






## Article

# Prediction of Rockfill Materials' Shear Strength Using Various Kernel Function-Based Regression Models—A Comparative Perspective

Mahmood Ahmad <sup>1,2</sup>, Ramez A. Al-Mansob <sup>1,\*</sup>, Irfan Jamil <sup>3</sup>, Mohammad A. Al-Zubi <sup>4</sup>,  
Mohamad Muayad Sabri Sabri <sup>5</sup> and Arnold C. Alguno <sup>6</sup>

- <sup>1</sup> Department of Civil Engineering, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak 50728, Selangor, Malaysia; ahmadm@iium.edu.my
- <sup>2</sup> Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu 28100, Pakistan
- <sup>3</sup> Department of Civil Engineering, University of Engineering and Technology Peshawar, Peshawar 25000, Pakistan; irfanuop@hotmail.com
- <sup>4</sup> Department of Mechanical Engineering, Hijjawai Faculty for Engineering, Yarmouk University, Irbid 21163, Jordan; mohammad.alzubi@yu.edu.jo
- <sup>5</sup> Peter the Great St. Petersburg Polytechnic University, 195251 Saint Petersburg, Russia; mohamad.m.sabri@gmail.com
- <sup>6</sup> Department of Physics, Mindanao State University-Iligan Institute of Technology, Iligan City 9200, Philippines; arnold.alguno@g.msuiit.edu.ph
- \* Correspondence: ramez@iium.edu.my



**Citation:** Ahmad, M.; Al-Mansob, R.A.; Jamil, I.; Al-Zubi, M.A.; Sabri, M.M.S.; Alguno, A.C. Prediction of Rockfill Materials' Shear Strength Using Various Kernel Function-Based Regression Models—A Comparative Perspective. *Materials* **2022**, *15*, 1739. <https://doi.org/10.3390/ma15051739>

Academic Editor: Prabir K. Sarker

Received: 26 January 2022

Accepted: 23 February 2022

Published: 25 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

**Abstract:** The mechanical behavior of the rockfill materials (RFMs) used in a dam's shell must be evaluated for the safe and cost-effective design of embankment dams. However, the characterization of RFMs with specific reference to shear strength is challenging and costly, as the materials may contain particles larger than 500 mm in diameter. This study explores the potential of various kernel function-based Gaussian process regression (GPR) models to predict the shear strength of RFMs. A total of 165 datasets compiled from the literature were selected to train and test the proposed models. Comparing the developed models based on the GPR method shows that the superlative model was the Pearson universal kernel (PUK) model with an R-squared ( $R^2$ ) of 0.9806, a correlation coefficient ( $r$ ) of 0.9903, a mean absolute error (MAE) of 0.0646 MPa, a root mean square error (RMSE) of 0.0965 MPa, a relative absolute error (RAE) of 13.0776%, and a root relative squared error (RRSE) of 14.6311% in the training phase, while it performed equally well in the testing phase, with  $R^2 = 0.9455$ ,  $r = 0.9724$ , MAE = 0.1048 MPa, RMSE = 0.1443 MPa, RAE = 21.8554%, and RRSE = 23.6865%. The prediction results of the GPR-PUK model are found to be more accurate and are in good agreement with the actual shear strength of RFMs, thus verifying the feasibility and effectiveness of the model.

**Keywords:** shear strength; rockfill materials; Gaussian functions; polynomial kernel; radial basis function; Pearson universal kernel

## 1. Introduction

In civil engineering projects, such as rockfill dams, slopes, and embankments, rockfill materials (RFMs) are often used as filling materials. RFMs consist of coarse gravels, cobbles, and boulders mined from rock quarries or riverbeds. Quarried materials are angular to sub-angular, whereas riverbed materials are rounded to sub-rounded. Mineral composition, particle size, shape, gradation, individual particle strength, void content, relative density, and surface roughness of the particles all influence the behavior of the RFMs utilized in the construction of rockfill dams. Several studies in geotechnical engineering have been carried out, such as that examining the contact between the soils and concrete used in earth and rockfill dams [1]. Inverse analysis provides an means to better understand dam

behavior [2], and offers measures of surface roughness, apparent porosity, apparent density, water absorption, and uniaxial compression strength (UCS), as well as an understanding of the influence of the heating rate and cooling process of gneiss stone [3].

RFMs are the preferred materials for the construction of high-embankment dams in seismically vulnerable areas because they provide structural flexibility and considerably reduce the issues caused by porewater pressure during and after construction. The mechanical behavior of the RFMs utilized in the dam's shell must be evaluated. However, because the materials may contain pieces larger than 500 mm in diameter, the characterization of RFMs with special reference to shear strength and deformation is challenging and expensive.

A number of studies have been carried on RFM behavior. Marsal [4], Mirachi et al. [5], Venkatachalam [6], Gupta [7], Abbas [8], and Honkanadavar and Sharma [9] conducted laboratory experiments on different RFMs and found that the behavior of stress–strain is nonlinear, inelastic, and dependent on the stress level. They also found that as the maximum particle size for riverbed RFM increases, so does the angle of internal friction, but quarry RFM shows the reverse trend. Frossard et al. [10] developed a logical method for evaluating RFMs' shear strength based on size effects. To associate the shear strength parameter to some riverbed RFM index features, Honkanadavar and Gupta [11] developed a power law. Due to the difficulty of performing large-scale strength tests and characterizing the mechanical behavior of RFMs, different approaches for predicting their behavior have been developed. In laboratory experiments, RFM with a large particle size (maximum particle size of 1200 mm) was determined to be incompatible [11]. Large-scale shear tests are time-consuming and difficult, and estimating the nonlinear shear strength function without an analytical method is complicated.

In the recent past, machine learning (ML) algorithms have achieved notable successes at efficiently solving real-world problems in different sectors, including civil and environmental engineering [12], geotechnical engineering [13–18], and other fields of science [19–26]. The artificial neural network (ANN) approach was found to be more efficient in predicting the shear strength of RFMs [27]. Zhou et al. [28] has recently shown that cubist and random forest regression algorithms are better at predicting RFM shear strength results than ANN and traditional regression models. To predict RFMs' shear strength, Ahmad et al. [29] used support vector machine, random forest, adaptive boosting, and k-nearest neighbor algorithms. This field, on the other hand, is still being researched and further explored.

