University of Zagreb

FACULTY OF AGRICULTURE

Ivan Brandić

DEVELOPMENT OF NEW NONLINEAR MATHEMATICAL MODELS FOR MODELLING THE HIGHER HEATING VALUE OF BIOMASS

DOCTORAL THESIS

Zagreb, 2024.

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Ivan Brandić

RAZVOJ NOVIH NELINEARNIH MATEMATIČKIH MODELA U MODELIRANJU GORNJE OGRJEVNE VRIJEDNOSTI BIOMASE

DOKTORSKI RAD

Zagreb, 2024.

University of Zagreb

FACULTY OF AGRICULTURE

Ivan Brandić

DEVELOPMENT OF NEW NONLINEAR MATHEMATICAL MODELS FOR MODELLING THE HIGHER HEATING VALUE OF BIOMASS

DOCTORAL THESIS

Supervisors:
Full prof. Neven Voća, PhD
Lato Pezo, PhD

Zagreb, 2024.

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Ivan Brandić

RAZVOJ NOVIH NELINEARNIH MATEMATIČKIH MODELA U MODELIRANJU GORNJE OGRJEVNE VRIJEDNOSTI BIOMASE

DOKTORSKI RAD

Mentori:
prof. dr. sc. Neven Voća
dr. sc. Lato Pezo

Zagreb, 2024.

Bibliography data

Scientific area: Biotechnical sciences
Scientific field: Agriculture
Branch of science: /
Institution: University of Zagreb Faculty of Agriculture
Supervisors of the PhD thesis:

1. Full. Prof. Neven Voća, PhD, University of Zagreb Faculty of Agriculture
2. Lato Pezo, PhD, University of Belgrade, Institute of General and Physical Chemistry

Number of pages: /
Number of figures: /
Number of tables: /
Number of appendices: /
Number of references: /
Date of oral examination:
The members of the PhD defence committee:

1.
2.
3.

Rad je pohranjen u:
Nacionalnoj i sveučilišnoj knjižnici u Zagrebu, Ulica Hrvatske bratske zajednice 4 p.p. 550, 10 000 Zagreb, Knjižnici Sveučilišta u Zagrebu Agronomskog Fakulteta, Svetošimunska cesta 25, 10 000 Zagreb.

The doctoral thesis is stored in:
National and University Library in Zagreb, Ulica Hrvatske bratske zajednice 4 p.p. 550, 10 000 Zagreb, Central Agricultural Library, University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, 10 000 Zagreb. The subject of the thesis was accepted at the session of the Faculty Council of the University of Zagreb Faculty of Agriculture, held on March 3, 2020, and approved at the session of the Senate of the University of Zagreb, held on June 02, 2020.

UNIVERSITY OF ZAGREB

FACULTY OF AGRICULTURE

DECLARATION OF ORIGINALITY

I, **Ivan Brandić**, declare that I have composed solely by myself the thesis titled:

DEVELOPMENT OF NEW NONLINEAR MATHEMATICAL MODELS FOR MODELLING THE HIGHER HEATING VALUE OF BIOMASS

I hereby declare that:

- I am the sole author of this doctoral thesis;
- Doctoral thesis is the original report of my research and that I have used no other sources in its preparation than those indicated therein;
- I am familiar with the provisions of the University of Zagreb Code of Ethics (Art. 19).

Zagreb,

_____

PhD Candidate signature

Doctoral thesis grade

Doctoral thesis was defended at the University of Zagreb, Faculty of Agriculture on _____, 2024 in front of the PhD defence committee comprised of:

Grade: _____

Members of the committee:

1.Assoc. prof. Nikola Bilandžija, PhD _____
2. Prof. Jerko Gunjača, PhD _____
3. Assoc. Prof. Biljana Lončar, PhD_____

Mentor's information

Prof. Neven Voća

Prof. Neven Voća, Ph.D., is a full-time professor in the Department of Sustainable Technologies and Renewable Energy Sources at the University of Zagreb, Faculty of Agriculture. Born and raised in Croatia, he completed his education at the University of Zagreb, Faculty of Agriculture. From 1992 to 1997, he graduated as an engineer in the field of agricultural engineering. He then continued his studies at the same university, obtaining a Master of Science degree in 2003 and a doctorate in 2007. He has been working at the University of Zagreb, Faculty of Agriculture since 1999. His scientific interests are mainly focused on biomass and biofuels as well as waste management. Throughout his career, Prof. Voća has also completed numerous professional trainings, such as a scientific visit to Kansas State University in 2003, where he learned about the production of animal feed and biofuels, and a professional training on management issues at Scuola di Management BIC in 1999. In addition to his academic and professional duties, Prof. Voća is also involved in various scientific and professional activities and has published more than 250 scientific and professional papers to date.
From 2011 to 2015, he was the Scientific Secretary of the Scientific Council for Nature Conservation at the Croatian Academy of Sciences and Arts. From 2012 to 2015, Prof. Voća was the Director of the Environmental Protection Agency and Croatia's representative on the Management Board of the European Environment Agency. He was also an external member of the Parliamentary Committee for Environmental Protection from 2008 to 2011.

Lato Pezo, PhD

Dr Lato Pezo, born 3 August 1971, is a distinguished full research professor specialising in the field of biotechnology. His academic path began at the Faculty of Mechanical Engineering, University of Belgrade, Serbia, where he obtained his B.Sc. in 1994, his M.Sc. in 1997 and finally his Ph.D. in 1999. His professional career began as a scientist at the Institute of General and Physical Chemistry, University of Belgrade, where he later worked as a Chemical Engineering Designer, Assistant Research Professor and Associate Research Professor before obtaining his current position as Full Research Professor. Throughout his career, Dr Pezo has been instrumental in numerous national projects ranging from the development of membrane, sorption and membrane transportation devices for the separation and purification of liquids and gases to the development of innovative chemical engineering processes for the reduction of environmental pollution. In particular, he has led several projects as manager, including the reactivation of idle capacity in the process industry and the development of an aquatic plant biomass filter to remove pollutants from wastewater. In addition, Dr Pezo has worked on European Union projects such as the Zero Food Waste education programme for European citizens and Erasmus+ programme, and COST: “Open Network on DEM Simulations” (ON-DEM) CA22132. His expertise covers a wide range, including mathematical modelling, numerical modelling, discrete element modelling, computational fluid dynamics and statistics in food science and biotechnology. He is fluent in English and has a basic knowledge of German.

The research was supported by the Croatian Science Foundation, under project No. IP-2018-01-7472 "Sludge management via energy crops production" within the project "Young Researchers' Career Development Project—Training of Doctoral Students", and co-financed by the European Union, under the OP "Efficient Human Resources 2014–2020" from the ESF funds.

ACKNOWLEDGMENT

XYZ

ABSTRACT

Biomass as a renewable energy source is becoming increasingly important due to rising prices and the environmental incompatibility of conventional energy sources. The most important parameter in assessing fuel quality is the higher heating value (HHV), which represents the amount of energy released per unit area. Existing linear mathematical models have a higher error in HHV prediction than nonlinear models, indicating the need to develop new nonlinear mathematical models for HHV biomass prediction. In this study, new nonlinear mathematical models in the form of higher-order polynomials (HOP), artificial neural networks (ANN), random regression forests (RFR) and support vector models (SVM) will be developed for HHV biomass modelling in agriculture and wood. In addition, the research will compare the developed nonlinear models to determine the smallest HHV modelling error for different sets of input variables. The data used to develop the model will be taken from the available literature and the non-linear models will be based on characteristic sets of variables from ultimate, proximate, and structural analysis. For all models developed, a data split of 70% for training and 30% for model testing is used. A statistical analysis of “goodness of fit" is performed to identify the model with the lowest error and to select the most appropriate model for modelling HHV biomass. The results obtained from the nonlinear regression models are compared with the experimentally obtained data from the literature to investigate the performance and effectiveness of the developed models for HHV modelling. After testing the reliability and performance of the model, a sensitivity analysis is performed to optimise the models and find the optimal sample and the relationship between the input variables and the output variable (HHV). For this purpose, Yoon's method is used, which is based on the sensitivity analysis of the neurons in the hidden layer of the ANN. Yoon's method makes it possible to evaluate the importance of individual input variables about the HHV output values and to determine the most important variables for accurate and reliable modelling of the HHV biomass. The sensitivity analysis will help to identify the most important parameters affecting HHV, optimise the model and reduce the prediction error. The results of this research will provide insight into different nonlinear mathematical models for modelling HHV biomass, such as HOP, ANN, RFR and SVM, and their success in predicting HHV based on different sets of input variables. The comparison of the developed models will allow the selection of the most appropriate model for reliable prediction of HHV biomass, which would reduce the time required for HHV determination and improve the use of biomass as a renewable energy source.

Keywords: biomass, higher heating value, mathematical modelling, artificial neural networks, higher degree polynomials, random forest regression, support vector machine.

EXTENDED ABSTRACT IN CROATIAN

Razvoj novih nelinearnih matematičkih modela u modeliranju gornje ogrjevne vrijednosti biomase

Biomasa kao obnovljivi izvor energije sve više dobiva na važnosti zbog porasta cijene i ekološke neprihvatljivosti konvencionalnih izvora energije. Glavni parametar u procjeni kvalitete goriva je ogrjevna vrijednost (HHV), koja predstavlja količinu energije koja se oslobađa po jedinici površine. Postojeći linearni matematički modeli (sa jednim ili dva parametra) imaju veću grešku u predviđanju HHV od nelinearnih modela, što ukazuje na potrebu za razvojem novih nelinearnih matematičkih modela koji nude veći stupanj nelinearnosti za predviđanje HHV biomase.

U ovom istraživanju, cilj je razviti nove nelinearne matematičke modele u obliku polinoma višeg stupnja (HOP), umjetnih neuronskih mreža (ANN), slučajnih šuma za regresiju (RFR) i modela potpornih vektora (SVM) za modeliranje HHV poljoprivredne i šumske biomase. Također, istraživanje će usporediti razvijene nelinearne modele kako bi se utvrdila najmanja pogreška modeliranja HHV s obzirom na različite setove ulaznih varijabli. Javno dostupni podaci korišteni za razvoj modela bit će prikupljeni iz raspoložive literature, a nelinearni modeli temeljit će se na karakterističnim setovima varijabli elementarne, fizikalno-kemijske i strukturalne analize. Za sve razvijene modele koristit će se podjela podataka od 70% za učenje i 30% za testiranje modela. Provest će se statistička analiza prikladnosti (eng. "Goodness of fit") kako bi se odredio model s najmanjom pogreškom za modeliranje HHV biomase. Rezultati dobiveni nelinearnim modelima usporedit će se s literaturno prikupljenim podacima kako bi se ispitale performanse i učinkovitost razvijenih modela za modeliranje HHV. Nakon ispitivanja pouzdanosti i performansi modela, bit će provedena analiza osjetljivosti kako bi se ispitao utjecaj ulaznih podataka na izlazne vrijednosti modela. U tu svrhu koristit će se Yoon-ova metoda, koja se temelji na analizi osjetljivosti neurona u skrivenom sloju ANN-a. Yoon-ova metoda omogućuje procjenu važnosti pojedinih ulaznih varijabli u odnosu na izlazne vrijednosti HHV, te utvrđivanje najvažnijih varijabli za točno i pouzdano modeliranje HHV biomase. Analiza osjetljivosti pomoći će u identificiranju najvažnijih parametara koji utječu na HHV te u optimizaciji modela i smanjenju pogreške u predviđanju. Rezultati ovog istraživanja pružit će uvid u različite nelinearne matematičke modele za modeliranje HHV biomase, kao što su HOP, ANN, RFR i SVM, te njihovu uspješnost u predviđanju HHV na temelju različitih setova ulaznih varijabli. Usporedba razvijenih modela omogućit će odabir najprikladnijeg modela za pouzdano predviđanje HHV biomase, čime bi se smanjilo vrijeme potrebno za utvrđivanje HHV i poboljšala upotreba biomase kao obnovljivog izvora energije.

Fokus provedenog istraživanja bio je na razvoju novih nelinearnih modela za modeliranje HHV biomase. Ispitani su novorazvijeni nelinearni modeli, a nakon međusobne usporedbe, odabran je model s najmanjom pogreškom u odnosu na setove ulaznih varijabli elementarne, fizikalno-kemijske i strukturalne analize. Podaci koji su korišteni za razvijanje modela prikupljeni su iz raspoložive literature prethodnih znanstvenih istraživanja, konkretno iz znanstvenih publikacija koji su citirani u bazama "*Web of Science*" i "*ScienceDirect*". Prikupljeni podaci su zatim podijeljeni na poljoprivrednu i šumsku biomasu. Nelinearni modeli temeljili su se na setovima podataka elementarne (eng. ultimate), fizikalno-kemijske (eng. proximate), strukturalne (eng. structural) i kalorimetrijske (eng. calorimetric) analize. Nakon što je prikupljen dovoljan broj podataka, proveden je postupak "čišćenja" podataka s obzirom na kompletnost prikupljene baze podataka. Završetkom postupka čišćenja, podaci su podijeljeni na dio za učenje i testiranje modela u omjeru 70% i 30%. Za kreiranje nelinearnih modela koristili su se softverski alati za matematičko modeliranje i statističku analizu. Razvijeni su različiti modeli, uključujući modele HOP, ANN, RFR i SVM. Svaki model testiran je na pouzdanost i performanse. Analiza osjetljivosti provedena je Yoon-ovom metodom. Po završetku analiza, rezultati su interpretirani kako bi se mogli usporediti svi novorazvijeni modeli u smislu modeliranju HHV biomase.

Statističkom analizom strukturalnog sastava biomase utvrđeno je da razlika u udjelu celuloze nije statistički značajna dok udio lignina, hemiceluloze i HHV je bio statistički značajno viši kod šumske biomase. Analizirajući razlike u fizikalno-kemijskom sastavu poljoprivredne i šumske biomase nisu utvrđene statistički značajne razlike u udjelu fiksiranog ugljika, hlapivih tvari i pepela, dok je kod varijabli elementarne analize utvrđen statistički značajno manji udio dušika (N) kod šumske biomase. Nakon kreiranja nelinearnih modela u svrhu modeliranja HHV biomase utvrđeno je da je ANN model imao manju razinu greške kod svih setova podataka (karakterističnih analiza) koji su korišteni kao setovi ulaznih podataka. Navedeno je dokazano pomoću statističke analize prikladnosti ("Goodness of fit") modela gdje je ANN model za set ulaznih podataka elementarne analize (sa svim varijablama seta podataka) ima iduće rezultate: Hi-kvadrat test (χ^2) = 1,02, korijen iz prosječne kvadratne pogreške (RMSE)=1,01, prosječne pogreške pristranosti (MBE) = 0,06, prosječne postotne pogreške (MPE) = 4,21, sumu kvadrata pogreške (SSE) = 251,19), prosječno apsolutno odstupanje (AARD) = 196,44, koeficijent determinacije (R^2) = 0,90, mjere asimetrije podataka (Skew) = -0,46, mjere spljoštenosti podataka (Kurt) = 1,23, standardna devijacija (SD) = 1,01 i varijanca = 1.01. Kod ulaznog seta fizikalno-kemijske analize (fiksirani ugljik i hlapive tvari) za ANN model izračunati su idući pokazatelji prikladnosti modela: χ^2 (0,41), RMSE (0,64), MBE (0,03), MPE (2,65), SSE (118,33), AARD (240,27), R^2 (0,96), Skew (-0,48), Kurt (1,53), standardna devijacija (0,64), varijanca (0,41). Kod ulaznog seta podataka sturkturalne analize statistički paramteri za model ANN pokazali su iduće vrijednosti: χ^2 (0,26), RMSE (0,51), MBE (0,00), MPE (2,30), SSE (74,81), AARD (193,72), R^2 (0,91), Skew (-1,09), Kurt (3,34), standardna devijacija (0,51), varijanca (0,26). Također za sve razvijene modele najmanju pogrešku u modeliranju su dobiveni kod ulaznog seta podataka fizikalno kemijske analize, gdje je kao glavni pokazatelj reprezentativnosti regresije uzet koeficijent determinacije (R^2). Yoonova metoda globalne osjetljivosti koristila se za optimizaciju ANN modela (najpogodnijeg modela) kako bi se pronašla relativna važnost ulaznih varijabli (%) na izlaznu vrijednost HHV. Kod ulaznog seta podataka elementarne analize na HHV najviše utječe povećanje C, N i S kao i smanjenje H i O. Model sa najmanje pogreške u modeliranju uključivao je kao prediktorske varijable fiksirani ugljik i hlapive tvari, Yoonovom metodom je utvrđeno da smanjenjem navedenih varijabli dolazi do povećanja HHV biomase, dok kod varijabli stukturalne analize kao ulaznih podataka povećanje udjela celuloze, lignina kao I smanjenje hemiceluloze utječe na ukupno veći HHV.

Table of Contents

List of tables:

List of abbervations:

1. Higher heating value (HHV)
2. High order polynomial (HOP)
3. Artificial neural networks (ANN)
4. Random forest regression (RFR)
5. Support vector machine (SVM)
6. Chi-squared test ($\chi 2$)
7. Root mean squared error (RMSE)
8. Mean bias error (MBE)
9. Mean percentage error (MPE)
10. Sum of squared error (SSE)
11. Coefficient of determination (R^2)
12. Skewness (Skew)
13. Kurtosis (Kurt)
14. Standard deviation (SD)
15. Renewable energy sources (RES)
16. European union (EU)
17. Deep learning (DL)
18. Machine learning (ML)
19. Artificial intelligence (AI)
20. Multilayer perceptron (MLP)
21. Carbon dioxide (CO_2)
22. Carbon (C)
23. Hydrogen (H)

24. Oxygen (O)
25. Nitrogen (N)
26. Sulphur (S)
27. Fixed carbon (FC)
28. Moisture content (MC)
29. Volatile matter (VM)
30. Lower heating value (LHV)
31. Step function (ReLu)
32. Support vector (SV)
33. Error insensitivity (ε)
34. Global sensitivity analysis (GSA)
35. Analysis of variance (ANOVA)
36. Honestly significant difference (HSD)
37. Levenberg–Marquardt model (LM)

Scientific papers	Base	Category	Quartile	Impact factor
1. **Brandić I**., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2022). Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis. Mathematics.	WoS	A1	Q1	2.4
2. **Brandić I**., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2023). Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass. Mathematics.	WoS	A1	Q1	2.4
3. **Brandić I**., Antonović A., Pezo L., Matin B., Krička T., Jurišić V., Špelić K., Kontek M., Kukuruzović J., Grubor M., Matin A. (2023). Energy Potentials of Agricultural Biomass and the Possibility of Modelling Using RFR and SVM Models. Energies.	WoS	A1	Q3	3.2
4. **Brandić I**., Pezo L., Voća N., Leto J., Šurić J., Peter A., Bilandžija N. (2024). Assessing the properties of Miscanthus x giganteus under varying levels of ash fertilization treatment and regression neural network insight into calorific value. Thermal Science.	WoS	A1	Q4	1.7
5. **Brandić I**., Voća N., Gunjača J., Lončar B., Bilandžija N., Peter A., Šurić J., Pezo L. (2024). Biomass higher heating value prediction: Machine Learning Insight into Ultimate, Proximate and Structural analysis datasets. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects.	WoS	A1	Q3	2.9

Explanation of the connection between research hypotheses and published research papers

Research hypotheses	Explanation of research hypotheses
H1. The models based on the set of input variables of proximate analysis have the lowest modelling	In Scientific Paper number 1, an artificial neural network (ANN) model for the estimation of the higher heating value (HHV) of biomass was created based on input data from ultimate analysis. This model was compared with existing prediction models in

error in all non-linear models examined compared to the sets of input variables of ultimate and structural analysis.	terms of modelling error. In scientific paper number 2, the artificial neural network (ANN), random forest regression (RFR), higher order polynomials (HOP) and support vector machine (SVM) models were developed for higher heating value (HHV) biomass modelling. Input data from the structural analysis of agricultural and wood biomass were used and the models were compared in terms of modelling error. The scientific paper number 3 deals with the development of random forest regression (RFR) and support vector machines (SVM) models based on the input data from the ultimate analysis of biomass for higher heating value (HHV) estimation. The performance of these models was compared in terms of error rate. In scientific paper number 4, the artificial neural network (ANN) model for higher heating value (HHV) modelling was developed based on the input variables of the ultimate analysis. The performance of the developed model was compared with models previously published in the scientific literature. In scientific paper number 5, artificial neural network (ANN), higher order polynomial (HOP), random forest regression (RFR) and support vector machines (SVM) models were developed for higher heating value (HHV) biomass modelling using a range of input data from ultimate, proximate and structural analysis. The models were compared in terms of modelling error, taking into account the model developed and the dataset used.
H2. The ANN models have a lower error in modelling HHV of biomass than HOP, RFR and SVM, regardless of the set of input variables.	In scientific paper number 5, models were created in the form of the artificial neural network (ANN), random forest regression (RFR), higher order polynomials (HOP) and support vector machine (SVM). These models are based on data sets from ultimate, structural and proximate analyses datasets of the biomass. After creation, all models were compared in terms of modelling error, taking into account the model type and the input dataset used.

1. INTRODUCTION

In the context of the growing energy crisis and the emphasised need to use renewable energy sources (RES), the role of biomass as a raw material for energy production is becoming increasingly important. This is particularly important given the rising prices of conventional fuels and their harmful impact on the environment. Biomass is one of the key elements for achievement of the energy independence and energy sustainability (Callejón-Ferre et al., 2014). Long-term energy strategies, especially those of the European Union (EU), emphasise the importance of the transition to renewable energy sources. The production of energy from renewable sources is one of the fundamental components for achieving the EU energy policy targets by 2030 (Scarlat Nicolae et al., 2019), while Mandley et al., (2022) states that energy produced from biomass could contribute to the necessary reduction in the use of conventional energy sources by up to 40%, as well as to an increase in the production of the biomass material itself in order to reduce negative climate changes. Biomass, as one of the most widely represented RES, offers a wide range of possibilities to produce various forms of useful energy. In addition to being used to produce biofuel, biomass is also used to generate electricity and thermal energy, which allows it to be integrated into various energy systems. Utilising biomass as an energy source can help to reduce the overall cost of energy production. This means that it is not only an environmentally friendly alternative to conventional fuels, but also an economically efficient one (Demirbas, 2017).

The term biomass usually refers to biodegradable residues resulting from various production processes, e.g. in agriculture and forestry, as well as organic waste from plant and animal production. Lignocellulosic biomass, which is rich in fibres and cellulose, stands out as one of the most efficient raw materials for energy production (Olatunji et al., 2018). However, to optimise its use, it is crucial to determine its physical and chemical properties. This includes ultimate, proximate, and structural analyses of the biomass. This information is necessary to accurately determine the potential of biomass for energy production and is usually determined through laboratory analyses (Callejón-Ferre et al., 2014; Demirbas, 2017; Baxter et al., 2014; Yi et al., 2017). Another important aspect that should be considered when evaluating the quality of biomass as a fuel is its higher heating value (HHV). This parameter indicates the amount of energy released per unit mass when biomass is completely combusted and is expressed in MJ kg^{-1}. In terms of modelling, the HHV is a key indicator for determining the energy efficiency of biomass and comparing it with other fuel types (Sheng and Azevedo, 2005).

The determination of key parameters of biomass fuel properties, such as the HHV, is often a time-consuming and cost-intensive process. This process traditionally requires the use of advanced laboratory equipment and empirical methods (Boumanchar et al., 2019). Therefore, there is a growing interest in developing mathematical models that would enable faster and more efficient prediction of these properties without the need for expensive and time-consuming laboratory analyses. Recently, the use of deep learning (DL) and machine learning (ML) principles has gained popularity as a means of developing models to predict HHV. These techniques, which use advanced ML algorithms, enable faster and more accurate modelling of biomass properties (Zhang et al., 2020). Various non-linear mathematical models are used for this purpose, including higher-order polynomial models (HOP), artificial neural networks (ANN), random forest regression (RFR) and support vector machine (SVM) models. Among the models mentioned above, ANN models are of particular interest. These models, which are part of the broader field of artificial intelligence (AI), can analyse large amounts of data and identify non-linear relationships between different variables, enabling a better understanding and more accurate prediction of biomass fuel properties (Giwa et al., 2015; Dashti et al., 2019).

ANN differ from traditional regression models in their ability to link dependent and independent variables in a non-linear way, facilitating the detection of complex interactions between them (Pattanayak et al., 2021). Although the application of ANNs in analysing biomass data is still in its early stages, interest in their application and development is growing (Vardiambasis et al., 2020). ANNs work based on a specific structure (architecture), a learning algorithm and a transfer function. By selecting and analysing input data, models can transform this data to produce the desired output, in this case, an estimation of HHV. The effectiveness of ANNs is confirmed by comparing literature-based and model-calculated values (Grossi and Buscema, 2007; Kartal and Oezveren, 2020). Among the different forms of ANN, the multi-layer perceptron (MLP) network is considered one of the most effective in prediction and estimation (Pattanayak et al., 2021).

SVM is based on theories of averaging and are algorithms that can be used for classification or regression through supervised learning. SVM, used as a regression models make predictions by splitting the data into parts for learning and testing the model and are suitable for predicting calorific value (García Nieto et al., 2019). In their research, the authors Dubey and Guruviah (2022) create an SVM model for the prediction and optimisation of HHV in agricultural biomass based on proximate analysis data. The SVM model resulted in a lower prediction error compared to other developed models, with a coefficient of determination (R^2) of 0.905.

RFRs are an effective tool in prediction and are defined as learning algorithms that use multiple random decision branches to predict the output value concerning the mean value of the processed data. They include prediction methods in the form of classification and regression and use a finite (fixed) number of random branches (Scornet et al., 2015; Biau and Scornet, 2016). The authors emphasize that for modelling it is important to determine intervals that contain values with a certain prediction probability. Zhang et al., (2020) and Callejón-Ferre et al. (2014) predict HHV biomass in their research using structural analysis data (content of cellulose, hemicellulose, and lignin). For the prediction, they used regression models based on the input data of the different types of biomass. Authors proved that there is a correlation between HHV and the analysed samples. Akdeniz et al. (2018) developed an algorithm for HHV prediction of different lignocellulosic raw materials based on lignin and extractives (residual materials of biomass) input data. The literature values were compared with the calculated values, and it was found that the developed algorithm showed a greater ability to estimate compared to the offered models. In the research paper titled "Estimation of the higher heating values for lignocellulosic biofuels", the author evaluated the possibility of modelling the HHV of lignocellulosic biomass and biofuels through predictive models based on the input data from various analyses, including proximate, ultimate, and structural analyses. The proposed models varied in terms of R^2 values for the input data from proximate analysis (R^2 = 0.35-0.85), ultimate analysis (0.35 – 0.71) and structural analysis (0.64 – 0.90). Through the statistical analysis conducted, the authors concluded that none of the proposed equations is universal and that it depends on the type of input data, while the highest degree of data agreement was achieved with models based on structural analysis (Petrova, 2021). In their research, Maksimuk et al. (2021) used 30 regression models for prediction using structural analysis data of agricultural biomass from literature sources. The regression correlation models predicted HHV with a lower model error compared to the offered developed models, which makes the model acceptable when applying the estimated energy value. Ozveren (2017) developed the ANN model for HHV biomass prediction in his research, which demonstrates the possibility of applying the model to estimate the calorific properties of biomass, for the ability to solve non-linear and complex calculation problems. Obafemi et al. (2019) used data from the ultimate analysis of biomass (content of carbon, hydrogen, nitrogen, sulphur, and oxygen) to create the ANN model. The model was used to predict the HHV and energy properties to assess the possibility of converting waste into useful energy. Through their research, they demonstrate the applicability of the created algorithm in predicting energy properties. In the scientific study by Kartal and Ozveren (2020), the ANN model was developed to predict the gasification effect of different types of biomass. The developed model successfully predicted the process and predicted HHV with minimal model error compared to the developed models offered in the literature.

The aim of this dissertation is to develop and evaluate non-linear mathematical models for the precise modelling of HHV biomass based on literature datasets. The expected scientific contribution is the creation of universal, reliable models that optimise the time and resources required for modelling, regardless of the type of biomass. The research aims to identify the models with the least error in modelling and enable comparison between HOP, ANN, RFR and SVM models to enrich the existing understanding and practice in this important field.

1.1. Hypotheses and research goals

1.1.1. Hypothesis:

1. The models based on the set of input variables of proximate analysis have the lowest modelling error in all non-linear models examined compared to the sets of input variables of ultimate and structural analysis.

2. The ANN models have a lower error in modelling HHV of biomass than HOP, RFR and SVM, regardless of the set of input variables.

1.1.2. Objectives:

1. Development of new non-linear mathematical models in the form of HOP, ANN, RFR and SVM for the modelling of HHV biomass based on input variables from laboratory analyses (obtained from the literature).

2. Comparison of the newly developed non-linear models and determination of the lowest error in HHV modelling concerning different sets of input variables.

2. REVIEW OF RELEVANT LITERATURE

2.1. Biomass

The use of current fossil fuels and their negative impact on the climate have searched sustainable, RES urgent. As Tun et al. (2019) emphasise, given the speed of climate change and environmental degradation, it is necessary to focus on renewable energy sources. Sustainable energy production is becoming increasingly important in the transition from traditional fossil fuels to renewable sources. Burg et al. (2018) state that biomass as a renewable source offers various applications, not only in the generation of electricity and thermal energy, but also as a fuel. Biomass is, therefore, versatile in the energy sector. In view of the growing demand for sustainable energy sources, biomass and its various applications are playing an increasingly important role. Moreover, as the use of biomass expands, it is crucial to understand the different types and potentials of different forms of agricultural biomass to make the most of this renewable resource (Perea-Moreno et al., 2019).

Agricultural biomass plays a particularly important role in the energy sector and has considerable production potential. Bilandžija et al. (2018) point to a steadily growing interest in the use of solid agricultural biomass for energy production, particularly in Europe. This trend points to the increasing role of biomass in the continent's sustainable energy future. According to Avcıoğlu et al. (2019), residues from agricultural production, such as straw residues, forest prunings and others, are also increasingly in focus worldwide as a RES. The authors add that the energy potential of biomass can vary depending on the different characteristics and proximate properties of the biomass itself. This opens the scope for further research and optimization to make better use of these renewable resources.

Considering the growing needs for energy production from biomass, the properties of which can be determined experimentally, the need to develop mathematical models for calorific value estimation is emphasized to save time and reduce process costs (Kalivodová et al., 2022). As numerous models offered in the literature have a certain degree of modelling errors, it is difficult to create a universal model, i.e. one that is suitable for biomass with different (variable) chemical properties. ML models offer the possibility of solving the estimation of the calorific value of biomass with a lower error rate, as they can solve complex non-linear relationships between the data (Afolabi et al., 2022).

2.2. Agricultural and wood biomass

Agricultural biomass as a source of renewable energy is a key element in efforts to reduce greenhouse gas emissions and promote sustainable development. In a global context increasingly characterised by environmental challenges, the inclusion of biomass in energy paradigms is of great importance. When analysing the benefits it offers, it can be concluded that agricultural biomass has several advantages compared to other RES. Primarily, agricultural biomass is widely available and can be produced locally, allowing countries and regions to utilise local resources, thereby reducing dependence on conventional energy sources. In addition, the utilisation of agricultural biomass does not require significant investment in infrastructure projects. The use of agricultural biomass can lead to a reduction of waste generated by agricultural activities, which has environmental benefits (Zhang et al., 2019; Saini et al., 2015). Agricultural biomass can be divided into biomass from energy crops (grown solely for energy production), agricultural residues, which includes crop residues such as straw and husks, and various organic wastes from agricultural production (Sivabalan et al., 2021). By processing various feedstocks from agricultural biomass such as corn stalks, wheat straw, bagasse, rice husks, fruit husks, corn cobs, husks, and various residues, the need for conventional fuel sources can be greatly reduced. In addition, the agricultural biomass has potential as a raw material for processing into value-added products (Kumar Sarangi et al., 2023).

On the other hand, although agricultural biomass has numerous benefits, it also comes with certain challenges. The collection and transport of biomass can incur significant costs, and the energy value of biomass can vary depending on the type of harvest and processing method. There is also a risk of side effects such as increased competition for land and water resources, which can lead to environmental and socioeconomic tensions. In terms of applications, agricultural biomass can be used in several areas, including the production of energy, bioproducts and biofuels. Forecasts indicate that the use of agricultural biomass will increase in the future in parallel with global efforts to reduce greenhouse gas emissions. However, the future of this energy segment will depend on several factors, including technological innovation, regulatory frameworks, and market demand (Searchinger et al., 2008; Mohanty et al., 2002).

Wood biomass is one of the most important renewable feedstocks for energy production due to its potential benefits in reducing greenhouse gas emissions and promoting a sustainable energy economy. Forest biomass can be divided into primary and secondary. Primary biomass includes materials obtained directly from the forest, such as whole logs and

trees, post-cutting residues, and thinner stems and saplings. Secondary forest biomass refers to the waste generated from the processing of wood in sawmills and other industrial facilities, including sawdust, crisps, bark, and boards and residues (Titus et al., 2021). The technical feasibility of its use as a raw material for bioenergy and bioproducts has been thoroughly investigated. Wood pellets, which are made from pressed sawdust and other wood residues, are a popular form of this type of biomass. However, despite the potential reduction in greenhouse gas emissions, unsustainable management of this biomass can have negative consequences for the environment, such as deforestation and other environmental problems (White, 2010).

2.3. Biomass properties

As one of the most important RES, biomass has been analysed in detail, including its proximate, ultimate, structural analysis and energy properties. The calorific value is the most important characteristic for determining fuel quality and indicates the amount of energy that is released when a raw material is completely burnt. It is expressed in MJ kg $^{-1}$. In general, the HHV varies between 17 and 23 MJ kg^{-1} for the lignocellulosic biomass, according to Esteves et al. (2023). The direct combustion of biomass produces minimal amounts of greenhouse gases and other potentially hazardous pollutant emissions compared to conventional (fossil) fuels. The emission of substances released during combustion of biomass varies considerably depending on the composition of the biomass, the type of combustion plant used and the settings of the combustion process. The most important factor influencing combustion efficiency and the number of emissions is the composition of the biomass burned. The properties of biomass are reflected in its composition, and the analysis of properties in terms of energy modelling includes ultimate (carbon - C, hydrogen - H, nitrogen - N, oxygen - O and sulphur - S), proximate (fixed carbon - FC, volatiles - VM and ash) and structural analysis (cellulose, hemicellulose, and lignin). It is important to point out that the presence of these elements in biomass varies greatly depending on the type of biomass itself and the conditions of its cultivation (Lisý et al., 2020). Biomass with a high moisture content (MC) has a lower density, which leads to a reduction in the energy required for water evaporation. The percentage of MC has a major impact on the transport, storage and processing of biomass, as well as on the amount of energy obtained from a given material (Schaffer et al., 2015). The chemical composition of biomass is used to characterise it in terms of fuel quality. Different groups of biomass are considered in the analysis, e.g. agricultural and wood biomass. The chemical elements in the biomass can be classified as

micro and macro elements according to their concentrations. It is important to emphasise that there are significant differences in the chemical composition of different types of biomass

2.3.1. Calorimetric analysis

The calorific value can be a higher heating value (HHV) or a lower heating value (LHV). The HHV refers to the heat (energy) extracted from the fuel during combustion with condensed (generated) vapour. LHV contains the latent heat of the water vapour that is produced during combustion by the condensation of water vapour into a liquid. The relationship between the HHV and LHV is such that LHV represents the correction of HHV due to moisture in the fuel (biomass). In general, the calorific value can be determined experimentally using an adiabatic bomb calorimeter, which measures the enthalpy change between reactants (fuel) and products (water). However, measuring the energy value is a complex and time-consuming process that requires the use of specialised equipment (Acar and Ayanoglu, 2012).

Analysing the raw material (biomass) is an important process to determine the quality and the possibility of using it as fuel. The heating value of biomass is measured with an adiabatic bomb calorimeter using the ASTM standard (White, 2017). In calorimetry, the biomass is burned in a bomb and the heat of combustion is transferred to the medium outside the tank. Thermometers were used to measure the temperature of the water and the energy value, which was then calculated based on the change in water temperature and the correction formula (Wang et al., 2021). Adiabatic bomb calorimeters are sometimes referred to as quasi-adiabatic due to the existing heat exchange with the environment. The exchanged heat is calculated using Newton's law of heat transfer or more complex equations describing heat transfer. Calorimeters that measure heat in the way described above are called passive adiabatic, i.e. isoperibolic, when they are in a constant temperature environment (Horvat, 2015). When measuring with an adiabatic calorimeter, different temperatures and energy units are used. To reduce the possibility of errors in the interpretation of the data, it is necessary to understand the relationships of the individual data within the process. The adiabatic bomb calorimeter is a standard instrument for measuring the energy value of solid fuels and liquid combustible samples. The heating value (combustion energy value) of a particular sample can be defined as the amount of energy released during complete combustion at a constant volume. The HHV is the most important parameter in the planning and modelling of energy operations for systems based on biomass fuels. In addition, the HHV is an important parameter in optimising the conversion of biomass into a useful fuel that can be subjected to various conditions. The experimental method of determination is always expensive and time-consuming, so it is recommended to use

different mathematical models in the evaluation (Xing et al., 2019). The importance of HHV modelling lies in the possibility of accelerated assessment, the design of energy systems and the possibility of using raw materials as fuel (Taki and Rohani, 2022).

2.3.2. Ultimate analysis

Ultimate analysis of the composition of biomass enables the precise determination of the content of key elements such as C, H, N, S and O in biomass. With this method, it is possible to obtain information about the composition of biomass on a dry and ash-free basis, allowing a deeper understanding of its chemical structure and potential applications. A device known as an elemental analyser is used to determine the elemental composition. The process involves taking a small sample of biomass, the weight of which is known in advance, in powder form. This sample is then burnt in a strictly controlled atmosphere. During this process, different elements in the biomass react and turn into gases: Carbon turns into CO_2, hydrogen turns into water (H_2O), nitrogen turns into (NO) and S turns into SO_2. After combustion, the released gases flow out of the chamber via a specific segment of heated high-purity copper. This phase of the process is crucial because the copper acts as an oxygen scavenger and simultaneously converts NO_x to N_2. The final reduction and quantification of the elements C, H, N and S is based on the identification and measurement of the formed gases CO_2, H_2O, N_2 and SO_2 (Shadangi et al., 2023).

Modelling the energy values of biomass based on ultimate analysis provides a highly relevant data set, as this method does not require analysis of the microelements, but focuses solely on the percentage of the elements C, H, N, S and O. This approach has long been at the centre of scientific interest, particularly in the theoretical estimation of the calorific value of coal, as demonstrated by the research of Ozyuguran et al. (2018). The ultimate analysis, defined as the chemical composition of a given fuel, allows the creation of models that are more precise and widely accepted compared to those based on microelements, according to the studies by Onochie et al. (2023) and Bychkov et al. (2017). This method is thus crucial for accurately determining the heating value of biomass with the aim of understanding and utilisation of this RES.

2.3.3. Structural analysis

Lignocellulosic biomass mostly consists of a structure comprising cellulose, hemicellulose, and lignin. The distribution of these components in the biomass is a complex structure and depends on the type of biomass. Knowing the composition of biomass, i.e. its structure, is a very important aspect for its use in industrial applications (Díez et al., 2020). Cellulose is a plant component composed of elongated chains of at least 500 glucose molecules, which is why it is categorised as a polysaccharide. Over time, cellulose has gained a significant role as a raw material in modern industry. It is used in various sectors, including the fibre and textile industries, paper production, animal nutrition, cosmetics, and pharmaceuticals. In the plant structure, cellulose is often associated with two main components: Hemicellulose and lignin, which are amorphous by nature. High concentrations of cellulose can be detected in agricultural residues such as palm chaff, cocoa pods, plantain peels, banana leaves, corncobs, wood, and sugar beet waste (Pinto et al., 2022). Hemicellulose is a type of polysaccharide found in plant cell walls and is the second most abundant renewable polymer in lignocellulosic materials after cellulose. It is characterised as a heteropolysaccharide with a complex structure containing various monosaccharides such as glucose, xylose, mannose, galactose, arabinose, rhamnose and acids such as glucuronic and galacturonic acid, the proportion of which varies depending on the source (Peng et al., 2012). Lignin contributes significantly to the recalcitrance of lignocellulosic biomass, which hinders its conversion into biofuels and other value-added products (Watkins et al., 2015). The ratio of cellulose to lignin is a crucial factor in determining the suitability of a particular plant species for energy production. For combustion processes, a lower proportion of cellulose and hemicellulose in the biomass is preferable, i.e. biomass with a higher lignin content is more suitable for direct combustion processes (Voća et al., 2021). The lignin content influences the increase in the energy value of biomass, which is why its valorisation is important (Beaucamp et al., 2022). Dorez et al., (2014) emphasises that the effect of increasing the flammability of the material is associated with an increased lignin content.

2.3.4. Proximate analysis

Proximate analyses provide information about the content of volatile matter (VM), moisture content (MC), ash, and fixed carbon (FC) in the biomass. A detailed procedure is required for the exact determination of these components. The equipment used for this process includes a temperature-controlled muffle furnace, a highly sensitive analytical balance, a porcelain dish, a temperature-controlled precision dryer and a desicator. The equipment ensures accuracy and consistency of results and provides a comprehensive understanding of the composition of the biomass. The biomass contains a high proportion of VM, which is calculated using a formula that considers the differences in the mass of the sample. The ash content in the biomass represents the inorganic residues that remain after combustion, the presence of which depends on the origin of the plant. Ash measurement provides information about the inorganic composition of the biomass (Shadangi et al., 2023). Higher-quality fuels produced from biomass have a lower ash content. With an increase in ash, the calorific value of the biomass decreases (Kwaghger and Iortyer, 2017). When modelling biomass and its potential applications in energy-related scenarios, the proximate analysis dataset is often used (Nimmanterdwong et al., 2021).

Various studies have demonstrated the feasibility of using primary analysis data for estimating the HHV of biomass (Chen et al., 2022). Estimation of HHV based on proximate analysis is suitable data set as it reduces the cost and time required for determination under laboratory conditions (Roy and Ray, 2020). This method streamlines the process of assessing biomass HHV for energy applications, making it an attractive option for researchers and practitioners looking for efficient and cost-effective analytical approaches.

2.4. Mathematical modelling

Mathematical modelling is a key element in scientific and technical research for the quantitative analysis and interpretation of phenomena. Regression models, as a subset of mathematical modelling methods, are particularly useful in engineering applications such as modelling the calorific value of different types of fuels or materials. These models allow the identification and quantification of relationships between dependent and independent variables, leading to reliable predictions and system optimization. In the last decade, non-linear regression models have been increasingly used, as they are able to model complex relationships between different variables that are often inherently non-linear. This type of modelling provides greater accuracy in predictions, which is critical in advanced engineering applications where small deviations can lead to significant changes in system performance. The importance of energy conservation and proper design is particularly evident in the context of power generation.

One of the increasingly applicable tools in the field of energy problem modelling is machine learning. In recent years, machine learning has emerged as a powerful tool for extracting insights from data and refers to a set of techniques that can automatically extract patterns in data, usually large amounts of data. Due to recent improvements in methods, computing infrastructure and data availability, these techniques have become ubiquitous in numerous applications (Donti and Kolter, 2021). Research in the field of computer architecture has focused on the energy efficiency of production processes for decades. The field of machine learning, on the other hand, has mostly focussed on creating highly accurate models, often without considering energy consumption. This is particularly pronounced in the field of deep learning, where the complexity and precision of the model take precedence over energy efficiency. Such models are becoming increasingly demanding in terms of computing and memory resources. Their creation requires considerable computing power, especially in the phase of training on large data sets. During implementation, on the other hand, these models are often used multiple times (García-Martín et al., 2019). The development of precise models enables better planning and optimization of the energy production process, resulting in higher energy efficiency and lower resource consumption. In addition, mathematical modelling offers the possibility to simulate different scenarios and conditions, facilitating informed decision-making. Integration with advanced machine learning algorithms can further improve the accuracy of predictions and thus ensure the sustainability of energy systems.

