Research Article

Using an SGB Decision Tree Approach to Estimate the Properties of CRM Made by Biomass Pretreated with Ionic Liquids

Ya-Qing Gu [iD],[1] **Tao Shu** [iD],[2] **Bin Ge** [iD],[3] **Ping Wang** [iD],[1] **Chen Gao** [iD],[4] **and Hamid Heydari** [iD][5]

[1]*College of Electrical Engineering, Shandong Huayu University of Technology, Dezhou, Shandong 253034, China*
[2]*Modern Information Technology Center, Sichuan Vocational and Technical College, Suining, Sichuan 629000, China*
[3]*Institute of Information Engineering, Shandong Huayu University of Technology, Dezhou, Shandong 253034, China*
[4]*Experiment Management Center, Dezhou University, Dezhou, Shandong 253033, China*
[5]*Institute of Petroleum Engineering, University of Tehran, Tehran, Iran*

Correspondence should be addressed to Tao Shu; st@scun.edu.cn and Hamid Heydari; h.heydari.2014@ut.ac.ir

The use of ionic liquids (ILs) for biomass pretreatment to produce cellulose-rich materials (CRMs) has been well proven. In this research, due to the wide range of applications and ease of using artificial intelligence procedures, on the basis of the algorithm of stochastic gradient boosting (SGB) decision tree, an artificial intelligence approach is proposed to estimate the properties of cellulose-rich materials (CRMs). That being the case, the dataset of the empirical output values was gathered and was randomly broken down into datasets for testing and training. These results show that the best forecasting tool for calculating the properties of CRMs is the developed model. Furthermore, the accuracy of the databank of the biodiesel target values has been examined. In contrast, the influences of model contributed variables on the output have been examined as a new issue. It reveals that the most influencing variable in determining the properties of CRMs is the cellulose enrichment factor. Therefore, this research provides an innovative and accurate tool for predicting the properties of CRMs and sensitivity investigation on effective parameters to help investigators developing the optimized process.

1. Introduction

The source of many environmental issues, including rising greenhouse gases, global warming, and air pollution, is the increasing consumption of fossil fuels in the world [1–3]. Biomass as carbon-neutral and low-emission fuels, which are considered renewable sources, has been considered as a suitable alternative to fossil fuels in recent years [4–6]. The components of lignocellulosic biomass include cellulose, hemicellulose, and lignin which contain about 40–60% w/w, 15–30% w/w, and 10–25% w/w, respectively. These three components weave together, forming a strong cell wall that leads to the formation of recalcitrant biomass. The main structure of the biomass skeleton is formed by the accumulation of hard cellulose microfibers, and the inside space is filled by amorphous hemicelluloses and lignin. Eventually, all the components are joined together by covalent and

hydrogen bonds [7]. Studies have shown that lignin is the main factor preventing the deconstruction of lignocellulosic biomass that leads to the conversion of fibrous cellulose into biochemicals and/or bio-fuels as well. When lignin is extracted and along with that, hemicellulose decreases, the available surface area in cellulose increases, creating areas favorable to reactions such as thermal conversion, bio digestibility, and hydrolysis [8, 9].

Ionic liquid solvents are commonly used to extract lignin from a large number of lignocellulosic biomass types in pretreatment processes [10]. The solid material obtained from the abovementioned process is a cellulose-rich material (CRM) and can be upgraded into high-quality biochemical or fuels through processes such as fermentation [11], torrefaction [12], and pyrolysis [13].

There are three different solvent systems for the ionic liquid solvents mentioned which include ionic liquids (ILs),

deep eutectic solvents (DESs), and IL-containing solvent [14]. Since IL is pure salt, it has a low melting point, so its physical and chemical properties can be optimized for the intended purpose by making small changes in its anion and cation composition [15]. Due to the special properties of IL solvent, various properties can be enumerated for it, such as low vapor pressure, nonflammability, and thermal and chemical stability [16].

Moreover, recent studies have revealed that ILs which contain organic solvents can reduce the viscosity of IL by promoting the dissociation of their anion and cation and lead to further solubility of lignocellulosic biomass [17, 18]. The chemical composition of DES includes a number of chemical components in the form of hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), and the melting point of DES is significantly lower than the HBA and HBD components that form it [19]. Also, the release of the biomass recalcitrance is facilitated by the strong hydrogen bond [20]. The number of papers published on ILs in the role of pretreatment of biomass has been increased in recent years [21].

Artificial intelligence methods for understanding complex and multidimensional relationships in biomass pretreatment processes using ILs have more advantages than regression and traditional modeling, and machine learning (ML) is highly regarded among the artificial intelligence methods [10, 22, 23].

