

## Article

# Biomass Higher Heating Value Estimation: A Comparative Analysis of Machine Learning Models

Ivan Brandić <sup>1</sup>, Lato Pezo <sup>2</sup>, Neven Voća <sup>1</sup> and Ana Matin <sup>1,\*</sup>

<sup>1</sup> Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia; nvoca@agr.hr (N.V.)

<sup>2</sup> Institute of General and Physical Chemistry, University of Belgrade, Studentski trg 12/V, 11000 Belgrade, Serbia

\* Correspondence: amatin@agr.hr

**Abstract:** The research conducted focused on the capabilities of various non-linear and machine learning (ML) models in estimating the higher heating value (HHV) of biomass using proximate analysis data as inputs. The research was carried out to identify the most appropriate model for the estimation of HHV, which was determined by a statistical analysis of the modeling error. In this sense, artificial neural networks (ANNs), support vector machine (SVM), random forest regression (RFR), and higher-degree polynomial models were compared. After statistical analysis of the modeling error, the ANN model was found to be the most suitable for estimating the HHV biomass and showed the highest specific regression coefficient, with an  $R^2$  of 0.92. SVM ( $R^2 = 0.81$ ), RFR, and polynomial models ( $R^2 = 0.84$ ), on the other hand, also exhibit a high degree of estimation, albeit with somewhat larger modelling errors. The study conducted suggests that ANN models are best suited for the non-linear modeling of HHV of biomass, as they can generalize and search for links between input and output data that are more robust but also more complex in structure.

**Keywords:** energy properties; biomass; machine learning; artificial neural networks; support vector machines; random forest regression; polynomials; higher heating value



**Citation:** Brandić, I.; Pezo, L.; Voća, N.; Matin, A. Biomass Higher Heating Value Estimation: A Comparative Analysis of Machine Learning Models. *Energies* **2024**, *17*, 2137. <https://doi.org/10.3390/en17092137>

Academic Editors: Anna Kozuch, Krzysztof Adamowicz and Miloš Gejdoš

Received: 20 March 2024

Revised: 24 April 2024

Accepted: 29 April 2024

Published: 30 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Renewable energy sources are gaining more attention, especially when they are used in sectors that want a stable and sustainable energy supply system. One of the most important of these sources is biomass, which includes organic matter of plant or animal origin that can be used directly as fuel or useful energy [1]. The basic products that can be derived from biomass include various feedstocks, fuels for transportation and the production of heat (energy) through direct combustion [2]. Biofuels derived from the biomass of agricultural and forestry production or from the cultivation of energy crops are considered promising sources for replacing existing conventional fuel sources and thus influencing the reduction in greenhouse gases and the increase in negative climate change [3]. For this biomass to be utilized as a fuel source through various conversions, the calorific value must be determined, which is the most important parameter in quality assessment but also promotes the use of feedstock as a fuel source [4].

The higher heating value is used in the design of energy systems of different sizes and is considered an extremely important parameter for the processes [5]. Aghel et al. (2023) [6] state that the relationship between the variables of biomass proximate analysis is not linear, so nonlinear modeling is a better alternative in finding a solution. In addition to the analysis of the HHV as a measure of fuel quality, proximate analysis can provide a detailed insight into the physicochemical composition of the biomass, which is of crucial importance for further modeling. Proximate analysis provides an insight into the physicochemical composition of the feedstock, considering the proportion of the main components of fixed carbon (FC), volatile matter (VM), and ash [4]. Velázquez Martí et al. (2023) [7]

state that proximate analysis is one of the least economically demanding (laboratory) analyses for the determination and characterization of biomass as a fuel. For this reason, the datasets of the above analysis are suitable for modeling and processing in a machine learning (ML) environment. Understanding proximate analysis enables the use of advanced analysis techniques, such as the application of deep learning in modeling. In the field of modeling, the application of deep learning (DL) is becoming increasingly popular as an ML approach, which achieves significantly better results in modeling compared to less complex models and whose main feature is learning from data [8]. ML models should be used when it comes to understanding the interaction and “intelligent” analysis of large datasets. ML algorithms can be supervised, semi-supervised, unsupervised, and reinforcement learning [9]. Using ML models, it is possible to develop more accurate and reliable predictive models that are used as regression models for calculating the desired output values [10]. Classical ML models such as artificial neural networks (ANNs), random forest regression (RFR), support vector machines (SVMs), extreme learning machines (ELMs), K-nearest neighbors (KNNs), and decision trees (DTs) find extensive use across various scientific domains for modeling purposes. Among these, SVM stands out as a widely employed discriminant technique rooted in statistical learning theory, renowned for its robust generalization capabilities. Achieving the optimal network involves striking a balance between model complexity and training error, as highlighted by Ma et al. (2022) [11]. ELM, on the other hand, constructs a single-layer feedforward network through the random generation of input weights and biases for hidden layers, as elucidated by Wang et al. (2022) [12]. For sequence data, a plethora of state-of-the-art machine learning techniques, including ensemble learning models like XGBoost [13], LightGBM [14], and CatBoost, offer viable options. XGBoost excels in prediction accuracy and interpretability, particularly for high-dimensional datasets. Conversely, LightGBM accommodates large datasets and facilitates GPU training, demonstrating superior accuracy and speed compared to XGBoost. Data fusion further enhances forecasting accuracy by integrating gradient boosting-based categorical attributes, a capability supported by the CatBoost algorithm, as discussed by Dutta and Roy (2022) [15].