2.4.1. Linear modelling

In the field of statistical modelling, regression analysis is used to evaluate relationships between variables. The focus is on the relationship between the dependent variable and one or more independent variables or predictors. In the above context, the main objective is to find a mathematical model that best explains the relationships being analysed. Regression analysis helps to understand how the dependent variable changes about changes in the independent variables, assuming that other variables are held constant. This approach enables the estimation of the values of the dependent variable based on the specified independent variables. The main objective is to formulate a regression function that provides accurate predictions and allows interpretation of the relationships between dependent and independent variables. In certain circumstances, regression analysis can provide a basis for concluding causal relationships between variables (Sarmento and Costa, 2017).

2.4.2. Non-linear modelling

Non-linear modelling is a statistical method for representing complex relationships between variables that do not follow a linear pattern. It is used in various fields to understand and predict non-linear phenomena. Although non-linear models can be more accurate than linear models when dealing with complex data, they are often more difficult to estimate and interpret. Nonlinear models can be computationally intensive and require more data compared to linear models. There are different types of non-linear models, which can be categorised according to the form of the function, the method of parameter estimation, the type and complexity of the model. These models use different methods to model non-linear relationships between variables (Abbasi Nozari et al., 2012; Guo and Ringwood, 2021; Mustapa et al., 2019).

Linear and non-linear mathematical models differ in the type of relationship they describe between independent and dependent variables. A linear model assumes that the relationship between the variables is linear, which means that a change in the independent variable causes a proportional change in the dependent variable. A non-linear model, on the other hand, describes a situation in which the relationship between variables cannot be predicted by a linear regression line (Paredes-Astudillo et al., 2022). Models are defined as linear if all operators of the mathematical model express linearity. For example, in a statistical linear model, the relationship is assumed to be linear for the parameters, but it does not have to be linear for the predictors. On the other hand, a differential equation is linear if it can be written

with linear differential operators, although it may also contain non-linear expressions. On the contrary, a mathematically programmed model is considered linear if the objective and constraint functions are completely represented by linear equations. If one or more objective or constraint functions are represented by a non-linear equation, then the model is said to be non-linear (Crooke et al., 2002). The main difference between linear and nonlinear models lies in their ability to model complex relationships. While linear models are simple and easier to interpret, they are often limited when it comes to describing real, often complex, phenomena. Linearity "in the parameters" means that although the predictors themselves can be transformed or combined in a non-linear way, the relationship between the transformed predictors and the expected response remains linear (Olive, 2017). In comparison, non-linear models can describe a much broader range of relationships, including those where changes in the independent variables cause non-linear, often unpredictable changes in the dependent variables (Crawley, 2012).

2.5. Non-linear and machine learning regression models

In the modern data environment, linear methods often do not provide satisfactory results due to the inherent complexity of the phenomena being analysed. Nonlinear regression models allow for more flexible and precise predictions and interpretations of data, although their implementation and analysis can be computationally more demanding and conceptually more complex. Artificial neural networks (ANNs) represent a highly adaptable class of models that simulate the way biological neurons process information. They can be used in a variety of tasks and areas, from image classification to financial market predictions. Support vector models (SVM) are another class of nonlinear models that focus on finding the optimal hyperplane for splitting or approximating data. Although they are often used for classification tasks, they can also be adapted to regression problems and provide accurate and robust ways of modelling nonlinear relationships. Random Forest Regression Models (RFR) are a method that combines the predictions of multiple underlying models to achieve higher accuracy and robustness. These models are particularly useful when working with data that is noisy, incomplete or contains many features. Higher degree polynomials offer a simpler but extremely useful way of modelling non-linear relationships. While they do not offer the same level of flexibility as the previously mentioned models, they are often easier to interpret and implement, making them suitable for quick analyses and research.

2.5.1. Artificial neural network (ANN) models

In the field of big data management and with the increase of computing capacity, ANN show high performance in solving classification and regression problems (Cao et al., 2018).

ANN can be defined as a system of interconnected artificial neurons on which the operations of this model are based. In general, models are built with three layers: Input, hidden and output. When the neurons receive information from different inputs, they generate nonlinearity through activation functions. It is important to emphasise that ANN models are highly dependent on the data on which they are trained (Chen et al., 2020). ANN drawing inspiration from the neurological structures of the human brain, serve as sophisticated information processing systems that excel in identifying and processing patterns. Notably, ANNs possess the capability to continuously enhance their performance by analysing outcomes from previously resolved tasks (Haglin et al., 2019). These models have gained widespread recognition as universal evaluation tools across a multitude of fields, such as pattern recognition, processing, function finding, and process simulation. Their suitability is particularly noted when the primary objectives encompass modelling, predicting, evaluating specific outcomes, and unravelling the complexities inherent in nonlinear data interactions (Puig-Arnavat and Bruno, 2015).

Machine learning methods, especially ANN, show significant capabilities and capacities in modelling and assessment (Masi et al., 2021). Choosing the right topology of a neural network is crucial for the optimal application of ANN and deep learning (DL) models. It is often recommended to choose a feed-forward network with a larger number of neurons, as numerous empirical indicators suggest that it is more efficient to optimise and adjust the weights while avoiding the risk of overfitting. The choice of the correct network topology is determined by trial and error (Mesroghli et al., 2009). The selection of the appropriate loss function is also an essential component for the efficiency of ANN and DL models. There are general guidelines for selecting an appropriate loss function depending on the type of response variable, whether it is continuous, binary, or categorical. (Montesinos López et al., 2022). The choice of the number of hidden layers is crucial for the configuration of a deep learning model. Although the universal approximation theorem suggests that a single hidden layer with many neurons can approximate any function, in practice multiple layers are often required, especially when the data is not linearly separable. If the data is linearly separable, deep learning models and ANN may not be necessary, but a single hidden layer is sufficient for most problems. The number of neurons in the network is crucial: too few neurons in the hidden layers can lead to underfitting, while too many neurons can cause severe overfitting. In the input layer, the number of neurons is determined by the number of input features,

while in the output layer, the number of neurons is determined by the number of output variables or the number of classes of the response variable if it is ordinal or categorical. The learning rate adjusts the weights and thresholds (bias) of the neural network so that a given input to the network produces the desired output. (Montesinos López et al., 2022).

In situations where huge amounts of information need to be processed, ANN become important tools, as conventional processing methods are often not efficient enough for such tasks. When using ANN, the desired output value is based on a comparison between predicted and actual data (Jat et al., 2018). One of the main advantages of using ANNs is their flexibility in modelling and the ability to efficiently deal with robust data sets. They enable the identification and modelling of complex nonlinear relationships between input and output variables. However, for the model to be effective, it is crucial to select and accurately adjust the network parameters for each specific application (Abdolrasol et al., 2021). One of the most popular and efficient types of ANN is the multilayer perceptron (MLP), which consists of input, hidden and output layers. The main advantage of this approach lies in the flexibility of customising the network architecture based on the input specifications and the target output. The intrinsic properties of the ANN offer the possibility of autonomous learning through iterative cycles. Through this process, the ANN continuously adapts to the available information and efficiently implements complex mathematical transformations for precise data processing. The structure of the ANN, including the number of neurons and hidden layers, is not static. To achieve the optimal network architecture that best fits a particular problem, it is often necessary to make a number of experimental adjustments (Ozveren, 2017). Depending on the architecture, topology and learning method, ANN models can be categorised in different ways (Mesroghli et al., 2009). The learning process can either be supervised, where the model has access to the target output data, or unsupervised, where the model must discover the data structure without clearly defined outputs (Cinar, 2020). The methodology of the ANN model includes at the beginning the definition and purpose of the model, the determination of the input data, the division of the data into learning and testing parts, the definition of the network structure, the selection of an appropriate algorithm, the standard methods for converting input variables into input data, the learning process and the testing process (Vanneschi and Silva, 2023).

The main advantage of the ANN model is the ability to find complex relationships between variables, with a high tolerance for missing data and the possibility of real-time evaluation (Bermejo et al., 2019). The evaluation of the training process is one of the most important operations for the general improvement of the modelling process and the possibility of real application of the ANN model (Lujano-Rojas et al., 2023). In addition to training the model, the ANN also requires the adjustment of the hyperparameters, i.e. the

optimisation of the parameters that determine the operation of the algorithm, so that the model must be trained in several iterations. The error and accuracy of the model are highly dependent on the optimisation of the model, therefore increasing the performance of the ANN model is based on the trial-and-error method (Matveeva and Bychkov, 2022). ANN consists of interconnected layers consisting of artificial neurons (nodes), connections, weight coefficients, bias thresholds, and activation and transfer functions. They are interconnected and their "strength" is determined by the weighting coefficient. These connections are analogous to the connections in the human brain (Casas, 2019). In recent years, research has focused heavily on improving artificial intelligence algorithms using various optimisation techniques. One of these methods is ANN with multiple hidden layers. ANN models are applicable in various fields. Due to their adaptability and software engineering, such algorithms are the preferred machine learning option in the field of learning and estimating desired output values (Abdolrasol et al., 2021). ANNs are considered nonlinear statistical models that work on the principle of a biological neuron and are considered effective, useful, and successful in the field of nonlinear problems and pattern recognition (Du and Swamy, 2006).

2.5.1.1. Application of mathematical functions in ANN modelling

From a feasibility point of view, nonlinear activation functions are key elements of hardware implementations in ANN. The research mainly focuses on various aspects such as accuracy, the use of approximation methods and the cost of implementing these systems, considering analogue and digital platforms. Accurate implementations of nonlinear activation functions have a significant impact on improving the learning and generalisation capabilities of ANNs (Shakiba and Zhou, 2021). Exponential functions are often used as activation functions in ANNs because they can model phenomena that increase or decrease at a constant rate, representing the activation of an artificial neuron in response to a stimulus. In addition, exponential functions can be used to model the probability of an event, which is particularly useful in sentiment analysis. These functions are often combined with other activation functions such as sigmoidal and step functions (ReLu) to create more complex neural network structures that can model a wider range of phenomena (Urban, 2018).

Sigmoidal functions are mathematical functions used in ANN models and generally in deep learning models. This function assigns an output range of 0-1 to each input value X. As its curve has the shape of the letter S, it is ideal for "squeezing" input values into this range. The properties of the sigmoidal function include the range, which includes all real numbers,

and the horizontal asymptotes to which the function tends when X tends towards positive or negative infinity. In addition, the function has an inflexion point at x=0, where the curve changes from concave to convex or vice versa. Because of these properties, the sigmoidal function is particularly useful in tasks that require output probabilities (Hamdan and Roach, 2022). An identity function is an activation function that produces an output value that is identical to its input value. Unlike other activation functions such as the sigmoidal or step function (ReLU), the identity function does not transform the input values, so it can be calculated and implemented quickly. However, due to its simplicity, it is often not useful in complex neural networks where a higher degree of significance is required. It has a constant derivative equal to 1, which facilitates the learning process, but can lead to problems such as vanishing or exploding gradients in deep neural networks. The domain and codomain of the function include all real numbers, so it can accept and generate any real number (Wanto et al., 2017).

In the context of ANN modelling, a tangential hyperbolic function is often used as an activation function. This function has the range of all real numbers in the interval from -1 to 1 (Shakiba and Zhou, 2021).

2.5.2. Support vector machine (SVM) models

SVM represent a robust machine learning algorithm that is crucial for pattern recognition. It is particularly useful for nonlinear modelling of high-dimensional data with a limited number of samples. SVM works by solving inner product operations in a high-dimensional feature space. This is achieved by using the kernel function in the low-dimensional space, which avoids the so-called chaos of dimensionality. The value of the kernel function measures the similarity between the feature vector and the training and prediction samples. There are three primary kernel functions within SVM, and the selection of the appropriate kernel function is critical for successful nonlinear classification or regression. The selection of the optimal decision function, which is essential for the SVM categorisation model, is performed as a solution to an inequality constrained quadratic optimisation problem. This optimal decision function, which is global and unique due to the convex nature of the optimisation problem, allows SVM to avoid local minima and provides a sparse solution with a simple geometric interpretation (Wang et al., 2019). The SVM decision function is actually an optimal hyperplane, which is used to separate observations belonging to one class from another based-on patterns of information about these observations, called features. This hyperplane is used to determine the most likely label for unknown data. The features used for hyperplane inference are usually not raw data, but mostly derived data resulting from some kind of interpolation during the feature selection phase. The features are also referenced by

coordinates that are based on their relationships to each other and form support vectors. As with other forms of machine learning, working with SVM is about reconciling two complementary goals: maximising the percentage of labels correctly assigned by the classifier to new examples (i.e. optimising its accuracy) and ensuring that the classifier is generalisable to new data (i.e. optimising its reproducibility). While the first point is limited by the informativeness of the features used, the second point is limited by the number of unique examples used to train the model (Pisner and Schnyer, 2019).

Three basic phases are distinguished in the SVM analysis:

1. Selection of features

2. Training and testing of regressors or classifiers

3. Evaluation of performance

In addition, Gholami and Fakhari (2017) mention the steps of SVM implementation in the field of nonlinear modelling:

- Sample matrix preparation: For regression analysis, the independent (input parameters) and dependent data (target parameters) are often presented in separate matrix columns. The data can then be further divided into current, test and validation parts.
- Selecting the kernel function: It is necessary to select the function that generates the hyperplane that is furthest away from all the data in the set.
- Selection of parameters: The use of SVM involves the selection of key parameters such as the kernel function, the trade-off parameter, and the error insensitivity (ε), which is achieved by applying a validation procedure to determine them.
- Performance of the learning algorithm: During model training, SVM uses formulas to compute Lagrange multipliers that identify support vectors to select the optimal hyperplane from the input and output data.
- Data classification and modelling: Lagrange multipliers and support vectors enable accurate classification and modelling of new data. A large modelling error may indicate a need for improvement in feature extraction, feature selection or parameter estimation.

In the feature selection phase, the original raw data is converted into a set of features that can be used as input for the SVM. The training and testing phase of the classifier involves the use of examples for which the classes are already known in advance. This enables supervised model learning, and the information is used to model or predict new

class labels. The final phase of evaluation is to check how accurately the model can classify the data (Pisner and Schnyer, 2019). Zoppis et al., (2018) state that MFAs have made an outstanding contribution to various scientific fields where statistical inference is currently applied to many important problems, as evidenced by the exponential growth in the application of the model. SVMs are intuitive models that, by identifying the maximum edge of the hyperplane separating positive from negative classes, classify new elements depending on the "half-space" in which they are located concerning the separating hyperplane.

2.5.3. Random forest regression (RFR) models

Random forest models for regression are a combination of tree predictors such that each tree depends on the values of a random vector, where the data is processed independently of the same distribution for all trees in the forest. In contrast to classification models, where trees are created based on categorical target variables, RFR models are based on numerical values of the target variables, resulting in output data in numerical values (Jaiswal and Samikannu, 2017). The margin of error for the above models converges towards the limit as the number of "trees" in the model increases. RFR enables the unique predictive accuracy and interpretability of the model as an ML model. With the help of random sampling and ensemble strategies, a more accurate prediction and generalisation of the model is achieved. The key features of the RFR model are the ability to predict the output value in different applications, measure the importance of each feature (input data), and identify the closeness of pairs in model training (Qi, 2012).

The error of the classifier depends on the weighting of the individual trees in the model as well as on the correlation between them (Scornet, 2015). RFR are easy to fit to nonlinearities in the data and therefore usually have a higher predictive power. In terms of regression and modelling, they are suitable for medium and large data sets. When the number of independent variables is larger than the number of observed variables, models such as linear and logistic regression do not show high performance as the number of parameters to be estimated exceeds the number of observed parameters. RFR show high effectiveness in such cases as they do not use all predictor variables at once. Fitting the RFR model involves choosing optimal values for key hyperparameters such as the number of trees, the maximum depth, the number of features at each split, and the minimum number of samples required for a split. The most effective fitting strategies include grid search, random search and model-based sequential optimisation (Probst et al., 2019). The RFR algorithm is based on grounded trees, which serve as its building blocks. In a tree-based model, a given data set is

recursively divided into two groups based on a specific criterion until a predetermined stopping condition is met. At the bottom of the decision trees are the so-called leaves or end nodes. These models work by choosing the best possible way to split the data at each step by analysing how the data behaves about the target variable. Each split creates two new nodes, and the process continues until a stopping condition is reached. This may be the maximum depth of the tree, the minimum number of samples required for further splitting, or when no improvement in prediction can be achieved (Schonlau and Zou, 2020).

2.5.4. High order polynomials

Approximation theory is a classical branch of mathematics that investigates the extent to which functions can be approximated by simpler functional forms. It has found numerous applications in computer science. Most of these applications of approximation theory focus on the approximation of functions to polynomials using a uniform norm (or infinity norm) (Bun and Steinke, 2015). The problem of finding the roots of a polynomial equation is important because many calculations in engineering and scientific computing can be traced back to it. Solving functional equations is a problem that often arises in practical applications. Many calculations in technical and scientific computing can be reduced to the problem of solving a polynomial equation. Polynomials can approximate or fit all continuous functions very well. Therefore, the study of solving polynomial equations is of great theoretical and practical importance. As we all know, the roots of a polynomial of degree higher than 4 cannot be expressed by a formula (Wang et al., 2011).

The approximation by a polynomial of a higher degree has several advantages over the approximation by a polynomial of a lower degree. For example, as the length of the interval increases, the approximation errors decrease. The number of segments on the contour is reduced and the information characters are calculated more accurately. The higher the degree of the polynomial, the smaller the number of mesh nodes and the better the accuracy and quality of the approximation. The higher the degree of the polynomial, the smaller the number of nodes and the better the final solution within the approximation interval. At the same time, increasing the degree of the polynomial usually leads to a loss of stability, a significant increase in the error of the predicted value during extrapolation, an increase in the order of the derivatives used, and an increase in the complexity of the calculation. Therefore, formulas for smoothing polynomials of higher degree are almost completely avoided. The approximation by a higher order polynomial makes it possible to optimise the number of segments on the contour and to obtain the analytical dependence of the curvature for a more

precise calculation of informative signs that are invariant with respect to geometric transformations. The effectiveness of polynomial approximation and smoothing depends on many factors, the most important form being the polynomial (Dikusar, 2018).

2.5.5. Model optimization

Prediction and optimisation form the basis for many real analytical problems that arise in various disciplines. Due to their complexity, they are usually treated sequentially in existing studies, with the prediction problem being solved first, followed by optimisation. In this paradigm, unknown parameters in an optimisation problem are predicted by a prediction model and then used in the optimisation model to make optimal decisions (Yan and Wang, 2022). Mathematical programming techniques include various methods and programme structures for solving linear and nonlinear models. The applicability of these methods depends on the mathematical structure of the model and the system being analysed. Each method has its advantages and limitations, so it is important to know the different types of optimisation methods (Loucks and van Beek, 2017). Once the model has been created, the next step is its optimisation, a process that is crucial for achieving optimal solutions and improving system performance (DiMaio and Chiu, 2016). Model optimisation involves the application of various mathematical and computational methods to maximise or minimise a particular function within the model to find the most efficient solution to a given problem. The use of these optimisation techniques can be greatly facilitated by integrating them into software decision support systems. This enables faster and more accurate analyses and provides insights into complex problems, improving the quality of decisions (Hewitt and Frejinger, 2020). This integrated approach enables users to effectively solve problems in various fields, from finance and logistics to engineering and resource management, using advanced algorithms and computer modelling to achieve optimal results. In the age of machine learning, performance based on accuracy and calculation time is one of the most important parameters in optimisation.

Numerous parameters associated with machine learning modelling methods are time-consuming, so standard optimisation methods are not considered suitable. In the ML model, two types of parameters are subject to optimisation: hyperparameters and model parameters. The hyperparameters are set before and during model training, while the model parameters are changed during learning. The quality of the prediction model depends on the configuration of all hyperparameters (Hossain and Timmer, 2021).

2.5.6. Sensitivity analysis

Sensitivity analysis in ML models is a multi-faceted approach that involves identifying important features in a data set, understanding the impact of each feature on the prediction of the model, and estimating the variance in model performance for different sizes of data sets. In the context of modelling, hyperparameter sensitivity analysis is applied to identify the correct parameters for model performance (Naik and Kiran, 2021; Kim et al., 2022). Global sensitivity analysis (GSA) plays a role in identifying the input variables that significantly affect the behaviour of the model under the conditions of nonlinearity. Sensitivity analysis is also key to interpreting ML models by analysing the effects of the input variables and their (relative) importance in determining the output value (La Rocca and Perna, 2022). The complexity of ML models, especially those with a larger number of parameters, poses a particular challenge in the application of sensitivity analysis due to the fundamental differences between the different model types (Razavi et al., 2021). In application, sensitivity analysis involves assessing the dependence of a model's output on its inputs, often using a relevance factor to measure the impact of each parameter on the model's output (Baghban et al., 2019).

Sensitivity analysis is a cornerstone of modelling and simulation and provides invaluable insight into the influence and relative importance of input parameters on model predictions. Using global sensitivity methods, as described in detail by Li et al. (2016), this approach quantitatively assesses how different input variables affect model outcomes, deepening the understanding of their importance in the overall modelling process. Khoshroo et al. (2018) also highlight the role of sensitivity analysis in setting research priorities by identifying and ranking the most influential factors that significantly improve the quality of results. This type of analysis is crucial for assessing the impact of variations in input variables on output results, an essential step in improving the accuracy and reliability of models. In the field of ANN, sensitivity analysis becomes a key tool to analyse and identify the input variables that significantly affect the performance of the network. This is particularly important for complex models where the accuracy and reliability of the results are of paramount importance. By identifying key variables, sensitivity analysis helps to optimise these models and leads to more accurate and reliable predictions. Machine learning (ML) algorithms are commonly used tools in engineering practice due to their enormous application potential in various nonlinear problems, especially in the management of high-dimensional data. Handling such data poses a challenge in analysing and making decisions. For this reason, the method of global sensitivity analysis is used, which reveals the meaning of individual variables and the relationship between input and output data to discover optimal patterns (Zhang, 2019). Puig-

Arnavat and Bruno (2015) emphasise the importance of using sensitivity analysis based on ANN models in biomass-related modelling processes to uncover the importance of individual variables. This integration of sensitivity analysis with ML and ANN models in biomass research and other high-dimensional data scenarios improves the effectiveness and accuracy of modelling efforts, making it an important component in modern engineering practice.

3. RESULTS AND DISCUSSION

3.1. Review of published qualification papers

3.1.1. Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis

Miscanthus is a perennial energy crop that produces high yields and has the potential to be converted into energy. The ultimate analysis determines the composition of the biomass and the energy value in terms of the higher heating value (HHV), which is the most important parameter in determining the quality of the fuel. In this study, an artificial neural network (ANN) model based on the principle of supervised learning was developed to predict the HHV of miscanthus biomass. The developed ANN model was compared with the models of predictive regression models (suggested from the literature) and the accuracy of the developed model was determined by the coefficient of determination. The paper presents data from 192 miscanthus biomass samples based on ultimate analysis and HHV. The developed model showed good properties and the possibility of prediction with high accuracy (R^2 = 0.77). The paper proves the possibility of using ANN models in practical application in determining fuel properties of biomass energy crops and greater accuracy in predicting HHV than the regression models offered in the literature.

Keywords: artificial neural network; prediction; miscanthus; energy potential

3.1.2. Energy Potentials of Agricultural Biomass and the Possibility of Modelling Using RFR and SVM Models

Agricultural biomass is one of the most important renewable energy sources. As a byproduct of corn, soybean and sunflower production, large amounts of biomass are produced that can be used as an energy source through conversion. In order to assess the quality and the possibility of the use of biomass, its composition and calorific value must be determined. The use of nonlinear models allows for an easier estimation of the energy properties of biomass concerning certain input and output parameters. In this paper, RFR (Random Forest Regression) and SVM (Support Vector Machine) models were developed to determine their capabilities in estimating the HHV (higher heating value) of biomass based on input parameters of ultimate analysis. The developed models showed good performance in terms of HHV estimation, confirmed by the coefficient of determination for the RFR (R^2 = 0.79) and SVM (R^2 = 0.93) models. The developed models have shown promising results in accurately predicting the HHV of biomass from various sources. The use of these algorithms for biomass energy prediction has the potential for further development.

Keywords: agricultural biomass; higher heating value; machine learning; estimation; energy potential.

3.1.3. Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass

The aim of this study was to investigate the potential of using structural analysis parameters for estimating the higher heating value (HHV) of biomass by obtaining information on the composition of cellulose, lignin, and hemicellulose. To achieve this goal, several nonlinear mathematical models were developed, including polynomials, support vector machines (SVMs), random forest regression (RFR) and artificial neural networks (ANN) for predicting HHV. The performed statistical analysis “goodness of fit” showed that the ANN model has the best performance in terms of coefficient of determination (R^2 = 0.90) and the lowest level of model error for the parameters X^2 (0.25), RMSE (0.50), and MPE (2.22). Thus, the ANN model was identified as the most appropriate model for determining the HHV of different biomass based on the specified input parameters. In conclusion, the results of this study demonstrate the potential of using structural analysis parameters as input for HHV modelling, which is a promising approach for the field of biomass energy production. The development of the model ANN and the comparative analysis of the different models provide important insights for future research in this field.

Keywords: structural analysis; support vector machine; artificial neural network; random forest regression; high order polynomials.

3.1.4. Assessing the properties of *Miscanthus x Giganteus* under varying levels of ash fertilization treatment and regression neural network insight into calorific value

The aim of the study was to investigate the changes in ultimate, proximate analysis and calorific properties of *Miscanthus x Giganteus* with three types of planting materials (two rhizomes - R1 and R2 - and one seedling – S) and three ash fertiliser treatments (P_0, P_2, and P_5) were included in the study. The research further examined their effects on crop yield, stem height and various chemical properties. The results showed that the maximum yield was obtained with the R1 x P_2 plant type, while the minimum yield was recorded with the R2 x P_2 plant type. In addition, the greatest mean stem height (3.34 m) was recorded for the R2 x P_5 plant type. Significant differences were also found in the chemical components between the plant types and treatments. For example, the highest ash content of 2.25% was found in plant type 'S' x P_5, while the highest coke content of 14.48 % was found in plant type R1 x P_5. The statistical analysis confirmed that planting material and ash fertilisation had significant influence on the physicochemical properties of Miscanthus x Giganteus. This consequently affects the calorific value, with the mean higher and lower heating value being 18.32 and 17.04 MJ/kg, respectively. The neural regression network models showed robust predictive performance for the higher (HHV) and lower heating value LHV, with low chi-square values (X^2) and high coefficients of determination (R^2).

Key words: Miscanthus x Giganteus, fertilisation, energy properties, artificial neural network, modelling.

3.1.5. Biomass higher heating value prediction: machine learning insights into ultimate, proximate, and structural analysis datasets

In this study machine learning (ML) models have been employed to predict the higher heating value (HHV) of biomass by utilizing input variables derived from ultimate, proximate, and structural analyses. In total, 180 models were developed, with 124 utilizing ultimate analysis data, 28 based on proximate analysis, and 28 relying on structural analysis. Various ML techniques, including polynomial models (SOP), support vector machines (SVM), random forest regression (RFR), and artificial neural networks (ANN), were employed for analysis. The study found that ANN models, when “fed” with FC and VM data, provided considerable accuracy in prediction results, with the best results obtained with 2-12-1 architecture ($R^2 = 0.96$). In addition, a separate model configuration that processed inputs on biomass constituents such as cellulose, lignin, and hemicellulose showed remarkable agreement with empirical data. Additional findings revealed that the models created using SOP ($R^2 = 0.95$), SVM ($R^2 = 0.95$), and RFR ($R^2 = 0.90$) demonstrated minimal discrepancies when predicting HHV. This study provides significant insights into the investigation of biomass analysis techniques employing ML tools, paving the way for future research aimed at constructing a robust tool for HHV prediction. Subsequent models may explore integrating inputs from diverse analysis methods and leveraging advanced machine learning techniques to enhance accuracy further.

Keywords: energy properties; mathematical modelling; ultimate analysis; proximate analysis; structural analysis.

3.2. Unified discussion

3.2.1. Basic characteristics of collected data after statistical analysis

The research began with the crucial step of data collection, followed by an important phase of data quality control to ensure the reliability and quality of the database - a process that is critical in data analysis due to the potential challenges in data quality (Hellerstein and Berkeley, 2008). The performance of ML models is highly dependent on data quality, which is ensured by the process of error labeling, imbalance, scoring, and data homogenisation, transformation, and cleaning (Collins et al., 2018).). The next step was to categorise the collected data into meaningful units, in particular agricultural and woody biomass, in order to identify possible differences. This categorisation facilitated the use of analysis of variance (ANOVA) and, if necessary, the Tukey post hoc HSD test to further explore the data. To determine the mean values of the data by category, the mean was used as a measure of the central tendency. A correlation analysis was also carried out to confirm the relationship between the variables analysed.

The articles present models processed on the basis of different types of input data: ultimate analysis is discussed in Articles 1, 2, 4 and 5, structural analysis in articles 3 and 5, and proximate analysis in Article 5.

In Article 1 (3.1.1.), the analysis focussed on 16 samples of Miscanthus. The data from the ultimate biomass analysis were statistically processed and the following ranges of mean values were determined: Nitrogen (N) (0.18-0.31%), Carbon (C) (50.90-51.76%), Sulphur (S) (0.09-0.21%), Hydrogen (H) (5.31-5.89%), Oxygen (O) (41.99-42.91%) and Higher Heating Value (HHV) (17.83 – 18.51 MJ kg^{-1}).

In Article 2 (3.1.2.), the research aimed to develop models using corn, soybean, and sunflower biomass data. This study determined mean values for elements and calorific values, such as C (49.99%), H (5.44%), N (1.44%), S (0.12%), O (40.08%) and HHV (19.94 MJ kg^{-1}). A correlation analysis was performed to assess the relationship and strength of association between the ultimate analysis variables and HHV. A significant positive correlation was found, especially between C, H, and HHV, which was also found in the literature (Noushabadi et al., 2021). Furthermore, research by Yang et al. (2023) confirmed these results and emphasised that the carbon content in biomass fuels significantly affects HHV.

In Article 4 (3.1.4.), a statistical analysis was presented using ANOVA that showed differences in the ultimate analysis of the biomass composition analysis of Miscanthus. The mean values were obtained for N (0.72%), C (51.18%), S (0.06%), H (5.84%), O (42.19%) and HHV (18.32 MJ kg^{-1}).

Conversely, in Article 5, the data were divided into agricultural and wood biomass and the mean percentage of C (47.61 and 48.13%), H (5.52 and 5.37%), N (1.27 and 0.70%), S (0.26 and 0.33%), O (39.87 and 42.75%) and HHV (18.73 and 18.89 MJ kg^{-1}) were analysed. The only statistically significant difference between the analysed groups was found in the nitrogen content, which was significantly higher in the agricultural biomass category.

In Article 3 (3.1.3.), a comprehensive correlation analysis was performed to examine the relationship between the structural components of biomass, namely hemicellulose, lignin, and cellulose, and their HHV. The analysis revealed remarkably high positive correlation coefficients with HHV: 0.74 for hemicellulose, 0.88 for lignin and 0.89 for cellulose. These results indicate that each of these structural elements of biomass - hemicellulose, lignin, and cellulose - significantly and positively influences the energy value of biomass, showing a strong correlation between the composition of biomass and its potential as an energy source. Esteves et al., (2023) state the influence of individual components of the structural composition of biomass on the energy value and conclude that the HHV is mainly influenced by the lignin content of the raw material. Gani et al., (2024) state that the energy value of biomass can be determined by analysing lignin and cellulose. The analysis showed that biomass with a higher cellulose content impairs flammability, while a higher lignin content increases the calorific value. Soomro et al., (2021) state that the HHV of biomass rich in cellulose is on average 17.28 – 18.58 MJ kg^{-1}, while the HHV of biomass rich in lignin is 19.07 – 22.50 MJ kg^{-1}.

Article number 5 (3.1.5.) of the study series focuses not only on the ultimate biomass analysis but also delves into both proximate and structural analyses. The proximate analysis, which assesses the basic characteristics of the biomass, did not reveal any statistically significant differences between agricultural and wood biomass. The mean values determined for the entire dataset are as follows: fixed carbon (FC) 14.66%, volatile matter (HT) 74.93% and ash content 18.21%. The structural analysis conducted in Article 5 focussed on determining the percentage of cellulose, hemicellulose, and lignin in the biomass samples. Mean values were determined for cellulose (46.67%), lignin (15.82%) and hemicellulose (23.45%).

3.2.2. Performance of the models in estimating the HHV based on ultimate analysis input data

In Article 1 (3.1.1.), data collection included 16 different samples of Miscanthus and resulted in a total of 192 data points. The focus was on quantifying the variables of the ultimate analysis, an essential step in the creation of an ANN model. This model was specifically developed to estimate the HHV of the biomass. To evaluate the efficiency of this ANN model, it was compared with a range of empirical equations commonly used to estimate the HHV. For this comparison, 10 different empirical models were selected from scientific databases reflecting the breadth of research in this field, including notable works by Sheng and Azevedo (2005), Nhuchhen and Afzal (2017) and Callejón-Ferre et al. (2011). In Article 1 (3.1.1.) the performance of both the developed ANN model and the empirical models was rigorously evaluated using the statistical method "Goodness of fit". This analysis involved the calculation of various statistical parameters to determine how well each model estimated the HHV of Miscanthus. Interestingly, the results showed that the models generally did not perform satisfactorily in estimating the HHV. The best-fitting model among them showed the following statistical indicators: X^2 = 0.17, RMSE = 0.01 and R^2 = 0.47. Despite the relatively low modelling error of this model, it did not provide satisfactory results, as evidenced by the low overlap between the actual and predicted data. In contrast, the ANN model developed in this study showed a greater ability to model HHV, as indicated by a higher R^2 value of 0.77. This result not only emphasises the potential of ANNs in accurately predicting biomass characteristics but also highlights the limitations of conventional empirical models in dealing with the complex nature of biomass fuels such as Miscanthus. This was also confirmed by the research conducted by Guleç et al., (2022), in which they obtained the highest accuracy model by ANN with a high R^2 for learning (0.96) and testing (0.92). The authors conclude that the ANN model is suitable and shows high performance in HHV modelling, which can be further improved by adjusting the activation functions, algorithms, number of neurons and randomisation. The superior performance of the ANN model in this context suggests that it is suitable for a more accurate estimation of HHV, thus providing a more reliable tool for biomass energy researchers and practitioners.

In the published paper number 2 (3.1.2.), a slightly different approach was taken, focusing on the comparison of RFR and SVM models in terms of their ability to model the HHV of biomass. In this paper, 51 data points from the ultimate analysis were used. The SVM model showed superior performance with a significant degree of overlap between the collected and estimated data, as indicated by a high R^2 of 0.93. In addition, the SVM model showed a relatively low modelling error, with a X^2 value of 0.82 and a RMSE of 0.90. In

contrast, the RFR model showed higher values for X^2 (5.99) and RMSE (2.43), associated with a lower R^2 value of 0.79. This comparison suggests that the SVM model was more effective in modelling HHV in this study, especially with smaller data sets, which is also confirmed by Ozyuğuran et al. (2018).

Furthermore, the effectiveness of an ANN model in estimating the HHV of biomass was investigated in Article 4 (3.1.4.). This study found that the ANN model had the highest performance in terms of overlap between real and modelled data, as evidenced by an impressive R^2 value of 0.96. In addition, the model had a remarkably low modelling error, with an X^2 value of 0.01 and an RMSE value of 0.03. These results led to the conclusion that the ANN model was best suited for modelling the energy value of biomass, outperforming the capabilities of both the RFR and SVM models in this application. The high accuracy and low error rates of the ANN model underline its potential as a powerful tool for predicting the energy value of biomass, especially under varying conditions of growing. This insight into the calorific value of Miscanthus through the ANN model opens new avenues for optimising biomass use for energy production and provides a more accurate and reliable method for estimating the potential energy yield of biomass fuels.

The research presented in article number 5 (3.1.5.) aims to test and determine the most appropriate model ANN, RFR, SVM and HOP for analysing biomass. This study focussed on using the input variables from the ultimate analysis to evaluate the effectiveness of each model. The ANN model proved to be the front-runner in terms of modelling ability with an X^2 value of 1.02, an RMSE value of 1.01 and a high R^2 value of 0.90 using all variables from the ultimate analysis. Higher order polynomials also showed their usefulness in nonlinear modelling with an X^2 of 1.77, an RMSE of 1.33 and an R^2 of 0.82, indicating their effectiveness in handling complex data structures. In comparison, the SVM and RFR models showed slightly lower modelling performance. In a parallel study conducted by Noushabadi et al. (2021), the focus was on estimating the HHV of different types of biomass fuels based on elemental analysis input variables. In this study, various machine learning techniques and hybrid equations were used as basic tools. The evaluation of the ANN model they developed yielded remarkable statistical results: an R^2 of 0.92, an RMSE of 1.08 and a standard deviation (SD) of 3.76. Although their model showed better performance in terms of R^2, the model constructed in Article 5 (3.1.5.) had a lower error in estimating the HHV biomass when modelling errors and residual analysis were considered. Olatunji et al. (2019) also contributed to the research on HHV biomass estimation through a multilayer perceptron network model based on the inputs of elemental analysis. They developed hybrid ANN models and used the RMSE as a measure of model error. The model error (RMSE) for their ANN model with elastic backpropagation was 3.59, while the ANN model with the Levenberg-

Marquardt algorithm had an RMSE of 3.10. In comprehensive studies by Xing et al. (2019), the R^2 (0.90) and RMSE (3.55) values for ANN models were determined based on elemental analysis input data. Comparing the ML models developed in these various studies with the empirical equations available for estimating the HHV, ANN models prove to be the most suitable for modelling, especially about the elemental analysis input dataset. This suitability is reflected in the lowest modelling error and the high degree of model agreement with the real data.

These results from Article number 5 and the corroborating studies emphasise the increasing reliability and precision of ANN models in biomass energy research and show that they are capable of handling complex data sets with higher accuracy than traditional methods. These advances in modelling techniques are crucial for improving the predictability and efficiency of biomass energy use and opening new avenues for sustainable energy research and development.

3.2.3. Performance of the models in estimating the HHV based on proximate analysis input data

In Article 5 (3.1.5.), a comprehensive analysis was conducted in which 28 nonlinear models were created based on the input data from proximate analyses. These models were evenly divided into four types: ANN, SVM, RFR and HOP. Using the combinatorial method, which ensures that members are not repeated, these models were systematically developed to cover all possible combinations of the three variables from the proximate analysis (FC, HT, and Ash). The performance of these models was rigorously evaluated using a range of statistical metrics, including X^2, RMSE, MBE, MPE, SSE, AARD, R^2, skewness, kurtosis, SD, and variance. This approach provides a holistic overview of the effectiveness and accuracy of each model. Seven different ANN models were developed as part of the study, focussing on different combinations of the input variables. Three models included either one or two variables, and one model used all three variables. Analysis of the statistical fit of these models revealed that the model with two input variables (FC and HT) performed best, achieving low modelling error (X^2 = 0.41; RMSE = 0.64) and high accuracy (R^2 = 0.96). Conversely, the model with only one input variable (HT) showed the worst performance in modelling HHV biomass (X^2 = 9.04; RMSE = 3.00; R^2 = 0.19). Further research by Veza et al. (2022) focused on the accurate estimation of HHV biomass using proximate analysis. Several ANN models were developed, each based on different algorithms. The Levenberg–Marquardt (LM) ANN model proved to be the most effective, as shown by a high R^2 (0.94)

and a low RMSE (0.99). Geleç et al (2022) also developed a regression ANN model for modelling HHV biomass using different activation functions such as Tansig, Logsig and Poslin. This model achieved satisfactory accuracy in both the test and training series (R^2 = 0.82 - 0.87). As for the SVM models in the study, the best fitting SVM model was developed using two input variables from the proximate analysis dataset (FC and PE). This model showed a low modelling error (X^2 = 0.62 and RMSE = 0.79) and a high degree of accuracy (R^2 = 0.95). The least useful model for modelling HHV biomass was created with a single input variable (VM) and showed a high modelling error (X^2 = 19.36 and RMSE = 4.39) and a low degree of overlap (R^2 = 0.02). However, the model that included all variables in the dataset had the lowest modelling error, albeit with a slightly lower degree of overlap (R^2 = 0.94). Keybondorian et al. (2017) undertook extensive research to estimate HHV biomass using proximate analysis, specifically the proportions of FC, VM and ash. Using the hybrid least SVM algorithm, they developed a model that proved to be robust and reliable in estimating the energy value of different biomass types, as evidenced by its RMSE (1.03) and R^2 (0.93). Xing et al. (2019) reported a slightly higher modelling error in their study (RMSE = 4.67 and R^2 = 0.85), highlighting the need for further advances in SVM models to assess the energy parameters of biomass. This collective investigation of different studies highlights the importance and potential of using different modelling approaches, in particular ANN and SVM, to accurately estimate the HHV of biomass based on proximate analyses. The results provide valuable insights into the effectiveness of different modelling types and configurations and pave the way for improved and more reliable methods for biomass energy assessment. The RFR performed impressively, especially when modelled with a combination of FC and ash and with the complete set of variables from the dataset. Despite a relatively high modelling error (X^2 = 2.50), the RFR model achieved a high level of accuracy, as indicated by a high R^2 value (0.90). However, the performance of the model decreased when only HT was used as an input variable, resulting in a higher modelling error (X^2 = 10.47; RMSE = 3.23) and a significantly lower R^2 (0.21). Afolabi et al. (2022) followed a comprehensive approach by integrating the input variables of elemental and physicochemical analysis. Their study included several models, including the RFR model. The RFR model was characterised by its low error rate (RMSE = 1.37), making it the most accurate model for estimating HHV biomass, mainly due to its minimal error values. Dubey and Guruviah, (2023) also compared the differences of RFR and ANN models in terms of HHV biomass modelling error by the input variables of the proximate analysis and determined the RMSE modelling error for ANN (0.186) and RFR model (0.101) but found that better model performance can be achieved by a metaheuristic optimisation approach.

The HOP models were also found to be effective for HHV estimation, especially for the proximate analysis data. The best-fitting HOP model using FC and ash as input variables showed a low modelling error (X^2 = 0.52; RMSE = 0.72) and a high R^2 (0.95). Models using VM and ash also showed high accuracy (R^2 = 0.94), and a model including all input variables achieved an R^2 of 0.93. However, the least effective HOP model using only VM as input showed significant limitations in estimating HHV biomass (X^2 = 10.09; RMSE = 3.17; R^2 = 0.08). These results highlight the importance of selecting appropriate modelling techniques to obtain accurate and reliable estimates of HHV biomass. While RFR models are advantageous for comprehensive data sets, HOP models are characterised by specific combinations of physicochemical variables. The integration of different data types, as shown by Afolabi et al. (2022), is crucial for improving the precision of biomass energy value estimates and provides a robust framework for accurate assessment of biomass energy parameters. The research in Article 5 suggests that ANN models with proximate analysis data have the highest modelling accuracy in terms of the ability to estimate HHV and a low modelling error compared to other models. It is also important to emphasise that the proximate analysis data set proved to be the most suitable for modelling the HHV for all model types.

