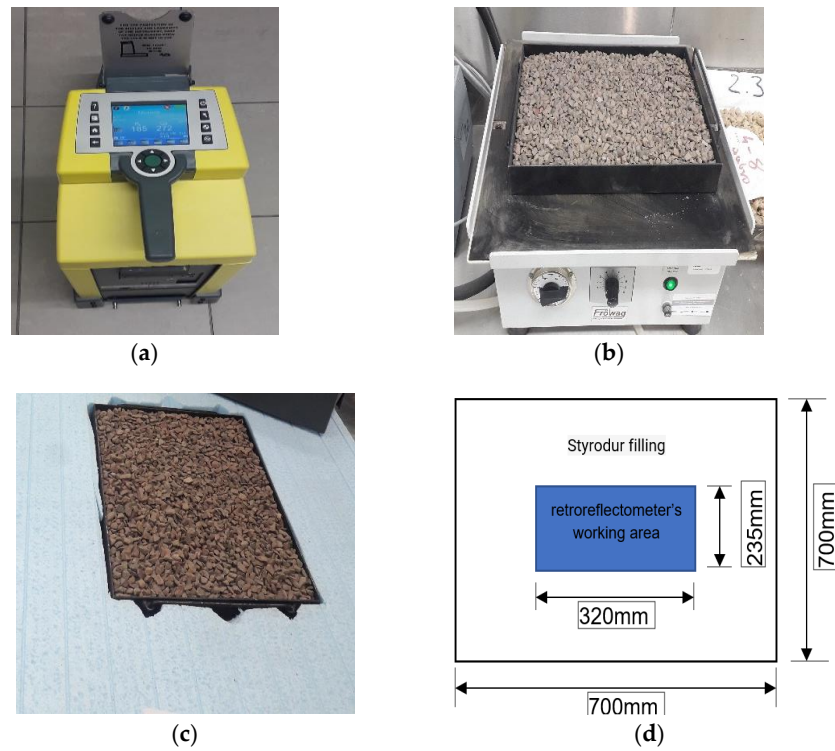
Road surfaces mostly have RL values in the range of 10 to 30  $\text{mcd m}^{-2} \text{lx}^{-1}$ . In comparison, those of road marking surfaces (excluding glass beads) can usually reach at least twice that amount.

#### 2.5. Test Stand

According to the requirements contained in Polish standards WT-2 [41] and PN-EN 1436 [42], the measurement of the luminance coefficient should be made at an angle of  $2.29^\circ$  (equivalent to viewing the road from a distance of 30 m from the position of a driver of a passenger vehicle). The measurement field should be a rectangle of 185 mm  $\times$  50 mm, with an angular span of observation of  $\pm 0.17^\circ$ . This requirement was met by the LTL-X Mark II device used to perform the tests. The device allows the determination two parameters,  $Q_d$  and RL, with the ability to import the results into a database. The device and test stands are presented in Figure 2.

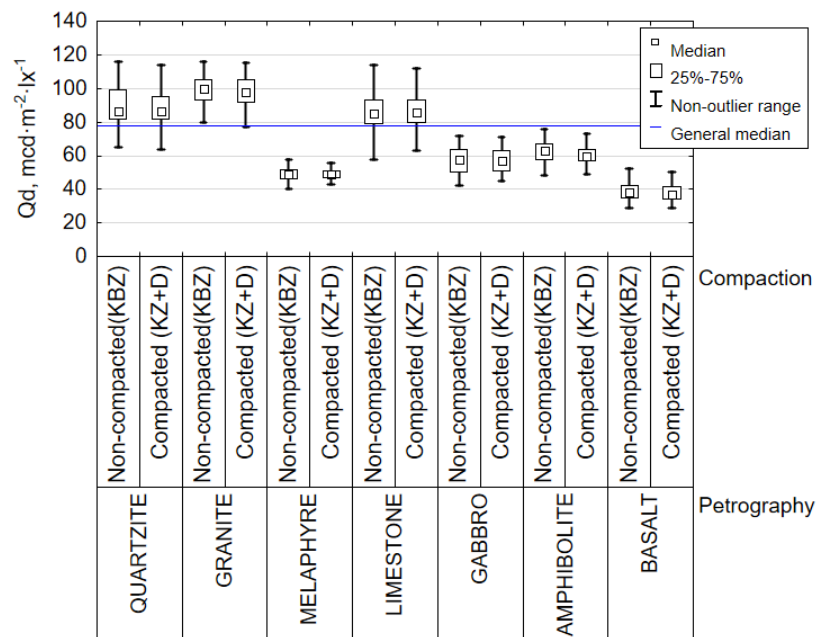
The test is performed in fixed diffused light. The sample is illuminated with a D65 standard light source of constant luminance, as specified in ISO/CIE 10526 using the aforementioned LTL-X Mark II device (Figure 2a). The measurements are performed while the ambient temperature is between 0 and 30 °C, and the temperature of the test sample/surface is between 5 and 40 °C.

The test form met the recommendations of WT-2/2014 [41], according to which the stable test area should be 700 mm  $\times$  700 mm (Figure 2c,d). Meanwhile, the empty space that is not subject to measurement should be filled with any rigid material (Figure 2c). The working area in which the aggregate was placed was sized to match the retroreflectometer's working area, i.e., 320 mm  $\times$  235 mm (Figure 2d). Each aggregate sample was tested 5 times, at different locations in the working area. Subsequently, another sample was taken from another location in the aggregate stockpile (series). Each aggregate was evaluated on the basis of a dataset of 5 series. Thus, every luminescence coefficient of aggregate was measured on subsamples taken from 5 different locations of stockpile in 5 places for each subsample. As a result, to determine coefficients  $Q_d$  and RL, 25 measurements for each grain size and aggregate type were performed. Such an amount of data was required in order to obtain a robust DM model. The number of aggregate datasets, as well as test repeatability (including reproducibility in the validation stage), should be big enough to cover the spectrum of potential cases in order to predict using a potential model.



**Figure 2.** The measuring instrument and test stand for luminance coefficient testing: (a) LTL-X Mark II road retroreflectometer; (b) laboratory compactor; (c) mold with test sample; (d) scheme test stand.

The research plan included conditions of sample surface preparations by taking into account the compaction of the sample. Therefore, in some cases, a vibrating table was used (Figure 2b). Test conditions are presented and explained in Section 3.1. Adequate repeatability, adopted as >10% (results range) [41], was obtained at a compaction time of 3 min and an amplitude of 3 mm. The use of vibrations in the mold of the sample preparation reduced the results of Qd variability up to  $3 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  for all aggregates. Summarizing the effect of compaction on variability of the Qd is presented in Figure 3.

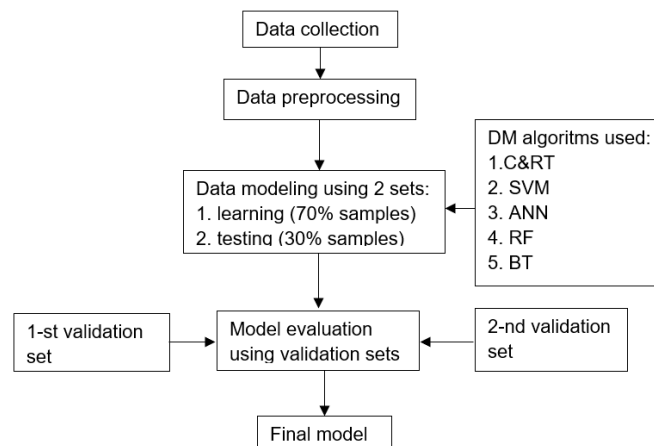


**Figure 3.** The influence of compaction on variability of Qd results.

Observing results in Figure 3, it should be noted that compaction had a slight but significant influence on decreasing the dispersion of Qd results for each aggregate type. Thereby, compaction was included as a factor in the DM model.

### 3. DM Evaluation

The analysis of the data and the prediction of the luminance coefficient Qd were divided into several stages according to the diagram in Figure 4.



**Figure 4.** Algorithm of data analysis using DM techniques to predict luminance coefficient Qd.

The first stage of the analysis involved collecting data, including the aggregate luminance coefficients (Qd, RL), and selecting factors such as petrography, grain size, and test conditions. After extracting the data, the results obtained required verification. For this reason, a pre-processing procedure was necessary. Data processing is a method of transforming raw data into a comprehensible format and assessing its representativeness. There are several techniques for accomplishing this, e.g., sampling, filling in missing data, and detecting and fixing errors [43]. After extracting the preprocessed data, several different machine learning algorithms capable of performing different prediction tasks were applied. This stage belongs to the modeling phase. To select a model whose predictive ability is the most optimal, evaluative metrics (error and correlation scores) were used to compare the model's predicted values with the observed results. To improve the efficiency of the use of DM techniques, various preliminary methods exist, such as: extracting different data, tuning the model's hyperparameters, or running the test with a different algorithm. After completing the previous steps, the model was evaluated using a test dataset. Data from two different sources were then used, with tests performed by different operators. This dataset was validating in nature and its result was crucial in formulating the final conclusion on the choice of modeling method.

