บทความวิจัย

การทดสอบความไม่คงเส้นคงวาในประสิทธิภาพของแบบจำลองค่าความร้อนชีวมวลตามการวิเคราะห์แบบแยกธาตุ

อัครา กิจการเจริญสิน* และ ศุภเชษฐ์ อินทร์เนตร
ภาควิชาวิศวกรรมคอมพิวเตอร์และเทคโนโลยีการเงิน คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 2697 6705 อีเมล: akara_kij@live4.utcc.ac.th DOI: 10.14416/j.kmutnb.2024.03.006
รับเมื่อ 5 เมษายน 2565 แก้ไขเมื่อ 7 มิถุนายน 2565 ตอบรับเมื่อ 21 มิถุนายน 2565 เผยแพร่ออนไลน์ 19 มีนาคม 2567

บทคัดย่อ

ชีวมวลเป็นหนึ่งในพลังงงานทดแทนที่สำคัญของแผนพัฒนาพลังงานทดแทนและพลังงานทางเลือกของประเทศไทย การวิเคราะห์แบบแยกธาตุบ่งบอกถึงค่าพลังงานความร้อนและคุณภาพของชีวมวล การทำนายค่าความร้อนที่มีความถูกต้องคงเส้นคงวาบนข้อมูลที่แบบจำลองไม่เคยพบเป็นสิ่งสำคัญต่อบางพื้นที่ซึ่งไม่มีข้อมูลเพียงพอที่จะให้แบบจำลองเรียนรู้ บทความใช้สามการทดลองเพื่อตรวจความไม่คงเส้นคงวาของแบบจำลองค่าความร้อน ข้อมูลที่ให้แบบจำลองเรียนรู้ คือ สิ่งทดลอง ส่วนผลลัพธ์ของการทดลองคือความไม่คงเส้นคงวาของแบบจำลอง บทความสร้างสถานการณ์จำลองเมื่อนำแบบจำลองค่าความร้อนมาทำนายข้อมูลอื่นที่แบบจำลองไม่เคยเรียนรู้ ผลการทดลองพบว่า แบบจำลองค่าความร้อนมีความถูกต้องที่ไม่คงเส้นคงวา บนข้อมูลที่แบบจำลองเคยพบนั้นแบบจำลองแต่ละอันให้ค่าความผิดพลาดเฉลี่ยที่ไม่ต่างกันเชิงสถิติ แต่มีโมเมนต์สูงต่างกัน ส่วนกรณีข้อมูลที่แบบจำลองไม่เคยพบนั้นการแจกแจงความความผิดพลาดจะต่างกันในทุกโมเมนต์ บนสถานการณ์จำลองแบบจำลองไม่สามารถรักษาระดับความถูกต้องอย่างที่เคยมีและไม่สามารถให้ผลการทำนายที่แม่นยำบนตัวอย่างชีวมวลของประเทศไทย ดังนั้นต้องนำข้อมูลชีวมวลในพื้นที่มาให้แบบจำลองเรียนรู้ใหม่จึงได้ผลการทำนายค่าความร้อนที่มีความแม่นยำ

คำสำคัญ: ชีวมวล ค่าความร้อน การวิเคราะห์แบบแยกธาตุ การทำนาย ความไม่คงเส้นคงวา

การอ้างอิงบทความ: อัครา กิจการเจริญสิน และ ศุภเชษฐ์ อินทร์เนตร, "การทดสอบความไม่คงเส้นคงวาในประสิทธิภาพของแบบจำลองค่าความร้อนชีวมวลตามการวิเคราะห์แบบแยกธาตุ," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 34, ฉบับที่ 2, หน้า 1–14, เลขที่บทความ 242-025954, เม.ย.–มิ.ย. 2567.

Research Article

Performance Inconsistencies in Biomass Higher Heating Value Models for Ultimate Analysis

Akara Kijkarncharoensin* and Supachate Innet
Department of Computer Engineering and Financial Technology, School of Engineering, University of the Thai Chamber of Commerce, Bangkok, Thailand

* Corresponding Author, Tel. 0 2697 6705, E–mail: akara_kij@live4.utcc.ac.th
DOI: 10.14416/j.kmutnb.2024.03.006
Received 5 April 2022; Revised 7 June 2022; Accepted 21 June 2022; Published online: 19 March 2024

Abstract

Biomass is a sustainable renewable energy that can replace fossil fuels, reduce greenhouse gas emissions, and be integrated into the energy structure of Thailand. Ultimate analysis measures the Higher Heating Value (HHV) of chemical elements to determine the energy quantity of a fuel. The accuracy and consistency of the out-of-sample data for the prediction model are essential for data-poor regions like Thailand. The present study conducted three experiments to verify the consistency of the HHV models using out-of-sample data. The published datasets were the treatments, with the accuracy stabilities being the responses. Multiple situations of out-of-sample implementation were simulated. The results confirmed accuracy inconsistencies in both linear and nonlinear HHV models. The models presented statistical indifferences in the average error of the in-sample performance, while the higher moments of error distribution remained distinct. All higher moments of the residuals of the model were different in the out-of-sample data. The simulated examples indicated that previous models could not maintain the accuracy of their training sets. Additionally, they could not provide an accurate prediction of the biomass data of Thailand. Therefore, a practical dataset is necessary to retrain the HHV models before implementation to ensure accurate HHV prediction in Thailand.