One of the advantages of ML is that it can learn and detect patterns automatically from a large dataset without the need for explicit programming. ML has made it simply possible to evaluate the importance of all input features which helps to better understand the system, leading to the model interpretation-based design of experiments. Recent studies have used ML in upgrading and conversion processes including hydrothermal, torrefaction, pyrolysis, and fermentation [24–30]. Gradient boosting (GB), support vector machines (SVMs), and random forest (RF) are among the algorithms that have been widely used in the literature. The algorithms performed well in predicting and provided desirable values of two important indicators, R^2 (regression coefficient) and RMSE (root mean square error). Low values of R^2 and RMSE have been reported at around 0.90. Nevertheless, the feature number was less than 15, which is different from the biomass pretreatment system with ILs. One of the challenges we faced in our system was also describing features specifically for ILs. Xu et al. used an explicit inner relationship between the main variables of DESs to describe the characteristics of each DES, which showed a clear effect on the pretreatment of lignocellulosic biomass [31]. Physical and chemical parameters related to hydrogen bonds in DESs were also analyzed by partial least square (PLS) and principal component analysis (PCA). Cellulose content and solid recovery in CRMs are among the most important properties in the biomass pretreatment process with ILs. However, the application of ML in evaluating and predicting feature importance for them has not been reported yet. Therefore, in this study, we intend to use machine learning algorithms to predict cellulose

enrichment factor and solid recovery in CRM, which is the result of biomass pretreatment with ILs. The dataset used in our research includes 23 features related to raw biomass characteristics, IL identities, IL treatment process conditions, and catalyst loading.

The SGB decision tree approach was considered and focused on for modeling this work. Visualization and interpretation of high-importance features by plotting their impact patterns in the form of statistical analyses led to a better understanding of the multidimensional relationships in the biomass pretreatment system with ILs. Finally, different sets of input features were examined to explore their effect on the predictive power of the model.

2. A History of the Stochastic Gradient Boosting (SGB) Decision Tree

An algorithmic boosting function enhances a predictive function's accuracy. This technique begins by repeatedly executing the function, followed by merging each function's throughput with weights to reduce prediction error. Boosting, which applies to both regression and classification issues, was developed during the last decade and is, thus, considered among the most robust and recent learning algorithms [32].

Friedman's SGB technique [33] is regarded as an advanced functional approach and statistical study variation. This technique's algorithmic function pertains to boosting regression trees and calculates the sequence of basic trees where residuals of predictors from its previous tree create the subsequent one. Tree intricacy is determined by a single split consisting of a single root node and two child nodes. Data are then partitioned incrementally using an SGB approach. After then, for each partition, the distinction among the measured values and remnants is determined. Following that, the residuals are fitted to the tree node to create a new partition, which will reduce the data's residual variance in the previously described tree sequence. By classifying each statement using the most common technique, as a result, all the trees produced throughout the process are accrued, which reduces the SGB's responsiveness for the imbalanced sets of data, deficient training set, and anomalies.

Additionally, ensemble education techniques such as bagging, boosting, and related methods aggregate projections derived from several designs. The techniques are significantly effective in artificial intelligence and the extraction of information [34]. The techniques include the following:

$$f(x) = a_0 + \sum_{k=1}^{k} a_k f_k(x), \tag{1}$$

where k and $f_k(x)$ denote each of the basic learner sizes and the ensemble, respectively, and numerous functions from x, being input variables originating from the training set.

Ensemble estimation $F(x)$ is a linear mixture from all projections from the basic ensemble learners, where the parameters specifying this linear combination are $\{a_k\}_0^k$.

Boosting predicts $F(x)$ derived from the preceding equation once the additive enhancement is used via the subsequent formula

$$f(x) = \sum_{k=0}^{k} \beta_k g(x; a_k), \qquad (2)$$

where $g(x; a_k)$ and $a = (a_1 a_2, \ldots)$ are denoted as simple functions of x. A progressive method is utilized to fit the parameters $\{a_k\}_0^k$ and to expand coefficients $\{\beta_k\}_0^k$ to the training data.