#### 3.1. Analysis of the Test Set and Its Preprocessing (Data Preprocessing)

The process of identifying the best data mining technique required performing a series of pre-estimation analyses according to the diagram in Figure 3. Their primary goal was to recognize the structure of the data contained in Table 1. It should be noted that the case selection of this set of aggregates was random. The analyzed samples were subjected to an assessment of the presence of outliers and redundancy resulting from the mutual correlation of the samples. Data mining techniques are not as conservative as when using a multiple regression model; nonetheless, certain steps should be taken to improve the quality of the input set [44]. For the qualitative variables (measurement conditions, grain size, and aggregate origin), Cramér's V method for correlation assessment was used. However, for quantitative features (Qd and RL), the traditional correlation coefficient was used. The value of 0.9 was assumed as a threshold value for both the correlation coefficients  $r$  and Cramér's V, which was intended to indicate an extremely high correlation

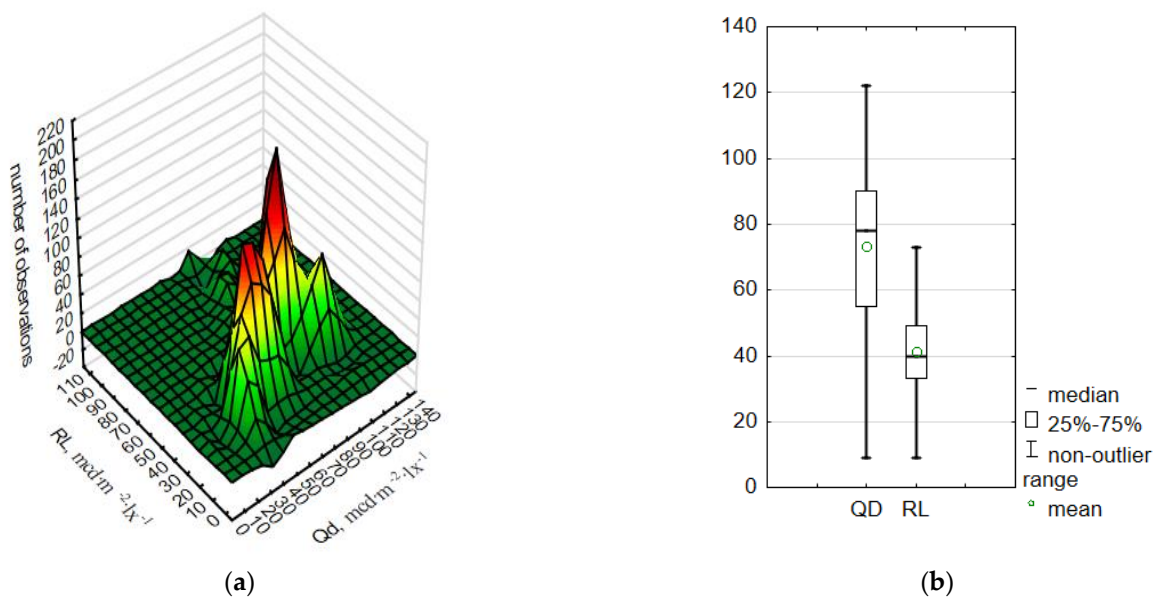


between predictors. Quantitative as well as qualitative variables were used in the analysis. Quantitative features represented parameters obtained during direct measurement using a reflectometer, whereas the qualitative features were related to the properties of the aggregate and the test conditions. A summary of the input data is presented in Table 2.

**Table 2.** Characteristics of input variables.

Quantitative Variable	
Qd, $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$	Luminance coefficient under diffused illumination
RL, $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$	Surface reflectance coefficient
Qualitative variable	
Petrography	7 types of aggregates (Table 1)
Grain size	(Table 1) Three classes: Uncompacted aggregate (KBZ)
Test conditions	Compacted aggregate (KZ) compacted aggregate with filling and leveling of the sample surface (KZ + D)

The distribution of the two quantitative variables (Qd and RL) used to build the DM model is presented in Figure 5.



**Figure 5.** Distribution of Qd and RL of tested aggregates: (a) histogram; (b) box and whisker plot.

It should be noted (Figure 5) that the largest number of cases relative to the Qd feature were in the range of two main intervals of results: the first is  $60 \div 70 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ , and the next is  $90 \div 110 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ . In contrast, the value of the  $R_L$  coefficient was in the range of  $40 \div 60 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ . The input factors studied were also evaluated for outliers. For quantitative variables, a homogeneous interval of  $\langle Q1-1.5IRQ; Q3 + 1.5IRQ \rangle$  was used. As a result, only 2% outliers were detected for the  $R_L$  feature. With regard to quality characteristics, grain size—5.3% and test conditions—3.2%. No outliers were detected in the set of other features. Such a low percentage of outliers was indicative of a balanced set; the structure and representativeness for the use of machine learning aimed at proper prediction of aggregate luminance Qd. A summary of basic statistical characteristics is presented in Table 3.

**Table 3.** Basic statistics of Qd and RL features.

Variable	Descriptive Statistics									
	Mean	Median	Minimum	Maximum	QI	QIII	SD	CV	Skewness	Kurtosis
Qd, $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$	73	78	9	122	55	90	21.8	29.7	−0.13	−0.97
RL, $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$	41	40	9	95	33	49	13.1	31.6	1.00	2.04

SD—standard deviation. CV—coefficient of variance.

The result of the high positive kurtosis of the RL feature is noteworthy. Its value suggests a clustering of its values around a central value much greater than that of the Qd feature. In addition, its results clustered around the central value with a lower standard deviation compared to the Qd feature. Road pavements typically achieve Qd values in the range of 55 to 90  $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  depending on the type of pavement and the color of the aggregate [13]. Observing the interquartile range of QIII and QI, it seems that the range of 50% of the middle abundance of the aggregates used obtained the level of the Qd coefficient in exactly the same range. In contrast, the quartile range of the RL feature from 33 to 49  $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  indicates that the contrast of the road marking with the surrounding pavement for the aggregates used is higher than the typical nighttime visibility value of the pavement RL from 10 to 30  $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  [13]. Thus, the use of aggregates from the adopted database has a significant effect on increasing the luminance coefficient of the pavement, and thus its brightness.

### 3.2. Analysis of the Selection of Regression Data Mining Technique

#### 3.2.1. Artificial Neural Network (ANN)

An artificial neural network is a type of mathematical model that learns to create and optimize a function (or distribution) that defines a set of input (learning) characteristics. The network’s learning process occurs through modifying the weighting parameters of the network of nodes, which is possible by applying certain network learning performance measures. Weight tuning parameters are generated by the learning function. The network consists of a set of neurons, each of which has an activation function (weight function) that processes the input data. The behavior of a neuron (and the entire neural network) is strongly dependent on the type of activation function used. These weights must be updated in the learning process by minimizing the estimation error. Layer node structures can be created in which data must flow in a certain direction. Different ways of connecting network nodes affect the network’s capabilities. Neurons in a network can be connected in various ways to form various topologies, which has a major impact on the network’s learning capabilities. One of the most popular types of networks in supervised learning is the multilayer, layered, forward-coupled perceptron network [45].

The optimal topology of the ANN was determined using an iterative algorithm [42]. As a result the optimal ANN topology was obtained as follows:

- ANN type: MLP 25-6-1 (multilayer perceptron of 25(input variables)-6(hidden layer perceptron)-1(output variable);
- Hidden layer activation function type: exponential;
- Output activation function type: exponential.

#### 3.2.2. Random Forests (RF)

Random forests are a set of parallel decision trees. Because the random forest classifier is based on two main random factors, it contains different decision trees. The data used to generate each tree are sampled with replacements from the learning set. The best distribution of all features is selected from a random subset of them [45]. Random forest works by fitting multiple decision trees to different subsamples of a dataset and applying averaging to increase predictive accuracy and monitor overtraining. Typical individual decision trees are characterized by high variance and a tendency to overgeneralize [46]. The randomiza-

tion is intended to reduce the variance of the estimator. Forest regression is believed to handle diverse tasks well and has the potential to deal with non-linear relationships [47].

For obtaining the best model, some parameters had to be set in order to reproduce a final model. In the experiment the following calculation principles was used:

- Number of additive terms(trees): 300;
- Random test data proportion: 0.3;
- Subsample proportion: 0.5;
- Minimum n in child node: 1;
- Maximum n of nodes: 3;
- Minimum n of cases: 5;
- Maximum n of levels: 10;
- Cycles to calculate mean error: 10;
- Percentage decrease in training error: 5.