Keywords: Biomass, Higher Heating Value, Ultimate Analysis, Prediction, Inconsistency

Please cite this article as: A. Kijkarncharoensin and S. Innet, "Performance inconsistencies in biomass higher heating value models for ultimate analysis," *The Journal of KMUTNB*, vol. 34, no. 2, pp. 1–14, ID. 242-025954, Apr.–Jun. 2024 (in Thai).

1. Introduction

The highest proportion of global greenhouse gases originate from heat and electricity generation [1]. Hence, renewable energy technologies are essential for reducing greenhouse gas emissions and mitigating climate change [2]. In a developing country like Thailand, the increased expenditure on sustainable and affordable energy is already included in the integrated energy blueprint policies, and hence, the proportion of renewable energy used in Thailand is projected to increase from 12% to 30% by 2036, of which 77.2% will be attributed to biomass resources [3]. Therefore, active research is being conducted in this field to realize the renewable energy goal [4], [5].

The energy value of fuels is usually defined as the Higher Heating Value (HHV), which is the total energy released during a combustion reaction, including the evaporation and condensation of water. Experimental procedures typically evaluate the HHV through proximate and ultimate analyses. The former method uses the fixed carbon, volatile matter, and ash to estimate the heating value, while the latter determines the energy of constituent chemical elements. Based on the in-sample data, the accuracy of ultimate analysis is superior to that of proximate analysis [6], [7].

In Thailand, where biomass datasets are unavailable, the model prediction power may not be the same as that reported in earlier studies [8]. Many factors affect the properties of biomass, which include the species, growing area, rain, sun exposure, and harvesting time. The HHV prediction of the biomass of out-of-sample non-native species may not be as accurate as that in earlier studies. Therefore, the ultimate analysis, which provides the highest in-sample accuracy, is essential for evaluating the biomass energy yield of native species. However, there are no studies on the performance consistencies of out-of-sample data.

The accuracy of HHV models is often reported differently in earlier studies. Boumanchar *et al.* [9] compared their model with the linear models of other studies [10]–[13]. However, the accuracies stated in earlier studies were actually lower than those reported. Additionally, the accuracy degradation of linear models was observed in some studies [14]–[16]. Therefore, these inconsistencies in the accuracy need to be investigated.

The performance variabilities of nonlinear models are challenging to clarify owing to insufficient model information, unpublished datasets, and unrevealed source codes for reproducing the results. However, examining the performance consistency of these models is necessary before implementing the models for regions with insufficient data.

This study conducted experiments to examine the performance consistency of the HHV models reported in earlier studies on ultimate analysis. The proxies of fifteen linear and nonlinear machine learning models were investigated for their accuracies through three experiments. First, random searching constructed the highest accuracy models for each learning algorithm. Then, these models were treated as proxies of the HHV models in earlier studies. The published datasets were used as experimental treatments, while model accuracy was evaluated as the response.

We assumed that the proxies were equivalent to the nonlinear HHV models in earlier studies.

Finally, the experiments were limited to the ultimate analysis of the published datasets and the minimum mean square error cost function. To the best of our knowledge, no formal study has been conducted on the consistency of the accuracy of HHV models. The main contribution of this study is the verification of the consistency and evaluation of HHV models.

The remainder of this study is organized as follows: Section 2 describes the experimental design, the model, and the datasets. The practical diagrams and dataset overview are illustrated here. Section 3 presents the results and discussions pertaining to arising inconsistencies. Finally, the performance of the HHV models is inferred by comparing the experimental and reported values. The last section presents the conclusions of the study.

2. Materials and Methods

2.1 Accuracy Measurement

The Mean Square Error (MSE) (1) is the average of the prediction error square. The square root of this value is known as the $RMSE$ (2). The minimum MSE objective function implies a convex hypothesis space solution and the minimum sum of error square, while the RMSE is a proxy of the standard error. Therefore, MSE and $RMSE$ can indicate the size and spread of the distribution simultaneously, and were implemented as the primary measurement tools in this study.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(HHV_i - \widehat{HHV}_i)^2 \tag{1}$$

$$RMSE = \sqrt{MSE} \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|HHV_i - \widehat{HHV}_i\right| \tag{3}$$

$$AAE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|HHV_i - \widehat{HHV}_i\right|}{HHV_i}\times 100 \tag{4}$$

$$ABE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(HHV_i - \widehat{HHV}_i\right)}{HHV_i}\times 100 \tag{5}$$

The study used three other measurement tools supporting the results; Mean Absolute Error (MAE), average Absolute Error (AAE), and Average Bias Error (ABE), exhibited in Equations (3)–(5). However, these indicators refer to the magnitude of the error in their mean. Therefore, they are inferior to MSE and $RMSE$.

2.2 Experimental Design

The study assumed that optimization schemes affect the convergence rate and computational time. Local optima typically arise from inappropriate initial conditions. Therefore, differences in schemes cannot affect the weight and bias vectors of the optimal global solution.

Random searching can mimic earlier models through MATLAB statistics and machine learning toolbox. The source codes [17] were set as "OptimizeHyperparameters" to "all." and the models with the lowest MSE values were considered as proxies of those in earlier studies.

