Research Article

Developing a Novel Method for Estimating the Speed of Sound in Biodiesel Known as Grey Wolf Optimizer Support Vector Machine Algorithm

Zhenzhen Lv, Ming Hu, Yixin Yang, and Jeren Makhdoumi

1School of Electrical and Information Engineering, Anhui University of Technology, Maanshan, Anhui 243002, China
2Department of Educational Science, Payame Noor University, Damghan, Iran

Correspondence should be addressed to Ming Hu; xiaoming-sudo@outlook.com and Jeren Makhdoumi; jerenmkhd@student.pnu.ac.ir

Received 4 June 2021; Revised 13 June 2021; Accepted 25 June 2021; Published 2 July 2021

Academic Editor: Alireza Baghban

Copyright © 2021 Zhenzhen Lv 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.

In the current study, our goal was to obtain a robust model to predict the speed of sound in biodiesel. For this purpose, an extensive databank has been extracted from previously published papers. Then, a Support Vector Machine (SVM) has been optimized by Grey Wolf Optimization (GWO) method to analyze these data and determine the correlation between speed of sound in biodiesel and its related properties including pressure, temperature, molecular weight, and normal melting point. The results were very satisfactory because the values of statistical parameters $R^2$ and RMSE were obtained 1 and 1.4024, respectively. Here, this is the first time that the sensitivity analysis is used to estimate this target value. This analysis shows that the pressure widely affects the output values with relevancy factor 87.92. Also, our proposed method is highly accurate than other machine learning methods used in papers employed for this objective.

1. Introduction

In the future, the use of petroleum and fossil fuels will be limited [1]. A large number of studies have recently investigated biodiesel utilization within engines [2]. The need for oil imports would be reduced by using animal or agricultural sources to produce methyl esters. This can improve energy security and the local economy and lead to a satisfactory carbon emission balance [3]. Moreover, biodiesel combustion within a diesel engine typically emits smaller quantities of carbon [3–5].

Biodiesel can be extracted from several chemical compositions of feedstock. Due to the dependence of fat/oil structures in fatty acids on the source of oil/fat, biodiesel highly varies in physicochemical properties, including the cold-flow properties and cetane number [3]. Hence, the biodiesel type is considerably important in combustion and emissions [6].

It is important to identify better fatty acid compositions in order to enhance engine performance and diminish emissions. This has been studied by numerous works [7–11]. It is rational to relate the properties of biodiesel to some important oil characteristics, e.g., fatty acid composition, chain length, number of double bonds, unsaturation degree, and molecular weight [12–16]. Earlier works related the fatty acid composition and cetane number through regression models [12]. The use of different methods of artificial intelligence has been widely used in various sciences [17–21], and the cetane number was studied using artificial neural networks (ANNs) and multiple linear regression models [13]. Furthermore, temperature and density were associated in previous studies [22]. Some researchers related the cetane number, viscosity, density, and increased heating value to the number of double bonds and molecular weight [23]. The cetane number, oxidative stability, cold filter plugging point, and iodine value were related to the long-chain saturated factor and methyl ester unsaturation degree [24]. Some studies related the number of double bonds and the number of C atoms in the fatty acid [25].
Biodiesel utilization in engines was broadly investigated [26] under transient conditions [27]. Furthermore, broad examinations were performed statistically to find and study the impacts of biodiesel feedstock on emissions of engines [28] and fuel properties [29].

As an important and practical property, the present study focuses on the prediction of speed of sound in biodiesel in order to develop an accurate predictive correlation formulation based on the fatty acid composition through a multiple linear regression known as SVM-GWO. This very important property of biodiesel has received less attention from researchers, so we were looking for an accurate model to be able to estimate this functional property with high accuracy. In this paper, an extensive database has been used and an attempt has been made to evaluate the accuracy of this model using various analyses.

2. Materials and Methods

2.1. Support Vector Machine (SVM). The SVM is one of the machine learning (ML) methods. Rather than other techniques, this technique works based on a minimum of structural risk expressed by statistical theory [30]. This technique, for the first time, is proposed by Vapnik in 1992 for classification problems [31]. Afterward, it was developed by Cortes and Vapnik, in 1995 and 1997, for regression problem adaptation [32, 33]. SVM can be employed for both linear and nonlinear problems, but for nonlinear problems, it must be improved by kernel functions. The SVM equations are given in the following [34]. In Equation (1), a sample dataset is used for training using the SVM regression method, where \( y_i \), \( x_i \), \( d \), and \( R \) are output, input, input space dimension, and output space, respectively.

\[
\begin{cases}
  x_i, y_i | x \in R^d, & y \in R, i = 1, 2, \ldots, n \end{cases}
\]  

(1)

In Equation (2), the input data is mapped from \( R^d \) space to a high-dimension one, \( R^k (k > d) \).