### 3.2.3. Boosted Trees (BT)

The aforementioned random forests are a combination of tree predictors. Therefore, each tree depends on the value of a random sampled vector. The generalization error for random forests converges to the limit, as the number of trees can reach a very large value. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using random feature selection to split each node gives error rates that are more favorable using the Adaboost algorithm [48] and are more reliable with respect to noise. The boosted tree algorithm evolved from the use of methods of boosting to regression trees. The main idea is to create a sequence of (very) simple trees, where each successive one is built to predict the residuals generated by the previous one. It can be shown that such a procedure of “additive weighted expansion” of trees will result in a perfect fit between predicted and observed values, even if the very nature of the relationship between the predictors and the dependent variable is very complex (e.g., non-linear). Thus, the gradient enhancement method—fitting weighted additive expansion of simple trees—is a very general and powerful machine learning algorithm. The decision tree is commonly used for classification and regression tasks because it offers many advantages, such as high efficiency, simplicity, and interpretability [49]. The boosting strategy was originally developed for classification problems, but demonstrates successful application to regression tasks. Random tree boosting is a technique that builds each regression tree step by step, using a predefined loss function to measure the error at each step and correct it with each subsequent step. Therefore, a predictive model is actually a collection of weaker predictive models [50]. Gradient amplification takes into account additive models, which are taught in stages of the form [51] (4):

$$F_m = F_{m-1}(x) + h_m(x) \quad (4)$$

where:  $h_m(x)$ —weak learning functions. When applying the gradient boosting strategy to regression trees, the small regression trees are the basis functions, and the entire regression model using the boosted tree method is the sum of them.

In the experiment for testing, the following calculation principles were used:

- Number of additive terms(trees): 300;
- Random test data proportion: 0.3;
- Subsample proportion: 0.5;
- Minimum n in child node: 1;
- Maximum n of nodes: 3;
- Minimum n of cases: 5;
- Maximum n of levels: 10.

### 3.2.4. Support Vector Machines (SVMs)

At the core of the support vector machines (SVMs) method is the concept of a decision space, which is divided by building boundaries separating objects with different classifications. The support vector machines method performs classification tasks by constructing hyperplanes in a multidimensional space that separate cases belonging to different classes. However, regression can also be performed, as well as both tasks, for multiple variables, both continuous and categorized. For each categorized variable, a set of variables is created with codes indicating the classification of each case (0 or 1). For example, a variable taking three values; A, B, and C; will be represented by three variables. A key element of the SVM method is the selection of the form of the kernel function and the model parameters coupled to it. There are several types of kernel functions to use in the SVM model: linear, polynomial, radial basis functions (RBF), and sigmoid function. The support vector machines (SVMs) method is a comprehensive machine learning model that enables both linear and non-linear classification, regression, and identification of outliers. SVR differs from simple regression, which aims to minimize the error between the predicted and actual values, because SVR aims to fit the error (usually the square of the residuals) into an interval with a certain threshold value. In this way, the SVR method attempts to find the closest fit between the actual data points and the function representing them [52].

In the experiment for analysis, the preliminary evaluation of some important parameters needed to be determined. To achieve this, a special script was elaborated in the R programming language and then was exported in the Statistica<sup>®</sup> program. The main parameters, the best for this dataset, were given below (in the Statistica<sup>®</sup> program):

- Regression type 1: (coefficients:  $C = 1$ ,  $\epsilon = 0.5$ );
- Kernel type: radial basis function (RBF,  $\gamma = 3$ );
- WN number 41.

### 3.2.5. Classification and Regression Trees (C&RT)

Analysis using C&RT involves recursively dividing observations into disjoint subsets. For regression analysis, the dependent variable must be interval in nature. In this type of analysis, as in other data mining methods, the relationships between variables are not necessarily linear. The authors of the C&RT algorithm recommend this type of analysis when the assumptions of multiple regression are met [53]. The distribution rules are related to the variance or average deviation. The key to obtaining a stable C&RT model is to build several trees for the same number of subsets in multiple cross-validation technique, but with different random number generator settings. Subsequently, the tree structures and their ranking are compared. Their convergence demonstrates the stability of the model. Algorithm C&RT needed some additional conditions to account for relations in a dataset. This algorithm is not so effective as RF or BT, but is still useful and simple:

- Number of additive terms: 1000;
- Bonferroni adjustment;
- Minimum n in child node: 1;
- Maximum n of nodes: 5;
- Minimum n of cases: 5;
- Probability for splitting: 0.05;
- Probability for merging: 0.05.

## 4. Analysis of the Research Results

### 4.1. Evaluation of the Effectiveness of Selected DM Techniques

The modeling process began by performing a series of simulations using the algorithms presented in Section 3, namely: C&RT, RF, BT, ANN, and SVM. In the case of ANN and SVM, some additional work was required to find the optimal parameters. In the case of ANN, a series of iterative steps were required to determine the optimal structure of the neural network in terms of the type of activation function and the number of neurons in

the hidden network. On the other hand, in the case of SVM, the parameters of the radial functions C are part of the Kernel function of the SVM method [50]. Qualitative evaluation of DM technique selection was performed using popular regression metrics and is provided in Table 4.

**Table 4.** Regression metrics of the DM method.

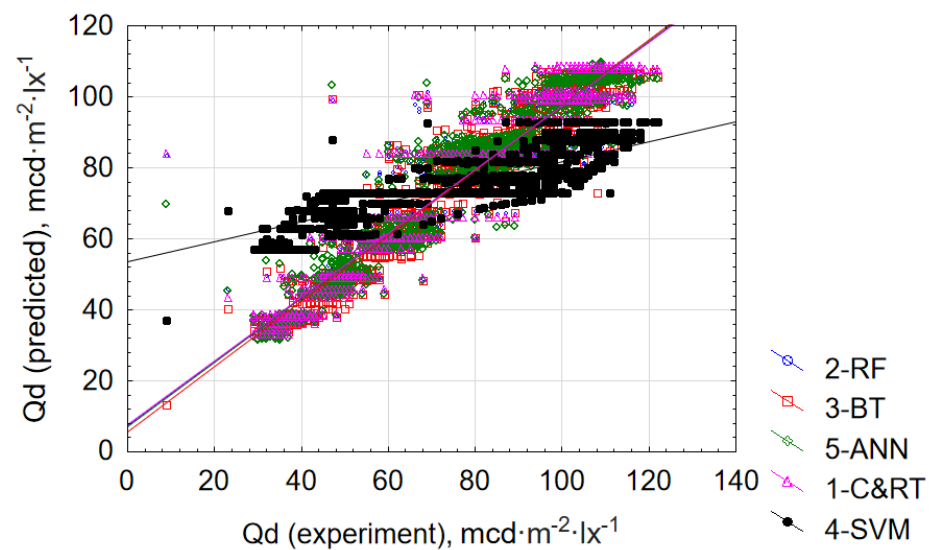
Metrics	Description	Form
R-squared	$R^2$	$\left(\frac{\sigma_{yy'}}{\sigma_y \cdot \sigma_{y'}}\right)^2$
Root mean square error	RMSE	$\sqrt{\frac{\sum_{i=1}^N (y_i - y_{\text{mean}})^2}{N}}$
Mean absolute error	MAE	$\frac{\sum_{i=1}^N  y_e - y_p }{N}$

$y_p$ —predicted value,  $y_e$ —experimental value,  $y_i$ —i-th observed value,  $y_{\text{mean}}$ —mean value,  $\sigma_y$ —variance of experimental data,  $\sigma_{y'}$ —variance of predicted data,  $\sigma_y \sigma_{y'}$ —covariance of predicted and experimental data.

The selection of the final DM technique required obtaining a minimum MAE or RMSE and a maximum  $R^2$  value. It should be kept in mind that the feature for which DM-predictive techniques were used was the luminance coefficient Qd. In the first stage, the dataset containing 1895 valid cases was divided into two subsets:

- a teaching set containing 70% of randomly selected cases;
- a test set containing 30% of randomly selected cases.

Following the stage of constructing a series of DM models, a collection of predicted values of the luminance coefficient Qd was obtained, which are shown collectively in Figure 6.