Fifteen models, exhibited in Table 1, were considered in the experiments. Numbers 1–12 are the mimic models, whose objective functions were the minimum MSE. However, the settings of the adaptive neuro-fuzzy inference system (ANFIS, number 13), multilayer perceptron (number 14), and radial basis function (GA-RBF, number 15) follow those in previous studies [18]–[20].

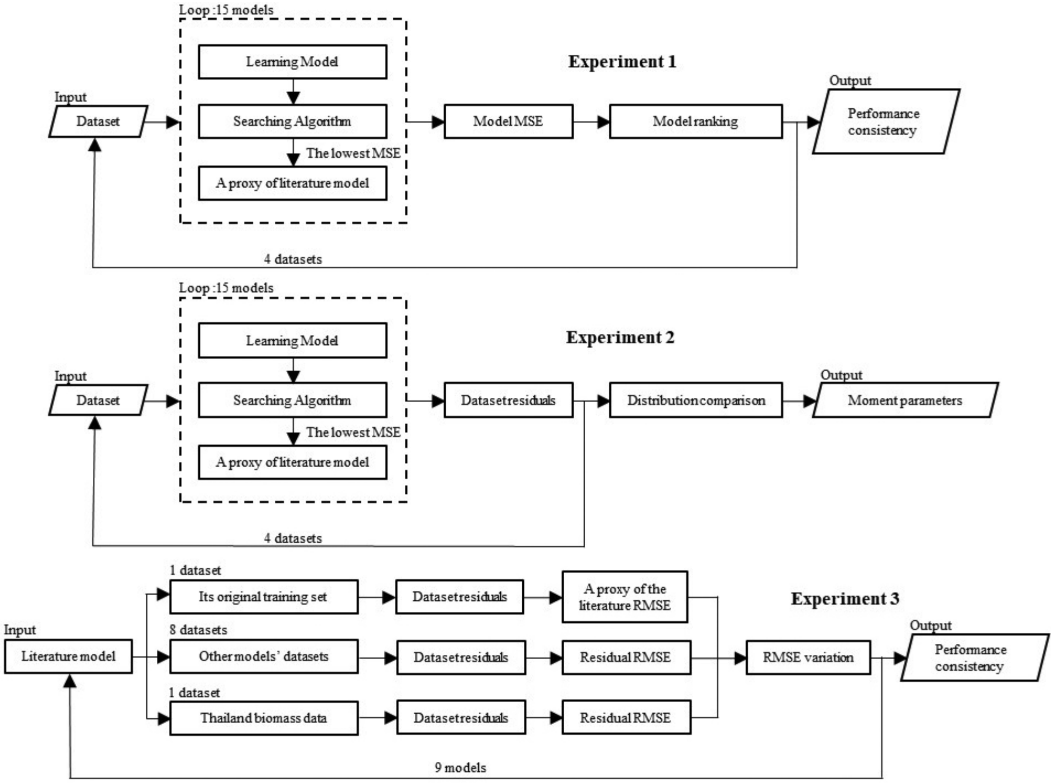

Figure 1 Schematic flowchart representing the experiment to test the performance variation for the out-of-sample data.

Table 1 Parameter settings for the biomass HHV models implemented in the 1st and 2nd experiments

No.	Name	Parameter	Proxy
1.	Stepwise regression	Quadratic	[21], [22]
2.	Robust regression	Quadratic	[23]
3.	Higher-Dimensional linear model	Quadratic	-
4.	Generalized linear model (GLM)	Quadratic	-
5.	Partial least square	Linear	-
6.	Ridge regression	Optimal λ, Quadratic	-
7.	Lasso regression	Optimal λ, Quadratic	-
8.	Lasso GLM	Optimal λ, Quadratic	-
9.	Support vector machine (SVM)	OptimizeHyperparameter	[24]
10.	Gaussian process regression (GP)	OptimizeHyperparameter	-
11.	Ensemble tree	OptimizeHyperparameter	[25]
12.	Regression tree	OptimizeHyperparameter	[26]
13.	ANFIS	ANFIS-SC5	[18]
14.	Multilayer perceptron	MLP-II	[19]
15.	Radial basis function	GA-RBF	[20]

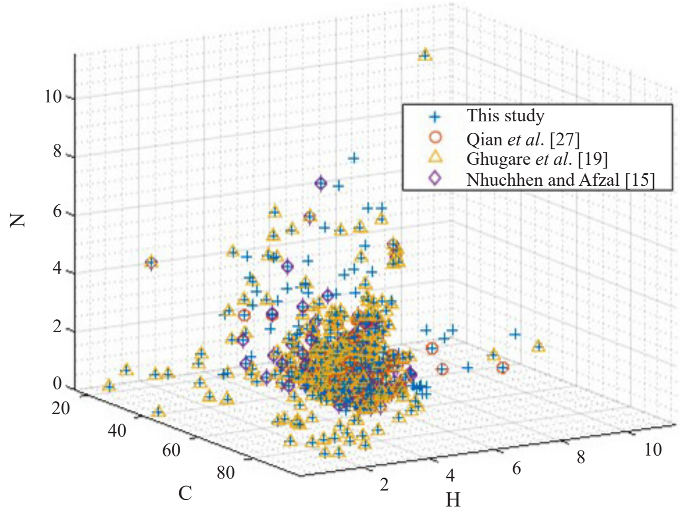

Figure 2 Plot illustrating the datasets of the 1st and 2nd experiments.

