Research Article

Comprehensive Modeling in Predicting Biodiesel Density Using Gaussian Process Regression Approach

Bingxian Wang [ID] [1] **and Issam Alruyemi** [ID] [2]

[1]*School of Mathematics and Statistics, Huaiyin Normal University, Huaian, Jiangsu 223300, China*
[2]*Fouman Faculty of Engineering, College of Engineering, University of Tehran, Fouman, Iran*

Correspondence should be addressed to Issam Alruyemi; essamkhudur@ut.ac.ir

In this study, four Gaussian process regression (GPR) approaches by various kernel functions have been proposed for the estimation of biodiesel density as the functions of pressure, temperature, molecular weight, and the normal melting point of fatty acid esters. Comparing the actual values with GPR outputs shows that these approaches have good accuracy, but the performance of the rational quadratic GPR model is better than others. In this GPR model, RMSE = 0.47, MSE = 0.22, MRE = 0.04, $R^2 = 1$, and STD is equal to 0.3. In addition, for the first time, this study shows that the effective parameters affect the biodiesel density. According to this analysis, it was shown that among the input parameters, pressure has the greatest effect on the target values with a relevancy factor of 0.59. This study can be used as a suitable and valuable work/tool for chemical and petroleum engineers who attempt environment protection and recovery improvement.

1. Introduction

Recently, in various countries, the issue of energy is deep and complex [1–3]. The greenhouse gas concentration in the atmosphere has been increased because the consumption of fossil fuels increased, and in that case, the earth's temperature increased too [4, 5]. For the solution to these problems, several agreements have been signed to reduce greenhouse gas emissions [6]. There are so many reasons but the main reasons for blooming the renewable energy resources are the aim of controlling the emission of pollutants and the implementation and credibility of these agreements [7, 8]. The two things that mainly can form biodiesel are the oils that come from vegetables and the fat of an animal's body [9]. Biodiesel is a combination of fatty acid alkyl esters [10]. Ethanol and methanol are some of the alcohols that the oils that come from vegetables and the fat of animal's body are transesterification catalytically by them [10]. Biodiesel is a clean fuel to burn, and it is not a toxic fuel because of low concentrations of sulfur. In that case, this fuel has the minimum bad effect on the emission of greenhouse gas changes, and also

unlike fossil fuels, the biodiesel fuel has a less negative effect on our environment too [11, 12]. Additionally, in diesel engines, we can use biodiesel fuels lonely or use this fuel with fossil diesel because it improves engine life [13], although this fuel costs more than petroleum-based diesel and has higher viscosity too. Some other disadvantages of this fuel are lower oxidation stability, higher cloud point, and lower energy content in comparison with petroleum-based diesel [14]. The observational and modeling studies for the establishment of these properties become important because the concentration on usages of the biofuels and their properties have been grown [15, 16]. For example, the density of biodiesel is the important property that has a significant matter in the thermophysical process. One of the important topics in diesel fuels is the investigation of density because diesel fuels have so many technical and economic parts for the usage of the fuel and also the environmental effects [17–19].

For estimating the properties of fossil fuels, lots of investigations have been carried out in the literature and also there are different advancements in this study, such as a new way for guessing many properties of fuels consisting of surface

Table 1: Obtained statistical parameters to evaluate the performance of the models.

Model	Phase	R^2	MRE (%)	MSE	RMSE	STD
GPR (exponential)	Train	1.000	0.01	0.01	0.11	0.10
	Test	1.000	0.05	0.38	0.61	0.45
	Total	1.000	0.02	0.10	0.61	0.29
GPR (Matern)	Train	1.000	0.04	0.18	0.42	0.27
	Test	1.000	0.04	0.25	0.50	0.33
	Total	1.000	0.04	0.19	0.50	0.28
GPR (squared exponential)	Train	1.000	0.04	0.21	0.46	0.30
	Test	1.000	0.04	0.23	0.48	0.32
	Total	1.000	0.04	0.22	0.48	0.30
GPR (rational quadratic)	Train	1.000	0.04	0.22	0.47	0.30
	Test	1.000	0.04	0.22	0.47	0.30
	Total	1.000	0.04	0.22	0.47	0.30

tension and viscosity has been proposed by Queimada and his colleagues [20]. Barati-Harooni and his colleagues estimate interfacial tension between oil and brine with developed least-squares support vector machine in terms of pressure, salinity, and temperature [21]. Rostami and his colleagues correlate interfacial tension of hydrocarbon and water using genetic programming (GP) with an R-squared of 0.910 [22]. The Kay's model is a model that can determine the density of biodiesels that was estimated by Pratas and his colleagues [17]. The alkane density with an average absolute relative deviation of 60% was predicted by Gahk and his colleagues [23]. Miraboutalebi with the help of his coworkers estimate cetane numbers with a root mean squared error (RMSE) of 2.530 and R-squared of 0.950 by implementing an artificial neural network (ANN) [24]. On the other hand, the cetane number in terms of fatty acid methyl ester (FAME) was predicted by Mostafaei with developing the adaptive neuro-fuzzy inference system [25].

For biodiesel properties, there are some investigational searches in the literature. For example, the density of biodiesels was measured by Paratas et al. at atmospheric pressure with ten individual samples in temperatures between 278.150 and 373.170 K [17]. The viscosities and densities of three individual mixtures of methylcyclohexane and fatty acid methyl esters were obtained by Li and his colleagues in the atmosphere in temperatures between 293.150 and 324.150 K [26]. The density of soybean oil biodiesel was experimentally determined by Aitbelale and his colleagues at temperatures between 298.150 and 393.150 K and pressures up to 140.0 MPa [27]. The surface tension values were measured by Aitbelale et al. for three different biodiesels in temperature and pressure of 473.0 K and 7.0 MPa [27].

The lack of sufficient accuracy and difficulties of computations cause more attention in the aforementioned literature. On the other hand, much attention has been paid to artificial intelligence methods to a precise solution in order to model different processes [28–33]. One of the things that are crucial for the process design and the same operation is the accuracy of thermophysical properties. The development of a precise and low-cost advancement towards the estima-

tion of biodiesel properties is worthy because of the necessity of these properties most importantly in the clean energy resource topics. For this job, the density of biodiesel fuels has been researched thoroughly. The development of the GPR algorithm model is the main purpose of this research for the estimation of biodiesel density. This algorithm is better for the estimation of individual properties compared to other models because of the independence of this algorithm from the outliers. In the process of development of this model, the reliability and accuracy of the collected dataset are important so for the first time for the identification of suspected data points of biodiesel density, a throughout analysis has been carried out. On the other hand, the effects of input variables on the output have been researched statistically as an important part of this work.