**Figure 6.** Fit of model results with experimental results.

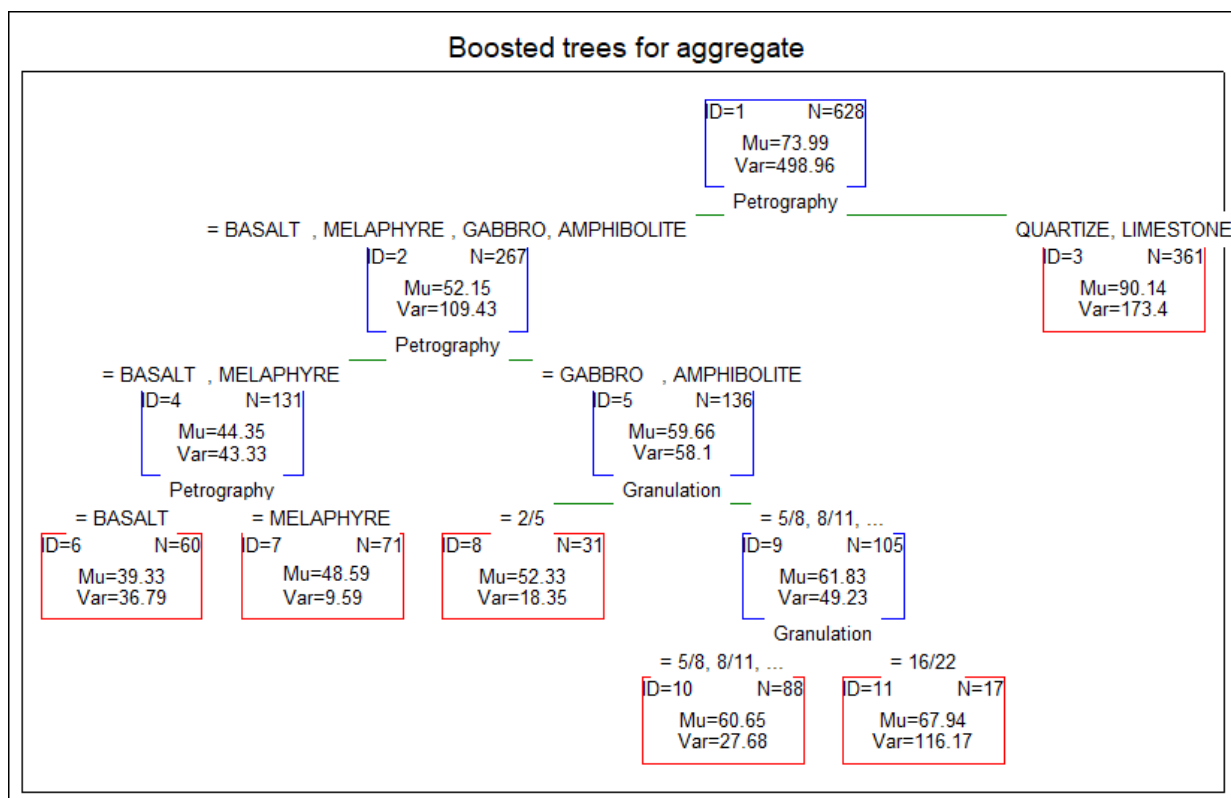
Observing the results in Figure 6, it should be noted that the least impressive DM technique, among those used, was SVM. In this case, the obtained model did not allow proper prediction of large and small values of the luminance coefficient Qd, which do not necessarily represent outliers. This is probably due to the fact that the qualitative variables were on a nominal rather than an ordinal scale. Other techniques have shown far greater effectiveness. Nevertheless, observing the influential values occurring (extreme values on the regression line), it can be expected that the BT technique was by far the most effective in relation to the problem in question, which was the effective prediction of the Qd feature. In order to make a final assessment as to the choice of the appropriate technique, the characteristics given in Table 4 were used, while their calculated values are shown in Table 5.



**Table 5.** Summary of regression metrics of fit quality.

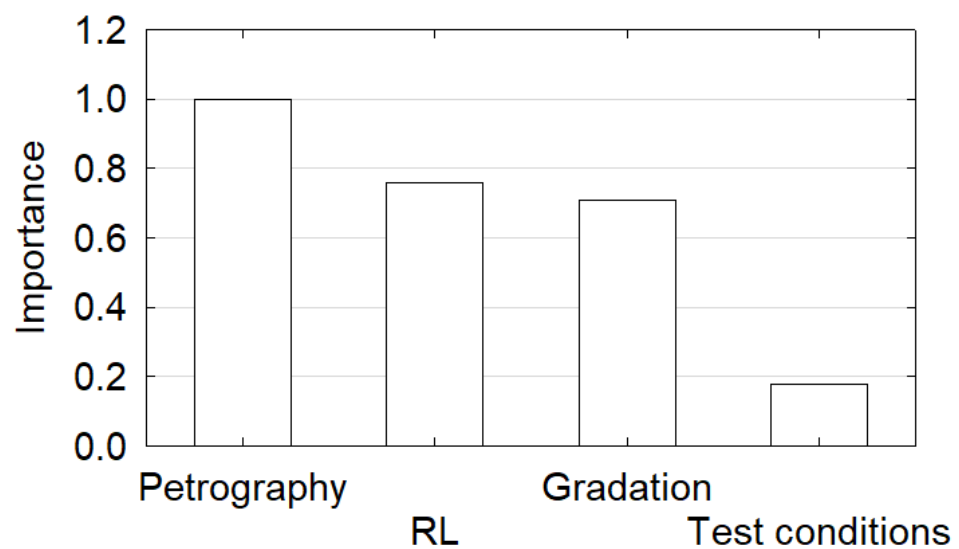
DM Method	MAE	RMSE	R <sup>2</sup>
RF	4.5	42.7	0.95
<b>BT</b>	<b>4.3</b>	<b>37.9</b>	<b>0.96</b>
ANN	4.7	44.6	0.95
C & RT	4.9	48.9	0.95
SVM	14.5	268.2	0.80

As previously speculated, the Boosted Trees (BT, highlighted) algorithm proved to be the best DM technique for predicting the value of the luminance coefficient Qd of an aggregate. During the construction of the BT model, a learning ratio of 0.1 was used [54], while the ratio for drawing learning samples in successive boosting steps was 0.5, based on the recommendations provided in the paper [55]. Its results were very similar relative to all other DM techniques, with the exception of SVM. The SVM method had by far the highest value of variability expressed by RMSE characteristics. Through this, the MAE value (estimation error) was also very large, at 14.5 mcd m<sup>-2</sup> lx<sup>-1</sup>. Such a value of variability in the SVM method excluded the possibility of correct discrimination of aggregates of similar color, for example: quartzite and limestone. To counteract this, the BT model also took into account the occurrence of the same aggregate, i.e., limestone and quartzite from two different mines in the Świętokrzyskie voivodeship, Poland. The rules for creating a boosted tree are easy to interpret, but the notation of the established scoring system assigned to the nodes is very complex. Therefore, the end result of using BT is a model encoded in the form of an XML (Extensible Markup Language) file, which was created to represent various data in a structured way that allows for easy implementation into other data analysis systems. It is currently the best language for data presentation, recommended by the W3C organization [56]. From the perspective of practical applications, the set of rules in graphical form is the most useful (Figure 7).



**Figure 7.** Structure (rules) of the results using the BT technique.

It should be noted that the average result of the luminance coefficient  $Q_d$  of the entire set ( $ID = 1$ ) is  $73.9 \text{ mcd m}^{-2} \text{ lx}^{-1}$ , whereas the scatter of the results (variance) is significant at  $499 \text{ mcd m}^{-2} \text{ lx}^{-1}$ . Thus, that a set of rules can be built to quickly distinguish  $Q_d$  values against different aggregates. A distinct group can be observed first, which are aggregates of limestone and quartzite, which have the greatest ability to brighten the potential pavement. Rules were defined for this “leaf” in the BT model, which did not further split ( $\text{Mean}(Q_d) = 90.1 \text{ mcd m}^{-2} \text{ lx}^{-1}$  for  $ID = 2$ ). In contrast, much greater variation (larger rule-set) was obtained for the other darker-colored aggregates ( $\text{Mean}(Q_d) = 52.1 \text{ mcd m}^{-2} \text{ lx}^{-1}$  for  $ID = 2$ ). As a result of the distribution, basalt aggregate achieved an average value of  $Q_d = 40 \text{ mcd m}^{-2} \text{ lx}^{-1}$ , while melaphyre achieved  $49 \text{ mcd m}^{-2} \text{ lx}^{-1}$  with the smallest scatter of results. In the cases of gabbro and amphibolite, the value of the predicted luminance was affected by additional rules related to grain size. Therefore, it can be presumed that the texture or shape of the aggregate affected the value of the luminance coefficient. Ylien A. et al., also came to similar conclusions in their paper [11]. Evaluating the structure of the tree (Figure 7) as a whole is quite complex, particularly when there are more “leaves”, however, some useful properties that can help with this can be derived. The most commonly used summary is the parameter characterizing the importance of a feature, indicating how important each feature was for decision processes in the regression tree. This is a number between 0 and 1 for each function, where 0 denotes “not used at all” and 1 denotes “perfectly predicts the objective”. The significance of features always adds up to 1. The validity of the features can be visualized, as shown in Figure 8.



**Figure 8.** Validity chart of BT model predictors.

The results in Figure 8 confirm the fact that luminance and related rules were most influenced by aggregate petrography. The second most important factor was nighttime visibility RL, with aggregate grain size being of equal importance. In contrast, test conditions that take into account the preparation of the aggregate surface during the test had a marginal effect on the prediction of the luminance coefficient  $Q_d$ . Therefore, the mineral composition and treatment of the aggregate and its potential texture had the most significant impact on the aggregate luminance coefficient  $Q_d$ . A summary of the effect of grain size variation and petrography on the  $Q_d$  result is shown in Figure 9.