This procedure begins with the calculation of an estimation $F_0(x)$, followed by $k = 1, 2, \ldots, K$.

$$(\beta_k, a_k) = \text{argmin}_{\beta,a} \sum_{i=1}^{N} L(y_i, f_{k-1}(x_i) + \beta g(x_i; a)), \quad (3)$$

$$f_k(x_i) = f_{k-1}(x_i) + \beta g(x_i; a_k). \qquad (4)$$

The solution to equation (1) is accomplished in two steps using gradient boosting for the differentiable and arbitrary loss functions. Initially, the basic learner function $g(x; a_k)$ is corresponded using the least square criterion.

$$a_k = \text{argmin}_{a,\rho} \sum_{i=1}^{N} (y_{ik} - \rho g(x_i; a))^2. \qquad (5)$$

To current pseudoresiduals,

$$y_{ik} = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{K-1}(X)}. \qquad (6)$$

The optimal coefficient value is then determined as

$$\beta_K = \text{argmin}_\beta \sum_{i=1}^{N} L(y_i, f_{k-1}(x_i) + \beta g(x_i; a)). \qquad (7)$$

As a result, equation (1)'s optimization of the complex function issue is substituted with equation (2), which employs smallest squares, succeeding to equation (3)'s sole variable optimization, which is centered on the loss criterion L. SGB is concerned with remedying observations that are proximate to decision-making limits determined through the data structure when performing the boosting operation [33]. During the boosting process, it is more likely that particular findings from a decision tree proximate to other classes will be identified and corrected [35].

3. Methodology

By utilizing machine learning techniques, the following six steps are required to develop an SGB tree model: (1) preparation of the dataset, (2) calculation of the descriptors, (3) feature selection and model training, (4) validation of the model, (5) recognition of the applicability domain, and (6) interpreting the model. The following subsections provide a more detailed description of the steps mentioned above.

3.1. The Dataset upon Which the Model Was Built.
A large set of 514 experimental data on CRM properties, including solid recovery (SR) and cellulose enrichment factor (CEF), which are a function of 23 effective variables, was collected from the literature. Further details on these parameters and data are provided elsewhere [10]. Three-quarters of these data was randomly used for the training phase, and the rest was retained for testing and evaluating the accuracy of the model.

3.2. Sensitivity Analysis.
Sensitivity studies were performed to ascertain each component of the input's effect on variables in the study. For quantitatively measuring the effect of each parameter, the following relevance factor was defined [36–38]:

$$r = \frac{\sum_{i=1}^{n} (X_{K,i} - \overline{X}_K)(Y_I - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_{K,i} - \overline{X}_K)^2 - \sum_{i=1}^{n} (Y_I - \overline{Y})^2}}. \qquad (8)$$

From the previous formula, n denotes an overall amount of the data points, $X_{k,i}$ denotes the ith value of the input for the kth criterion, and Y_i denotes the ith output value. $\overline{X}_k$ and $\overline{Y}$ denote the mean values of the kth input and output parameters, respectively. The relevancy factor has a value between the numbers −1 and +1, with more significant absolute numbers indicating the criterion corresponded to possess a farther significant effect on the variables in the study. A positive or negative value of the factor of relevancy indicates the straight or opposite consequence of criteria corresponded to, which suggests that the variable increments when particular input criteria with a positive relevance element are increased. At the same time, it lowers when a particular input criterion with a negative relevance factor increases.

Twenty-four different criteria for the input have been investigated in this study, all of which directly impacted the corresponding results. Figures 1 and 2 show the results of the sensitivity analysis. As shown, the largest eigenvalue n belongs to Lg, with the positive relevance factor calculated as 0.59.

3.3. The Preanalysis Phase.
In this work, five distinct statistical approaches were used to estimate and validate the CEF and SR generated from the SGB tree model. The created model was run using the MATLAB program version 2018. The data gathered in the experiment phase of this study were utilized for training the model, while approximately 25% of the data obtained was used to test the models. Additionally, the data were normalized [39–41].

$$D_K = 2 \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1. \qquad (9)$$

In equation (9), x denotes the number from the nth parameter, and as forecasted, the absolute number of D_k calculates as lower than 1. The CEF and SR are predicted through the primary output, whereas the remaining variables are utilized as a source of data for the SGB tree model.

