

## Research Article

# Developing a Robust Model Based on the Gaussian Process Regression Approach to Predict Biodiesel Properties

Inna Pustokhina <sup>1</sup>, Amir Seraj <sup>2</sup>, Hafsan Hafsana <sup>3</sup>, Seyed Mojtaba Mostafavi <sup>4</sup>,  
and S. M. Alizadeh <sup>5</sup>

<sup>1</sup>Department of Propaedeutics of Dental Diseases, I. M. Sechenov First Moscow State Medical University, Sechenov University, Moscow, Russia

<sup>2</sup>Department of Instrumentation and Industrial Automation, Ahwaz Faculty of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

<sup>3</sup>Biology Department, Faculty of Science and Technology, Universitas Islam Negeri Alauddin Makassar, Sultan Alauddin Street, Gowa 92118, South Sulawesi, Indonesia

<sup>4</sup>HiTech Institute of Theoretical and Computational Chemistry, India

<sup>5</sup>Petroleum Engineering Department, Australian College of Kuwait, West Mishref, Kuwait City, Kuwait

Correspondence should be addressed to Amir Seraj; [amir.seraj@afp.put.ac.ir](mailto:amir.seraj@afp.put.ac.ir)

Received 28 April 2021; Accepted 26 May 2021; Published 7 June 2021

Academic Editor: Alireza Baghban

Copyright © 2021 Inna Pustokhina et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Biodiesel is assumed a renewable and environmentally friendly fuel that possesses the potential to substitute petroleum diesel. The basic purpose of the present study is to design a precise algorithm based on Gaussian Process Regression (GPR) model with several kernel functions, i.e., Rational Quadratic, Squared Exponential, Matern, and Exponential, to estimate biodiesel properties. These properties include kinematic viscosity (KV), pour point (PP), iodine value (IV), and cloud point (CP) as a function of fatty acid composition. In order to develop this model, some variables are assumed, such as molecular weight, carbon number, double bond numbers, monounsaturated fatty acids, polyunsaturated fatty acid, weight percent of saturated acid, and temperature. The performance and efficiency of the GPR model are measured through several statistical criteria and the results are summarized in root mean square error (RMSE) and coefficients of determination ( $R^2$ ).  $R^2$  and RMSE are sorted as 0.992 & 0.15697, 0.998 & 0.96580, 0.966 & 1.38659, and 0.968 & 1.56068 for four properties such as KV, IV, CP, and PP, respectively. It is worth to mention this point that the kernel function Squared Exponential shows a great performance for IV and PP and kernel functions Exponential and Matern indicate appropriate efficiency for CP and KV properties, respectively. On the other hand, the results of the offered GPR models are compared with those of the previous models, LSSVM-PSO and ANFIS. The outcomes proved the superiority of this model over two former models in point of estimating the biodiesel properties.

## 1. Introduction

In the age of increasing greenhouse gases, a sharp fall in oil resources and rising fossil fuel prices forced the authorities to attend to biomass resources much more than previous [1–4]. All of these convincing reasons make the biofuels, such as biodiesel and bioethanol, suitable and major alternatives for fossil fuels [2, 5, 6]. Biodiesel has high adaptability to the environment and on the other hand is reproducible fuels [7, 8]. For these convincing reasons, this fuel is a

suitable replacement for petroleum diesel [9–11]. Some fuels such as biodiesel can be acquired by alcoholics and chain reactions of animal fats and vegetables with several light alcohols involving ethanol and methane while acidic catalysts or alkaline is applied to them [12]. These chain reactions mainly diminish the oil viscosity. The general properties of biodiesel mainly depend on the construction of employed oil [13]. Many different and determinative properties such as density, IV, KV, flash, PP, and CP are mentioned for the quality of biodiesel [14, 15]. The fact is that the experimental

evaluation of these properties is simple, but it required expansive time and cost. The offering model should predict the fuel properties, and on the other hand, it should enhance the quality of biofuel; therefore, it is essential to propose an accurate and reliable model. To approximate the Cetane number about biodiesel, Ramadhas et al. proposed a multilayer feed forward model [16]. To estimate the Cetane number and fatty acid methyl esters (FAME), Hansen and Bamgboye suggested a novel correlation [17]. To measure the iodine and saponification value of different types of biodiesel, Gopinath et al. suggested a multiple linear regression model. They could reduce estimation error to about 3.4% [18]. Phankosol and coworkers expanded an experimental model based on double bond and carbon numbers in various temperature span to evaluate biodiesel viscosity. The Average Absolute Deviation (AAD) for this algorithm is estimated to be about 6.95% [19]. Rocabrano-Valdés and coworkers performed an artificial neural network (ANN) to predict the number of cetane, biodiesel density, and dynamic viscosity. MSE for validation set of this model is about  $1.842 \times 10^{-3}$  [20]. Talebi et al. formed a modern system to analyze and evaluate biodiesel features according to profile of FAME [21]. Miraboutalebi and coworkers handled an ANN model to evaluate cetane numbers. By analyzing statistical data, it can be understood that RMSE and  $R^2$  are about 2.53 and 0.95, respectively [22]. Hong and coworkers employed fatty acid methylparaben esters profile to approximate the biodiesel features. The range of Average Absolute Error (AAE) in Hong work was between 0.14 and 7.5 percent [23]. Giwa and coworkers employed a multilayer perceptron neural network (MLP) to approximate the number of cetane in biodiesel features. Their model is inspired from five fatty acids [24]. Hosseinpour and coworkers predicted the number of cetane by mixture of ANN and partial least square. In this method, the percent error (PE),  $R^2$ , and MSE are about 1.06, 0.99, and 0.72, respectively [25]. Mostafaei proposed a system known as ANFIS to predict the number of cetane in biodiesel [26]. It is noteworthy to mention that, because of the deficiency of worthy and useful experimental data, expanding an accurate model can be helpful to researchers. In recent years, a minority of estimated models such as artificial intelligence methods have been executed to evaluate material properties and processes in different applications [27–33]. Recently, with extensive development in technology and science, some novel and smart methods are suggested such as GMDH (Group Method of Data Handling), GPR, ANFIS, ANN, and LSSVM; by means of these useful methods, many complex and nonlinear problems can be modelled in many different branches [34–38].

Based on the study, approximately there are no studies around biodiesel properties such as PP, CP, KV, and IV; in other words, there is not any accurate and smart model to be able to model these four parameters. This work aimed to extend a detailed model to approximate the biodiesel properties as mentioned above in point of fatty acid methyl ester utilizing the GPR algorithm. To achieve this goal, an extensive dataset is utilized and has been evaluated for the model accuracy and precession by statistical parameters.

