Biomass higher heating value (HHV) modeling on the basis of proximate analysis using iterative network-based fuzzy partial least squares coupled with principle component analysis (PCA-INFPLS)

Soleiman Hosseinpour¹,⁎, Mortaza Aghbashlo²,⁎, Meisam Tabatabaei³,⁎

¹ Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
² Microbial Biotechnology Department, Agricultural Biotechnology Research Institute of Iran (ABRII), P.O. Box: 31535-1897, Agricultural Research, Education, and Extension Organization (AREEO), Karaj, Iran
³ Biofuel Research Team (BRTeam), Karaj, Iran

ARTICLE INFO

Keywords:
Adaptive network-based fuzzy inference system
Biomass fuels
Higher heating value
Iterative partial least squares
Principle component analysis
Proximate analysis

ABSTRACT

In this study, a novel iterative network-based fuzzy partial least squares coupled with principle component analysis (PCA-INFPLS) was proposed to predict the HHV of biomass fuels as a function of fixed carbon (FC), volatile matter (VM), and ash content. In this methodology, the PCA analysis was used to eliminate the collinearity of experimental data for providing the required background to the INFPLS model. In the INFPLS structure, adaptive network-based fuzzy inference system (ANFIS) was applied to correlate the inputs and the outputs of iterative PLS score vectors. Furthermore, the capability of the PCA-INFPLS approach in estimating the biomass fuels HHV was compared with those of the PLS, ANFIS, NFPLS, and INFPLS models. Generally, the PCA-INFPLS approach was much more efficient than the other applied methods in modeling the biomass fuels HHV. More specifically, the developed model predicted the HHV of biomass fuels with an $R^2 > 0.96$, an $MSE < 0.51$, and an $MAPE < 2.5\%$. Therefore, this approach could be utilized for reliable and accurate approximation of the HHV of biomass feedstocks based on the proximate analysis instead of lengthy laboratorial measurements. The PCA-INFPLS approach was then embedded into a simple and user-friendly software for estimating the biomass fuels HHV based on the proximate analysis.

1. Introduction

Massive utilization of conventional fossil fuel energy resources has led to growing concerns over the uncertainties of future energy supplies and more importantly deleterious atmospheric effects related to their direct combustion [1–3]. Hence, there is a surge of interest in the use of environmentally-benign alternative energy sources to protect the living environment [4]. In the past few decades, biofuels production from biomass using thermochemical conversion has attracted increasing attention as a biorenewable pathway to substitute fossil fuels since it offers considerable environmental benefits. Although biomass thermochemical transformation is among the most-studied biofuels conversion technologies to produce various hydrocarbons, being overwhelming about biomass large-scale production can lead to controversial competition over land and water for fuel or food. Therefore, only the application of waste-oriented biomass materials such as agricultural residues, sewage, animal manure, food waste, and municipal solid wastes could be suggested for biofuel production as an environmentally-acceptable and sustainable solution for the future [5,6].

Generally, biomass materials used as feedstocks in thermochemical conversion exhibit a great deal of variations in their chemical and structural compositions, leading to profound differences in their energy contents. Therefore, measuring the higher heating value (HHV) of different biomass fuels a priori can aid engineers and researchers to design biomass thermochemical conversion systems more efficiently. HHV is one of the most important quality indicators of biomass fuels as it represents their thermochemical characteristics. The HHV of a given biomass is usually experimentally determined using a bomb calorimeter [7,8]. Although experimental measurement of the HHV of biomass fuels is relatively easy and inexpensive, it is a very daunting, time-consuming, and tedious task. Accordingly, an accurate model would be beneficial in measuring the heating value of biomass feedstocks for...
It is interesting to note that determining the HHV a priori not only can give valuable insights regarding the propriety of a bio-based feedstock for thermochemical conversion but also can relieve the laborious experimental measurement. It is well-documented that the ultimate and proximate analyses of biomass fuels have a good agreement with their HHV. Therefore, estimating the HHV of biomass fuels using their ultimate and proximate analyses is regarded as an attractive alternative to the experimental procedures. It should be noted that ultimate analysis could result in more informative outcomes regarding elemental composition of biomass fuels compared with proximate analysis. However, proximate analysis is more popular for developing mathematical models to predict the HHV of biomass fuels because of its simplicity and the fact that it is less costly. This analysis is often used to determine the gross bio-based solid feedstocks components such as fixed carbon (FC), volatile matter (VM) and ash content [9,10].