More specialized machine learning methods, such as artificial neural networks (ANNs), enable more complex analyses and predictions in the energy sector [16]. ANNs, as one of the ML techniques, can be successfully used for predictions and work on the principle of the human brain. This enables the processing of a large amount of data and the analysis of more complex relationships [17]. Although ANN models have numerous advantages in modeling, the optimization of the model is particularly challenging due to the complexity of the structure [18], but due to the ability to generalize and implement multidimensional nonlinearity, the system can also be used in real-time predictions [19].

Besides ANN models that handle complex data, random forest regression (RFR) appears to be a simpler but effective alternative in regression modeling and is recognized as one of the most effective tools in the field of regression modeling. Due to its simplicity, it is easier to optimize, which facilitates the learning process of models and predictions [20]. Despite its practical application, RFR shows high efficiency compared to other ML models [21]. Although RFR offers simplicity and efficiency, support vector machines (SVMs) represent a different approach that focuses on data classification in energy modeling [22]. SVM is a method for classifying data into two groups and searches for the hypersurface that best separates the data into two classes, positive and negative [23]. The support vector machine (SVM) method shows high performance when the edge of the class separation is clear. It is more effective in spaces with large dimensions, and it is also widely used in forecasting methods for reasons of “efficient” storage [24]. Polynomials are often used in regression models because they can approximate a variety of functions and describe complex patterns in data. In the context of regression, polynomials enable model flexibility as they can model non-linear relationships between variables [25]. The application of polynomial regression is particularly useful in situations where simple linear models do not provide sufficiently

accurate predictions, making polynomial regression a powerful tool in data analysis and machine learning [26,27].

Qian et al. (2018) [28] developed different regression models in their research for predicting the calorific value based on the inputs of proximal analysis. The authors find that the most accurate model contains a combination of linear and polynomial expressions and interaction effects. The developed method showed a high degree of accuracy ( $R^2 > 0.91$ ). In the study conducted by Aghel et al. (2023) [6], a dynamic Elman recurrent neural network (ENN) was developed to predict biomass HHV based on the input data from the terminal and proximate analysis. It was found that the ENN model built from a hidden layer with four nodes was the most accurate and had a relatively low modeling error. Matveeva and Bychkov (2022) [29] highlight the effectiveness of using MLP ANNs for datasets with pronounced heterogeneity, achieving an  $R^2$  value of  $0.880 \pm 0.025$ .

A structure with the rectified linear unit (*ReLU*) activation function and the Adam training algorithm is recommended. It is emphasized that the amount of data and their preprocessing, including the rejection of dependent and noisy values, significantly improves the accuracy of the predictions.

This research aims to build several ML models (ANNs, RFR, and SVM) and a higher order polynomial regression model to determine the possibility of estimating the HHV value based on the biomass proximate analysis dataset. Reliability metrics such as the coefficient of determination ( $R^2$ ), root mean square error (*RMSE*), and mean absolute error (*MAE*) are used to compare models to determine accuracy and modeling error.

## 2. Materials and Methods

### 2.1. Data Collection and Preprocessing

Data on FC, VM, ash, and HHV were collected from the available database of scientific literature [30–34]. Data were collected in semi-structured form, and a total of 872 data points were used for modeling, with all missing data and duplicate values removed. Data processing included cleaning, normalization, coding, and feature selection to improve the data structure and remove any obstacles that could affect the model's ability to learn effectively. The collected data can be found in Supplementary Table S1.

### 2.2. Data Analysis

In the context of the ML pipeline used, data analysis involved performing descriptive statistics, visualizing the data distribution through graphs, and creating correlation diagrams to understand the relationships between variables before proceeding with feature engineering and model building. The Python programming language was used for data analysis and the creation of ML models, including the *Jupyter notebook IDE* environment with the associated software packages.

### 2.3. Model Selection

Four machine learning models were selected for modeling. An ANN was used due to its ability to model complex non-linear relationships and interactions between features [35]. The basic structure of this model consists of layers with artificial neurons (nodes) and activation functions between them. The model is configured with multiple layers to adapt to the specifics of the dataset. The ANN model was created as a sequential model with 2 layers. The first layer consisted of 10 artificial neurons and a ReLU function, while the second layer consisted of 1 neuron and a linear function. To optimize the model, the optimizer "Adam" and the mean square error (*MSE*) were used as loss functions. The model was trained with 4000 epochs and a batch size of 100, with the remaining dataset used to measure performance.

SVM is also included for robustness and efficiency in regression tasks. The model used radial basis function (RBF) to transform the data multidimensionally for easier separation (regression). The input data for the SVM model are split in a ratio of 70% for training and 30% for testing, with a "random seed" of 42. At the beginning of the pipeline, the Python

standardization function (“*StandardScaler*”) and “*GridSearchCV*” were used to automatically adjust the parameters of the model  $C$ ,  $\gamma$ , and kernel by cross-validation (5-fold). The values 1, 0.1, 0.01, and 0.001 were tried out for the  $\gamma$  value.

The RFR model uses an ensemble of multiple decision trees, which provides greater precision and stability in the estimation. The model also evaluates the importance of certain features in the modeling process, which provides additional insights during the analysis. An important hyperparameter in the random forest (RF) algorithm is the size of the feature set used to determine the optimal partitioning rule at each tree node [36]. The RFR model was set up with 200 trees, with a maximum depth of 20 per tree and a seed generator of 42 to ensure the reproducibility of the results.