## 2. GPR Method

**2.1. A Summary of Gaussian Process Regression.** Gaussian Process (GP) is described as a complex of random variables, in which some variables have a multivariable distribution of Gaussian. GPRs are nonprometric probabilistic models which are based on kernel. Suppose a training set,  $\{(x_i, y_i); i = 1, 2, \dots\}$  that  $y_i$  and  $x_i \in R^d$  and both of them are from unknown distributions. A trained GPR model predicts the value of  $y^{\text{new}}$  which its input is the matrix  $x^{\text{new}}$ . Suppose a linear regression function,  $y = x^T \beta + \varepsilon$ , which  $\varepsilon \sim N(0, 0^2)$ . The GPR method tends to explain  $y$  by presenting hidden variables which can be shown by  $l(x_i)$  that  $i = 1, 2, 3, \dots, n$ , which starts from a Gaussian Process (GP) while the common distribution of  $l(x_i)$ 's is a Gaussian function and fundamental function,  $b$ .  $l(x_i)$ 's is a covariance function that catches the smoothness of  $y$ . The base function has to project  $x$  in a feature space. The dimension of feature space is  $p$ . Covariance and mean are the principal parameters by which a GP is described. Suppose the mean function of  $l(x)$  is  $m(x) = E(l(x))$  and its covariance function is  $k(x, x') = \text{Cov}[l(x), l(x')]$ . Now assume the GPR model is like  $y = b(x)^T \beta + l(x)$ , which  $b(x) \in R^d$  and  $l(x) \sim \text{GP}(0, k(x, x'))$ .  $\theta$  is a hyperparameter of  $k(x, x')$  and therefore can be expressed like  $k(x, x'|\theta)$ . Generally, several algorithms approximate  $\theta$ ,  $\sigma^2$ , and  $\beta$  for training a suitable model and allocate initial values and some specifications such as  $k$  and  $b$  as parameters. This study investigates four disparate and important kernel functions such as Rational Quadratic, Matern, Exponential, and Squared Exponential. Equations (1)–(4) present these specifications, respectively, and in these equations,  $\sigma_l$  is the scale of characteristics length which means how  $x$ 's can be far from  $y$ 's to become uncorrelated.  $\sigma_f$  is standard deviation,  $r = \sqrt{(x_i - x_j)^T (x_i - x_j)}$ , and on the other hand,  $\alpha$  is a positive parameter with complexity in scale. It is important to mention that  $\sigma_f$  and  $\sigma_l$  should be greater than zero [39]. This could be possible only through  $\theta$  which included two parameters  $\theta_1 = \log \sigma_l$  and  $\theta_2 = \log \sigma_f$ .

$$k(x_i, x_j|\theta) = \sigma_f^2 \exp\left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha}, \quad (1)$$

$$k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \frac{\sqrt{5} r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left(-\frac{\sqrt{5} r}{\sigma_l}\right), \quad (2)$$

$$k(x_i, x_j|\theta) = \sigma_f^2 \left(-\frac{\sqrt{5} r}{\sigma_l}\right), \quad (3)$$

$$k(x_i, x_j|\theta) = \sigma_f^2 \exp\left[-\frac{1}{2} \frac{(x_i - x_j)^T (x_i - x_j)}{\sigma_l^2}\right]. \quad (4)$$

In such previous equations, four base functions are studied here as well; these functions include constant, empty, linear, and pure quadratic as can be seen in equations (5)–(8), respectively. The specifications of base functions are as follows:

$$X = (x_1, x_2, \dots, x_n)^T,$$

$$X^2 = \begin{pmatrix} x_{11}^2 & \dots & x_{1d}^2 \\ \vdots & \ddots & \vdots \\ x_{n1}^2 & \dots & x_{nd}^2 \end{pmatrix}, \quad (5)$$

$$B = I_{n \times 1}, \quad (6)$$

$$B = \text{empty matrix},$$

$$B = [1, X], \quad (7)$$

$$B = [1, X, X^2], \quad (8)$$

which  $B = (b(x_1), b(x_2), \dots, b(x_n))^T$ .

For estimating  $\theta$ ,  $\sigma^2$ , and  $\beta$ , the function known as marginal log-likelihood that mentioned in equation (9) should be maximized.  $K(X, X|\theta)$  in matrix  $\begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{pmatrix}$  is covariance function. The final goal is maximizing the equation  $\log\{P(y|X, \beta(\theta, \sigma^2), \sigma^2, \theta)\}$  based on  $\sigma^2$  and  $\beta$ . The function in the log is known as the likelihood function. Firstly, algorithm calculates  $\beta(\theta, \sigma^2)$ ; this should be maximized in respect to  $\sigma^2$ ,  $\beta$ , and  $\theta$ . This assists to obtain  $\beta$  - profile likelihood that mentioned above and as mentioned previously, it should be maximized by two parameters  $\sigma^2$  and  $\beta$ .

$$\log P(y|X, \beta\sigma^2, \theta) = -\frac{1}{2} \left\{ (y - B\beta)^T [K(X, X|\theta) + \sigma^2 I_n]^{-1} (y - B\beta) \right\} - \frac{n}{2} \log 2\pi - \frac{1}{2} \log |K(X, X|\theta) + \sigma^2 I_n|. \quad (9)$$

### 3. Data Gathering

56 laboratory data of IV were extracted from previously reported sources [18, 21]. 25 and 44 experimental data were employed for PP and CP, respectively [40, 41]. In addition, 59 data were utilized for KV of biodiesel [19, 24]. Data are separated into two categories: 75 percent of the total data are randomly selected for training and the rest are categorized into a testing set for validating model. The input data for evaluation of KV are double bounds number ( $d_n$ ), carbon number ( $C_n$ ), and temperature ( $K$ ). To estimate the IV, the input data are weight percent of poly unsaturated fatty acid (PU), monounsaturated fatty acid (MU), and, moreover, the number of double bonds. To estimate CP, input data include weight percent of saturated fatty acid ( $C_0$ ), carbon number ( $C_n$ ), and molecular weight ( $M_w$ ). For evaluation of PP, the input data include number of double bond, molecular weight, and carbon number.  $C_n$ ,  $d_n$ , and  $M_w$  are expressed in the following formulas:

$$C_n = \sum C_i X_i, \quad (10)$$

$$d_n = C_1 + 2 \times C_2 + 3 \times C_3, \quad (11)$$

$$M_w = 14 \times (C_n + 1) - 2 \times d_n + 32, \quad (12)$$

$$C_0 = \sum C_0 X_i. \quad (13)$$

In these formulations  $C_3$ ,  $C_2$ , and  $C_1$  represent the sum of triunsaturated fatty acid methyl esters, di, and mono, respectively.  $X_i$  indicates mass fraction [26].

### 4. Results and Discussion

This study introduced a new algorithm known as GPR to predict the Biodiesel properties. The principle goal of this section is the graphical and statistical analysis of the GPR algorithm. Figures 1–4 show the graphical view between

predicted and experimental output; in other terms, these figures, which are known as data index charts, attempt to compare the experimental output with predicted values graphically. Usually in experimental papers, there is a slight difference between the actual values and the modelled values [42–44]. Figures 1–4 represent the data index for four different biodiesel properties such as IV, KV, PP, and CP and each figure is divided into four subfigures that compare the performance of the network in disparate kernel functions. Subfigures a, b, c, and d represent disparate kernel functions of GPR, as follows: Exponential, Matern, Squared Exponential, and Rational Quadratic, respectively. Carefully in these figures, it can be understood that all kernel functions have a nice cover between experimental and predicted outputs and have a good prediction. As can be observed, the estimated and experimental values nicely overlapped which show the proposed GPR algorithm has great efficiency in the prediction of disparate properties of biodiesel.

Figures 5–8 indicate the regression plot for each biodiesel property and each figure compares the performance of the GPR model among four different GPR kernel functions. As can be seen in the four figures, the majority of test data accumulated near  $y = x$  line. As can be understood from the regression plot, increasing in test data accumulation rate near  $y = x$  line resulted in high accuracy in the GPR model. As mentioned earlier, these charts represent a graphical and general view, and it generally can be seen that all test data are near  $y = x$  line. Based on these charts, the result of the GPR model has an acceptable performance in all biodiesel properties.