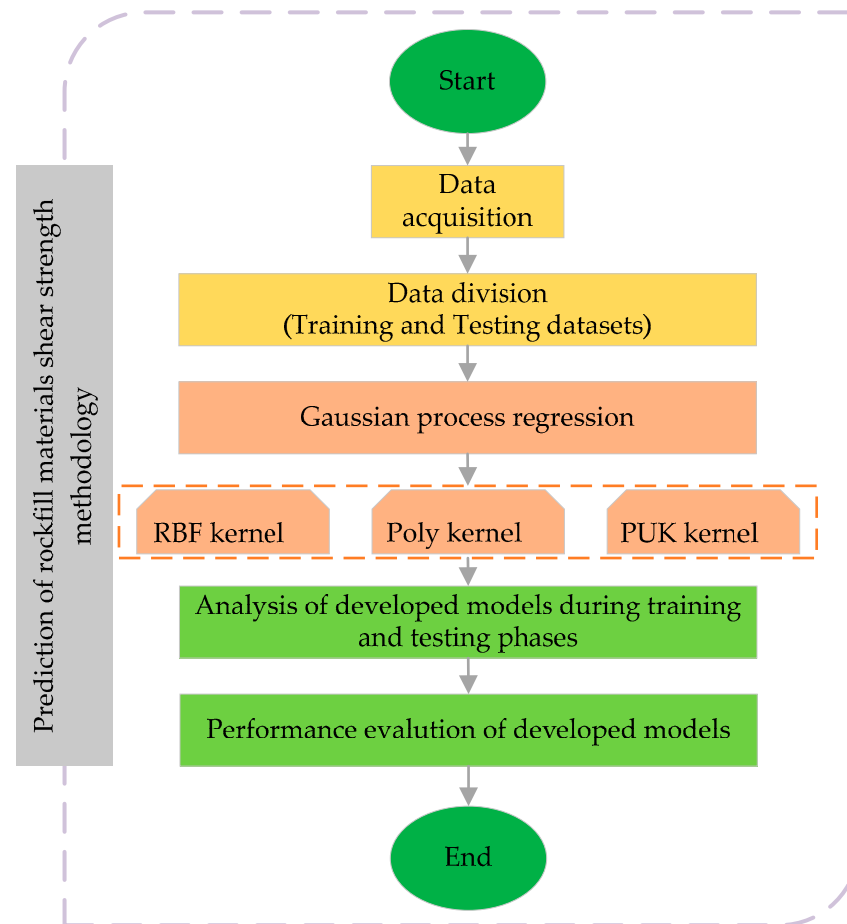
No prior study was found to simultaneously evaluate the effectiveness of a range of kernel function-based Gaussian process regression (GPR) models. The goal of this research is to evaluate and compare the efficacy of various kernel function-based GPR models based on the results of reliable experimental tests of shear strength prediction modeling for rock materials. The aims of this research include the following:

- To examine the capability of various kernel function-based GPR computing techniques, namely, radial basis function kernel, polynomial kernel, and Pearson universal kernel, in the area of predicting the shear strength of RFMs;
- To undertake a comparative study of the shear strength prediction of rockfill materials and select the best outcomes provided by the developed GPR models based on the performance metrics;
- To conduct sensitivity analyses to determine the effect of each input parameter on the RFMs' shear strength.

The remainder of the paper is organized as follows. In Section 2, the details of three different types of kernel function-based GPR computing techniques for predicting the shear strength of rockfill materials are presented. In Section 3, the details of the data catalog and correlation analysis are presented. Performance evaluation measures are presented in Section 4. Section 5 reports the developed models' results. Based on the observations and results of these models, Section 5 draws conclusions and future research directions.

## 2. Gaussian Process Regression

Gaussian process regression (GPR) is a suitable and recently described method that has been used in a variety of machine learning applications [30]. The GPR model's probabilistic solution leads to the identification of general kernel regression problems. The applied regressor's training process can be categorized as Bayesian, and the model relations are assumed to follow a Gaussian distribution to encode the previous output function information [31]. The Gaussian process is defined by a set of variables, each of which has a joint Gaussian allocation [32]. Figure 1 illustrates the scheme of development of the selected methodology.



**Figure 1.** Flowchart of the proposed methodology.

### 2.1. Radial Basis Function Kernel

The RBF kernel is a typical kernel function used in several machine learning algorithms for kernel learning. It is widely used in vector machinery classification. RBF kernels are typically a decent first option. Instead of the usual kernel, this kernel transfers samples to higher-dimensional areas, allowing a nonlinear relationship between class labels and attributes to be dealt with [33]. Furthermore, with the linear kernel and normalized poly kernel, RBF is a unique situation because the linear kernel with a penalty parameter  $C$  operates with the same parameters as the RBF kernel.

In comparison, kernel values can be used to calculate  $\gamma X_i^T + r > 1$  for polynomial kernels up to infinity as long as the grade is high [34]. Furthermore, under such parameters, the sigmoid kernel is not true, i.e., it is not an internal two-vector product. The RBF kernel is not appropriate in several cases. The linear kernel can be employed in particular when the number of functions is quite large [35].

## 2.2. Polynomial Kernel

The polynomial kernel (poly kernel) is a kernel function often used with support-vector machines (SVMs) and other kernel-coded models that relate to parallel vectors in the machine learning language. The polynomial kernel appears to achieve similarity not only on the input samples' stated functions, but also in their combinations. In the context of regression analysis, such groupings were categorized as interaction features. These groupings have been identified as interaction features in the context of regression analysis. The enclosed polynomial kernel space is the same as a polynomial regression, but it is an educated sum of parameters that does not include a combined blow-up. The functionality links to logical function connections if the function data input are binary [36]. The polynomial kernel for degree polynomials is well-defined as

$$K(x, Y) = (x^T y + C)^d \quad (1)$$

where  $x$  and  $y$  are vectors in the input space, i.e., feature vectors found from training or trial samples, and  $C \geq 0$  is an unconstrained parameter that trades off higher-order vs. lower-order polynomial definitions. When  $C = 0$ , the kernel is said to be homogeneous.

## 2.3. Pearson Universal Kernel

The Pearson universal kernel (PUK) is a machine learning programming tool that aids in the comprehensive interpretation and understanding of various data types. The Pearson VII function is considered to have a general form for curve suitably, and is assumed by

$$f(x) = H / \left[ 1 + 2(x - x_0) \sqrt{2 \left( \frac{1}{\omega} - 1 \right) / \delta} \right]^{2\omega} \quad (2)$$

where  $H$  is the top tallness at the middle  $x_0$  of the peak, and  $x$  represents the self-determining variable. The parameters  $\sigma$  and  $\omega$  regulate the half-width and the following factor of the peak. A function  $\omega$ , on the other hand, belongs to the class of effective kernel functions. The kernel matrix might be symbolic and positively semi-definite; to show that PUK is certainly resolving these situations, Uestuen [37] redrafted Equation (3) into a function of both vectors:

$$K(x_i, x_j) = \frac{1}{\left[ 1 + \left( 2 \sqrt{|x_i - x_j|^2} \sqrt{2 \left( \frac{1}{\omega} - 1 \right) / \sigma} \right)^2 \right]^{2\omega}} \quad (3)$$

## 3. Data Catalog and Correlation Analysis

The dataset collected from Kaunda [27] was separated into training (80 percent of total data) and testing (20 percent of the remaining data) datasets for this investigation. The database has been presented in Kaunda [27] in detail. Table 1 summarizes the 165-sample dataset, which includes numerous shear strength of RFM tests, and the input and output variables' minimum (min), maximum (max), mean, and standard deviation (SD). As can be seen in the table, the database includes input parameters, i.e., particle material size (or sieve) gradation, fineness modulus, gradation modulus, material hardness, relative density, and confining (normal) stress, and one output parameter, i.e., shear strength.

To choose the most resilient representation, a statistical study of input and output variables of the training and testing data was performed (see Table 2). It was accomplished through the use of a trial-and-error strategy. Previous studies show that the shear strength ( $\tau$ ) of RFM is a function of  $D_{10}$ ,  $D_{30}$ ,  $D_{60}$ , and  $D_{90}$ , which correspond to the 10%, 30%, 60%, and 90% passing sieve sizes, while  $UCS_{min}$  and  $UCS_{max}$  (MPa) indicate the minimum and maximum uniaxial compressive strengths (MPa), the  $FM$  and  $GM$  parameters describe fineness modulus and gradation modulus, respectively,  $\gamma$  is the dry unit weight ( $kN/m^3$ ),

$\sigma_n$  is the normal stress (MPa), and  $R$  shows the International Society of Rock Mechanics (ISRM) hardness rating [27,29]. As a result, the current study’s GPR models are constructed using these input variables.