2. Material and Methods

2.1. Gaussian Process Regression.

In recent years, the trend of neural networks, which is a branch of artificial intelligence, has made significant progress in solving problems related to the engineering field. The main disadvantage of this method is overfitting, which of course can be improved by adjusting the weight [34, 35].

Of course, setting these parameters is also complex and difficult, and to solve this problem, a conventional mathematical method called Bayesian network is used. It should be noted that this method is probabilistic and uses Bayesian interference to calculate the probability [36]. This network plots each variable graphically, and these variables are connected by an arc, and each variable shows its knowledge content as a distribution of probabilities. It should be noted that the potential specificity of BNs is of great importance for assessing uncertainty. High distribution leads to more uncertainty. One of the factors influencing the increase of complex prior distribution on functions in the Bayesian method to neural networks is the prior overweight [37].
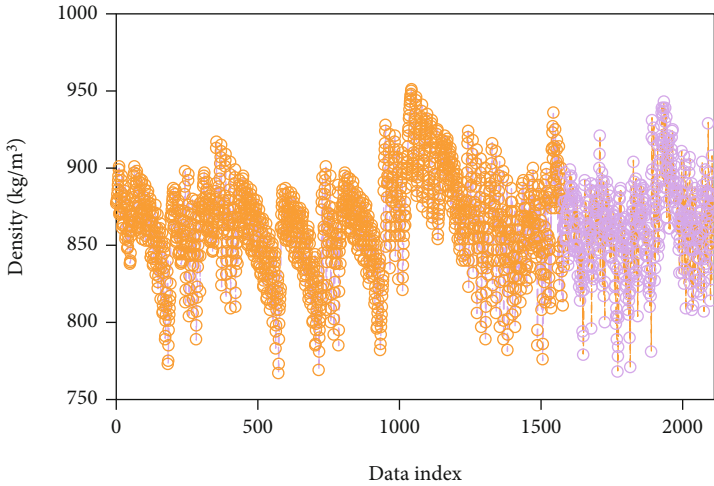
Gaussian process regression is a developed method for the abovementioned problem. It should be noted that this method is nonparametric. The advantages of the GPR
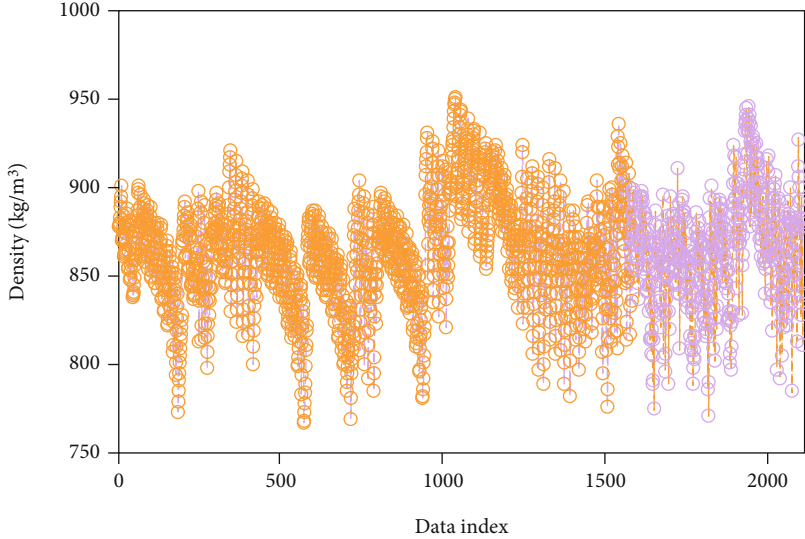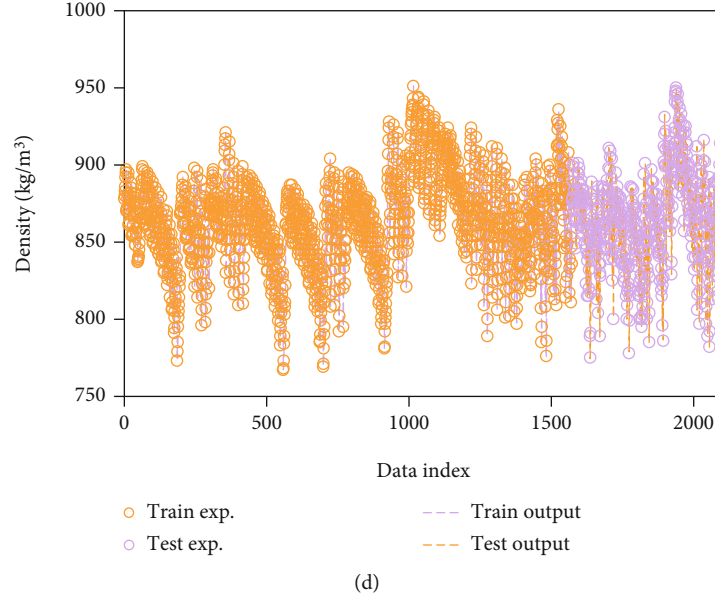
(a)
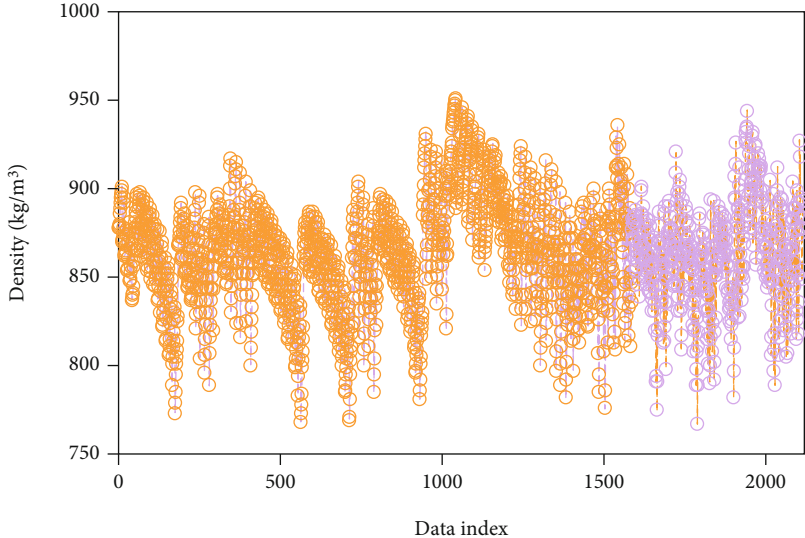
(b)

(c)

Figure 1: Continued.

(d)

FIGURE 1: The performance of the values predicted against their corresponding actual values using different kernel functions including (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic.

algorithm include measuring uncertainty for predictions and working well on small data. It is worth noting that the GPR method has important advantages over Bayesian, including simplicity, nonlinearity, easy generalization, and has several dimensions. The model parameters are determined using the sample training information in the GPR method.