In order to acquire a better understanding of regression charts and more explanations, Tables 1–4 are prepared. According to Table 1 which is related to KV property, if consider closely  $R^2$  field, the test value for the Matern kernel function is 0.992 which is more appropriate in comparison with other kernel functions. The RMSE value for this kernel function is 0.15697. If consider Table 2 carefully, for IV, the best  $R^2$  value for test data is 0.998 while the RMSE value is

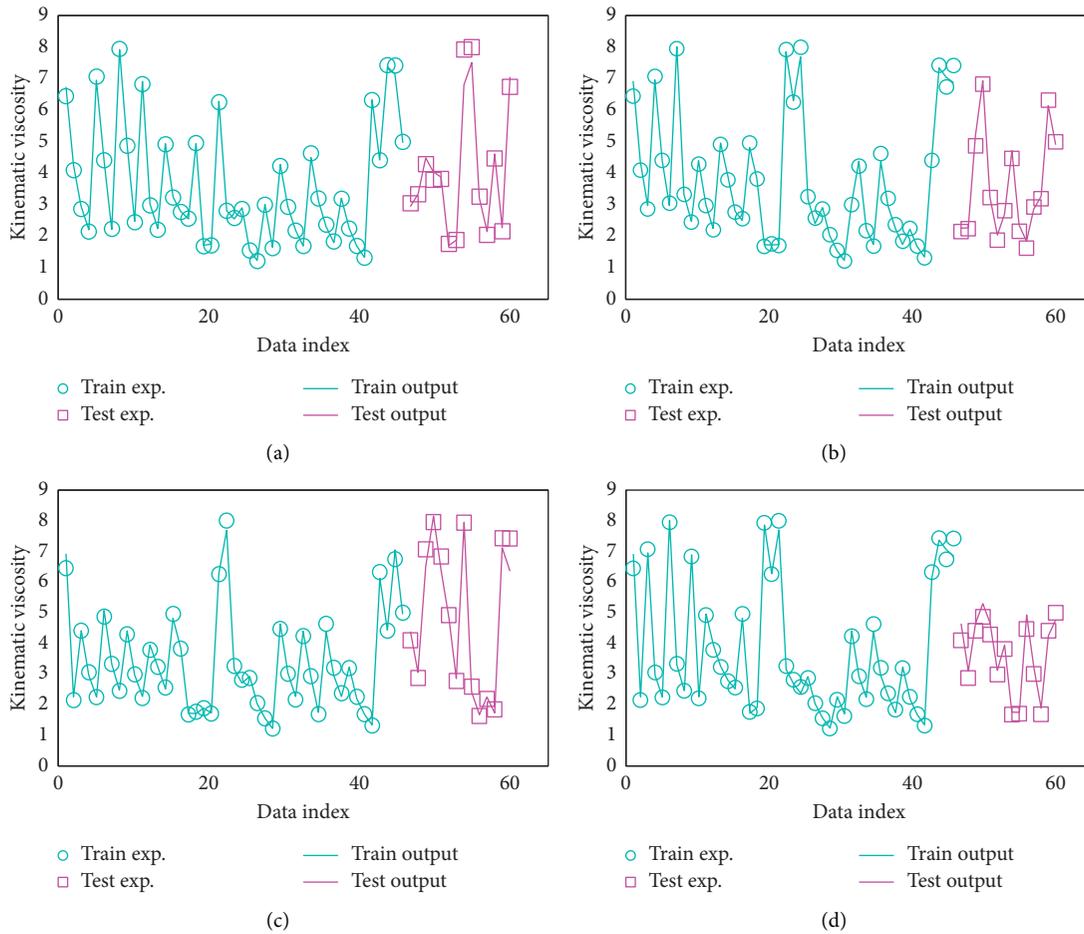


FIGURE 1: Data index graph for train and test data for KV property based on different kernel functions of GPR.

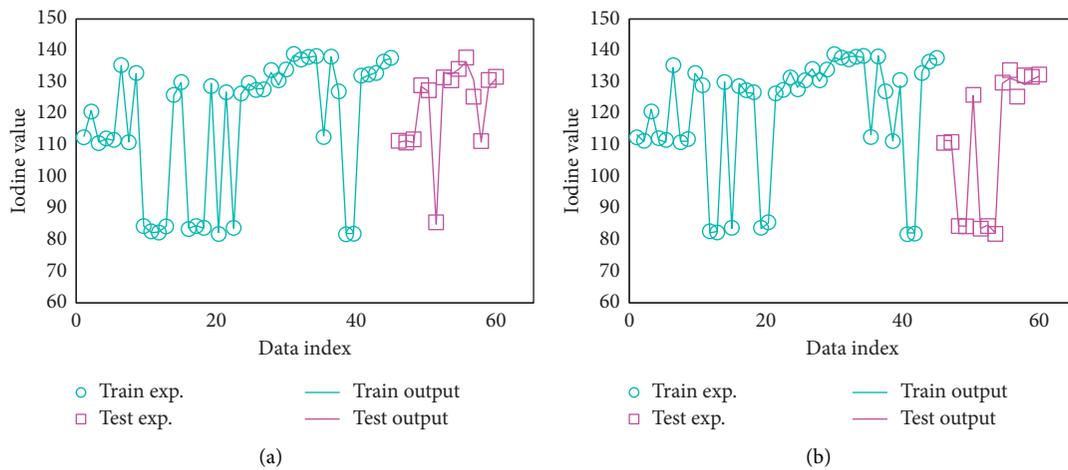


FIGURE 2: Continued.

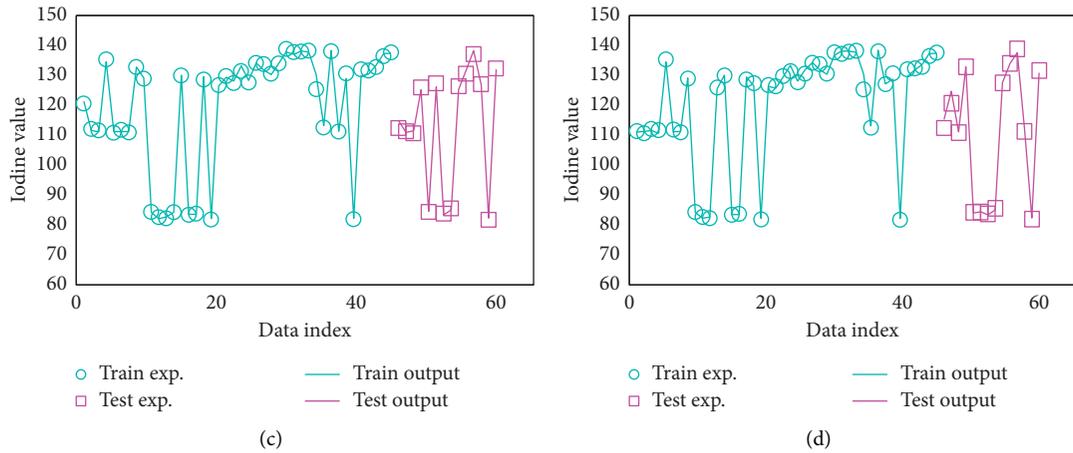


FIGURE 2: Data index graph for train and test data for KV property based on different kernel functions of GPR.