3.2.4. Performance of the models in estimating the HHV based on structural analysis input data

Article number 3 (3.1.3.) of the study series focused on the investigation of nonlinear machine learning models (ANN, RFR, SVM and HOP) about input data sets from structural analysis, in particular the percentage of cellulose, hemicellulose, and lignin. Among these models, ANN proved to be the most effective, as indicated by a low modelling error (X^2 = 0.25, RMSE = 0.50) and a high degree of accuracy (R^2 = 0.90). However, the other models, namely RFR (X^2 = 0.29, RMSE = 0.54, R^2 = 0.89), SVM (X^2 = 0.35, RMSE = 0.59, R^2 = 0.86) and HOP (X^2 = 0.32, RMSE = 0.56, R^2 = 0.87), showed slightly weaker modelling performance.

A comprehensive approach was followed in Article 5 (3.1.5.), where 28 nonlinear models were created based on structural analysis input data. This dataset included seven models each for ANN, SVM, RFR and HOP. Using a combinatory to calculate the total number of models without repetition, the study analysed performance across all possible combinations of the three structural analysis variables. Statistical analysis in this article revealed that the best-fitting model for estimating HHV was constructed using all input variables, namely

cellulose, hemicellulose, and lignin. This model had the lowest modelling error (X^2 = 0.26; RMSE = 0.51), combined with a high degree of data overlap (R^2 = 0.91), highlighting its high reliability and precision. In the SVM analysis, the model that included all input variables proved to be the most efficient in terms of statistical parameters, characterised by the lowest X^2 (0.79) and RMSE (0.89), indicating high accuracy and efficiency.

Park et al., (2023) use models to predict the HHV of lignocellulosic biomass based on structural analysis data. When considering the entire data set, the authors report R^2 = 0.54 – 0.58 and RMSE = 1.38 to 1.54 and state that the model does not have good universal performance. On the other hand, after using two categories of data (woody and herbaceous biomass), results with a lower modelling error were obtained, with the best R^2 being 0.78 – 0.84 for woody biomass and 0.83 – 0.87 for herbaceous biomass. Akdeniz et al., (2018) develop an algorithm for estimating the HHV of various lignocellulosic biomasses, considering the lignin and extract content as input data. The models are based on 11 different lignocellulosic materials whose data were mainly obtained experimentally. The developed hybrid algorithm showed a satisfactory performance in terms of data overlap with R^2 = 0.92, while the modelling errors SSE (0.30), MSE (0.30) and RMSE (0.55) showed a low error level.

The models developed in Articles 3 (3.1.3.) and 5 (3.1.5.) have a higher generalisation ability and universality, as evidenced by the high level of the specific indicator of representativeness of the regression (R^2) and a lower modelling error. This is an indication of the ability of the ML model to estimate the HHV based on the entire dataset, regardless of the categories of forest or agricultural biomass.

The alignment of the analysis for the prediction of HHV biomass based on the input variables of the structural analysis is considered complex due to the high variability between the individual samples analysed, which is confirmed by Maksimuk et al. (2021), and they recommend the inclusion of additional variables and ultimate analysis to increase accuracy or reduce the error in modelling. From all this it can be concluded that ML models have a high generalisation ability and the possibility of a relatively good estimation of the HHV. However, to obtain a better model, further research should focus on adding input variables of ultimate and proximal analysis and determine which input parameters can provide an optimal model, i.e. the one with a lower modelling error.

3.2.5. Sensitivity analysis based on ANN model

The practical use of sensitivity analysis is illustrated in paper number 1, where an ANN model was developed, and different prediction models were tested. Here, the Yoon model was used for sensitivity analysis to determine the relative importance of the input variables in predicting HHV. This analysis covered a range from -1 to 1 and examined the effects of variables such as C, H, N, S and O. The results showed that for the optimal sample (highest HHV value) there was an increase in S (68.89%) and C (3.13%) and a decrease in N (-20.64%), H (-3.07%) and O (-4.27%). Taki and Rohani, (2022) state that the variables C, O and S have no effect on the change in HHV, while the decrease in H and N affects the increase in energy value. The different results of the sensitivity analysis in this case are due to the different data set used in the HHV modelling, in their research Taki and Rohani, (2022) besides the ultimate analysis used both ash and H_2O percentage as input variables.

Similarly, Article 5 (3.1.5.) examined the effects of input variables on HHV output using an ANN model based on the ultimate analysis inputs. This study emphasised that HHV output is significantly influenced by the variables C, H, N, S and O. The sensitivity analysis showed that increasing the C, N and S content while reducing H and O significantly improved HHV output. The authors Kujawska et al. (2023), in research aimed at finding an improved HHV biomass prediction method using ANN models, come to slightly different conclusions, but also determine the most important variables as input data, i.e. those that would have the greatest (relative) influence on the output value. The authors indicate that the variables that have a significant influence on the HHV are C, H and S. The differences mentioned above are due to a greater variability of the individual components depending on the data set. In addition, the study investigated an ANN model that used proximate analysis data such as fixed carbon (FC) and volatile matter (VM) and showed high predictive accuracy for HHV. In the case of structural analysis, the ANN model showed that changes in cellulose (Cel) and lignin (Lig) together with a decrease in hemicellulose (Hem) significantly affected the HHV value. García Nieto et al. (2019) extended this research paradigm by focusing on predictive modelling of HHV biomass for energy process applications using different models of simulation units (SU). Their model included variables such as HT, FC and ratios of O/C and H/C as well as the temperature and duration of process reactivity. A comprehensive sensitivity analysis (0-100%) was performed on their hybrid model. FC and VM were found to have the greatest relative importance, confirming their crucial role in predictive modelling of HHV biomass. These studies emphasise the importance of sensitivity analysis for understanding model dynamics and improving prediction accuracy, especially in the context of biomass research. This scientific approach not only helps in model optimisation, but also

contributes significantly to the progress of predictive modelling in various fields, especially in energy-related applications.

3.2.6. Research limitations and future approach

Recent research in the field of HHV biomass modelling using different mathematical models (HOP, ANN, RFR, SVM) for different data sets (ultimate, proximate, and structural analysis) requires a critical examination of both their limitations and future research opportunities. Considering that ML algorithms and techniques are still emerging and constantly evolving, there is a continuing trend towards developing new learning methods and reducing computational time. ML also offers new opportunities in the fields of energy and agriculture (Pugliese et al., 2021). Modelling and prediction are a standard method for finding relationships between individual components in a large amount of data (Huang et al., 2020). Lai et al., (2020) states that in the future, the development of artificial intelligence models and adapted hybrid models will be widely used in the modelling and design of energy systems in the field of renewable energy sources.

Table 1 shows a summary of the performance of the significant models in terms of R^2 in relation to the type of model, the data set through 5 published scientific papers within the research.

Table 1. Summary of the performance (R^2) of the most significant models in relation to the type of data set processed by 5 published scientific papers.

Input data ↓	Model ↓	Article 1 (3.1.1.)	Article 2 (3.1.2.)	Article 3 (3.1.3.)	Article 4 (3.1.4.)	Article 5 (3.1.5.)
Ultimate analysis input data	ANN	R^2=0.77	-	-	R^2=0.96	R^2=0.90
	SVM	-	R^2=0.93	-	-	R^2=0.81
	RFR	-	R^2=0.79	-	-	R^2=0.76
	HOP	-	-	-	-	R^2=0.82
Proximate analysis input data	ANN	-	-	-	-	R^2=0.96
	SVM	-	-	-	-	R^2=0.95
	RFR	-	-	-	-	R^2=0.90
	HOP	-	-	-	-	R^2=0.95
Structural analysis input data	ANN	-	-	R^2=0.90	-	R^2=0.91
	SVM	-	-	R^2=0.86	-	R^2=0.74
	RFR	-	-	R^2=0.89	-	R^2=0.82
	HOP	-	-	R^2=0.87	-	R^2=0.79

The models created in the research based on the input data of the proximate analysis achieved the highest representativeness of the regression (Table 1). compared to the models based on the input data of the ultimate and structural analysis. Of all the models created in the research, the ANN model based on the input data from the proximity analysis showed the highest data agreement as determined by the coefficient of determination (R^2=0.96).

The models showed promising results, however, their applicability seems to be limited to certain biomass types and environmental conditions, indicating the need for broader validation. The specificity and generalisability of these models are crucial, especially for adaptation to different biomass types and ecological scenarios. An important limitation is the dependence of model accuracy on the quality and extent of the input data. In situations where certain biomass attributes are inadequately represented, model accuracy can suffer significantly, emphasising the importance of more comprehensive and representative data sets. In contrast to empirical equations tailored to specific biomass types, nonlinear ML models offer broader modelling benefits (Dashti et al., 2019).

The complex interactions within biomass components pose significant challenges and affect the nuanced accuracy of the models. Future research should aim to expand the variables in the models to cover a wider range of biomass properties and thus reduce modelling errors (Brandić et al., 2023). When increasing R^2, there is a possibility of potential overfitting and the stated requirement of using an adjusted R^2 in order to take into account the number of variables and preserve the robustness of the model. Such an extension could improve the accuracy of the models and extend their applicability to different types of biomass, with better results being achieved by novel and hybrid models (Nhuchhen and Afzal, 2017). Testing the models under different ecological and geographical conditions is crucial for validating their effectiveness in different environments. This approach not only serves to refine and adapt current models but also promotes the development of more versatile tools for biomass modelling and thus makes an important contribution to energy modelling, especially to the sustainable management and use of biomass. The scientific contribution of the proposed research is significant, especially through the development of new, more reliable nonlinear mathematical models as universal tools for HHV prediction and modelling in agricultural and wood biomass. These models represent a significant advance in biomass research as they provide a deeper and more accurate understanding of the properties and effects of biomass. In addition, identifying the most efficient nonlinear model form that minimises error while reducing input variables, regardless of the type of biomass, provides a way to streamline HHV prediction and modelling processes. By reducing the amount of time and variables required, this research proposes a more efficient, cost-effective

approach to biomass modelling that has practical implications for simplifying and accelerating biomass analysis in various applications.

After developing nonlinear mathematical models for modelling HHV biomass and analysing the results, it is crucial to understand their fundamental differences, i.e. their advantages and disadvantages in the context of modelling and prediction. In this research, ANN models proved to be the most efficient, whose main advantage lies in their ability to learn and understand deep interactions. This makes them ideal for complex tasks that require many numerical computations. However, inadequate tuning of the architecture can lead to overfitting and model cluttering, which increases modelling error (Hoang, 2023; Hosseinzadeh et al., 2020). On the other hand, RFR uses multiple trees for more robust prediction and shows advantages when working with unbalanced and missing data, allowing for better interpretability compared to more complex models. The main disadvantage of RFR is its lower efficiency on high-dimensional data (Xue et al., 2021; Fouedjio, 2020). SVM models that use a hyperplane for class separation are extremely efficient for high-dimensional data where nonlinearity is present. Despite its high efficiency in solving nonlinear problems, SVM requires precise parameter tuning and optimisation, including kernel selection and learning hyperparameter tuning (Wang et al., 2019; Shao et al., 2020). As regression models, HOP models have a high ability to estimate and solve fewer complex tasks with a lower error level. However, the increasing amount and multidimensionality of data can lead to overfitting and poor generalisation on the test set (Morala et al., 2021). Ultimately, the choice of model depends on the specific requirements of the problem, the availability of data sets and the interpretability of the model.

CONCLUSIONS

The results of this doctoral thesis provide scientific proof of the feasibility of modelling the higher heating value (HHV) of agricultural and wood biomass. During the research, various models were developed using artificial neural networks (ANN), random forest regression (RFR), high order polynomial (HOP) and support vector machine (SVM), all based on input datasets from ultimate, proximate, and structural analyses. The aim of the work was to develop new nonlinear mathematical models of different types for modelling the HHV of biomass based on sets of input variables from laboratory analyses. Furthermore, these newly developed models were to be compared and the smallest error in modelling the output value (HHV) concerning different sets of input variables was to be determined. On this basis, two hypotheses were initially formulated:

1. Models based on the proximate analysis input dataset (FC, VM and ash) achieved the smallest modelling error compared to the ultimate analysis (C, H, N, S and O) and structural analysis (cellulose, hemicellulose, and lignin) datasets. This was demonstrated for all nonlinear models developed. During the research, the collected data were categorised into agricultural and wood biomass. Among the nonlinear models based on the input variables of the proximate analysis, the ANN model achieved the lowest modelling error (χ^2=0.41, RMSE =0.64) and the highest data agreement (R^2 = 0.96), while the SVM (χ^2 = 0.61, RMSE =0.78, R^2 = 0.94), RFR (χ^2 = 6.56, RMSE = 2.56, R^2 = 0.90) and HOP (χ^2 = 0.52, RMSE = 0.72, R^2 = 0.95) showed significant overlap, but with slightly higher error values. The analysis also assessed models based on ultimate and structural analyses, with the most appropriate models showing a slightly higher error in modelling HHV and a lower level of data overlap (R^2 = 0.74 – 0.91). Based on the results of this study, hypothesis 1 is accepted, which states that all model types (ANN, RFR, SVM and HOP) have the lowest error when created with input variables from proximate analysis.

2. ANN models have a lower error in modelling the higher heating value (HHV) of biomass compared to HOP models, RFR models and SVM models, regardless of the set of input variables. Throughout the study, the ANN models showed the highest level of agreement between the actual data and the model predictions (R^2 = 0.90 – 0.96) and the lowest modelling error in terms of χ^2 (0.26 – 1.02) and RMSE (Root Mean Square Error, 0.51 – 1.01). On the other hand, the HOP models achieved a slightly lower degree of data overlap (R^2 = 0.82 – 0.95) and a higher level of error in terms of χ^2 (0.52 – 1.77) and RMSE (0.72 – 1.33). SVM models generally achieve satisfactory performance (R^2 = 0.74 – 0.95; χ^2 = 0.62 – 2.68 and RMSE = 0.79 – 1.63). The RFR models show a slightly higher error in relation to the modelling (R^2 = 0.76 – 0.90; χ^2 = 0.54 – 6.56 and RMSE = 0.73 – 2.56). On this basis,

the hypothesis was tested by comparing the newly developed nonlinear models using statistical parameters and it was proved that ANN models have the lowest modelling error compared to the other developed models.

Additional conclusions were drawn based on the scope of this dissertation:

- No statistically significant differences were found between the analysed collected data of the forestry and agricultural biomass categories in the ultimate and proximate analysis datasets, indicating the homogeneity of the data within both categories.
- Between the variables of the structural analysis, there is a difference in the proportion of lignin, hemicellulose and HHV within the biomass categories, which are statistically significantly higher in the forest biomass category, while there is no statistically significant difference in the proportion of cellulose.
- Using the method of combining without repeating the members and comparing the models with each other in terms of modelling error, the most appropriate models in terms of the number of input variables in the models were determined for each data set and each type of model developed:
    - Ultimate analysis: The most appropriate model is developed with all input variables (C, H, N, S and O)
    - Proximate analysis: The best-fit model was developed with 2 input variables (FC and VM)
    - Structural analysis: The best-fitting model was developed with all input variables (cellulose, hemicellulose, and lignin).

- After performing a sensitivity analysis using Yoon's relative importance model, the influence of the input variables on the HHV was determined, so that in the ultimate analysis the optimal sample was found with an increase in C, N and S and a decrease in H and O. In the proximate analysis data set, the highest value of HHV was found in the samples with reduced FC and HT values, while in the structural analysis data set, an optimal pattern was determined with an increase in cellulose, hemicellulose, and lignin.

LITERATURE

Abbasi Nozari H., Dehghan Banadaki H., Mokhtare M., Hekmati Vahed S. (2012). Intelligent non-linear modelling of an industrial winding process using recurrent local linear neuro-fuzzy networks. J Zhejiang Univ Sci C 13 (6): 403–412. doi:10.1631/jzus.C11a0278

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. (2021). Artificial neural networks based optimization techniques: A review. Electron 10 (21). doi:10.3390/electronics10212689

Acar S., Ayanoglu A. (2012). Determination of higher heating values (HHVs) of biomass fuels. Energy Educ Sci Technol Part A Energy Sci Res 28 (2): 749–758

Afolabi I.C., Epelle E.I., Gunes B., Güleç F., Okolie J.A. (2022). Data-Driven Machine Learning Approach for Predicting the Higher Heating Value of Different Biomass Classes. Clean Technol 4 (4): 1227–1241. doi:10.3390/cleantechnol4040075

Akdeniz F., Biçil M., Karadede Y., Özbek F.E., Özdemir G. (2018). Application of real valued genetic algorithm on prediction of higher heating values of various lignocellulosic materials using lignin and extractive contents. Energy 160: 1047–1054. doi:10.1016/j.energy.2018.07.053

Avcıoğlu A.O., Dayıoğlu M.A., Türker U. (2019). Assessment of the energy potential of agricultural biomass residues in Turkey. Renew Energy 138: 610–619. doi:10.1016/j.renene.2019.01.053

Baghban A., Kahani M., Nazari M.A., Ahmadi M.H., Yan W.M. (2019). Sensitivity analysis and application of machine learning methods to predict the heat transfer performance of CNT/water nanofluid flows through coils. Int J Heat Mass Transf 128: 825–835. doi:10.1016/j.ijheatmasstransfer.2018.09.041

Baxter X.C., Darvell L.I., Jones J.M., Barraclough T., Yates N.E., Shield I. (2014). Miscanthus combustion properties and variations with Miscanthus agronomy. Fuel 117 (PART A): 851–869. doi:10.1016/j.fuel.2013.09.003

Beaucamp A., Muddasar M., Amiinu I.S., Moraes Leite M., Culebras M., Latha K., Gutiérrez M.C., Rodriguez-Padron D., del Monte F., Kennedy T., Ryan K.M., Luque R., Titirici M.M., Collins M.N. (2022). Lignin for energy applications - state of the art, life cycle, technoeconomic analysis and future trends. Green Chem 24 (21): 8193–8226. doi:10.1039/d2gc02724k

Bermejo J.F., Fernández J.F.G., Polo F.O., Márquez A.C. (2019). A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. Appl Sci 9 (9). doi:10.3390/app9091844

Biau G., Scornet E. (2016). A random forest guided tour. Test 25 (2): 197–227. doi:10.1007/s11749-016-0481-7

Bilandzija N., Voca N., Jelcic B., Jurisic V., Matin A., Grubor M., Kricka T. (2018). Evaluation of Croatian agricultural solid biomass energy potential. Renew Sustain Energy Rev 93 (May): 225–230. doi:10.1016/j.rser.2018.05.040

Boumanchar I., Charafeddine K., Chhiti Y., M'hamdi Alaoui F.E., Sahibed-dine A., Bentiss F., Jama C., Bensitel M. (2019). Biomass higher heating value prediction from ultimate analysis using multiple regression and genetic programming. Biomass Convers Biorefinery 9 (3): 499–509. doi:10.1007/s13399-019-00386-5

Brandić I., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2023). Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass. Mathematics 11 (9). doi:10.3390/math11092098

Bun M., Steinke T. (2015). Weighted polynomial approximations: Limits for learning and pseudorandomness. Leibniz Int Proc Informatics, LIPIcs 40: 625–644. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.625

Burg V., Bowman G., Erni M., Lemm R., Thees O. (2018). Analyzing the potential of domestic biomass resources for the energy transition in Switzerland. Biomass and Bioenergy 111 (February): 60–69. doi:10.1016/j.biombioe.2018.02.007

Bychkov A.L., Denkin A.I., Tikhova V.D., Lomovsky O.I. (2017). Prediction of higher heating values of plant biomass from ultimate analysis data. J Therm Anal Calorim 130 (3): 1399–1405. doi:10.1007/s10973-017-6350-0

Callejón-Ferre A.J., Carreño-Sánchez J., Suárez-Medina F.J., Pérez-Alonso J., Velázquez-Martí B. (2014). Prediction models for higher heating value based on the structural analysis of the biomass of plant remains from the greenhouses of Almería (Spain). Fuel 116: 377–387. doi:10.1016/j.fuel.2013.08.023

Callejón-Ferre A.J., Velázquez-Martí B., López-Martínez J.A., Manzano-Agugliaro F. (2011). Greenhouse crop residues: Energy potential and models for the prediction of their higher heating value. Renew Sustain Energy Rev 15 (2): 948–955. doi:10.1016/j.rser.2010.11.012

Cao W., Wang X., Ming Z., Gao J. (2018). A review on neural networks with random weights. Neurocomputing 275: 278–287. doi:10.1016/j.neucom.2017.08.040

Casas I. (2019). Networks, Neural, Second Edi. Edition, International Encyclopedia of Human Geography, Second Edition. Elsevier, 381–385 pp. doi:10.1016/B978-0-08-102295-5.10410-X

Chen Jiefeng, Ding L., Wang P., Zhang W., Li J., Mohamed B.A., Chen Jie, Leng S., Liu T., Leng L., Zhou W. (2022). The Estimation of the Higher Heating Value of Biochar by Data-Driven Modeling. J Renew Mater 10 (6): 1555–1574.

doi:10.32604/jrm.2022.018625

Cinar A.C. (2020). Training Feed-Forward Multi-Layer Perceptron Artificial Neural Networks with a Tree-Seed Algorithm. Arab J Sci Eng 45 (12): 10915–10938. doi:10.1007/s13369-020-04872-1

Collins E., Rozanov N., Zhang B. (2018). Evolutionary data measures: Understanding the difficulty of text classification tasks. CoNLL 2018 - 22nd Conf Comput Nat Lang Learn Proc 380–391. doi:10.18653/v1/k18-1037

Crawley M.J. (2012). Applied Linear Regression Models:Non-linear Regression. Book 661–684

Crooke P.S., Hotchkiss J.R., Marini J.J. (2002). Linear and nonlinear mathematical models for noninvasive ventilation. Math Comput Model 35 (11–12): 1297–1313. doi:10.1016/S0895-7177(02)00086-9

Dashti A., Noushabadi A.S., Raji M., Razmi A., Ceylan S., Mohammadi A.H. (2019). agri. Fuel 257 (March): 115931. doi:10.1016/j.fuel.2019.115931

Demirbas A. (2017). Higher heating values of lignin types from wood and non-wood lignocellulosic biomasses. Energy Sources, Part A Recover Util Environ Eff 39 (6): 592–598. doi:10.1080/15567036.2016.1248798

Díez D., Urueña A., Piñero R., Barrio A., Tamminen T. (2020). and Lignin Content in Di ff erent Types of Biomasses by Thermogravimetric Analysis and Pseudocomponent Kinetic Model. Processes 8 (1048): 1–21

Dikusar N. (2018). Shape Approximation Based on Higher-Degree Polynomials. EPJ Web Conf 173: 1–8. doi:10.1051/epjconf/201817301003

DiMaio F., Chiu W. (2016). Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps, 1st Edition, Methods in Enzymology. Elsevier Inc., 255–276 pp. doi:10.1016/bs.mie.2016.06.003

Donti P.L., Kolter J.Z. (2021). Machine Learning for Sustainable Energy Systems. Annu Rev Environ Resour 46: 719–747. doi:10.1146/annurev-environ-020220-061831

Dorez G., Ferry L., Sonnier R., Taguet A., Lopez-Cuesta J.M. (2014). Effect of cellulose, hemicellulose and lignin contents on pyrolysis and combustion of natural fibers. J Anal Appl Pyrolysis 107: 323–331. doi:10.1016/j.jaap.2014.03.017

Du K.L., Swamy M.N.S. (2006). Neural networks in a softcomputing framework. Neural Networks a Softcomputing Framew 1–566. doi:10.1007/1-84628-303-5

Dubey R., Guruviah V. (2022). Machine learning approach for categorical biomass higher heating value prediction based on proximate analysis. Energy Sources, Part A Recover Util Environ Eff 44 (2): 3381–3394. doi:10.1080/15567036.2022.2065386

Dubey R., Guruviah V. (2023). Predictive Modeling of Higher Heating Value of Biomass Using Ensemble Machine Learning Approach. Arab J Sci Eng 48 (7): 9329–9338.

doi:10.1007/s13369-022-07346-8

Esteves B., Sen U., Pereira H. (2023). Influence of Chemical Composition on Heating Value of Biomass: A Review and Bibliometric Analysis. Energies 16 (10): 1–17. doi:10.3390/en16104226

Fouedjio F. (2020). Exact Conditioning of Regression Random Forest for Spatial Prediction. Artif Intell Geosci 1 (January): 11–23. doi:10.1016/j.aiig.2021.01.001

Gani A., Erdiwansyah, Desvita H., Munawar E., Mamat R., Nizar M., Darnas Y., Sarjono R.E. (2024). Comparative analysis of HHV and LHV values of biocoke fuel from palm oil mill solid waste. Case Stud Chem Environ Eng 9 (October 2023): 100581. doi:10.1016/j.cscee.2023.100581

García Nieto P.J., García-Gonzalo E., Paredes-Sánchez J.P., Bernardo Sánchez A., Menéndez Fernández M. (2019). Predictive modelling of the higher heating value in biomass torrefaction for the energy treatment process using machine-learning techniques. Neural Comput Appl 31 (12): 8823–8836. doi:10.1007/s00521-018-3870-x

García-Martín E., Rodrigues C.F., Riley G., Grahn H. (2019). Estimation of energy consumption in machine learning. J Parallel Distrib Comput 134: 75–88. doi:10.1016/j.jpdc.2019.07.007

Gholami R., Fakhari N. (2017). Support Vector Machine: Principles, Parameters, and Applications, 1st Edition, Handbook of Neural Computation. Elsevier Inc., 515–535 pp. doi:10.1016/B978-0-12-811318-9.00027-2

Giwa S.O., Adekomaya S.O., Adama K.O., Mukaila M.O. (2015). Prediction of selected biodiesel fuel properties using artificial neural network. Front Energy 9 (4): 433–445. doi:10.1007/s11708-015-0383-5

Grossi E., Buscema M. (2007). Introduction to artificial neural networks. Eur J Gastroenterol Hepatol 19 (12): 1046–1054. doi:10.1097/MEG.0b013e3282f198a0

Güleç F., Pekaslan D., Williams O., Lester E. (2022). Predictability of higher heating value of biomass feedstocks via proximate and ultimate analyses – A comprehensive study of artificial neural network applications. Fuel 320 (January). doi:10.1016/j.fuel.2022.123944

Guo B., Ringwood J. V. (2021). Non-linear modeling of a vibro-impact wave energy converter. IEEE Trans Sustain Energy 12 (1): 492–500. doi:10.1109/TSTE.2020.3007926

Haglin J.M., Jimenez G., Eltorai A.E.M. (2019). Artificial neural networks in medicine. Health Technol (Berl) 9 (1). doi:10.1007/s12553-018-0244-4

Hamdan M.H., Roach D.C. (2022). The Sigmoid Neural Network Activation Function and its Connections to Airy's and the Nield-Kuznetsov Functions. Proof 2 (April): 108–114.

doi:10.37394/232020.2022.2.13

Hellerstein J.M., Berkeley U.C. (2008). DataCleaning-ucb(2). United Nations Econ Comm Eur 1–42

Hewitt M., Frejinger E. (2020). Data-driven optimization model customization. Eur J Oper Res 287 (2): 438–451. doi:10.1016/j.ejor.2020.05.010

Hoang N.D. (2023). Compressive Strength Estimation of Rice Husk Ash-Blended Concrete Using Deep Neural Network Regression with an Asymmetric Loss Function. Iran J Sci Technol - Trans Civ Eng 47 (3): 1547–1565. doi:10.1007/s40996-022-01015-4

Horvat G. (2015). Reakcijska kalorimetrija. Kem u Ind 64 (11–12): 633–644. doi:10.15255/kui.2014.023

Hossain R., Timmer D.D. (2021). Machine learning model optimization with hyper parameter tuning approach. Glob J Comput Sci Technol 21 (2): 7–13

Hosseinzadeh A., Baziar M., Alidadi H., Zhou J.L., Altaee A., Najafpoor A.A., Jafarpour S. (2020). Application of artificial neural network and multiple linear regression in modeling nutrient recovery in vermicompost under different conditions. Bioresour Technol 303 (January): 122926. doi:10.1016/j.biortech.2020.122926

Huang Y.F., Lo S.L. (2020). Predicting heating value of lignocellulosic biomass based on elemental analysis. Energy 191 (xxxx): 116501. doi:10.1016/j.energy.2019.116501

Jaiswal J.K., Samikannu R. (2017). Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. Proc - 2nd World Congr Comput Commun Technol WCCCT 2017 (February 2017): 65–68. doi:10.1109/WCCCT.2016.25

Jat D.S., Dhaka P., Limbo A. (2018). Applications of statistical techniques and artificial neural networks: A review. J Stat Manag Syst 21 (4): 639–645. doi:10.1080/09720510.2018.1475073

Kalivodová M., Baláš M., Milčák P., Lisá H., Lisý M., Lachman J., Kracík P., Križan P., Vejražka K. (2022). the Determination of Higher Heating Value By Calculation Based on Elemental Analysis. Paliva 14 (1): 8–20. doi:10.35933/paliva.2022.01.02

Kartal F., Özveren U. (2020). A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. Energy 209. doi:10.1016/j.energy.2020.118457

Keybondorian E., Zanbouri H., Bemani A., Hamule T. (2017). Estimation of the higher heating value of biomass using proximate analysis. Energy Sources, Part A Recover Util Environ Eff 39 (20): 2025–2030. doi:10.1080/15567036.2017.1400609

Khoshroo A., Emrouznejad A., Ghaffarizadeh A., Kasraei M., Omid M. (2018). Sensitivity analysis of energy inputs in crop production using artificial neural networks. J Clean

Prod 197: 992–998. doi:10.1016/j.jclepro.2018.05.249

Kim H.S., Choi D., Yoo D.G., Kim K.P. (2022). Hyperparameter Sensitivity Analysis of Deep Learning-Based Pipe Burst Detection Model for Multiregional Water Supply Networks. Sustain 14 (21). doi:10.3390/su142113788

Kujawska J., Kulisz M., Oleszczuk P., Cel W. (2023). Improved Prediction of the Higher Heating Value of Biomass Using an Artificial Neural Network Model Based on the Selection of Input Parameters. Energies 16 (10). doi:10.3390/en16104162

Kumar Sarangi P., Subudhi S., Bhatia L., Saha K., Mudgil D., Prasad Shadangi K., Srivastava R.K., Pattnaik B., Arya R.K. (2023). Utilization of agricultural waste biomass and recycling toward circular bioeconomy. Environ Sci Pollut Res 30 (4): 8526–8539. doi:10.1007/s11356-022-20669-1

Kwaghger A., Iortyer L.A.&. (2017). the Development of Equations for Estimating High Heating Values From Proximate and Ultimate Analysis for Some Selected Indigenous Fuel Woods. Eur J Eng Technol 5 (3): 21–33

La Rocca M., Perna C. (2022). Opening the Black Box: Bootstrapping Sensitivity Measures in Neural Networks for Interpretable Machine Learning. Stats 5 (2): 440–457. doi:10.3390/stats5020026

Lai J.P., Chang Y.M., Chen C.H., Pai P.F. (2020). A survey of machine learning models in renewable energy predictions. Appl Sci 10 (17). doi:10.3390/app10175975

Li S., Yang B., Qi F. (2016). Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. Combust Flame 168: 53–64. doi:10.1016/j.combustflame.2016.03.028

Lisý M., Lisá H., Jecha D., Baláš M., Križan P. (2020). Characteristic properties of alternative biomass fuels. Energies 13 (6): 1–17. doi:10.3390/en13061448

Loucks D.P., van Beek E. (2017). Water resource systems planning and management: An introduction to methods, models, and applications, Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications. 1–624 pp. doi:10.1007/978-3-319-44234-1

Lujano-Rojas J.M., Dufo-López R., Artal-Sevil J.S., García-Paricio E. (2023). Searching for Promisingly Trained Artificial Neural Networks. Forecasting 5 (3): 550–575. doi:10.3390/forecast5030031

Maksimuk Y., Antonava Z., Krouk V., Korsakova A., Kursevich V. (2021). Prediction of higher heating value (HHV) based on the structural composition for biomass. Fuel 299 (March): 120860. doi:10.1016/j.fuel.2021.120860

Mandley S., Wicke B., Junginger H., van Vuuren D., Daioglou V. (2022). Integrated assessment of the role of bioenergy within the EU energy transition targets to 2050. GCB Bioenergy 14 (2): 157–172. doi:10.1111/gcbb.12908

Masi F., Stefanou I., Vannucci P., Maffi-Berthier V. (2021). Thermodynamics-based Artificial Neural Networks for constitutive modeling. J Mech Phys Solids 147 (October 2020): 104277. doi:10.1016/j.jmps.2020.104277

Matveeva A., Bychkov A. (2022). How to Train an Artificial Neural Network to Predict Higher Heating Values of Biofuel. Energies 15 (19): 1–13. doi:10.3390/en15197083

Mesroghli S., Jorjani E., Chehreh Chelgani S. (2009). Estimation of gross calorific value based on coal analysis using regression and artificial neural networks. Int J Coal Geol 79 (1–2): 49–54. doi:10.1016/j.coal.2009.04.002

Mohanty A.K., Misra M., Drzal L.T. (2002). Sustainable Bio-Composites from renewable resources: Opportunities and challenges in the green materials world. J Polym Environ 10 (1–2): 19–26. doi:10.1023/A:1021013921916

Montesinos López O.A., Montesinos López A., Crossa J. (2022). Multivariate Statistical Machine Learning Methods for Genomic Prediction, Multivariate Statistical Machine Learning Methods for Genomic Prediction. doi:10.1007/978-3-030-89010-0

Morala P., Cifuentes J.A., Lillo R.E., Ucar I. (2021). Towards a mathematical framework to inform neural network modelling via polynomial regression. Neural Networks 142: 57–72. doi:10.1016/j.neunet.2021.04.036

Mustapa R.F., Dahlan N.Y., Yassin I.M., Mohd Nordin A.H., Zabidi A. (2019). Energy consumption prediction through linear and non-linear baseline energy model. Indones J Electr Eng Comput Sci 17 (1): 102–109. doi:10.11591/ijeecs.v17.i1.pp102-109

Naik D.L., kiran R. (2021). A novel sensitivity-based method for feature selection. J Big Data 8 (1). doi:10.1186/s40537-021-00515-w

Nhuchhen D.R., Afzal M.T. (2017). HHV predicting correlations for torrefied biomass using proximate and ultimate analyses. Bioengineering 4 (1). doi:10.3390/bioengineering4010007

Nimmanterdwong P., Chalermsinsuwan B., Piumsomboon P. (2021). Prediction of lignocellulosic biomass structural components from ultimate/proximate analysis. Energy 222: 119945. doi:10.1016/j.energy.2021.119945

Noushabadi A.S., Dashti A., Ahmadijokani F., Hu J., Mohammadi A.H. (2021). Estimation of higher heating values (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation. Renew Energy 179: 550–562. doi:10.1016/j.renene.2021.07.003

Obafemi O., Stephen A., Ajayi O., Nkosinathi M. (2019). A survey of artificial neural network-based prediction models for thermal properties of biomass. Procedia Manuf 33: 184–191. doi:10.1016/j.promfg.2019.04.103

Olatunji O., Akinlabi S., Oluseyi A., Peter M., Madushele N. (2018). Experimental investigation of thermal properties of Lignocellulosic biomass: A review. IOP Conf Ser

Mater Sci Eng 413 (1). doi:10.1088/1757-899X/413/1/012054

Olive D.J. (2017). Theory for Linear Models. Linear Regres 313–342. doi:10.1007/978-3-319-55252-1_11

Onochie U.P., Ofomatah A.C., Onwurah C., Tyopine A.A., Akingba O.O., Kubeynje B.F., Aluma C.C., Alozie C. (2023). Potentials of Biomass Waste Resources with Respect to their Calorific Value, Proximate and Ultimate Analysis for Energy Utilization. IOP Conf Ser Earth Environ Sci 1178 (1). doi:10.1088/1755-1315/1178/1/012012

Ozveren U. (2017). An artificial intelligence approach to predict gross heating value of lignocellulosic fuels. J Energy Inst 90 (3): 397–407. doi:10.1016/j.joei.2016.04.003

Özyuğuran A., Yaman S., Küçükbayrak S. (2018). Prediction of calorific value of biomass based on elemental analysis. Int Adv Res Eng J 02 (03): 254–260

Paredes-Astudillo Y.A., Botta-Genoulaz V., Montoya-Torres J.R. (2022). Comparing linear and non-linear modelling approaches of learning effects in 2-stage flow-shop scheduling problems. IFAC-PapersOnLine 55 (10): 842–847. doi:10.1016/j.ifacol.2022.09.517

Park S., Kim S., Kim H.E., Oh K.C., Kim S.J., Cho L.H., Jeon Y.K., Kim D.H. (2023). Calorific Value Prediction model using structure composition of Heat-Treated Lignocellulosic biomass. Energies 16 (23): 7896. doi:10.3390/en16237896

Pattanayak S., Loha C., Hauchhum L., Sailo L. (2021). Application of MLP-ANN models for estimating the higher heating value of bamboo biomass. Biomass Convers Biorefinery 11 (6): 2499–2508. doi:10.1007/s13399-020-00685-2

Peng F., Peng P., Xu F., Sun R.C. (2012). Fractional purification and bioconversion of hemicelluloses. Biotechnol Adv 30 (4): 879–903. doi:10.1016/j.biotechadv.2012.01.018

Perea-Moreno M.A., Samerón-Manzano E., Perea-Moreno A.J. (2019). Biomass as renewable energy: Worldwide research trends. Sustain 11 (3). doi:10.3390/su11030863

Petrova T. (2021). Estimation of the higher heating values for lignocellulosic biofuels. Proc 2021 6th Int Symp Environ Energies Appl EFEA 2021 14–18. doi:10.1109/EFEA49713.2021.9406264

Pinto E., Aggrey W.N., Boakye P., Amenuvor G., Sokama-Neuyam Y.A., Fokuo M.K., Karimaie H., Sarkodie K., Adenutsi C.D., Erzuah S., Rockson M.A.D. (2022). Cellulose processing from biomass and its derivatization into carboxymethylcellulose: A review. Sci African 15 (March): e01078. doi:10.1016/j.sciaf.2021.e01078

Pisner D.A., Schnyer D.M. (2019). Support vector machine, Machine Learning: Methods and Applications to Brain Disorders. Elsevier Inc., 101–121 pp. doi:10.1016/B978-0-12-815739-8.00006-7

Probst P., Wright M.N., Boulesteix A.L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov 9 (3): 1–15. doi:10.1002/widm.1301

Pugliese R., Regondi S., Marini R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. Data Sci Manag 4 (December): 19–29. doi:10.1016/j.dsm.2021.12.002

Puig-Arnavat M., Bruno J.C. (2015). Artificial Neural Networks for Thermochemical Conversion of Biomass. Recent Adv Thermochem Convers Biomass 133–156. doi:10.1016/B978-0-444-63289-0.00005-3

Qi Y. (2012). Random forest for bioinformatics. Ensemble Mach Learn Methods Appl 307–323. doi:10.1007/9781441993267_10

Razavi S., Jakeman A., Saltelli A., Prieur C., Iooss B., Borgonovo E., Plischke E., Lo Piano S., Iwanaga T., Becker W., Tarantola S., Guillaume J.H.A., Jakeman J., Gupta H., Melillo N., Rabitti G., Chabridon V., Duan Q., Sun X., Smith S., Sheikholeslami R., Hosseini N., Asadzadeh M., Puy A., Kucherenko S., Maier H.R. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. Environ Model Softw 137 (December 2020). doi:10.1016/j.envsoft.2020.104954

Roy R., Ray S. (2020). Development of a non-linear model for prediction of higher heating value from the proximate composition of lignocellulosic biomass. Energy Sources, Part A Recover Util Environ Eff 00 (00): 1–14. doi:10.1080/15567036.2020.1817191

Sarmento R., Costa V. (2017). Introduction to Linear Regression (April): 140–147. doi:10.4018/978-1-68318-016-6.ch006

Scarlat Nicolae, Dallemand Jean-Francois, Taylor Nigel, Banja Manjola, Sanchez Lopez Javier, Avraamides Marios. (2019). Brief on biomass for energy in the European Union | EU Science Hub. EC Publ 1–8

Schaffer B.E., Nordbotten J.M., Rodriguez-Iturbe I. (2015). Plant biomass and soil moisture dynamics: Analytical results. Proc R Soc A Math Phys Eng Sci 471 (2183). doi:10.1098/rspa.2015.0179

Schonlau M., Zou R.Y. (2020). The random forest algorithm for statistical learning. Stata J 20 (1): 3–29. doi:10.1177/1536867X20909688

Scornet E., Biau G., Vert J.P. (2015). Consistency of random forests. Ann Stat 43 (4): 1716–1741. doi:10.1214/15-AOS1321

Scornet E., Biau G., Vert J.P. (2015). Consistency of random forests. Ann Stat 43 (4): 1716–1741. doi:10.1214/15-AOS1321

Searchinger T., Heimlich R., Houghton R.A., Dong F., Elobeid A., Fabiosa J., Tokgoz S., Hayes D., Yu T.H. (2008). Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science (80-) 319 (5867): 1238–

1240. doi:10.1126/science.1151861

Shadangi K.P., Sarangi P.K., Behera A.K. (2023). Characterization techniques of biomass: physico-chemical, elemental, and biological, Bioenergy Engineering: Fundamentals, Methods, Modelling, and Applications. Elsevier Ltd, 51–66 pp. doi:10.1016/B978-0-323-98363-1.00022-3

Shakiba F.M., Zhou M. (2021). Novel Analog Implementation of a Hyperbolic Tangent Neuron in Artificial Neural Networks. IEEE Trans Ind Electron 68 (11): 10856–10867. doi:10.1109/TIE.2020.3034856

Shao M., Wang X., Bu Z., Chen X., Wang Y. (2020). Prediction of energy consumption in hotel buildings via support vector machines. Sustain Cities Soc 57 (March): 102128. doi:10.1016/j.scs.2020.102128

Sheng C., Azevedo J.L.T. (2005). Estimating the higher heating value of biomass fuels from basic analysis data. Biomass and Bioenergy 28 (5): 499–507. doi:10.1016/j.biombioe.2004.11.008

Sivabalan K., Hassan S., Ya H., Pasupuleti J. (2021). A review on the characteristic of biomass and classification of bioenergy through direct combustion and gasification as an alternative power supply. J Phys Conf Ser 1831 (1). doi:10.1088/1742-6596/1831/1/012033

Soomro A., Chen S., Sun Z., Ma S., Xiang W. (2021). Chemometric modelling on element compositions and product distributions of cellulose and lignin. Biomass Convers Biorefinery 11 (5): 2233–2246. doi:10.1007/s13399-020-00638-9

Taki M., Rohani A. (2022). Machine learning models for prediction the Higher Heating Value (HHV) of Municipal Solid Waste (MSW) for waste-to-energy evaluation. Case Stud Therm Eng 31 (January): 101823. doi:10.1016/j.csite.2022.101823

Titus B.D., Brown K., Helmisaari H.S., Vanguelova E., Stupak I., Evans A., Clarke N., Guidi C., Bruckman V.J., Varnagiryte-Kabasinskiene I., Armolaitis K., de Vries W., Hirai K., Kaarakka L., Hogg K., Reece P. (2021). Sustainable forest biomass: a review of current residue harvesting guidelines. Energy Sustain Soc 11 (1): 1–32. doi:10.1186/s13705-021-00281-w

Tun M.M., Juchelkova D., Win M.M., Thu A.M., Puchor T. (2019). Biomass energy: An overview of biomass sources, energy potential, and management in Southeast Asian countries. Resources 8 (2). doi:10.3390/resources8020081

Urban, S. (2018). Neural Network Architectures and Activation Functions: A Gaussian Process Approach. (P. van der Smagt, D. Cremers, & B. Bischl, Eds.) Universitätsbibliothek der TU München.