The GP model is obtained by linking previous knowledge to the current process of modeling and integrating real laboratory data.

One of the salient and important differences between the old methods of machine learning and GPR is not finding the best approximation with experimental points and is a complete BN core. GPR works periodically by obtaining posterior distributions on models. In the following, we will explain how to create GP regression [38].

Randomly selected points $T = \{x_{L,i}, y_{L,i}\}$ and $L = \{x_{T,i}, y_{T,i}\}$, $i = 1, 2, 3, \cdots, n$, which are test and learning data from a particular distribution, are assumed as follows [39]:

$$T = \left\{x_{L,i}, y_{L,i}\right\}, L = \left\{x_{T,i}, y_{T,i}\right\}, i = 1, 2, \cdots, n. \quad (1)$$

As an important point, we remind that the model parameters are adjusted based on the learning data [40].

As the input and goal data, respectively, x and y have been assumed that noise has affected them.

The general formula of the GPR described as follows [41]:

$$y_{L,i} = f(x_{L,i}) + e_{L,i}, n = 1, 2, 3, \cdots, n. \quad (2)$$

In the abovementioned equation, X_L and Y_L denote independent and objective variables of the training data, and subsequently, the $\varepsilon \sim N(0, \sigma_{noise}^2 l_n)$ denotes for the observation noise with the independent Gaussian distribu-

tion that I_{noise} symbolizes and σ_{noise}^2 the variance of noise and unit array [42].

Note: GP assumes the output $f(x)$ is random. So the following equation is obtained:

$$y_{T,i} = f(x_{T,i}) + \varepsilon_{T,i}, i = 1, 2, 3, \cdots, n. \quad (3)$$

In the above equation, y_T and x_T represent the goals and independent variables and the $f(x)$ represents a Gaussian process with covariance function $k(x, x')$ and mean function $m(x)$.

$$f(x_{L,i}) \sim GP\left(m(x), k\left(x.x'\right)\right). \quad (4)$$

But in practice, the exact determination of $m(x)$ can be complicated, so the value of $m(x)$ is taken to be zero to make the calculations easier, so we have the following [43]:
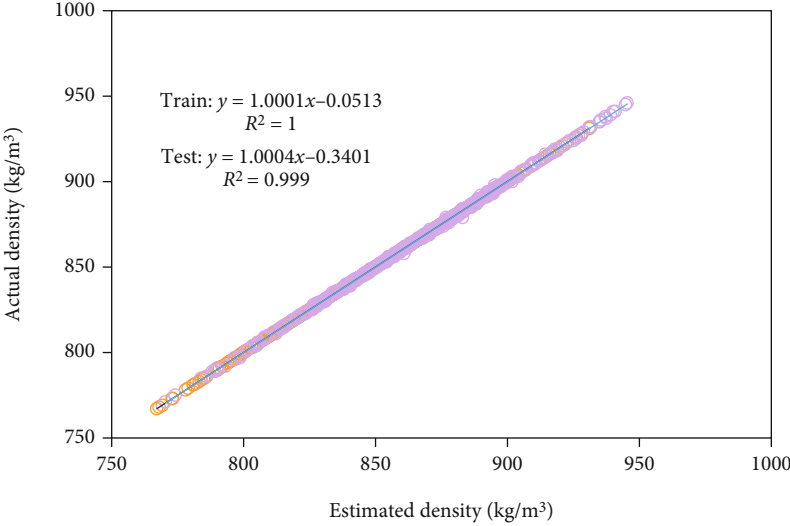
$$f(x_{L,i}) \sim GP\left(0, k\left(x.x'\right)\right). \quad (5)$$

From Eqs. (2) and (5), we can conclude the following equation:
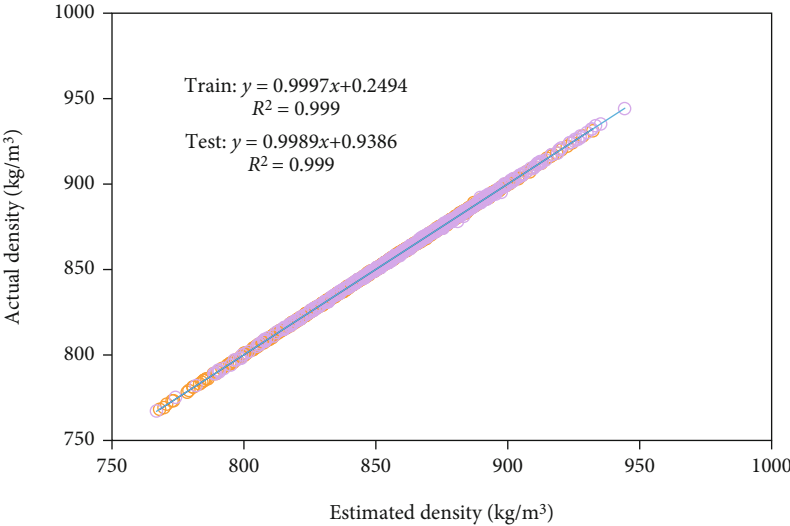
$$y \sim N\left(0, k\left(x.x'\right) + \sigma_{noise}^2 I_n\right). \quad (6)$$

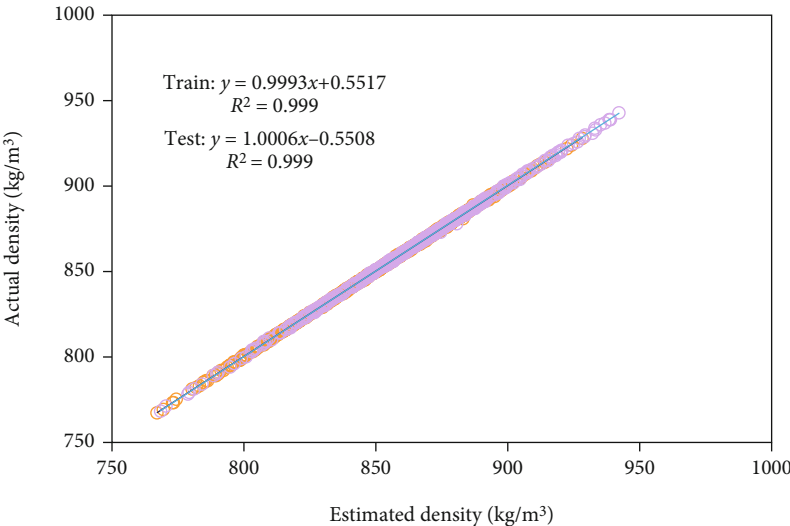A better representation of the variables mentioned in the text above can be provided as follows [44]:

$$\begin{bmatrix} \overrightarrow{f_L} \\ \overrightarrow{f_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} k(x_L.x_L) & k(x_L.x_T) \\ k(x_T.x_L) & k(x_T.x_T) \end{bmatrix}\right), \quad (7)$$