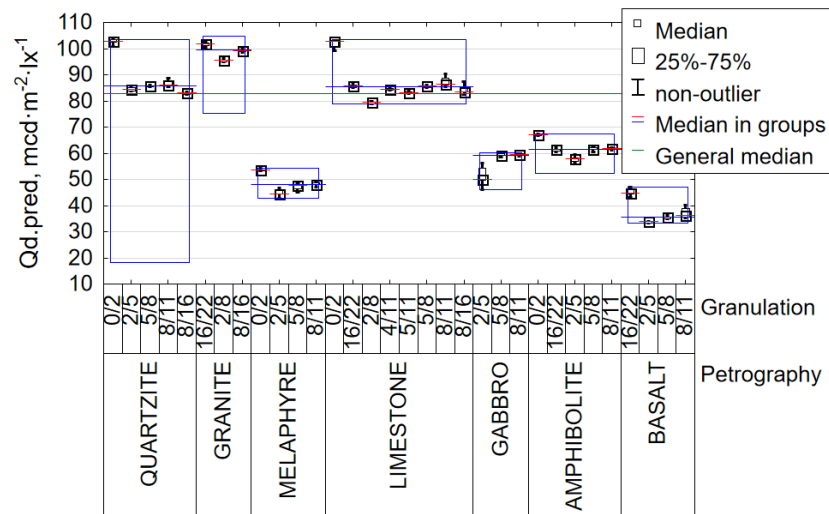


Figure 9. Plot of Qd variation against aggregate type and grain size.

The similarity between quartzite, granite, and limestone aggregate is observed in the range of the finest 0/2 fraction. The coarser the aggregates, the more luminance coefficient values begin to diverge. The median luminance coefficient of all results was about  $80 \text{ mcd m}^{-2} \text{ lx}^{-1}$ , which was higher than the Polish requirement to use aggregates for pavement brightening with  $Q_d > 60 \text{ mcd m}^{-2} \text{ lx}^{-1}$  as per WT-2/2014 [41]. In terms of Polish recommendations, the recommended fractions for luminance coefficient testing are 4/8, 5/8 or 8/11. Nevertheless, numerous other aggregate fractions were used for the purpose of teaching the model. It should be noted that 5/8 and 8/11 fractions in the models yielded very similar values of the median luminance coefficient of the tested aggregates (Figure 9). Therefore, the luminance coefficient test should be performed for coarse aggregate  $> 4 \text{ mm}$ , which confirms the validity of the proposed fractions in the adopted recommendations, although the tests performed for the 8/11 aggregate fraction were characterized by a larger interquartile range ( $Q_3 - Q_1$ ). In addition, outliers were observed representing the  $Q_d$  results of aggregate with 0/2 grain size in all aggregate cases. The  $Q_d$  value for this fraction obtained a much higher value than the other cases of a given aggregate. Therefore, this type of fraction should not be used as representative, whereas its presence in the tests was intended to increase the effectiveness of the BT model rules. A summary of the distribution of the luminance coefficient is shown in Figure 10.

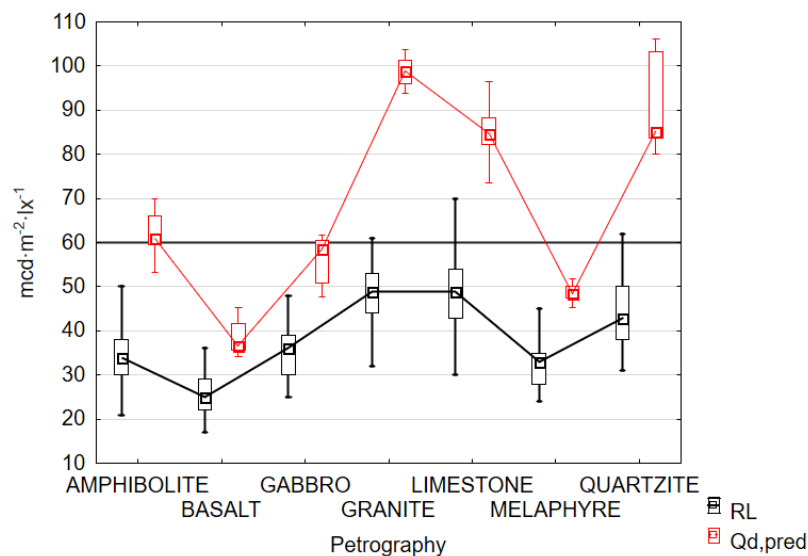
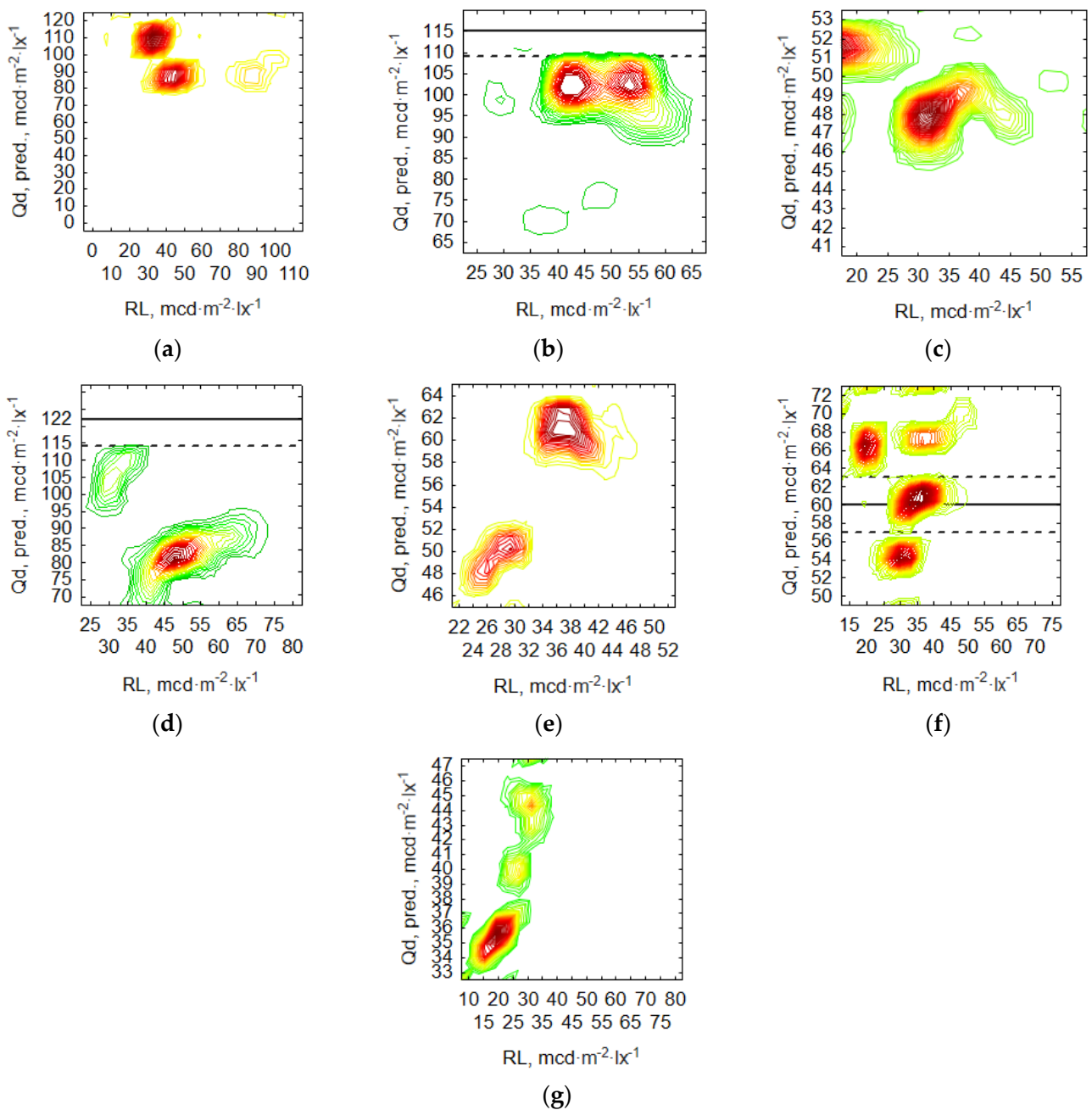


Figure 10. Distribution of  $Q_d, \text{pred}$  (predicted) and RL (box: 25–75%; whisker: non-outlier).