3.4. Methodology for Modelling and Verifying.
Verification of the model and its output and its accuracy is a critical step in the model development process. Validation is
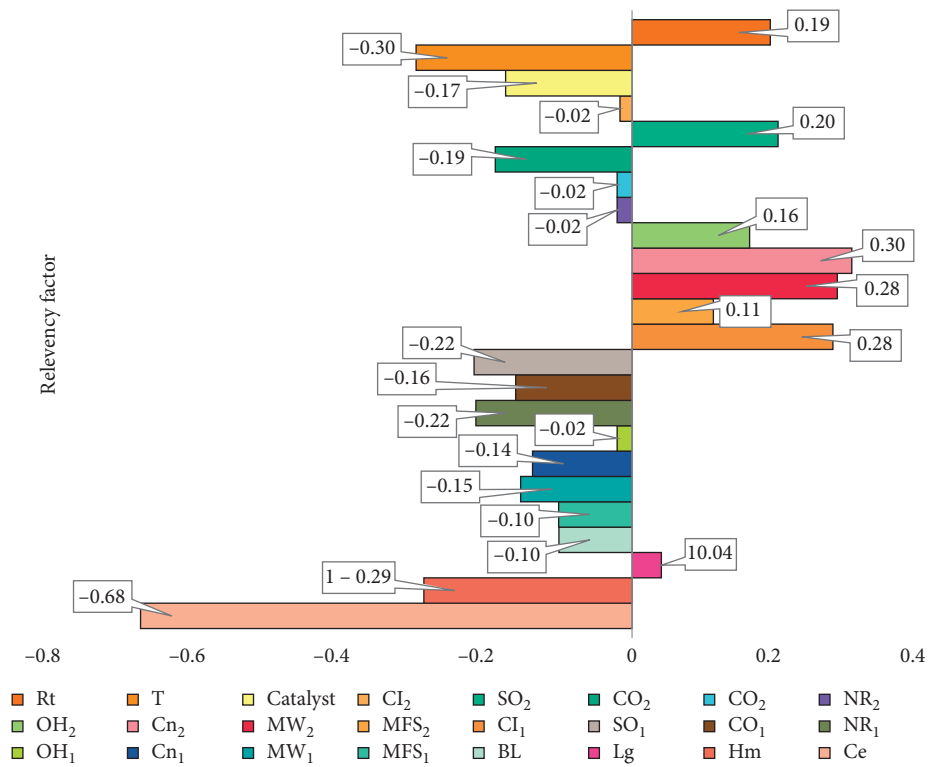
Figure 1: Sensitivity of parameters affecting the output parameter SR.
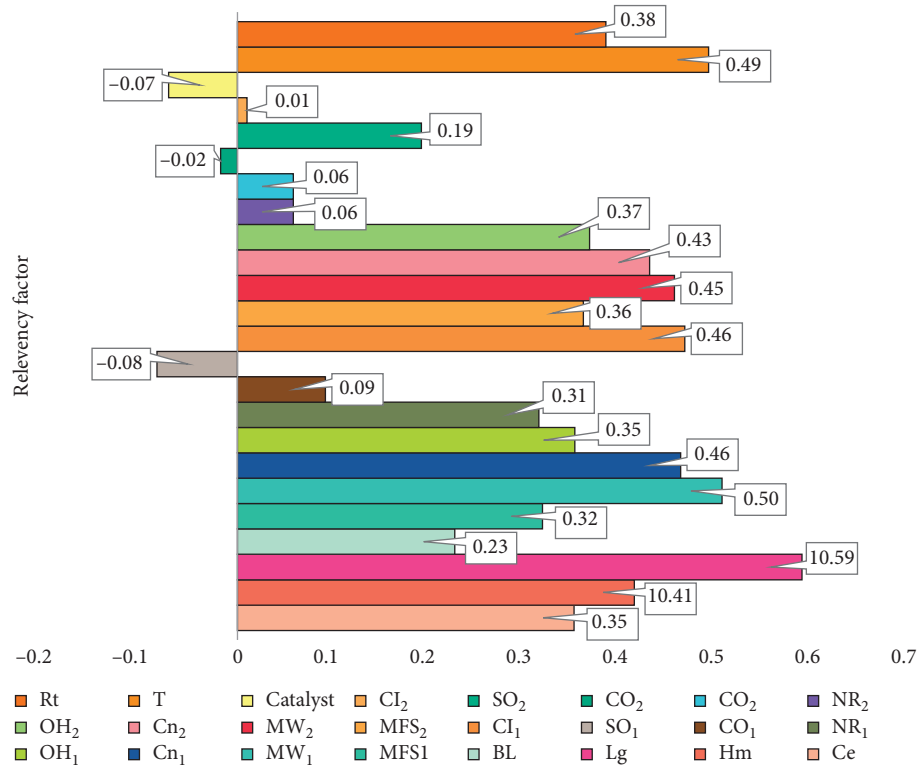

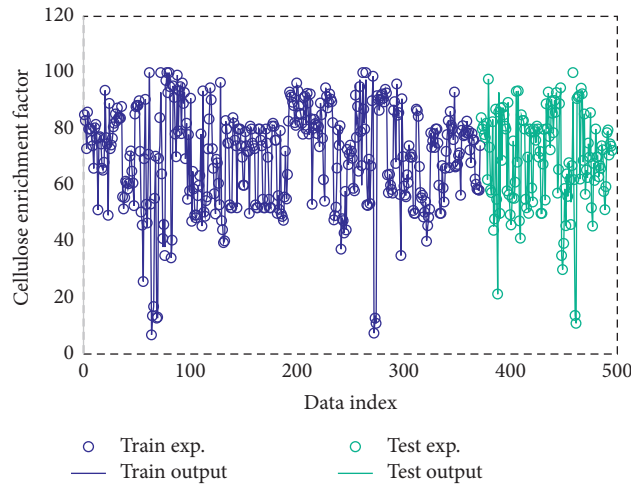Figure 2: Sensitivity of parameters affecting the output parameter CEF.