Train: $y = 1.0001x - 0.0513$
$R^2 = 1$

Test: $y = 1.0004x - 0.3401$
$R^2 = 0.999$

(a)

Train: $y = 0.9997x + 0.2494$
$R^2 = 0.999$

Test: $y = 0.9989x + 0.9386$
$R^2 = 0.999$

(b)

Train: $y = 0.9993x + 0.5517$
$R^2 = 0.999$

Test: $y = 1.0006x - 0.5508$
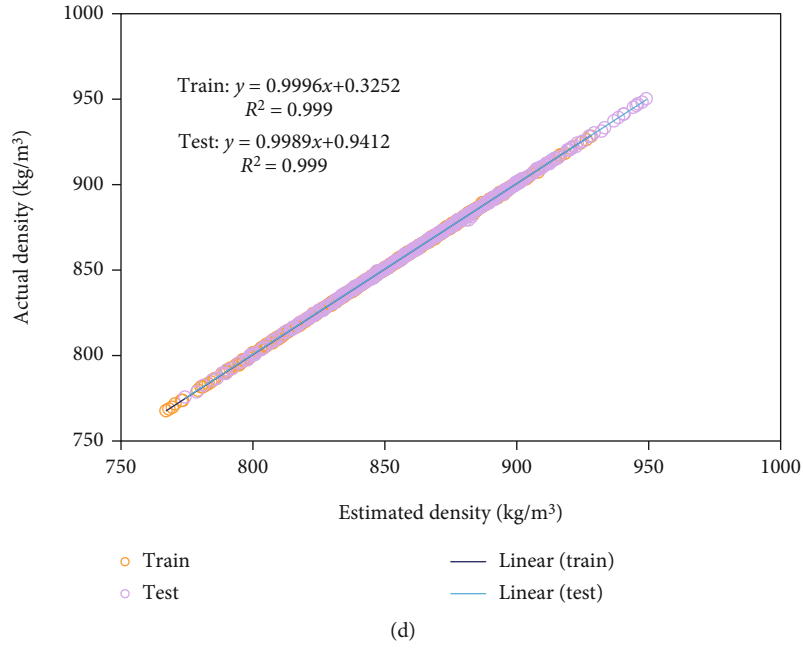$R^2 = 0.999$

(c)

FIGURE 2: Continued.

(d)

FIGURE 2: Cross plots of predicted output values for different kernel functions including (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic.

$$\begin{bmatrix} \overrightarrow{\varepsilon_L} \\ \overrightarrow{\varepsilon_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{\text{noise}}^2 I_n & 0 \\ 0 & \sigma_{\text{noise}}^2 I_n \end{bmatrix}\right). \tag{8}$$

The following Gaussian function is obtained by adding the sum of Eqs. (7) and (8):

$$\begin{bmatrix} \overrightarrow{y_L} \\ \overrightarrow{y_T} \end{bmatrix} \sim N\left(0, \begin{bmatrix} k(x_L.x_L) + \sigma_{\text{noise}}^2 I_n & k(x_L.x_T) \\ k(x_T.x_L) & k(x_T.x_T) + \sigma_{\text{noise}}^2 I_n \end{bmatrix}\right). \tag{9}$$

Therefore, the previous distribution of Y_T is obtained from Gaussian conditions as below:

$$\left(\overrightarrow{y_T} \mid \overrightarrow{y_L}\right) \sim N\left(\mu_T, \Sigma_T + \sigma_{\text{noise}}^2 I_n\right). \tag{10}$$

And values Σ_T and μ_T are assumed as follows:

$$\Sigma_T = k(x_T.x_T) = k(x_T.x_T) + \sigma_{\text{noise}}^2 I_n \\ - k(x_T.x_L) \cdot \left(k\left(x_L.x_L + \sigma_{\text{noise}}^2 I_n\right)^{-1} k(x_L.x_T)\right),$$

$$\mu_T = m\left(\overrightarrow{y_T}\right) = k(x_T.x_L) \cdot \left(k\left(x_L.x_L + \sigma_{\text{noise}}^2 I_n\right)^{-1} \overrightarrow{y_T}. \tag{11}$$

In GPR modeling, the following theoretical concept is obtained by predicting the output of experimental data through independent variables and training data. From the above equations, it can be concluded that the covariance and the mean function of both together with the Gaussian distribution represent a GP.

To better predict the goals of the developed GPR model, the selection of the core function in the training phase is of great importance. Therefore, in this research, to find the best kernel function, we use four different and conventional kernel functions, of course, with changes.

These functions are described as follows [45]:

(i) Rational quadratic covariance function

$$k_{\text{RQ}}\left(x.x'\right) = \sigma^2 \left(1 + \frac{x - x'2}{2al^2}\right)^{-a}. \tag{12}$$

In rational quadratic covariance function equation $\sigma^2, 1,$ $a > 0$ represents the variance, length, and weight scale changes.
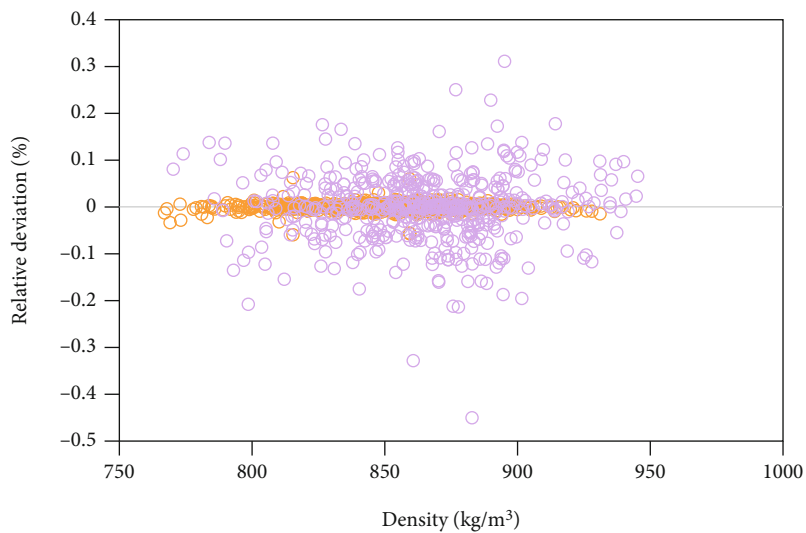
(ii) Squared exponential covariance function