This study set the range of λ between 0.01 and 1,000 and then searched for the optimal value.

The accuracy consistency of the models was verified through the three experiments illustrated in Figures 1. Initially, the datasets considered the independent variables assigned to the models in Table 1. Then, three methodologies distinctly inspected the consistency.

The first experiment returned the MSE rankings of the dataset. Ranking similarity implied performance consistency. The second experiment produced the residual distributions. Distribution equivalence indicated no differences in accuracy.

The last experiment examined the $RMSE$ variation in the out-of-sample data. It simulated the situation when implementing the HHV model to the out-of-sample data. However, earlier studies reported the accuracy through various measurements. The present study used their models and an original training set to generate an $RMSE$ as a proxy for the one reported in earlier studies. Then, the $RMSE$ was computed for eight other datasets and a Thailand biomass dataset to check the variability.

2.3 Experimental Datasets and Models

Figure 2 describes the four datasets involved in the 1st and 2nd experiments. The first was the dataset used by Qian *et al.* [27], containing 78 raw datapoints for proximate and ultimate analyses. This dataset was a sub-dataset of another dataset published previously [28]. The original dataset had 86 records. The dataset from Qian *et al.* [27] is relatively compact and has been used in many studies [24], [29], [30]. The second involved 536 raw datapoints published by Ghugare *et al.* [19]. These data were collected from an available biomass database to train a biomass HHV model. Next was a collection of published 206 raw datapoints from Nhuchhen and Afzal [15] who used it to train their HHV model after torrefaction; all three datasets present values in percentage of dry basis.

The last dataset was the dataset of an ultimate analysis study [31]. It comprises 908 raw dry datapoints from 22 published works. Duplicate and mismatched records were cleaned by treating the records published earlier as the correct data. Therefore, the dataset used in this experiment included the three datasets described above and new records from the published work. Since the dataset used in this study spread throughout the domain, it is suitable to support biomass data.

Table 2 displays nine models and ten datasets involved in the 3rd experiment. It included a Thailand biomass sample and six additional earlier studies from the past decade stacked with the three datasets mentioned earlier.

3. Results and Discussion

3.1 Ranking Inconsistency

Table 3 ascendingly lists the five most accurate prediction models according to the 20-fold cross-validation. They indicate the most precise prediction models for the four datasets. The table also indicates that, in each treatment, the prediction model with the highest-accuracy was different. Therefore, there was no model with the highest accuracy for all datasets. Consequently, the model prediction power was inconsistent depending on the treatment.

Most of the models listed in Table 3 are nonlinear models. Therefore, the HHV models were nonlinear. Stepwise linear regression performed moderately well according to Qian *et al.* [27] because this dataset has a sample size of only 78, compared with the sample size of 908 in this study. Figure 3 and Table 4 indicate the reason for the linear model not providing a Unique Minimum Variance Unbiased Estimator (UMVUE).

Table 2 Models and datasets for the performance variant examination in the 3rd experiment

Model	HHV Model	Dataset		
		Sample Size	Year	Source
I	$0.3399C + 0.2590H$	170	2020	[32]
II	$32.9C + 162.7H - 16.2O - 954.4S + 1.408$	33	2020	[21]
III	$0.2328C + 6.9703$	171	2019	[9]
IV	$-4.9140 + 0.2611N + 0.4114C$	39	2018	[33]
V	$32.7934 + 0.0053C^2 - 0.5321C$	206	2017	[15]
VI	$847.08\left(\frac{C}{3} + H\right)$	78	2016	[27]
VII	$0.367C + \frac{53.883O}{2\,131C^2 - 93\,299} + \frac{CH - 115.971}{10.472H + 0.129CO} - \frac{91.531}{35.299 + N} + \frac{232.698}{77.545 + S}$	563	2014	[19]
VIII	$0.2949C + 0.8250H$	53	2011	[10]
IX	$430.2C - 186.7H - 127.4N + 178.6S + 184.2O - 2379.9$	20	2005	[8]
Thailand Biomass	-	7	2006	[34]

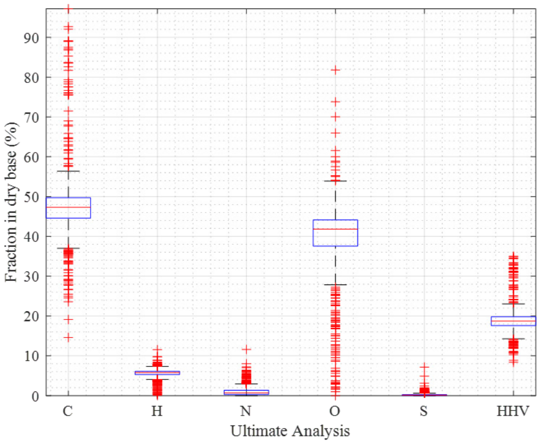

Figure 3 Box plot of the article biomass data; carbon (C), hydrogen (H), nitrogen (N), oxygen (O), sulfur (S), and higher heating value (HHV).