A preliminary comparison of both luminance coefficients ( $Q_{d,pred}$  and RL) in Figure 10 suggests that the most suitable aggregates (median) for pavement brightening are granite ( $Q_d = 99 \text{ mcd m}^{-2} \text{ lx}^{-1}$ ), limestone ( $Q_d = 85 \text{ mcd m}^{-2} \text{ lx}^{-1}$ ), and quartzite ( $Q_d = 86 \text{ mcd m}^{-2} \text{ lx}^{-1}$ ). All offer definitely greater luminance coefficient values than the recommended baseline, i.e.,  $60 \text{ mcd m}^{-2} \text{ lx}^{-1}$  [41]. Aggregates that yielded approximately borderline results are amphibolite ( $Q_d = 61 \text{ mcd m}^{-2} \text{ lx}^{-1}$ ) and gabbro ( $Q_d = 59 \text{ mcd m}^{-2} \text{ lx}^{-1}$ ). Therefore, they can be used conditionally as alternative pavement-brightening solutions. The others do not achieve the intended pavement-brightening effect. An interesting issue is the value of the RL coefficient, which does not fully correlate with the values of the predicted  $Q_{d,pred}$ . Visibility at night probably depends on the presence of certain minerals in the aggregate, which allow significant luminance of the aggregate when illuminated by vehicle lighting. This is a similar case to a situation when horizontal markings containing glass beads in the paint are evaluated. Typically, the introduction of beads increases nighttime visibility (RL). However, it significantly reduces the  $Q_d$  value, which reflects the perception of pavement illuminated by sunlight, i.e., illumination incident at a different angle than that of vehicle lighting. Quartzite and granite aggregates, despite their high  $Q_{d,pred}$  values, did not achieve the same range of RL as limestone aggregate. Therefore, it can be concluded that the best solution for brightening pavement is to use common limestone aggregate. Very relevant to a given analysis are the results of the interaction between the predicted value of the luminance coefficient  $Q_{d,pred}$  and the accompanying RL value, as shown in Figure 11.

Figure 11 is essentially a collection of histograms. It allows the determination of the local peaks (frequencies) of the luminance coefficient  $Q_{d,pred}$ . The quantitative nighttime visibility parameter RL was used as an accompanying variable. It should be clearly emphasized that the use of a single central measure to describe a set such as the mean or median does not allow an exhaustive characterization of the distribution of the luminance coefficient (as in Figure 10). This is related to a number of phenomena among which grain size and the associated alignment of aggregate grains during testing (importance = 0.7 based on Figure 8) can be mentioned. Nevertheless, it can be assumed that the occurrence of numerous local modal values of the distribution of  $Q_{d,pred}$  results can also be attributed to the typical randomness of the mineralogical composition on the aggregate surface, resulting from the process of its crushing or the deposit occurrence. Aggregate was taken from various locations of the heap, so in all likelihood this effect could have been significant. It should be kept in mind that the large number of random effects influences the use of the final efficiency of the boosted tree model. Amphibolite aggregate has four modal values (Figure 11f), while quartzite and granite have two distinct modal values. By far the largest discrepancies between modal values were observed for gabbro aggregate (Figure 11e). Comparing two aggregates, quartzite and limestone, both modal values have almost identical  $Q_d$  at the corresponding RL level. Nevertheless, by far the majority of cases are in the value range of  $80 \div 85 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  with  $RL_{45} \div 50 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ , whereas, in the case of quartzite, both modal values have almost identical values. Therefore, the use of this BT model has the advantage of being able to describe and identify such peculiarities in the value of the luminance coefficient  $Q_d$ . When using traditional estimators, the variability of  $Q_d$  results for quartzite and limestone aggregate would be considered insignificantly different. For additional validation, the luminance coefficient measurement results of aggregates from the northern region of Poland published by Wasilewska et al., were compared [57]. The results of the average values (solid line) are highlighted in Figure 11 with the range of the results (dashed line). The obtained results of these comparisons suggest that the presented model and results from other regions of Poland allow prediction for aggregates from other sources. It should be mentioned that the basic structure of the BT model was created on the basis of aggregates from one region, and further “fine-tuning” of its predictive capabilities requires successive additions to the result base.



**Figure 11.** Histogram of Qd,pred against RL: (a) Quartzite; (b) Granite (compared to results [52]); (c) Melaphyre; (d) Limestone (compared to results [52]); (e) Gabbro; (f) Amphibolite (compared to results [52]); (g) Basalt.

In conclusion, some critical issues should not be overlooked. As already mentioned, limestone, granite, and quartzite aggregates had high Qd and RL. In the case of limestone aggregate, it should be noted that it is characterized by a high LA and the lowest PSV (Table 3) by what, in some countries, may be a limitation in its wider use for wearing courses for heavy-traffic routes. Quartzite and granite aggregates, on the other hand, are known for their high  $\text{SiO}_2$  content and thus offer low adhesion with asphalt despite the fact that they have a high and favorable PSV. Therefore, the ability to correctly predict Qd is valuable information that can be further taken into account in the optimal design of the bituminous mixture composition.



#### 4.2. Validation of the Adopted Boosted Trees Model

The final validation of the adopted DM model is crucial to verifying the reasoning adopted and the correctness of the rules adopted. Model validation was performed using an additional two validation sets containing drawn aggregate types. Each harvest was subjected to Qd and RL determination by two independent operators and included in its entirety the step of preparing aggregates and performing a series of tests again. Such an action was intended to take into account the effect of randomness due to the human factor, randomness due to the location on the aggregate collection heap, and to confirm the stability of the obtained DM model. Based on previous experiments, it was determined that it would be most appropriate to prepare the samples with surface compaction and supplement the sample level with additional aggregate, caused by the compression of the compaction process (KZ + D based on Table 2). Batches of aggregates with the characteristics listed in Table 6 were drawn for verification analysis.

**Table 6.** Selection of aggregates for the validation stage.

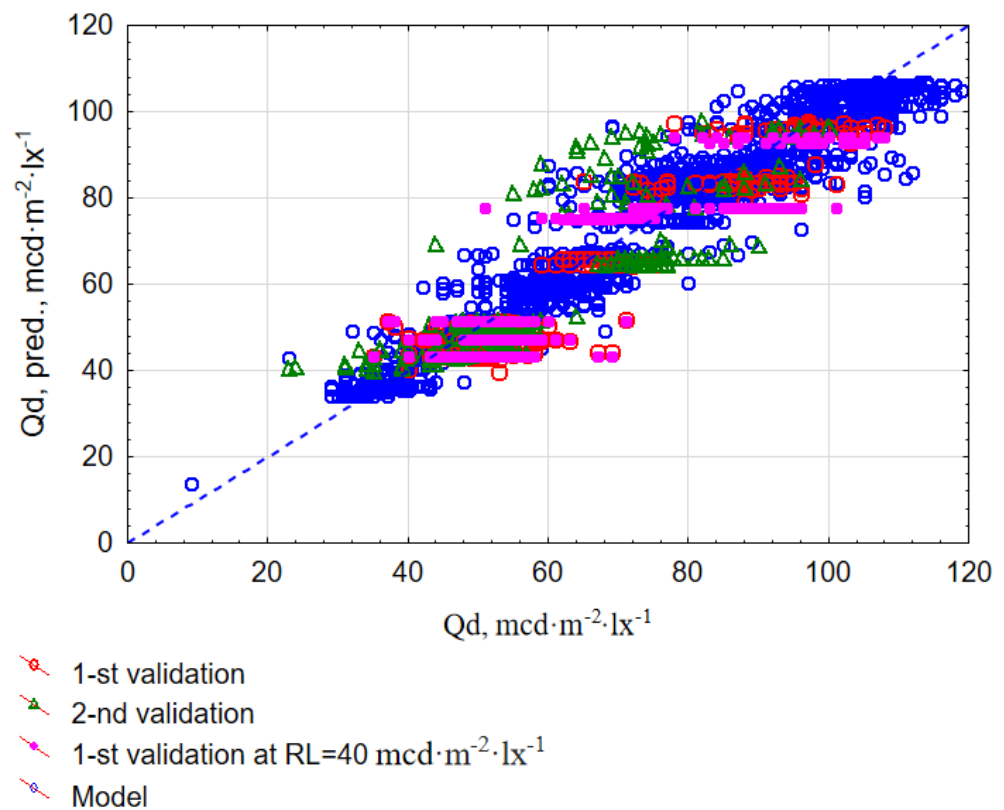
No.	Type of Aggregate	Grain Size
1.	<b>Amphibolite</b>	<b>0/2, 5/8, 8/11</b>
2.	<b>Basalt</b>	2/5, 5/8, 8/11, <b>16/22</b>
3.	<b>Gabbro</b>	<b>2/5, 5/8, 8/11</b>
4.	<b>Granite</b>	2/8, <b>8/16</b> , 16/22
5.	Quartzite	0/2, 2/5, 5/8, 8/11, 8/16,
6.	<b>Melaphyre</b>	0/2, 2/5, <b>5/8</b> , 8/11, 8/16,
7.	<b>Limestone</b>	0/2, 2/8, 4/11, <b>5/11</b> , 5/8, 8/11, 8/16, 16/22

In Table 6, the aggregates selected, along with the grain size, are highlighted in bold font. In the next step, retroreflectometer tests were performed in the same way as when the results were obtained to build the learning and testing set. The accumulated database of results was transferred in order to determine the value of the predicted luminance coefficient Qd according to the set of rules in Figure 7. As a result, using the metrics given in Table 4, the following fits between model and experimental data were obtained (Table 7).

**Table 7.** Summary of regression metrics of quality of fit of validation stage of Qd [ $\text{mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ ].