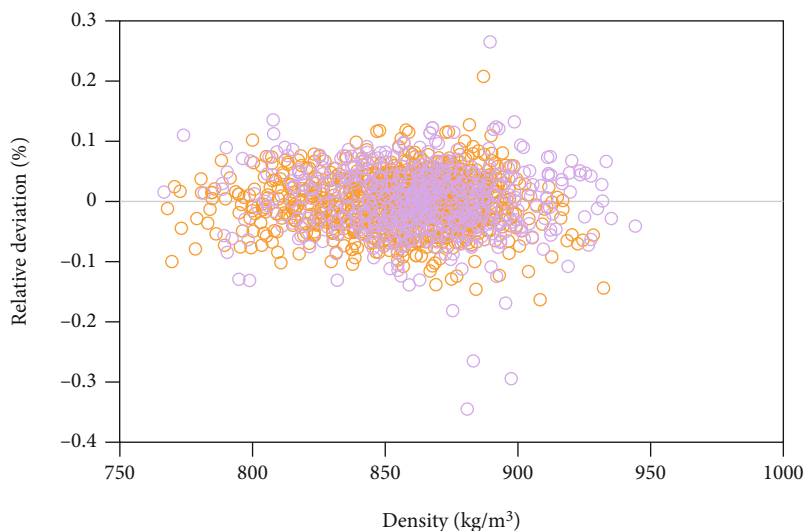$$k_{\text{SE}}\left(x.x'\right) = \sigma^2 \left(-\frac{x - x'2}{l^2}\right). \tag{13}$$

(iii) Exponential covariance function

$$k_E\left(x.x'\right) = \sigma^2 \exp\left(-\frac{x - x'}{l}\right). \tag{14}$$
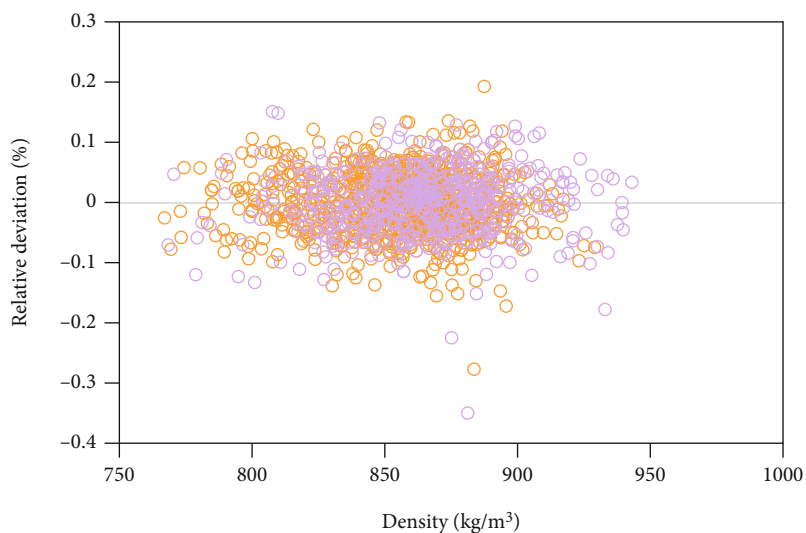
(iv) Matern covariance function

(a)
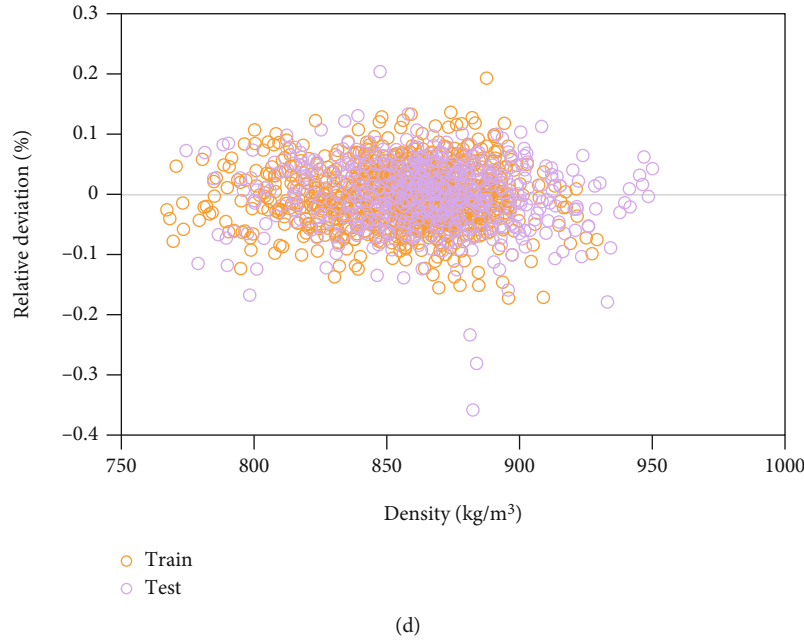

(b)

(c)

FIGURE 3: Continued.

(d)

FIGURE 3: The relative deviation between the real density and GPR outputs using different kernel functions including (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic.

$$k_M\left(x.x'\right) = \sigma^2 \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2v}\frac{x-x'}{l}\right)^v K_v\left(\sqrt{2v}\frac{x-x'}{l}\right).$$

(15)

The symbol represents the gamma function and x, y are positive parameters, and the adjusted K_v is the Bessel function [46].

In the Matern equation, exponential covariance and quadratic are important functions. The Matern function is an exponential when the value of $V = 0.5$ and also when the exponential is a square when V is inclined to infinity. Since the Matern equation has a greater degree of freedom, it also performs better than the other two [47].

Since the GPR method is nonparametric, the learning stage tries a lot to modify the parameters of the above equations.

2.2. Data Collection. The dataset containing 2117 real density points has been collected from different resources. References to this data have been reported elsewhere [48]. These points are in the pressure range of 0.1-129.78 MPa, melting point of 238.15-304.15 K, molecular weight of 186.291-310.514 g/mol, and temperature range of 278.36-413.15 K. The density values are different between 769.4 and 951.3 kg/m^3 in terms of these conditions.

3. Results and Discussion

As mentioned before, here, four various GPR algorithms including kernel functions in Matern, rational quadratic, exponential, and square exponential forms are used for estimating the density of biodiesels. To evaluate the precision

of these algorithms, statistical analysis of parameters is determined as follows:

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n} \frac{\left|y_{\exp.,i} - y_{\text{pred}.,i}\right|}{y_{\text{pred}.,i}},$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} \left(y_{\exp.i} - y_{\text{pred}.i}\right)^2,$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_{\exp.i} - y_{\text{pred}.i}\right)^2},$$

$$\text{STD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left(\frac{y_{\exp.i} - y_{\text{pred}.i}}{y_{\exp.i}}\right)^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{\text{pred}.i} - y_{\exp.i}\right)^2}{\sum_{i=1}^{n} \left(y_{\text{pred}.i} - \overline{y_{\exp.i}}\right)^2}.$$

(16)

As you see in Table 1, R^2 values of rational quadratic, Matern, exponential, and square exponential forms are equal to 1. According to other statistical parameters, the rational quadratic form depicts a better performance than the other kernel functions. In this kernel function form, RMSE, MSE, MRE, and STD are obtained 0.47, 0.22, 0.04, and 0.30, respectively. These values exhibit the rational quadratic formability in the forecast of density values.