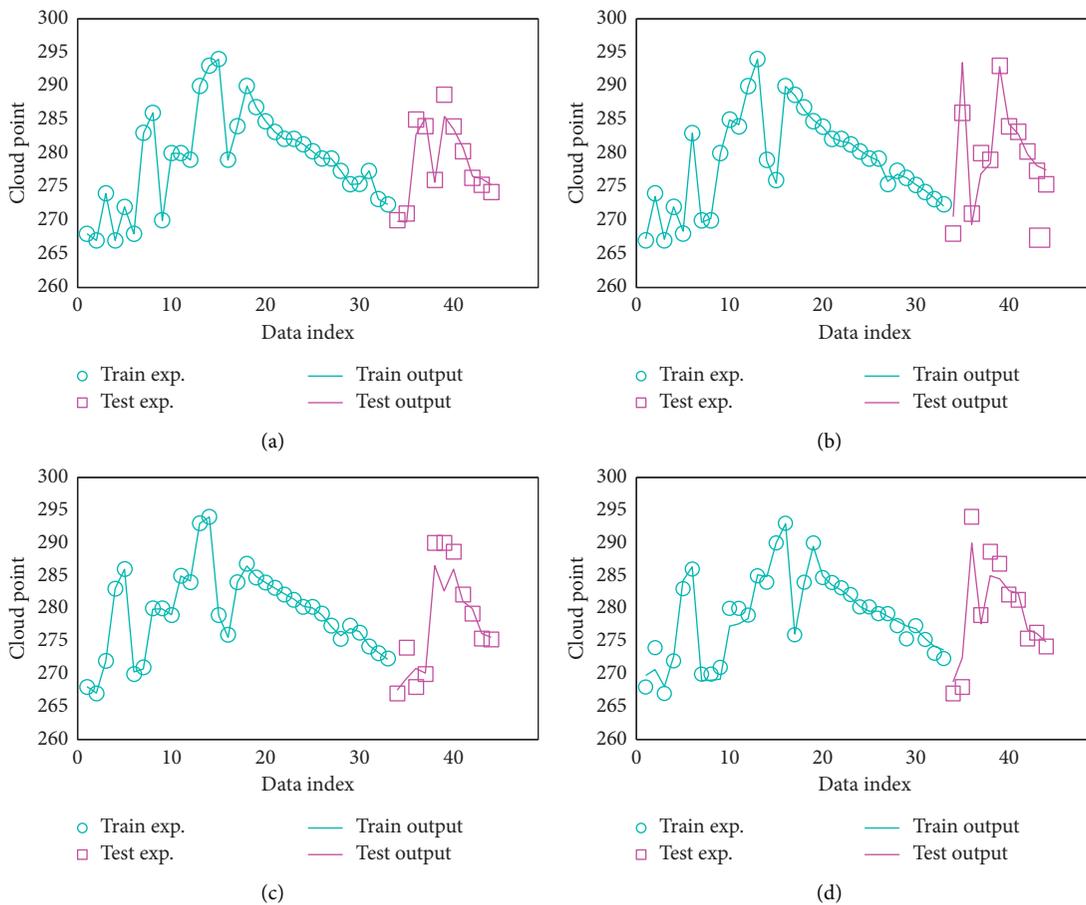


FIGURE 3: Data index graph for train and test data for KV property based on different kernel functions of GPR.

0.96580 and this belongs to the Squared Exponential kernel function. As can be seen in Table 3, the best  $R^2$  value in test data for CP property belongs to exponential function with a value about 0.966 while Squared Exponential kernel function for PP property in Table 4 has the best value for  $R^2$  about 0.968. The RMSE values for CP and PP are 1.38659 and

1.56068, respectively. As can be seen in regression plots (diagrams 5–8) and the reported data in Tables 1–4, the fitting line for each property is near  $y=x$  line. It can be deduced that the GPR algorithm has appropriate accuracy for the prediction of biodiesel properties. In order to check out the proposed model more meticulously with the

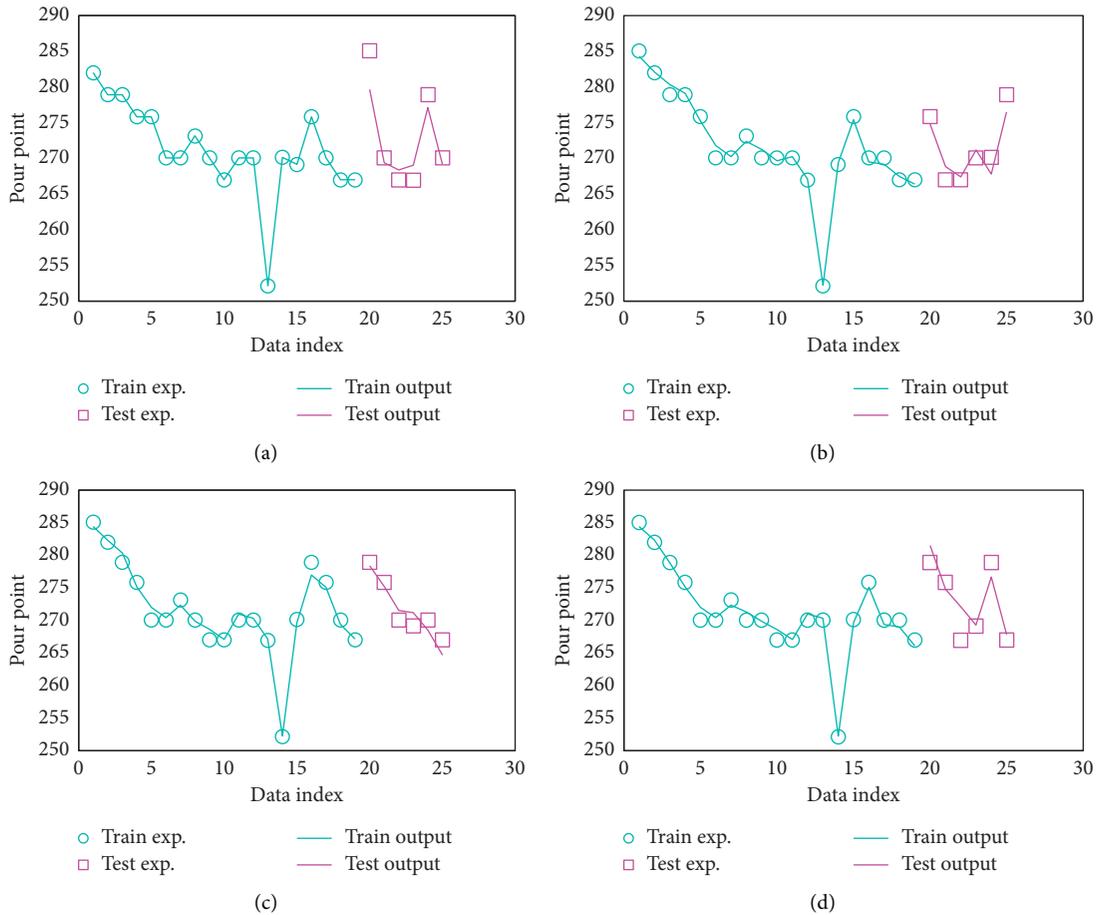


FIGURE 4: Data index graph for train and test data for KV property based on different kernel functions of GPR.

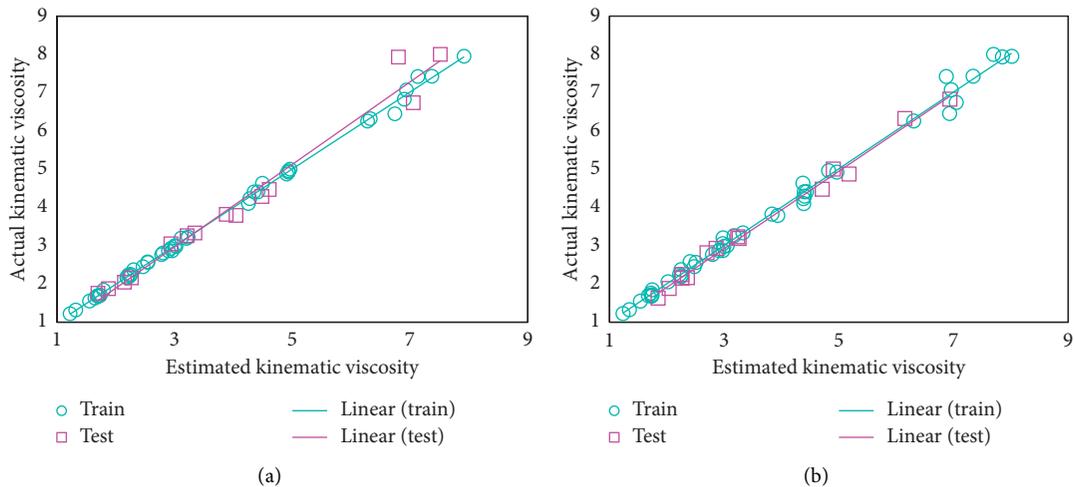


FIGURE 5: Continued.