Vanneschi L., Silva S. (2023). Artificial Neural Networks. Nat Comput Ser 161–204. doi:10.1007/978-3-031-17922-8_7

Vardiambasis I.O., Kapetanakis T.N., Nikolopoulos C.D., Trang T.K., Tsubota T., Keyikoglu R., Khataee A., Kalderis D. (2020). Hydrochars as emerging biofuels: Recent advances and application of artificial neural networks for the prediction of heating values. Energies 13 (17): 1–20. doi:10.3390/en13174572

Veza I., Irianto, Panchal H., Paristiawan P.A., Idris M., Fattah I.M.R., Putra N.R., Silambarasan R. (2022). Improved prediction accuracy of biomass heating value using proximate analysis with various ANN training algorithms. Results Eng 16 (September): 100688. doi:10.1016/j.rineng.2022.100688

Voća N., Leto J., Karažija T., Bilandžija N., Peter A., Kutnjak H., Šurić J., Poljak M. (2021). Energy properties and biomass yield of miscanthus x giganteus fertilized by municipal sewage sludge. Molecules 26 (14). doi:10.3390/molecules26144371

Wang D., Peng J., Yu Q., Chen Y., Yu H. (2019). Support vector machine algorithm for automatically identifying depositional microfacies usingwell logs. Sustain 11 (7). doi:10.3390/su11071919

Wang L., Wu X., Gao X., Wu Y., Cen K. (2021). A method for in-situ measurement of calorific value of coal: a numerical study. Thermochim Acta 703 (April): 179011. doi:10.1016/j.tca.2021.179011

Wang L., Zhou G., Wang K., Cui Y. (2011). Numerical methods for solving high order polynomial equations. Proc - 4th Int Conf Inf Comput ICIC 2011 150–153. doi:10.1109/ICIC.2011.83

Wanto A., Windarto A.P., Hartama D., Parlina I. (2017). Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density. IJISTECH (International J Inf Syst Technol 1 (1): 43. doi:10.30645/ijistech.v1i1.6

Watkins D., Nuruddin M., Hosur M., Tcherbi-Narteh A., Jeelani S. (2015). Extraction and characterization of lignin from different biomass resources. J Mater Res Technol 4 (1): 26–32. doi:10.1016/j.jmrt.2014.10.009

White E.M. (2010). Woody biomass for bioenergy and biofuels in the united states- a briefing paper. USDA For Serv - Gen Tech Rep PNW-GTR (825): 1–45

White W.M. (2017). Analytical techniques. Encycl Earth Sci Ser 1–5. doi:10.1007/978-3-319-39193-9_55-1

Xing J., Luo K., Wang H., Gao Z., Fan J. (2019a). A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy 188: 116077. doi:10.1016/j.energy.2019.116077

Xing J., Wang H., Luo K., Wang S., Bai Y., Fan J. (2019b). Predictive single-step kinetic model of biomass devolatilization for CFD applications: A comparison study of empirical correlations (EC), artificial neural networks (ANN) and random forest (RF). Renew Energy 136: 104–114. doi:10.1016/j.renene.2018.12.088

Xue L., Liu Yuetian, Xiong Y., Liu Yanli, Cui X., Lei G. (2021). A data-driven shale gas production forecasting method based on the multi-objective random forest regression. J Pet Sci Eng 196 (July 2020): 107801. doi:10.1016/j.petrol.2020.107801

Yan R., Wang S. (2022). Integrating prediction with optimization: Models and applications in transportation management. Multimodal Transp 1 (3): 100018. doi:10.1016/j.multra.2022.100018

Yang X., Li H., Wang Y., Qu L. (2023). Predicting Higher Heating Value of Sewage Sludges via Artificial Neural Network Based on Proximate and Ultimate Analyses. Water (Switzerland) 15 (4). doi:10.3390/w15040674

Yi L., Feng J., Qin Y.H., Li W.Y. (2017). Prediction of elemental composition of coal using proximate analysis. Fuel 193: 315–321. doi:10.1016/j.fuel.2016.12.044

Zhang F., Li C., Yu Y., Johnson D.M. (2019). Resources and future availability of agricultural biomass for energy use in Beijing. Energies 12 (10): 1–14. doi:10.3390/en12101828

Zhang H., Zimmerman J., Nettleton D., Nordman D.J. (2020). Random Forest Prediction Intervals. Am Stat 74 (4): 392–406. doi:10.1080/00031305.2019.1585288

Zhang P. (2019). A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. Appl Soft Comput J 85: 105859. doi:10.1016/j.asoc.2019.105859

Zoppis I., Mauri G., Dondi R. (2018). Kernel methods: Support vector machines, Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. Elsevier Ltd., 503–510 pp. doi:10.1016/B978-0-12-809633-8.20342-7

CURRICULUM VITAE

Ivan Brandić, born on 7 December 1995 in Mali Lošinj, Croatia, successfully graduated from the General Secondary School in Mali Lošinj in 2015. In the same year, he began his studies at the Faculty of Agriculture at the University of Zagreb, specialising in agricultural technology. In 2018, he obtained the title of Bachelor of Agricultural Technology and defended his thesis entitled "Modern transport technology and short supply chains for agricultural products with an overview of the islands". After completing his undergraduate studies, Brandić enrolled in postgraduate studies in mechanisation in 2018, which he completed with the highest distinction (summa cum laude) by defending his thesis on "Innovative mechanisation and remote sensing application in energy crop cultivation", thus obtaining the title of Master of Engineering in Agronomy. In addition, in 2020 he enrolled at the Algebra Open School for Education for the programme "Project manager of projects financed from EU funds", which he successfully completed. During his academic career, he participated in several international and national symposia with oral presentations in Croatia, Czech Republic and Lithuania. In November 2023, he participated in the PhD conference "MendelNet" in Brno, Czech Republic, "where he was awarded for the best presentation in the section "Techniques and Technology". As an author or co-author, he has been involved in numerous A1 and A2 category papers and conference presentations. He is proficient in the Python programming language and speaks fluent English, as well as basic German and Italian.

List of publications:

Scientific papers in journals:

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

Brandić I., Voća N., Leto J., Bilandžija N. (2024). Modelling the Yield and Estimating the Energy Properties of Miscanthus x Giganteus in Different Harvest Periods. AgriEngineering 6 (1): 423–437. doi:10.3390/agriengineering6010026

Brandić I., Pezo L., Voća N., Leto J., Šurić J., Peter A., Bilandžija N. (2024). Assessing the properties of Miscanthus x giganteus under varying levels of ash fertilization treatment and regression neural network insight into calorific value. Thermal Science.

Brandić I., Antonović A., Pezo L., Matin B., Krička T., Jurišić V., Špelić K., Kontek M., Kukuruzović J., Grubor M., Matin A. (2023). Energy potentials of agricultural biomass and the possibility of modelling using RFR and SVM models. Energies 16 (2): 690. doi:10.3390/en16020690

Brandić I., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2023). Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass. Mathematics 11 (9). doi:10.3390/math11092098

Brandić I., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2022). Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis. Mathematics 10 (20): 1–12. doi:10.3390/math10203732

Brandić I., Voća N., Gunjača J., Lončar B., Bilandžija N., Peter A., Šurić J., Pezo L. (2024). Biomass higher heating value prediction machine learning insights into ultimate, proximate, and structural analysis datasets. Energy Sources, Part A Recover Util Environ Eff 46 (1): 2842–2854. doi:10.1080/15567036.2024.2309303

Brandić I., Sito S., Kovačev I., Fabijanić G., Čopec K., Bilandžija N. (2022). Trends in mechanization development in collecting biomass of energy crops. Glasnik Zaštite Bilja, 45. (4.), 70-80. https://doi.org/10.31727/gzb.45.4.7

Matin, A., **Brandić, I.,** Majdak, T., Metić, L. i Matin, B. (2024). Flax processing and storage. Glasnik Zaštite Bilja, 47. (3.), 87-87

Špelić K., Panjičko M., Zupančić G.D., Lončar A., **Brandić I.**, Tomić I., Matin A., Krička T., Jurišić V. (2024). Towards a sustainable energy future: Evaluating Arundo donax L. in continuous anaerobic digestion for biogas production. GCB Bioenergy 16 (5): 1–14. doi:10.1111/gcbb.13135

Demir Hasan, Demir Hande, Lončar B., Pezo L., **Brandić I.,** Voća N., Yilmaz F. (2023). Optimization of Caper Drying Using Response Surface Methodology and Artificial Neural Networks for Energy Efficiency Characteristics. Energies 16 (4). doi:10.3390/en16041687

Galić A., Pliestić S., Kovačev I., Čopec K., **Brandić I.** (2023). Application of machine learning models in the feed mixture production. Krmiva, 65 (2), 69-76. https://doi.org/10.33128/k.65.2.2

Kontek M., Brezinščak L., Jurišić V., **Brandić I.,** Antonović A., Matin B., Špelić K., Krička T., Matin A. (2023). Mitigating the Energy Crisis: Utilization of Seed Production Wastes for Energy Production in Continental Croatia. Energies 16 (2): 1–11. doi:10.3390/en16020738

Kukuruzović J., Matin A., Kontek M., Krička T., Matin B., **Brandić I.,** Antonović A. (2023). The Effects of Demineralization on Reducing Ash Content in Corn and Soy Biomass with the Goal of Increasing Biofuel Quality. Energies 16 (2). doi:10.3390/en16020967

Lubura J., Pezo L., Sandu M.A., Voronova V., Donsì F., Šic Žlabur J., Ribić B., Peter A., Šurić J., **Brandić I.,** Klõga M., Ostojić S., Pataro G., Virsta A., Oros A.E., Micić D., Đurović S., De Feo G., Procentese A., Voća N. (2022). Food Recognition and Food Waste Estimation Using Convolutional Neural Network. Electron 11 (22): 1–14. doi:10.3390/electronics11223746

Matin A., **Brandić I.,** Antonović A. (2024). Effect of conduction drying on nutrient and fatty acid profiles : a comparative analysis of hazelnuts and walnuts (February). doi:10.3389/fsufs.2024.1351309

Matin A., **Brandić I.,** Voća N., Bilandžija N., Matin B., Jurišić V., Antonović A., Krička T. (2023a). Changes in the Properties of Hazelnut Shells Due to Conduction Drying. Agric 13 (3): 1–15. doi:10.3390/agriculture13030589

Matin A., **Brandić I.,** Voća N., Bilandžija N., Matin B., Jurišić V., Špelić K., Antonović A., Grubor M., Krička T. (2023b). Influence of Conduction Drying on the Physical and Combustion Properties of Hazelnut Shell. Energies 16 (3): 1–12. doi:10.3390/en16031297

Matin A., Krička T., Grubor M., Jurišić V., Majdak T., Špelić K., Majdak J., Matin B., Jovičić N., Antonović A**., Brandić I.** (2023c). Energy properties of agricultural biomass after the pyrolysis. J Process Energy Agric 27 (1): 39–44. doi:10.5937/jpea27-43553

Matin B., Leto J., Antonović A., **Brandić I.,** Jurišić V., Matin A., Krička T., Grubor M., Kontek M., Bilandžija N. (2023). Energetic Properties and Biomass Productivity of Switchgrass (Panicum virgatum L.) under Agroecological Conditions in Northwestern Croatia. Agronomy 13 (4). doi:10.3390/agronomy13041161

Šurić J., **Brandić I.,** Peter A., Bilandžija N., Leto J., Karažija T., Kutnjak H., Poljak M., Voća N. (2022a). Wastewater Sewage Sludge Management via Production of the Energy Crop Virginia Mallow. Agronomy 12 (7). doi:10.3390/agronomy12071578

Šurić J., Šic Žlabur J., Peter A., **Brandić I.,** Voća S., Dujmović M., Leto J., Voća N. (2022b). Energy vs. Nutritional Potential of Virginia Mallow (Sida hermaphrodita L.) and Cup Plant (Silphium perfoliatum L.). Plants 11 (21). doi:10.3390/plants11212906

Šurić J., Voća N., Peter A., Bilandžija N., **Brandić I.,** Pezo L., Leto J. (2023). Use of Artificial Neural Networks to Model Biomass Properties of Miscanthus (Miscanthus × giganteus) and Virginia Mallow (Sida hermaphrodita L.) in View of Harvest Season. Energies 16 (11). doi:10.3390/en16114312

Scientific papers in conference proceedings:

Tomić I., **Brandić I.,** Jurišić V., Špelić K., Grubor M., Matin A. (2024). Zbrinjavanje komine masline sušenjem // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 577-582.

Đurović A., Matin B., **Brandić I**., Tomić I., Matin A., Antonović A. (2024). Influence of soybean (Glycine max) and hemp (Cannabis sativa) stalks biomass to glycerol solvent ratio on liquefied properties // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 507-513.

Matin A., **Brandić I.**, Antonović A., Grubor M., Jurišić V., Majdak T., Špelić K., Matin B., Krička T. (2024). Optimizacija energetske učinkovitosti procesa sušenja zrna kukuruza na osnovu temperaturnih, psihrometrijskih i energetskih parametara // Zbornik radova 59.

Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 483-488.

Brandić I., Pezo L., Krička T., Voća N., Matin A. (2024). Primjena modela umjetnih neuronskih mreža u modeliranju procesa sušenja zrna kukuruza // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 536-541.

Špelić K., **Brandić I.**, Matin A., Tomić I., Jurišić V. (2024). Utjecaj procesa anaerobne fermentacije na sadržaj lignina u proizvodnji silaže i sjenaže miskantusa // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 542-547.

Grubor M., Matin A., Antonović A., Matin B., Kovačev I., **Brandić I.** (2024). Poljoprivreda 4.0 u oblikovanju pametne poljoprivrede // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 570-576.

Brandić I., Voća N., Leto J., Bilandžija N. (2023). Application of machine learning models in miscanthus x giganteus yield estimation // Proceedings of the 49th International Symposium "Actual Tasks on Agricultural Engineering". 2023. str. 193-201.

Brandić I., Pezo L., Šic Žlabur J., Peter A., Šurić J., Voća N. (2023). Primjena modela umjetnih neuronskih mreža u predviđanju otpada od hrane na području Europske unije. 2023. str. 416-421.

Šurić J., Peter A., Šic Žlabur J., **Brandić I.,** Voća N. (2023). Uzroci i mogućnosti sprječavanja nastajanja otpada od hrane // Zbornik sažetaka 58. hrvatskog i 18. međunarodnog simpozija agronoma. 2023. str. 411-415.

Peter A., Šurić J., Dujmović Purgar D., **Brandić I.,** Voća N. (2023). Impacts, threats, and potential benefits of invasive plant species. // Zbornik sažetaka 58. hrvatskog i 18. međunarodnog simpozija agronoma. 2023. str. 462-466.

Brandić I.,Peter A., Šurić J., Pezo L., Voća N. (2022). Umjetne neuronske mreže kao statistički alat u predviđanju donje ogrjevne vrijednosti biomase // Zbornik radova / Majić, Ivana ; Antunović, Zvonko (ur.). Vodice, 2022. str. 619-623.

Šurić J., Peter A., **Brandić I.,** Leto J., Bilandžija N., Voća N. (2022). Razlike u lignoceluoloznom sastavu energetskih kultura u dva roka žetve // Zbornik radova / Ivana Majić, Zvonko Antunović (ur.). Osijek, 2022. str. 671-675

Abstracts in conference proceedings:

Jurišić V., Panjičko M., Zupančić GD., Lončar A., **Brandić I.,** Tomić I., Matin A., Špelić K. (2024). Biogas production from Arundo donax in a continuous digestion system // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 251-251

Jurišić V., Čorba M., Panjičko M., Zupančić GD., Lončar A., **Brandić I.,** Tomić I., Matin A., Špelić K. (2024). Biodegradability of Arundo donax during a continuous mesophilic anaerobic digestion // Zbornik radova 59. Hrvatski i 19. Međunarodni simpozija agronoma. Zagreb: Sveučilište u Zagrebu Agronomski fakultet, Zagreb, Hrvatska, 2024. str. 250-250

Galić A., Pliestić S., Kovačev I., Čopec K., **Brandić I.** (2023). Primjena modela strojnog i dubokog učenja u tehnologiji proizvodnje krmnih smjesa // Zbornik sažetaka 28. međunarodnog savjetovanja Krmiva 2023 / Zvonko, Antunović ; Zlatko, Janječić (ur.). Zagreb: Krmiva, 2023. str. 50-50

Matin B., Matin A., **Brandić I.,** Đurović A., Ištvanić J., Antonović A. (2023). Lignocellulose composition, proximate analysis and heat value of certain forest and energy crop biomasses and their potential as raw materials for the production of solid biofuels. Ohrid, sjeverna makedonija: ss. Cyril and methodius university in skopje faculty of design and technologies of furniture and interior – Skopje Republic of North Macedonia.

APPENDICES

Article 1. Brandić I., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2022). Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis. Mathematics.

Article 2. Brandić I., Antonović A., Pezo L., Matin B., Krička T., Jurišić V., Špelić K., Kontek M., Kukuruzović J., Grubor M., Matin A. (2023). Energy Potentials of Agricultural Biomass and the Possibility of Modelling Using RFR and SVM Models. Energies.

Article 3. Brandić I., Pezo L., Bilandžija N., Peter A., Šurić J., Voća N. (2023). Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass. Mathematics.

Article 4. Brandić I., Pezo L., Voća N., Leto J., Šurić J., Peter A., Bilandžija N. (2024). Assessing the properties of Miscanthus x giganteus under varying levels of ash fertilization treatment and regression neural network insight into calorific value. Thermal Science.

Article 5. Brandić I., Voća N., Gunjača J., Lončar B., Bilandžija N., Peter A., Šurić J., Pezo L. (2024). Biomass higher heating value prediction: Machine Learning Insight into Ultimate, Proximate and Structural analysis datasets. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects.

mathematics

MDPI

Article

Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis

Ivan Brandić [1], Lato Pezo [2], Nikola Bilandžija [1,*], Anamarija Peter [1], Jona Šurić [1] and Neven Voća [1]

[1] Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia
[2] Institute of General and Physical Chemistry, University of Belgrade, Studentski trg 12/V, 11000 Belgrade, Serbia
* Correspondence: nbilandzija@agr.hr

Abstract: Miscanthus is a perennial energy crop that produces high yields and has the potential to be converted into energy. The ultimate analysis determines the composition of the biomass and the energy value in terms of the higher heating value (HHV), which is the most important parameter in determining the quality of the fuel. In this study, an artificial neural network (ANN) model based on the principle of supervised learning was developed to predict the HHV of miscanthus biomass. The developed ANN model was compared with the models of predictive regression models (suggested from the literature) and the accuracy of the developed model was determined by the coefficient of determination. The paper presents data from 192 miscanthus biomass samples based on ultimate analysis and HHV. The developed model showed good properties and the possibility of prediction with high accuracy (R^2 = 0.77). The paper proves the possibility of using ANN models in practical application in determining fuel properties of biomass energy crops and greater accuracy in predicting HHV than the regression models offered in the literature.

Keywords: artificial neural network; prediction; miscanthus; energy potential

MSC: 49M37

Citation: Brandić, I.; Pezo, L.; Bilandžija, N.; Peter, A.; Šurić, J.; Voća, N. Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis. *Mathematics* **2022**, *10*, 3732. https://doi.org/10.3390/math10203732

Academic Editor: Krzysztof Ejsmont

Received: 1 September 2022
Accepted: 8 October 2022
Published: 11 October 2022

1. Introduction

Recently, energy crops have been increasingly used as raw materials for energy production. Cultivation of energy crops is possible on neglected (marginal) agricultural land that is not used for growing food crops. The production of thermal energy from biomass is highly efficient and sustainable. The main advantage of using biofuel from biomass is the reduction of greenhouse gases due to the neutrality of carbon dioxide. Research on energy crops for biomass production shows the possibility of environmental protection and economic production efficiency and provides a sustainable way of energy production [1]. By using biomass as an energy source, a significant reduction in greenhouse gas emissions can be achieved. For this reason, biomass is considered a good substitute for fossil fuels and has been increasingly studied recently [2]. According to the European Commission (European Commission, Joint Research Centre), biomass is one of the most important renewable energy sources in the EU and can provide the possibility of a reliable energy supply. Miscanthus is an energy crop used to produce biomass, and its cultivation provides high yields per unit area. Miscanthus is a perennial energy crop with low agrotechnical requirements and can be grown on marginal soils. The quality of biomass-derived fuels is influenced by the physical and chemical properties of the biomass. The content of carbon, hydrogen, nitrogen, sulfur, and oxygen determined by ultimate analysis are important chemical parameters that affects the quality of the fuel [3].

Ultimate analysis is important in determining the fuel properties [4]. The heating value indicates the heat energy generated during combustion. HHV is an important energy property of fuels that defines the energy efficiency of feedstock use and it is influenced by

the chemical composition of the raw material. HHV is an important aspect in evaluating the energy properties of biomass [5]. Biomass is composed of various elements, but carbon, hydrogen, and oxygen make up a majority (97–99%) of the biomass content [5]. Empirical methods for determining the composition and energy properties of biomass are time-consuming and costly, so mathematical models have recently been developed that can facilitate the prediction process. In determining the combustible properties of energy crops, prescribed laboratory methods are used that provide high precision of the final results. Recently, machine learning techniques have been increasingly used in the prediction of HHV biomass. In research by Dai et al., [6] extreme learning machine (ELM) was used as techniques with a signal forward neural network architecture to determine HHV biomass. The model used shows high accuracy in predicting biomass fuel values. Knowing this, ANN can be used as a mathematical tool for predicting the energy properties of biomass [7]. ANN as a form of non-linear models can calculate the HHV of miscanthus biomass, based on ultimate analysis, with high precision and are recognized as a potential method for predicting biomass heating value and reducing the time and cost of the process [8].

ANN belong to the field of artificial intelligence and have recently been increasingly used as a mathematical tool that enables predictions with great precision. ANN have several advantages over regression-based models. They can handle a large amount of aggregated data and can detect nonlinear relationships between dependent and independent variants as well as possible interactions between variables [9]. The application of ANN as a model for biomass research is still at an early stage, but over time there is growing interest in its use [10]. Özveren [11] conducted research in which an ANN model was developed as an artificial intelligence model for predicting biomass with higher heating values. The research shows the practical use of applying the ANN model as a method for predicting the energy values of biomass. Olatunji et al. [12] used ultimate analysis data of different types of waste in their research and developed the ANN model to predict the HHV. The model was used to predict energy properties to evaluate the possibility of converting waste into useful energy. Research has shown that algorithms can be successfully used in determining these properties. In a study conducted by Kartal & Özveren [13], an ANN model was developed to predict the gasification performance of different types of biomass. The developed model successfully simulated the vegetation process with an acceptable margin of error. The model also proved successful in predicting the calorific value of different biomass samples. Before creating an ANN, the data used for model must be divided into sets for training, testing, and validation. In several studies conducted using ANN models for prediction, the authors divided data sets in the ratio of 70% for training, 15% for testing, and 15% for model validation [14,15].

The aim of this work was to develop a ANN model for predicting HHV of miscanthus biomass based on ultimate analysis. In addition, already developed regression models for prediction of HHV were collected from the literature and used for the calculations. The input data used for ANN and the predictive regression models were based on the ultimate analysis and included data on the percentage of nitrogen (N), carbon (C), sulfur (S), hydrogen (H), and oxygen (O). ANN was developed using the principle of supervised learning and compared the obtained data on predicted HHV with the experimentally obtained data on HHV. Yoon's interpretation method was used to determine the relative importance of the input parameters in the ANN model calculations. Dashti et al. [8] states that it is of great importance to determine the factors of relevance (influence) of input variables on the target result. Noushabadi et al. [16] states that the relevance factor shows the influence of the elements of ultimate analysis (C, H, N, S, and O) on HHV. Positive and negative values of each parameter are the result of an increase or decrease of the input parameter on the output. The main objective of the study was to obtain an empirical model for predicting HHV values based on the input data of the ultimate analysis and to compare the R^2 values with existing regression models.

2. Materials and Methods

2.1. Crop Establishment and Data Collection

Voća et al. [17] stated that the planting of miscanthus was established in 2011 at the Grassland Center (Medvednica). It was harvested in March 2020, at the beginning of the next growing season. The testing of Miscanthus biomass samples was performed in the laboratory of the Faculty of Agriculture in Zagreb. The samples were dried in a laboratory dryer. After drying, the samples were ground in a laboratory mill. Each sample was analyzed three times to ensure accurate analysis. The percentages of C, H, N, and S were determined simultaneously using the dry combustion method CHNS analyzer. The calorific value was determined using an oxygen bomb calorimeter, given in MJ/kg in dry mass. Data from ultimate analysis and HHV data for miscanthus were collected from the literature and are presented in Table S1. Data on N, C, S, H, and O were collected for each sample, and the values were N (0.031–0.769%), C (49.45–53.42), S (0.055–1.28%), H (5.21–6.27%), O (39.91–48.92%), and HHV (15.53–19.25 MJ/kg). According to literature data, the value of HHV of miscanthus varies between 18.18–18.66 MJ/kg, N 0.28–0.39%, C 46.75–50%, S 0.13–0.19%, and H 5.76–6.09% [18], which shows that the presented data are in range with the data from the literature.

2.2. Statistical Analysis

Statistical processing was performed using the software package TIBCO STATISTICA 13.3.0 (StatSoft TIBCO Software Inc., Palo Alto, CA, USA). The analyzed data are presented as means with standard deviation. Analysis of variance (ANOVA) with Tukey's HSD post hoc test to compare sample means was used to examine variation in observed parameters.

To show the performance of the developed ANN model and predictive regression models for calculating HHV with ultimate analysis inputs (N, C, S, H, and O), it is necessary to calculate statistical parameters: reduced chi-square (x^2) (Equation (1)), root mean square error (RMSE) (Equation (2)), coefficient of determination (R^2), mean bias error (MBE) (Equation (3)), mean percentage error (MPE) (Equation (4)), and sum of squared estimate of errors (SSE) (Equation (5)). The RMSE shows the efficiency of the model by comparing the predicted values with the already measured values. The value obtained by the MBE is used as an indicator of the standard deviation of the predicted values from the measured values [19]. The listed parameters are given by the following formula [20].

$$x^2 = \frac{\sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2}{N - n} \quad (1)$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2 \right]^{1/2} \quad (2)$$

$$MBE = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i}) \quad (3)$$

$$MPE = \frac{100}{N} \cdot \sum_{i=1}^{N} \left(\frac{|x_{pre,i} - x_{\exp,i}|}{x_{\exp,i}} \right) \quad (4)$$

$$SSE = \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2 \quad (5)$$

where $x_{exp,i}$ stands for the experimental values and $x_{pre,i}$ is the predicted values calculated by the model, N and n are the number of observations and constants, respectively.

Yoon's method of global sensitivity (Equation (6)) was used to calculate the direct influence of the input parameters on the output variables, corresponding to the weighting coefficients within the ANN model [21]:

$$RI_{ij}(\%) = \frac{\sum_{k=0}^{n} (w_{ik} \cdot w_{kj})}{\sum_{i=0}^{m} \left| \sum_{k=0}^{n} (w_{ik} \cdot w_{kj}) \right|} \cdot 100\% \quad (6)$$

where w—denotes the weighting factor in the ANN model, i—input variable, j—output variable, k—hidden neuron, n—number of hidden neurons, m—number of inputs.

2.3. ANN Modeling

ANN are among the most researched areas of neurocomputing. A multilayer perceptron (MLP) is a neural network with hidden layers Figure 1. ANN can adapt its internal structure depending on the input data and the final goal of the function. The basic characteristics of ANN are the ability to learn independently, the ability to adapt the system to the available information and data processing, and to perform complex mathematical operations at high speed. The number of neurons and hidden layers in ANN can vary and is determined by the trial-and-error method [11]. Neural networks are categorized by their architecture, topology, and learning mode [22]. Neural networks take inputs, compute them, and convert them into outputs. This process is called the learning process of the network. The learning process of ANN can be supervised and unsupervised. In supervised learning mode, the model has access to output data for computations, while in unsupervised mode, there is no output data [23].

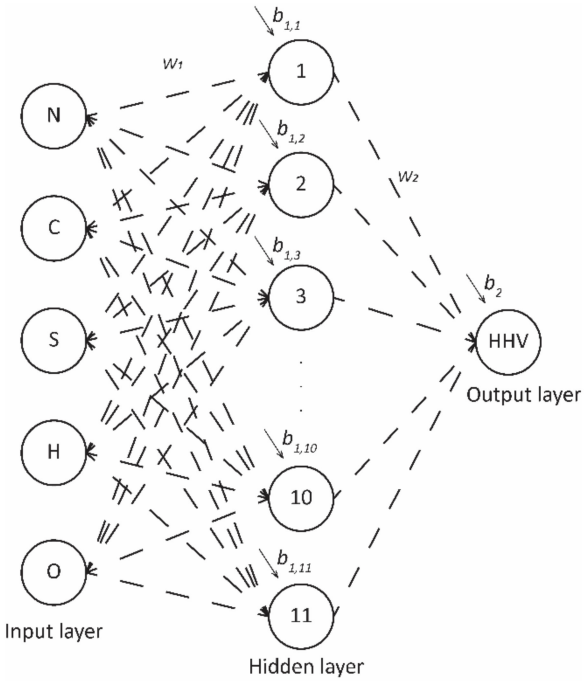

Figure 1. Structure of 5-11-1 ANN.

Models of ANN can provide a link between input and output data without using a complicated type of computational method. MLP ANN is recognized as the most effective type of ANN [9,24]. ANN is a mathematical structure developed from the motivation of the learning process in the human brain. ANN is a promising modeling technique for datasets with nonlinear relationships. Multilayer feedforward networks (MLP-ANN) consist of interdependent units (neurons). These neurons are arranged in the form of layers (input,

hidden, and output layers). The number of neurons and hidden layers varies and can be determined by the method of trial and error so that the model error is minimal [11].

The data used for ANN were collected from the literature and were randomly divided into sets for training (70%), testing (15%) and validation (15%). ANN model was trained 100,000 times with a random number of hidden layers (1–20), and duration of creating the model was 27 min. The model was created on a computer with an Intel i5 processor (12th Gen Intel(R) Core(TM), i5-12400F, 2.50 GHz) and 8 GB of RAM.

Different transfer functions and random values for weighting coefficients and bias were used. Training of the network data was set up during the ANN learning cycle to determine the number of neurons and adjust the weight coefficients in each neuron [25]. The biases and weight coefficients related to the hidden and the output layers of the model are represented in the matrices and vectors W_1 and B_1 and W_2 and B_2, respectively [26]. The neural network model can be represented in matrix notation: Equation for calculating the output data (Equation (7)) of the neural network [27]:

$$Y = f_1(W_2 \cdot f_2(W_1 \cdot X + B_1) + B_2) \quad (7)$$

where Y represents the output value, f_1 and f_2 represent the transfer function in the hidden and output layer, X represents the matrix of the input layer [28].

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was used for the calculations. The BFGS algorithm is one of the most effective algorithms for optimization and can be successfully used for the optimization of multivariate problems [29].

2.4. Regression Models

Table 1 presents the models of the proposed equations for the calculation of HHV biomass found in the literature [5,30,31]. The models are based on establishing relations between variables based on ultimate analysis and HHV output values.

Table 1. List of equations for calculating HHV.

Sr.no.	Proposed Equations from the Literature	References
1	$HHV = a + b \cdot C$	[31]
2	$HHV - a + b \cdot H$	[31]
3	$HHV - a + b \cdot O$	[31]
4	$HHV = a + b \cdot \frac{O}{C}$	[31]
5	$HHV = a + b \cdot \frac{H}{C}$	[31]
6	$HHV = a + b \cdot C + c \cdot H + d \cdot C^2 + e \cdot H^2$	[31]
7	$HHV = a + b \cdot \frac{O}{C} + c \cdot \frac{H}{C} + d \cdot \left(\frac{O}{C}\right)^2 + e \cdot \left(\frac{H}{C}\right)^2$	[31]
8	$HHV = a \cdot C - b$	[5]
9	$HHV = a + b \cdot C^2 + c \cdot C + d \cdot H + e \cdot C \cdot H + g \cdot N$	[30]
10	$HHV = a + b \cdot (C)^2$	[31]

3. Results

Table 2 shows the mean values of the variables of the ultimate analysis and HHV with standard deviation and Tukey's HSD test of miscanthus.

Table 2 shows the differences in the percentages of nitrogen, carbon, hydrogen, sulfur, and oxygen and is expressed as mean and standard deviation. The prefabricated statistical analysis shows that the observed values between the samples are not statistically significant (statistically significant at $p \leq 0.05$). Higher content of C and H components leads to a higher total value of HHV [32]. The sample MxG 6 has the highest average content of N (0.31%) and H (5.89%) and a higher value of HHV (18.45 MJ/kg). The sample MxG 14 has the lowest average percentage value of elements S (0.09%) and H (5.31%) and the lowest value of HHV (17.83 MJ/kg). The average values in the paper are: 0.22% N, 51.42% C, 0.13% S, 5.80% H, 42.42% O and 18.18 MJ/kg for HHV.

Table 2. Average values of nitrogen, carbon, sulfur, hydrogen, and oxygen of investigated biomass of miscanthus.

Sample	N	C	S	H	O	HHV
MxG1	0.24 ± 0.15 [a]	51.49 ± 0.58 [a]	0.11 ± 0.04 [a]	5.82 ± 0.32 [a]	42.33 ± 0.74 [a]	18.21 ± 0.54 [a]
MxG2	0.19 ± 0.13 [a]	51.3 ± 0.53 [a]	0.14 ± 0.06 [a]	5.85 ± 0.2 [a]	42.53 ± 0.42 [a]	18.16 ± 0.37 [a]
MxG3	0.23 ± 0.15 [a]	50.9 ± 0.96 [a]	0.14 ± 0.06 [a]	5.82 ± 0.2 [a]	42.91 ± 0.89 [a]	18.22 ± 0.58 [a]
MxG4	0.22 ± 0.13 [a]	51.75 ± 0.73 [a]	0.21 ± 0.3 [a]	5.83 ± 0.33 [a]	41.99 ± 0.69 [a]	18.24 ± 0.64 [a]
MxG5	0.2 ± 0.09 [a]	51.38 ± 0.8 [a]	0.14 ± 0.08 [a]	5.84 ± 0.33 [a]	42.44 ± 0.82 [a]	18.37 ± 0.48 [a]
MxG6	0.31 ± 0.21 [a]	51.53 ± 0.86 [a]	0.21 ± 0.2 [a]	5.89 ± 0.23 [a]	42.06 ± 0.88 [a]	18.45 ± 0.61 [a]
MxG7	0.2 ± 0.13 [a]	51.33 ± 0.93 [a]	0.12 ± 0.09 [a]	5.82 ± 0.33 [a]	42.54 ± 0.92 [a]	17.97 ± 0.73 [a]
MxG8	0.2 ± 0.11 [a]	51.65 ± 1.33 [a]	0.11 ± 0.06 [a]	5.88 ± 0.27 [a]	42.16 ± 1.18 [a]	18.23 ± 0.64 [a]
MxG9	0.21 ± 0.12 [a]	51.76 ± 0.77 [a]	0.13 ± 0.05 [a]	5.85 ± 0.35 [a]	42.05 ± 0.86 [a]	18.35 ± 0.32 [a]
MxG10	0.18 ± 0.1 [a]	51.48 ± 0.97 [a]	0.11 ± 0.05 [a]	5.83 ± 0.33 [a]	42.4 ± 0.85 [a]	18.1 ± 0.61 [a]
MxG11	0.22 ± 0.16 [a]	51.09 ± 1.14 [a]	0.11 ± 0.05 [a]	5.85 ± 0.27 [a]	42.74 ± 1.12 [a]	18.06 ± 0.33 [a]
MxG12	0.19 ± 0.11 [a]	51.6 ± 0.82 [a]	0.12 ± 0.06 [a]	5.86 ± 0.35 [a]	42.24 ± 0.97 [a]	18.51 ± 0.5 [a]
MxG13	0.27 ± 0.22 [a]	51.15 ± 0.8 [a]	0.15 ± 0.09 [a]	5.79 ± 0.36 [a]	42.64 ± 0.86 [a]	18.24 ± 0.6 [a]
MxG14	0.2 ± 0.14 [a]	51.53 ± 0.82 [a]	0.09 ± 0.03 [a]	5.31 ± 1.71 [a]	42.86 ± 2.12 [a]	17.83 ± 0.81 [a]
MxG15	0.25 ± 0.15 [a]	51.73 ± 0.99 [a]	0.12 ± 0.05 [a]	5.83 ± 0.38 [a]	42.08 ± 1.19 [a]	18.09 ± 0.4 [a]
MxG16	0.18 ± 0.12 [a]	51.11 ± 1.12 [a]	0.11 ± 0.05 [a]	5.83 ± 0.33 [a]	42.77 ± 1.15 [a]	18.02 ± 0.43 [a]

N—Nitrogen (%); C—Carbon (%); S—Sulfur (%); H—Hydrogen (%); O—Oxygen (%). The means in the same row (various samples), with different lowercase superscripts, are statistically different ($p \leq 0.05$), according to Tukey's HSD test.

The correlation analysis of the parameters of ultimate analysis and HHV was performed via Rstudio and related packages (corrplot).

The diagram of the correlation matrix shows the correlation coefficients between the variables. Positive values of the correlation coefficient are shown in blue, while negative values are shown in red. The intensity of the color in the circle is proportional to the correlation coefficient. In Figure 2, it can be observed that the elements O, S, and N are positively correlated with the value of HHV, while C and H are negative. It can be seen that variable S has the highest positive correlation coefficient, i.e., a significant influence on HHV, while variables N and O also have positively correlated values, but less influence on HHV. The variable H in the correlation graph shown has a negative correlation value on HHV. Based on Figure 2, HHV is best correlated with the concentrations of H, S, and N (when the blue color is shown, it is a positive correlation).

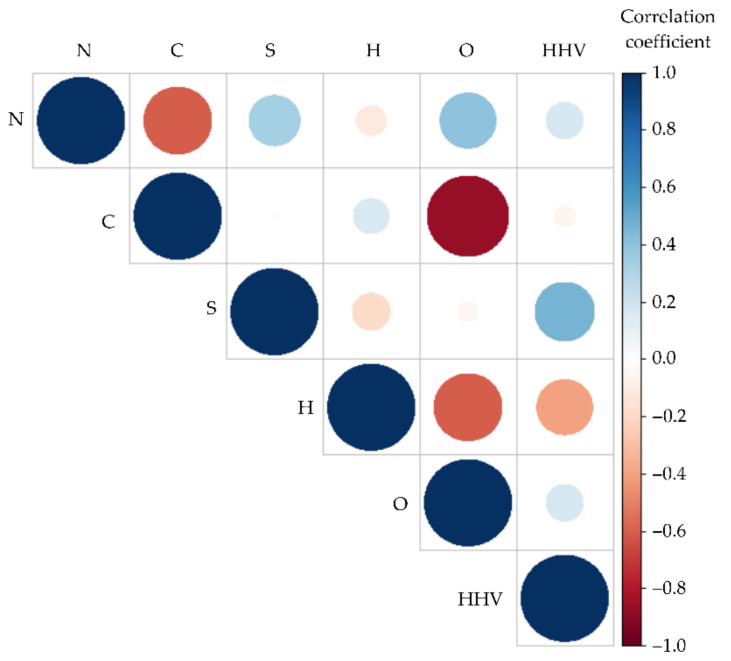

Figure 2. Correlation plot of observed values.

After determining the mean of all parameters, the correlations of the variables and their contribution were determined. The influence of the variables (N, C, S, H, O, and HHV) and the samples are combined graphically.

Principal Component Analysis (PCA) is used in the search for orthogonal directions of greatest dispersion of given data with the task of finding patterns in the distribution of individual data with respect to the original data defined in a space with multiple dimensions [33]. The analysis is also used to build predictive models, and it is easy to interpret the impact of individual variables on a given value. In Figure 3, the right side of the diagram shows sample 6, which is significant and has the highest values for HHV, H, N, and S. The upper part of the diagram shows samples 9, 8, 12, 15, 4 with the highest content of C. On the left side of the diagram are samples 14, 16, 11, 3, which have the highest content of O. According to the PCA analysis, the parameters N, S, H and HHV have the greatest influence on the data.

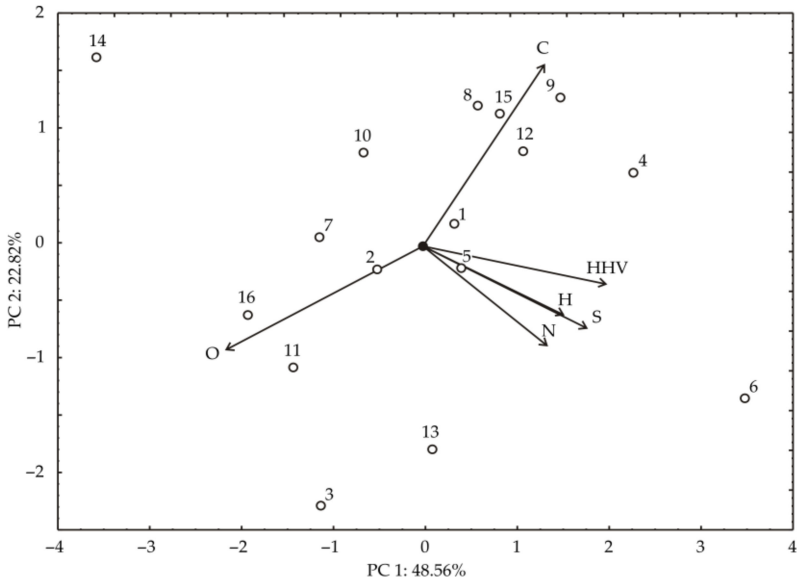

Figure 3. PCA of observed values.

4. Discussion

4.1. Prediction of HHV Using Developed Regression Models

Table 3 shows the calculated statistical test of "goodness of fit" for the proposed models to calculate the HHV value based on ultimate analysis.

Table 3. Statistical test goodness of fit (developed regression models).

Model	x^2	RMSE	MBE	MPE	SSE	R^2	Skewness	Kurtosis	SD	Variance
Model 1	0.31	0.01	0.01	854.24	59.66	0.00	−0.63	1.47	0.56	0.31
Model 2	0.19	0.01	0.01	578.45	35.56	0.40	−1.32	9.26	0.43	0.19
Model 3	0.30	0.01	0.01	810.35	57.95	0.03	−0.67	2.10	0.55	0.30
Model 4	0.31	0.01	0.01	830.08	58.92	0.02	−0.65	1.81	0.56	0.31
Model 5	0.19	0.01	0.01	623.98	36.77	0.38	−1.19	6.78	0.44	0.19
Model 6	0.22	0.01	0.01	582.07	42.91	0.36	−1.53	12.54	0.47	0.22
Model 7	0.19	0.01	0.01	585.55	35.77	0.40	−1.28	8.67	0.43	0.19
Model 8	0.31	0.01	0.01	854.24	59.66	0.00	−0.63	1.47	0.56	0.31
Model 9	0.17	0.01	0.01	526.96	31.57	0.47	−1.79	13.09	0.41	0.17
Model 10	0.31	0.01	0.01	853.32	59.64	0.00	−0.63	1.49	0.56	0.31
ANN	0.07	0.27	-0.03	1.10	13.74	0.77	0.53	2.29	0.27	0.07

x^2—reduced chi-square, RMSE—root mean square error, R^2—coefficient of determination, MBE—mean bias error and MPE—mean percentage error, ANN—artificial neural network.