Comparing the results of R^2 from this table with the results published by Abooali et al. who used SGB and GP models to predict biodiesel density, it was concluded that the models presented by us have a higher ability to
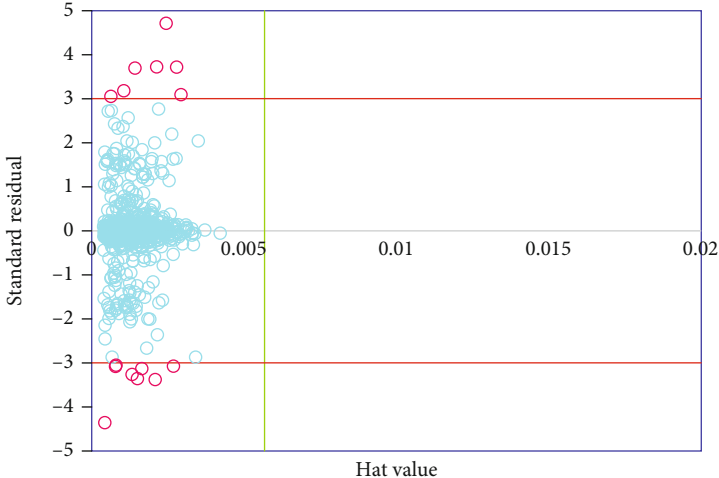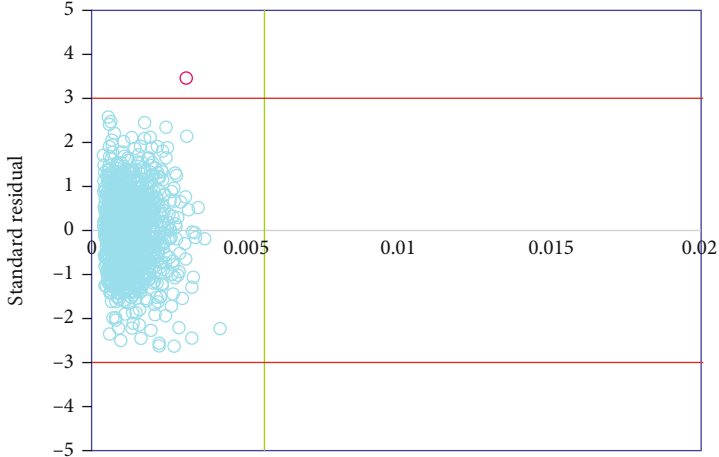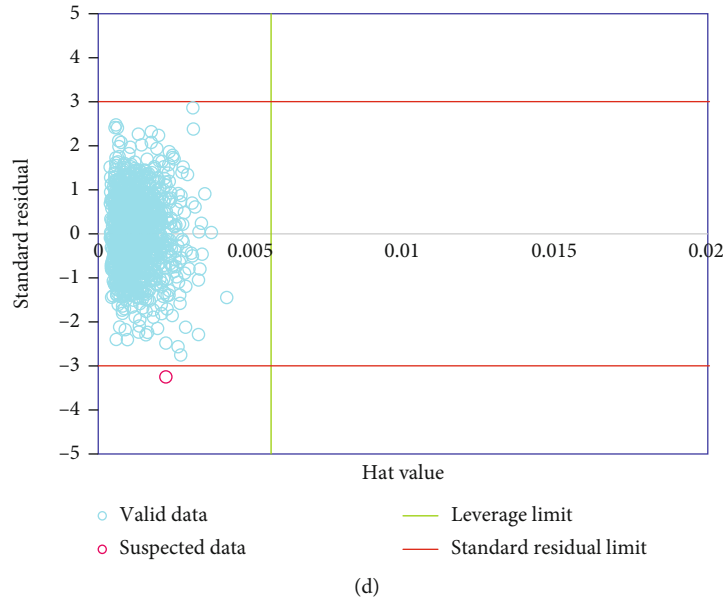
(a)
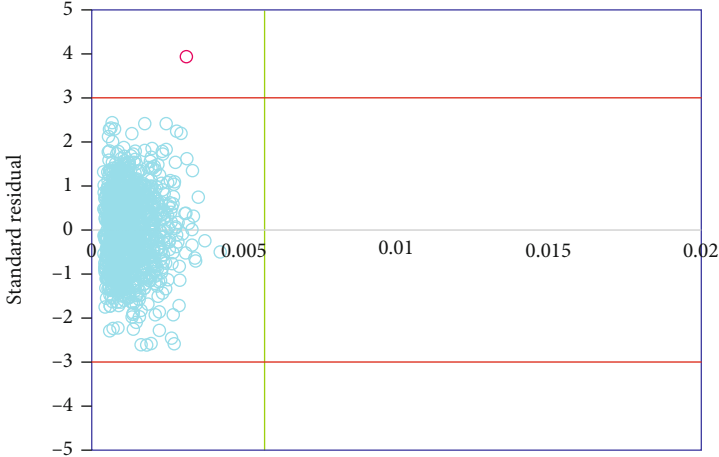
(b)

(c)

Figure 4: Continued.

(d)

FIGURE 4: Determination of suspected data point for the various kernel functions including (a) exponential, (b) Matern, (c) squared exponential, and (d) rational quadratic.
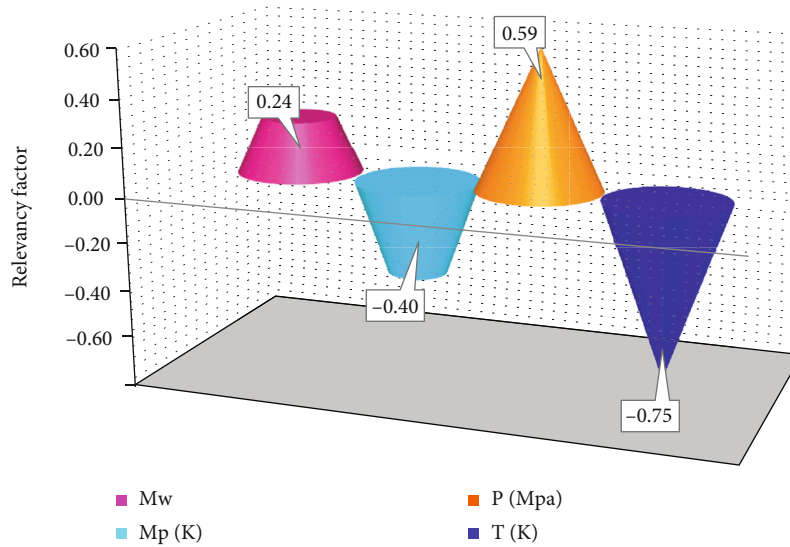

FIGURE 5: Sensitivity analysis on various input parameters.

predict the target values because the R^2 values for these two models were obtained 0.99988 and 0.99635, respectively [48].