\[
\psi(x) = (\phi(x_1), \phi(x_2), \ldots, \phi(x_n)).
\]  

(2)

Equation (3) introduces the prediction model for SVM as follows:

\[
f(x) = \omega^T \phi(x) + b, \quad \omega \in R^k, \ldots, \phi(x_n),
\]  

(3)

where \( b, \omega, \) and \( f(x) \) are bias constant, weight, and a nonlinear mapping function, respectively, and \( \omega \) and \( b \) are defined by Equation (4) with minimal structural risk.

\[
\min R = \frac{1}{2} ||\omega||^2 + c \times R_{\text{emp}},
\]  

(4)

where \( ||\omega||^2 \) is used to handle the difficulty of the model, \( c \) is the regularization coefficient, and \( R_{\text{emp}} \) is a function for handling errors. Also, for optimizing the objective function, \( R_{\text{emp}} \) is defined as the linear term of the error of SVM. So, Equation (4) can be changed into Equation (5) using the relaxation factors, \( \xi_i \), and \( \xi^*_i \), and insensitivity loss function, \( \epsilon \):

\[
\min \ J = \left( \frac{1}{2} ||\omega||^2 + c \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \xi_i^* \right) \left\{ \begin{array}{l} y_i - \omega^T \phi(x_i) - b \leq \epsilon + \xi_i, \vspace{0.2cm} \\
\omega^T \phi(x_i) + b - y_i \leq \epsilon + \xi_i, \newline
\xi_i, \xi_i^* \geq 0 (i = 1, 2, \ldots, n). \end{array} \right\}
\]  

(5)

Also, the Lagrange function is given by Equation (6) to solve the SVM error.

\[
L(\omega, \xi_i, \xi^*_i, \alpha, \alpha^*, \epsilon, \beta, \beta^*) = \frac{1}{2} ||\omega||^2 + c \sum_{i=1}^{n} (\xi_i + \xi^*_i) - \sum_{i=1}^{n} \alpha_i [\omega^T \phi(x_i) + b - y_i + \epsilon + \xi_i] - \sum_{i=1}^{n} \alpha^*_i [\omega^T \phi(x_i) - b + \epsilon + \xi^*_i] - \sum_{i=1}^{n} (\beta_i \xi_i + \beta^*_i \xi^*_i), \alpha_i, \alpha^*_i, \beta_i, \beta^*_i > 0,
\]  

(6)

where \( \alpha_i, \alpha^*_i, \beta_i, \beta^*_i, c \) are defined as Lagrange factors.

According to Karush-Kuhn-Tucker optimization conditions (Equation (7)) and symmetric kernel function (Equation (8)), the optimization problem can be obtained as Equation (9).

\[
\frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) \phi(x_i),
\]  

(7)

\[
\frac{\partial L}{\partial b} = 0 \rightarrow \omega \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) = 0,
\]  

\[
\frac{\partial L}{\partial \xi_i} = 0 \rightarrow c - \alpha_i - \beta_i = 0,
\]  

\[
\frac{\partial L}{\partial \xi^*_i} = 0 \rightarrow c - \alpha^*_i - \beta^*_i = 0,
\]  

\[
K(x_i, x_j) = \phi(x_i)^T \phi(x_j),
\]  

(8)

\[
\text{max } W(\alpha_i, \alpha^*_i) = -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha^*_i) (\alpha_j - \alpha^*_j) K(x_i, x_j) + \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) y_i - \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) \epsilon
\]  

s.t. \( \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) = 0 \) \( 0 \leq \alpha_i, \alpha^*_i \leq c. \)

(9)

So, the SVM regression function is given by

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) K(x_i, x) + b.
\]  

(11)
where \( r \) iteration whose calculations are given by the distance between the prey and grey wolf, the position vector \( \sigma \) (Equation (12)), where \( \alpha \) is representative of the width parameter of this function.

\[
K(x_i, x_j) = \exp \left( -\frac{||x_i - x_j||^2}{2\sigma^2} \right).
\] (12)

2.2. Grey Wolf Optimization (GWO). The GWO algorithm was introduced by Mirjalili et al. in 2014 as a novel metaheuristic algorithm inspired by the social hunting of grey wolves [35]. Generally, this algorithm follows four classes including (1) decision-making is performed by alpha (\( \alpha \)) wolves about everything, (2) the alpha wolves are supported/consulted by beta (\( \beta \)) wolves, (3) the delta (\( \delta \)) wolves must surrender to \( \alpha \) and \( \beta \) wolves, and finally in (4) the other wolves are defined by omega (\( \omega \)), which have to follow \( \alpha \) and \( \beta \) orders. The wolves must help others whenever required [36, 37]. So, the hierarchy of power reduces from \( \alpha \) to \( \omega \). In four classes, a specific optimization issue is defined by solutions of the GWO algorithm. So, \( \alpha \), \( \beta \), and \( \omega \) are the best solutions in this algorithm, and others are considered as \( \omega \). With this definition, the algorithm is updated in every iteration. The process of the algorithm follows these rules for prey: searching, surrounding, chasing, and attacking. The surrounding is given as follows:

\[
X(t + 1) = X_p(t) - A \times D,
\] (13)

where \( A, D, X_p(t), X(t + 1), \) and \( t \) are the matrix coefficient, the distance between the prey and grey wolf, the position vector of each wolf, the next position of a grey wolf, and the current iteration whose calculations are given by

\[
D = |C \times X_p(t) - X(t)|,
\]
\[
A = 2ar_1 - a,
\]
\[
C = 2ar_2,
\] (14)

where \( r_1, r_2 \) are the random vectors from 0 to 1.

In the hypersphere form, the relocation around the prey is feasible with the help of these equations. So, the \( \omega \) wolves can update their positions as follows:

\[
X(t + 1) = X_1 + X_2 + X_3 \overline{3},
\] (15)

where \( X_1, X_2, \) and \( X_3 \) are defined as the following:

\[
X_1 = X_{\alpha}(t) - A_1 \times D_{\alpha},
\]
\[
X_2 = X_{\beta}(t) - A_2 \times D_{\beta},
\]
\[
X_3 = X_{\delta}(t) - A_3 \times D_{\delta},
\]
\[
D_{\alpha} = |C_1 \times X_{\alpha}(t) - X(t)|,
\]
\[
D_{\beta} = |C_2 \times X_{\beta}(t) - X(t)|,
\]
\[
D_{\delta} = |C_3 \times X_{\delta}(t) - X(t)|.
\] (16)

2.3. Designing the GWO-SVM Model. Concerning the previous discussion, \( C, \varepsilon, \) and \( \gamma \) are used to handle the SVM performance. So, the GWO can be optimized by these factors. Table 1 depicts the characteristics of the GWO-SVM algorithm.

2.4. Gathering Data and Selecting Features. In this study, the database containing 1048 data with various variables related to the test system of the speed of sound in biodiesel—i.e., temperature, melting point, pressure, and molecular weight, has been collected from previously published papers. The source and range of inputs and output data are given elsewhere [38]. Three-quarters of the data are selected as training phase data and one-quarter of them are randomly separated as testing phase data.

3. Results and Discussion

In this section, we evaluate the ability of the proposed model to predict the target parameter, which is followed by various analyses.

3.1. Sensitivity Analysis (SA). In terms of exploring the impact of input data on the output, SA is defined as a mathematical method and used to determine useful priorities after the recognition of methodological errors and vital regions [39]. There are two forms for SA including local and global. The assessment of an input effect on results, while others are constant, is performed by local SA whereas the global SA evaluates the effect which stemmed from inputs on the outcome whenever changed [40].

The impact of input parameters on the speed of sound has been shown in Figure 1 that the most effective one is related to pressure with the relevancy factor of 87.92%.

Also, the relative factors of temperature, melting point, and molecular weight, with scores of \(-29.56\%, -25.63\%, \) and 15.18%, are not so big.

3.2. Outlier Analysis. Another statistical method used in this study is outlier diagnosis. This method is considered a fundamental method applied to determine datasets with different behavior from all data [41, 42]. It uses leverage statistical technique to find the outliers having parameters such as
standardized residuals ($R$), critical leverage limit ($H^*$), and Hat indices ($H$) [43, 44]. $H$ and $H^*$ are defined as follows:

$$H = X \left( X' X \right)^{-1} X', \tag{17}$$

$$H^* = \frac{3n}{(p+1)},$$

where $X$ together with $t$ are the two-dimensional ($n \times k$) matrix and transpose matrix, respectively. Also, $p$ and $n$ are the numbers of input parameters and training points, respectively. Here, the likely Hat solutions include the main diagonal space of $H$. Also, Williams’ plot, defined by $R$ versus $H$, is used to determine the outlying candidates. Then, the feasible data region is introduced as a squared area to limit the warning leverage value on the horizontal and vertical axes and cut-off value, which is usually $\pm 3$, respectively. $R$ and $H$ are placed out of the valid area—i.e., $[-3, 3]$ and $[0, H^*]$—and classified as the outliers. Figure 2 depicts Williams’ plot of GWO-SVM outputs. It is observed that most of the data values are placed in the valid area and the rest of them,
including 16 points, have a higher value than $H^*$. So it is demonstrated that the useful GWO-SVM algorithm can detect the inherent relationships between the speed of sound value and input parameters in addition to having a much more acceptable approach.

3.3. Model Assessment. The model assessment is performed by the speed of sound values that resulted in training and testing of the proposed model. Figure 3 shows these values versus the data index. It is proved that this model has the considerable capability to predict the speed of sound in biodiesel.