The regression models offered from the literature use a different number of input variables of the ultimate analysis. Models 1,2,3,8,10 use one, while models 4,5,6 use two input variables. The highest number of input data is used by regression models 7 and 9, where the number of input variables is three. In contrast to the equations offered to calculate the HHV, ANN uses all five input variables of the ultimate analysis and shows the highest accuracy in prediction. The presented models in the calculations did not show sufficient accuracy and precision to be used as a reliable method for predicting HHV biomass of miscanthus. The coefficient of determination (R^2) was used as the most important statistical parameter to evaluate the suitability of the mathematical models, which was lowest for model 1, 8 and model 10 (R^2 = 0.00) and highest for model 9 (R^2 = 0.47) in the calculations for 10 different models. The reliability of the regression models and ANN is ensured by the parameters MPE, SSE and R^2, but other parameters (for most models) also show good performance. The calculated statistical parameter x^2 shows good performance in models 2,5,7 (0.19) and in model 9 (0.17). For the above-mentioned reason, it is necessary to consider several statistical parameters when evaluating performance of the model.

4.2. ANN Model

In developing the model ANN, the input variables (N, C, S, H, and O) and the output value (HHV) had to be determined. The weights and biases were determined randomly by looking for values that would make the model accurate enough to predict the output.

The ANN model developed for the prediction of HHV showed a good ability to generalize data and predict. The model showed the best performance with 11 neurons in the hidden layer within the network, where a high R^2 value (0.77 overall) and an overall low sum of squares value (SOS) were achieved during the training cycle (Table 4).

Table 4. Weights and biases of input and output layer.

Input Layer						Output Layer	
Weight					Bias	Weight	Bias
N	C	S	H	O		HHV	
−1.74	10.34	−30.08	−7.41	1.90	1.62	−1.76	2.06
−0.28	3.37	1.69	2.99	−4.61	−0.37	1.18	
2.58	−0.83	−5.02	−0.40	−0.37	−1.56	0.20	
4.23	−0.73	−6.78	−0.79	0.10	−1.40	−0.31	
10.55	−3.52	−15.48	12.93	−0.96	−2.58	−1.91	
1.08	1.54	1.49	−2.03	−3.48	−1.57	−0.44	
3.56	−2.75	−8.74	1.54	1.03	−2.09	−1.30	
−3.67	1.30	2.92	2.02	−2.74	−0.87	0.48	
−2.72	−0.49	6.47	−0.49	0.34	0.75	−0.60	
2.25	2.01	6.90	−5.15	1.20	3.20	0.47	
−1.14	1.93	2.53	0.98	−1.49	0.71	−1.56	

MLP-ANN (Multi-layer perceptron Artificial neural network) is one of the forms of ANN that are mostly used in applications for solving nonlinear equations [34]. Table 4 shows the weight coefficients and biases of the developed MLP-ANN network model. It can be seen that the best results were obtained with a hidden layer with a number of 11 hidden neurons, where the experimental values of HHV best match the values of HHV calculated with the ANN model.

Table 5 shows the training, test and validation performance of the model ANN, expressed by the coefficient of correlation and by the training (0.042), test (0.026), and validation (0.021) error of the model. Table 3 shows the results of the statistical test indicating the deviations between the observed values and the expected values. The values shown indicate the ability of the algorithm to predict according to the given model data.

Table 5. Summary of ANN.

Net. Name	Train. Perf.	Test Perf.	Valid. Perf.	Train. Error	Test Error	Valid. Error	Train. Algorithm	Error Function	Hidden Activation	Output Activation
MLP 5-11-1	0.861	0.902	0.951	0.042	0.026	0.021	BFGS 71	SOS	Tanh	Logistic

Train.—Training; Perf.—Performance; Valid.—Validation.

Therefore, the ANN structural model MLP 5-11-1 proved to be sufficiently accurate to predict HHV based on the N, C, S, H, and O contents. The training (0.861), test (0.902) and validation (0.951) performance values shows that the model is able to predict values almost equal to the measured values.

Scatterplot is one of the most common visualization techniques, and displays and displays the behavior of the entered data [35,36]. Figure 4. shows the data of the predicted HHV versus the target HHV, which largely shows the overlap.

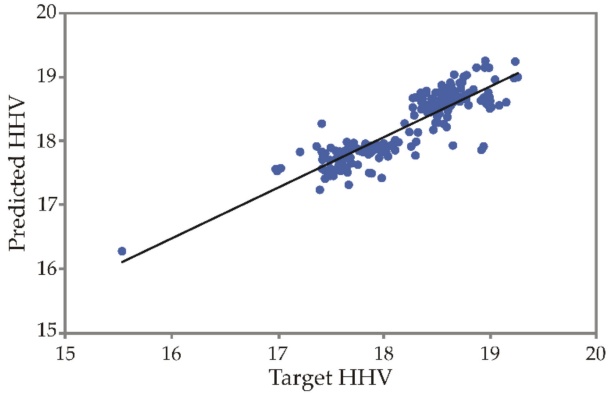

Figure 4. Predicted HHV vs. target HHV.

The calculated parameters comprising the statistical test "goodness of fit" are shown in Table 3. The reported values of x^2 (0.07), RMSE (0.27), MBE (−0.03), MPE (1.10), SSE (13.74), and R^2 (0.77). The residual analysis also yielded the parameters skewness (0.534), kurtosis (2.293), standard deviation (0.269), and variance (0.072). Conducted analysis shows that the model has good predictive accuracy.

The range in which the relevance factor is determined is between −1 and +1. The increase of HHV is mainly influenced by the increase of the parameter S. The influence of input variables was studied according to Yoon's interpretation method for parameters N, C, S, H, O. In Figure 5 is the influence of variables N (−20.64%), C (3.13%), S (68.89%), H (−3.07%), and O (−4.27%) on the target value of HHV. In Figure 5, it can be seen that the parameters C and S have positive values of relative importance for the variable HHV, while the values of N, H, and O have a negative influence and are not factors of relative importance in determining the value of HHV. Looking at the calculation of the input data carried out according to Yoon's method of interpretation based on the ultimate analysis the positive variable S have the greatest influence on the determination of the output values of HHV.

The predictive regression models offered in the literature are used as nonlinear models to predict HHV biomass of miscanthus. As shown in the paper, the use of the predictive models does not provide suitability and sufficient accuracy in determining the HHV miscanthus with respect to the input parameters. Using ANN as a nonlinear model to determine the HHV value provides a more convenient way of prediction and provides more accurate weighting coefficients and biases, which are the basis for establishing relations between input parameters and output data.

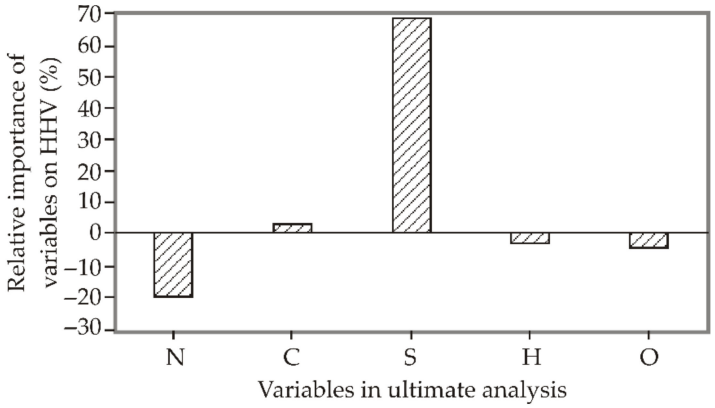

Figure 5. Relative importance of variables on HHV.

5. Conclusions

The use of ANN models to predict the energy properties of biomass has been increasingly explored recently. The main point of the study is the creation of an improved model (in the form of ANN) compared to existing literature regression models, as evidenced by a higher R^2 value. The calculations performed according to the proposed non-linear mathematical models are not suitable enough to predict the HHV biomass of miscanthus ($R^2 \leq 0.47$). Incorporating available data from the ultimate analysis of miscanthus the developed neural network model showed high accuracy in predicting the higher heating value (overall $R^2 = 0.77$). The factors N, C, S, H, and O influence the value of HHV. In the developed model, the increase in HHV is mainly influenced by the increase in the values of the parameter S. Although these models are not yet widely used as mathematical models for prediction (especially for variables that have nonlinear relationships), they offer the possibility of obtaining the desired result with less time, lower cost, and satisfactory accuracy, which can replace existing empirical methods. The developed model will be able to make more accurate predictions as more input data is collected. Future plans are to expand the database (literature sources and experimental data) and the development of new models such as Random Forest Regression and Support Vector Machine.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math10203732/s1, Table S1. Result of ultimate analysis and HHV of miscanthus biomass.

Author Contributions: Conceptualization, I.B., L.P., N.B. and N.V.; methodology, I.B., A.P. and J.Š.; Software, I.B.; validation, L.P., N.B. and N.V.; draft preparation, I.B., L.P., A.P. and J.Š.; writing—review and editing, I.B., L.P., N.B., A.P., J.Š. and N.V.; visualization, I.B. and L.P.; supervision, L.P., N.B. and N.V.; project administration, N.V.; funding acquisition, N.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Croatian Science Foundation, under project No. IP-2018-01-7472 "Sludge management via energy crops production" and within the project "Young Researchers' Career Development Project—Training of Doctoral Students", co-financed by the European Union, under the OP "Efficient Human Resources 2014–2020" from the ESF funds.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The publication was supported by the Open Access Publication Fund of the University of Zagreb Faculty of Agriculture.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Koçar, G.; Civaş, N. An overview of biofuels from energy crops: Current status and future prospects. *Renew. Sustain. Energy Rev.* **2013**, *28*, 900–916. [CrossRef]
2. Long, H.; Li, X.; Wang, H.; Jia, J. resources and their bioenergy potential estimation: A review. *Renew. Sustain. Energy Rev.* **2013**, *26*, 344–352. [CrossRef]
3. Baxter, X.C.; Darvell, L.I.; Jones, J.M.; Barraclough, T.; Yates, N.E.; Shield, I. Miscanthus combustion properties and variations with Miscanthus agronomy. *Fuel* **2014**, *117*, 851–869. [CrossRef]
4. Yi, L.; Feng, J.; Qin, Y.H.; Li, W.Y. Prediction of elemental composition of coal using proximate analysis. *Fuel* **2017**, *193*, 315–321. [CrossRef]
5. Sheng, C.; Azevedo, J.L.T. Estimating the higher heating value of biomass fuels from basic analysis data. *Biomass Bioenergy* **2005**, *28*, 499–507. [CrossRef]
6. Dai, Z.; Chen, Z.; Selmi, A.; Jermsittiparsert, K.; Denić, N.M.; Nešić, Z. Machine learning prediction of higher heating value of biomass. *Biomass Convers. Biorefin.* **2021**, 1–9. [CrossRef]
7. Giwa, S.O.; Adekomaya, S.O.; Adama, K.O.; Mukaila, M.O. Prediction of selected biodiesel fuel properties using artificial neural network. *Front. Energy* **2015**, *9*, 433–445. [CrossRef]
8. Dashti, A.; Noushabadi, A.S.; Raji, M.; Razmi, A.; Ceylan, S.; Mohammadi, A.H. Estimation of biomass higher heating value (HHV) based on the proximate analysis: Smart modeling and correlation. *Fuel* **2019**, *257*, 115931. [CrossRef]
9. Pattanayak, S.; Loha, C.; Hauchhum, L.; Sailo, L. Application of MLP-ANN models for estimating the higher heating value of bamboo biomass. *Biomass Convers. Biorefin.* **2021**, *11*, 2499–2508. [CrossRef]
10. Vardiambasis, I.O.; Kapetanakis, T.N.; Nikolopoulos, C.D.; Trang, T.K.; Tsubota, T.; Keyikoglu, R.; Khataee, A.; Kalderis, D. Hydrochars as emerging biofuels: Recent advances and application of artificial neural networks for the prediction of heating values. *Energies* **2020**, *13*, 4572. [CrossRef]
11. Özveren, U. An artificial intelligence approach to predict gross heating value of lignocellulosic fuels. *J. Energy Inst.* **2017**, *90*, 397–407. [CrossRef]
12. Olatunji, O.O.; Akinlabi, S.; Madushele, N.; Adedeji, P.A.; Felix, I. Multilayer perceptron artificial neural network for the prediction of heating value of municipal solid waste. *AIMS Energy* **2019**, *7*, 944–956. [CrossRef]
13. Kartal, F.; Özveren, U. A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. *Energy* **2020**, *209*, 118457. [CrossRef]
14. Ahmed, A.A.M. Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). *J. King Saud Univ. Eng. Sci.* **2017**, *29*, 151–158. [CrossRef]
15. Alizadeh, M.J.; Kavianpour, M.R. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Mar. Pollut. Bull.* **2015**, *98*, 171–178. [CrossRef]
16. 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]
17. Voća, N.; Leto, J.; Karažija, T.; Bilandžija, N.; Peter, A.; Kutnjak, H.; Šurić, J.; Poljak, M. Energy properties and biomass yield of miscanthus x giganteus fertilized by municipal sewage sludge. *Molecules* **2021**, *26*, 4371. [CrossRef] [PubMed]
18. Meehan, P.G.; Finnan, J.M.; Donnell, K.P.M. The effect of harvest date and harvest method on the combustion characteristics of Miscanthus X giganteus. *GCB Bioenergy* **2013**, *5*, 487–496. [CrossRef]
19. Khalil, S.A.; Shaffie, A.M. A comparative study of total, direct and diffuse solar irradiance by using different models on horizontal and inclined surfaces for Cairo, Egypt. *Renew. Sustain. Energy Rev.* **2013**, *27*, 853–863. [CrossRef]
20. Arsenović, M.; Pezo, L.; Stanković, S.; Radojević, Z. Factor space differentiation of brick clays according to mineral content: Prediction of final brick product quality. *Appl. Clay Sci.* **2015**, *115*, 108–114. [CrossRef]
21. Yoon, Y.; Swales, G.; Margavio, T.M. A Comparison of Discriminant Analysis versus Artificial Neural Networks. *J. Oper. Res. Soc.* **1993**, *44*, 51–60. [CrossRef]
22. Mesroghli, S.; Jorjani, E.; Chehreh Chelgani, S. Estimation of gross calorific value based on coal analysis using regression and artificial neural networks. *Int. J. Coal Geol.* **2009**, *79*, 49–54. [CrossRef]
23. Cinar, A.C. Training Feed-Forward Multi-Layer Perceptron Artificial Neural Networks with a Tree-Seed Algorithm. *Arab. J. Sci. Eng.* **2020**, *45*, 10915–10938. [CrossRef]
24. Darvishan, A.; Bakhshi, H.; Madadkhani, M.; Mir, M.; Bemani, A. Application of MLP-ANN as a novel predictive method for prediction of the higher heating value of biomass in terms of ultimate analysis. *Energy Source Part A* **2018**, *40*, 2960–2966. [CrossRef]
25. Grieu, S.; Faugeroux, O.; Traoré, A.; Claudet, B.; Bodnar, J.L. Artificial intelligence tools and inverse methods for estimating the thermal diffusivity of building materials. *Energy Build.* **2011**, *43*, 543–554. [CrossRef]
26. Voca, N.; Pezo, L.; Peter, A.; Suput, D.; Loncar, B.; Kricka, T. Modelling of corn kernel pre-treatment, drying and processing for ethanol production using artificial neural networks. *Ind. Crop. Prod.* **2021**, *162*, 113293. [CrossRef]
27. Pezo, L.L.; Ćurčić, B.L.; Filipović, V.S.; Nićetin, M.R.; Koprivica, G.B.; Mišljenović, N.M.; Lević, L.B. Artificial neural network model of pork meat cubes osmotic dehydratation. *Hem. Ind.* **2013**, *67*, 465–475. [CrossRef]
28. Kollo, T.; von Rosen, D. *Advanced Multivariate Statistics with Matrices*; Springer: Dordrecht, The Netherlands, 2005.

29. Garijo, D. A Bernstein Broyden–Fletcher–Goldfarb–Shanno collocation method to solve non-linear beam models. *Int. J. Nonlinear Mech.* **2021**, *131*, 103672. [CrossRef]
30. Nhuchhen, D.R.; Afzal, M.T. HHV predicting correlations for torrefied biomass using proximate and ultimate analyses. *Bioengineering* **2017**, *4*, 7. [CrossRef]
31. Callejón-Ferre, A.J.; Velázquez-Martí, B.; López-Martínez, J.A.; Manzano-Agugliaro, F. Greenhouse crop residues: Energy potential and models for the prediction of their higher heating value. *Renew. Sustain. Energy Rev.* **2011**, *15*, 948–955. [CrossRef]
32. Bilandžija, N.; Voća, N.; Leto, J.; Jurišić, V.; Grubor, M.; Matin, A.; Krička, T. Yield and Biomass Composition of Miscanthus x Giganteus in the Mountain Area of Croatia. In *Transactions of FAMENA*; University of Zagreb: Zagreb, Croatia, 2018; Volume 42, pp. 51–60.
33. Rutledge, D.N. Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis. *J. Anal. Test.* **2018**, *2*, 235–248. [CrossRef]
34. Agahian, S.; Akan, T. Battle royale optimizer for training multi-layer perceptron. *Evol. Syst.* **2021**, 123456789. [CrossRef]
35. Geladi, P.; Manley, M.; Lestander, T. Scatter plotting in multivariate data analysis. *J. Chemometr.* **2003**, *17*, 503–511. [CrossRef]
36. Keim, D.A.; Hao, M.C.; Dayal, U.; Janetzko, H.; Bak, P. Generalized scatter plots. *Inform. Visual.* **2010**, *9*, 301–311. [CrossRef]

energies

MDPI

Article

Energy Potentials of Agricultural Biomass and the Possibility of Modelling Using RFR and SVM Models

Ivan Brandić [1], Alan Antonović [2,*], Lato Pezo [3], Božidar Matin [2], Tajana Krička [1], Vanja Jurišić [1], Karlo Špelić [1], Mislav Kontek [1], Juraj Kukuruzović [1], Mateja Grubor [1] and Ana Matin [1]

[1] Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia
[2] Faculty of Forestry and Wood Technology, University of Zagreb, Svetošimunska Cesta 23, 10000 Zagreb, Croatia
[3] Institute of General and Physical Chemistry, University of Belgrade, Studentski trg 12/V, 11000 Belgrade, Serbia
* Correspondence: alan.antonovic@gmail.com

Abstract: Agricultural biomass is one of the most important renewable energy sources. As a byproduct of corn, soybean and sunflower production, large amounts of biomass are produced that can be used as an energy source through conversion. In order to assess the quality and the possibility of the use of biomass, its composition and calorific value must be determined. The use of nonlinear models allows for an easier estimation of the energy properties of biomass concerning certain input and output parameters. In this paper, RFR (Random Forest Regression) and SVM (Support Vector Machine) models were developed to determine their capabilities in estimating the *HHV* (higher heating value) of biomass based on input parameters of ultimate analysis. The developed models showed good performance in terms of *HHV* estimation, confirmed by the coefficient of determination for the RFR (R^2 = 0.79) and SVM (R^2 = 0.93) models. The developed models have shown promising results in accurately predicting the *HHV* of biomass from various sources. The use of these algorithms for biomass energy prediction has the potential for further development.

Keywords: agricultural biomass; higher heating value; machine learning; estimation; energy potential

Citation: Brandić, I.; Antonović, A.; Pezo, L.; Matin, B.; Krička, T.; Jurišić, V.; Špelić, K.; Kontek, M.; Kukuruzović, J.; Grubor, M.; et al. Energy Potentials of Agricultural Biomass and the Possibility of Modelling Using RFR and SVM Models. *Energies* **2023**, *16*, 690. https://doi.org/10.3390/en16020690

Academic Editor: David Herak

Received: 30 November 2022
Revised: 3 January 2023
Accepted: 4 January 2023
Published: 6 January 2023

1. Introduction

With increasing population growth and negative climate change trends, there is a need to create sustainable systems of energy production and bioeconomy [1]. The use of green technologies, i.e., biofuels, is among the most effective ways for reducing greenhouse gasses that directly affect global warming. Biomass is a renewable energy source and has high potential in energy production [2]. Large amounts of lignocellulosic biomass in the world allow its utilization and conversion into an alternative fuel source [3]. Calorific value is the most important parameter in assessing the possibility of using biomass as a fuel [4]. Higher heating value (*HHV*) is an important factor in describing the quality of fuel and the possibility of using biomass for energy conversion. The use of biomass as an alternative fuel source is considered environmentally and economically viable and offers the possibility of replacing current fossil fuels. One of the fundamental characteristics of biomass is its chemical composition, i.e., the ultimate analysis, which includes a percentage of carbon (C), hydrogen (H), nitrogen (N), sulfur (S) and oxygen (O) [5]. Large amounts of biomass are produced as a by-product in agricultural production and most of it is unused. Biomass from agricultural products such as corn residues, straw and sunflower stalks represents an easily available energy source [6]. The authors also state that corn cobs are an economically viable and environmentally accessible source of biomass that can be used for agricultural energy production. Soybean is a legume that has economic value in terms of seed, straw and biomass yield. In the study by Krisnawati and Adie [7], the soybean was mentioned for its great prospect as a biomass energy source and is especially recognized as a biomass

source for fuel production due to its favorable energy properties [8]. The sunflower is an extremely important feedstock grown on a large area that can produce a significant amount of biomass per hectare [9]. The authors also note that biomass from sunflower cultivation is a suitable feedstock for the production of second-generation biofuels.

In order to assess the possibility of using biomass as a fuel, its composition must be determined by various laboratory analyses. The elements of organic matter in biomass are carbon (C), hydrogen (H), oxygen (O), nitrogen (N) and sulfur (S) which can be measured by the ultimate analysis [10,11]. Higher heating value (*HHV*) is a specific characteristic of biomass that can be used to evaluate the possibility of using biomass for energy through conversion (in the form of heat, fuel, etc.) [12]. In the study of Roman et al. [13], the investigation aimed to evaluate the mechanical and energetic properties of shredded pine forest residues during the briquetting process. The shredded fragments of the forest residues were compacted by the principal stresses with a determination of the energy value consumed during the briquetting process.

In the conducted research by Ibikunle et al. [14], a model was developed to predict the *HHV* of municipal waste using the input data of the ultimate analysis. The models used to estimate *HHV* were either in linear or quadratic form. The best-fitting model in the study showed good performance in terms of prediction and the coefficient of determination (R^2 = 0.97) was used as the main evaluation parameter. SVM models for estimating *HHV* are applicable to different types of biomass, thus providing a good solution to the problem of estimating *HHV* [15]. A machine learning (ML) model created to estimate *HHV* of biomass was based on the input parameters of proximate analysis data (percentage of fixed carbon, ash and volatile matter). An extreme learning machine (ELM) method proved to be very practical in estimating the *HHV*, as evidenced by the high coefficient of determination for the input parameters of fixed carbon (0.972), volatiles (0.989) and ash (0.968) [12]. Bychkov et al. [16] investigated developed models for predicting the *HHV* of plant biomass from ultimate analysis data. In the conducted study, authors used 150 models of which 8 were selected for model testing, with 3 showing good performance in estimating the *HHV* of biomass with small deviations from actual values.

The aim of this paper is to determine the possibility of the mathematical modelling of *HHV* for corn, soybean and sunflower biomass using machine learning techniques, such as the support vector machine (SVM) and random forest regression (RFR) models for regression, where all data are divided into two parts, such that one part is for training and one part for testing the model in a ratio of 70–30(%). The statistical test "goodness of fit" is used as the main evaluation method for the model performance. A comparison of the SVM and RFR models will show which model is better for predicting *HHV* biomass based on the input parameters of the ultimate analysis.

2. Materials and Methods

2.1. Data Collection

The data for creating the SVM and RFR models were obtained from the literature [17–28]. The data of the ultimate analysis and *HHV* for 51 biomass samples, including 27 samples for corn, 15 for soybean and 19 for sunflower biomass, including the biomass and biomass products, were collected for model development.

2.2. Nonlinear Modelling

After the collection dataset, the data were divided into a part for training and a part for testing the model, in the ratio of 70% and 30%, for SVM and RFR models. Statistical analysis was performed using TIBCO STATISTICA 13.3.0 software (StatSoft TIBCO Software Inc., Palo Alto, CA, USA). The analyzed data were presented in the form of mean and standard deviation. Analysis of variance (ANOVA) and Tukey's HSD (honestly significant difference) test were used to compare the samples to show the variability of the observed data.

SVM, as a supervised learning model, is based on statistical theories and can be used for clustering and regression [15]. The SVM model created is based on the input data

from the ultimate analysis, as type 1 regression models with Kelner type (Radial Basis Function—RDF) and 9 support vectors. The total number of iterations of the SVM model is 10,000.

As nonlinear models, RFRs are suitable for predictions with medium and large data sets [28]. The RFR models were also based on the input data from the final analysis with the data split into 70% for training and 30% for testing the model. The models were built using 10,000 random trees. For each internal node within the decision trees, entropy is calculated using the formula (Equation (1)) [29]:

$$E = -\sum_{i=1}^{c} p_i \times \log(p_i) \tag{1}$$

where c represents the number of unique classes and p_i prior probability of each given class. In order to solve the nonlinear problem, the Kelner function is used to map the input vectors into a multidimensional vector proctor that is used to find the hyperplane [30]. The equation used to create the SVM model is shown in the following equation (Equation (2)) [31]:

$$\gamma = \omega^T \theta(\chi) + b \tag{2}$$

where γ is target value, w is the weight vector, b is the threshold, θ is the nonlinear function of the model and χ is input vector.

2.3. Models Verification

To show the performance of the developed SVM and RFR models with respect to the input variables of the ultimate analysis, the following statistical parameters must be calculated: x^2 (reduced chi-square) from Equation (3), RMSE (root mean square error) from Equation (4), MBE (mean bias error) from Equation (5), MPE (mean percentage error) from Equation (6), and SSE (sum of squared estimate error) from Equation (7). "Goodness of fit" is calculated using the above statistical parameters to find the model with the lowest error, and they are represented by the following equations [32]:

$$x^2 = \frac{\sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2}{N - n} \tag{3}$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2 \right]^{1/2} \tag{4}$$

$$MBE = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i}) \tag{5}$$

$$MPE = \frac{100}{N} \cdot \sum_{i=1}^{N} \left(\frac{\left| x_{pre,i} - x_{\exp,i} \right|}{x_{\exp,i}} \right) \tag{6}$$

$$SSE = \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2 \tag{7}$$

3. Results

Table 1 shows the average values of the ultimate analysis and the *HHV* of the observed biomass.

Table 1 shows the mean and standard deviation of the variables of the ultimate analysis of corn, soybean and sunflower biomass. Sunflower biomass has the highest value for C (average 54.16%), H (average 6.51%), N (2.22%) and *HHV* (average 22.46 MJ/kg), and the lowest value for S (average 0.06%) and O (35.62%). Higher proportions of C and H influenced the increase in the *HHV* value [33]. Soybean biomass has the lowest average

value of C (47.37%), H (4.53%) and *HHV* (17.93 MJ/kg), while it has the highest average value of S (0.22%). The corn biomass sample has an average value of C (45.45%), H (5.29%), N (0.70%), S (0.08%), O (38.11%) and *HHV* (19.42 MJ/kg). All observed samples showed a significant difference. Thus, the variables C and S are statistically significant at a significance level of $\alpha = 0.05$, while the observed variables H, N, O and *HHV* are statistically significant at a significance level of $\alpha = 0.01$.

Table 1. Ultimate analysis of corn, soybean and sunflower biomass.

Sample	C (%)	H (%)	N (%)	S (%)	O (%)	*HHV* (MJ/kg)
Corn biomass	48.45 ± 7.3^{a}	5.29 ± 1.11^{b}	0.70 ± 0.50^{a}	0.08 ± 0.09^{a}	38.11 ± 11.22^{a}	19.42 ± 2.72^{a}
Soybean biomass	47.37 ± 3.48^{a}	4.53 ± 0.73^{a}	1.38 ± 2.32^{b}	0.22 ± 0.3^{b}	46.5 ± 6.83^{b}	17.93 ± 1.04^{a}
Sunflower biomass	54.16 ± 8.9^{b}	6.51 ± 0.9^{c}	2.22 ± 1.64^{c}	0.06 ± 0.13^{a}	35.62 ± 9.25^{a}	22.46 ± 4.41^{b}
Significance	**	*	*	**	*	*
Minimum	47.37	4.53	0.70	0.06	35.62	17.93
Maximum	54.16	6.51	2.22	0.22	46.50	22.46
Average	49.99	5.44	1.44	0.12	40.08	19.94

C—concentration of carbon; H—concentration of hydrogen; N—concentration of nitrogen; S—concentration of sulfur; O—concentration of oxygen; *HHV*—higher heating value. The means in the same column (various samples) with different lowercase superscripts are statistically different ($p < 0.05$), according to Tukey's HSD test. Statistical significance; * $p \leq 0.01$; ** $p \leq 0.05$.

Figure 1 shows the correlation diagram of the observed variables of the ultimate analysis and *HHV* of the average values of corn, soybean and sunflower biomass. The correlation of the observed values is shown in the range −1 to 1, which corresponds to the color intensity. Variables C (r = 0.98), H (r = 0.99) and N (r = 0.70) are positively correlated with the *HHV* value, while variables S (r = −0.83) and O (r = −0.88) are negatively correlated with the *HHV* value.

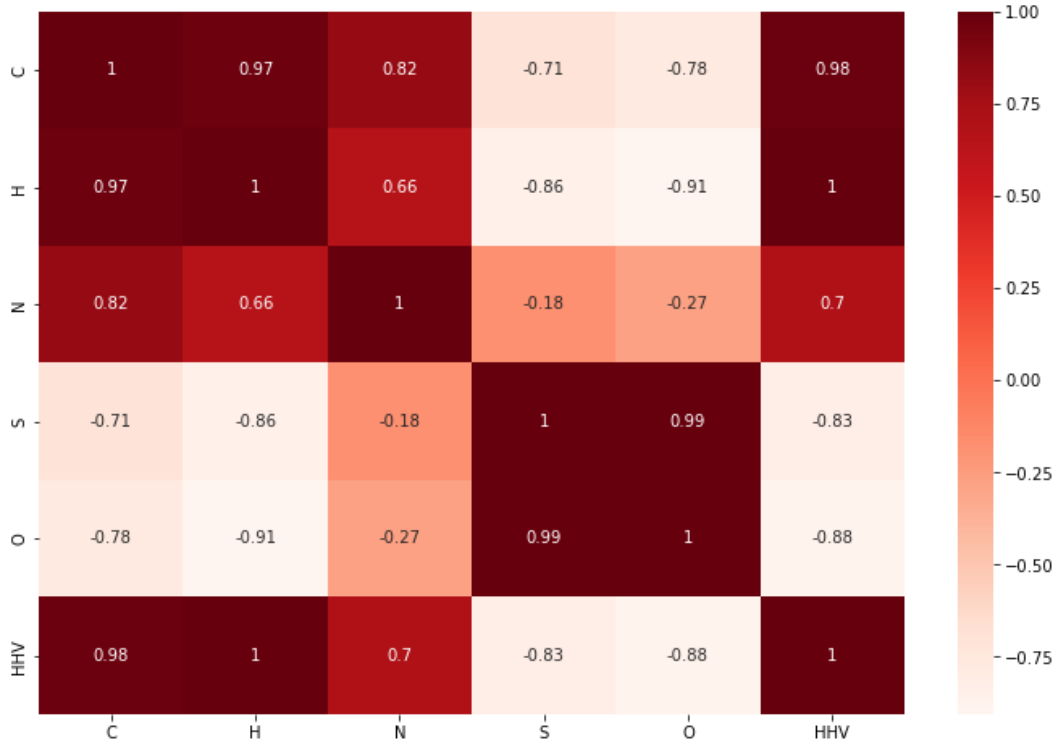

Figure 1. Correlation plot of observed variables of ultimate analysis and *HHV*.

Figure 2 shows the PCA of the observed biomass samples in relation to the variables of the ultimate analysis and the *HHV*. The PCA method is used to simplify a complex array of

data into more understandable groups for presentation by grouping all data with minimal losses into meaningful units [34,35]. The sunflower biomass group has the highest value for H, C, N and *HHV*, while the soybean biomass group has the lowest value for C, H and *HHV*. The observed corn biomass group has the lowest value for the variable N.

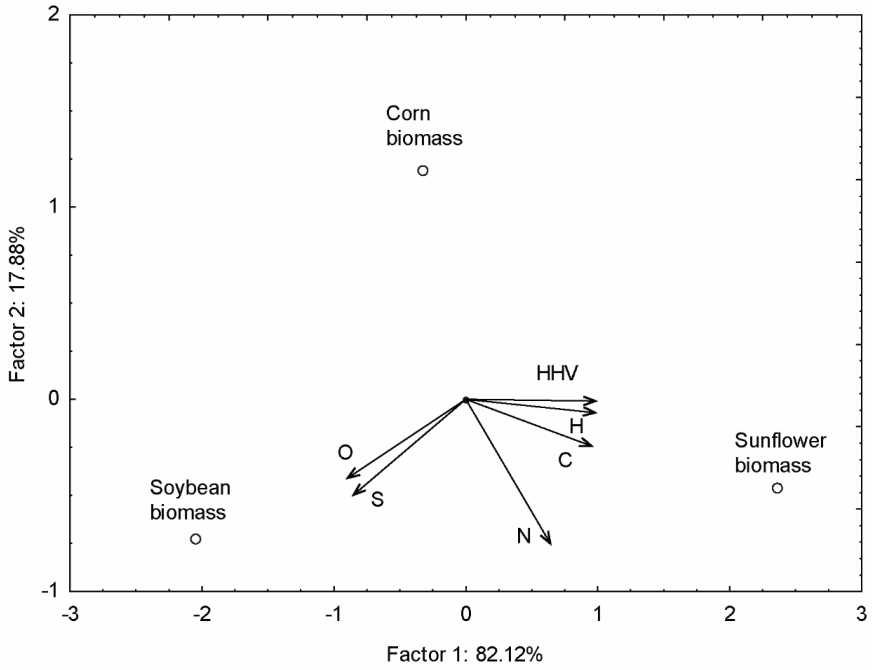

Figure 2. Principal component analysis (PCA) of observed corn, soybean and sunflower biomass.

4. Discussion

4.1. Support Vector Machine (SVM)

SVM models can be used for prediction in the form of regression due to their ability to generalize to different sample sizes and the possibility of nonlinear modeling [36]. The standard form of SVM belongs to the supervised form of learning and offers numerous advantages in terms of optimization and solution finding in nonlinear modeling [30]. Table 2 shows the values of the vector, weighting coefficients and decision constant with respect to the input variables in the SVM model.

Table 2. Vector values of the developed SVM model.

Vector No.	Weights	Support Vector					Decision Constant
		C	H	N	S	O	
1	9.00	0.22	0.63	0.07	0.07	0.73	−0.09
2	−7.34	0.20	0.68	0.07	0.12	0.70	
3	−0.21	0.95	0.00	0.21	0.02	0.00	
4	−1.01	0.35	0.49	0.00	0.32	0.96	
5	0.34	0.25	0.71	0.11	0.47	0.83	
6	−9.00	0.28	0.67	0.17	0.09	0.72	
7	6.14	0.51	0.70	0.16	0.00	0.64	
8	−1.95	0.77	0.96	0.57	0.00	0.36	
9	4.02	1.00	0.97	0.59	0.00	0.20	

Table 2 presents vector values of the developed SVM model.

4.2. Random Forest Regression (RFR)

Random Forest Regression (RFR) can easily adapt to nonlinear relationships between data and shows better predictive ability than linear regression models. RFRs are considered a reliable tool for predicting performance [29,37].

Figure 3 shows the importance of the predictor in the value 0 to 1 on the output value of *HHV* in the RFR model. In modeling, the highest predictor value is for O (1.00), followed by C (0.99), N (0.95), S (0.78) and H (0.57). With regard to the presented Figure 3, it can be concluded that the input parameters O, C and N have the greatest influence on the output value of *HHV*.

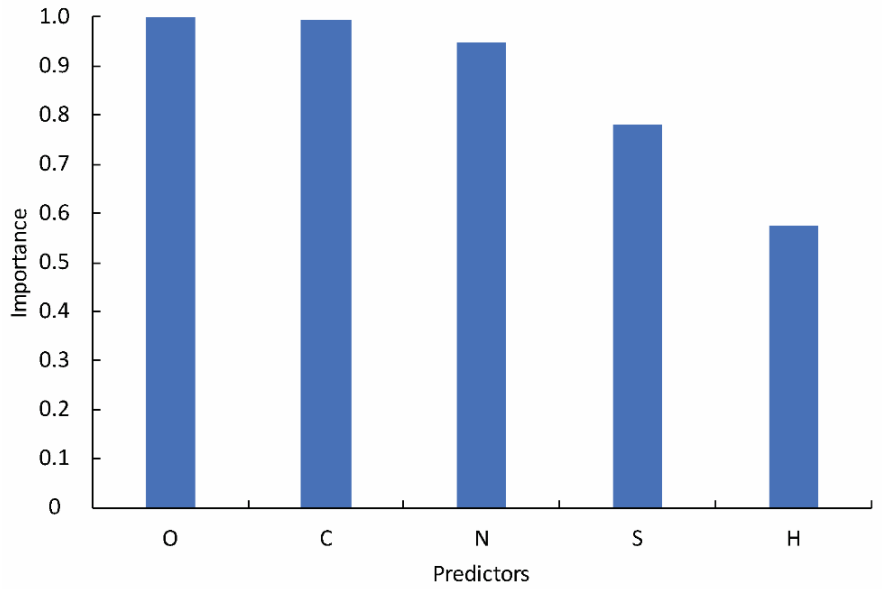

Figure 3. Importance of predictors (ultimate analysis) on *HHV* (RFR model).

4.3. Goodness of Fit

Table 3 shows the statistical analysis, "goodness of fit", which presents the ability of the developed models to predict the *HHV* values.

Table 3. "Goodness of fit" for developed SVM and RFR models.

								Residual Analysis			
Model	**x^2**	**RMSE**	**MBE**	**MPE**	**SSE**	**AARD**	**R^2**	**Skewness**	**Kurtosis**	**SD**	**Variance**
SVM model	0.82	0.90	−0.03	3.07	49.28	44.12	0.93	−3.04	14.32	0.91	0.82
RFR model	5.99	2.43	−0.01	8.32	359.53	103.30	0.79	0.94	2.08	2.45	5.99

Table 3 shows the statistical test, "Goodness of fit", which shows the performance of the developed SVM and RFR models. The calculated values of x^2 (0.82), RMSE (0.90), MBE (−0.03), SSE (49.28), AARD (44.12) and R^2 (0.93), and the residual analysis skewness (−3.04), kurtosis (14.32), SD (0.91) and Var (0.82) show the low level of error of the SVM model. The values of x^2 (5.99), RMSE (2.43), MBE (−0.01), MPE (8.32), SSE (359.53), AARD (103.30), and R^2 (0.79) were calculated for the RFR model in the table. The skewness (0.94), kurtosis (2.08), SD (2.45) and variance (5.99) parameters were determined by the residual analysis. Both developed models showed satisfactory performance in modeling the *HHV* values. Considering that, R^2 is used as the main indicator of the model's ability for estimation.

The scatterplot visualization technique was used to analyze correlations of variables on the x and y axes and to detect associations and anomalies in a multidimensional dataset [38]. Figure 4 shows the parity plot of the predicted and targeted *HHV* of the developed SVM

and RFR models for training and test data. Both models show an extensive overlap of the data, with a coefficient of determination of R^2 = 0.79 for the RFR model, while the overlap values in the SVM model have a higher coefficient of determination (R^2 = 0.93). In the study conducted by Xing et al. [38], machine learning models were built to estimate the *HHV* value of biomass based on the input parameters of ultimate and proximate analysis. The RFR model in the study shows a great fit for prediction (R^2 > 0.94), while the SVM model also shows good performance (R^2~0.90). The RFR models give good results in terms of performance. The parameters R^2, MAPE and RMSE are calculated at 0.94, 0.57 and 2.56, respectively, while for the SVM model the parameters R^2, MAPE and RMSE are 0.90, 0.76 and 3.53, respectively. The study also showed the relative importance of the input parameters of ultimate analysis in the models for C (61.6%), H (20%), O (9.6%) and N (8.8%). Considering everything, it can be concluded that the developed models are suitable for estimating the *HHV* based on the input parameters of the ultimate analysis. Using the performed statistical test "Goodness of Fit", the parameters showed a low level of error in estimating the *HHV*, while the SVM model shows a higher level of performance in modeling.

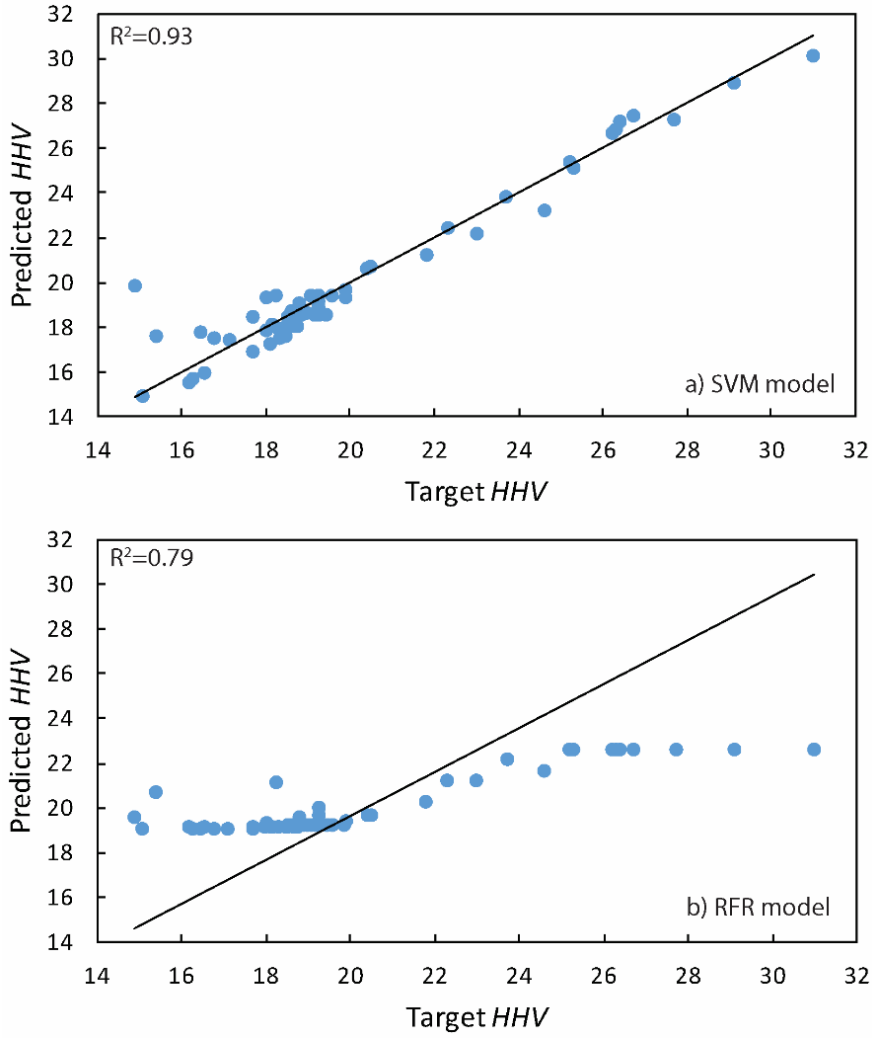

Figure 4. Parity plot of predicted *HHV* vs. Target *HHV* for developed (**a**) SVM and (**b**) RFR models.