To make a better decision about all these models, Figure 1 shows the experimental and estimated density values, simultaneously. In this figure, there is a good agreement between the real density values and GPR outputs.

Also, the cross/regression plot of predicted and real density has been shown in Figure 2.

To express the quality of GPR outputs, we refer to the density data located on bisector lines in the analysis. Moreover, Figure 3 shows the relative deviation between the real density and GPR outputs.

The accuracy of these density point data affects the validity of models. In this examination, too many data points have been used. It is important to know that these data may have errors due to measurements done in laboratories. So, these types of data are separated from the other data points. In this regard, some strongly developed strategies are required to remove these data and enhance the model accuracy. Here, the separation of these suspected data is accomplished by the Leverage method. In this method, after the determination of residual values, a Hat matrix is created as follows [49]:

$$H = A\left(A^T A\right)^{-1} A^T, \qquad (17)$$

where A is an $i \times j$ dimensional matrix and i and j are defined for the model parameters and training points, respectively. Then, the critical leverage limit is computed by i and j as follows:

$$H^* = \frac{3(b+1)}{a}. \tag{18}$$

According to William's plot, as shown in Figure 4, we can separate the suspected data obtained as residuals from Hat values. Here, the leverage limits in the green line and two red lines are considered as residuals. So, the data points located outside these lines are considered as suspected data. As you see, for exponential, Matern, squared exponential, and rational quadratic forms, 14, 1, 1, and 1 points are considered as the suspected data, respectively, among 2217 points.

In this study, the suggested GPR algorithms create a relationship between density and inputs. So, sensitivity analysis is utilized to show this relationship affects the output. In this regard, to determine the most efficient variable for density, the relevancy factor, r, in the range of -1 and 1, is used. If the absolute value of r is large, it can affect the density further. The less and more relation to density is shown by negative and positive r values, respectively. The r is calculated as follows [50]:

$$r = \frac{\sum_{i=1}^{n}(x_{K,i} - \bar{x}_k)(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_{K,i} - \bar{x}_k)^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{19}$$

where $X_{k,i}$ and Y_i are inputs and outputs, respectively. The $\overline{X_k}$ and $\bar{Y}$ are the average of inputs and that of outputs, respectively. Figure 5 shows that the higher temperature and melting point lead to less density. Also, the temperature is considered the most efficient parameter for density. In addition, parameters such as pressure and molecular weight have a direct relationship with this target.

4. Conclusion

In this work, four various kernel functions including Matern, rational quadratic, exponential, and square exponential functions have been used for GPR algorithms to compute the density of biodiesels. To prepare and validate these algorithms, a large database containing 2217 actual data is gathered. It is concluded that the proposed models have highly precise to predict real data. The rational quadratic GPR model has shown greater performance compared with other models. In this model, the calculations show that RMSE = 0.47, MSE = 0.22, MRE = 0.04, R^2 = 1, and STD is equal to 0.3. Other analyses also confirmed the accuracy of this model, which indicates that this attractive and simple model can be used in biodiesel-related industries.

Data Availability

The data are stated in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. D. Nguyen, R. Daneshfar, A. H. S. Dehaghani, and C. H. Su, "The effect of shear rate on aggregation and breakage of asphaltenes flocs: experimental study and model-based analysis," *Journal of Molecular Liquids*, vol. 325, p. 114861, 2021.

[2] M. Ghadiri, A. Marjani, R. Daneshfar, and S. Shirazian, "Model order reduction of a reservoir simulation by SOD-DEIM," *Journal of Petroleum Science and Engineering*, vol. 200, p. 108137, 2021.

[3] 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.

[4] 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.

[5] S. A. Montzka, E. J. Dlugokencky, and J. H. Butler, "Non-CO$_2$ greenhouse gases and climate change," *Nature*, vol. 476, no. 7358, pp. 43–50, 2011.

[6] L. Cao, G. Bala, K. Caldeira, R. Nemani, and G. Ban-Weiss, "Importance of carbon dioxide physiological forcing to future climate change," *Proceedings of the National Academy of Sciences*, vol. 107, no. 21, pp. 9513–9518, 2010.

[7] T. Sandler, "Environmental cooperation: contrasting international environmental agreements," *Oxford Economic Papers*, vol. 69, no. 2, pp. 345–364, 2017.

[8] S. Jafarmadar and P. Nemati, "Multidimensional modeling of the effect of exhaust gas recirculation on exergy terms in a homogenous charge compression ignition engine fueled by diesel/biodiesel," *Journal of Cleaner Production*, vol. 161, pp. 720–734, 2017.

[9] A. Espinoza, S. Bautista, P. C. Narváez, M. Alfaro, and M. Camargo, "Sustainability assessment to support governmental biodiesel policy in Colombia: a system dynamics model," *Journal of Cleaner Production*, vol. 141, pp. 1145–1163, 2017.

[10] A. Demirbas and S. Karslioglu, "Biodiesel production facilities from vegetable oils and animal fats," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 29, no. 2, pp. 133–141, 2007.

[11] J. J. Torres, N. E. Rodriguez, J. T. Arana, N. A. Ochoa, J. Marchese, and C. Pagliero, "Ultrafiltration polymeric membranes for the purification of biodiesel from ethanol," *Journal of Cleaner Production*, vol. 141, pp. 641–647, 2017.

[12] A. Datta and B. K. Mandal, "Use of Jatropha biodiesel as a future sustainable fuel," *Energy Technology & Policy*, vol. 1, no. 1, pp. 8–14, 2014.

[13] G. Knothe and K. R. Steidley, "Lubricity of components of biodiesel and petrodiesel. The origin of biodiesel lubricity," *Energy & Fuels*, vol. 19, no. 3, pp. 1192–1200, 2005.

[14] L. F. R. Verduzco, "Density and viscosity of biodiesel as a function of temperature: empirical models," *Renewable and Sustainable Energy Reviews*, vol. 19, pp. 652–665, 2013.

[15] E. Sann, M. Anwar, R. Adnan, and M. A. Idris, "Biodiesel for gas turbine application-an atomization characteristics study," *Advances in Internal Combustion Engines and Fuel Technologies*, pp. 213–242, IntechOpen, 2013.

[16] Z. J. West, T. Yamada, C. R. Bruening et al., "Investigation of water interactions with petroleum-derived and synthetic aviation turbine fuels," *Energy & Fuels*, vol. 32, no. 2, pp. 1166–1178, 2018.