Table 3 Top five ranking prediction models for each dataset based on 20-fold cross-validation

This Study	Ghugare *et al.* [19]	Nhuchhen and Afzal [15]	Qian *et al.* [27]
Lasso (***1.6747***)	GP (***1.3879***)	Lasso GLM (***0.9022***)	PLS (***0.0001***)
Lasso GLM (1.9435)	Lasso (1.4704)	Lasso (0.9358)	GP (0.0001)
Ridge (2.062)	Ridge (1.4745)	Ridge (0.97854)	Stepwise (0.0001)
Ensemble tree (2.0775)	Ensemble tree (1.5799)	GP (0.99576)	Robust (0.0001)
SVM (2.1303)	GLM (1.6867)	PLS (1.03888)	Ridge (0.0002)

Note: Values in the parenthesis are the MSE. The values in italics and bold are the minimum values of the MSE in each dataset.

Figure 3 shows box plots of the ultimate analysis for the biomass data used in this study. All predictors (C, H, N, O, and S) and response (HHV) variables did not present normal distributions. Therefore, the residuals from the linear combination could not have a normal distribution.

Table 4 tabulates the parameter correlation matrix. Collinearities existed among the features in the ultimate analysis. These collinearities may decrease or increase the regression coefficients and their variances asymmetrically [35]. Therefore, hypotheses tested using the regression coefficient are likely to be biased.

Table 4 Correlation matrix of the parameters in the ultimate analysis of the study dataset

	C	H	N	O	S	HHV
C	1.0000	–0.1555	–0.1907	–0.5196	0.0015	0.8733
H	–0.1555	1.0000	–0.0265	0.4455	–0.1716	–0.0707
N	–0.1907	–0.0265	1.0000	–0.2802	0.2613	–0.1241
O	–0.5196	0.4455	–0.2802	1.0000	–0.2909	–0.4510
S	0.0015	–0.1716	0.2613	–0.2909	1.0000	0.0449
HHV	0.8733	–0.0707	–0.1241	–0.4510	0.0449	1.0000

Correlation coefficients exhibit the level of a linear relationship between two variables. The correlation coefficient between H and HHV was relatively low (–0.0707). Therefore, the linear relationship between these two parameters is weak. The association of biomass HHV was nonlinear. The correlation matrix in Table 4 also supports the nonlinear relationship reported previously [19], [36]. Additionally, the evidence of non-linearity contrasts with the linear model reported in earlier studies [10], [18], [36]. Therefore, in the ultimate analysis approach, the classical linear regression could not provide the UMVUE of the biomass HHV model.

3.2 Residual Distribution Dissimilarity

The influence of datasets on the model prediction power can be inspected through the

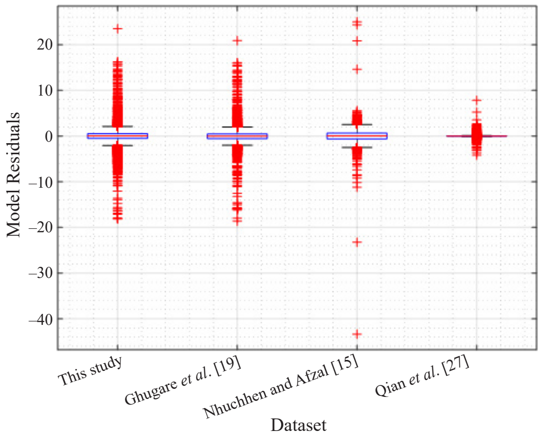

Figure 4 Box plots of the residual distribution generated from 4 treatments.

residual boxplots in Figure 4. Unfortunately, the severity in outliers indicated that none of the residuals had a normal distribution. As the spread of the outliers was different, the treatment residuals had distinct variances. Therefore, the experiment indicated the residual non-normality and variance heteroscedasticity.

Table 5 presents the descriptive statistics of the model residuals up to the fourth moment. It presents the number of data, mean, median, skewness, kurtosis, and standard error (S.E.). The confidence interval (CI) is referred to as the mean interval. The results show that the residuals in all datasets have a statistically zero mean at α = 0.01 and an equal median. However, the skew is non-zero, and kurtosis is more significant than 3.0. A high degree of kurtosis causes a fat tail distribution, which is a risk factor for financial investments. This distribution indicates a high probability that the actual HHV will differ from the predicted value.

The experimental models were conditioned on the given learning algorithms and training data. The results indicate that if the models are accurate for these conditions, then the average errors will not different statistically. However, the treatments affected the higher moment of the residual distribution.

3.3 Out-of-sample Performance Variability

Table 6 reports the performance variability between the in-sample and out-of-sample data examined in the 3rd experiment. Table 2 presents the experimental models and datasets.

The results show that the accuracy is volatile when implementing the models on the out-of-sample data. Additionally, the models in the literature are not accurate for Thai biomass. Therefore, the experimental results indicate the necessity of creating a Thai biomass model [8]. The supplementary material contains the experimental results on the MSE, MAE, AAE, ABE, R^2, and the Kolmogorov–Smirnov test.