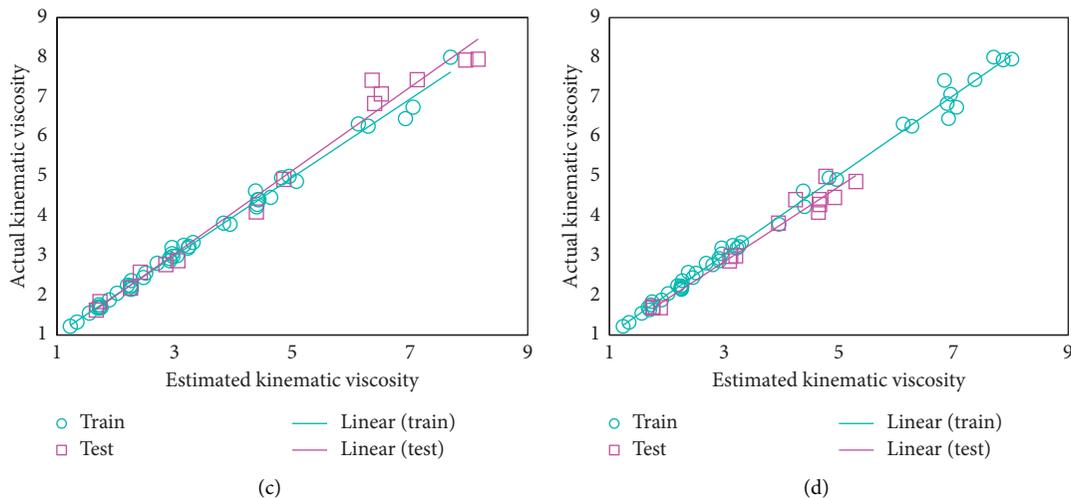


FIGURE 5: Regression plot for KV property based on different kernel functions of GPR.

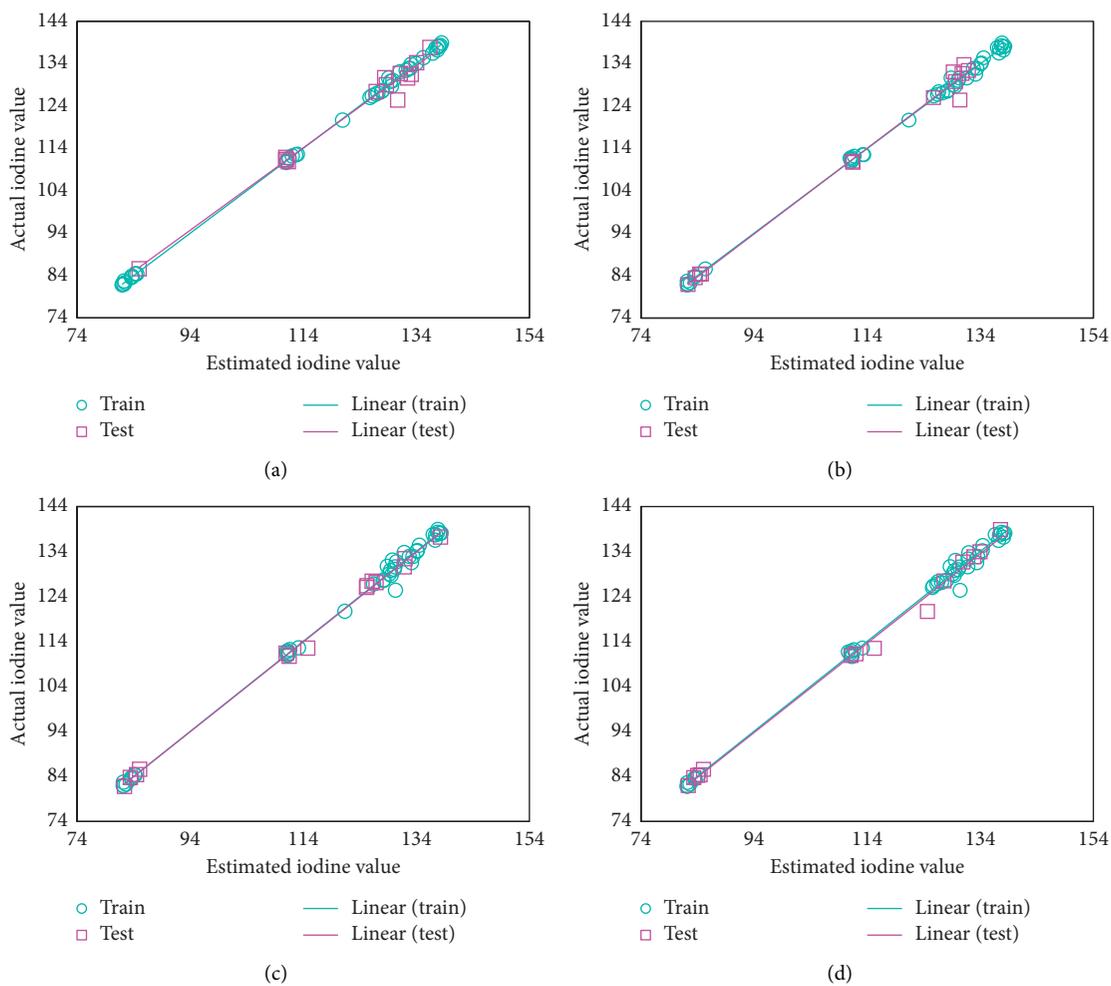


FIGURE 6: Regression plot for IV property based on different kernel functions of GPR.