DM Method	MAE	RMSE	R <sup>2</sup>
<b>Boosted trees (teaching and testing)</b>	<b>4.3</b>	<b>37.9</b>	<b>0.96</b>
1st validation	7.8	76.1	0.76
2nd validation	5.7	53.2	0.85

The validation process confirmed the stability of the model with regard to the aggregate base available in the Świętokrzyskie voivodeship, Poland. The mean absolute error (MAE) of the residuals was larger by up to  $3.5 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  (1st validation) than the value obtained by the learning and test sets. Moreover, it is less than half of the recommended maximum Qd result range of 10% of its mean value. Therefore, it can be concluded that results received from the validation set did not differ from the adopted requirements of WT-2/2014 [41], and the error discrepancy was negligible in comparison to the obtained machine learning model. Finally, the values of single readings of the luminance coefficient Qd validation process, along with its prediction, were projected in Figure 12.



**Figure 12.** Fit of experimental results against the Qd feature model.

Green and red colors indicate the set of validation results. According to the results in Table 7, the “2nd validation” set was characterized by a slightly higher dispersion of results than the “1st validation” set. Nevertheless, the vast majority of results did not deviate from the base set (blue color), on the basis of which the rules of the random tree model were built. Therefore, the efficiency of the BT model built for Qd prediction was satisfactory, and the prediction results were stable.

However, obtaining an adequate prediction using BT rules requires the use of a popular retroreflector with the ability to determine the RL coefficient. The variant of the device that realizes only the measurement of the RL coefficient is a cheaper, and at the same time very popular, option used to assess the quality of horizontal markings in terms of adequate nighttime visibility. Given the high kurtosis of the RL results and their low variability, an additional simplified simulation of the luminance coefficient Qd was performed. The case of unavailability of a retroreflector and taking the central value from the study (median) of the RL coefficient =  $40 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  was considered. The result was determined using the results of Qd introducing a constant value of  $\text{RL} = 40 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  from the “1st validation” set projected in Figure 12 as “1.1st validation”. The value of the calculated MAE for the Qd,pred factor was  $9.2 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ . Applying this simplification slightly increased the estimation error by  $2 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  and resulted in some weakening of the prediction ability against aggregates with  $\text{Qd} > 80 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ . Nevertheless, the efficiency of the model remains high enough that accepting simulation results with such an established simplification can be satisfactory in certain cases.

In conclusion, thanks to the Boosted Trees (BT) algorithm, it is possible to quickly create an effective model for predicting the luminance coefficient Qd and use it to evaluate the brightness of aggregate. Further replenishment of the aggregate base, including their mixture in varying proportions, will allow the model’s predictive capabilities to improve dramatically and is the goal of further research. In addition, work is underway to implement the model rules into the C++ or Python programming languages.

## 5. Conclusions

Based on the research and analyses performed, the following conclusions were formulated:

- Based on the preliminary analysis, the highest efficiency was achieved by the Boosted Tree (BT) model, which, despite the similar value of the coefficient of determination  $> 0.95$ , against the neural network technique had the lowest average absolute error  $< 4.3 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ .
- In the analysis presented here, it was observed that one aggregate can have several modal values of Qd, which depend on their mineralogical composition. Therefore, only DM techniques are able to correctly describe the variability resulting from the aforementioned random factors, taking into account the numerous sources of aggregates.
- The significant overestimation of Qd for aggregate with a grain size of 0/2, regardless of its petrographic origin.
- Obtained results are in line with WT-2/214, which suggests that coarse aggregate fractions with a grain size of  $>4 \text{ mm}$  should be used to determine the luminance coefficient.
- The presence of certain minerals of a given aggregate causes the luminance coefficient that characterizes daytime and nighttime visibility to reach significant value variations. In this study, it was proven that the petrographic description has the greatest impact on Qd.
- The lack of a full correlation between Qd and RL suggests that nighttime visibility is related to the presence of certain “inclusions” of individual minerals that enhance the contrast of the pavement when illuminated by vehicle lighting.
- The best option for pavement brightening was obtained by limestone aggregate. The luminance coefficient determining daytime visibility (median) has a value of  $Qd = 85 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ , nighttime visibility  $RL = 49 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ , and a very good asphalt affinity. Other suitable aggregates for pavement brightening, but with lower RL values, were quartzite ( $Qd = 86 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ ,  $RL = 48 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ ) and granite ( $Qd = 99 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ ,  $RL = 43 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$ ). Aggregates that reached the threshold value of  $60 \text{ mcd}\cdot\text{m}^{-2}\cdot\text{lx}^{-1}$  were amphibolite and gabbro. The other aggregates in the data set did not allow for an acceptable level of pavement brightness enhancement.

**Author Contributions:** Conceptualization, G.M.; Methodology, G.M.; Validation, P.B.-P.; Formal analysis, G.M.; Investigation, P.B.-P.; Resources, P.B.-P.; Data curation, P.B.-P.; Writing—review and editing, G.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by the program of the Minister of Science and Higher Education under the name: Regional Initiative of Excellence in 2019–2022, project number 025/RID/2018/19, financing amount PLN 12,000,000.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research was conducted as a part of the project “The use of recycled materials” under the RID (The Development of Road Engineering Innovations, number DZP/RID-I-06/1/NCBR/2016), project of the National Center for Research and Development—Poland, and the Polish General Directorate for National Roads and Motorways. The work is supported by the program of the Minister of Science and Higher Education under the name: Regional Initiative of Excellence in 2019–2022, project number 025/RID/2018/19, financing amount PLN 12,000,000.

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

## References

1. Filipczyk, M.; Kukielska, D. Bright and Bleached Surfaces. Theory and Practice. *Min. Sci.* **2016**, *23*, 17–22. [[CrossRef](#)]
2. CIE. *Road Lighting Calculations*; CIE Central Bureau: Vienna, Austria, 2000; ISBN 978-3-901906-54-1.
3. Strbac-Hadzibegovic, N.; Strbac-Savic, S.; Kostic, M. A New Procedure for Determining the Road Surface Reduced Luminance Coefficient Table by On-Site Measurements. *Light. Res. Technol.* **2019**, *51*, 65–81. [[CrossRef](#)]