To approximate the HHV of bio-based fuels quickly and reliably, various empirical models have been developed based on the proximate analysis by several researchers [9,11–15]. Overall, the precision of the reported models for estimating the HHV of various biomass fuels is acceptable. However, owing to the importance of this quality factor of biomass feedstocks, there is still a need to more accurately approximate HHV by using heuristically soft computing methods like artificial neural network (ANN). In line with this, Ghugare et al. [16] employed multilayer perceptron (MLP) ANN model and genetic programming (GP) for developing two biomass HHV prediction models with an \( R^2 \) higher than 0.95. Estiati et al. [17] very recently developed a multilayer perceptron ANN based on the proximate analysis (FC, VM, and ash content) to compute the HHV of biomass fuels with an \( R^2 \) higher than 0.96. Recently, Hosseinpour et al. [18] applied iterative neural network-adapted partial least squares (INFPLS) to predict the HHV of biomass fuels as a function of FC, VM, and ash content with an \( R^2 > 0.95 \).

In addition to ANN technique as a powerful tool for approximating the HHV of biomass fuels based on the proximate analysis, fuzzy logic approach has also been known to be a promising candidate technique for dealing with models suffering from nonlinearities and complexities [19]. However, it is very laborious to construct fuzzy rules manually for complex and nonlinear systems such as HHV of biomass fuels as a function of the proximate analysis. Fortunately, this problem can be overcome if fuzzy rules are developed using ANN technology, resulting in a hybrid approach called adaptive neuro-fuzzy inference system (ANFIS). In fact, ANFIS is one of the most employed neuro-fuzzy systems for modeling many problems of various disciplines because of its capability to handle large complex imprecise, uncertain, noisy, and rowdy dataset.

Moreover, estimating the HHV of biomass fuels can be effectively carried out using multivariate statistical data analysis methods such as partial least squares (PLS) as well. However, the standard PLS method is unsuitable for modeling complex and nonlinear biological systems due to its fixed inner linear function. This inherent shortcoming of the PLS method can be markedly prevailed using the combined biologically-inspired and knowledge-based soft computing techniques such as ANIFS approach by correlating its inputs and outputs score vectors nonlinearly. Interestingly, iterative training strategy can profoundly enhance the ability of the standard PLS method for modeling such highly nonlinear and complex systems. In better words, this extended version of PLS technique iteratively updates its weight vectors based on interior relation functions. It is worth quoting that various pre-processing statistical techniques such as Principle Component Analysis (PCA) can also be used to improve the capability of internal ANFIS model by eliminating the co-linearity between inputs before introducing to the PLS structure for interrelating the inputs and the outputs score vectors. In fact, PCA converts a set of observations of possibly correlated information into a set of values of linearly uncorrelated variables (PCs) in order to eliminates/reduces the co-linearity/dimensionality of complex modeling systems.

Therefore, the aim of this investigation was to present a novel iterative network-based fuzzy partial least squares coupled with principle component analysis (PCA-INFPLS) for exact estimation of biomass fuels HHV on the basis of the proximate analysis for the first time. In this method, the co-linearity of the experimental data collected from the published literature was eliminated by PCA analysis in order to prepare the required background to the INFPLS model. ANFIS model was then used to correlate the input and the outputs of the iterative PLS score vectors internally for approximating the HHV of biomass fuels using 350 dataset. The accuracy of the proposed model for modeling the HHV of biomass fuels was also compared with those of the PLS, ANFIS, NFPLS, and INFPLS models. Moreover, an easy-to-use computer model was also provided for facilitating the application of the developed model by researchers and by engineers, making it more applicable to the biofuel industry.