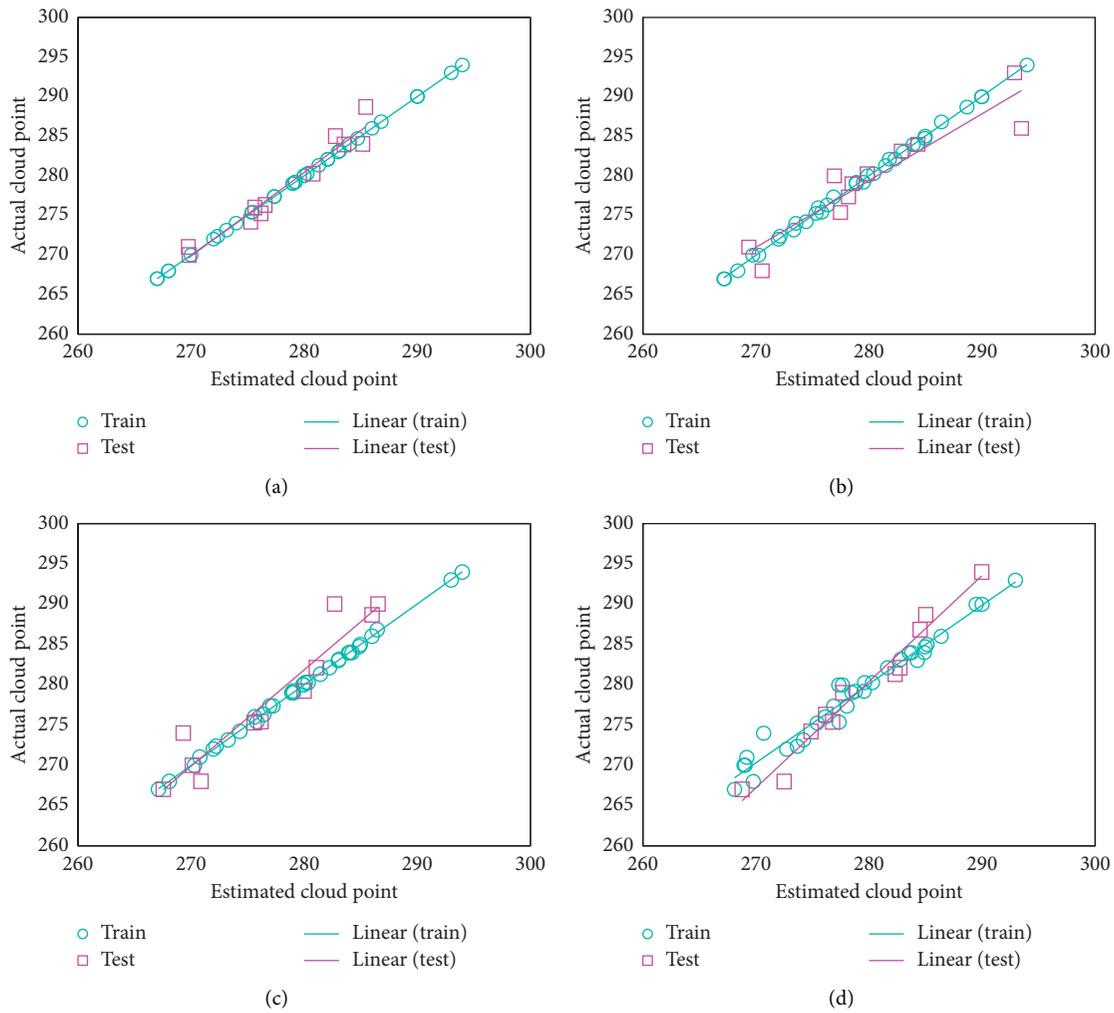


FIGURE 7: Regression plot for CP property based on different kernel functions of GPR.

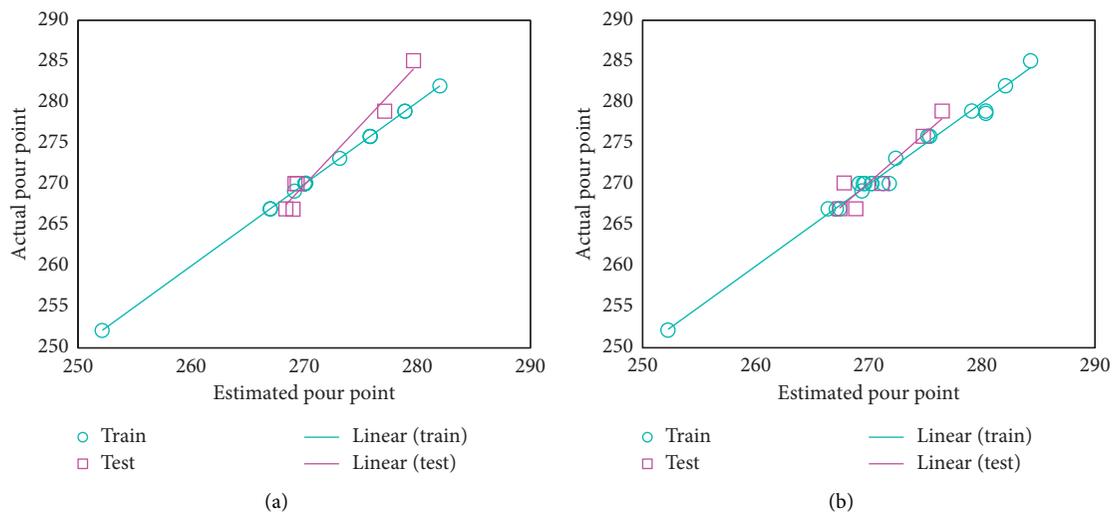


FIGURE 8: Continued.

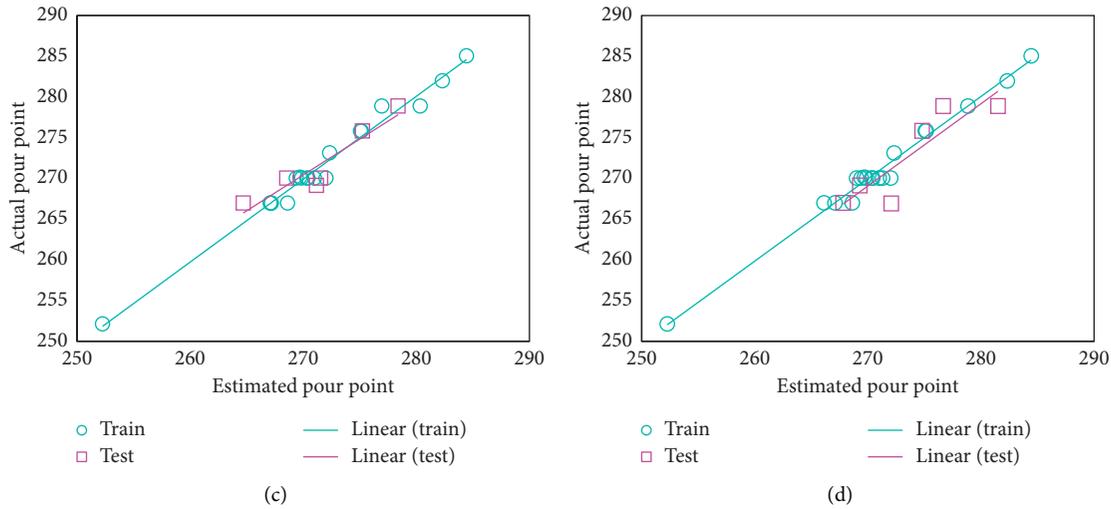


FIGURE 8: Regression plot for PP property based on different kernel functions of GPR.

TABLE 1: The comparison between four kernel functions for KV property from the point of view of statistical quantities.

Model	Phase	$R^2$	MRE (%)	MSE	RMSE	STD
GPR (Exponential)	Train	0.998	2.61001	0.00622	0.07884	0.06408
	Test	0.975	4.50131	0.12657	0.35577	0.29305
	Total	0.991	3.25054	0.03573	0.35577	0.16650
GPR (Matern)	Train	0.994	1.19836	0.02510	0.15841	0.12080
	Test	0.992	4.29964	0.02464	0.15697	0.08520
	Total	0.993	2.04742	0.02624	0.15697	0.11351
GPR (Squared Exponential)	Train	0.993	2.40750	0.01830	0.13529	0.10158
	Test	0.981	5.07691	0.13892	0.37272	0.28103
	Total	0.991	2.38574	0.03773	0.37272	0.16620
GPR (Rational Quadratic)	Train	0.994	2.77862	0.02566	0.16019	0.12024
	Test	0.969	6.90077	0.08160	0.28565	0.14865
	Total	0.991	3.50671	0.03681	0.28565	0.14139

TABLE 2: The comparison between four kernel functions for IV property from the point of view of statistical quantities.

Model	Phase	$R^2$	MRE (%)	MSE	RMSE	STD
GPR (Exponential)	Train	1.000	0.23024	0.15939	0.39923	0.29438
	Test	0.984	0.96971	3.15731	1.77688	1.35545
	Total	0.998	0.27729	0.79448	1.77688	0.83169
GPR (Matern)	Train	0.999	0.44408	0.44046	0.66367	0.39484
	Test	0.993	0.83689	3.02336	1.73878	1.45060
	Total	0.997	0.46640	0.98545	1.73878	0.81850
GPR (Squared Exponential)	Train	0.997	0.55285	1.18464	1.08841	0.85742
	Test	0.998	0.69364	0.93277	0.96580	0.56754
	Total	0.997	0.62869	1.22526	0.96580	0.81806
GPR (Rational Quadratic)	Train	0.996	0.60686	1.30529	1.14249	0.87268
	Test	0.996	0.74534	1.93710	1.39180	1.14923
	Total	0.996	0.64906	1.56170	1.39180	0.99212

outcomes of ANFIS and LSSVM models reported in previous literature [45], Tables 1–4 are prepared as numerical.