The developed SVM and RFR models show good ability in estimating the *HHV* of corn, soybean and sunflower biomass. In developing the model, the input parameters of the ultimate analysis (percent concentration of C, H, N, S and O) were used to estimate *HHV*. In the case of the RFR model, the output value is most influenced by the variables in the following order: O (1.00), C (0.99), N (0.94), S (0.78) and H (0.57). The SVM model

with a number of five independent input parameters, as a regression model, shows the best performance with nine support vectors using RDF (Radial Basis Function) as a Kelner type. Data from 51 different biomass samples were used for the study (27 for corn biomass, 15 for soybean biomass and 19 for sunflower biomass). The SVM model showed better performance than the RFR model because the mentioned model generalizes and covers the data better for a small and medium data set, while the RFR model shows better estimation performance for a medium and large data set [39,40].

5. Conclusions

Agricultural biomass generated from corn, sunflower and soybean production has a great potential as a feedstock for energy production. Large amounts of biomass are produced as a byproduct during agricultural production and are often unused. In order to assess the possibility of using biomass as a fuel, its energy properties and composition must be determined. Nonlinear models offer the possibility to estimate the *HHV* of biomass with high accuracy. The developed nonlinear models in the form of Random Forest Regression (RFR) and Support Vector Machine (SVM) were determined as successful tools in estimating biomass *HHV*. Thus, this work shows the satisfactory ability of the SVM (R^2 = 0.93) and RFR models (R^2 = 0.79) in estimating *HHV* based on the input parameters of the ultimate analysis of observed agricultural biomass.

Author Contributions: Conceptualization, I.B., L.P., A.M. and A.A.; methodology, I.B., M.K. and B.M.; software, I.B., L.P. and K.Š.; validation, L.P., T.K., V.J. and A.M.; formal analysis, I.B. and L.P.; investigation, J.K., K.Š., M.G. and M.K.; writing—original draft preparation, I.B.; writing—review and editing, L.P., T.K., A.M. and A.A.; visualization, I.B. and L.P.; supervision, A.M.; project administration, A.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund, under the Operational program competitiveness and cohesion 2014–2022, project no. KK 01.2.1.02.0286, "*Development of innovative pellets from forest and/or agricultural biomass—INOPELET*".

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Antar, M.; Lyu, D.; Nazari, M.; Shah, A.; Zhou, X.; Smith, D.L. Biomass for a sustainable bioeconomy: An overview of world biomass production and utilization. *Renew. Sustain. Energy Rev.* **2021**, *139*, 110691. [CrossRef]
2. Biswas, B.; Pandey, N.; Bisht, Y.; Singh, R.; Kumar, J.; Bhaskar, T. Pyrolysis of agricultural biomass residues: Comparative study of corn cob, wheat straw, rice straw and rice husk. *Bioresour. Technol.* **2017**, *237*, 57–63. [CrossRef] [PubMed]
3. Cai, J.; Wu, W.; Liu, R. An overview of distributed activation energy model and its application in the pyrolysis of lignocellulosic biomass. *Renew. Sustain. Energy Rev.* **2014**, *36*, 236–246. [CrossRef]
4. Erol, M.; Haykiri-Acma, H.; Küçükbayrak, S. Calorific value estimation of biomass from their proximate analyses data. *Renew. Energy* **2010**, *35*, 170–173. [CrossRef]
5. Khunphakdee, P.; Korkerd, K.; Soanuch, C.; Chalermsinsuwan, B. Data-driven correlations of higher heating value for biomass, waste and their combination based on their elemental compositions. *Energy Rep.* **2014**, *8*, 36–42. [CrossRef]
6. Asonja, A.; Desnica, E.; Radovanović, L.Z. Energy efficiency analysis of corn cob used as a fuel. *Energy Sources Part B Econ. Plan. Policy* **2017**, *12*, 1–7. [CrossRef]
7. Krisnawati, A.; Adie, M.M. Variability of Biomass and Harvest Index from Several Soybean Genotypes as Renewable Energy Source. *Energy Procedia* **2015**, *65*, 14–21. [CrossRef]
8. Muchlish Adie, M.; Krisnawati, A. Soybean Opportunity as Source of New Energy in Indonesia. *Int. J. Renew. Energy Dev.* **2014**, *3*, 37–43. [CrossRef]
9. Antonopoulou, G.; Dimitrellos, G.; Beobide, A.S.; Vayenas, D.; Lyberatos, G. Chemical Pretreatment of Sunflower Straw Biomass: The Effect on Chemical Composition and Structural Changes. *Waste Biomass Valorization* **2015**, *6*, 733–746. [CrossRef]
10. Qian, H.; Zhu, W.; Fan, S.; Liu, C.; Lu, X.; Wang, Z.; Huang, D.; Chen, W. Prediction models for chemical exergy of biomass on dry basis from ultimate analysis using available electron concepts. *Energy* **2017**, *131*, 251–258. [CrossRef]
11. 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]

12. Dai, Z.; Chen, Z.; Selmi, A.; Jermsittiparsert, K.; Denić, N.M.; Nešić, Z. Machine learning prediction of higher heating value of biomass. *Biomass-Convers. Biorefinery* **2021**, *11*, 1–9. [CrossRef]
13. Roman, K.; Barwicki, J.; Rzodkiewicz, W.; Dawidowski, M. Evaluation of Mechanical and Energetic Properties of the Forest Residues Shredded Chips during Briquetting Process. *Energies* **2021**, *14*, 3270. [CrossRef]
14. Ibikunle, R.A.; Lukman, A.F.; Titiladunayo, I.F.; Akeju, E.A.; Dahunsi, S.O. Modeling and robust prediction of high heating values of municipal solid waste based on ultimate analysis. *Energy Sources Part A Recover. Util. Environ. Eff.* **2020**, 1–18. [CrossRef]
15. Nieto, P.G.; Garcia-Gonzalo, E.; Lasheras, F.S.; Paredes–Sánchez, J.; Fernández, P.R. Forecast of the higher heating value in biomass torrefaction by means of machine learning techniques. *J. Comput. Appl. Math.* **2019**, *357*, 284–301. [CrossRef]
16. Bychkov, A.L.; Denkin, A.I.; Tikhova, V.D.; Lomovsky, O.I. Prediction of higher heating values of plant biomass from ultimate analysis data. *J. Therm. Anal. Calorim.* **2017**, *130*, 1399–1405. [CrossRef]
17. Grubor, M.; Krička, T.; Bilandžija, N.; Jurišić, V.; Voća, N.; Antonović, A.; Matin, A. Maize straw as an energy source for solid fuel production. In *53. Hrvatski i 13. Međunarodni Simpozij Agronoma*; Rozman, V., Antunović, Z., Eds.; Sveučilišta Josipa Jurja Strossmayera u Osijeku: Osijek, Croatia, 2018.
18. Matin, A.; Krička, T.; Bilandžija, N.; Jurišić, V.; Grubor, M.; Antonović, A.; Lakić, J.; Voća, N. Varieties influence on soybean straw biomass combustible and noncombustible properties. In *53. Hrvatski i 13. Međunarodni Simpozij Agronoma*; Rozman, V., Antunović, Z., Eds.; Sveučilišta Josipa Jurja Strossmayera u Osijeku: Osijek, Croatia, 2018.
19. Demirbaş, A. Fuel Characteristics of Olive Husk and Walnut, Hazelnut, Sunflower, and Almond Shells. *Energy Sources* **2002**, *24*, 215–221. [CrossRef]
20. Krička, T.; Matin, A.; Voća, N.; Pospišil, A.; Grubor, M.; Šaronja, I.; Jurišić, V. Changes in nutritional and energy properties of soybean seed and hull after roasting. *Res. Agric. Eng.* **2018**, *64*, 96–103. [CrossRef]
21. Tumuluru, J.S.; Kremer, T.; Wright, C.T.; Boardman, R.D. Proximate and Ultimate Compositional Changes in Corn Stover during Torrrefaction using Thermogravimetric Analyzer and Microwaves. In Proceedings of the 2012 ASABE Annual International Meeting, Dallas, TX, USA, 29 July–1 August 2012; pp. 1348–1365. [CrossRef]
22. Tahir, M.H.; Irfan, R.M.; Hussain, M.B.; Alhumade, H.; Al-Turki, Y.; Cheng, X.; Karim, A.; Ibrahim, M.; Rathore, H.A. Catalytic Fast Pyrolysis of Soybean Straw Biomass for Glycolaldehyde-Rich Bio-oil Production and Subsequent Extraction. *ACS Omega* **2021**, *6*, 33694–33700. [CrossRef]
23. Ahmad, S.; Zhu, X.; Wei, X.; Zhang, S. Influence of process parameters on hydrothermal modification of soybean residue: Insight into the nutrient, solid biofuel, and thermal properties of hydrochars. *J. Environ. Manag.* **2021**, *283*, 111981. [CrossRef]
24. Riva, G.; Foppa Pedretti, E.; Toscano, G.; Duca, D.; Rossini, G.; Mengarelli, C. Torrefaction of residues and by-products from sunflower chain. *J. Agric. Eng.* **2013**, *44*, 601–606. [CrossRef]
25. Liu, X.; Hua, W.; Wu, S. Characterization of Thermo-Chemical Degradation and Pyrolysis Properties for Three Kinds of Biomass Residues. *BioResources* **2016**, *11*, 8806–8819. [CrossRef]
26. Tortosa Masiá, A.A.; Buhre, B.J.P.; Gupta, R.P.; Wall, T.F. Characterising ash of biomass and waste. *Fuel Process. Technol.* **2007**, *88*, 1071–1081. [CrossRef]
27. Turzyński, T.; Kluska, J.; Ochnio, M.; Kardaś, D. Comparative Analysis of Pelletized and Unpelletized Sunflower Husks Combustion Process in a Batch-Type Reactor. *Materials* **2021**, *14*, 2484. [CrossRef] [PubMed]
28. Özyuğuran, A.; Yaman, S.; Küçükbayrak, S. Prediction of calorific value of biomass based on elemental analysis. *Int. Adv. Res. Eng. J.* **2018**, *2*, 254–260. Available online: https://www.dergipark.gov.tr (accessed on 2 September 2022).
29. Schonlau, M.; Zou, R.Y. The random forest algorithm for statistical learning. *Stata J.* **2020**, *20*, 3–29. [CrossRef]
30. Ding, S.; Zhu, Z.; Zhang, X. An overview on semi-supervised support vector machine. *Neural Comput. Appl.* **2017**, *28*, 969–978. [CrossRef]
31. Zendehboudi, A.; Baseer, M.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. *J. Clean. Prod.* **2018**, *199*, 272–285. [CrossRef]
32. Arsenović, M.; Pezo, L.; Stanković, S.; Radojević, Z. Factor space differentiation of brick clays according to mineral content:Prediction of final brick product quality. *Appl. Clay Sci.* **2015**, *115*, 108–114. [CrossRef]
33. Bilandžija, N.; Voća, N.; Leto, J.; Jurišić, V.; Grubor, M.; Matin, A.; Krička, T. Yield and Biomass Composition of Miscanthus x Giganteus in the Mountain Area of Croatia. In *Transactions of FAMENA*; University of Zagreb: Zagreb, Croatia, 2018; Volume 42, pp. 51–60.
34. Rutledge, D.N. Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis. *J. Anal. Test.* **2018**, *2*, 235–248. [CrossRef]
35. Beattie, J.R.; Esmonde-White, F.W.L. Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. *Appl. Spectrosc.* **2021**, *75*, 361–375. [CrossRef] [PubMed]
36. Wang, X.; Huang, F.; Cheng, Y. Computational performance optimization of support vector machine based on support vectors. *Neurocomputing* **2016**, *211*, 66–71. [CrossRef]
37. Han, S.; Kim, H. Optimal Feature Set Size in Random Forest Regression. *Appl. Sci.* **2021**, *11*, 3428. [CrossRef]
38. Nguyen, Q.V.; Lin Huang, M.; Simoff, S. Enhancing Scatter-plots with Start-plots for Visualising Multi-dimensional Data. In Proceedings of the 2020 24th International Conference Information Visualisation (IV), Melbourne, Australia, 7–11 September 2020; pp. 80–85. [CrossRef]

39. Xing, J.; Luo, K.; Wang, H.; Gao, Z.; Fan, J. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* **2019**, *188*, 116077. [CrossRef]
40. Kramer, K.A.; Hall, L.O.; Goldgof, D.B.; Remsen, A.; Luo, T. Fast Support Vector Machines for Continuous Data. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2009**, *39*, 989–1001. [CrossRef] [PubMed]

mathematics

MDPI

Article

Comparison of Different Machine Learning Models for Modelling the Higher Heating Value of Biomass

Ivan Brandić [1,*], Lato Pezo [2], Nikola Bilandžija [1], Anamarija Peter [1], Jona Šurić [1] and Neven Voća [1]

[1] Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia
[2] Institute of General and Physical Chemistry, University of Belgrade, Studentski trg 12/V, 11000 Belgrade, Serbia
* Correspondence: ibrandic@agr.hr

Abstract: The aim of this study was to investigate the potential of using structural analysis parameters for estimating the higher heating value (HHV) of biomass by obtaining information on the composition of cellulose, lignin, and hemicellulose. To achieve this goal, several nonlinear mathematical models were developed, including polynomials, support vector machines (SVMs), random forest regression (RFR) and artificial neural networks (ANN) for predicting HHV. The performed statistical analysis "goodness of fit" showed that the ANN model has the best performance in terms of coefficient of determination (R^2 = 0.90) and the lowest level of model error for the parameters X^2 (0.25), RMSE (0.50), and MPE (2.22). Thus, the ANN model was identified as the most appropriate model for determining the HHV of different biomasses based on the specified input parameters. In conclusion, the results of this study demonstrate the potential of using structural analysis parameters as input for HHV modeling, which is a promising approach for the field of biomass energy production. The development of the model ANN and the comparative analysis of the different models provide important insights for future research in this field.

Keywords: structural analysis; support vector machine; artificial neural network; random forest regression; high order polynomials

MSC: 49M37

Citation: 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. https://doi.org/10.3390/math11092098

Academic Editor: Konstantin Kozlov

Received: 3 April 2023
Revised: 25 April 2023
Accepted: 26 April 2023
Published: 28 April 2023

1. Introduction

With the increasing use of renewable energy sources to meet the growing demand for energy, biomass will play a central role in the coming years. This is particularly important given the rising cost of conventional fuels and the need to mitigate the exacerbation of climate change. [1]. Biomass, which refers to biodegradable residues from agricultural production, various types of organic waste, residues of biological origin from agriculture and forestry, and biological residues from plant and animal production [2], is one of the most common renewable sources for energy production [3]. Research has shown that the use of energy crops for energy production can replace existing conventional fuels and slow down negative climate change [4]. Lignocellulosic biomass is widely recognized as an effective and efficient renewable resource for energy production [5]. In recent times, the use of renewable energy resources has become a critical component of energy security. A crucial factor in determining fuel quality and conducting tests is the higher heating value (HHV), but its determination requires a long time and the use of special laboratory equipment. For this reason, various mathematical models are developed to predict the HHV value depending on the input parameters of different analyses [6]. In addition to mathematical models, numerous machine learning techniques (ML) facilitate the creation of more efficient forecasting models. Recently, artificial neural networks (ANNs) have been increasingly used in the field of modeling and represent a suitable tool for the investigation and evaluation of biomass energy parameters [7]. The basic features of an ANN are its

structure, "learning" algorithm, and activation function used to transfer the computed values from one neuron to another. By selecting the input parameters, the models in the form of ANNs can fit and generalize the data according to the desired output value. The effectiveness of ANN is determined by comparing experimental and computed data. There are several models of ANN, but in application, especially for regression, the most efficient models are multilayer perceptrons (MLP-ANN) [8–10]. RFR models, in addition to anticipating, allow the size of the interval to be determined by estimation (instead of using a separate dataset required for calibration). They also provide reliability and a much higher level of effectiveness for prediction than existing linear methods [11]. RFR is used as a useful tool for predicting desired output values. When using the RFR model, it is important to determine intervals that contain values with a certain prediction probability. Similar to other approaches to forecasting, these models are typically used for finding "points" that are not accompanied by actual value deviation data [12]. SVM as a model for regression analysis, uses hyperplane classifiers by mapping the input data into a multidimensional space and comparing it to the output data [13]. Garcia Nieto et al. [14] conducted a study to predict the higher heating value of biomass and compared the performance of different models. The cubic-SVM model showed the strongest correlation between predicted and actual values, with the highest R^2 value of 0.94 and the lowest RMSE and MAE values of 0.39 and 0.32, respectively. Its MBE value was close to zero (0.0012), indicating minimal bias in the predictions. In contrast, the random forest model showed the weakest performance with the lowest R^2 value (0.59) and the highest RMSE (1.06) and MAE values (0.86), indicating lower accuracy and precision in predicting the higher heating value of biomass. In his study, Afolabi et al. [15] (2022) compared different ML models, including decision tree (DT), random forest (RF), and artificial neural networks (ANN) using criteria of statistical measures. The table shows the mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) for each model, allowing an evaluation of their performance. The DT model has an MAE of 1.48, an MSE of 4.36, and an RMSE of 2.09. The RF model has an MAE of 1.01, an MSE of 1.87, and an RMSE of 1.37, indicating better performance compared with the DT model. The ANN model has an MAE of 1.21, an MSE of 2.43, and an RMSE of 1.56, indicating better performance compared to the DT model, but slightly worse than the RF model. Overall, the random forest model shows the best performance among the three models based on the lowest values of MAE, MSE, and RMSE. Liu et al., 2022 [16], used a RFR to predict the HHV of torrefied biomass. The model RF was trained with 10-fold cross-validation to fit the hyperparameters. The model achieved high prediction accuracy, with an R^2 value of 0.91 for the test dataset. The results of this study show that the random forest model (RF) has the ability to estimate the higher heating value (HHV) of torrefied biomass with a high degree of precision. Dubey and Guruviah (2022) [17] constructed a support vector machine (SVM) model for predicting and optimizing the prediction of HHV of agricultural biomass based on proximate analysis data. The main objective of this paper is to develop nonlinear machine learning models in the form of higher-degree polynomials, SVM, RFR, and ANN, and to investigate the possibility of HHV modeling in terms of the input parameters of biomass structural analysis, which include the variables cellulose, hemicellulose, and lignin. For a better understanding of the overall concept of the work, Figure 1 shows the flowchart of the research conducted with the aim of determining the most appropriate machine learning model in terms of predictive ability, but also in terms of the error rate of each model. The present study is characterized by the comparative evaluation of different machine learning approaches, which include polynomial functions, support vector machines (SVM), random forest regressors (RFR), and artificial neural networks (ANN), with the aim of predicting the higher heating value (HHV) of biomass through structural analysis parameters.

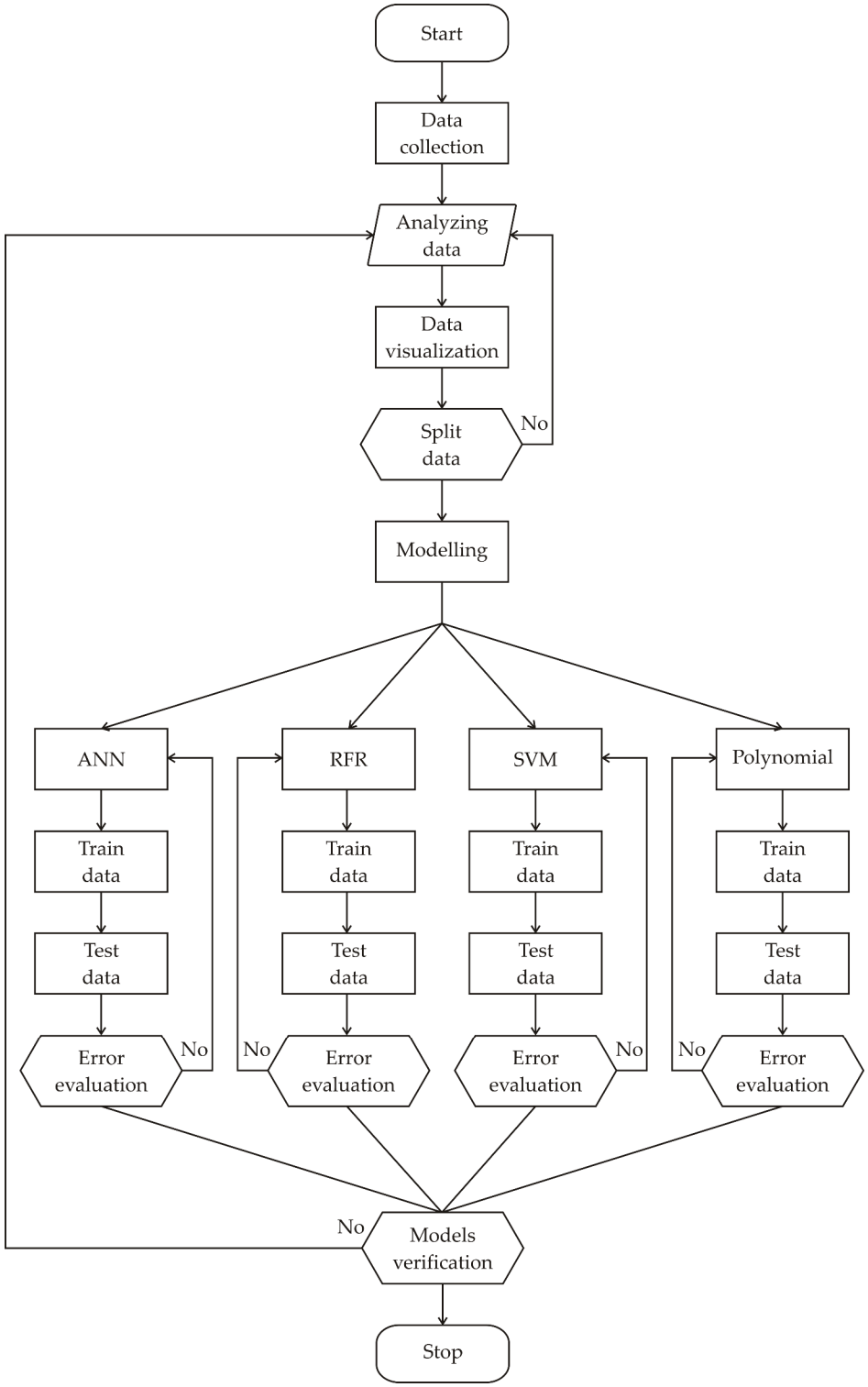

Figure 1. Flowchart of the conducted research.

By means of Yoon's global sensitivity method (based on the ANN model), the influence of the input variables on the output values is investigated. In order to compare the above models in terms of the different modeling errors, literature and calculated data are compared with the model. The coefficient of determination is taken as the main parameter of the model comparison, through which the most appropriate model for estimating the HHV biomass is determined.

2. Materials and Methods

2.1. Data Collection

Data from different types of biomasses were used for the study, including input variables, percentages of cellulose, lignin, hemicellulose, and HHV on a dry basis. The data used for the analysis were taken from published papers [1,18,19] and are listed in Table S1 (overall, 235 samples). The large differences between the minimum and maximum values also allow for the construction of a more universal model for the prediction of HHV from the above variables.

2.2. Data Processing

The Python (Python 3.10) [20] libraries Pandas, Seaborn, Matplotlib.pyplot, and NumPy [21–24] were imported to create a pair plot for the dataset used. These libraries were used extensively for data analysis and plotting [25]. The *corr()* function, which calculates the pairwise correlation between all columns, was used to calculate the correlations between the features of the data frame. To display only the upper triangle of the heatmap, a triangular mask was created using the NumPy module. The heatmap was created with the Seaborn *heatmap()* function, the correlation matrix as input, and the mask used to annotate the correlation values. To demonstrate the distribution of each feature and the pairwise correlations, a pair plot was created using Seaborn, and the two plots were combined into a single figure using the Matplotlib library. To evaluate the effectiveness and performance of the support vector machine (SVM), random forest regression (RFR), polynomial, and artificial neural network (ANN) models in calculating the higher heating value (HHV) based on input data from the structural analysis components, several statistical parameters were calculated. These included the reduced chi-square (X^2) (1), root mean square error (RMSE) (2), coefficient of determination (R^2) (3), mean systematic error (MBE) (4), mean percentage error (MPE) (5), total squared error (SSE) (6), and average absolute relative deviation (AARD) (7). The RMSE values indicate the efficiency of the model by comparing calculated values with experimentally measured values. The MBE values are used to determine the standard deviation between the predicted and measured values [26]. These statistical parameters were calculated using equations [27–29]. In addition, Yoon's method of global sensitivity (8) was used to evaluate the direct influence of the input parameters on the output variables, which corresponds to the weighting coefficients (w) within the ANN model [30].

$$x^2 = \frac{\sum_{i=1}^{N} (x_{predicted,i} - x_{experimental,i})^2}{N - n} \quad (1)$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N} (x_{predicted,i} - x_{experimental,i})^2 \right]^{1/2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left[x_i^{predicted} - x_i^{experimental} \right]^2}{\sum_{i=1}^{n} \left[x_i^{predicted} - x^m \right]^2}, x_m = \frac{\sum_{i=1}^{n} x_i^{experimental}}{n} \quad (3)$$

$$MBE = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{predicted,i} - x_{experimental,i}) \quad (4)$$

$$MPE = \frac{100}{N} \cdot \sum_{i=1}^{N} \left(\frac{\left| x_{predicted,i} - x_{experimental,i} \right|}{x_{experimental,i}} \right) \quad (5)$$

$$SSE = \sum_{i=1}^{N} (x_{predicted,i} - x_{experimental,i})^2 \quad (6)$$

$$AARD = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|x_i^{predicted} - x_i^{experimental}\right|}{x_i^{experimental}} \quad (7)$$

$$RI_{ij}(\%) = \frac{\sum_{k=0}^{n}(w_{ik}\cdot w_{kj})}{\sum_{i=0}^{m}\left|\sum_{k=0}^{n}(w_{ik}\cdot w_{kj})\right|}\cdot 100\% \quad (8)$$

where N is population size, n is sample size, $x^{predicted}$ indicates predicted value, $x^{experimental}$ is experimental value. In this study, the C++ programming language [31] was used to implement ML models (ANN, SVM, and RFR). C++ was chosen because of its high performance and efficient memory management, which are essential for processing large and complex datasets [32]. The low-level control of the language also enabled the optimization of algorithms and the implementation of advanced techniques in the ML models. Before the models were created, all the data were divided into a part for training and a part for testing the model in a 70:30 ratio.

2.3. SVM Modelling

SVM models are based on theories of averaging and are algorithms that can be used for supervised learning for regression. SVM as regression models make predictions by splitting the data into a part for learning and testing the model and are suitable for predicting the HHV of biomass. In models for nonlinear applications, the input low-dimensional space vectors must first be transformed with a nonlinear function (9) (Φ) [14]:

$$f(x) = w^T\Phi(x) + b \quad (9)$$

where w and b represent weight vector and intercept of the model.

The SVM model was created to predict HHV biomass based on the input parameters of the structural analysis. The model was created as regression type 1 with a training constant of 10. The epsilon measure of the model is set to 0.1, while the radial basis function (gamma value) is set to 1.00. The total number of model iterations is 10,000.

2.4. Polynomial Regression Model

The created polynomial model relies on relations between variables based on structural analysis of biomass and HHV values of output data. To adjust for possible causes of variation, a statistical experimental design was used to examine the effects of three variables (factors) on an outcome variable while controlling for a grouping variable (block). The polynomial model of this experimental design was as follows:

$$HHV = \beta_0 + \beta_1\cdot Cel + \beta_2\cdot Cel^2 + \beta_3\cdot Lig + \beta_4\cdot Lig^2 + \beta_5\cdot Hem + \beta_6\cdot Hem^2 + \beta_7\cdot Cel\cdot Lig + \beta_8\cdot Cel\cdot Hem + \beta_9\cdot Lig\cdot Hem$$

where HHV is the response variable, Cel, Lig, and Hem are the three factors, $\beta_{/1}$ is the intercept, $\beta_{/1}$–$\beta_{/1}$ are the main effects of the factors, $\beta_{/1}$–$\beta_{/1}$ are the two-way interactions between the factors.

2.5. RFR Modeling

RFR are learning algorithms that combine multiple random decision trees and make anticipation based on the average value; they include methods of classification and regression, and use a certain number of random trees [33–35]. During RFR model evaluation, the number of trees was set to values of 100, 200, 300, 400, 500, and 10,000.

2.6. ANN Modelling

MLP is one of the types of ANN and is widely used in computing. The main advantage of this type of network is that it can "learn" to make connections between input and output data, which is very useful in predicting nonlinear problems in areas where large amounts of data have to be processed [36–38]. Number of artificial neurons in the hidden layer may vary based on the error and trial methods. The learning process of a neural network involves processing input data, which is then converted to the desired output data [39]. The two basic types of network learning processes are supervised and unsupervised. In the supervised learning process, the model is provided with ANN output data, based on which it compares the values obtained [40]. The developed ANN model was trained 100,000 times with a random number of neurons in the hidden layer (5–20). Different activation functions and random values for weighting coefficients and bias were used. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm was used to solve the nonlinear optimization during the modeling process of ANN [41]. The neural network model, written in matrix notation, contains biases and weight coefficients for the hidden and output layers, represented by matrices and vectors $W_{/1}$, $B_{/1}$, $W_{/1}$, and $B_{/1}$, respectively. Y is the output value, while $f_{/1}$ and $f_{/1}$ denote the transfer functions for the hidden and output layers, respectively. X represents the matrix of the input layer [42]:

$$Y = f_1(W_2 \cdot f_2(W_1 \cdot X + B_1) + B_2) \quad (10)$$

3. Results

3.1. Data Distribution

According to the literature data obtained, the cellulose concentration ranged between 10.66 and 56.62%, while lignin percent scoped between 2.39 and 22%; hemicellulose spanned across the range of 5.97–37%, whereas the obtained HHV value was between 12.54 and 19.25 MJ kg^{-1}. The extreme values of the collected data varied greatly due to the different types of biomasses, which are characterized by different chemical compositions.

Figure 2 shows the distribution of the individual variables and the correlation between the observed variables and the HHV biomass.

The correlation coefficient between the observed variables is positive and statistically significant at a coefficient of $p \leq 0.01$. The correlation coefficient (r) of HHV and the variable Hem is 0.74, while it is 0.88 for lignin and 0.89 for cellulose. When looking at the distribution, certain behavioral patterns of the observed values become visible. In the dataset used to build the model, the largest part consisted of data on the energy crop of Miscanthus (192 samples), whose (for example) calorific value varies from 15.53 to 19.52 MJ kg^{-1}, while the other biomass samples have a lower average calorific value (12.54–17.07 MJ kg^{-1}), which can explain the uneven distribution of the data.

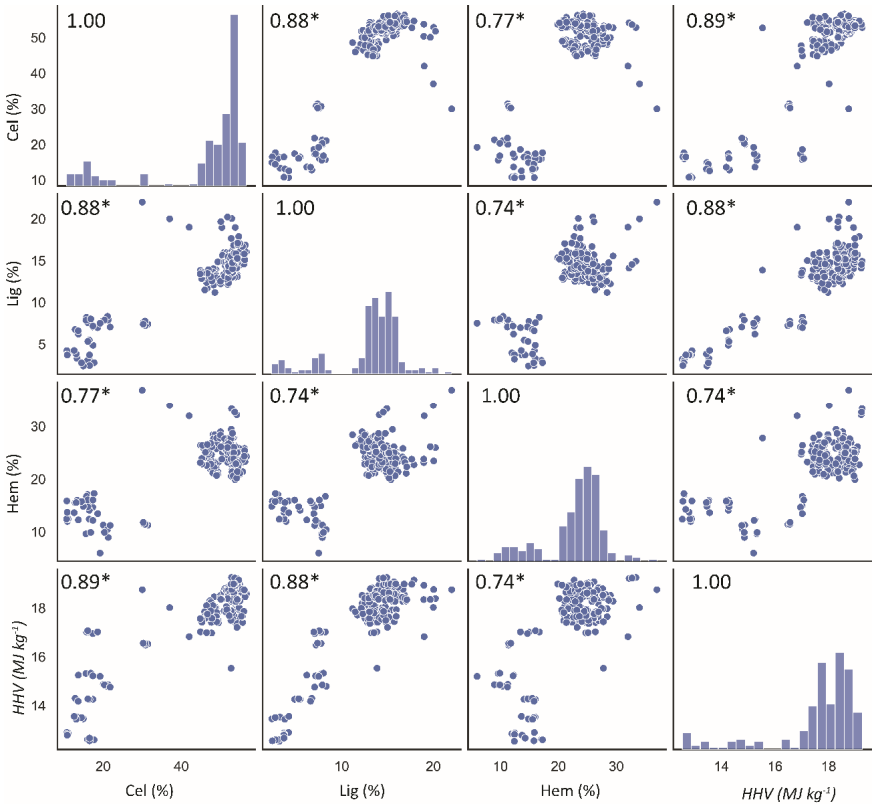

Figure 2. Pair plot with correlation coefficient of the observed values from the structural analysis of the biomass (statistical significance: * $p \leq 0.01$).

3.2. Polynomial Regression Model

With regard to the proposed higher-degree polynomial model, the intercept values, the main effect on the factors, and the interaction of the effects on several factors were calculated and presented (Table 1).

Table 1. Estimated effects of factors on HHV output.

Factor	Effect	ε
β_0	17.38	0.18
β_1	0.71	0.54
β_2	0.48	0.76
β_3	5.07	0.76
β_4	−1.61	1.95
β_5	0.16	0.79
β_6	0.44	1.67
β_7	−2.57	2.29
β_8	1.51	1.67
β_9	−2.28	2.29

β_0—intercept value; β_1–β_9—main effects; ε—standard error.

3.3. RFR Model

During the development of the RFR model to predict HHV values, a large number of decision trees were constructed (1940). The data for RFR were split into a random sample of 30% and a subsample of 50%.

3.4. ANN Model

The proposed ANN model consists of an input layer, a hidden layer, and an output layer with architecture 3-4-1 (number of artificial neurons in the input, hidden, and output layers). The weights and biases (Table 2) were determined by randomly searching for values that would make the model sufficiently accurate in modeling the output values.

Table 2. Weights and biases of input and output layers of the developed ANN model.

Artificial Neuron Number	Input Layer				Output Layer	
	Weight Coefficient			Bias	Weight Coefficient	Bias
	Cel	Lig	Hem		HHV	
1	8.70	−5.06	−1.02	−3.82	−0.47	1.46
2	−3.00	−1.80	−1.91	1.55	−0.24	
3	3.35	−2.33	0.37	−0.53	1.72	
4	2.38	−1.87	0.45	−0.02	−1.74	

Cel—cellulose; Lig—lignin; Hem—hemicellulose; HHV—higher heating value.

The model presented showed the greatest predictive ability in the architecture with four neurons in the values.

3.5. Model Performance

The results (Table 3) show that the MLP neural network model outperforms the other three models in terms of RMSE, AARD, and R^2. The low RMSE value indicates that the ANN model predicts the output variables with high accuracy. The high AARD value indicates that the ANN model has a low relative error and is therefore acceptable for practical applications. Several measures, such as R^2 and MBE, indicate that the RFR model performs quite well. The low MBE value indicates that the RFR model predicts the output variables with low bias. Nevertheless, the RMSE and AARD values show that the model has more error than the MLP model. For most measurements, the SVM model performed worse than the other models. The SVM model has more relative errors than the other models, indicating that it is less suitable for real-world applications. The low R^2 value indicates that the input variables explain a smaller proportion of the variation in the output variables. The skewness value indicates that the error distribution of the SVM model' is nearly symmetric.

Table 3. Performance of developed ML models.

Model	Net. Name	Training Perf.	Test Perf.	Training Error	Test Error	Training Algorithm	Error Function	Hidden Activation	Output Activation
ANN	MLP 3-4-1	0.88	0.95	0.15	0.07	BFGS 82	SOS	Exponential	Identity
RFR	-	0.89	0.92	-	-	-	-	-	-
SVM		0.85	0.89						
Polynomial		0.85	0.92						

ANN—artificial neural network; RFR—random forest regression; SVM—support vector machine.

The ANN was trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization technique with an exponential hidden layer activation function and an identity output activation function. With training and testing errors of 0.15 and 0.07, respectively, the ANN achieved a training performance of 0.88 and a testing performance of 0.95. The training performance of the RFR model was 0.89, and the test performance was 0.92. SVM and polynomial regression models achieved training and test values of 0.85 and 0.89, respectively. Overall, it was found that the MLP neural network performed better than the other models with the specified architecture and training conditions.

The scatter plot is one of the most commonly used types of visualization that shows the behavior of data on the x-y axis [43,44]. Figure 3 shows the scatter plot of the overlap of the predicted data with the experimentally determined values for the developed models in the form of polynomials, SVM, RFR, and ANN.

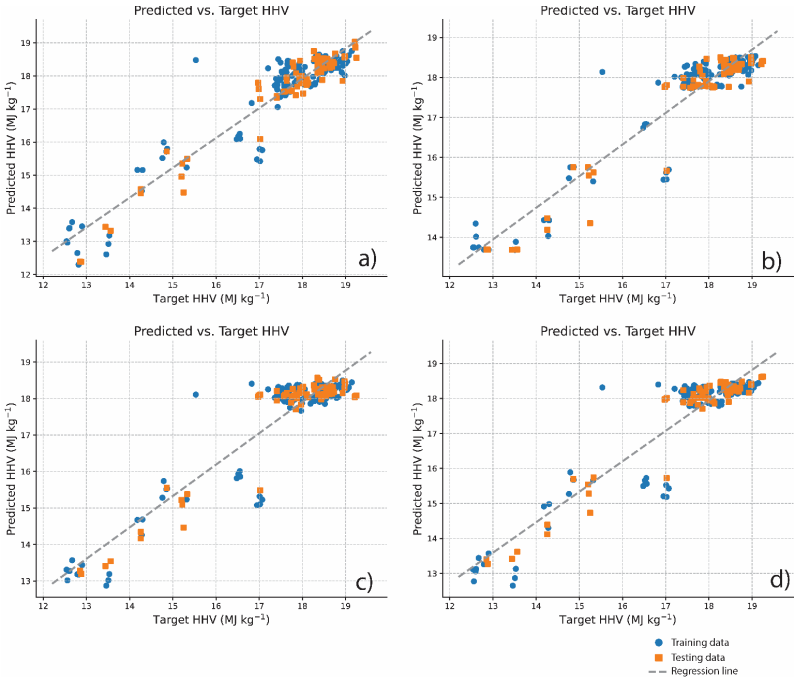

Figure 3. Comparison of experimentally obtained values with nonlinear models ((**a**) ANN, (**b**) RFR, (**c**) SVM, and (**d**) polynomial) predicted values for training and testing data.

Figure 4 shows the ability of the models to predict HHV biomass as a function of the number of observations. As can be seen, the models generally agree well with the observed data, with the highest agreement observed between observations 50 and 150, i.e., when the data are clustered and there is little variability. The model ANN had the lowest estimation error with respect to the number of observed samples of the models listed.

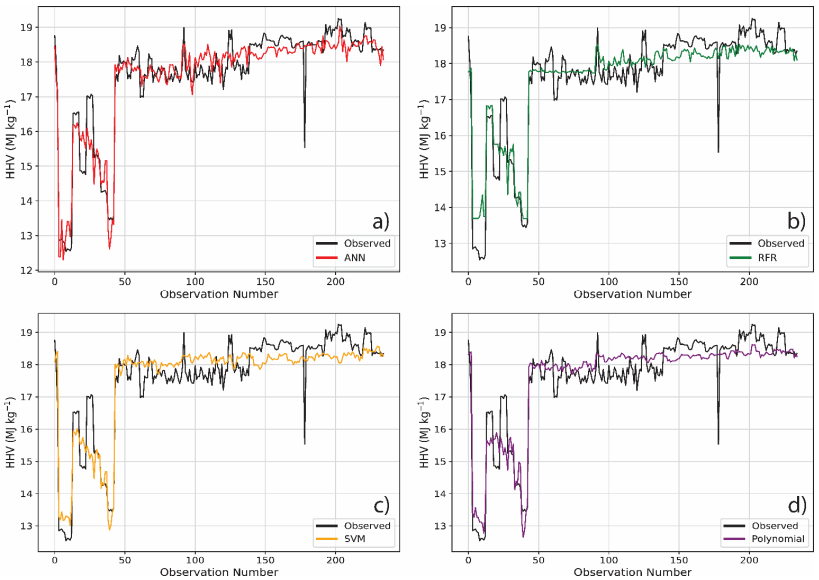

Figure 4. Comparison of nonlinear models ((**a**) ANN, (**b**) RFR, (**c**) SVM, and (**d**) polynomial) in the estimation of HHV biomass regarding the observation number.

3.6. Global Sensitivity Analysis of the Developed ANN Model

The global sensitivity analysis is performed according to the Yoon method, which calculates the direct influence of the input parameters on the output values. Figure 5 shows the relative importance (%) of each variable in the structural analysis of biomass for the output value of HHV. The range in which the relevance factor is shown is between −1 and 1. The influence of HHV is affected by increasing the value of lignin and hemicellulose and decreasing the input value of cellulose.

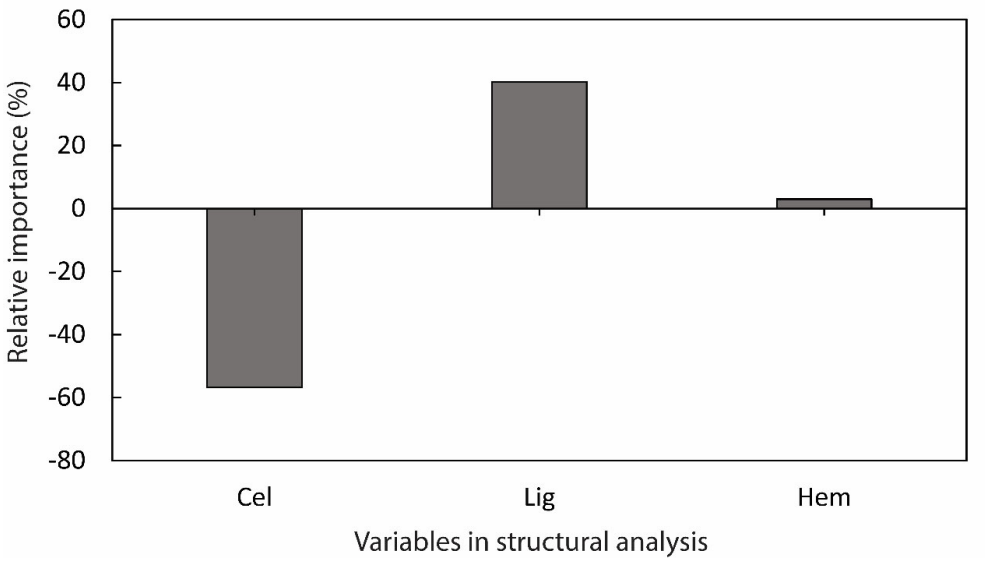

Figure 5. Relative importance (%) of structural analysis variables on HHV.