Table 5 Descriptive statistics of the model residuals in each dataset at α = 0.01

Dataset	Count	Mean	Median	Skewness	Kurtosis	S.E.	Confidence Interval (CI)	
This Study	13,620	−0.0305	0.0458	0.1541	25.5470	0.0146	−0.0681	0.0070
Ghugare *et al.* [19]	8,040	−0.0241	−0.0410	0.5913	34.2030	0.0191	−0.0735	0.0252
Nhuchhen and Afzal [15]	3,090	−0.0154	0.0230	−3.6176	180.59000	0.0310	−0.0952	0.0644
Qian *et al.* [27]	1,170	0.0214	0.0008	2.3558	49.6550	0.0166	−0.0215	0.0642

Table 6 Variation in the model RMSE for the 3rd experiment on the entire sample data

RMSE	Dataset									
Model	I	II	III	IV	V	VI	VII	VIII	IX	Thai Biomass
I	**2.0544**	2.3605	1.9548	1.4095	2.2698	0.7384	1.6528	1.1757	1.5404	2.5352
II	4.8129	**1.4999**	5.5912	4.8214	3.0600	2.0984	5.8466	2.3816	11.1104	4.3861
III	1.9605	4.3366	**1.6994**	1.6249	3.0289	1.2408	2.0561	1.4952	2.7263	2.8605
IV	2.4765	1.2479	2.5629	**0.4218**	1.8251	0.5496	1.5973	2.3465	2.3355	2.9690
V	2.4490	1.5919	2.4193	1.2581	**1.6360**	0.8173	1.8299	2.3859	10.1084	3.8905
VI	2.2675	4.4418	2.2381	0.7379	2.4430	**0.1694**	1.8319	1.0063	0.9755	3.0934
VII	2.0567	0.7731	2.2862	0.9093	1.9108	0.4353	**1.0660**	1.1034	2.6735	3.2024
VIII	2.2095	4.3160	2.1697	0.8118	2.4539	0.1785	1.8069	**0.9878**	1.0397	2.9956
IX	7.0166	2.0597	7.3475	5.0215	4.7715	5.4259	5.6177	5.5348	**0.6593**	7.2395

Note: The bold values refer to the in-sample performance.

Table 7 Performance inconsistency among the reported and re-evaluated values collected from the literature

No.	Originally Reported						Boumanchar *et al.* [9]			This Study Dataset					
	MSE	RMSE	MAE	AAE	ABE	R^2	AAE	ABE	$\rho_{HHV,\widehat{HHV}}$	MSE	RMSE	MAE	AAE	ABE	R^2
1.						0.7529*	8.2162	6.1784	0.5292	2.8955	2.8955	1.0005	5.4722	1.0834	0.7023
2.	0.0270	0.1630		0.7930		0.9890	7.8195	4.8710	0.4574	3.8139	1.9529	1.1011	6.1132	-0.0709	0.6079
3.	0.0230	0.1520		0.7290		0.9907	21.6091	21.3792	0.4789	4.5675	2.1372	1.1884	6.5293	0.8501	0.5304
4.	0.0210	0.1460		0.7190		0.9917	7.6070	2.1376	0.4271	5.0209	2.2407	1.2501	6.9498	1.6248	0.4838
5.	0.0210	0.1460		0.7600		0.9915	7.8014	1.3687	0.4187	5.4473	2.3339	1.3136	7.3118	2.1383	0.4399
6.	0.0250	0.1590		0.8250		0.9896	8.6486	5.6098	0.4387	12.8125	3.5795	1.6621	7.3206	-2.1333	-0.3173
7.	0.0220	0.1500		0.7300		0.9910	7.6964	3.9647	0.4443	36.9891	6.0819	1.4357	6.5461	-0.1908	-2.8030
8.				8.5700	1.1400	0.7340	12.3974	12.0221	0.4362	3.9082	1.9769	1.3510	7.0209	-3.5111	0.5982
9.			1.2000	5.3300	1.0000		12.6002	-9.7611	0.4574	8.5571	2.9252	2.6332	16.5784	15.8988	0.1202
10.			1.0610	5.3100	1.1900		7.0133	2.5392	0.4487	2.5063	1.5831	1.0238	5.5408	1.7348	0.7423
11.			0.8400	5.4400	0.6200		7.2522	2.2382	0.4308	4.3959	2.0967	1.4602	8.1904	6.1484	0.5480
12.			0.5580	5.9600	0.5000		7.4715	0.5270	0.4943	7.7760	2.7885	1.8419	9.4442	0.9167	0.2005

*Computed in this article. The literature were incorrectly computed by setting “Constant as zero” during data analysis in Excel.

1. HHV = 0.2949C + 0.8250H [10]	7. HHV = −5.290 + 0.493C + 5.052/H [11]
2. HHV = −3.147 + 0.468C [11]	8. HHV = 0.3198C + 0.0803O + 0.4704N − 1.4502S + 0.9364 [12]
3. HHV = −2.907C + 0.491C + 0.261H [11]	9. HHV = (338.4C + 244.2)/1000 [13]
4. HHV = −3.393 + 0.507C − 0.341H + 0.067N [11]	10. HHV = ($1.59C^2$ + 154.5C + 7464)/1000 [13]
5. HHV = −3.440 + 0.517(C + N) − 0.433(H + N) [11]	11. HHV = (303.81C + 81.62O − 490.68S + 159.92)/1000 [13]
6. HHV = 5.736 + $0.006C^2$ [11]	12. HHV = (−150.6O + 24660)/1000 [13]

Comparisons with the earlier results in the literature support the experimental results of the study. Table 7 compares the performance of the models from the three sources; the original report, evaluation by Boumanchar *et al.* [9], and the evaluations in this study. The initially reported values are representatives of the in-sample performance. The table indicates that AAEs and RMSEs were different among the assessments. Therefore, the experiments of the study clearly indicate the volatile accuracy for out-of-sample data [9].