Table 5 summarizes  $R^2$  and RMSE values among three different models for four biodiesel properties. As it can be understood from Table 5, the GPR model indicates the best

performance for IV property in comparison with ANFIS and LSSVM-PSO with a value of about 0.998 in  $R^2$  while for other properties ANFIS and LSSVM show more appropriate performance in terms of the view of  $R^2$  and RMSE.

TABLE 3: The comparison between four kernel functions for CP property from the point of view of statistical quantities.

Model	Phase	$R^2$	MRE (%)	MSE	RMSE	STD
GPR (Exponential)	Train	0.967	0.00069	0.00001	0.00247	0.00158
	Test	0.966	0.37589	1.92262	1.38659	0.93954
	Total	0.955	0.09442	0.48066	1.38659	0.64771
GPR (Matern)	Train	0.999	0.07276	0.06294	0.25089	0.15196
	Test	0.874	0.62866	7.30808	2.70335	2.15533
	Total	0.944	0.37087	2.73100	2.70335	1.30838
GPR (Squared Exponential)	Train	0.999	0.04991	0.03376	0.18375	0.12258
	Test	0.904	0.79289	9.58471	3.09592	2.24565
	Total	0.950	0.27415	2.50823	3.09592	1.40005
GPR (Rational Quadratic)	Train	1.000	0.31099	1.38429	1.17656	0.81684
	Test	0.950	0.69381	5.80117	2.40856	1.48864
	Total	0.990	0.39714	2.40895	2.40856	1.10468

TABLE 4: The comparison between four kernel functions for PP property from the point of view of statistical quantities.

Model	Phase	$R^2$	MRE (%)	MSE	RMSE	STD
GPR (Exponential)	Train	0.982	0.00056	0.00000	0.00195	0.00124
	Test	0.878	0.73386	6.64170	2.57715	1.73628
	Total	0.967	0.17651	1.59401	2.57715	1.18811
GPR (Matern)	Train	0.989	0.21290	0.54286	0.73679	0.46621
	Test	0.889	0.55839	2.85624	1.69004	0.80875
	Total	0.973	0.30712	1.15081	1.69004	0.68377
GPR (Squared Exponential)	Train	1.000	0.26262	0.86125	0.92803	0.60432
	Test	0.968	0.52306	2.43573	1.56068	0.72847
	Total	0.966	0.33494	1.33190	1.56068	0.72451
GPR (Rational Quadratic)	Train	0.984	0.25484	0.73098	0.85497	0.51735
	Test	0.788	0.73120	6.62593	2.57409	1.78533
	Total	0.954	0.31634	1.90128	2.57409	1.09967

TABLE 5:  $R^2$  and RMSE values for ANFIS, LSSVM-PSO, and GPR models among different properties.

Biodiesel properties	Testing phase					
	ANFIS [45]		LSSVM-PSO [1]		GPR	
	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
KV	0.997	0.08	0.9986	0.08	0.992	0.15697
IV	0.998	1.02	0.9962	1.192	0.998	0.96580
CP	0.957	1.54	0.9998	0.086	0.966	1.38659
PP	0.996	0.44	0.9999	0.031	0.968	1.56068

## 5. Conclusion

In this paper, a Gaussian Process Regression model using four different kernel functions such as Exponential, Matern, Squared Exponential, and Rational Quadratic was proposed. This model has the ability to estimate the physical and chemical features of the biodiesel material where these properties include KV, PP, CP, and IV. A valuable dataset was collected from different sources for these biodiesel properties. On the other hand, the results of offered GPR model are compared with two previous models, ANFIS and LSSVM-PSO results. The graphical and statistical approaches indicated the GPR model obtained high efficiency in terms of estimation and evaluation of biodiesel properties. The proposed GPR algorithm is easy to apply and researchers can open an account on this algorithm from the

point of view of simplicity and usefulness. This model can be helpful for those who desire to work with biodiesel fuels.

## Data Availability

The data used to support the findings of this study are provided within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] R. Razavi, A. Bemani, A. Baghban, A. H. Mohammadi, and S. Habibzadeh, "An insight into the estimation of fatty acid methyl ester based biodiesel properties using a LSSVM model," *Fuel*, vol. 243, pp. 133–141, 2019.
- [2] A. Bemani, Q. Xiong, A. Baghban, S. Habibzadeh, A. H. Mohammadi, and M. H. Doranehgard, "Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO-LSSVM models," *Renewable Energy*, vol. 150, pp. 924–934, 2020.
- [3] Z. Liu and A. Baghban, "Application of LSSVM for biodiesel production using supercritical ethanol solvent," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 39, no. 17, pp. 1869–1874, 2017.