**Table 1.** The inputs and output of the present study.

S. No.	$D_{10}$ (mm)	$D_{30}$ (mm)	$D_{60}$ (mm)	$D_{90}$ (mm)	$C_C$	$C_U$	$GM$	$FM$	$R$	$UCS_{min}$ (MPa)	$UCS_{max}$ (MPa)	$\gamma$ (KN/m <sup>3</sup> )	$\sigma_n$ (MPa)	$\tau$ (MPa)
1	0.4	2.9	9.7	31	2.17	24.25	3.41	5.57	4	50	100	20.8	1.142	0.93
2	0.44	1.5	6.99	27.5	0.73	15.89	3.82	5.16	4	50	100	18.7	0.159	0.189
3	0.4	3.3	10.3	33.3	2.64	25.75	3.32	5.64	4	50	100	21.8	0.344	0.357
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
163	0.02	0.94	4	18	11.05	200	4.78	4.19	1	1	5	15.4	0.044	0.025
164	20.7	26.7	32.8	53	1.05	1.58	0.89	8.2	5	100	250	15.3	0.115	0.191
165	0.4	2.3	12.2	44.4	1.08	30.5	3.3	5.69	4	50	100	18.7	0.794	0.577
Min	0.01	0.56	1.2	2.6	0.1	1.36	0.2	3	1	1	5	9.32	0.002	0.005
Max	33.9	42.4	80.1	100	22.27	1040	6	8.8	6	250	400	38.9	4.205	3.921
Mean	4.463	7.86	18.28	39.927	2.404	69.561	2.903	6.142	4.327	73.691	168.455	20.799	0.734	0.662
SD	8.875	10.335	14.42	22.432	3.414	193.628	1.278	1.298	0.957	37.975	87.844	4.861	0.785	0.652

**Table 2.** Statistics of parameters of the training and testing datasets.

Statistical Parameter	Dataset	Input Variable											Output Variable		
		$D_{10}$ (mm)	$D_{30}$ (mm)	$D_{60}$ (mm)	$D_{90}$ (mm)	$C_C$	$C_U$	$GM$	$FM$	$R$	$UCS_{min}$ (MPa)	$UCS_{max}$ (MPa)	$\gamma$ (KN/m <sup>3</sup> )	$\sigma_n$ (MPa)	$\tau$ (MPa)
Minimum	Training	0.010	0.560	1.200	2.600	0.100	1.360	0.200	3.000	1.000	1.000	5.000	9.320	0.002	0.005
	Testing	0.010	0.560	1.200	2.600	0.100	1.470	0.200	3.000	1.000	1.000	5.000	9.320	0.021	0.024
Maximum	Training	33.900	42.400	80.100	100.000	22.270	1040.000	6.000	8.800	6.000	250.000	400.000	38.900	4.205	3.921
	Testing	33.900	42.400	50.000	99.000	22.270	1040.000	6.000	8.800	5.000	100.000	250.000	38.900	3.223	2.492
Mean	Training	4.867	8.465	19.287	40.386	2.199	53.324	2.788	6.250	4.364	75.045	170.682	20.766	0.729	0.660
	Testing	2.887	5.442	14.252	38.091	3.226	134.510	3.365	5.709	4.182	68.273	159.545	20.932	0.756	0.668
Standard deviation	Training	9.179	10.577	15.135	22.018	3.075	156.064	1.243	1.261	0.910	39.230	88.010	4.605	0.780	0.662
	Testing	7.453	9.050	10.349	24.289	4.492	194.958	1.331	1.374	1.131	32.444	87.967	5.854	0.816	0.619

Understanding the relationship between each input and result can definitely facilitate the development of a proper prediction model. Among the numerous correlation coefficients described thus far, the correlation coefficient technique has proven to be more common. As stated in Equation (4), the correlation coefficient is equal to the product of the covariance of two parameters divided by their standard deviation.  $\rho_{m,n} \approx 1$  represents the high degree of interdependence between two variables, while  $\rho_{m,n} \approx 0$  stands for a linear relationship between two variables  $m$  and  $n$  that are independent of one another. The Pearson correlation coefficients for the variable inputs and the target output are reported in Table 3.

$$\rho(m, n) = \frac{\text{cov}(m, n)}{\sigma_m \sigma_n} \tag{4}$$

Obviously, the values provided in Table 3 reveal that  $\sigma_n$  and  $D_{90}$  have the most significant influence on the  $\tau$ , while  $FM$ ,  $C_u$ , and  $UCS_{max}$  affect the output non-considerably.

**Table 3.** Pearson correlation coefficients for variable inputs and the target output.

	$D_{10}$ (mm)	$D_{30}$ (mm)	$D_{60}$ (mm)	$D_{90}$ (mm)	$C_c$	$C_u$	$GM$	$FM$	$R$	$UCS_{min}$ (MPa)	$UCS_{max}$ (MPa)	$\gamma$ (KN/m <sup>3</sup> )	$\sigma_n$ (MPa)	$\tau$ (MPa)
$D_{10}$ (mm)	1													
$D_{30}$ (mm)	0.972	1												
$D_{60}$ (mm)	0.652	0.802	1											
$D_{90}$ (mm)	0.304	0.451	0.758	1										
$C_c$	-0.203	-0.229	-0.251	-0.214	1									
$C_u$	-0.171	-0.207	-0.171	-0.032	0.567	1								
$GM$	-0.784	-0.866	-0.849	-0.686	0.345	0.267	1							
$FM$	0.770	0.846	0.824	0.644	-0.357	-0.273	-0.959	1						
$R$	0.297	0.357	0.457	0.304	-0.607	-0.176	-0.465	0.459	1					
$UCS_{min}$ (MPa)	0.291	0.411	0.644	0.352	-0.358	-0.169	-0.444	0.428	0.838	1				
$UCS_{max}$ (MPa)	0.376	0.443	0.536	0.254	-0.363	-0.193	-0.454	0.433	0.864	0.945	1			
$\gamma$ (KN/m <sup>3</sup> )	-0.277	-0.237	-0.079	0.080	0.459	0.189	0.203	-0.305	-0.448	-0.227	-0.292	1		
$\sigma_n$ (MPa)	-0.249	-0.154	0.146	0.413	-0.164	-0.069	-0.086	0.058	0.098	0.089	-0.034	0.195	1	
$\tau$ (MPa)	-0.238	-0.130	0.210	0.472	-0.187	-0.080	-0.118	0.092	0.150	0.156	0.030	0.185	0.966	1

#### 4. Performance Evaluation Measures

The coefficient of determination ( $R^2$ ), Pearson's correlation coefficient ( $r$ ), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) are used to evaluate the data-driven modeling in this study. These parameters can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i_p} - y_{i_o})^2}{\sum_{i=1}^n (y_{i_o} - \bar{y}_o)^2} \quad (5)$$

$$r = \frac{\sum_{i=1}^n [(y_{i_o} - \bar{y}_p)(y_{i_o} - \bar{y}_p)]}{\sqrt{\sum_{i=1}^n (y_{i_o} - \bar{y}_p)^2} \sqrt{\sum_{i=1}^n (y_{i_o} - \bar{y}_p)^2}} \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |y_{i_o} - y_{i_p}| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_{i_o} - y_{i_p})^2} \quad (8)$$

$$RAE = \frac{\sum_{i=1}^n |y_{i_p} - y_{i_o}|}{\sum_{i=1}^n |y_{i_o} - \bar{y}_o|} \quad (9)$$

$$RRSE = \frac{\sqrt{\sum_{i=1}^n (y_{i_p} - y_{i_o})^2}}{\sqrt{\sum_{i=1}^n (y_{i_o} - \bar{y}_o)^2}} \quad (10)$$

where  $y_{i_o}$  and  $y_{i_p}$  represents the actual measurement and predicted shear strength of RFM, respectively,  $\bar{y}_o$  is the average of the reference samples' values, and  $n$  is the defined amount of data.