Figure 5 shows the influence of Cel (−56.78%), Lig (40.27%), and Hem (2.94%). Considering the presented influencing factors and their relative importance, it can be concluded that they affect the total output value of HHV by increasing the input values of lignin and hemicellulose and decreasing cellulose.

The global sensitivity analysis showed that increasing the value of lignin and hemicellulose and decreasing the value of cellulose has an impact on the effects of HHV. Considering the proportional relevance of the presented influencing factors, it can be concluded that they affect the overall output value of HHV by increasing the input values of lignin and hemicellulose and decreasing the input value of cellulose. Overall, the MLP neural network model performed best in predicting the HHV of different types of biomasses. Global sensitivity analysis revealed that the most important parameters affecting HHV were lignin and hemicellulose. Future research could focus on improving model accuracy by adding more diverse datasets and conducting controlled experiments to reduce the effects of external influences.

3.7. Goodness of Fit

To determine the ability of the model to predict HHV biomass with respect to the input parameters of the structural analysis, it is necessary to calculate statistical parameters to assess the model's ability to predict, as well as to compare the individual models to make it clear which nonlinear model is the most accurate in forecasting.

Table 4 shows the calculated statistical test "goodness of fit" in relation to the polynomials, SVM, RFR, and ANN models based on the calculation of the HHV value in relation to the input parameters of the structural analysis of biomass.

Table 4. Statistical test "Goodness of fit".

Model	X^2	RMSE	MBE	MPE	SSE	AARD	R^2	Skew	Kurt	SD	Var
ANN	0.25	0.50	0.03	2.22	57.98	100.87	**0.90**	−0.90	5.20	0.50	0.25
RFR	0.29	0.54	0.01	2.45	68.26	113.90	**0.89**	−0.50	2.47	0.54	0.29
SVM	0.35	0.59	0.03	2.74	80.97	158.04	**0.86**	−0.03	1.72	0.59	0.35
Polynominal	0.32	0.56	0.00	2.62	74.89	230.35	**0.87**	−0.23	2.37	0.57	0.32

X^2-chi-squared test; RMSE—root mean square error; MBE—mean bias error; MPE—mean percentage error; SSE—sum squared error; AARD—average absolute relative deviation; R^2—coefficient of determination; Skew—skewness; Kurt—kurtosis; SD—standard deviation; Var—variance.

The ANN model had the best performance based on several metrics, such as X^2, RMSE, AARD, and R^2. The low RMSE value indicates that the ANN model has high accuracy in predicting the output variables. The high AARD value indicates that the ANN model has a low relative error, making it suitable for practical applications. The R^2 value indicates that a high proportion of the variance in the output variable can be explained by the input variables. The RFR model had relatively good performance based on some metrics, such as R^2 and MBE. The low MBE value indicates that the RFR model has a low bias in predicting the output variables. However, the RMSE and AARD values indicate that the model has higher errors compared to the model ANN. The model, in the form of SVM, had lower performance compared to the other models based on most metrics. The high AARD value indicates that the SVM model has higher relative errors compared to the other models, making it less suitable for practical applications. The low R^2 value indicates that a smaller proportion of the variance in the output variable can be explained by the input variables. The skewness value indicates that the error distribution in the SVM model is nearly symmetric, while the low X^2 value and RMSE indicate good agreement between the predicted and actual values. The MBE of 0.00 shows that the predictions of the models do not tend to overestimate or underestimate the actual values. The developed SVM model showed the best results, with 39 support and weight vectors.

4. Discussion

Analyzing the composition of biomass and its calorific value is crucial for understanding and optimizing its use as a renewable energy source. Dai et al. 2021 [45] state that there is a need to develop a model for predicting the energy properties of biomass based on various analyses that enable the use of biomass resources in energy applications. Considering the specifics of the connection between input parameters and output values of HHV biomass, the ML model proved to be more accurate in application than the existing linear models. Xing et al. (2019) [46] conducted a study in which they examined the possibility of applying the ML model to the estimation of HHV biomass based on ultimate and proximate analysis. Unlike empirical models, which showed a lower predictive ability ($R^2 < 0.70$), ML models (ANN, SVM, and RF) showed better performance in HHV biomass modeling ($R^2 > 0.90$). Afolabi et al. (2022) [15] conducted a study in which they used different ML models to estimate the HHV of different biomass classes. They used MAE, MSE, and RMSE as statistical measures of model error. Among others, RF and ANN models were created, which had satisfactory performance in terms of modeling error for the statistically calculated parameters MAE, MSE, and RMSE (1.01, 1.87, 1.37, 1.21, 2.43, and 1.56). Chen et al. (2022) [47] provided research results on the evaluation of HHV biochar. Gradient-boosting regression (GBR), RF, SVM algorithms, and linear regression methods were developed through modeling. For the development, 52 samples were collected, and 97 were taken from published literature sources so that the models could be optimized. Based on 52 experimental data points, the machine learning (ML) methods showed better predictive capabilities (training $R^2 \geq 0.96$) for the higher heating value (HHV) of biochar compared to multiple linear regression (MLR) (training $R^2 < 0.94$). The gradient boosting regression (GBR) algorithm successfully predicted the HHV of biochar (test dataset) using finite and proximal analysis, with $R^2 = 0.98$, MAE = 0.83, and RMSE = 1.08 when trained with the

experimental dataset. The random forest (RF) and support vector machine (SVM) models performed similarly well in predicting HHV, with R^2 = 0.97, MAE = 0.93, RMSE = 1.22, and R^2 = 0.97, MAE = 0.93, and RMSE = 1.23, respectively. Hosseinpour et al. (2018) [48] developed a new network of fuzzy partial least squares combined with principal component analysis (PCA-INFPLS) to estimate HHV biomass based on the input parameters of solid carbon, volatile matter, and ash content. The developed model PCA-INFPLS shows high performance in predicting the HHV, with modeling errors R^2 > 0.96, MSE < 0.51, and MAPE < 2.5%, so it can be concluded that the proposed model is suitable for modeling. Ghugare et al. (2014) [49] developed a genetic program (GP) and a multilayer perceptron (MLP) model to evaluate the fuel properties of solid biomass. The GP and MLP models showed good predictive performance in terms of accuracy and generalization. They achieve high correlation coefficients (>0.95) and low MAPE (<4.5%) when comparing experimental and model-predicted higher heating values (HHV).

In this research, four models were developed and compared: polynomial regression, support vector machines (SVM), random forest regression (RFR), and artificial neural networks (ANN). The ANN model, specifically a multilayer neural perceptron (MLP), showed superior performance in terms of RMSE (0.50), AARD (100.87), and R^2 (0.90) compared to the other models, indicating high accuracy and low relative error. The RFR model also performed well, but with higher errors than the ANN model. The SVM model showed lower performance with higher relative errors and lower R^2 values (0.86), indicating that it is less suitable for practical applications. A global sensitivity analysis showed that the most important parameters affecting HHV are lignin and hemicellulose, while cellulose has a negative influence. These results have implications for optimizing biomass composition to achieve higher heating values. The model ANN, which uses the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization method and an exponential hidden layer activation function, showed the best performance in predicting the HHV for different types of biomasses.

Future research should focus on improving the accuracy of the models by including more diverse datasets and conducting controlled experiments to minimize the influence of external factors. In addition, exploring other modeling techniques and refining current models, especially the ANN, can help develop more accurate predictive tools for biomass heating value. Such improvements can facilitate better decision-making for the efficient use of biomass as a renewable energy source, help address energy challenges, and mitigate climate change.

5. Conclusions

- Recently, more attention has been paid to the development of various models for predicting the energy parameters of biomass fuels. The factors cellulose, hemicellulose, and lignin influence the HHV.
- Using Yoon's method of global sensitivity, the increase in HHV biomass was found to be influenced by the increase in the parameters lignin and hemicellulose and the decrease in cellulose content.
- Four developed nonlinear models showed high performance in estimating HHV biomass: ANN (R^2 = 0.90), RFR (R^2 = 0.89), SVM (R^2 = 0.86), and polynomial (R^2 = 0.87).
- Using the statistical test "goodness of fit", the ANN model showed the smallest errors in estimating HHV and was determined based on the calculated parameters X^2, RMSE, MBE, MPE, SSE, and AARD.
- Among the developed models, ANN showed the best ability to summarize, generalize data, and predict.
- To reduce the error rate in the development of the ML model for estimating energy values of biomass, the expansion of the database, the categorization of the data, and the development of new algorithms are required for future research.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math11092098/s1, Table S1: supporting information: data and codes.

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

Funding: The publication was supported by the Open Access Publication Fund of the Croatian Academic and Research Libraries Consortium (CARLC).

Data Availability Statement: Not applicable.

Acknowledgments: The publication was supported by the Croatian Science Foundation, under project No. IP-2018-01-7472 "Sludge management via energy crops production" within the project "Young Researchers' Career Development Project—Training of Doctoral Students", and cofinanced by the European Union, under the OP "Efficient Human Resources 2014–2020" from the ESF funds.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Callejón-Ferre, A.; Carreño-Sánchez, J.; Suárez-Medina, F.; Pérez-Alonso, J.; Marti, B.V. Prediction models for higher heating value based on the structural analysis of the biomass of plant remains from the greenhouses of Almería (Spain). *Fuel* **2014**, *116*, 377–387. [CrossRef]
2. Demirbas, A. Higher heating values of lignin types from wood and non-wood lignocellulosic biomasses. *Energy Sources Part A Recovery Util. Environ. Eff.* **2017**, *39*, 592–598. [CrossRef]
3. Scarlat, N.; Dallemand, J.F.; Taylor, N.; Banja, M.; Sanchez, L.J.; Avraamides, M. *Brief on Biomass for Energy in the European Union | EU Science Hub*; EC Publication: Luxembourg, 2019; pp. 1–8.
4. Koçar, G.; Civaş, N. An overview of biofuels from energy crops: Current status and future prospects. *Renew. Sustain. Energy Rev.* **2013**, *28*, 900–916. [CrossRef]
5. Olatunji, O.; Akinlabi, S.; Oluseyi, A.; Peter, M.; Madushele, N. Experimental investigation of thermal properties of Lignocellulosic biomass: A review. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *413*, 012054. [CrossRef]
6. Boumanchar, I.; Charafeddine, K.; Chhiti, Y.; Alaoui, F.E.M.; Sahibed-Dine, A.; Bentiss, F.; Jama, C.; Bensitel, M. Biomass higher heating value prediction from ultimate analysis using multiple regression and genetic programming. *Biomass-Convers. Biorefin.* **2019**, *9*, 499–509. [CrossRef]
7. Obafemi, O.; Stephen, A.; Ajayi, O.; Nkosinathi, M. A survey of Artificial Neural Network-based Prediction Models for Thermal Properties of Biomass. *Procedia Manuf.* **2019**, *33*, 184–191. [CrossRef]
8. Grossi, E.; Buscema, P.M. Introduction to artificial neural networks. *Eur. J. Gastroenterol. Hepatol.* **2007**, *19*, 1046–1054. [CrossRef]
9. Kartal, F.; Özveren, U. A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. *Energy* **2020**, *209*, 118457. [CrossRef]
10. Pattanayak, S.; Loha, C.; Hauchhum, L.; Sailo, L. Application of MLP-ANN models for estimating the higher heating value of bamboo biomass. *Biomass-Convers. Biorefin.* **2020**, *11*, 2499–2508. [CrossRef]
11. Johansson, U.; Boström, H.; Löfström, T.; Linusson, H. Regression conformal prediction with random forests. *Mach. Learn.* **2014**, *97*, 155–176. [CrossRef]
12. Zhang, H.; Zimmerman, J.; Nettleton, D.; Nordman, D.J. Random Forest Prediction Intervals. *Am. Stat.* **2019**, *74*, 392–406. [CrossRef]
13. Sun, J.; Zhang, J.; Gu, Y.; Huang, Y.; Sun, Y.; Ma, G. Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression. *Constr. Build. Mater.* **2019**, *207*, 440–449. [CrossRef]
14. Nieto, P.J.G.; García-Gonzalo, E.; Paredes-Sánchez, J.P.; Sánchez, A.B.; Fernández, M.M. Predictive modelling of the higher heating value in biomass torrefaction for the energy treatment process using machine-learning techniques. *Neural Comput. Appl.* **2018**, *31*, 8823–8836. [CrossRef]
15. 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]
16. Liu, X.; Yang, H.; Yang, J.; Liu, F. Application of Random Forest Model Integrated with Feature Reduction for Biomass Torrefaction. *Sustainability* **2022**, *14*, 16055. [CrossRef]
17. Dubey, R.; Guruviah, V. Machine learning approach for categorical biomass higher heating value prediction based on proximate analysis. *Energy Sources Part A Recovery Util. Environ. Eff.* **2022**, *44*, 3381–3394. [CrossRef]
18. Mansor, A.M.; Lim, J.S.; Ani, F.N.; Hashim, H.; Ho, W.S. Characteristics of cellulose, hemicellulose and lignin of MD2 pineapple biomass. *Chem. Eng. Trans.* **2019**, *72*, 79–84. [CrossRef]

19. Voća, N.; Leto, J.; Karažija, T.; Bilandžija, N.; Peter, A.; Kutnjak, H.; Šurić, J.; Poljak, M. Energy Properties and Biomass Yield of Miscanthus x Giganteus Fertilized by Municipal Sewage Sludge. *Molecules* **2021**, *26*, 4371. [CrossRef]
20. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
21. McKinney, W. Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; Volume 445, pp. 51–56.
22. Seaborn. Available online: https://seaborn.pydata.org/ (accessed on 20 January 2023).
23. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
24. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
25. El Hachimi, C.; Belaqziz, S.; Khabba, S.; Chehbouni, A. Data Science Toolkit: An all-in-one python library to help researchers and practitioners in implementing data science-related algorithms with less effort. *Softw. Impacts* **2022**, *12*, 100240. [CrossRef]
26. Demirbas, A. Relationships between Heating Value and Lignin, Moisture, Ash and Extractive Contents of Biomass Fuels. *Energy Explor. Exploit.* **2002**, *20*, 105–111. [CrossRef]
27. Khalil, S.A.; Shaffie, A. A comparative study of total, direct and diffuse solar irradiance by using different models on horizontal and inclined surfaces for Cairo, Egypt. *Renew. Sustain. Energy Rev.* **2013**, *27*, 853–863. [CrossRef]
28. Arsenović, M.; Pezo, L.; Stanković, S.; Radojević, Z. Factor space differentiation of brick clays according to mineral content: Prediction of final brick product quality. *Appl. Clay Sci.* **2015**, *115*, 108–114. [CrossRef]
29. Dashti, A.; Bahrololoomi, A.; Amirkhani, F.; Mohammadi, A.H. Estimation of CO_2 adsorption in high capacity metal−organic frameworks: Applications to greenhouse gas control. *J. CO2 Util.* **2020**, *41*, 101256. [CrossRef]
30. Aćimović, M.; Pezo, L.; Tešević, V.; Čabarkapa, I.; Todosijević, M. QSRR Model for predicting retention indices of Satureja kitaibelii Wierzb. ex Heuff. essential oil composition. *Ind. Crops Prod.* **2020**, *154*, 112752. [CrossRef]
31. Stroustrup, B. *The C++ Programming Language*; Addison-Wesley: Reading, MA, USA, 1986.
32. Rassokhin, D. The C++ programming language in cheminformatics and computational chemistry. *J. Cheminform.* **2020**, *12*, 10. [CrossRef]
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
34. Scornet, E.; Biau, G.; Vert, J.-P. Consistency of random forests. *Ann. Stat.* **2015**, *43*, 1716–1741. [CrossRef]
35. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]
36. Aitkin, M.; Foxall, R. Statistical modelling of artificial neural networks using the multi-layer perceptron. *Stat. Comput.* **2003**, *13*, 227–239. [CrossRef]
37. Talebi, N.; Nasrabadi, A.M.; Mohammad-Rezazadeh, I. Estimation of effective connectivity using multi-layer perceptron artificial neural network. *Cogn. Neurodyn.* **2017**, *12*, 21–42. [CrossRef] [PubMed]
38. Cao, W.; Wang, X.; Ming, Z.; Gao, J. A review on neural networks with random weights. *Neurocomputing* **2018**, *275*, 278–287. [CrossRef]
39. Brandić, I.; Pezo, L.; Bilandžija, N.; Peter, A.; Šurić, J.; Voća, N. Artificial Neural Network as a Tool for Estimation of the Higher Heating Value of Miscanthus Based on Ultimate Analysis. *Mathematics* **2022**, *10*, 3732. [CrossRef]
40. Ozveren, U. An artificial intelligence approach to predict gross heating value of lignocellulosic fuels. *J. Energy Inst.* **2017**, *90*, 397–407. [CrossRef]
41. Rajković, D.; Jeromela, A.M.; Pezo, L.; Lončar, B.; Zanetti, F.; Monti, A.; Špika, A.K. Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models. *Agronomy* **2021**, *12*, 58. [CrossRef]
42. Pezo, L.L.; Ćurčić, B.L.; Filipović, V.S.; Nićetin, M.R.; Koprivica, G.B.; Mišljenović, N.M.; Lević, L.B. Artificial neural network model of pork meat cubes osmotic dehydratation. *Hem. Ind.* **2013**, *67*, 465–475. [CrossRef]
43. Geladi, P.; Manley, M. Scatter plotting in multivariate data analysis. *J. Chemom.* **2003**, *17*, 503–511. [CrossRef]
44. Keim, D.A.; Hao, M.C.; Dayal, U.; Janetzko, H.; Bak, P. Generalized Scatter Plots. *Inf. Vis.* **2009**, *9*, 301–311. [CrossRef]
45. Dai, Z.; Chen, Z.; Selmi, A.; Jermsittiparsert, K.; Denić, N.M.; Nešić, Z. Machine learning prediction of higher heating value of biomass. *Biomass-Convers. Biorefin.* **2021**, *13*, 3659–3667. [CrossRef]
46. Xing, J.; Luo, K.; Wang, H.; Gao, Z.; Fan, J. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* **2019**, *188*, 116077. [CrossRef]
47. Chen, J.; Ding, L.; Wang, P.; Zhang, W.; Li, J.; Mohamed, B.A.; Chen, J.; Leng, S.; Liu, T.; Leng, L.; et al. The Estimation of the Higher Heating Value of Biochar by Data-Driven Modeling. *J. Renew. Mater.* **2022**, *10*, 1555–1574. [CrossRef]
48. Ghugare, S.B.; Tiwary, S.; Elangovan, V.; Tambe, S.S. Prediction of Higher Heating Value of Solid Biomass Fuels Using Artificial Intelligence Formalisms. *Bioenergy Res.* **2013**, *7*, 681–692. [CrossRef]
49. Hosseinpour, S.; Aghbashlo, M.; Tabatabaei, M. 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). *Fuel* **2018**, *222*, 1–10. [CrossRef]

ASSESSING THE PROPERTIES OF *MISCANTHUS X GIGANTEUS* UNDER VARYING LEVELS OF ASH FERTILIZATION TREATMENT AND REGRESSION NEURAL NETWORK INSIGHT INTO CALORIFIC VALUE

Ivan BRANDIĆ[1], Lato PEZO[2*], Neven VOĆA[1] Josip LETO[1], Jona ŠURIĆ[1], Anamarija PETER[1], Nikola BILANDŽIJA[1]

[1] Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia

[2] University of Belgrade Institute of General and Physical Chemistry, Studentski trg 12, Belgrade, Serbia

[*] Corresponding author; E-mail: latopezo@yahoo.co.uk

The aim of the study was to investigate the changes in ultimate, proximate analysis and calorific properties of Miscanthus x Giganteus with three types of planting materials (two rhizomes - R1 and R2 - and one seedling – S) and three ash fertiliser treatments (P_0, P_2, and P_5) were included in the study. The research further examined their effects on crop yield, stem height and various chemical properties. The results showed that the maximum yield was obtained with the R1 x P_2 plant type, while the minimum yield was recorded with the R2 x P_2 plant type. In addition, the greatest average stem height (3.34 m) was recorded for the R2 x P_5 plant type. Significant differences were also found in the chemical components between the plant types and treatments. For example, the highest ash content of 2.25% was found in plant type 'S' x P_5, while the highest coke content of 14.48 % was found in plant type R1 x P_5. The statistical analysis confirmed that planting material and ash fertilisation had significant influence on the physicochemical properties of Miscanthus x Giganteus. This consequently affects the calorific value, with the average higher and lower heating value being 18.32 and 17.04 MJ/kg, respectively. The neural regression network models showed robust predictive performance for the higher (HHV) and lower heating value LHV, with low chi-square values (X^2) and high coefficients of determination (R^2).

Key words*: Miscanthus x Giganteus, fertilisation, energy properties, artificial neural network, modelling.*

1. Introduction

Energy derived from biomass plays a crucial role in achieving the European Union's renewable energy targets for 2030 and beyond. However, this promising sector must manage the complexity of producing, processing and using biomass in a way that is both sustainable and efficient. Key to this strategy is achieving a balance that optimises greenhouse gas mitigation and preserves ecosystem services [1-3]. Compared to seed propagation, vegetative propagation of triploid *Miscanthus x Giganteus* is cost-intensive, making rhizomes the preferred choice for planting material due to their

integral role in vegetative propagation [4]. This biomass source not only has the potential to reduce greenhouse gas and pollutant emissions generally associated with increased fossil fuel use [5], but also offers exceptional opportunities for energy production due to its dense growth [6]. However, to realise the full potential of *Miscanthus x Giganteus* as a bioenergy feedstock, effective nitrogen and harvest management strategies are required [7], highlighting the need for targeted cultivation practises. Importantly, *Miscanthus x Giganteus* also has the added benefit of sequestering carbon in the soil, further contributing to climate change mitigation [8]. Morozova et al., (2020) [9] in research reports the average values of *Miscanthus x Giganteus* in the range of 20.5 – 30.4 t DM/ha in relation to different harvest periods. The increase in the global use of biomass for energy generation has implications for waste management, particularly in terms of the escalating volumes of biomass ash produced. Conventional landfill disposal methods are not only costly, but also result in potentially valuable resources being thrown away [10]. As an alternative, use biomass ash can as a fertiliser [11], which enriches agricultural soils with valuable nutrients, especially if mineral fertilisers are not used. This approach is not only resource-efficient, but also environmentally conscious and carries minimal risk of harmful environmental impacts [12]. Application of fly ash not exceeding 25% of soil weight can strengthen plant biomass while maintaining lower metal(loid) concentrations, potentially improving agricultural yields [13]. As a fertiliser, wood ash provides readily available nutrients such as phosphorus, calcium, magnesium, potassium and boron. It can increase soil pH and concentrations of the main nutrients while reducing the availability of aluminium and less important elements. It also reduces manganese toxicity, which could improve crop yields [14]. Ash in composting improves humification of organic matter and nutrient content, improving compost quality and plant health. It also helps to reduce volatile solids and improve the stability of the compost, increasing its marketability [15]. Ma et al., (2021) [16] notes that *Miscanthus* × *giganteus* shows inconsistent responses to nitrogen fertiliser, possibly influenced by environmental factors, soil types, nitrogen sources, plant age and timing of fertilisation. Fertilisation may possibly affect the associated microbial community in the soil, but the exact mechanisms remain unknown. Smith & Slater, (2010) [17] conducted a study on the effects of organic (cattle and pig manure, chicken litter and unlimed and limed sewage) and inorganic fertiliser (NPK) application on energy crops in Wales, including *Miscanthus x Giganteus*, *Arundo donax* and *Phalaris arundinacea*. The study found that *Miscanthus x Giganteus* responded with increased growth in the second year to all fertilisers applied, with inorganic nitrogen applications being more effective than organic fertilisers. Adjuik et al., (2020) [18] investigated the effects of different fertiliser treatments on biomass yield and greenhouse gas emissions of *Miscanthus x Giganteus* grown on set-aside agricultural land. No significant differences were found between the treatments, which included digestate from the biogas plant, synthetic fertiliser (urea), hydrochar and a control. Due to its robust combustion properties, *Miscanthus x Giganteus* can be used as a biofuel, especially in the form of pellets or briquettes [19]. In recent years, machine learning techniques have gained prominence in the renewable energy production sector, particularly in the area of modelling and prediction [20]. These computational strategies, such as artificial neural networks have been used to improve the prediction of biomass gasification process outcomes [21]

In view of the evidence presented in the above findings, it is intended to further investigate the effects of different planting patterns and different ash treatments on the physicochemical composition

and energy potential of *Miscanthus x Giganteus* biomass. The feasibility of implementing artificial neural network regression models to estimate calorific value will also be evaluated.

2. Materials and methods

2.1. Establishment of the crop and application of ash fertiliser measures

At the University of Zagreb experimental site (Zagreb, Croatia), an experimental field was established to investigate the impact of ash fertilization on the growth dynamics of *Miscanthus x Giganteus*. Three types of plant material were used for the experiment: rhizomes of the Croatian genotype (R1), rhizomes of the English origin (genotype R2) and seedlings of the Polish genotype (S). The rhizomes and seedlings are planted in plots of 4 m x 10 m (40 m^2), while the seedlings are planted in plots of 2.4 m x 10 m (24 m^2). A distance of 3 m is maintained both between plots and between replicates. The experimental design followed a split-split plot scheme with three repetitions, resulting in a total of 27 primary plots. The main factor in the experiment is the type of planting material (R1, R2, S), the sub-factor is the ash fertilisation (P_0, P_2, P_5).

2.2. Physicochemical and calorimetric analysis

From the experimental point of view, the analysis of *Miscanthus x Giganteus* biomass was performed in the laboratory of the University of Zagreb, Faculty of Agriculture, according to standard testing methods. Within the scope of the study, several analyses were performed on the sample. Dry matter analysis (DM) was performed using a Memmert laboratory dryer [22] according to the procedure specified in CEN /TS 14774-2:2009 [23]. Proximate Analysis, which included the evaluation of ash, coke, volatile matter (VM), and Fixed Carbon (FC) concentration, was performed using the method of burning the oven-dry sample in a crucible in a muffle furnace [24] according to EN ISO 18122:2015 [25] and CEN /TS 15148:2009 [26]. Ultimate Analysis encompassed the measurement of carbon (C), hydrogen (H), nitrogen (N), oxygen (O) and sulphur (S) using a Vario Macro CHNS analyzer [27] as described in the standards EN 15104:2011 [28] and EN 15289:2011 [29]. The heating value, in particular the HHV, was determined using an adiabatic bomb calorimeter [30] according to the method CEN /TS 14918:2005 [31].

2.3. Data processing

After the laboratory analyses, the data obtained were analysed using TIBCO Statistica 13.3.0 software (Palo Alto, CA, USA; 2017) [32]. In addition to basic statistical methods, principal component analysis (PCA) was also performed to reduce the dimensionality of the data and identify the most significant variability within the dataset, allowing for a better understanding of hidden structural patterns [33]. In parallel with the previously described methods, a univariate analysis was carried out to determine the influence of parameters such as the type of planting material, ash treatment and their interactive effects on the changes in biomass properties of *Miscanthus x Giganteus*.

2.4. Regression neural network modelling of calorific value

The last part of the research involved building a regression model in the form of an artificial neural network to estimate the energy values (HHV and LHV) of *Miscanthus x Giganteus* biomass based on the input parameters of the ultimate analysis. The first step was to split the data into 70% for learning and 30% for testing the model, which is considered a standard data split [34]. After data preparation, the regression models were built according to the following equation (1) [35]:

$$Y = f_1(W_2 \cdot f_2(W_1 \cdot X + B_1) + B_2) \quad (1)$$

Where Y is the output value; f_1, f_2 - transfer functions of the hidden and output layers; $W_{1,2}$ - weight coefficients of the hidden and output layers; $B_{1,2}$ - hidden and output layer biases.

After calculating the output values, statistical error tests and residual analyses were performed, including Chi-square test (X^2) (2), Root Mean Square Error (RMSE) (3), Mean Bias Error (MBE) (4), Mean Percentage Error (MPE) (5), Sum Squared Error (SSE) (6), Average Absolute Relative Error (AARD) (7) and Coefficient of Determination (R^2) (8) [36,37]:

$$x^2 = \frac{\sum_{i=1}^{N}(x_{p,i} - x_{e,i})^2}{N - n} \quad (2)$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N}(x_{p,i} - x_{e,i})^2\right]^{1/2} \quad (3)$$

$$MBE = \frac{1}{N} \cdot \sum_{i=1}^{N}(x_{p,i} - x_{e,i}) \quad (4)$$

$$MPE = \frac{100}{N} \cdot \sum_{i=1}^{N}\left(\frac{\left|x_{p,i} - x_{e,i}\right|}{x_{e,i}}\right) \quad (5)$$

$$SSE = \sum_{i=1}^{N}(x_{p,i} - x_{e,i})^2 \quad (6)$$

$$AARD = \frac{1}{N} \cdot \sum_{i=1}^{N}\left|\frac{x_{e,i} - x_{p,i}}{x_{e,i}}\right| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[x_i^p - x_i^e\right]^2}{\sum_{i=1}^{n}\left[x_i^p - x^m\right]^2}, x_m = \frac{\sum_{i=1}^{n} x_i^e}{n} \quad (8)$$

Where p in the index and exponent stands for predicted values; e in the index and exponent for experimentally determined values.

The last part of optimizing neural network regression models involved the method of global sensitivity based on the data obtained by artificial neural networks to find the optimal pattern. The Yoon's global sensitivity method was used according to the following formula (9) [38]:

$$RI_{ij}(\%) = \frac{\sum_{k=0}^{n}(w_{ik} \cdot w_{kj})}{\sum_{i=0}^{m}\left|\sum_{k=0}^{n}(w_{ik} \cdot w_{kj})\right|} \cdot 100\% \quad (9)$$

3. Results

3.1. Yield

Figure 1 shows a graphical representation of the yield and average plant height of *Miscanthus x Giganteus* in the study conducted.

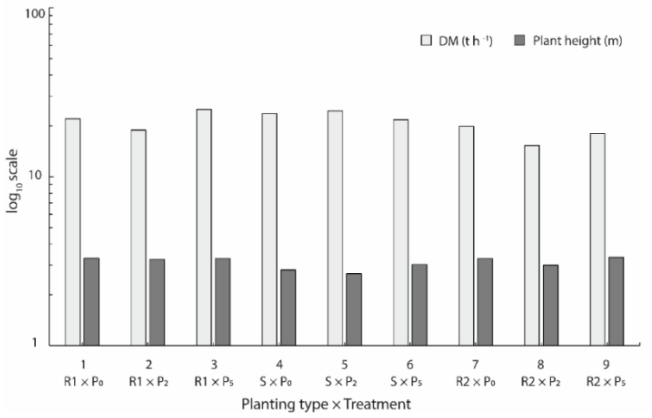

Figure 1. *Miscanthus x Giganteus* yield and plant height regarding planting type and ash treatment

To facilitate plotting yield and plant height variables on the y-axis, a logarithmic scale was used as a method to adjust the resolution of the data in the plot [39].

3.2. Ultimate analysis

The table 1 presents the results of a study examining the impact of different treatments (P_0, P_2, P_5) on three different planting types (R1, S, R2), assessing ultimate analysis.

Table 1. Ultimate analysis of studied biomass of different *Miscanthus x Giganteus* in relation to different planting material and ash treatment

No.	Planting type	Treatment	N (%)	C (%)	S (%)	H (%)	O (%)
1		P_0	0.57 ± 0.17^{a}	50.8 ± 0.55^{a}	0.05 ± 0.02^{a}	5.77 ± 0.06^{a}	42.8 ± 0.61^{c}
2	R1	P_2	0.74 ± 0.15^{ab}	51.17 ± 0.55^{ab}	0.05 ± 0.02^{a}	5.81 ± 0.14^{a}	42.22 ± 0.61^{abc}
3		P_5	0.56 ± 0.16^{a}	51.03 ± 0.79^{ab}	0.11 ± 0.06^{b}	5.86 ± 0.06^{a}	42.45 ± 0.95^{bc}
4		P_0	0.81 ± 0.09^{b}	50.97 ± 0.54^{ab}	0.06 ± 0.03^{ab}	5.85 ± 0.07^{a}	42.31 ± 0.64^{abc}
5	S	P_2	0.8 ± 0.14^{b}	50.96 ± 0.67^{a}	0.07 ± 0.02^{ab}	5.76 ± 0.28^{a}	42.41 ± 0.69^{abc}
6		P_5	0.76 ± 0.07^{ab}	50.87 ± 0.8^{a}	0.07 ± 0.03^{ab}	5.81 ± 0.08^{a}	42.49 ± 0.88^{bc}
7		P_0	0.75 ± 0.18^{ab}	51.52 ± 0.17^{ab}	0.06 ± 0.03^{ab}	5.92 ± 0.04^{a}	41.74 ± 0.28^{ab}
8	R2	P_2	0.71 ± 0.1^{ab}	51.54 ± 0.18^{ab}	0.06 ± 0.02^{a}	5.91 ± 0.01^{a}	41.79 ± 0.15^{ab}
9		P_5	0.75 ± 0.15^{ab}	51.79 ± 0.09^{b}	0.06 ± 0.01^{ab}	5.91 ± 0.06^{a}	41.5 ± 0.21^{a}
Significance			*	*	**	n.s.	*
Min			0.56	50.80	0.05	5.76	41.50
Max			0.81	51.79	0.11	5.92	42.80
Average			0.72	51.18	0.06	5.84	42.19

R1 - Rhizomes of the *Croatian* genotype; R2 - Rhizomes of the *English* genotype; S - seedlings of the *Polish* genotype; P_0= Ash fertilization treatment (0 t/ha); P_2= Ash fertilization treatment (2 t/ha); P_5= Ash fertilization treatment (5 t/ha); Different letters (in columns) indicate difference according to Tukey HSD post hoc test ($p \leq 0.05$); Statistical significance; * $p \leq 0.01$; ** $p \leq 0.05$.

The analysis of the main components of the ultimate analysis variable is shown in figure 2.

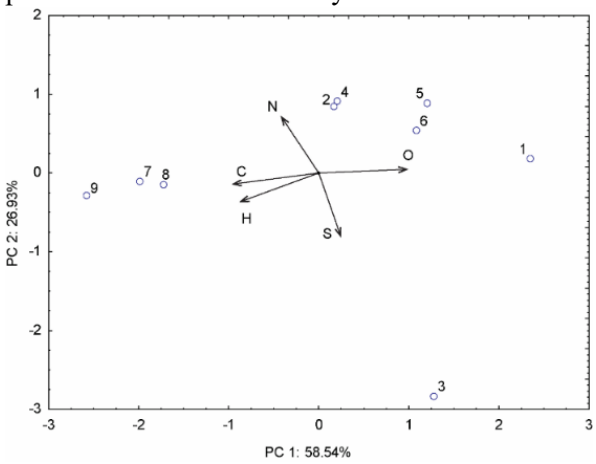

Figure 2. Prinicipal component analysis (PCA) of ultimate analysis

3.3. Proximate analysis and calorific values

Table 2 shows the results of a study that examined the effects of different treatments (P_0, P_2, P_5) on three different types of plants (R1, S, R2), assessing proximate analysis and calorific values.

Table 2. Proximate analysis and calorific values of studied biomass of different *Miscanthus x Giganteus* in relation to different planting material and ash treatment

No.	Planting type	Treatment	Ash (%)	Coke (%)	FC (%)	VM (%)	HHV (MJ/kg)	LHV (MJ/kg)
1		P_0	1.7 ± 0.09^{a}	12.97 ± 0.43^{ab}	10.12 ± 0.41^{a}	79.35 ± 0.62^{b}	18.21 ± 0.34^{a}	16.95 ± 0.33^{a}
2	R1	P_2	1.79 ± 0.09^{ab}	12.21 ± 1.11^{a}	9.34 ± 0.88^{a}	80.06 ± 1.55^{b}	18.42 ± 0.29^{ab}	17.15 ± 0.27^{b}
3		P_5	2.1 ± 0.11^{c}	14.48 ± 2.8^{b}	9.7 ± 0.92^{a}	70.52 ± 13.36^{a}	18.2 ± 0.25^{a}	16.92 ± 0.24^{a}
4		P_0	1.98 ± 0.31^{abc}	13.2 ± 0.57^{ab}	10.08 ± 0.63^{a}	79.3 ± 0.7^{b}	18.29 ± 0.3^{ab}	17.01 ± 0.29^{ab}
5	S	P_2	2.01 ± 0.09^{bc}	13.23 ± 0.54^{ab}	10.03 ± 0.51^{a}	79.01 ± 0.79^{b}	18.28 ± 0.27^{ab}	17.02 ± 0.28^{ab}
6		P_5	2.25 ± 0.38^{c}	12.62 ± 0.98^{ab}	9.21 ± 1.28^{a}	79.36 ± 0.74^{b}	18.11 ± 0.26^{a}	16.84 ± 0.25^{a}
7		P_0	1.81 ± 0.16^{ab}	12.83 ± 0.41^{ab}	9.86 ± 0.46^{a}	79.25 ± 0.29^{b}	18.46 ± 0.22^{ab}	17.16 ± 0.21^{ab}
8	R2	P_2	1.8 ± 0.06^{ab}	12.65 ± 0.84^{ab}	9.75 ± 0.76^{a}	79.83 ± 0.94^{b}	18.28 ± 0.16^{ab}	17 ± 0.16^{ab}
9		P_5	1.78 ± 0.12^{ab}	12.28 ± 1.32^{a}	9.41 ± 1.13^{a}	79.91 ± 1.25^{b}	18.64 ± 0.08^{b}	17.35 ± 0.08^{b}
Significance			*	**	n.s.	*	*	*
Minimum			1.70	12.21	9.21	70.52	18.11	16.84
Maximum			2.25	14.48	10.12	80.06	18.64	17.35
Average			1.91	12.94	9.72	78.51	18.32	17.04

R1 - Rhizomes of the *Croatian* genotype; R2 - Rhizomes of the *English* genotype; S - seedlings of the *Polish* genotype; P_0= Ash fertilization treatment (0 t/ha); P_2= Ash fertilization treatment (2

t/ha); P_5= Ash fertilization treatment (5 t/ha); Different letters (in columns) indicate difference according to Tukey HSD post hoc test ($p \leq 0.05$); Statistical significance; * $p \leq 0.01$; ** $p \leq 0.05$.

Principal component analysis (PCA) of proximate analysis and calorific values for *Miscanthus x Giganteus* is shown in the figure 3.

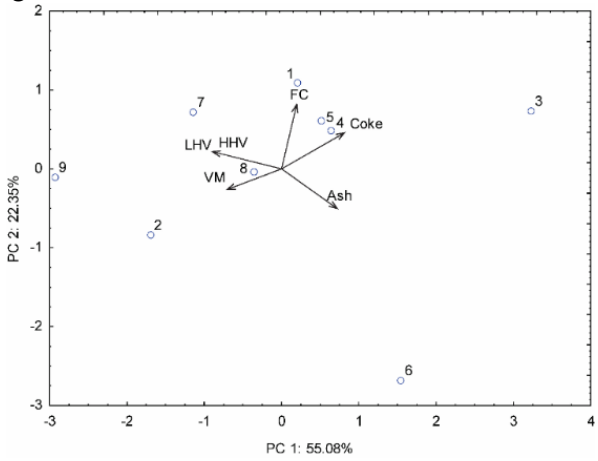

Figure 3. Prinicipal component analysis (PCA) of proximate analysis and calorific values of *Miscanthus x Giganteus*

3.4. Effect of planting material and treatment on changes in the biomass composition of Miscanthus

To study the influence of the parameters of planting type, ash treatment and their interactions on the composition and energy value of biomass, a univariate analysis with the values of the sum of squares for each variable and their statistical significance according to the p coefficient is presented in Table 3.

Table 3. Univariate analysis of the influence of the parameters type of planting material, ash treatment and their interactions on the change in biomass properties *Miscanthus x Giganteus*.

		SoS										
Effect	Df	Ash	Coke	FC	VM	N	C	S	H	O	HHV	LHV
Type	2	1.22*	5.67	0.13	143.88**	0.39*	7.61*	0.00	0.18*	10.82*	0.82*	0.67*
Treatment	2	0.69*	2.65	4.54**	149.99**	0.05	0.29	0.01**	0.01	0.36	0.00	0.01
Type × Treatment	4	0.49**	24.74*	3.47	361.99*	0.15	0.78	0.01**	0.05	1.74	1.02*	1.03*
Error	72	2.56	107.67	49.28	1483.83	1.40	21.40	0.07	1.00	27.43	4.58	4.38

SoS – Sum of squares; Df – Degrees of freedom; FC – content of fixed carbon; VM – content of volatile matter; N – content of nitrogen; C – content of carbon; S – content of sulfur; H – content of hydrogen; O – content of oxygen; HHV – HHV; LHV – lower heating value; Statistical significance; * $p \leq 0.01$; ** $p \leq 0.05$.

3.5. Modelling the heating value of biomass

Tables 4 and 5 show the basic characteristics and performance of the developed models

Table 4. Basic information about the performance of the developed regression model

	Performance		Model error				Activation function	
Output	Train	Test	Train	Test	Train algorithm	Error function	Hidden	Output
HHV	0.999	0.999	0.001	0.002	BFGS 8043	SOS	Tanh	Exp.
LHV	0.972	0.999	0.001	0.004	BFGS 0	SOS	Log.	Iden.

Table 5. Statistical error test and residual analysis of the developed regression models

Model	Output	X^2	RMSE	MBE	MPE	SSE	AARD	R^2	Skew	Kurt	StDev	Var
rNN	*HHV*	0.001	0.032	-0.010	0.063	0.008	0.102	0.964	-2.963	8.846	0.032	0.001
	LHV	0.002	0.043	-0.014	0.152	0.015	0.471	0.927	-1.887	4.951	0.043	0.002

rNN – regression neural network; *HHV* – HHV; *LHV* – lower heating value; X^2- chi squared test; RMSE – root mean square error; MBE – mean bias error; MPE – mean percentage error; SSE – sum squared error; AARD – average absolute relative deviation; R^2 – coefficient of determination; Skew – skewness; Kurt – kurtosis; SD – standard deviation; Var – variance

The rNN regression models for predicting *HHV* and *LHV* show robust performance. Prominent indicators include remarkably low chi-squared (X^2) values (0.001 for *HHV*, 0.002 for *LHV*) and substantial coefficients of determination (R^2 = 0.964 for *HHV*, 0.927 for *LHV*).

After conducting Yoon's sensitivity analysis to determine the relative importance of the input variables on the output values of *HHV* and *LHV*, the influence of each variable of the ultimate analysis on the output value was determined (Figure 4).