The experimental models listed in Tables 6 and 7 are not the proxies of the highest accurate model for the training sets; the error averages were thus distinct. Even though the models were precise for their training sets, the out-of-sample accuracy could not be guaranteed. Therefore, retraining the model is recommended to ensure statistical accuracy.

4. Conclusions

The experiments in this study indicate that

1) Performance inconsistencies arise in the biomass HHV models for ultimate analysis.

2) The models cannot maintain the *MSE, RMSE, MAE, AAE, ABE,* and R^2 when using out-of-sample data

3) Retraining the model using a practical dataset is recommended before implementation.

Additionally, the present study presents evidence of the accuracy degradation of nonlinear biomass HHV models. Other contributions include the procedures to inspect the accuracy consistency.

The experiments assumed the models with the lowest MSE constructed from the random parameter searching would have an accuracy equivalent to those of earlier models. Furthermore, the study was limited to the minimum MSE objective function for the ultimate analysis. Further research to construct a domestic dataset is essential to enhance the accuracy of ultimate analysis models and promote the use of biomass energy to mitigate climate change and reduce the usage of fossil fuel.

References

[1] R. K. P. and L. A. M. (eds. Core Writing Team), *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Geneva, Switzerland, 2015.

[2] S. Jenniches, “Assessing the regional economic impacts of renewable energy sources - A literature review,” *Renewable Sustainable Energy Reviews*, vol. 93, pp. 35–51, 2018.

[3] AEDP (Department of Alternative Energy Development and Efficiency). (2020, October). The 10-Year Alternative Energy Development Plan. Ministre of Energy. Bangkok, Thailand. [Online]. (in Thai). Available: http://www.eppo.go.th/images/Infromation_service/public_relations/PDP2018/PDP2018Rev1.pdf

[4] A. K. Gupta and S. Kerdsuwan, “Efficient energy conversion of wastes and fuels in power systems,” *KMUTNB International Journal of Applied Science Technology*, vol. 7, no. 2, pp. 1–26, 2014.

[5] C. Areepak, T. Jiradechakorn, S. Chuetor, V. Champreda, and N. Laosiripojana, “Improvement of lignocellulosic pretreatment efficiency by combined chemo - Mechanical pretreatment for energy consumption reduction and biofuel production,” *Renewable Energy*, vol. 182, pp. 1094–1102, 2022.

[6] I. Estiati, F. B. Freire, J. T. Freire, R. Aguado, and M. Olazar, “Fitting performance of artificial neural networks and empirical correlations to estimate higher heating values of biomass,” *Fuel*, vol. 180, pp. 377–383, 2016.

[7] J. Xing, K. Luo, H. Wang, Z. Gao, and J. Fan, "A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analyses with machine learning approaches," *Energy*, vol. 188, pp. 116077, 2019.

[8] P. Thipkhunthod, V. Meeyoo, P. Rangsunvigit, B. Kitiyanan, K. Siemanond, and T. Rirksomboon, "Predicting the heating value of sewage sludges in Thailand from proximate and ultimate analyses," *Fuel*, vol. 84, no. 7–8, pp. 849–857, 2005..

[9] I. Boumanchar, K. Charafeddine, Y. Chhiti, F. E. M'hamdi Alaoui, A. Sahibed-dine, F. Bentiss, C. Jama, and M. Bensitel, "Biomass higher heating value prediction from ultimate analysis using multiple regression and genetic programming," *Biomass Conversion and Biorefinery*, vol. 9, no. 3, pp. 499–509, 2019.

[10] C.-Yang Yin, "Prediction of higher heating values of biomass from proximate and ultimate analyses," *Fuel*, vol. 90, no. 3, pp. 1128–1132, 2011.

[11] A. J. Callejón-Ferre, B. Velázquez-Martí, J. A. López-Martínez, and F. Manzano-Agugliaro, "Greenhouse crop residues: Energy potential and models for the prediction of their higher heating value," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 2, pp. 948–955, 2011.

[12] H. L. Choi, S. I. A. Sudiarto, and A. Renggaman, "Prediction of livestock manure and mixture higher heating value based on fundamental analysis," *Fuel*, vol. 116, pp. 772–780, 2014.

[13] R. García, C. Pizarro, A. G. Lavín, and J. L. Bueno, "Spanish biofuels heating value estimation. Part I: Ultimate analysis data," *Fuel*, vol. 117, no. PARTB, pp. 1130–1138, 2014.

[14] X. Wen, S. Jian, and J. Wang, "Prediction models of calorific value of coal based on wavelet neural networks," *Fuel*, vol. 199, pp. 512–522, 2017.

[15] D. R. Nhuchhen and M. T. Afzal, "HHV predicting correlations for torrefied biomass using proximate and ultimate analyses," *Bioengineering*, vol. 4, no. 1, 2017.

[16] Z. Ceylan, E. Pekel, S. Ceylan, and S. Bulkan, "Biomass higher heating value prediction analysis by ANFIS, PSO-ANFIS and GA-ANFIS," *Global Nest Journal*, vol. 20, no. 3, pp. 589–597, 2018.