### 2. Materials and methods

#### 2.1. Data preparation

Database used for the development of the PLS, ANFIS, NFPLS, INFPLS, and PCA-INFPLS models were compiled from the published literature on 350 different biomass fuels [9,13]. The 350 data patterns collected from the previous papers were divided into 245 (≈70%), 52 (≈15%), and 53 (≈15%) datasets for the training, cross-validation, and testing of the developed models, respectively. The theoretical fundamentals of the PLS, ANFIS, NFPLS, INFPLS, and PCA-INFPLS approaches and their implementation procedures for modeling the HHV of biomass fuels have been thoroughly explained in the Supplementary Data.

#### 2.2. Development and evaluation of modeling systems

As mentioned earlier, the 350 patterns were divided into 245 (70%), 52 (≈15%), and 53 (≈15%) data sets for the training, cross-validation, and testing of the developed models, respectively. In the classical PLS and ANFIS methods, the FC, VM, and ash content were considered as input variables, while the HHV value was taken into account as the only output parameter. In the NFPLS modeling approach, the score vectors (\( t \)) and (\( u \)) were used as input and output vectors of the ANFIS models, respectively, for creating the inner relationships. In the INFPLS and PCA-INFPLS approaches, (\( t \)) and (\( u \)) were considered as input and

### Table 1

Statistical performance parameters of the developed approaches for modeling the HHV of biomass fuels.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training dataset</th>
<th>Test dataset</th>
<th>All dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>MSE</td>
<td>MAPE (%)</td>
</tr>
<tr>
<td>PLS</td>
<td>0.8695</td>
<td>1.6978</td>
<td>5.58</td>
</tr>
<tr>
<td>ANFIS</td>
<td>0.9178</td>
<td>1.6699</td>
<td>4.49</td>
</tr>
<tr>
<td>FNPLS</td>
<td>0.9200</td>
<td>1.0411</td>
<td>4.17</td>
</tr>
<tr>
<td>INFPLS</td>
<td>0.9569</td>
<td>0.8202</td>
<td>3.67</td>
</tr>
<tr>
<td>PCA-INFPLS</td>
<td>0.9565</td>
<td>0.5654</td>
<td>2.58</td>
</tr>
</tbody>
</table>
Fig. 1. Predicted HHV of biomass fuels vs. actual HHV using the linear PLS method for training (a), testing (b), and all datasets (c).

Fig. 2. Prediction error vs. actual HHV using the linear PLS method for training (a), testing (b), and all datasets (c), respectively.
output vectors for function \( f \), while \( \mathbf{u} \) and \( \mathbf{t} \) were regarded as input and output vectors for function \( g \), respectively. It is worth mentioning that the three obtained PCs were taken into account as model inputs in the PCA-INFPLS model. MATLAB computer program (ver. R2016a) was used to develop, train, and test the above-mentioned modeling systems. Maximum coefficient of determination \( (R^2) \), minimum mean-squared error \( (MSE) \), and mean absolute percentage error \( (MAPE) \) were used as statistical criteria to assess the capability of the developed models for modeling the HHV of biomass fuels on the basis of the proximate analysis. These statistical parameters are as follows [20,21]:

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_{\text{calc}}^i - y_{\text{exp}}^i)^2}{\sum_{i=1}^{N} (y_{\text{exp}}^i - \overline{y}_{\text{exp}})^2} \\
MSE = \frac{1}{N} \sum_{i=1}^{N} (y_{\text{calc}}^i - y_{\text{exp}}^i)^2 \\
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_{\text{calc}}^i - y_{\text{exp}}^i}{y_{\text{exp}}^i} \right| \times 100
\]