Using polynomial regression, it is possible to fit data with non-linear trends by transforming features in the equation. Due to its “flexibility”, the model can be adapted to different degrees of polynomials depending on the complexity of the data. Regularized polynomials are implemented using a pipeline that integrates polynomial features up to the fourth degree together with ridge regression that applies L2 regularization with a strength parameter  $\alpha = 0.5$ , effectively controlling model complexity and preventing overtraining.

#### 2.4. Model Evaluation

The model evaluation procedure assessed the performance of the models created. The statistical metrics chosen were indicators of the coefficient of determination ( $R^2$ ) (1), which is used as a measure of how much variation in the real data can be represented by the model. The root mean square error (RMSE) (2) was used to measure the mean square error of the prediction. The mean absolute error (MAE) (3) parameter is used to calculate the absolute value of the error in the modeling. The calculated metrics were used to compare the models in terms of their predictive ability. In addition, the execution time of the “execution code” was measured for all models to determine the modeling speed. The formulas mentioned are presented as follows [37]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (Y - Y_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_p)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_u| \quad (3)$$

#### 2.5. Model Optimization and Hyperparameter Tuning

Model optimization included adjustments to achieve the “best possible” performance and included calibration of the algorithm, adjustments to the learning process, and changes to the model architecture. By adjusting the hyperparameters, Tziachris et al. (2020) [38] found the settings that led to the best model performance on the test set. By optimizing the hyperparameters of the model, it is possible to achieve greater efficiency in modeling [38]. The configuration of the hyperparameters included a change in the learning rate, the number of trees in the RFR model, the number of layers and artificial neurons in the ANN model, and a change in the  $\gamma$  value in the SVM model. Methods were also used to systematically search for parameters and determine the optimal combination.

### 3. Results

Figure 1 shows the distribution diagram and the descriptive statistics of the observed variables of the samples of the proximate analysis and the HHV biomass.

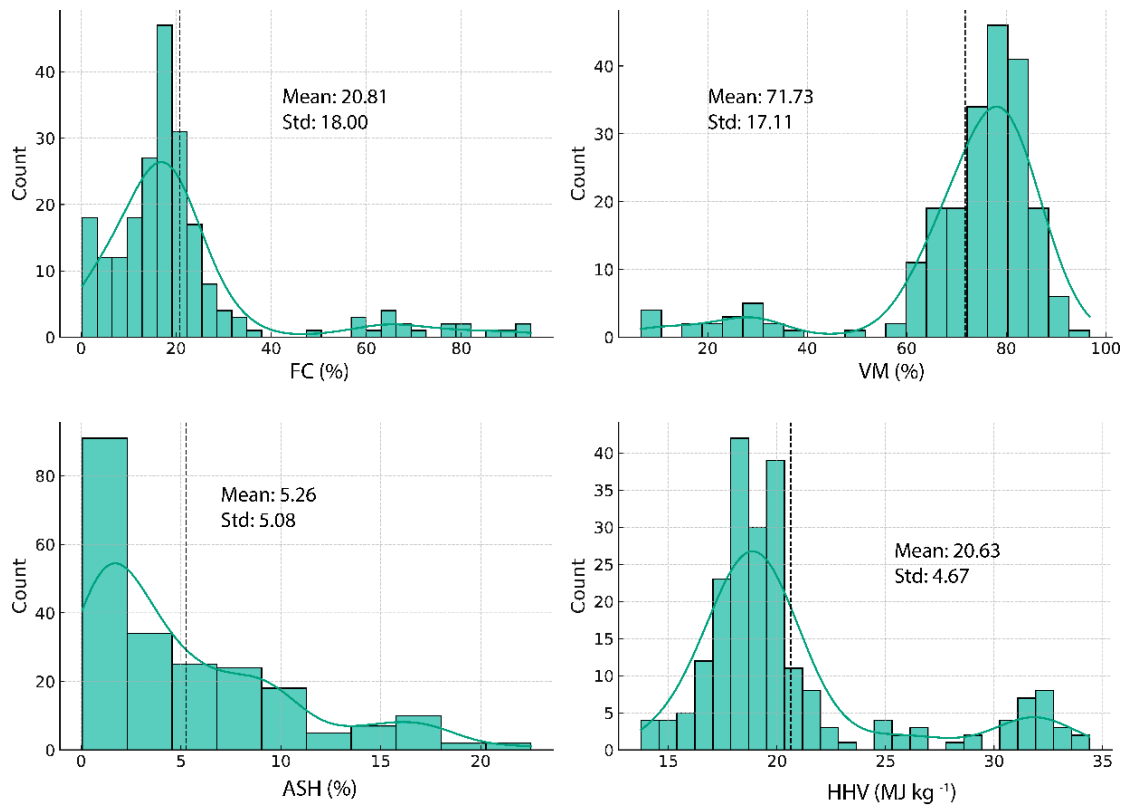


Figure 1. Distribution and descriptive statistics of the variables analyzed.

Figure 2 shows a scatter plot to determine the relationship between the input variables (FC, VM, and ash) and the output variables of the model (HHV).