FIGURE 3: Simultaneous viewing of real and modeled CEF output data.

necessary as ranges of parameters are expanded and the experiments are enhanced. The equations mentioned below are utilized to assess the proposed model's accuracy [42]:

$$\text{Mean relative error (MRE)} = \frac{100}{N} \sum_{i=1}^{N} \left(\frac{x_i^{\text{actual}} - x_i^{\text{predicted}}}{x_i^{\text{actual}}} \right),$$

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(x_i^{\text{actual}} - x_i^{\text{predicted}} \right)^2},$$

$$\text{Standard deviations (STDs)} = \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(\text{error} - \overline{\text{error}} \right) \right)^{0.5}, \quad (10)$$

$$\text{Mean squared error (MSE)} = \frac{1}{N} \sum_{i=1}^{N} \left(x_i^{\text{actual}} - x_i^{\text{predicted}} \right)^2,$$

$$R-\text{squared}\left(R^2 \right) = 1 - \frac{\sum_{i=1}^{N} \left(x_i^{\text{actual}} - x_i^{\text{predicted}} \right)^2}{\sum_{i=1}^{N} \left(x_i^{\text{actual}} - x^{\text{actual}} \right)^2}.$$

4. Results and Discussion

In Figures 3 and 4, the experimental values are compared to the CEF and SR calculated via the SGB tree for both the testing and training datasets.

The presented model can accurately estimate outputs, as seen in the figures. Several statistical and graphical techniques were employed to evaluate the suggested model's validity. The regression plots in Figures 5 and 6 demonstrate the proposed model's capacity to project targets, seen by a dense clustering of the data points surrounding the $Y = X$ line.

The proposed SBB tree's error plot is depicted in Figures 7 and 8. It plots relative error versus the CEF and SR through both training and testing sets of data, respectively.

Furthermore, estimated errors were also generally centered on the line of zero deviation. The mean of the relative error of the suggested model was worked out to be less than 17%, which displays the prediction accuracy through the suggested SGB tree model.

Table 1 contains statistical parameters derived from the suggested model. The suggested model exhibits low MSE, STD, RMSE, and MRE% values and an elevated R^2, showing that it accurately predicts outputs.

Phromphithak et al. used the random forest method to estimate the two output parameters of CEF and SR [10] and concluded that their model was able to predict these parameters with an accuracy of $R^2 = 0.94$ and 0.84, respectively, that is weaker than our proposed model.
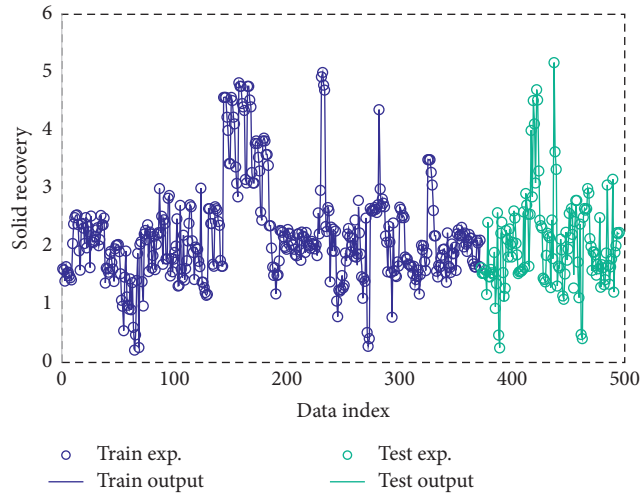
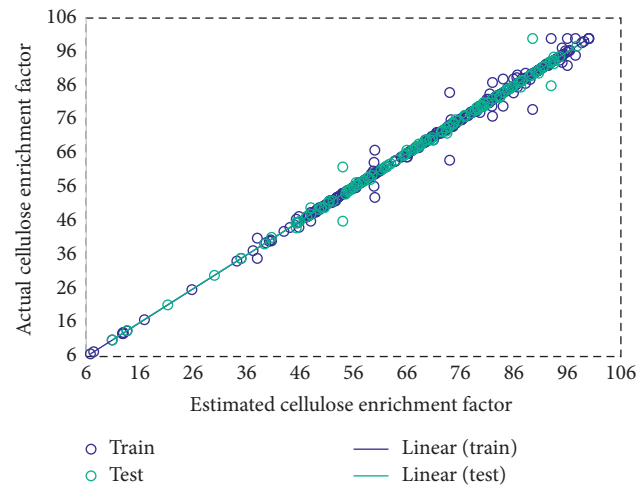FIGURE 4: Simultaneous viewing of real and modeled SR output data.