3. Results and discussions

3.1. PLS modeling

The statistical performance parameters, i.e., \( R^2 \), MSE, and MAPE of the developed approaches (PLS, ANFIS, FNPLS, IFNPLS, and PCA-INFPLS) for approximating the biomass fuels HHV as a function of the proximate analysis in the training and testing steps as well as for all the datasets used are tabulated in Table 1. The MSE values of the linear PLS for modeling the biomass fuels HHV in the training and testing steps were determined at 1.6978 and 1.1885, whereas the MAPE values were computed at 5.58% and 5.73%, respectively. Furthermore, the \( R^2 \) values were found to be 0.8695 and 0.9413 during the training and testing phases of the linear PLS, respectively. The validation of the linear PLS model was also performed by comparing the actual data with the estimated values during the training and testing steps as well as by considering all data patterns (Fig. 1a-c). It is obvious from the findings presented in Fig. 1 that the data points were not well located around a straight line with a slope equal to 1, showing the inadequacy and unreliability of the linear PLS in approximating the HHV of biomass fuels based on the proximate analysis.

The capability of the linear PLS model for biomass fuels HHV modeling was also visually assessed by representing prediction error...
3.2. ANFIS modeling

After some trial-and-error, it was found that the fuzzy system with 216 inference rules and 6 Gaussian membership functions (Fig. 3) for each input could successfully approximate the HHV of biomass fuels on the basis of the proximate analysis.

The MSE values of the ANFIS model for approximating the biomass fuels HHV in the training and testing steps were computed at 1.0699 and 0.7305, whereas the MAPE values were determined at 4.49% and 3.43%, respectively. In addition, the R² values were found to be 0.9178 and 0.9639 in the training and testing phases of the ANFIS model, respectively. A comparison between the statistical performance parameters of the linear PLS and ANFIS models showed the promising capability of the ANFIS model to predict the HHV of biomass fuels on the basis of the proximate analysis. In better words, the learning capability of the ANFIS model was superior and its generalization ability was remarkably better than that of the linear PLS model. The validation of the ANFIS model was also carried out by plotting the actual data against the predicted values in the training and testing steps as well as for all data patterns (Fig. 4(a-c)). Clearly, the ANFIS predicted values very close to the actual values, proving its accuracy for computing the HHV of biomass fuels on the basis of the proximate analysis.

Fig. 5a-c shows the prediction error (residual) against actual data of the ANFIS model in order to visually assess its capability for predicting the HHV of biomass fuels. The residuals of the ANFIS model were tightly distributed along the horizontal axis compared with the linear PLS model. This further demonstrated the consistency and reliability of the ANFIS model for determining the HHV of biomass fuels as a function of the proximate analysis. Therefore, ANFIS technique could be an appropriate tool for modeling the complex biological systems such as HHV of biomass fuels over the linear PLS or even the standard ANN model. It should be noted that the classical ANN models have "black-box" structure since their internal specifications are hidden. This inherent shortcoming of the ANN paradigm can be overcome to a great extent by ANFIS modeling systems. On the other hand, the fuzzy system cannot learn by itself about a given system, while ANN has learning abilities. Therefore, fuzzy system has the learning ability and ANN is more transparent in the ANFIS structure.

3.3. NFPLS modeling

The specifications of the best-performing ANFIS models in the NFPLS, INPFLS, and PCA-INPFLS structures for each h are summarized in Table 2. In the NFPLS model, three fuzzy systems with 20 Gaussian combination membership functions for each h could successfully approximate the HHV of biomass fuels on the basis of the proximate analysis composition.

It is obvious from the statistical parameters reported in Table 1 that the FNPLS model could slightly outperform the standard ANFIS model for modeling the biomass fuels HHV. The MSE values for modeling the HHV of biomass fuels using the FNPLS in the training and testing steps were found to be 1.0411 and 0.6316, while the MAPE values were calculated at 4.17% and 3.81%, respectively. Moreover, the R² values were determined at 0.9200 and 0.9688 in the training and testing phases of the FNPLS model, respectively. This could be attributed to the unique facets of this modeling system in integrating the PLS and ANFIS backgrounds into a single structure and thereby receiving the benefits of both methods concurrently.