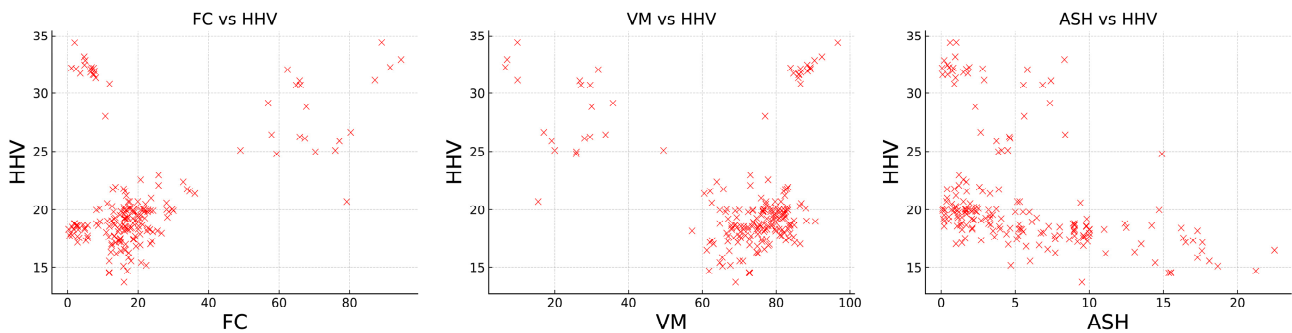
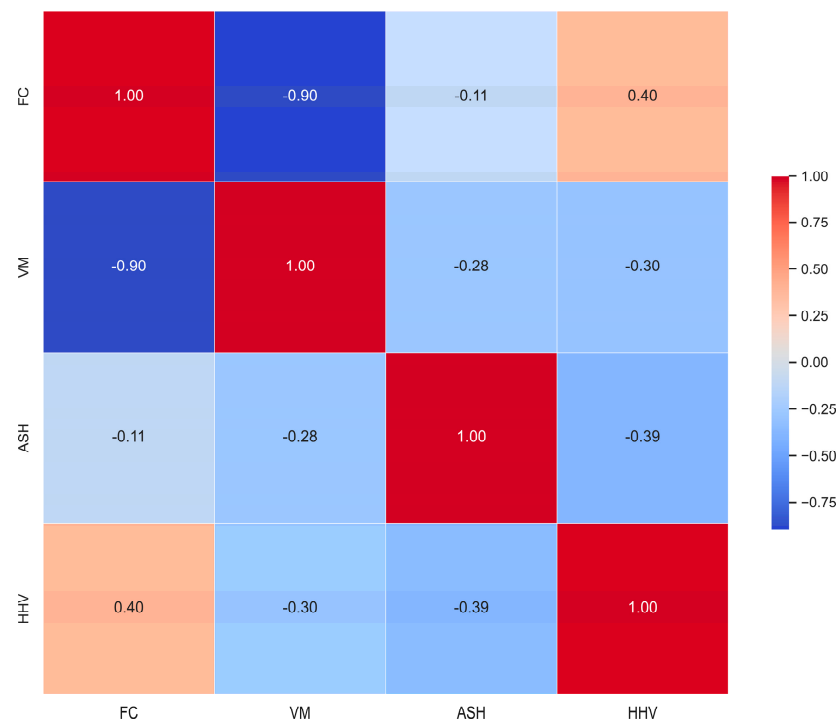


Figure 2. Scatterplot of the relationship between input and output variables in the dataset used.

Figure 3 shows the correlation matrix of the analyzed variables with the corresponding Pearson correlation coefficients of the input and output variables of the model.

The correlation matrix (Figure 3) shows the relationships (strength of the relationship according to Pearson’s correlation coefficient) between the four variables examined. For research purposes, it is particularly important to examine the correlation between the HHV and the variables of the proximate analysis.

Table 1 shows the results of the statistical analysis of the ML model in terms of modeling error (*MSE* and *MAE*) and the degree of data overlap ( $R^2$ ).



**Figure 3.** The correlation matrix of the variables in the dataset used.

**Table 1.** Performance of developed ML models used in this research.

Model Type	$R^2$	$MSE$	$MAE$
ANN	0.92	1.33	0.77
RFR	0.84	1.61	1.03
SVM	0.81	1.75	1.25
Polynomial	0.72	2.14	1.53

ANNs—artificial neural networks; RFR—random forest regression; SVM—support vector machine;  $R^2$ —coefficient of determination;  $MSE$ —mean squared error;  $MAE$ —mean average error.