4. CIE. *Road Surfaces and Lighting: Joint Technical Report CIE/PIARC*; Photocopy ed. 2008; CIE Central Bureau: Vienna, Austria, 2008; ISBN 978-3-901906-72-5.
5. Bodmann, H.W.; Schmidt, H.J. Road Surface Reflection and Road Lighting: Field Investigations. *Light. Res. Technol.* **1989**, *21*, 159–170. [[CrossRef](#)]
6. Iwański, M.; Mazurek, G.; Buczyński, P. Bitumen Foaming Optimisation Process on the Basis of Rheological Properties. *Materials* **2018**, *11*, 1854. [[CrossRef](#)] [[PubMed](#)]
7. CIE. *Road Lighting as an Accident Countermeasure*, 1st ed.; Internationale Beleuchtungskommission, Ed.; CIE Central Bureau: Vienna, Austria, 1992; ISBN 978-3-900734-30-5.
8. Van Tichelen, P.; Jansen, B.; Geerken, T.; Vanden Bosch, M.; Van Hoof, V.; Vanhooydonck, L.; Vercalsteren, A. *Final Report Lot 9: Public Street Lighting*; Ökopol: Hamburg, Germany, 2007.
9. CIE. *Calculation and Measurement of Luminance and Illuminance in Road Lighting: Computer Program for Luminance, Illuminance and Glare*, 2nd ed.; Technical Report, Internationale Beleuchtungskommission, Eds.; CIE Central Bureau: Vienna, Austria, 1990; ISBN 978-92-9034-030-0.
10. Dumont, E.; Paumier, J.L.; Ledoux, V. Are Standard R-Tables Still Representative of Road Surface Photometric Characteristics in France? In Proceedings of the CIE International Symposium on Road Surface Photometric Characteristics, Année, France, 9 July 2008; p. 8.
11. Ylinen, A.-M.; Pellinen, T.; Valtonen, J.; Puolakka, M.; Halonen, L. Investigation of Pavement Light Reflection Characteristics. *Road Mater. Pavement Des.* **2011**, *12*, 587–614. [[CrossRef](#)]
12. PKN-CEN/TR 13201-1:2016; Road Lighting—PART 1: Guidelines on Selection of Lighting Classes. SAI Global Standards: Chicago, IL, USA, 2016.
13. Sørensen, K. Performance of Road Markings and Road Surfaces. 2011. Available online: <https://nmfv.dk/wp-content/uploads/2012/03/Performance-of-road-markings-and-roadsurfaces.pdf> (accessed on 11 February 2023).
14. CIE. *Design Methods for Lighting of Roads*; Internationale Beleuchtungskommission, Ed.; CIE Central Bureau: Vienna, Austria, 1999; ISBN 978-3-900734-92-3.
15. Sørensen, K.; Nielsen, B. *Road Surfaces in Traffic Lighting*; The National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 1974.
16. Rebelo, F.J.P.; Martins, F.F.; Silva, H.M.R.D.; Oliveira, J.R.M. Use of Data Mining Techniques to Explain the Primary Factors Influencing Water Sensitivity of Asphalt Mixtures. *Constr. Build. Mater.* **2022**, *342*, 128039. [[CrossRef](#)]
17. Gong, H.; Sun, Y.; Mei, Z.; Huang, B. Improving Accuracy of Rutting Prediction for Mechanistic-Empirical Pavement Design Guide with Deep Neural Networks. *Constr. Build. Mater.* **2018**, *190*, 710–718. [[CrossRef](#)]
18. Guo, X.; Hao, P. Using a Random Forest Model to Predict the Location of Potential Damage on Asphalt Pavement. *Appl. Sci.* **2021**, *11*, 10396. [[CrossRef](#)]
19. Fakhri, M.; Dezfoulian, R.S. Pavement Structural Evaluation Based on Roughness and Surface Distress Survey Using Neural Network Model. *Constr. Build. Mater.* **2019**, *204*, 768–780. [[CrossRef](#)]
20. Gopalakrishnan, K.; Agrawal, A.; Ceylan, H.; Kim, S.; Choudhary, A. Knowledge Discovery and Data Mining in Pavement Inverse Analysis. *Transport* **2013**, *28*, 1–10. [[CrossRef](#)]
21. Nazemi, M.; Heidaripناه, A. Support Vector Machine to Predict the Indirect Tensile Strength of Foamed Bitumen-Stabilised Base Course Materials. *Road Mater. Pavement Des.* **2016**, *17*, 768–778. [[CrossRef](#)]
22. Bashar, M.Z.; Torres-Machi, C. Performance of Machine Learning Algorithms in Predicting the Pavement International Roughness Index. *Transp. Res. Rec.* **2021**, *2675*, 226–237. [[CrossRef](#)]
23. Tong, Z.; Gao, J.; Sha, A.; Hu, L.; Li, S. Convolutional Neural Network for Asphalt Pavement Surface Texture Analysis: Convolutional Neural Network. *Comput. Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1056–1072. [[CrossRef](#)]
24. Corte-Valiente, A.; Castillo-Sequera, J.; Castillo-Martinez, A.; Gómez-Pulido, J.; Gutierrez-Martinez, J.-M.; Corte-Valiente, A. An Artificial Neural Network for Analyzing Overall Uniformity in Outdoor Lighting Systems. *Energies* **2017**, *10*, 175. [[CrossRef](#)]
25. Kazanasmaz, T.; Günaydin, M.; Binol, S. Artificial Neural Networks to Predict Daylight Illuminance in Office Buildings. *Build. Environ.* **2009**, *44*, 1751–1757. [[CrossRef](#)]
26. Kayakuş, M.; Üncü, İ.S. Basketbol Salonlarının Parıltısının Makina Öğrenme Yöntemleriyle Tahmini. *Düzce Üniv. Bilim Teknol. Derg.* **2020**, *8*, 2468–2479. [[CrossRef](#)]
27. Qin, L.; He, S.; Yang, D.; Leon, A.S. Proposal for a Calculation Model of Perceived Luminance in Road Tunnel Interior Environment: A Case Study of a Tunnel in China. *Photonics* **2022**, *9*, 870. [[CrossRef](#)]
28. GDDKiA. *WT-1 Kruszywa Do Mieszanek Mineralno-Asfaltowych i Powierzchniowych Utrwaień Na Drogach Krajowych*; GDDKiA Generalna Dyrekcja Dróg Krajowych i Autostrad: Warsaw, Poland, 2014.
29. *PN-EN 13043:2004/Ap1:2010*; Kruszywa Do Mieszanek Bitumicznych i Powierzchniowych Utrwaień Stosowanych Na Drogach, Lotniskach i Innych Powierzchniach Przeznaczonych Do Ruchu. Polski Komitet Normalizacyjny: Warsaw, Poland, 1 March 2010.
30. *EN 1097-2:2020*; Tests for Mechanical and Physical Properties of Aggregates—Part 2: Methods for the Determination of Resistance to Fragmentation. CEN: Brussels, Belgium, 15 April 2020.
31. *EN 1097-8:2020*; Tests for Mechanical and Physical Properties of Aggregates—Part 8: Determination of the Polished Stone Value. CEN: Brussels, Belgium, 15 April 2020.

32. EN 1097-6:2013; Tests for Mechanical and Physical Properties of Aggregates—Part 6: Determination of Particle Density and Water Absorption. CEN: Brussels, Belgium, 3 July 2013.
33. EN 933-9:2022; Tests for Geometrical Properties of Aggregates—Part 9: Assessment of Fines—Methylene Blue Test. BSI: London, UK, 31 March 2022.
34. EN 933-1:2012; Tests for Geometrical Properties of Aggregates—Determination of Particle Size Distribution. Sieving Method. CEN: Brussels, Belgium, 11 January 2012.
35. EN 933-3:2012; Tests for Geometrical Properties of Aggregates—Part 3: Determination of Particle Shape—Flakiness Index. BSI: London, UK, 18 January 2012.
36. Wasilewska, M.; Gardziejczyk, W.; Gierasimiuk, P. Effect of Aggregate Graining Compositions on Skid Resistance of Exposed Aggregate Concrete Pavement. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *356*, 012001. [CrossRef]
37. Boyce, P.R. *Human Factors in Lighting*; CRC Press: Boca Raton, FL, USA, 2003; ISBN 978-0-429-22113-2.
38. Van Bommel, W.J.M.; de Boer, J.B. *Road Lighting*; Palgrave Macmillan: London, UK, 1980; ISBN 978-1-349-05802-0.
39. Huerne ter, H.L.; Hetebrij, D.; Elfring, J. Design of Reflective Pavements for Roads. In Proceedings of the 6th Eurasphalt & Eurobitume Congress, Prague, Czech Republic, 1–3 June 2016.
40. Üncü, I.S.; Kayakus, M. Analysis of Visibility Level in Road Lighting Using Image Processing Techniques. *Sci. Res. Essays* **2010**, *5*, 2279–2785.
41. General Directorate for National Roads and Motorways. *WT-2 Technical Guidelines 2: Asphalt Pavements for National Roads*; Part I: Asphalt Mixes; General Directorate for National Roads and Motorways: Warsaw, Poland, 2014.
42. PN-EN 1436:2018-02; Road Marking Materials—Road Marking Performance for Road Users and Test Methods. Polski Komitet Normalizacyjny: Warsaw, Poland, 9 February 2018.
43. Cano-Ortiz, S.; Pascual-Muñoz, P.; Castro-Fresno, D. Machine Learning Algorithms for Monitoring Pavement Performance. *Autom. Constr.* **2022**, *139*, 104309. [CrossRef]
44. Watroba, J. Statistic I Data Mining in scientific research. In *Materials from the Seminar Organized by StatSoft*; StatSoft Polska, Ed.; StatSoft Polska: Kraków, Poland, 2006; ISBN 978-83-88724-31-2. (In Polish)
45. Harty, J. *Advanced Machine Learning with Python: Solve Challenging Data Science Problems by Mastering Cutting-Edge Machine Learning Techniques in Python*; Packt Open Source Community Experience Distilled; Packt Publishing: Birmingham, UK; Mumbai, India, 2016; ISBN 978-1-78439-863-7.
46. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
47. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
48. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
49. Zhang, H.; Yang, Q.; Shao, J.; Wang, G. Dynamic Streamflow Simulation via Online Gradient-Boosted Regression Tree. *J. Hydrol. Eng.* **2019**, *24*, 04019041. [CrossRef]
50. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
51. Persson, C.; Bacher, P.; Shiga, T.; Madsen, H. Multi-Site Solar Power Forecasting Using Gradient Boosted Regression Trees. *Sol. Energy* **2017**, *150*, 423–436. [CrossRef]
52. Mahmud, K.; Azam, S.; Karim, A.; Zobaed, S.; Shanmugam, B.; Mathur, D. Machine Learning Based PV Power Generation Forecasting in Alice Springs. *IEEE Access* **2021**, *9*, 46117–46128. [CrossRef]
53. Breiman, L. (Ed.) *Classification and Regression Trees*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1998; ISBN 978-0-412-04841-8.
54. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
55. Hill, T.; Lewicki, P. *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry and Data Mining*; StatSoft: Tulsa, OK, USA, 2006; ISBN 978-1-884233-59-3.
56. Bray, T.; Paoli, J.; Sperberg-McQueen, C.M.; Maler, E. Extensible Markup Language (XML) 1.0, 5th ed. 2008. Available online: <http://www.w3.org/TR/2008/PER-xml-20080205> (accessed on 11 February 2023).
57. Wasilewska, M.; Grzyb, D.; Gardziejczyk, W. Ocena Właściwości Fizycznych Kruszyw Grubych Do Warstw Ścieralnych Nawierzchni Drogowych. *Constr. Mater.* **2022**, *1*, 78–80. [CrossRef]

**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.