Fig. 6a-c indicates the comparison between the experimental HHV of biomass fuels and those predicted by the FNPLS model. The results presented in this figure revealed that the plots between the predicted and experimental HHV were almost straight line for both training and testing stages. Therefore, the developed FNPLS model had a good generalization capability in estimating the HHV of biomass fuels on the
basis of the proximate analysis. Compared with the PLS and ANFIS models, the FNPLS model performed better because of the high non-linearity and complexity of the system. The superiority of the FNPLS model over the PLS and ANFIS models for modeling the HHV of biomass fuels was also evaluated by depicting prediction error (residual) vs. actual data (Fig. 7a-c). The compact distribution of the prediction error of the FNPLS model along the horizontal axis proved its promising potentiality in estimating the HHV of biomass fuels as a function of the proximate analysis.

### 3.4. INFPLS modeling

According to Table 2, three fuzzy systems with 20, 19, and 16 Gaussian combination membership functions for \( h = 1, 2, \) and 3 in the INFPLS structure, respectively, could satisfactorily approximate the HHV of bio-based fuels with respect to the proximate analysis. The MSE values of this approach in the training and testing phases were found to be 0.8202 and 0.5864, whereas the MAPE values were determined at 3.67% and 3.07%, respectively. Additionally, the \( R^2 \) values were determined at 0.9369 and 0.9710 in the training and testing phases, respectively. Clearly, the INFPLS model could slightly outperform the FNPLS model in approximating the HHV of biomass fuels due to the iteratively weight updating scheme of the iterative PLS method. The prediction performance of the INFPLS model in the training and testing phases as well as for all datasets is shown in Fig. 8a-c by comparing the experimental and predicted HHV of biomass fuels. It is obvious from this data presented in the figure that the data points were evenly and tightly distributed around the ideal unity-slope line, further demonstrating the suitability of this model for predicting the HHV of biomass fuels.

The prediction error (residual) of the INFPLS model against actual HHV of biomass fuels during training and testing steps as well as by considering all data patterns is depicted in Fig. 9(a-c). Obviously, the data points were well scattered on either side of the horizontal line with a slope of zero, revealing the suitability of the INFPLS model in approximating the biomass fuels HHV.

### 3.5. PCA-INFPLS modeling

According to Table 2, the inner ANFIS models had the best performance on which the numbers of Triangular membership functions in the INFPLS structure were fixed at 20, 20, and 18 for \( h = 1, 2, \) and 3, respectively. It is clear from the data reported in Table 1 that the MSE and MAPE were reasonably low, while the \( R^2 \) values were remarkably high in the training and testing steps of the PCA-INFPLS model. Although no straightforward relationship could be observed between the biomass fuels HHV and the proximate analysis, the statistical parameters of the PCA-INFPLS model were found in an acceptable range during both the training and testing phases. This meant that the PCA-INFPLS model had an excellent prediction ability for estimating the HHV of biomass fuels as a function of the proximate analysis. Fig. 10a-c compares the experimental data with the output data generated by the PCA-INFPLS model during the training and testing as well as for all datasets. It is clear from this figure that the estimated values of the PCA-INFPLS model were very close to the desired values and data points were excellently accumulated around a line with a slope equal to 1. In better words, the PCA-INFPLS model did satisfactorily approximate the HHV of biomass fuels as a function of the proximate analysis. Overall, the PCA-INFPLS model developed herein represented an important improvement for modeling the HHV of biomass fuels compared with the PLS, ANFIS, FNPLS, and INFPLS techniques. Although the PCA-INFPLS model employed lower numbers of rules compared with the FNPLS and INFPLS models, its prediction accuracy was better than these models. This occurred because of the fact that the ANFIS model could profoundly improve the learning capability of the iterative PLS model for the databases pre-processed by the PCA.