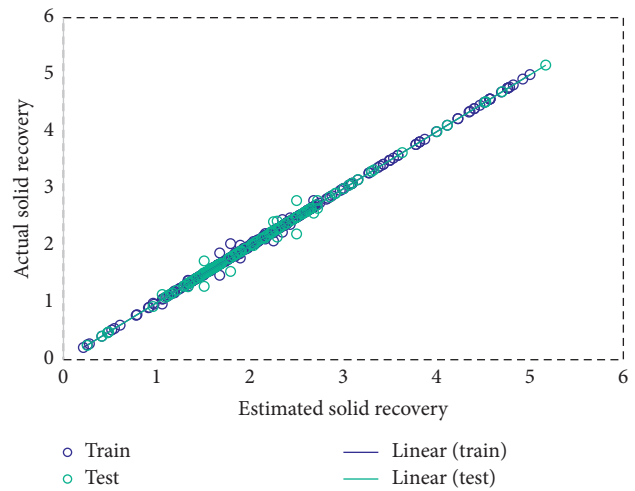FIGURE 5: Cross plot on the CEF output values.

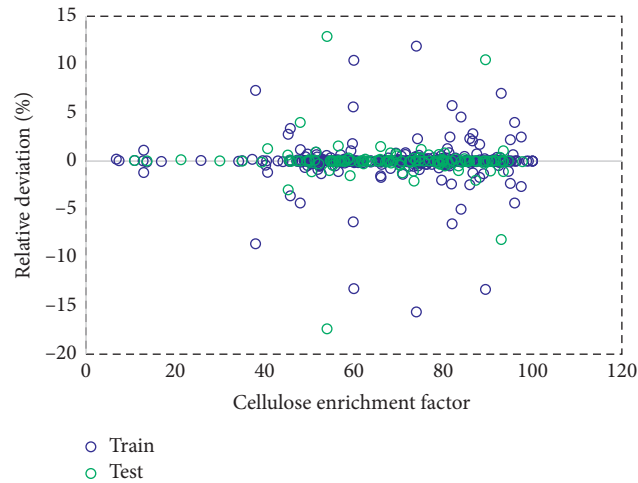FIGURE 6: Cross plot on the SR output values.

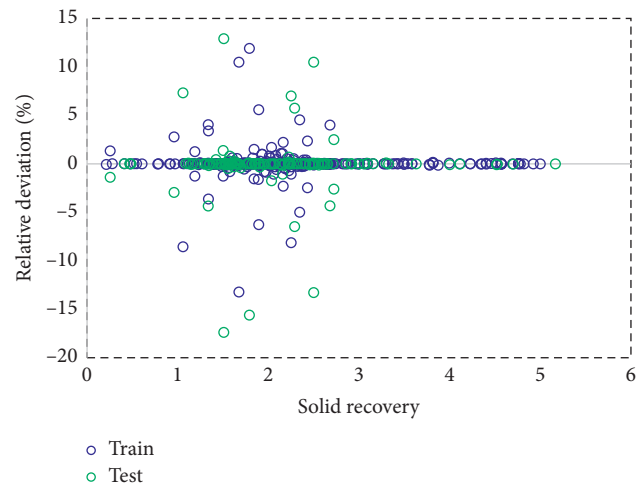Figure 7: The relative deviation analysis on the CEF output values.

Figure 8: The relative deviation analysis on the SR output values.

Table 1: Various statistical analyses to evaluate the proposed model.

Output	Phase	R^2	MRE (%)	MSE	RMSE	STD
Solid recovery	Training set	0.999	0.37	$6.87E-04$	$2.62E-002$	$2.53E-002$
	Testing set	0.995	0.98	$3.39E-003$	$5.82E-002$	$5.55E-002$
	Total set	0.998	0.52	$1.36E-003$	$3.69E-002$	$3.56E-002$
Cellulose enrichment factor	Training set	0.994	0.61	$1.83E+00$	$1.35E+00$	$1.28E+00$
	Testing set	0.992	0.71	$2.43E+00$	$1.56E+00$	$1.49E+00$
	Total set	0.994	0.64	$1.98E+00$	$1.41E+00$	$1.34E+00$