The degree of collinearity between predicted and measured data is described by the coefficient of determination ( $R^2$ ) and correlation coefficient ( $r$ ). The correlation coefficient, which varies from 1 to  $-1$ , is a measure of the degree to which observed and predicted data are linearly related. There is no linear relationship if  $r = 0$ . A perfect positive or negative linear relationship arises if  $r = 1$  or  $-1$ . Similarly,  $R^2$  denotes the percentage of variance in the measured data that the model can explain.  $R^2$  spans from 0 to 1, with higher values suggesting less error variation, and values above 0.5 are usually regarded as acceptable [38,39]. The MAE represents the average value of the predicted and actual values. When the MAE is close to 0, the adjustment has a better effect, implying that the prediction model more accurately describes the set of training data [40]. As a single measure of predictive power, the RMSE is the average magnitudes of the errors in predictions for

all observations. The RMSE is greater than or equal to 0, with 0 indicating a statistically perfect fit for the observed data. The relative absolute error (RAE) is the difference between expected and actual values that is calculated by dividing the mean difference by the arithmetic mean. It ranges from 0 to infinite, with being 0 the best value. The RRSE criteria measure the model's percentage error, which ranges from 0 to 100. Consequently, the better the model, the lower the values of these criteria are. Furthermore, visual inspections, such as scatter plots, were used to compare the performance of the developed models.

## 5. Results and Discussion

### 5.1. Comparative Performance

In this paper, various kernel function-based Gaussian process regression techniques were implemented using Waikato environment for knowledge analysis (WEKA). WEKA is an open-source software which consists of a collection of machine learning algorithms for data mining tasks. Most machine learning algorithms have hyperparameters that must be tuned. The critical hyperparameters in the GPR-RBF, GPR-Poly, and GPR-PUK models are tuned in this study as shown in Table 4. The values for the models' tuning parameters were chosen first, and then varied in the trials until the best fitness measures in Table 4 were obtained.

**Table 4.** Different regression models' optimal tuning parameters.

Model	Parameters for Optimal Tuning
RBF kernel	{noise = 0.25, gamma = 0.02}
Poly kernel	{noise = 0.5}
PUK kernel	{noise = 0.3, omega = 0.85, sigma = 0.9}

Table 5 shows the  $R^2$ ,  $r$ , MAE, RMSE, RAE, and RRSE values for shear strength estimation after hyperparameter tuning for the training and testing phases, respectively. At a glance, for both tables, GPR-PUK is the top-ranked model. For the training results, based on the  $R^2$  (0.9224, 0.9430, and 0.9806),  $r$  (0.9604, 0.9711, and 0.9903), MAE (0.1565, 0.1015, and 0.0646), RMSE (0.2219, 0.1605, and 0.0965), RAE (31.7002%, 20.5572%, and 13.0776%), and RRSE (31.7002%, 20.5572%, and 13.0776%), respectively, for GPR-RBF, GPR-Poly, and GPR-PUK models, the GPR-PUK outputs are verified to be the most compatible with actual RFM shear strength values. Following that, GPR-Poly demonstrated a high level of accuracy. Furthermore, the findings obtained for the testing dataset demonstrate GPR-PUK's superior performance. Considering the respective values of  $R^2$  (0.9334, 0.9411, and 0.9455),  $r$  (0.9661, 0.9701, and 0.9724), MAE (0.1395, 0.1092, and 0.1048), RMSE (0.1781, 0.1508, and 0.1443), RAE (29.0963%, 22.783%, and 21.8554%), and RRSE (29.2377%, 24.7641%, and 23.6865%), respectively, for the GPR-RBF, GPR-Poly, and GPR-PUK models, GPR-Poly is the second most precise model, and GPR-PUK has outperformed both GPR-RBF and GPR-Poly. In comparison with the ANN model ( $R^2 = 0.9386$ ) and linear regression method ( $R^2 = 0.836$ ) reported by Kaunda [27] and Andjelkovic et al. [41], respectively, for the test data, the proposed GPR-PUK ( $R^2 = 0.9455$ ) has better prediction capacity. In general, the generalization and reliability of the GPR-PUK perform well, and larger datasets can yield better prediction results.

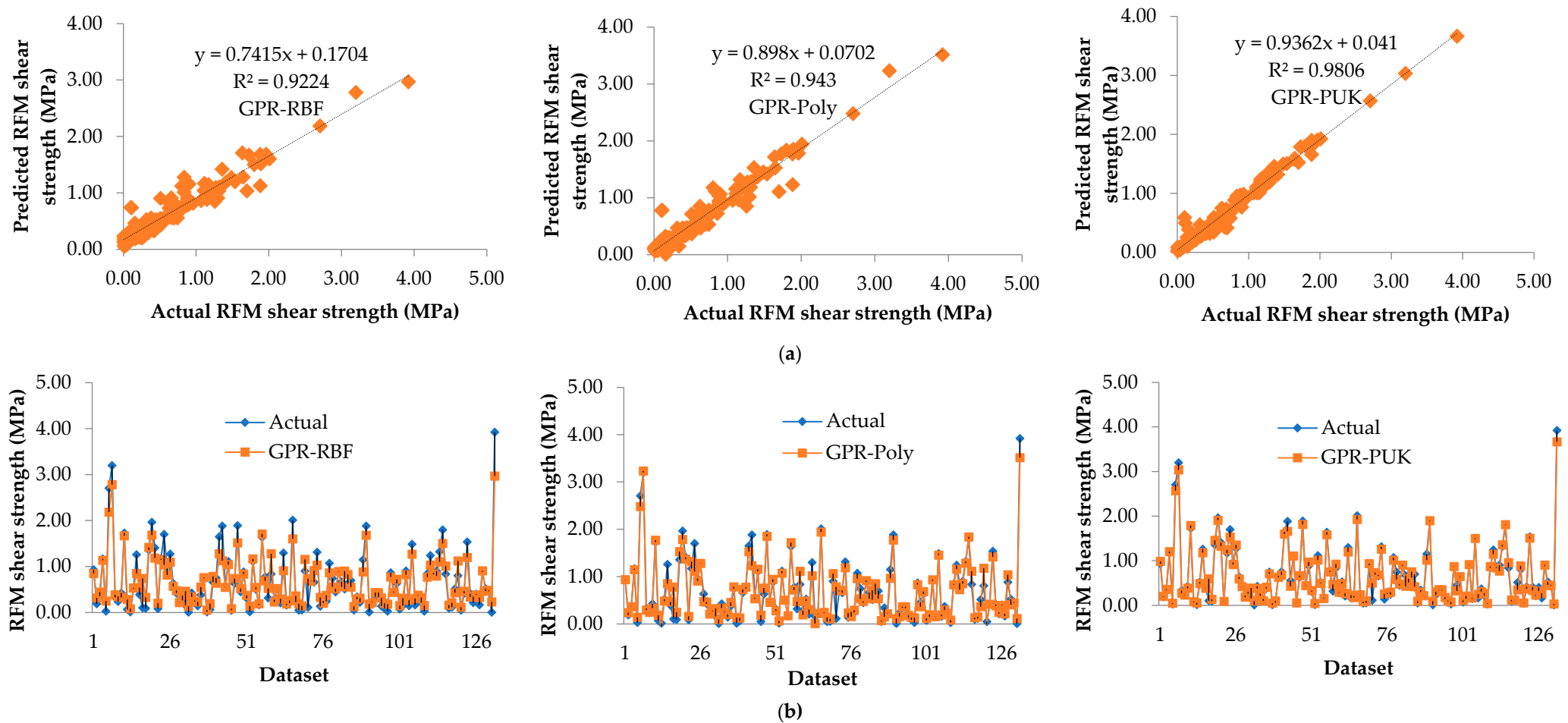
**Table 5.** Statistical indices and error measures between experimental/actual and predicted shear strength of three different models used in present study for RFMs.