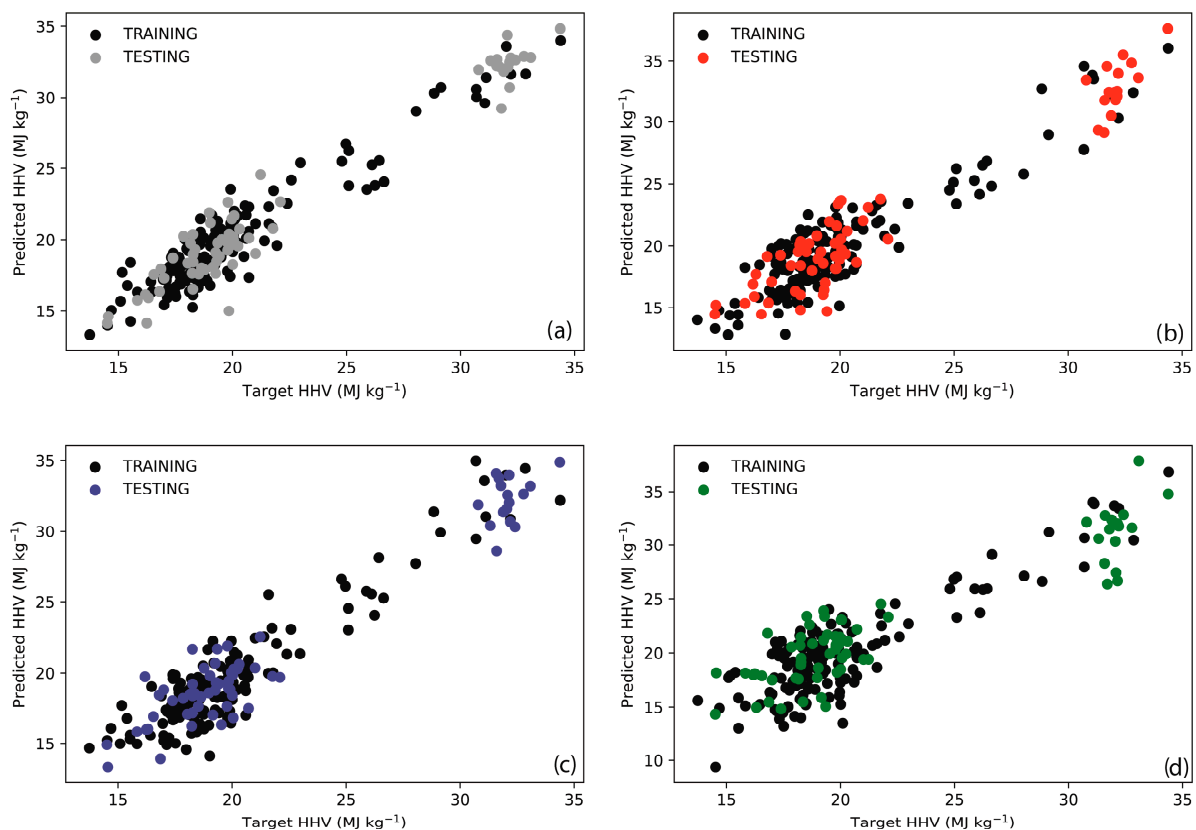
Table 2 shows the execution time required for modeling the exposed values of the models created.

**Table 2.** Execution time for developed models.

Model	Execution Time (s)
ANN	222.33
SVM	3.82
RFR	0.13
Polynomial	0.03

ANNs—artificial neural networks; SVM—support vector machine; RFR—random forest regression.

Figure 4 shows a scatter plot of the developed models in terms of regression representativeness.



**Figure 4.** Scatter plots of the overlap of actual and predicted HHV values for (a) ANNs, (b) SVM, (c) RFR, and (d) polynomial.

#### 4. Discussion

After statistically analyzing the variables needed to build the ML models, the mean values and standard deviation for FC, VM, ash, and HHV biomass were determined. The results of the above values can be seen in Figure 1. From this analysis, it can be concluded that there is considerable variability between the samples. The average value of FC is 20.81% and the standard deviation is 18.00%, indicating a wide range of values in the samples. For VM, most of the samples have a significant share, with a mean value of 71.73% and a standard deviation of 17.11%. The ash content varies with a mean value of 5.26% and a standard deviation of 5.08%, which indicates less variation compared to other measurements. For the HHV, the average is 20.63 MJ kg<sup>-1</sup> with a standard deviation of 4.67 MJ/kg, indicating moderate variability between samples. For the development of new models for the prediction of biomass components, Park et al. (2023) [39] gave the values for VM in the range of 59.07–87.3%, FC 10.44–38.64%, and ash 0.25–10.03%, indicating that the dataset collected is consistent with the range in the literature. For the HHV values, Harun et al. (2018) [40] reported values in the range of 17.58–19.15 MJ kg<sup>-1</sup>. The average HHV value in this research is slightly higher. The reason for this is that certain biomass samples (from the collected data) were subjected to different treatments that affected the change in energy value. The displayed scatter diagrams (Figure 2) illustrate the relationships between the input characteristics (FC, VM, and ash) and the HHV within the analyzed dataset. Plotting FC against HHV indicates a potential correlation, suggesting that FC content may be an important factor in predicting HHV. Plotting VM against HHV shows more scattered points, suggesting a weaker relationship between VM and HHV. The plot of variable ash with HHV also shows a scattered pattern, suggesting that ash content may not have a strong or clear influence on HHV. To develop the model, Qian et al. (2018) [28] also analyzed the relationship between individual components of the proximate analysis and the HHV. The scatter plot showed that FC has the highest coefficient of determination

( $R^2 = 0.62$ ) with HHV, while the variables ash and VM have a weaker relationship with HHV. Observing the strength of the relationship using the Pearson correlation (Figure 3), the variables ash ( $-0.39$ ) and VM ( $-0.40$ ) are negatively correlated, while the variable FC ( $0.40$ ) is positively correlated with HHV. The creation of correlation matrices [41] shows that the variables VM and FC show no significant statistical correlation with HHV, while the variables VM and FC show a strong linear relationship. The best-fit ML model was the ANN model developed with 3 artificial neurons in the input layer and 10 in the hidden layer, and the ReLU activation function was used because it is considered the most suitable activation function for neural networks [42]. A linear activation function was used to calculate the output data. The model was optimized with the Adam optimizer using the *MSE* parameter as the loss function. The SVM model created used the RBF (radial basis function) kernel. With the RBF function, it is possible to achieve better results to minimize the computer load by separating the samples from different classes [43]. The model also included a regularization parameter ( $C = 100$ ) and a gamma parameter. The RFR regression model was created with 200 estimators and a maximum depth of the trees of 20, and the model is based on the construction of multiple decision trees that calculate the output data based on the average prediction of each tree [5]. The polynomial regression was created with three levels, and power regularization was included ( $\alpha = 0.5$ ) to prevent overfitting of the model to the data. Overfitting occurs when a model learns from the training set and tests improvements using errors from the training samples, leading to a disproportionate impact on the model's performance with original data [44]. When comparing the performance of four ML models on the regression task of predicting the value of the HHV variable, the ANN achieved the best results in all metrics. An  $R^2$  value of 0.92, an *RMSE* of 1.33, and an *MAE* of 0.77 show that the ANN can predict the value of the HHV variable well. The SVM also performed well with an  $R^2$  of 0.81, an *RMSE* of 1.75, and an *MAE* of 1.25. RFR and polynomial showed satisfactory performance with an  $R^2$  of 0.84, an *RMSE* of 1.61, and an *MAE* of 1.03 and an  $R^2$  of 0.72, an *RMSE* of 2.14, and an *MAE* of 1.53, respectively. These results show that the ANN is the best model for predicting the value of the HHV variable. Aghel et al. (2023) [6] reported the model accuracy for various developed ANN models as  $R^2$  0.83–0.88 and *MAE* 0.66–0.85, while Afolabi et al. (2022) [5] reported 1.21 for the ANN model and 1.01 for the RF model.