Also, to assess the accuracy of results with real values, the determination coefficient, $R^2$, is used and varies from 0 to 1. The $R^2$ values for testing and training the GWO-SVM dataset are 1 and 1, respectively. Thus, the accuracy of the predicted model is verified. The diagram of real values versus predicted values is shown in Figure 4.

The main part of the speed of sound in biodiesel values situates along the bisector line which shows how the GWO-SVM model is able to do prediction with high accuracy. Also, Figure 5 depicts the percentage of deviation for the GWO-SVM model which is not more than 0.6% that demonstrates the precision of the model.
Also, Table 2 shows statistical analyses of the SVM-GWO model and verifies the accuracy of this model for predicting the speed of sound in biodiesel.

### Table 2: Statistical analyses based on the proposed SVM-GWO model.

| Set    | $R^2$ | MRE (%) | MSE        | RMSE       | STD        |
|--------|-------|---------|------------|------------|------------|
| Train  | 1.000 | 0.061   | 1.925445498| 1.3876     | 1.0321     |
| Test   | 1.000 | 0.061   | 1.966602271| 1.4024     | 1.0542     |
| Total  | 1.000 | 0.061   | 1.935734691| 1.4024     | 1.0371     |

### Table 3: Comparison between the accuracy of different models in predicting target outcomes.

| Statistical parameter | SGB model | GP model | SVM-GWO model |
|-----------------------|-----------|----------|---------------|
| $R^2$                 | 0.99996   | 0.99803  | 1.000         |
| RMSE                  | 1.55      | 8.81907  | 1.4024        |

3.4. **Comparison with Literature.** Table 3 shows the comparison done between previously developed models (SGB and GP) by Abooali et al. [38] and our model for predicting the speed of sound in biodiesel. As it turns out, the model proposed in this study has a higher ability to predict output values because it has more $R^2$ and less RMSE compared to other models.

### 4. Conclusions

In this study, the SVM-GWO model has been proposed to investigate the effect of structural features on the performance of the speed of sound in biodiesel. The database containing large experimental data has been collected from previously published papers. Comparing all of the ML models, our model showed the best accuracy. So it has great capability to assist in the objective design of the speed of sound in biodiesel. Furthermore, it was shown that the pressure has the highest impact on the output values. In conclusion, according to the obtained maximum value of the coefficient of determination and minimum RMSE, our model is considered the most precise model to predict the speed of sound in biodiesel; therefore, it can be used to estimate this important property in related processes.

### Data Availability

Data references are described in the text of the article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

[1] 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, p. 100202, 2019.

[2] 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.

[3] A. K. Agarwal, “Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines,” Progress in Energy and Combustion Science, vol. 33, no. 3, pp. 233–271, 2007.

[4] A. Demirbas, “Biodiesel production from vegetable oils via catalytic and non-catalytic supercritical methanol transesterification methods,” Progress in Energy and Combustion Science, vol. 31, no. 5-6, pp. 466–487, 2005.
and signal processing,” *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.

[34] S. Dai, D. Niu, and Y. Han, “Forecasting of power grid investment in China based on support vector machine optimized by differential evolution algorithm and grey wolf optimization algorithm,” *Applied Sciences*, vol. 8, no. 4, p. 636, 2018.

[35] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[36] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. S. Coelho, “Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization,” *Expert Systems with Applications*, vol. 47, pp. 106–119, 2016.

[37] A. Naserbegi, M. Aghaie, and A. Zolfaghar, “Implementation of grey wolf optimization (GWO) algorithm to multi-objective loading pattern optimization of a PWR reactor,” *Annals of Nuclear Energy*, vol. 148, p. 107703, 2020.

[38] D. Aboooli, 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.

[39] A. Baghban, M. Kahani, M. A. Nazari, M. H. Ahmadi, and W. M. Yan, “Sensitivity analysis and application of machine learning methods to predict the heat transfer performance of CNT/water nanofluid flows through coils,” *International Journal of Heat and Mass Transfer*, vol. 128, pp. 825–835, 2019.

[40] A. Baghban, A. H. Mohammadi, and M. S. Taleghani, “Rigorous modeling of CO2 equilibrium absorption in ionic liquids,” *International Journal of Greenhouse Gas Control*, vol. 58, pp. 19–41, 2017.

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

[42] 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.

[43] S. Alizadeh, I. Alruyemi, R. Daneshfar, M. Mohammadi-Khanposhtani, and M. Nasri, “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.

[44] F. Mousazadeh, M. H. T. Naeem, R. Daneshfar, B. S. Soulhani, and M. Nasri, “Predicting the Condensate Viscosity near the Wellbore by ELM and ANFIS-PSO Strategies,” *Journal of Petroleum Science and Engineering*, vol. 204, p. 108708, 2021.