Model	Dataset	R <sup>2</sup>	<i>r</i>	MAE (MPa)	RMSE (MPa)	RAE (%)	RRSE (%)
GPR-RBF	Training	0.9224	0.9604	0.1565	0.2219	31.7002	33.6238
	Testing	0.9334	0.9661	0.1395	0.1781	29.0963	29.2377
GPR-Poly	Training	0.9430	0.9711	0.1015	0.1605	20.5572	24.3232
	Testing	0.9411	0.9701	0.1092	0.1508	22.783	24.7641
GPR-PUK	Training	0.9806	0.9903	0.0646	0.0965	13.0776	14.6311
	Testing	0.9455	0.9724	0.1048	0.1443	21.8554	23.6865

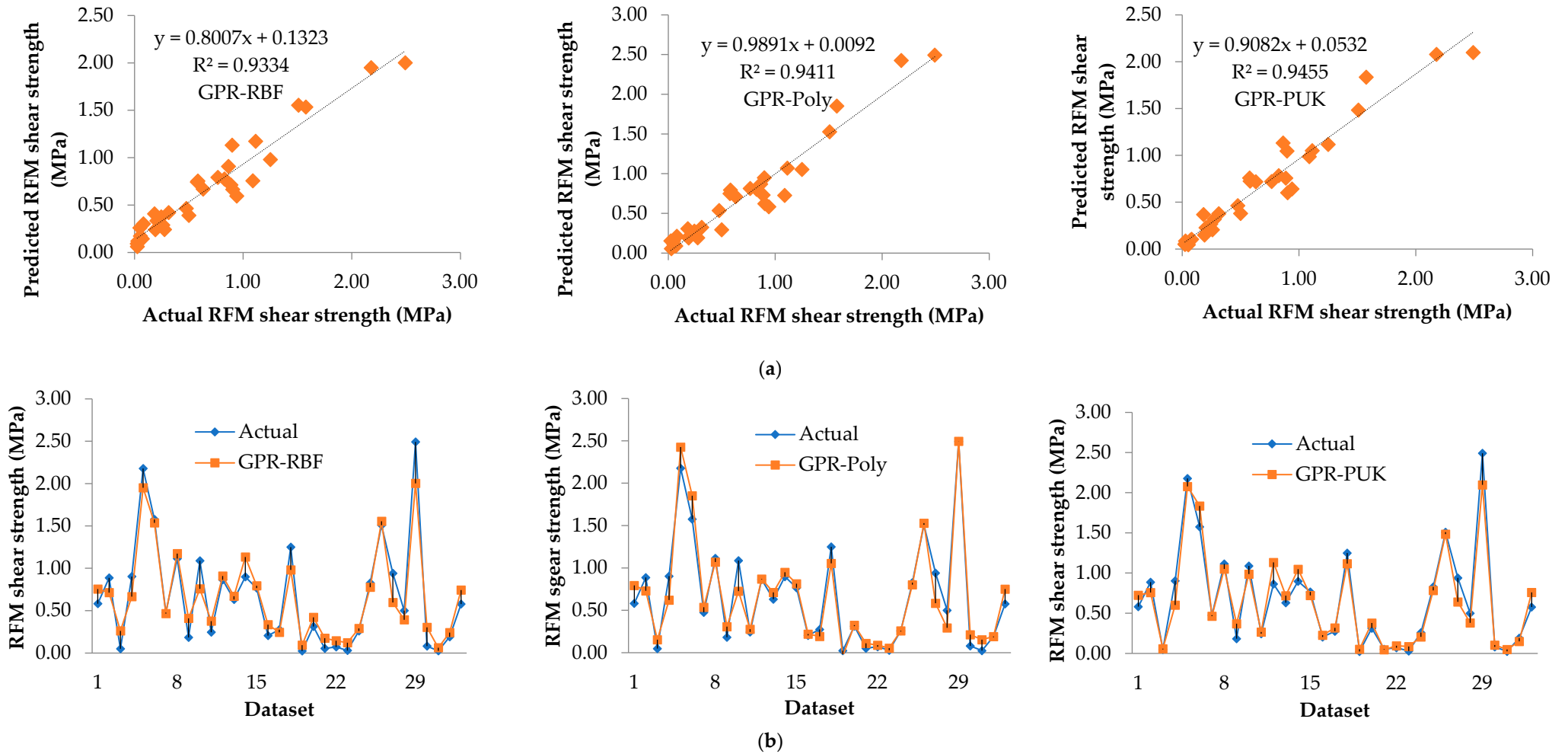
In addition, for the GPR-RBF, GPR-Poly, and GPR-PUK models, the graphical correlation between measured (on the horizontal axis) and predicted (on the vertical axis) shear strength is presented in Figure 2a for the training dataset and Figure 3a for the testing dataset, respectively. The trend line for GPR-PUK has been drawn by comparing the observed regression in Figures 2a and 3a, and the GPR-PUK findings have the maximum inclination to the line of  $y = x$  (i.e.,  $R^2 = 0.9806$ ).

The accuracy of all the developed models i.e., GPR-RBF, GPR-Poly, and GPR-PUK in predicting RFM shear strength is illustrated in Figure 2b for the training dataset and Figure 3b for the testing dataset, respectively. As seen in this graph, the closer one moves to the y axis (i.e., the lower the error), the higher the accuracy in both the training and testing datasets. Here, the GPR-PUK model has presented the most reliable prediction. This is evident by the higher aggregation of the results around the y axis ( $y = 0$ ), except for a few noise points. In comparison to the other models, i.e., GPR-RBF and GPR-Poly, the comparison findings are sufficiently consistent, which is adequate for the proposed GPR-PUK model to predict RFM shear strength values.





**Figure 2.** Comparison of the results in the training dataset of the various kernel function-based Gaussian process regression (GPR) models (a) measured vs. predicted RFM shear strength, (b) showing the accuracy of the models in predicting RFM shear strength.



**Figure 3.** Comparison of the results in the testing dataset of the various kernel function-based Gaussian process regression (GPR) models (a) measured vs. predicted RFM shear strength, (b) showing the accuracy of the models in predicting RFM shear strength.