Based on the statistical indicators used, the ANN model showed the highest efficiency, which indicates high precision and reliability in prediction. The ANN model had the longest execution time (222.33 s), which is a consequence of the complexity of the model, the number of neurons, and the learning process, which involves a multi-layered structure and optimization. The shortest execution time is seen for third-degree polynomials due to the relative simplicity and non-complex regression tasks. The SVM and RFR models also have a shorter execution time due to the lower complexity of the model and the complexity of the optimization process. The experiments conducted by Guimarães et al. (2023) [45] demonstrate that predicting the training time of models like decision trees and neural networks is achievable with reasonable accuracy. The average prediction error is 0.103 s for decision trees and 21.263 s for ANNs, which is acceptable considering their training times span up to 14 s and over 1400 s, respectively. The application of the ML model in the estimation of HHV biomass shows clear advantages in terms of accuracy and reliability of predictions. Models such as ANNs often achieve high  $R^2$  values ( $>0.90$ ) and lower *RMSE* and *MAE* values, demonstrating their ability to estimate by managing the data [46,47]. On the other hand, it should be noted that although high accuracy and lower modeling errors are achieved, longer runtimes are often required due to higher learning complexity [48], which can be a limiting factor when computational resources are limited. On the other hand, other ML models that have simpler structures provide faster results due to their (relative) simplicity, although they are not as accurate in estimation [48].



## 5. Conclusions

The artificial neural network (ANN) demonstrated superior performance across all evaluated metrics in this scientific study. The model achieved an  $R^2$  value of 0.92, an RMSE (root mean square error) of 1.33, and an MAE (mean absolute error) of 0.77, indicating its high accuracy in predicting the HHV variable. This proficiency is attributed to the ANN's complex structure, which involves a multi-layered configuration and an optimization process, encompassing numerous neurons and a comprehensive learning procedure. However, it is noteworthy that the ANN model had the longest execution time, clocking in at 222.33 s. This extended duration is primarily due to the model's complexity, the extensive number of neurons, and the intricate learning process that includes a multi-layer structure and optimization. In contrast, the polynomial regression model of the third degree demonstrated the shortest execution time at 0.03 s. This efficiency stems from its relative simplicity and the less complex nature of the regression tasks it handles. Despite this, the polynomial model exhibited the lowest accuracy and the highest error in modeling, with an  $R^2$  value of 0.72, an RMSE of 2.14, and an MAE of 1.54. The most effective ANN model was achieved with a configuration of 4000 epochs, with 3 artificial neurons in the input layer and 10 in the hidden layer. Additionally, the ReLU (rectified linear unit) activation function was employed in this model. Furthermore, a diagram and correlation analysis concerning the fixed carbon (FC) content of the HHV suggested a potential correlation. This indicates that the content of fixed carbon might be a significant factor in predicting the higher heating value, highlighting an area for further exploration in the field.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en17092137/s1>, Table S1. Collected data for modeling.

**Author Contributions:** Conceptualization, I.B. and L.P.; methodology, A.M.; software, I.B.; validation, N.V.; data curation, L.P.; writing—original draft preparation, I.B.; writing—review and editing, N.V.; supervision, N.V.; project administration, N.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data used to create the model can be found in Table S1 along with the Python code used to create the machine learning model.

**Acknowledgments:** When writing the manuscript, certain AI-based tools were used to improve parts of the text that are exclusively related to correcting grammar and sentence structure. The following AI-based tools were used: *Quillbot* when searching for synonyms to avoid redundancy and repetition in writing. *Grammarly* and *Instatext* were used to correct the grammatical part of the text, i.e., to change the structure of certain sentences and to remove/add grammatical and corresponding characters. The *Adobe Illustrator* program was used to improve the image resolution (1000 dpi).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Helal, M.A.; Anderson, N.; Wei, Y.; Thompson, M. A Review of Biomass-to-Bioenergy Supply Chain Research Using Bibliometric Analysis and Visualization. *Energies* **2023**, *16*, 1187. [CrossRef]
2. Tshikovhi, A.; Motaung, T.E. Technologies and Innovations for Biomass Energy Production. *Sustainability* **2023**, *15*, 12121. [CrossRef]
3. Jha, S.; Okolie, J.A.; Nanda, S.; Dalai, A.K. A Review of Biomass Resources and Thermochemical Conversion Technologies. *Chem. Eng. Technol.* **2022**, *45*, 791–799. [CrossRef]
4. Basu, P. *Biomass Characteristics*, 1st ed.; Elsevier Inc.: Amsterdam, The Netherlands, 2010; ISBN 9780123749888.
5. Afolabi, I.C.; Epelle, E.I.; Gunes, B.; Güleç, F.; Okolie, J.A. Data-Driven Machine Learning Approach for Predicting the Higher Heating Value of Different Biomass Classes. *Clean Technol.* **2022**, *4*, 1227–1241. [CrossRef]
6. Aghel, B.; Yahya, S.I.; Rezaei, A.; Alobaid, F. A Dynamic Recurrent Neural Network for Predicting Higher Heating Value of Biomass. *Int. J. Mol. Sci.* **2023**, *24*, 5780. [CrossRef] [PubMed]
7. Velázquez Martí, B.; Gaibor-Chávez, J.; Franco Rodríguez, J.E.; López Cortés, I. Biomass Identification from Proximate Analysis: Characterization of Residual Vegetable Materials in Andean Areas. *Agronomy* **2023**, *13*, 2347. [CrossRef]