[17] M. J. Pratas, S. V. D. Freitas, M. B. Oliveira, S̷. C. Monteiro, A̷. S. Lima, and J̷. A. P. Coutinho, "Biodiesel density: experimental measurements and prediction models," *Energy & Fuels*, vol. 25, no. 5, pp. 2333–2340, 2011.

[18] S. Phankosol, K. Sudaprasert, S. Lilitchan, K. Aryusuk, and K. Krisnangkura, "Estimation of density of biodiesel," *Energy & Fuels*, vol. 28, no. 7, pp. 4633–4641, 2014.

[19] A. T. Hoang, "Prediction of the density and viscosity of biodiesel and the influence of biodiesel properties on a diesel engine fuel supply system," *Journal of Marine Engineering & Technology*, vol. 17, pp. 1–13, 2018.

[20] A. Queimada, L. Rolo, A. Caco, I. Marrucho, E. Stenby, and J. Coutinho, "Prediction of viscosities and surface tensions of fuels using a new corresponding states model," *Fuel*, vol. 85, no. 5-6, pp. 874–877, 2006.

[21] A. Barati-Harooni, A. Soleymanzadeh, A. Tatar et al., "Experimental and modeling studies on the effects of temperature, pressure and brine salinity on interfacial tension in live oil-brine systems," *Journal of Molecular Liquids*, vol. 219, pp. 985–993, 2016.

[22] A. Rostami, H. Ebadi, M. Arabloo, M. K. Meybodi, and A. Bahadori, "Toward genetic programming (GP) approach for estimation of hydrocarbon/water interfacial tension," *Journal of Molecular Liquids*, vol. 230, pp. 175–189, 2017.

[23] A. A. Gakh, E. G. Gakh, B. G. Sumpter, and D. W. Noid, "Neural network-graph theory approach to the prediction of the physical properties of organic compounds," *Journal of Chemical Information and Computer Sciences*, vol. 34, no. 4, pp. 832–839, 1994.

[24] 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.

[25] M. Mostafaei, "Prediction of biodiesel fuel properties from its fatty acids composition using ANFIS approach," *Fuel*, vol. 229, pp. 227–234, 2018.

[26] D. Li, M. Guo, X. Wang, S. Lin, W. Jia, and G. Wang, "Measurement and correlation of density and viscosity of binary mixtures of fatty acid (methyl esters+ methylcyclohexane)," *The Journal of Chemical Thermodynamics*, vol. 137, pp. 86–93, 2019.

[27] R. Aitbelale, Y. Chhiti, F. E. M.'. Alaoui, A. Sahib Eddine, N. Muñoz Rujas, and F. Aguilar, "High-pressure soybean oil biodiesel density: experimental measurements, correlation by Tait equation, and perturbed chain SAFT (PC-SAFT) model-ing," *Journal of Chemical & Engineering Data*, vol. 64, no. 9, pp. 3994–4004, 2019.

[28] 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.

[29] 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.

[30] R. Daneshfar, F. Keivanimehr, M. Mohammadi-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. 706–712, 2020.

[31] R. Setiawan, R. Daneshfar, O. Rezvanjou, S. Ashoori, and M. Naseri, "Surface tension of binary mixtures containing environmentally friendly ionic liquids: insights from artificial intelligence," *Environment, Development and Sustainability*, pp. 1–22, 2021.

[32] 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 108708, 2021.

[33] S. M. Alizadeh, I. Alruyemi, 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. 1–14, 2021.

[34] C. K. Williams, *Prediction with Gaussian processes: from linear regression to linear prediction and beyond, in Learning in graphical models*, Springer, 1998.

[35] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecological Modelling*, vol. 203, no. 3-4, pp. 312–318, 2007.

[36] K. H. Reckhow, "Water quality prediction and probability network models," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 56, no. 7, pp. 1150–1158, 1999.

[37] S. Asante-Okyere, C. Shen, Y. Yevenyo Ziggah, M. Moses Rulegeya, and X. Zhu, "Investigating the predictive performance of Gaussian process regression in evaluating reservoir porosity and permeability," *Energies*, vol. 11, no. 12, p. 3261, 2018.

[38] S. Sniekers and A. van der Vaart, "Adaptive Bayesian credible sets in regression with a Gaussian process prior," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 2475–2527, 2015.

[39] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *The Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.

[40] Q. Fu, W. Shen, X. Wei, P. Zheng, H. Xin, and C. Zhao, "Prediction of the diet nutrients digestibility of dairy cows using Gaussian process regression," *Information Processing in Agriculture*, vol. 6, no. 3, pp. 396–406, 2019.

[41] M. Zhu, S. Liu, and S. Gu, "Short-term tide level forecasting based on Gaussian process regression," in *The Eleventh ISOPE Pacific/Asia Offshore Mechanics Symposium*, Shanghai, China, 2014.

[42] Y. Zhang, G. Su, and L. Yan, "Gaussian process machine learning model for forecasting of karstic collapse," in *Communications in Computer and Information Science*, Springer, 2011.

[43] Y. Zhang and X. Xu, "Predicting doped MgB_2 superconductor critical temperature from lattice parameters using Gaussian process regression," *Physica C: Superconductivity and Its Applications*, vol. 573, p. 1353633, 2020.

[44] A. S. Alghamdi, K. Polat, A. Alghoson, A. A. Alshdadi, and A. A. Abd el-Latif, "Gaussian process regression (GPR) based non-invasive continuous blood pressure prediction method from cuff oscillometric signals," *Applied Acoustics*, vol. 164, p. 107256, 2020.

[45] S. A. Aye and P. Heyns, "An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission," *Mechanical Systems and Signal Processing*, vol. 84, pp. 485–498, 2017.

[46] W. Gao, M. Karbasi, M. Hasanipanah, X. Zhang, and J. Guo, "Developing GPR model for forecasting the rock fragmentation in surface mines," *Engineering with Computers*, vol. 34, no. 2, pp. 339–345, 2018.

[47] S. Hong, Z. Zhou, C. Lu, B. Wang, and T. Zhao, "Bearing remaining life prediction using Gaussian process regression with composite kernel functions," *Journal of Vibroengineering*, vol. 17, no. 2, pp. 695–704, 2015.

[48] D. Abooali, R. Soleimani, and S. Gholamreza-Ravi, "Characterization of physico-chemical properties of biodiesel components using smart data mining approaches," *Fuel*, vol. 266, p. 117075, 2020.

[49] 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*, 2021.

[50] D. Ahangari, R. Daneshfar, M. Zakeri, S. Ashoori, and B. S. Soulgani, "On the prediction of geochemical parameters (TOC, S1 and S2) by considering well log parameters using ANFIS and LSSVM strategies," *Petroleum*, 2021.