[17] A. Kijkarncharoensin. (2022, January). *Performance inconsistency of the Biomass Higher Heating Value (HHV) Models derived from Ultimate Analysis*. [Online]. Available: https://codeocean.com/capsule/9079586/tree/v1

[18] E. Akkaya, "ANFIS based prediction model for biomass heating value using proximate analysis components," *Fuel*, vol. 180, pp. 687–693, 2016.

[19] S. B. Ghugare, S. Tiwary, V. Elangovan, and S. S. Tambe, "Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms," *Bioenergy Research*, vol. 7, no. 2, pp. 681–692, 2014.

[20] A. Dashti, A. S. Noushabadi, M. Raji, A. Razmi, S. Ceylan, and A. H. Mohammadi, "Estimation of biomass higher heating value (HHV) based on the proximate analysis: Smart modeling and

correlation," *Fuel*, vol. 257, pp. 115931, 2019.

[21] C. Qian, Q. Li, Z. Zhang, X. Wang, J. Hu, and W. Cao, "Prediction of higher heating values of biochar from proximate and ultimate analyses," *Fuel*, vol. 265, pp. 116925, 2020.

[22] C. Telmo, J. Lousada, and N. Moreira, "Proximate analysis, backwards stepwise regression between gross calorific value, ultimate and chemical analysis of wood," *Bioresource Technology*, vol. 101, no. 11, pp. 3808–3815, 2010.

[23] R. A. Ibikunle, A. F. Lukman, I. F. Titiladunayo, E. A. Akeju, and S. O. Dahunsi, "Modeling and robust prediction of high heating values of municipal solid waste based on ultimate analysis," *Energy Sources, Part A: Recover Utilization Environmental Effects*, 2020.

[24] M. Duan, Z. Liu, D. Yan, W. Peng, and A. Baghban, "Application of LSSVM algorithm for estimating higher heating value of biomass based on ultimate analysis," *Energy Sources, Part A: Recover Utilization Environmental Effects*, vol. 40, no. 6, pp. 709–715, 2018.

[25] S. H. Samadi, B. Ghobadian, and M. Nosrati, "Prediction of higher heating value of biomass materials based on proximate analysis using gradient boosted regression trees method," *Energy Sources, Part A: Recover Utilization Environmental Effects*, vol. 43, no. 6, pp. 672–681, 2021.

[26] F. Vallejo, L. A. Díaz-Robles, R. Vega, and F. Cubillos, "A novel approach for prediction of mass yield and higher calorific value of hydrothermal carbonization by a robust multilinear model and regression trees," *Journal of the Energy Institute*, vol. 93, no. 4, pp. 1755–1762, 2020.

[27] H. Qian, X. Guo, S. Fan, K. Hagos, X. Lu, C. Liu, and D. Huang, "A simple prediction model for higher heat value of biomass," *Journal of Chemical and Engineering Data*, vol. 61, no. 12, pp. 4039–4045, 2016.

[28] S. V. Vassilev, D. Baxter, L. K. Andersen, and C. G. Vassileva, "An overview of the chemical composition of biomass," *Fuel*, vol. 89, no. 5, pp. 913–933, 2010.

[29] A. Darvishan, H. Bakhshi, M. Madadkhani, M. Mir, and A. Bemani, "Application of MLP-ANN as a novel predictive method for prediction of the higher heating value of biomass in terms of ultimate analysis," *Energy Sources, Part A: Recover Utilization Environmental Effects*, vol. 40, no. 24, pp. 2960–2966, 2018.

[30] J. O. Ighalo, A. G. Adeniyi, and G. Marques, "Application of linear regression algorithm and stochastic gradient descent in a machine-learning environment for predicting biomass higher heating value," *Biofuels, Bioproducts and Biorefining*, vol. 14, no. 6, pp. 1286–1295, 2020.

[31] A. Kijkarncharoensin. (2022, January). *Ultimate_Analysis*. [Online]. Available: https://data.mendeley.com/datasets/t8x96g88p3/2 (accessed Jan. 28, 2022).

[32] A. O. Balogun, O. A. Lasode, and A. G. McDonald, "Higher heating value prediction model from proximate and ultimate analysis data," *Journal of Multidisciplinary Engineering Science and Technology*, vol. 7, no. 8, 2020.

[33] A. Ozyuguran, A. Akturk, and S. Yaman, "Optimal use of condensed parameters of ultimate

analysis to predict the calorific value of biomass," *Fuel*, vol. 214, pp. 640–646, 2018.

[34] S. Garivait, U. Chaiyo, S. Patumsawad, and J. Deakhuntod, "Physical and chemical properties of thai biomass fuels from agricultural residues," *The 2nd Joint International Conference on "Sustainable Energy and Environment (SEE 2006)"* 2006, pp. 48–51.

[35] C. F. Mela and P. K. Kopalle, "The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations," *Applied Economics*, vol. 34, no. 6, pp. 667–677, 2002.

[36] S. U. Patel, B. Kumar, Y. Badhe, B. Sharma, S. Saha, S. Biswas, A. Chaudhury, S. Tambe, and B. Kulkarni, "Estimation of gross calorific value of coals using artificial neural networks," *Fuel*, vol. 86, no. 3, pp. 334–344, 2007.