5. Conclusions

The present work provides a novel perspective on the prediction of CEF and SR. We have developed a precise stochastic gradient-boosting decision tree model to this end. The model has been constructed and validated using the test and training sets. Additionally, extrinsic assessment sets were used to evaluate the actual projective capabilities of the model. This design is based solely on effective inputs. Thus, the estimated CEF and SR produced via the suggested method can fill in the investigational observations by assuming unknown or missing values. Furthermore, the proposed prediction mechanism may point researchers in the direction of a novel successful measurement method. This work quantified error analysis for various inputs. Forward to the results, the strategy outperformed earlier

models in terms of generalizability, validity, and accuracy. Finally, the suggested method can be an efficient tool to analyze and design more effective related units.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] A. Shahsavari and M. Akbari, "Potential of solar energy in developing countries for reducing energy-related emissions," *Renewable and Sustainable Energy Reviews*, vol. 90, pp. 275–291, 2018.

[2] A. Karapekmez and I. Dincer, "Comparative efficiency and environmental impact assessments of a solar-assisted combined cycle with various fuels," *Applied Thermal Engineering*, vol. 164, p. 114409, 2020.

[3] K. Kaygusuz, "Energy and environmental issues relating to greenhouse gas emissions for sustainable development in Turkey," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 253–270, 2009.

[4] M. Umar, X. Ji, D. Kirikkaleli, and A. A. Alola, "The imperativeness of environmental quality in the United States transportation sector amidst biomass-fossil energy consumption and growth," *Journal of Cleaner Production*, vol. 285, p. 124863, 2021.

[5] Z. A. Elum and A. S. Momodu, "Climate change mitigation and renewable energy for sustainable development in Nigeria: a discourse approach," *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 72–80, 2017.

[6] P. K. Swain, L. M. Das, and S. N. Naik, "Biomass to liquid: a prospective challenge to research and development in 21st century," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4917–4933, 2011.

[7] S. Wang, G. Dai, H. Yang, and Z. Luo, "Lignocellulosic biomass pyrolysis mechanism: a state-of-the-art review," *Progress in Energy and Combustion Science*, vol. 62, pp. 33–86, 2017.

[8] C. Zhang, J. Mo, Q. Fu et al., "Wood-cellulose-fiber-based functional materials for triboelectric nanogenerators," *Nano Energy*, vol. 81, p. 105637, 2020.

[9] B. Duan, Y. Huang, A. Lu, and L. Zhang, "Recent advances in chitin based materials constructed via physical methods," *Progress in Polymer Science*, vol. 82, pp. 1–33, 2018.

[10] S. Phromphithak, T. Onsree, and N. Tippayawong, "Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents," *Bioresource Technology*, vol. 323, p. 124642, 2021.

[11] A. S. Dotsenko, G. S. Dotsenko, O. V. Senko et al., "Complex effect of lignocellulosic biomass pretreatment with 1-butyl-3-methylimidazolium chloride ionic liquid on various aspects of ethanol and fumaric acid production by immobilized cells within SSF," *Bioresource Technology*, vol. 250, pp. 429–438, 2018.

[12] T. Onsree, N. Tippayawong, T. Williams et al., "Torrefaction of pelletized corn residues with wet flue gas," *Bioresource Technology*, vol. 285, p. 121330, 2019.

[13] N. Khuenkaeo, B. MacQueen, T. Onsree, S. Daiya, N. Tippayawong, and J. Lauterbach, "Bio-oils from vacuum ablative pyrolysis of torrefied tobacco residues," *RSC Advances*, vol. 10, no. 58, pp. 34986–34995, 2020.

[14] J. M. Klein, H. Squire, W. Dean, and B. E. Gurkan, "From salt in solution to solely ions: solvation of methyl viologen in deep eutectic solvents and ionic liquids," *The Journal of Physical Chemistry B*, vol. 124, no. 29, pp. 6348–6357, 2020.

[15] S. Phromphithak, P. Meepowpan, S. Shimpalee, and N. Tippayawong, "Transesterification of palm oil into biodiesel using ChOH ionic liquid in a microwave heated continuous flow reactor," *Renewable Energy*, vol. 154, pp. 925–936, 2020.

[16] D. Z. Troter, Z. B. Todorović, D. R. Đokić-Stojanović, O. S. Stamenković, and V. B. Veljković, "Application of ionic liquids and deep eutectic solvents in biodiesel production: a review," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 473–500, 2016.

[17] X. Zhang, W. Zhao, Y. Li, C. Li, Q. Yuan, and G. Cheng, "Synergistic effect of pretreatment with dimethyl sulfoxide and an ionic liquid on enzymatic digestibility of white poplar and pine," *RSC Advances*, vol. 6, no. 67, pp. 62278–62285, 2016.