8. Taye, M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* **2023**, *12*, 91. [[CrossRef](#)]
9. Tufail, S.; Riggs, H.; Tariq, M.; Sarwat, A.I. Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms. *Electronics* **2023**, *12*, 1789. [[CrossRef](#)]
10. Calix, R.A.; Ugarte, O.; Okosun, T.; Wang, H. Machine Learning-Based Regression Models for Ironmaking Blast Furnace Automation. *Dynamics* **2023**, *3*, 636–655. [[CrossRef](#)]
11. Ma, H.; Ding, F.; Wang, Y. A Novel Multi-Innovation Gradient Support Vector Machine Regression Method. *ISA Trans.* **2022**, *130*, 343–359. [[CrossRef](#)]
12. Wang, X.; Zhang, C.; Li, L.; Fritsche, S.; Endrigkeit, J.; Zhang, W.; Long, Y.; Jung, C.; Meng, J. Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (*Brassica napus* L.). *PLoS ONE* **2012**, *7*, e50038. [[CrossRef](#)]
13. Su, J.; Wang, Y.; Niu, X.; Sha, S.; Yu, J. Prediction of Ground Surface Settlement by Shield Tunneling Using XGBoost and Bayesian Optimization. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105020. [[CrossRef](#)]
14. Mahmood, J.; Mustafa, G.-E.; Ali, M.A. Accurate Estimation of Tool Wear Levels during Milling, Drilling and Turning Operations by Designing Novel Hyperparameter Tuned Models Based on LightGBM and Stacking. *Measurement* **2022**, *190*, 110722. [[CrossRef](#)]
15. Dutta, J.; Roy, S. OccupancySense: Context-Based Indoor Occupancy Detection & Prediction Using CatBoost Model. *Appl. Soft Comput.* **2022**, *119*, 108536. [[CrossRef](#)]
16. Babatunde, D.E.; Anozie, A.; Omoleye, J. Artificial Neural Network and Its Applications in the Energy Sector—An Overview. *Int. J. Energy Econ. Policy* **2020**, *10*, 250–264. [[CrossRef](#)]
17. Kufel, J.; Bargieł-Lączek, K.; Kocot, S.; Koźlik, M.; Bartnikowska, W.; Janik, M.; Czogalik, Ł.; Dudek, P.; Magiera, M.; Lis, A.; et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. *Diagnostics* **2023**, *13*, 2582. [[CrossRef](#)]
18. Yang, X.; Li, H.; Wang, Y.; Qu, L. Predicting Higher Heating Value of Sewage Sludges via Artificial Neural Network Based on Proximate and Ultimate Analyses. *Water* **2023**, *15*, 674. [[CrossRef](#)]
19. Jwo, D.J.; Biswal, A.; Mir, I.A. Artificial Neural Networks for Navigation Systems: A Review of Recent Research. *Appl. Sci.* **2023**, *13*, 4475. [[CrossRef](#)]
20. Dudek, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies* **2022**, *15*, 7547. [[CrossRef](#)]
21. Tyrallis, H.; Papacharalampous, G.; Langousis, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water* **2019**, *11*, 910. [[CrossRef](#)]
22. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Tree-Based Ensemble Methods for Predicting PV Power Generation and Their Comparison with Support Vector Regression. *Energy* **2018**, *164*, 465–474. [[CrossRef](#)]
23. Orrù, P.F.; Zoccheddu, A.; Sassu, L.; Mattia, C.; Cozza, R.; Arena, S. Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry. *Sustainability* **2020**, *12*, 4776. [[CrossRef](#)]
24. Cemiloglu, A.; Zhu, L.; Arslan, S.; Xu, J.; Yuan, X.; Azarafza, M.; Derakhshani, R. Support Vector Machine (SVM) Application for Uniaxial Compression Strength (UCS) Prediction: A Case Study for Maragheh Limestone. *Appl. Sci.* **2023**, *13*, 2217. [[CrossRef](#)]
25. Abd-Elhameed, W.M.; Amin, A.K. New Formulas and Connections Involving Euler Polynomials. *Axioms* **2022**, *11*, 743. [[CrossRef](#)]
26. Khan, W.A.; Acikgoz, M.; Duran, U. Note on the Type 2 Degenerate Multi-Poly-Euler Polynomials. *Symmetry* **2020**, *12*, 1691. [[CrossRef](#)]
27. Abd-Elhameed, W.M.; Ahmed, H.M.; Napoli, A.; Kowalenko, V. New Formulas Involving Fibonacci and Certain Orthogonal Polynomials. *Symmetry* **2023**, *15*, 736. [[CrossRef](#)]
28. Qian, X.; Lee, S.; Soto, A.M.; Chen, G. Regression Model to Predict the Higher Heating Value of Poultry Waste from Proximate Analysis. *Resources* **2018**, *7*, 39. [[CrossRef](#)]
29. Matveeva, A.; Bychkov, A. How to Train an Artificial Neural Network to Predict Higher Heating Values of Biofuel. *Energies* **2022**, *15*, 7083. [[CrossRef](#)]
30. ÖzyüçÜran, A.; Yaman, S. Prediction of Calorific Value of Biomass from Proximate Analysis. *Energy Procedia* **2017**, *107*, 130–136. [[CrossRef](#)]
31. Espina, R.U.; Barroca, R.B.; Abundo, M.L.S. The Energy Yield of the Torrefied Coconut Shells. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, 1187. [[CrossRef](#)]
32. Krishnan, R.; Hauchhum, L.; Gupta, R.; Pattanayak, S. Prediction of Equations for Higher Heating Values of Biomass Using Proximate and Ultimate Analysis. In Proceedings of the 2018 2nd International Conference on Power, Energy and Environment: Towards Smart Technology (ICEPE), Shillong, India, 1–2 June 2018; IEEE: Piscataway Township, NJ, USA, 2019. [[CrossRef](#)]
33. Pattanayak, S.; Hauchhum, L.; Loha, C.; Sailo, L. Selection Criteria of Appropriate Bamboo Based Biomass for Thermochemical Conversion Process. *Biomass Convers. Biorefinery* **2020**, *10*, 401–407. [[CrossRef](#)]
34. Noushabadi, A.S.; Dashti, A.; Ahmadijokani, F.; Hu, J.; Mohammadi, A.H. Estimation of Higher Heating Values (HHVs) of Biomass Fuels Based on Ultimate Analysis Using Machine Learning Techniques and Improved Equation. *Renew. Energy* **2021**, *179*, 550–562. [[CrossRef](#)]
35. Abdolrasol, M.G.M.; Suhail Hussain, S.M.; Ustun, T.S.; Sarker, M.R.; Hannan, M.A.; Mohamed, R.; Ali, J.A.; Mekhilef, S.; Milad, A. Artificial Neural Networks Based Optimization Techniques: A Review. *Electronics* **2021**, *10*, 2689. [[CrossRef](#)]
36. Han, S.; Kim, H. Optimal Feature Set Size in Random Forest Regression. *Appl. Sci.* **2021**, *11*, 3428. [[CrossRef](#)]