The capability of the PCA-INFPLS model for biomass HHV approximation was visually scrutinized by presenting prediction error (residual) vs. actual data (Fig. 11a-c). Obviously, the residuals were closely scattered around the horizontal line and showed no regular trends towards any clear patterns in the prediction error. In another word, the model had not only lower numbers of overestimated data points but also limited numbers underestimated data points. These results implied that the PCA-INFPLS model was satisfactorily capable of learning the relationships between the HHV and the proximate analysis of biomass fuels.

In order to further reveal the suitability of the developed approaches herein, their statistical results were compared with those reported in the published literature (Table 3). The MAPE and MSE of PCA-INFPLS approach were found to be 2.50% and 0.50. However, the MAPE values were 4.68% and 4.84% as well as the MSE values were 1.08 and 1.01 for MLP ANN and GP approaches developed by Ghugare et al. [16] as one of the best models reported in the literature for predicting the HHV of biomass fuels. Interestingly, the PCA-INFPLS approach developed throughout this study outperformed our previous INNPLS model which predicted the HHV of biomass fuels with an MAPE value of 2.97% and an MSE value of 0.61 [18]. In addition, almost all models developed throughout this study were much more precise than the empirical correlations and soft-computing approaches reported in the prestigious journals.

Overall, the HHV of biomass fuels could be satisfactorily predicted using the PCA-INFPLS approach as a function of the proximate analysis. In order to facilitate the application of this model by researchers and engineers, a simple and user-friendly software was developed and supplemented to this article (available at https://drive.google.com/file/d/0B3tMEhyhPUuVEAYRNpsFjJmc/view). Briefly, the percentage of FC, VM, and ash content contained in a given biomass could be easily keyed into obtain its HHV with an acceptable accuracy of higher than 97%.
Fig. 6. Predicted HHV of biomass fuels vs. actual HHV using the NFPLS method for training (a), testing (b), and all datasets (c).

Fig. 7. Prediction error vs. actual HHV using the NFPLS model for training (a), testing (b), and all datasets (c), respectively.
Fig. 8. Predicted HHV of biomass fuels vs. actual HHV using the INFPLS method for training (a), testing (b), and all datasets (c).

Fig. 9. Prediction error vs. actual HHV using the INFPLS model for training (a), testing (b), and all datasets (c), respectively.
Fig. 10. Predicted HHV of biomass fuels vs. actual HHV using the PCA-INFPLS method for training (a), testing (b), and all datasets (c).

Fig. 11. Prediction error vs. actual HHV using the PCA-INFPLS model for training (a), testing (b), and all datasets (c), respectively.
4. Conclusions

In this study, a new soft-computing approach called PCA-INFPLS was proposed and applied for estimating the HHV of biomass fuels based on the proximate analysis. The database required for training and testing the proposed model was gathered from the scientific publications. Furthermore, a comparison was performed in order to assess the prediction accuracy of the proposed approach over the standard PLS and ANFIS models as well as the NPLS and INFPLS approaches. Generally, the PCA-INFPLS method developed herein for biomass fuels HHV prediction was capable of yielding excellent results compared with the other models. The results obtained by the PCA-INFPLS approach showed a high accuracy with regard to the actual data. More specifically, the proposed approach successfully predicted the biomass fuels HHV on the basis of the proximate analysis with an R² of higher than 0.96 and an MSE and MAPE values lower than 0.51 and 2.50%, respectively. Hence, the PCA-INFPLS approach could be considered as an attractive alternative not only to the empirical correlations but also to the classical soft-computing techniques for modeling the biomass fuels HHV based on their proximate analysis. Finally, a user-friendly software package was also developed to estimate the HHV of biomass feedstocks with respect to the proximate analysis using the PCA-INFPLS approach for making it more applicable by researchers and engineers in the biofuel industry.

Acknowledgements

The authors gratefully acknowledge financial support from the Iran National Science Foundation (Grant no. 96005466).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fuel.2018.02.126.

References