### 5.2. Rank Analysis

Rank analysis assigns score values to statistical parameters, using their ideal values as a benchmark, based on the number of models utilized. The model with the best performance receives the highest score, and vice versa. Table 6 shows how the total efficiency ranking of the developed models is interpreted using a rank analysis (i.e., the summation of the ranks of  $R^2$ ,  $r$ , MAE, RMSE, RAE, and RRSE into a single ranking score for the training and testing datasets). Based on the obtained total ranking scores of 12, 24, and 36 (respectively for GPR-RBF, GPR-Poly, and GPR-PUK), the superiority of GPR-PUK can be concluded (i.e., it has the most significant total rank). The GPR-Poly model has commonly been labeled as the second most accurate model. The most significant point about Table 6 is that the GPR-RBF, GPR-Poly, and GPR-PUK predictive models achieved the same rank for each measure. The GPR-PUK and GPR-RBF, for instance, had the highest (3) and lowest (1) level of accuracy, based on the  $R^2$ ,  $r$ , MAE, RMSE, RAE, and RRSE indices simultaneously.

**Table 6.** The results of the employed models' rank analysis.

Model Parameter	GPR-RBF		GPR-Poly		GPR-PUK	
	Training	Testing	Training	Testing	Training	Testing
$R^2$	1	1	2	2	3	3
$r$	1	1	2	2	3	3
MAE	1	1	2	2	3	3
RMSE	1	1	2	2	3	3
RAE	1	1	2	2	3	3
RRSE	1	1	2	2	3	3
Rank Score	6	6	12	12	18	18
Total Ranking Score (Training and Testing)	12		24		36	
Total Rank	3		2		1	

### 5.3. Sensitivity Analysis

Yang and Zang's [42] sensitivity analysis was used to analyze the developed models' ability to analyze the impact of input variables on the shear strength of rockfill material. This method has been used in several research studies [43–46], and is as follows,

$$r_{ij} = \frac{\sum_{m=1}^n (y_{im} \times y_{om})}{\sqrt{\sum_{m=1}^n y_{im}^2 \sum_{m=1}^n y_{om}^2}} \quad (11)$$

where  $n$  is the number of data values, and  $y_{im}$  and  $y_{om}$  are the input and output parameters. For each input parameter, the  $r_{ij}$  value varied from zero to one, with the highest values indicating the most efficient output parameter (which was  $\tau$  in this study). To estimate the relationship between input and output variables, the value of  $r_{ij}$  must be close to 1. Figure 4 shows the degree of importance of the input variables based on the experimental actual and predicted values of the shear strength. As it can be seen, the importance of different parameters can be displayed as  $\sigma_n > D_{90} > \gamma > R > FM > UCS_{min} > D_{60} > GM > UCS_{max} > D_{30} > C_c > C_u > D_{10}$ . In other words, the  $\sigma_n$  is the most important parameter, and the  $D_{10}$  is the least important parameter for predicting the shear strength of the RFMs. Furthermore, Table 3 shows that the normal stress  $\sigma_n$  has the highest  $\rho$  of 0.966 in all other parameters, validating the sensitivity analysis results.

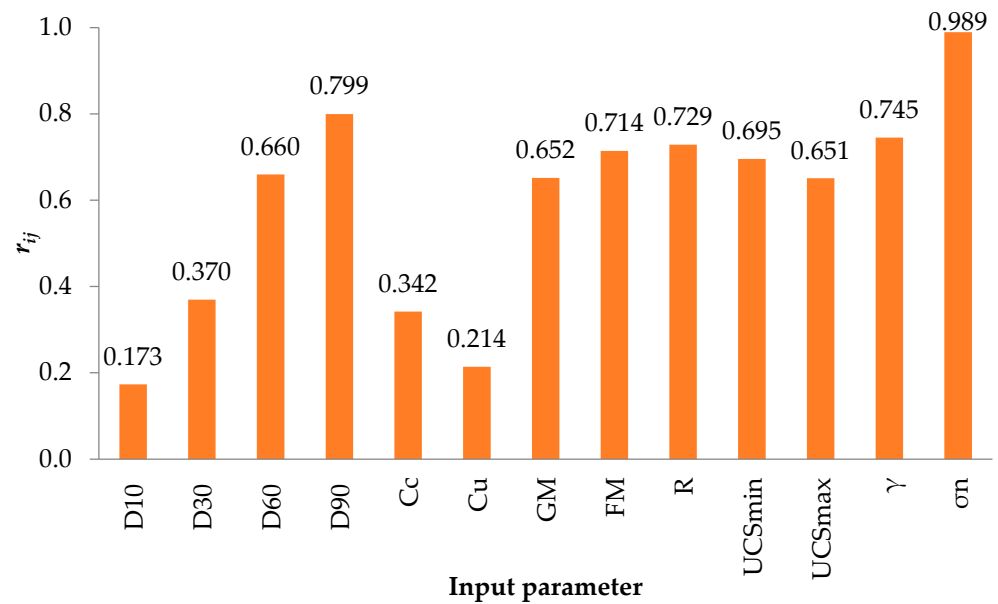


Figure 4. Sensitivity analysis of the input parameter.

5.4. Taylor Diagram

The Taylor diagram [47] is a simple visual depiction of a model’s performance compared to other models. Three indices are represented in the Taylor diagram: the correlation coefficient, the standard deviation, and the root mean square difference (RMSD). The model outcomes are compared in the Taylor diagram displayed in Figure 5 for a more in-depth examination of the results. The Taylor diagram, which compares the standard deviation (vertical and horizontal axes), correlation coefficient (radial lines), and RMSD, is a valuable tool for illustrating the accuracy of prediction models (green circular lines). The most accurate model, indicated by a pink dot (i.e., GPR-PUK), is introduced as having a similar standard deviation, higher correlation, and reduced RMSD when evaluating real values in the training and testing datasets. Figure 5 shows that the GPR-PUK model is closer to the red dot (actual/reference values) than the other GPR-RBF and GPR-poly models, indicating that this model is accurate.

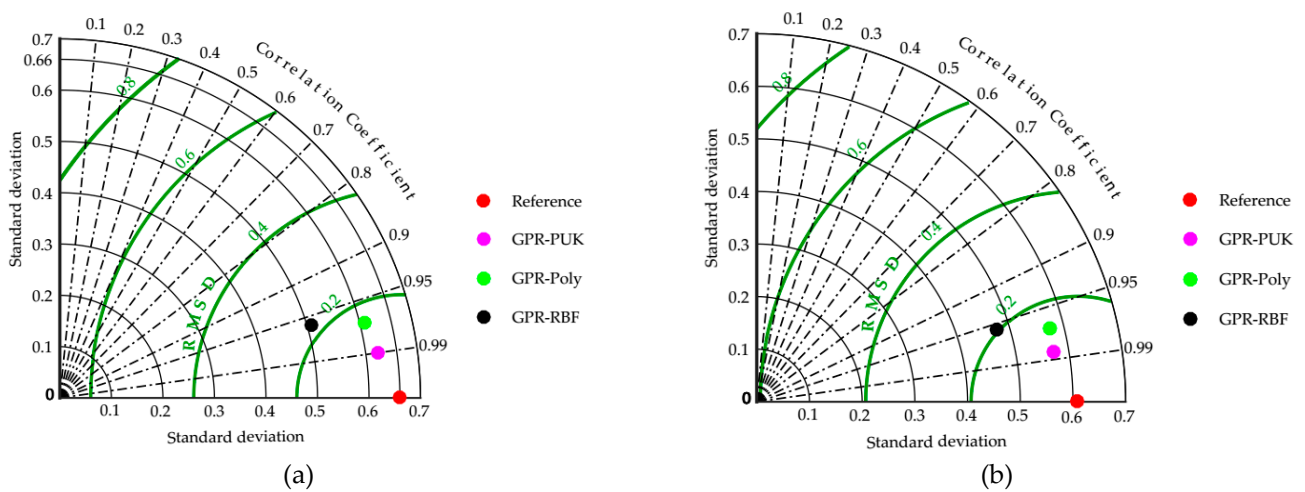


Figure 5. Taylor diagram for (a) training dataset (b) testing dataset, comparing all the models proposed in the present study.