37. Hodson, T.O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [[CrossRef](#)]
38. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M.; Doukas, I.J.D. Comparing Machine Learning Models and Hybrid Geostatistical Methods Using Environmental and Soil Covariates for Soil PH Prediction. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 276. [[CrossRef](#)]
39. Park, S.Y.; Oh, K.C.; Kim, S.J.; Cho, L.H.; Jeon, Y.K.; Kim, D.H. Development of a Biomass Component Prediction Model Based on Elemental and Proximate Analyses. *Energies* **2023**, *16*, 5341. [[CrossRef](#)]
40. Harun, N.Y.; Parvez, A.M.; Afzal, M.T. Process and Energy Analysis of Pelletizing Agricultural and Woody Biomass Blends. *Sustainability* **2018**, *10*, 1770. [[CrossRef](#)]
41. Villegas, J.M.; Avila, H. Quick-Scan Estimating Model of Higher Heating Value of Oil Palm Empty Fruit Bunches Based on Ash from Proximate Analysis Data. *Ing. Investig.* **2014**, *34*, 33–38. [[CrossRef](#)]
42. Wang, Y.; Li, Y.; Song, Y.; Rong, X. The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition. *Appl. Sci.* **2020**, *10*, 1897. [[CrossRef](#)]
43. Razaque, A.; Ben Haj Frej, M.; Almi'ani, M.; Alotaibi, M.; Alotaibi, B. Improved Support Vector Machine Enabled Radial Basis Function and Linear Variants for Remote Sensing Image Classification. *Sensors* **2021**, *21*, 4431. [[CrossRef](#)]
44. Shanmugavel, A.B.; Ellappan, V.; Mahendran, A.; Subramanian, M.; Lakshmanan, R.; Mazzara, M. A Novel Ensemble Based Reduced Overfitting Model with Convolutional Neural Network for Traffic Sign Recognition System. *Electronics* **2023**, *12*, 926. [[CrossRef](#)]
45. Guimarães, M.; Carneiro, D.; Palumbo, G.; Oliveira, F.; Oliveira, Ó.; Alves, V.; Novais, P. Predicting Model Training Time to Optimize Distributed Machine Learning Applications. *Electronics* **2023**, *12*, 871. [[CrossRef](#)]
46. Brandić, I.; Pezo, L.; Bilandžija, N.; Peter, A.; Šurić, J.; Voća, N. Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass. *Mathematics* **2023**, *11*, 2098. [[CrossRef](#)]
47. Šurić, J.; Voća, N.; Peter, A.; Bilandžija, N.; Brandić, I.; Pezo, L.; Leto, J. Use of Artificial Neural Networks to Model Biomass Properties of *Miscanthus × giganteus* and Virginia Mallow (*Sida hermaphrodita* L.) in View of Harvest Season. *Energies* **2023**, *16*, 4312. [[CrossRef](#)]
48. Abdollahi, S.A.; Ranjbar, S.F.; Razeghi Jahromi, D. Applying Feature Selection and Machine Learning Techniques to Estimate the Biomass Higher Heating Value. *Sci. Rep.* **2023**, *13*, 1–13. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.