- [4] A. H. S. Dehaghani and R. Daneshfar, "How much would silica nanoparticles enhance the performance of low-salinity water flooding?" *Petroleum Science*, vol. 16, no. 3, pp. 591–605, 2019.
- [5] J. Guo and A. Baghban, "Application of ANFIS strategy for prediction of biodiesel production using supercritical methanol," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 39, no. 17, pp. 1862–1868, 2017.
- [6] A. Baghban, "Computational modeling of biodiesel production using supercritical methanol," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 41, no. 1, pp. 14–20, 2019.
- [7] W. F. d. C. Rocha and D. A. Sheen, "Determination of physicochemical properties of petroleum derivatives and biodiesel using GC/MS and chemometric methods with uncertainty estimation," *Fuel*, vol. 243, pp. 413–422, 2019.
- [8] A. Dashti, A. S. Noushabadi, M. Raji, A. Razmi, S. Ceylan, and A. H. Mohammadi, "Estimation of biomass higher heating value (HHV) based on the proximate analysis: smart modeling and correlation," *Fuel*, vol. 257, Article ID 115931, 2019.
- [9] X. Feng, C. Yu, Z. Shu et al., "Rapid and non-destructive measurement of biofuel pellet quality indices based on two-dimensional near infrared spectroscopic imaging," *Fuel*, vol. 228, pp. 197–205, 2018.
- [10] M. F. Ferrão, M. d. S. Viera, R. E. P. Pazos, D. Fachini, A. E. Gerbase, and L. Marder, "Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions," *Fuel*, vol. 90, no. 2, pp. 701–706, 2011.
- [11] H. Sharon, R. Jayaprakash, M. Karthigai Selvan, D. R. Soban Kumar, A. Sundaresan, and K. Karupphasamy, "Biodiesel production and prediction of engine performance using SIMULINK model of trained neural network," *Fuel*, vol. 99, pp. 197–203, 2012.
- [12] S. Pradhan, C. S. Madankar, P. Mohanty, and S. N. Naik, "Optimization of reactive extraction of castor seed to produce biodiesel using response surface methodology," *Fuel*, vol. 97, pp. 848–855, 2012.
- [13] L. F. Chuah, S. Yusup, A. R. A. Aziz, J. J. Klemeš, A. Bokhari, and M. Z. Abdullah, "Influence of fatty acids content in non-edible oil for biodiesel properties," *Clean Technologies and Environmental Policy*, vol. 18, no. 2, pp. 473–482, 2016.
- [14] R. Wang, B. Song, W. Zhou et al., "A facile and feasible method to evaluate and control the quality of *Jatropha curcus* L. seed oil for biodiesel feedstock: gas chromatographic fingerprint," *Applied Energy*, vol. 88, no. 6, pp. 2064–2070, 2011.
- [15] M. Balat and H. Balat, "Progress in biodiesel processing," *Applied Energy*, vol. 87, no. 6, pp. 1815–1835, 2010.
- [16] A. S. Ramadhas, S. Jayaraj, C. Muraleedharan, and K. Padmakumari, "Artificial neural networks used for the prediction of the cetane number of biodiesel," *Renewable Energy*, vol. 31, no. 15, pp. 2524–2533, 2006.
- [17] A. Bamgboye and A. C. Hansen, "Prediction of cetane number of biodiesel fuel from the fatty acid methyl ester (FAME) composition," *International Agrophysics*, vol. 22, no. 1, p. 21, 2008.
- [18] A. Gopinath, S. Puhan, and G. Nagarajan, "Theoretical modeling of iodine value and saponification value of biodiesel fuels from their fatty acid composition," *Renewable Energy*, vol. 34, no. 7, pp. 1806–1811, 2009.
- [19] S. Phankosol, K. Sudaprasert, S. Lilitchan, K. Aryusuk, and K. Krisnangkura, "An empirical equation for estimation of kinematic viscosity of fatty acid methyl esters and biodiesel," *Journal of the American Oil Chemists' Society*, vol. 92, no. 7, pp. 1051–1061, 2015.
- [20] C. I. Rocabrano-Valdés, L. F. Ramírez-Verduzco, and J. A. Hernández, "Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel," *Fuel*, vol. 147, pp. 9–17, 2015.
- [21] A. F. Talebi, M. Tabatabaei, and Y. Chisti, "BiodieselAnalyzer: a user-friendly software for predicting the properties of prospective biodiesel," *Biofuel Research Journal*, vol. 1, no. 2, pp. 55–57, 2014.
- [22] S. M. Miraboutalebi, P. Kazemi, and P. Bahrami, "Fatty acid methyl ester (FAME) composition used for estimation of biodiesel cetane number employing random forest and artificial neural networks: a new approach," *Fuel*, vol. 166, pp. 143–151, 2016.
- [23] I. K. Hong, G. S. Jeon, and S. B. Lee, "Prediction of biodiesel fuel properties from fatty acid alkyl ester," *Journal of Industrial and Engineering Chemistry*, vol. 20, no. 4, pp. 2348–2353, 2014.
- [24] S. O. Giwa, S. O. Adekomaya, K. O. Adama, and M. O. Mukaila, "Prediction of selected biodiesel fuel properties using artificial neural network," *Frontiers in Energy*, vol. 9, no. 4, pp. 433–445, 2015.
- [25] S. Hosseinpour, M. Aghbashlo, M. Tabatabaei, and E. Khalife, "Exact estimation of biodiesel cetane number (CN) from its fatty acid methyl esters (FAMEs) profile using partial least square (PLS) adapted by artificial neural network (ANN)," *Energy Conversion and Management*, vol. 124, pp. 389–398, 2016.
- [26] M. Mostafaei, "ANFIS models for prediction of biodiesel fuels cetane number using desirability function," *Fuel*, vol. 216, pp. 665–672, 2018.
- [27] F. Mousazadeh, M. H. T. Naeem, R. Daneshfar, B. S. Soulgani, and M. Naseri, "Predicting the condensate viscosity near the wellbore by ELM and ANFIS-PSO strategies," *Journal of Petroleum Science and Engineering*, vol. 204, Article ID 108708, 2021.
- [28] S. M. Alizadeh, I. Alrueyemi, R. Daneshfar, M. Mohammadi-Khanaposhtani, and M. Naseri, "An insight into the estimation of drilling fluid density at HPHT condition using PSO-, ICA-, and GA-LSSVM strategies," *Scientific Reports*, vol. 11, no. 1, pp. 7033–7114, 2021.
- [29] S. Anbazhagan, V. Thiruvengatam, and K. Kulanthai, "Adaptive neuro-fuzzy inference system and artificial neural network modeling for the adsorption of methylene blue by novel adsorbent in a fixed-bed column method," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 39, no. 6, pp. 75–93, 2020.
- [30] R. Maachou, A. Lekfir, A. Bermad, and S. Abdelaziz, "Energy consumption modeling in activated sludge process using coupling PCA-ANFIS approach," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 38, no. 6, pp. 261–273, 2019.
- [31] A. Lekomtsev, A. Keykhosravi, M. B. Moghaddam, R. Daneshfar, and O. Rezvanjou, "On the prediction of filtration volume of drilling fluids containing different types of nanoparticles by ELM and PSO-LSSVM based models," *Petroleum*, inpress, 2021.
- [32] R. Setiawan, R. Daneshfar, O. Rezvanjou, and S. Ashoori, "Surface tension of binary mixtures containing environmentally friendly ionic liquids: insights from artificial intelligence," *Environment, Development and Sustainability*, pp. 1–22, 2021, inpress.

- [33] D. Ahangari, R. Daneshfar, M. Zakeri, and S. Ashoori, "On the prediction of geochemical parameters (TOC, S1 and S2) by considering well log parameters using ANFIS and LSSVM strategies," *Petroleum*, inpress, 2021.
- [34] N. Nabipour, R. Daneshfar, O. Rezvanjou et al., "Estimating biofuel density via a soft computing approach based on intermolecular interactions," *Renewable Energy*, vol. 152, pp. 1086–1098, 2020.
- [35] R. Daneshfar, A. Bemani, M. Hadipoor et al., "Estimating the heat capacity of non-Newtonian ionanofluid systems using ANN, ANFIS, and SGB tree algorithms," *Applied Sciences*, vol. 10, no. 18, p. 6432, 2020.
- [36] R. Daneshfar, F. Keivanimehr, M. M. Khanaposhtani, and A. Baghban, "A neural computing strategy to estimate dew-point pressure of gas condensate reservoirs," *Petroleum Science and Technology*, vol. 38, no. 10, pp. 1–7, 2020.
- [37] M. B. Vanani, R. Daneshfar, and E. Khodapanah, "A novel MLP approach for estimating asphaltene content of crude oil," *Petroleum Science and Technology*, vol. 37, no. 22, pp. 2238–2245, 2019.
- [38] A. Baghban and M. Adelizadeh, "On the determination of cetane number of hydrocarbons and oxygenates using Adaptive Neuro Fuzzy Inference System optimized with evolutionary algorithms," *Fuel*, vol. 230, pp. 344–354, 2018.
- [39] Z. Yun and X. Xu, "Predicting doped MgB2 superconductor critical temperature from lattice parameters using Gaussian process regression," *Physica C: Superconductivity and its Applications*, vol. 573, Article ID 1353633, 2020.
- [40] H. Imahara, E. Minami, and S. J. F. Saka, "Thermodynamic study on cloud point of biodiesel with its fatty acid composition," *Fuel*, vol. 85, no. 12-13, pp. 1666–1670, 2006.
- [41] J. F. Sierra-Cantor, C. A. Guerrero-Fajardo, and S. E. Reviews, "Methods for improving the cold flow properties of biodiesel with high saturated fatty acids content: a review," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 774–790, 2017.
- [42] A. H. S. Dehaghani, M. S. Taleghani, M. H. Badizad, and R. Daneshfar, "Simulation study of the Gachsaran asphaltene behavior within the interface of oil/water emulsion: a case study," *Colloid and Interface Science Communications*, vol. 33, Article ID 100202, 2019.
- [43] M. Mohammadi-Khanaposhtani, Y. Kazemzadeh, and R. Daneshfar, "Positive coupling effect in gas condensate flow: role of capillary number, Scheludko number and Weber number," *Journal of Petroleum Science and Engineering*, vol. 203, Article ID 108490, 2021.
- [44] D. D. Nguyen, R. Daneshfar, A. H. S. Dehaghani, and C.-H. Su, "The effect of shear rate on aggregation and breakage of asphaltene flocs: experimental study and model-based analysis," *Journal of Molecular Liquids*, vol. 325, Article ID 114861, 2021.
- [45] M. Mostafaei, "Prediction of biodiesel fuel properties from its fatty acids composition using ANFIS approach," *Fuel*, vol. 229, pp. 227–234, 2018.