[18] T. G. Weldemhret, A. B. Bañares, K. R. M. Ramos et al., "Current advances in ionic liquid-based pre-treatment and depolymerization of macroalgal biomass," *Renewable Energy*, vol. 152, pp. 283–299, 2020.

[19] Y. Chen and T. Mu, "Application of deep eutectic solvents in biomass pretreatment and conversion," *Green Energy & Environment*, vol. 4, no. 2, pp. 95–115, 2019.

[20] Q. Yu, A. Zhang, W. Wang et al., "Deep eutectic solvents from hemicellulose-derived acids for the cellulosic ethanol refining of Akebia' herbal residues," *Bioresource Technology*, vol. 247, pp. 705–710, 2018.

[21] V. G. Zuin and L. Z. Ramin, "Green and sustainable separation of natural products from agro-industrial waste: challenges, potentialities, and perspectives on emerging approaches," *Topics in Current Chemistry Collections*, Springer, Cham, Switzerland, pp. 229–282, 2018.

[22] M. Smuga-Kogut, T. Kogut, R. Markiewicz, and A. Słowik, "Use of machine learning methods for predicting amount of bioethanol obtained from lignocellulosic biomass with the use of ionic liquids for pretreatment," *Energies*, vol. 14, no. 1, p. 243, 2021.

[23] J. Chapman, V. K. Truong, A. Elbourne et al., "Combining chemometrics and sensors: toward new applications in monitoring and environmental analysis," *Chemical Reviews*, vol. 120, no. 13, pp. 6048–6069, 2020.

[24] P. J. García Nieto, E. García–Gonzalo, F. Sánchez Lasheras, J. P. Paredes–Sánchez, and P. Riesgo Fernández, "Forecast of the higher heating value in biomass torrefaction by means of

machine learning techniques," *Journal of Computational and Applied Mathematics*, vol. 357, pp. 284–301, 2019.

[25] T. Onsree and N. Tippayawong, "Machine learning application to predict yields of solid products from biomass torrefaction," *Renewable Energy*, vol. 167, pp. 425–432, 2021.

[26] X. Chen, H. Zhang, Y. Song, and R. Xiao, "Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis," *Chemical Engineering and Processing-Process Intensification*, vol. 130, pp. 36–42, 2018.

[27] L. Wang, F. Long, W. Liao, and H. Liu, "Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms," *Bioresource Technology*, vol. 298, p. 122495, 2020.

[28] X. Zhu, Y. Li, and X. Wang, "Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions," *Bioresource Technology*, vol. 288, p. 121527, 2019.

[29] J. Li, L. Pan, M. Suvarna, Y. W. Tong, and X. Wang, "Fuel properties of hydrochar and pyrochar: prediction and exploration with machine learning," *Applied Energy*, vol. 269, p. 115166, 2020.

[30] A. Pathy, S. Meher, and P. Balasubramanian, "Predicting algal biochar yield using extreme gradient boosting (XGB) algorithm of machine learning methods," *Algal Research*, vol. 50, p. 102006, 2020.

[31] H. Xu, Y. Kong, J. Peng et al., "Multivariate analysis of the process of deep eutectic solvent pretreatment of lignocellulosic biomass," *Industrial Crops and Products*, vol. 150, p. 112363, 2020.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, Berlin, Germany, 2001.

[33] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Berlin, Germany, 2006.

[35] D. G. Whiting, J. V. Hansen, J. B. McDonald, C. Albrecht, and W. S. Albrecht, "Machine learning methods for detecting patterns of management fraud," *Computational Intelligence*, vol. 28, no. 4, pp. 505–527, 2012.

[36] A. Rostami, A. Baghban, A. H. Mohammadi, A. Hemmati-Sarapardeh, and S. Habibzadeh, "Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: smart modeling and correlation development," *Fuel*, vol. 236, pp. 110–123, 2019.

[37] A. Bahadori, A. Baghban, M. Bahadori et al., "Computational intelligent strategies to predict energy conservation benefits in excess air controlled gas-fired systems," *Applied Thermal Engineering*, vol. 102, pp. 432–446, 2016.

[38] 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.

[39] A. Baghban and A. Khoshkharam, "Application of LSSVM strategy to estimate asphaltene precipitation during different production processes," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1855–1860, 2016.

[40] A. Baghban, T. Kashiwao, M. Bahadori, Z. Ahmad, and A. Bahadori, "Estimation of natural gases water content using adaptive neuro-fuzzy inference system," *Petroleum Science and Technology*, vol. 34, no. 10, pp. 891–897, 2016.

[41] 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.

[42] M. H. Ahmadi, A. Baghban, M. Sadeghzadeh et al., "Evaluation of electrical efficiency of photovoltaic thermal solar collector," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 545–565, 2020.