## 6. Summery and Conclusions

In this research, efforts have been made to create various kernel function-based regression models, i.e., GPR-RBF, GPR-Poly, and GPR-PUK, that may be used to predict the shear strength of RFMs. To train and test the development models, a database from the published literature with different values of influential parameters on RFMs, including  $D_{10}$ ,  $D_{30}$ ,  $D_{60}$ ,  $D_{90}$ ,  $UCS_{min}$ ,  $UCS_{max}$ ,  $FM$ ,  $GM$ ,  $\gamma$ ,  $\sigma_n$ , and  $R$ , is considered. The data are split into two categories: training set (80%) and testing set (20%). The output shear strength ( $\tau$ ) of the developed models was evaluated using statistical parameters, including  $R^2$ ,  $r$ , MAE, RMSE, RAE, and RRSE. Furthermore, visual inspection, such as with scatter plots, was also used to assess the effectiveness of the developed models. The applications for the aforementioned models for predicting the shear strength of RFMs were compared and discussed. The following conclusions are made based on the obtained results of this study:

1. GPR-PUK achieved an R-squared ( $R^2$ ) of 0.9806, a correlation coefficient ( $r$ ) of 0.9903, a mean absolute error (MAE) of 0.0646 MPa, a root mean square error (RMSE) of 0.0965 MPa, a relative absolute error (RAE) of 13.0776%, and a root relative squared error (RRSE) of 14.6311% in the training phase. In the testing phase, it performed equally well, with  $R^2 = 0.9455$ ,  $r = 0.9724$ , MAE = 0.1048 MPa, RMSE = 0.1443 MPa, RAE = 21.8554%, and RRSE = 23.6865%. The GPR-PUK model was found to be more accurate and stable than the other models. Furthermore, the PUK kernel model had a superior agreement to the observed data based on the scatter plots of actual and predicted values, indicating that it has the potential for wider applications in RFM properties prediction.
2. The results of the sensitivity analysis show that the degree of importance of different input parameters for predicting the shear strength of RFMs is as follows:  $\sigma_n > D_{90} > \gamma > R > FM > UCS_{min} > D_{60} > GM > UCS_{max} > D_{30} > C_c > C_u > D_{10}$ .
3. The developed PUK kernel model makes predictions as accurate as those made by other soft computing techniques. This research also points out that these machine learning techniques can be a potential approach for estimating basic soil parameters, such as the soil permeability coefficient.

GPR-PUK can be used to predict the shear strength of RFMs with high accuracy, according to this study. The sample size is, however, limited. As a result, this study should be extended to include a larger sample size. Furthermore, future studies using other algorithms, such as XGBoost, evolutionary polynomial regression, and gene expression programming, should be utilized to evaluate the algorithms' effectiveness and gain a comprehensive understanding of the techniques used for predicting the shear strength of RFMs.

**Author Contributions:** Conceptualization, M.A. and R.A.A.-M.; methodology, M.A. and M.M.S.S.; software, M.A.; validation, M.A., R.A.A.-M., A.C.A. and M.A.A.-Z.; formal analysis, M.A.; investigation, M.A., A.C.A. and I.J.; resources, M.M.S.S.; data curation, I.J.; writing—original draft preparation, M.A.; writing—review and editing, M.A., M.A.A.-Z. and I.J.; visualization—review and editing, M.M.S.S.; supervision, R.A.A.-M.; project administration, M.M.S.S.; funding acquisition, M.M.S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program 'Priority 2030' (Agreement 075-15-2021-1333 dated 30 September 2021).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data 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. Gu, H.; Liu, K. Influence of Soil Heterogeneity on the Contact Problems in Geotechnical Engineering. *Appl. Sci.* **2021**, *11*, 4240. [[CrossRef](#)]
2. Das, R.; Soulamani, A. Non-deterministic methods and surrogates in the design of rockfill dams. *Appl. Sci.* **2021**, *11*, 3699. [[CrossRef](#)]
3. Costa, K.; Xavier, G.; Marvila, M.; Alexandre, J.; Azevedo, A.; Monteiro, S. Influence of high temperatures on physical properties and microstructure of gneiss. *Bull. Eng. Geol. Environ.* **2021**, *80*, 7069–7081. [[CrossRef](#)]
4. Marsal, R.J. Large scale testing of rockfill materials. *J. Soil Mech. Found. Div.* **1967**, *93*, 27–43. [[CrossRef](#)]
5. Marachi, N.D. *Strength and Deformation. Characteristics of Rockfill Materials*; Report No. TE-69-5 to State of California Department of Water Resources; University of California: Berkeley, CA, USA, 1969.
6. Venkatachalam, K. Prediction of Mechanical Behaviour of Rockfill Materials. Ph.D. Thesis, Indian Institute of Technology—Delhi, New Delhi, India, 1993.
7. Gupta, A.K. Constitutive Modelling of Rockfill Materials. Ph.D. Thesis, Indian Institute of Technology—Delhi, New Delhi, India, 2000.
8. Abbas, S.; Varadarajan, A.; Sharma, K. Prediction of Shear Strength Parameter of Prototype Rockfill Material. *IGC-2003 Roorkee* **2003**, *1*, 5–8.
9. Honkanadavar, N.; Sharma, K. Testing and modeling the behavior of riverbed and blasted quarried rockfill materials. *Int. J. Geomech.* **2014**, *14*, 04014028. [[CrossRef](#)]
10. Frossard, E.; Hu, W.; Dano, C.; Hicher, P.-Y. Rockfill shear strength evaluation: A rational method based on size effects. *Géotechnique* **2012**, *62*, 415–427. [[CrossRef](#)]
11. Honkanadavar, N.; Gupta, S. Prediction of shear strength parameters for prototype riverbed rockfill material using index properties. In Proceedings of the Indian Geotechnical Conference—2010, Mumbai, India, 16–18 December 2010; pp. 335–338.
12. Froemelt, A.; Dürrenmatt, D.J.; Hellweg, S. Using data mining to assess environmental impacts of household consumption behaviors. *Environ. Sci. Technol.* **2018**, *52*, 8467–8478. [[CrossRef](#)] [[PubMed](#)]
13. Ahmad, M.; Tang, X.-W.; Qiu, J.-N.; Gu, W.-J.; Ahmad, F. A hybrid approach for evaluating CPT-based seismic soil liquefaction potential using Bayesian belief networks. *J. Cent. South Univ.* **2020**, *27*, 500–516.
14. Ahmad, M.; Tang, X.-W.; Qiu, J.-N.; Ahmad, F. Evaluating Seismic Soil Liquefaction Potential Using Bayesian Belief Network and C4.5 Decision Tree Approaches. *Appl. Sci.* **2019**, *9*, 4226. [[CrossRef](#)]
15. Ahmad, M.; Tang, X.; Qiu, J.; Ahmad, F.; Gu, W. LLDV—a Comprehensive Framework for Assessing the Effects of Liquefaction Land Damage Potential. In Proceedings of the 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Dalian, China, 14–16 November 2019; pp. 527–533.
16. Ahmad, M.; Tang, X.-W.; Qiu, J.-N.; Ahmad, F.; Gu, W.-J. A step forward towards a comprehensive framework for assessing liquefaction land damage vulnerability: Exploration from historical data. *Front. Struct. Civ. Eng.* **2020**, *14*, 1476–1491. [[CrossRef](#)]
17. Ahmad, M.; Tang, X.; Ahmad, F. Evaluation of Liquefaction-Induced Settlement Using Random Forest and REP Tree Models: Taking Pohang Earthquake as a Case of Illustration. In *Natural Hazards-Impacts, Adjustments & Resilience*; IntechOpen: London, UK, 2020.
18. Ahmad, M.; Al-Shayea, N.A.; Tang, X.-W.; Jamal, A.; Al-Ahmadi, H.M.; Ahmad, F. Predicting the Pillar Stability of Underground Mines with Random Trees and C4.5 Decision Trees. *Appl. Sci.* **2020**, *10*, 6486. [[CrossRef](#)]
19. Mosavi, A.; Shirzadi, A.; Choubin, B.; Taromideh, F.; Hosseini, F.S.; Borji, M.; Shahabi, H.; Salvati, A.; Dineva, A.A. Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. *IEEE Access* **2020**, *8*, 145968–145983. [[CrossRef](#)]
20. Mosavi, A.; Golshan, M.; Janizadeh, S.; Choubin, B.; Melesse, A.M.; Dineva, A.A. Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: A priority assessment of sub-basins. *Geocarto Int.* **2020**, 1829101. [[CrossRef](#)]
21. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Goodarzi, M.; Dineva, A.A. Groundwater salinity susceptibility mapping using classifier ensemble and Bayesian machine learning models. *IEEE Access* **2020**, *8*, 145564–145576. [[CrossRef](#)]
22. Mosavi, A.; Sajedi-Hosseini, F.; Choubin, B.; Taromideh, F.; Rahi, G.; Dineva, A.A. Susceptibility mapping of soil water erosion using machine learning models. *Water* **2020**, *12*, 1995. [[CrossRef](#)]
23. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Abdolshahnejad, M.; Gharechae, H.; Lahijanzadeh, A.; Dineva, A.A. Susceptibility Prediction of Groundwater Hardness Using Ensemble Machine Learning Models. *Water* **2020**, *12*, 2770. [[CrossRef](#)]
24. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Taromideh, F.; Ghodsi, M.; Nazari, B.; Dineva, A.A. Susceptibility mapping of groundwater salinity using machine learning models. *Environ. Sci. Pollut. Res.* **2021**, *28*, 10804–10817. [[CrossRef](#)]
25. Mosavi, A.; Hosseini, F.S.; Choubin, B.; Goodarzi, M.; Dineva, A.A.; Sardooi, E.R. Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction. *Water Resour. Manag.* **2021**, *35*, 23–37. [[CrossRef](#)]
26. Choubin, B.; Borji, M.; Hosseini, F.S.; Mosavi, A.; Dineva, A.A. Mass wasting susceptibility assessment of snow avalanches using machine learning models. *Sci. Rep.* **2020**, *10*, 18363. [[CrossRef](#)]
27. Kaunda, R. Predicting shear strengths of mine waste rock dumps and rock fill dams using artificial neural networks. *Int. J. Min. Miner. Eng.* **2015**, *6*, 139–171. [[CrossRef](#)]
28. Zhou, J.; Li, E.; Wei, H.; Li, C.; Qiao, Q.; Armaghani, D.J. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Appl. Sci.* **2019**, *9*, 1621. [[CrossRef](#)]

29. Ahmad, M.; Kamiński, P.; Olczak, P.; Alam, M.; Iqbal, M.J.; Ahmad, F.; Sasui, S.; Khan, B.J. Development of Prediction Models for Shear Strength of Rockfill Material Using Machine Learning Techniques. *Appl. Sci.* **2021**, *11*, 6167. [[CrossRef](#)]
30. Zhao, K.; Popescu, S.; Meng, X.; Pang, Y.; Agca, M. Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sens. Environ.* **2011**, *115*, 1978–1996. [[CrossRef](#)]
31. Pasolli, L.; Melgani, F.; Blanzieri, E. Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 464–468. [[CrossRef](#)]
32. Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006; Volume 38, pp. 715–719.
33. Keerthi, S.S.; Lin, C.-J. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput.* **2003**, *15*, 1667–1689. [[CrossRef](#)] [[PubMed](#)]
34. Lin, H.-T.; Lin, C.-J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Submitt. Neural Comput.* **2003**, *3*, 16.
35. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995.
36. Goldberg, Y.; Elhadad, M. splitSVM: Fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications. In Proceedings of the ACL-08: HLT, Short Papers, Columbus, OH, USA, 16–17 June 2008; pp. 237–240.
37. Üstün, B.; Melssen, W.J.; Buydens, L.M. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. *Chemom. Intell. Lab. Syst.* **2006**, *81*, 29–40. [[CrossRef](#)]
38. Santhi, C.; Arnold, J.G.; Williams, J.R.; Dugas, W.A.; Srinivasan, R.; Hauck, L.M. Validation of the swat model on a large rwer basin with point and nonpoint sources 1. *JAWRA J. Am. Water Resour. Assoc.* **2001**, *37*, 1169–1188. [[CrossRef](#)]
39. Van Liew, M.; Arnold, J.; Garbrecht, J. Hydrologic simulation on agricultural watersheds: Choosing between two models. *Trans. ASAE* **2003**, *46*, 1539. [[CrossRef](#)]
40. Lin, S.; Zheng, H.; Han, C.; Han, B.; Li, W. Evaluation and prediction of slope stability using machine learning approaches. *Front. Struct. Civ. Eng.* **2021**, *15*, 821–833. [[CrossRef](#)]
41. Andjelkovic, V.; Pavlovic, N.; Lazarevic, Z.; Radovanovic, S. Modelling of shear strength of rockfills used for the construction of rockfill dams. *Soils Found.* **2018**, *58*, 881–893. [[CrossRef](#)]
42. Yang, Y.; Zhang, Q. A hierarchical analysis for rock engineering using artificial neural networks. *Rock Mech. Rock Eng.* **1997**, *30*, 207–222. [[CrossRef](#)]
43. Faradonbeh, R.S.; Armaghani, D.J.; Abd Majid, M.; Tahir, M.M.; Murlidhar, B.R.; Monjezi, M.; Wong, H. Prediction of ground vibration due to quarry blasting based on gene expression programming: A new model for peak particle velocity prediction. *Int. J. Environ. Sci. Technol.* **2016**, *13*, 1453–1464. [[CrossRef](#)]
44. Chen, W.; Hasanipanah, M.; Rad, H.N.; Armaghani, D.J.; Tahir, M. A new design of evolutionary hybrid optimization of SVR model in predicting the blast-induced ground vibration. *Eng. Comput.* **2019**, *37*, 1455–1471. [[CrossRef](#)]
45. Rad, H.N.; Bakhshayeshi, I.; Jusoh, W.A.W.; Tahir, M.; Foong, L.K. Prediction of flyrock in mine blasting: A new computational intelligence approach. *Nat. Resour. Res.* **2020**, *29*, 609–623.
46. Ahmad, M.; Hu, J.-L.; Ahmad, F.; Tang, X.-W.; Amjad, M.; Iqbal, M.J.; Asim, M.; Farooq, A. Supervised Learning Methods for Modeling Concrete Compressive Strength Prediction at High Temperature. *Materials* **2021**, *14*, 1983. [[CrossRef](#)]
47. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [[CrossRef](#)]