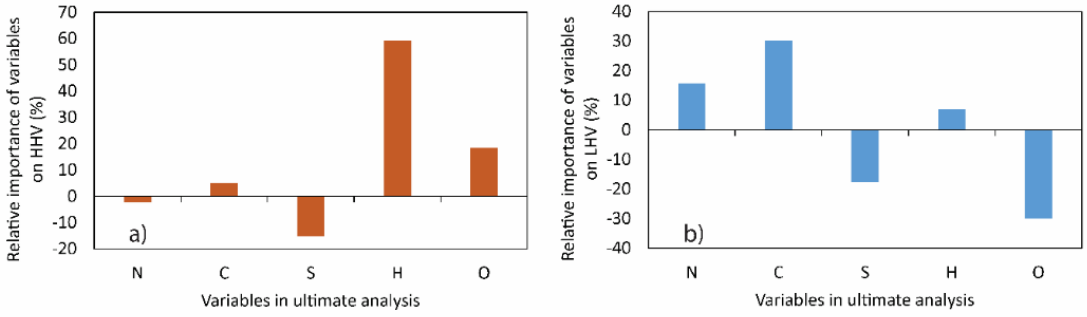

Figure 4. Relative importance (%) of ultimate analysis on the output value of a) *HHV* and b) *LHV*

4. Discussion

The average DM value of the tested samples was 21 t/h, while the average stem height was 3.10 m. In general, ash provides plants with vital substances that can improve plant metabolism, promote root development and improve plant health. The use of ash as fertiliser can increase both fresh mass

and dry mass yield of the plants [40]. With regard to the results, it can be seen that the application of R2 in interaction with P_5 had the greatest effect on stem height and was significantly above the average (3.34 m). Šurić et al. (2022) [41] found in their study that the use of sewage sludge as fertiliser increased the yield of the energy crop Virginia mallow. The application of 6.64 t/h sewage sludge increased the average stem height and dry matter yield from 3.12 m; 6.53 t/ ha to 3.28 m; 8.85 t/ha, compared to the control treatment. The two-year study conducted by Saletnik et al., (2018) [42] showed an 8-68% increase in energy crop yields when biochar, biomass ash and their combination are used as soil amendments to replace classical mineral fertilisers and reinforce organic practises. To determine the properties of the input raw material in the production process, it was necessary to study the physico-chemical and chemical properties of the biomass [43]. The highest proportion of C (51.79%) and H (5.92%) was found in R2 rhizomes in all fertiliser treatments. Voća et al (2021) [44] reported the values for elements of the ultimate analysis *Miscanthus x Giganteus* for C (51.65%), H (6.09%), N (0.18%), S (0.08%), O (42.00%) after laboratory analysis. When comparing the results of the analysis, it was found that the values obtained were within the range of the literature researched. The lowest sulphur content (0.05%) was found when rhizomes from Croatia (R1) were used, i.e. when no ash was used (P_0). Considering the negative impact of sulphur on the environment, it is recommended to use fuel with a lower sulphur content [45]. Anshariah et al., (2020) [46] states that there is a strong correlation between the proportion of FC and the increase in calorific value, i.e. that the increase in FC directly affects the increase in energy values. Although in the study the highest proportion of FC (10.12%) shows that R1 without applying any fertiliser treatment does not have the highest calorific value and is even lower than the average (18.21 MJ/kg), which is also influenced by other variables in the proximate analysis [47]. The highest ash content was found in plant type 'S' under treatment P_5 (2.25%). Gismatulina et al., (2022) [49] gives ash values in the range of 0.90-2.95%. The highest HHV and LHV values (18.64; 17.35 MJ/kg) were found in R2 with the P_5 treatment of ash fertilisation (5 tonnes per hectare). Significant differences were found in VM content between samples, which reached a maximum of 80.06% in R1 plant after P_2 treatment. This result highlights the significant influence of plant type and treatment on critical properties of the plant material, which has potential implications for energy production and various industrial applications. Šurić et al. (2022) [41] reported that no significant differences in ash, coke, fixed carbon and calorific value were found after the application of different sewage sludge fertiliser treatments. However, the application of sewage sludge treatment at a rate of 1.66 t/h resulted in a significant increase in VM. Osman et al. (2018) [50] reported a volatile matter (VM) value of 72.5% and ash content of 3.38% after analysis. The study by Yorgun and Şimşek (2003) [51] reported a biomass composition of 71.4% volatile matter, 18.5% solid carbon, 3.3% ash and 6.8% moisture.

In the final step of the study, an artificial neural network regression model was developed to model the HHV of biomass *Miscanthus x Giganteus*.

When validating the regression neural network (rNN) model, the data was split as standard into 70% for training and 30% for testing to ensure a comprehensive assessment of the models predictive accuracy. The robustness of the model was confirmed by various statistical error tests and residual analyses, including the chi-square test, root mean square error (RMSE) and coefficient of determination (R^2), demonstrating its effectiveness in predicting the higher and lower heating values (HHV and LHV) of *Miscanthus x Giganteus* biomass.

The model used to estimate the HHV showed better performance in training and testing (0.999 and 0.999) in contrast to the model used to estimate the LHV (0.972 and 0.999). Comparing the predictive performance of the rNN model developed in this study with that reported by Noushabadi et al. (2021) [52], a number of observations become clear. The coefficient of determination (R^2) achieved by the rNN model for HHV (0.964) and LHV (0.927) indicates a better fit to the data than the maximum R^2 of 0.96.

The study limitations include a limited sample size and diversity, focusing on specific *Miscanthus x Giganteus* species and ash fertilisation treatments. Its regional focus may not fully represent the different geographical contexts. Future research should investigate how different climates and soils affect Miscanthus x Giganteus, assess the long-term environmental impacts of ash fertilisation, and use advanced technologies to better understand plant-environment interactions. These steps are critical to understanding the plant's role in sustainable biomass production and its environmental impact.

5. Conclusions

In this study of *Miscanthus x Giganteus*, different planting materials and ash fertilizers were found to have different effects on crop yields, growth, and composition. Notably, sample 3 had the highest yield, sample 8 had the lowest yield, and sample 9 had exceptional development with the greatest average stem height. Unfertilized seedlings had elevated nitrogen levels, while R1 types had low sulphur levels under certain conditions. Ash formation was notable in 'S' x P5 plants, while R1 x P5 combinations had high carbon content as evidenced by high coke levels. Energy content, as measured by HHV and LHV, varied in all cases, illustrating the effects of treatments. The artificial neural network (ANN) regression model showed high efficiency in predicting the higher heating value (HHV) and lower heating value (LHV) of *Miscanthus x Giganteus*. The model showed excellent performance metrics with robust coefficients of determination, indicating its potential as a reliable tool for estimating the energy content of biomass.Acknowledgement
The publication was supported by the Croatian Science Foundation, under project No. IP-2018-01-7472 “Sludge management via energy crops production” within the project “Young Researchers’ Career Development Project—Training of Doctoral Students”, and co-financed by the European Union, under the OP “Efficient Human Resources 2014–2020” from the ESF funds.

References

[1] Scarlat, N., *et al*., Brief on biomass for energy in the European Union, *EU Science Hub. EC Publication*, (2019), pp. 1–8.

[2] Ben, F. N., *et al.,* Miscanthus in the European bio-economy: A network analysis, *Industrial Crops and Products*, (2020), 148

[3] Peni, D.*, et al.,* Green biomass quality of perennial herbaceous crops depending on the species, type and level of fertilization, *Industrial Crops and Products*, (2022), 184

[4] Głowacka, K., *et al.,* Genetic variation in Miscanthus × giganteus and the importance of estimating genetic distance thresholds for differentiating clones, *GCB Bioenergy*, (2015), pp. 386–404.

[5] Nakajima, T., *et al.,* Carbon sequestration and yield performances of Miscanthus × giganteus and Miscanthus sinensis, *Carbon Management*, (2018), pp. 415–423
[6] Bilandžija, N., *et al.,* Harvest systems of *Miscanthus x Giganteus* biomass: A review, *Journal of Central European Agriculture*, (2020), pp. 159–167.
[7] Parrish, A. S., *et al.,* Miscanthus × giganteus Responses to Nitrogen Fertilization and Harvest Timing in Illinois, USA, *Bioenergy Research*, (2021), pp. 1126–1135
[8] Littleton, E. W., *et al.,* Uncertain effectiveness of Miscanthus bioenergy expansion for climate change mitigation explored using land surface, agronomic and integrated assessment models, *GCB Bioenergy*, (2023), pp. 303–318
[9] Morozova, I., *et al.,* Assessment of areal methane yields from energy crops in Ukraine, best practices, *Applied Sciences (Switzerland)*, (2020), 10(13)
[10] Silva, F. C., *et al.,* Use of biomass ash-based materials as soil fertilisers: Critical review of the existing regulatory framework, *Journal of Cleaner Production*, (2019), pp. 112–124.
[11] Wójcik, M., *et al.,* The Use of Wood Biomass Ash in Sewage Sludge Treatment in Terms of Its Agricultural Utilization, *Waste and Biomass Valorization*, (2020), pp. 753–768
[12] Maschowski, C., Claudine, M., Bottom ash of trees from Cameroon as fertilizer, *Applied Geochemistry* (2016), pp. 88–90
[13] Yu, C. L., *et al.,* Effects of fly ash application on plant biomass and element accumulations: a meta-analysis, *Environmental Pollution*, (2019), pp. 137–142
[14] Demeyer, A., *et al.,* Characteristics of wood ash and influence on soil properties and nutrient uptake: An overview, *Bioresource Technology*, (2001), pp. 287–295
[15] Asquer, C., *et al.,* Biomass ash reutilisation as an additive in the composting process of organic fraction of municipal solid waste, *Waste Managemen*t, (2017), pp. 127–135
[16] Ma, L., *et al.,* The impact of stand age and fertilization on the soil microbiome of miscanthus × giganteus, *Phytobiomes Journal*, (2021) , pp. 51–59
[17] Smith, R., Slater, F. M., The effects of organic and inorganic fertilizer applications to Miscanthus×giganteus, Arundo donax and Phalaris arundinacea, when grown as energy crops in Wales, UK, *GCB Bioenergy*, (2010), no-no
[18] Adjuik, T., *et al.,* Application of hydrochar, digestate, and synthetic fertilizer to a Miscanthus x Giganteus crop: Implications for biomass and greenhouse gas emissions, *Applied Sciences (Switzerland),* (2020), pp. 1–21
[19] Oros, A. E. D., *et al.,* Miscanthus Giganteus Biomass for Sustainable Energy in Small Scale Heating Systems, *Agriculture and Agricultural Science Procedia*, (2015), pp. 538–544
[20] Donti, P. L., Kolter, J. Z. Machine Learning for Sustainable Energy Systems, *Annual Review of Environment and Resources*, (2021), pp. 719–747
[21] Elmaz, F., *et al.,* Predictive modeling of biomass gasification with machine learning-based regression methods, *Energy*, (2020), 191, 116541.
[22] Memmert model 30-1060, Memmert GmbH + Co. KG, Schwabach, Germany, https://www.memmert.com/
[23] CEN/TS 14774-2:2009, Solid Biofuels—Determination of Moisture Content; CEN: Brussels, Belgium, (2009), https://standards.iteh.ai/
[24] Muffle furnace, Nabertherm Controller B170, Lilienthal, Germany, https://nabertherm.com/

[25] EN ISO 18122:2015, Solid Biofuels, Determination of Ash Content. ISO: Geneva, Switzerland, 2015, https://www.iso.org/
[26] CEN/TS 15148:2009, Solid Biofuels—Determination of the Content of Volatile Matter. CEN, Brussels, Belgium, 2009, https://standards.iteh.ai/
[27] Vario Macro CHNS analyzer, Elementar Analysensysteme GmbH, Meerbusch, Germany, https://www.elementar.com/
[28] EN 15104:2011, Solid Biofuels, Determination of Total Content of Carbon, Hydrogen and Nitrogen—Instrumental Methods, SIS, Stockholm, Sweden, (2011), https://standards.iteh.ai/
[29] EN 15289:2011, Solid Biofuels, Determination of Total Content of Sulfur and Chlorine, ISO, Geneva, Switzerland, 2011, https://standards.iteh.ai/
[30] Adiabatic bomb calorimeter, IKA Analysentechnik GmbH, Meerbusch, Germany, https://www.ika.com/
[31] CEN/TS 14918:2005, Solid Biofuels—Method for the Determination of Calorific Value, CEN, Brussels, Belgium, (2005), https://standards.iteh.ai/
[32] TIBCO Statistica, v. 13.3.0, TIBCO Software Inc, Palo Alto, CA, USA, 2017, https://www.tibco.com/products/tibco-statistica
[33] Jolliffe, I. T., Cadima, J. Principal component analysis: a review and recent developments, *Philos Trans A Math Phys Eng Sci*, (2016), 13,374.
[34] Nguyen, Q. H., *et al.,* Influence of data splitting on performance of machine learning models in prediction of shear strength of soil, *Mathematical Problems in Engineering*, (2021)
[35] Pezo, L. L., *et al.,* Artificial neural network model of pork meat cubes osmotic dehydratation, *Hemijska Industrija,* (2013), pp. 465–475
[36] Brandić, I., *et al.,* Comparison of Different Machine Learning Models for Modelling the HHV of Biomass, *Mathematics*, (2023), 11(9)
[37] Rajković, D., *et al.,* Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models, *Agronomy*, (2022), 12(1)
[38] Yoon, Y., *et al.,* A comparison of discriminant analysis versus artificial neural networks, J. *Oper. Res. Soc.,* (2017), pp. 51–60
[39] Urzhumtsev, A., *et al.,* On the use of logarithmic scales for analysis of diffraction dana, *Acta Crystallographica Section D: Biological Crystallography*, (2009), pp. 1283–1291
[40] Wierzbowska, J., *et al.,* Environmental application of ash from incinerated biomass, *Agronomy*, (2020), 10(4)
[41] Šurić, J., *et al.,* Wastewater Sewage Sludge Management via Production of the Energy Crop Virginia Mallow, *Agronomy*, (2022) , 12(7)
[42] Saletnik, B., *et al.,* Biochar and biomass ash as a soil ameliorant: The effect on selected soil properties and yield of giant miscanthus (Miscanthus x giganteus), *Energies*, (2018), 11(10)
[43] Esteves, B., *et al.,* Influence of Chemical Composition on Heating Value of Biomass: A Review and Bibliometric Analysis, Energies, (2023), pp. 1–17
[44] Voća, N., *et al.,* Energy properties and biomass yield of *Miscanthus x Giganteus* fertilized by municipal sewage sludge, Molecules, (2021), 26(14)
[45] Choi, J.S., *et al.,* Improving the Analysis of Sulfur Content and Calorific Values of Blended Coals with Data Processing Methods in Laser-Induced Breakdown Spectroscopy, *Applied Sciences (Switzerland)*, (2022), 12(23)

[46] Anshariah, I. A. M., *et al.,* Correlation of fixed carbon content and calorific value of South Sulawesi Coal, Indonesia. IOP Conference Series: Earth and Environmental Science, 2020, Vol. 473(1)
[47] Nhuchhen, D.R., Afzal, M.T., HHV predicting correlations for torrefied biomass using proximate and ultimate analyses, *Bioengineering*, (2017), 4(1)
[49] Gismatulina, Y.A., *et al.,* Evaluation of Chemical Composition of Miscanthus × giganteus Raised in Different Climate Regions in Russia, *Plants*, (2022), 11(20)
[50] Osman, A. I., *et al.,* Physicochemical characterization of miscanthus and its application in heavy metals removal from wastewaters, *Environmental Progress and Sustainable Energy*, (2018), pp. 1058–1067
[51] Yorgun, S., Şimşek, Y. E., Fixed-bed pyrolysis of Miscanthus x Giganteus: Product yields and bio-oil characterization, *Energy Sources*, (2003), pp. 779–790
[53] Noushabadi, A. S., *et al.,* Estimation of HHVs (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation, *Renewable Energy*, (2021), pp. 550-562

ENERGY SOURCES, PART A: RECOVERY, UTILIZATION, AND ENVIRONMENTAL EFFECTS
2024, VOL. 46, NO. 1, 2842–2854
https://doi.org/10.1080/15567036.2024.2309303

Biomass higher heating value prediction machine learning insights into ultimate, proximate, and structural analysis datasets

Ivan Brandić[a], Neven Voća[a], Jerko Gunjača[b], Biljana Lončar[c], Nikola Bilandžija[d], Anamarija Peter[a], Jona Šurić[a], and Lato Pezo[e]

[a]Department of Sustainable Technologies and Renewable Energy, University of Zagreb Faculty of Agriculture, Zagreb, Croatia; [b]Department of Plant Breeding, Genetics and Biometrics, University of Zagreb Faculty of Agriculture, Zagreb, Croatia; [c]Department of Chemical Engineering, University of Novi Sad Faculty of Technology, Novi Sad, Serbia; [d]Department of Mechanization and Autonomous Systems in Agriculture, University of Zagreb Faculty of Agriculture, Zagreb, Croatia; [e]Engineering Department, University of Belgrade Institute of General and Physical Chemistry, Belgrade, Serbia

ABSTRACT

In this study machine learning (ML) models have been employed to predict the higher heating value (*HHV*) of biomass by utilizing input variables derived from ultimate, proximate, and structural analyses. In total, 180 models were developed, with 124 utilizing ultimate analysis data, 28 based on proximate analysis, and 28 relying on structural analysis. Various ML techniques, including polynomial models (SOP), support vector machines (SVM), random forest regression (RFR), and artificial neural networks (ANN), were employed for analysis. The study found that ANN models, when "fed" with FC and VM data, provided considerable accuracy in prediction results, with the best results obtained with 2-12-1 architecture ($R^2 = 0.96$). In addition, a separate model configuration that processed inputs on biomass constituents such as cellulose, lignin, and hemicellulose showed remarkable agreement with empirical data. Additional findings revealed that the models created using SOP ($R^2 = 0.95$), SVM ($R^2 = 0.95$), and RFR ($R^2 = 0.90$) demonstrated minimal discrepancies when predicting *HHV*. This study provides significant insights into the investigation of biomass analysis techniques employing ML tools, paving the way for future research aimed at constructing a robust tool for *HHV* prediction. Subsequent models may explore integrating inputs from diverse analysis methods and leveraging advanced machine learning techniques to enhance accuracy further.

ARTICLE HISTORY
Received 27 November 2023
Revised 8 January 2024
Accepted 19 January 2024

KEYWORDS
Energy properties; mathematical modeling; ultimate analysis; proximate analysis; structural analysis

Introduction

Agriculturally sourced biomass is gaining recognition as a significant renewable energy potential with substantial production capabilities. Its prominence in the energy sector is growing, as indicated by Bilandzija et al. (2018) and Cholewa et al. (2022). Biomass-derived energy stands as a crucial alternative energy source, as emphasized by Kołodziej, Pudełko, and Mańkowski (2023). By harnessing biomass in solid, liquid, or gaseous states, there is potential to substitute traditional energy sources and consequently diminish exhaust emissions, as highlighted by Gironès et al. (2017) and Tursi (2019). Lignocellulosic biomass, which consists mainly of agricultural residues, forestry by-products and various organic wastes, is at the forefront of renewable energy solutions due to its role in energy production (Das et al. 2021). The conversion of this biomass, especially from agricultural and forestry sources, into fuel is facilitated by a variety of technological processes (Li et al. 2021). A key parameter that defines the energy

CONTACT Neven Voća nvoca@agr.hr Department of Sustainable Technologies and Renewable Energy, University of Zagreb Faculty of Agriculture, Svetosimunska cesta 25, Zagreb, Croatia
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15567036.2024.2309303

potential of biomass fuel is its HHV, which measures the total energy output during combustion (Soponpongpipat, Sittikul, and Sae-Ueng 2015). ML algorithms are used for modeling the energetic properties of biomass and the calorific value, especially when input variables derived by empirical methods and obtained by chemical analysis are used (Dai et al. 2021). Access to extensive datasets is pivotal for data cleaning and analysis. Consequently, the key objective of the algorithms mentioned above is to enhance the efficiency of specific tasks, such as modeling, by utilizing pre-trained data. A fundamental feature and benefit of machine learning algorithms is their ability to establish meaningful (non-linear) correlations between input and output data, facilitating predictive analysis, as highlighted by Benos et al. (2021). Considering the emergence and need to analyze large datasets, "Big data" is emerging, which as more complex datasets represent a revolution in predictive analytics with more applications in agricultural practice (Ang and Seng 2021). Over time, many ML tools have been developed, among which the RFR models stand out as robust solutions for modeling multidimensional datasets. The algorithm works on the principle of nonlinear predictors that use a statistical technique to predict the desired output value (Matin and Chelgani 2016). Support vector machines (SVM) are used as machine learning models with the main objective of performing multitasking tasks while finding the optimal solution (Ranganayaki and Deepa 2019. As a learning method, SVM can be used for regression through its associated algorithm (Taki and Rohani 2022). As for regression, Bayesian schemes and kernel methods are be-coming increasingly popular as statistical tools and machine learning techniques (Martino and Read 2021). In supervised learning, models undergo training using input data and predetermined output values, which can be continuous in the context of regression models. The objective is to develop a model with minimal errors, subsequently validating it to assess its reliability, as highlighted by Tipping (2001). Artificial Neural Network (ANN) models excel at analyzing and modeling a diverse array of datasets. They are trained to discern patterns that result in non-linear relationships and correlations between inputs and outputs, as noted by Li et al. (2021). As data-driven, nonparametric models, ANNs can capture nuanced, unfamiliar functional relationships within empirical data. ANN serves as a computational approach mirroring the learning capabilities of the human brain and stands as a dependable tool for predictive modeling applications. Among ANN architectures, Multilayer Perceptron (MLP) models have proven to be particularly effective in handling intricate data due to their capacity for learning through non-linear connections. In practical applications, they have demonstrated remarkable efficacy in estimating biomass energy values, as evidenced by studies conducted by Çakman, Gheni, and Ceylan (2021) and Ighalo, Adeniyi, and Marques (2020). The mentioned model type showed high performance in modeling HHV biomass in research, reaching $R^2 = 0.77$ (Brandić et al. 2022). The authors also compared the performance of this model with 10 linear models used in various earlier studies.

Statistical parameters like the chi-square test, RMSE, MBE, and MPE indicate a lower error level in the developed model compared to the referenced ones. Matin and Chelgani (2016) employed the random forest method to estimate the calorific value of fuels, achieving satisfactory results, particularly when using the final analysis dataset, which proved to be the best fit. García Nieto et al. (2019) developed diverse models for HHV biomass estimation, underscoring the significance of incorporating hybrid SVM models and optimization techniques to enhance modeling performance. In a recent study, Xiaorui, Jiamin, and Longji (2023) improved the model's efficiency by integrating the radial basis function (RBF) into the SVM model, achieving an impressive R^2 value exceeding 0.91.

In the context of the literature review, this research aims to discern disparities in HHV modeling across various ML models, including SOP, SVM, RFR, and ANN. The study seeks not only to identify the most suitable dataset but also to enhance model performance through the refinement and consolidation of input variables in these models. A key goal is to examine the effectiveness of these ML models in estimating HHV biomass, focusing on identifying the most appropriate dataset in relation to the collected data. Additionally, the research will compare differences in accuracy or modeling error when input variables are reduced and

combined without repeating input variables, further contributing to the understanding of ML model optimization in HHV biomass estimation. This study addresses the notable gap in existing research regarding the comprehensive comparison and optimization of machine learning models for the prediction of biomass HHV, focusing on the integration of different input parameters from terminal, proximate and structural analyses. Despite the advances in ML applications for biomass analysis, there is a clear need for a systematic investigation to identify the most effective model configurations and datasets.

Materials and methods

Data gathering

The ML model was developed using data obtained from scientific sources, specifically the structural, ultimate, and proximate analyses gathered from the Web of Science database. Detailed information, including composition and understanding of the energetic properties of biomass, is presented in Supplementary Table S1, with references provided. The data is categorized into agricultural and forest biomass, providing a clear description of their composition, as well as their energetic properties, as outlined by Collard and Blin (2014). A total of 292 biomass samples were collected for proximate analysis, 249 samples for ultimate and 286 samples for structural analysis. The higher heating value served as the output parameter for all the models developed in the study.

Data manipulation

The statistical analysis involved calculating the mean and standard deviation of the categorized data, while the analysis of variance was employed to assess the differences between the observed categories. Data processing was carried out using the TIBCO STATISTICA program package and the Python programming language, along with the corresponding data.

Preparation of data for non-linear modeling

Data preparation for non-linear modeling followed the initial research phase, where collected data underwent a cleaning process to eliminate incomplete entries and standardize it for modeling, as discussed by Cvetkov-Iliev, Allauzen, and Varoquaux (2022) and Khayyaty et al. (2015). Ge et al. (2017) state that data analysis and cleaning is a key stage in discovering the relationship between each data in a set, but also a preparation for processing with different algorithms (Ge et al. 2017; Qiu et al. 2016). The research employed a combinatorial approach without repeated linkages to determine the total number of possible combinations of input variables in the models and the corresponding number of models created. The quantity of models was determined by considering the total number of variables and the potential combinations of specific chemical analyses related to biomass, such as ultimate (C, H, N, S, and O), proximate (FC, VM, and Ash), and structural analysis (Lig, Cel, Hem). In total, 180 non-linear models were developed using different techniques, namely SOP, ANN, RFR, and SVM. Among these models, 124 were based on ultimate input data, 28 on proximate analysis, and 28 on structural analysis. The initial values for all models were derived from *HHV* data (refer to Table 1). Prior to constructing the models, the data for each model were split into training (70%) and testing (30%) sets, using a neural network over 100,000 cycles. To enhance comprehension of the research findings and the potential application of diverse non-linear models for estimating biomass *HHV*, a schematic representation of the formulated models is provided in Supplementary Table S2.

Table 1. Mean values and standard deviations of the variables from structural, proximate and ultimate analysis of agricultural and wood biomass.

Dataset	Variable	Agricultural biomass	Wood biomass	Statistical significance
Structural analysis	Cel (%)	44.70 ± 13.40	44.04 ± 7.63	n.s.
	Lig (%)	15.08 ± 7.87	27.64 ± 7.04	*
	Hem (%)	23.21 ± 5.81	27.25 ± 4.41	*
	HHV (MJ kg $^{-1}$)	17.67 ± 1.66	19.68 ± 0.62	*
Proximate analysis	FC (%)	14.60 ± 13.90	20.10 ± 6.52	n.s.
	VM (%)	75.13 ± 17.49	74.36 ± 16.71	n.s.
	ASH (%)	5.79 ± 12.13	5.59 ± 17.15	n.s.
	HHV (MJ kg $^{-1}$)	18.23 ± 3.3	18.98 ± 3.41	n.s.
Ultimate analysis	C (%)	47.61 ± 7.76	48.13 ± 9.35	n.s.
	H (%)	5.52 ± 1.47	5.37 ± 1.5	n.s.
	N (%)	1.27 ± 1.12	0.70 ± 0.57	**
	S (%)	0.26 ± 0.22	0.33 ± 0.46	n.s.
	O (%)	39.87 ± 13.36	42.75 ± 12.46	n.s.
	HHV (MJ kg $^{-1}$)	18.73 ± 3.04	18.89 ± 3.47	n.s.

Cel – cellulose; Lig – lignin; Hem – Hemicellulose; HHV – higher heating value; FC – fixed carbon; VM – volatile matter; C – carbon; H – hydrogen; N – nitrogen; S – sulfur; O – oxygen; n.s. – not significant. Statistical significance: * p ≤ .01; ** p ≤ .05.

Artificial neural networks (ANNs)

Artificial neural networks exhibit excellent accuracy in predicting biomass higher heating value (*HHV*) by optimizing weight coefficients and biases corresponding to specific input variables (Darvishan et al. 2018). The number of nodes in the hidden layer is randomly determined to identify the optimal solution for the output value. The fundamental equation governing ANN output is represented in Equation 1, as detailed in studies by Pezo et al. (2013) and Kollo and Von Rosen (2005):

$$Y = f_1(W_2 \cdot f_2(W_1 \cdot X + B_1) + B_2) \tag{1}$$

where Y represents the output value, f_1 and f_2 correspond to the transfer function in the hidden and output layer, X represents the input matrix for initial layer.

In this study, an ANN model was developed and fine-tuned through 100,000 iterations to avoid random correlations and achieve optimal performance with nine hidden neurons. The ANN was trained for 100 epochs, achieving stable accuracy before the 50th epoch to prevent overfitting.

Random forest regression (RFR)

RFR, a type of machine learning algorithm, is utilized in supervised regression tasks and excels in nonlinear modeling, demonstrating its effectiveness even in the presence of outliers, as highlighted in studies by Li et al. (2018) and Scornet, Biau, and Vert (2015). Schonlau and Zou (2020) found that RFR models were suitable for predictions with medium and large data sets. The RFR models were constructed with the data split 70% for training and 30% for testing the model and 10,000 random trees. Additionally, entropy was calculated for each internal node within the decision trees using the formula (Eq. 2):

$$E = -\sum_{i=1}^{c} p_i \times \log\,(p_i) \tag{2}$$

where c represents the number of unique classes and pi prior probability of each given class.

Second order polynomials

Models in the form of second-order polynomials (SOP) are a suitable tool for finding relationships between variables by estimating the desired output value (Ostertagová 2012). Non-linear models in the form of second-order polynomials, incorporating quadratic, linear, and combined components, are developed to predict the higher heating value of agricultural and forest biomass based on specific input variables.

Statistical tests of the model fit

Statistical tests are employed to assess the model's accuracy concerning the alignment between real data and the data predicted by the model. Utilizing the fit allows for a comprehensive comparison of all models in terms of their performance, enabling the identification of the best-fitting model, as demonstrated in studies by Bakshaev and Rudzkis (2017) and Maydeu-Olivares (2017). When evaluating the effectiveness of higher heating value modeling, several key statistical parameters were computed, including the Chi-square test(X^2) (Equation 4), mean bias error (MBE) (Equation 5), mean percentage error (MPE) (Equation 6), root mean square error (RMSE) (Equation 7) and absolute average relative deviation (AARD) (Equation 8) (Arsenović et al. 2015):

$$x^2 = \frac{\sum_{i=1}^{N} \left(x_{pre,i} - x_{\exp,i}\right)^2}{N - n} \tag{3}$$

$$MBE = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(x_{pre,i} - x_{\exp,i}\right) \tag{4}$$

$$MPE = \frac{100}{N} \cdot \sum_{i=1}^{N} \left(\frac{\left| x_{pre,i} - x_{\exp,i} \right|}{x_{\exp,i}} \right) \tag{5}$$

$$RMSE = \left[\frac{1}{N} \cdot \sum_{i=1}^{N} \left(x_{pre,i} - x_{\exp,i}\right)^2 \right]^{1/2} \tag{6}$$

$$AARD = \frac{1}{N} \cdot \sum_{i=1}^{N} \left| \frac{x_{\exp,i} - x_{pre,i}}{x_{\exp,i}} \right| \tag{7}$$

where $x_{exp,i}$ is the experimental values and $x_{pre,i}$ is the predicted values calculated by the model, N and n are the number of observations and constants, respectively.

Yoon method of global sensitivity

Sensitivity analysis offers a deeper understanding of the ANN model, shedding light on the interactions between input and output variables, as emphasized by Fernández-Navarro et al. (2016). In this study, the sensitivity method proposed by Yoon (Eq. 9) was employed, following the approach outlined by Yoon, Swales, and Margavio (2017), to discern these relationships.:

$$RI_{ij}(\%) = \frac{\sum_{k=0}^{n} (w_{ik} \cdot w_{kj})}{\sum_{i=0}^{m} \left| \sum_{k=0}^{n} (w_{ik} \cdot w_{kj}) \right|} \cdot 100\% \quad (8)$$

where w - denotes the weighting factor in the ANN model, i - input variable, j - output variable, k - hidden neuron, n - number of hidden neurons, m - number of inputs.

The architecture (a) and flowchart (b) of the research conducted are shown in Supplementary Figure S1.

Results

Table 1 show the results of the composition and characteristic analysis of agricultural and wood biomass after statistical processing.

Table 1 displays the outcomes of the statistical analysis conducted on the composition, chemical characteristics, and heating value of biomass. The statistical analysis revealed that the disparity in cellulose content between agricultural and wood biomass was not statistically significant. However, substantial differences were observed in lignin and hemicellulose content between the two types of biomasses. Interestingly, the proximate analysis of agricultural and wood biomass did not show any statistically significant differences in the observed variables. Similarly, when comparing the ultimate analysis variables between agricultural and woody biomass, no significant statistical distinctions were found, except for the nitrogen content. Specifically, agricultural biomass exhibited a significantly higher average proportion (1.27%) at a significance level of $p \leq .05$ in this case.

Table 2 shows the performance of the most efficient ANN models in terms of characteristic analyses and the number of input parameters in the models.

Figure 1 shows a series of scatter plots comparing the predicted HHV with the target HHV for different biomass samples, using different modeling approaches. The rows correspond to the different types of datasets used to train and predict the model: structural analysis dataset, proximate analysis dataset and ultimate analysis dataset.

Table 3 displays the results of the model fit statistical tests for the most effective models, including ANN, SOP, SVM, and RFR, considering typical input variables and the number of inputs derived from ultimate, proximate, and structural analyses.

As shown in Table 3, the ANN model achieved the best results based on the ultimate analysis dataset, as shown by the R^2 (0.90), followed by SOP ($R^2 = 0.82$), SVM ($R^2 = 0.81$) and RFR with R^2 .76).

Table 2. Performance of developed ANN, RFR and SVM models based on different input parameters.

Model	Inputs	Net. name	Train. perf (R^2).	Test perf. (R^2)	Train. error	Test error	Train. algorithm	Error function	Hidden act.	Output act.
ANN	CHNSO	MLP 5-11-1	0.87	0.93	0.49	0.54	BFGS 123	SOS	Tanh	Tanh
	FC VM	MLP 2-12-1	0.96	0.96	0.19	0.23	BFGS 187	SOS	Tanh	Tanh
	Cel Lig Hem	MLP 3-9-1	0.92	0.86	0.12	0.16	BFGS 254	SOS	Tanh	Exponential
RFR	CH	-	0.71	0.83	-	-	-	-	-	-
	FC VM ASH	-	0.91	0.89	-	-	-	-	-	-
	Lig Hem	-	0.83	0.79	-	-	-	-	-	-
SVM	CHNSO	-	0.77	0.89	-	-	-	-	-	-
	FC ASH	-	0.96	0.95	-	-	-	-	-	-
	Cel Lig	-	0.75	0.71	-	-	-	-	-	-

ANN – Artificial neural network; MLP – Multi Layer Perceptron; RFR – Random forest regression; SVM – Support vector machine; C – Content of carbon (%); H – content of hydrogen (%); N – Content of nitrogen; S – Content of sulfur; O – Content of oxygen; FC – Content of fixed carbon; VM – Content of volatile matter; Cel – Content of cellulose; Hem – Content of hemicellulose; Lig – content of lignin; Net. – Network; Train – Training; Test. – Testing.

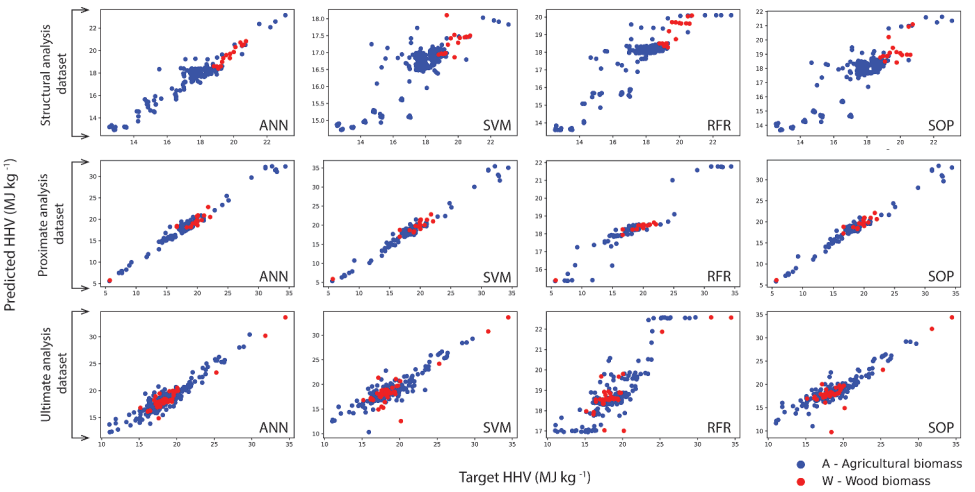

Figure 1. X-Y scatterplot for models based on input variables of structural, proximate and ultimate analysis for ANN, SVM, RFR and SOP model.

Table 3. Goodness of fit for the most efficient models.

										Residual analysis			
Analysis	Model	Inputs	χ2	RMSE	MBE	MPE	SSE	AARD	R^2	Skew	Kurt	SD	Var
Ultimate	ANN	CHNSO	1.02	1.01	0.06	4.21	251.19	196.44	0.90	−0.46	1.23	1.01	1.01
	SOP	CHNO	1.77	1.33	0.10	5.08	437.11	237.51	0.82	0.94	8.10	1.33	1.76
	SVM	CHNSO	1.86	1.36	0.01	5.24	462.32	238.01	0.81	0.65	5.49	1.37	1.86
	RFR	CH	4.49	2.11	−0.03	8.05	1113.54	379.07	0.76	1.07	5.76	2.12	4.49
Proximate	ANN	FC VM	0.41	0.64	0.03	2.65	118.33	240.27	0.96	−0.48	1.53	0.64	0.41
	SOP	FC ASH	0.52	0.72	0.02	3.01	150.43	339.65	0.95	0.08	2.26	0.72	0.52
	SVM	FC ASH	0.62	0.79	0.12	3.47	177.33	366.24	0.95	−1.17	2.99	0.78	0.61
	RFR	FC VM ASH	6.56	2.56	0.03	8.80	1907.50	505.41	0.90	0.85	9.09	2.56	6.55
Structural	ANN	Cel Lig Hem	0.26	0.51	0.00	2.30	74.81	193.72	0.91	−1.09	3.34	0.51	0.26
	RFR	Lig Hem	0.54	0.73	0.00	3.19	153.37	275.97	0.82	−0.46	2.91	0.73	0.54
	SOP	Cel Lig Hem	0.61	0.78	−0.01	3.22	172.78	262.14	0.79	−0.51	3.37	0.78	0.61
	SVM	Cel Lig	2.68	1.63	1.16	8.13	376.45	255.21	0.74	−0.80	2.16	1.15	1.32

C – content of carbon; H – content of hydrogen; N – content of nitrogen; S – content of sulfur; O – content of oxygen; FC – content of fixed carbon; VM – content of volatile matter; Cel – content of cellulose; Hem – content of hemicellulose; Lig – content of lignin; ANN – Artificial neural network; SVM – Support vector machine; RFR – Random forest regression; Skew – Skewness; Kurt – Kurtosis; SD – Standard deviation; Var – Variance.

Looking at the amount of modeling error (RMSE and MBE), the ANN model also showed the best performance in the proximate analysis.

The SOP model had the second highest R^2 value of 0.95, followed by the SVM model with an R^2 value of 0.95 and the RFR model with an R^2 value of 0.90. In the structural analysis, the ANN model had the best results with an R^2 value of 0.91 and the lowest RMSE value of 0.51. The RFR model had the second highest R^2 value of 0.82, followed by the SOP model with an R^2 value of 0.79 and the SVM model with an R^2 value of 0.74.

Supplementary Table S3 shows the statistical tests of the model fit for all models developed.

In Figure 2, the diagram illustrates Yoon's global sensitivity method for ANN models, emphasizing the relative importance of input parameters from ultimate analysis (C, H, N, S, and O), proximate analysis (FC, VM), and structural analysis (Cel, Lig, Hem) concerning the output value of higher heating value.

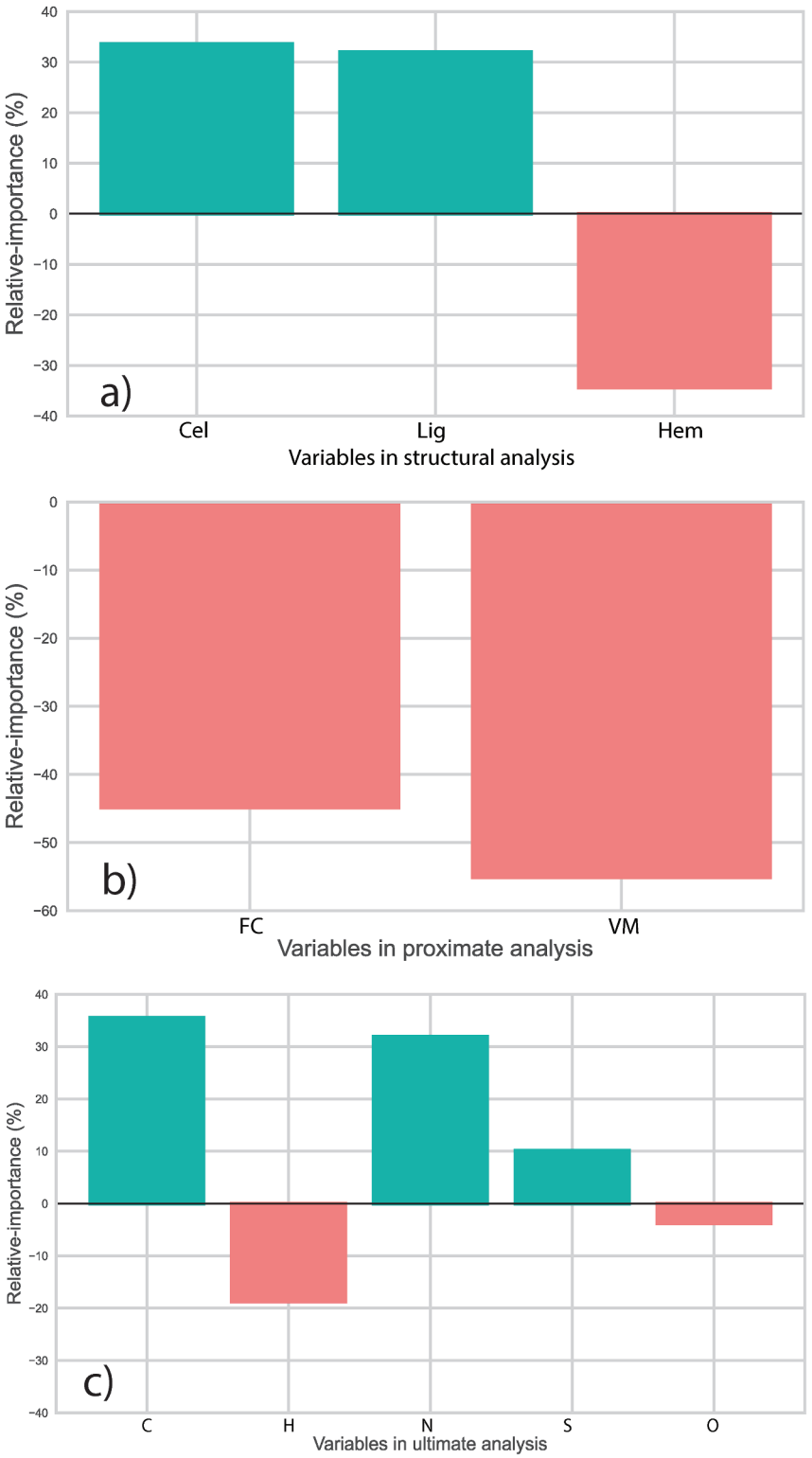

Figure 2. Yoon's method of global sensitivity of ANN models based on input parameters (a) ultimate analysis (b) proximate analysis and (c) structural analysis.

Discussion

For agricultural biomass the average share of cellulose is 44.70%, and for wood biomass 44.04%. The proportion of lignin and hemicellulose is larger on average for wood biomass (27.64% and 27.25%), while the proportion is lower for agricultural biomass (15.08% and 23.21%). Forest biomass has a *HHV* of 19.68 MJ kg −1, while agricultural biomass is 17.67 MJ kg $^{-1}$. Mansor et al. (2019) gives the values of hemicellulose 32–37%, lignin 19–22%, and cellulose 30–42% whereas agricultural biomass has an *HHV* value of 17.67 MJ kg^{-1}, while Callejón-Ferre et al. (2014) mention average *HHV* values in the range of 12.60–17.01 MJ kg $^{-1}$ for different types of biomasses.

The share of FC in agricultural biomass is 14.60%, while in wood biomass it is 20.10% in wood biomass. The average percentage of VM and ash is higher in agricultural biomass (75.13% and 5.79%, respectively), while the percentage is lower in wood biomass (74.36% and 5.59%, respectively). The *HHV* of agricultural biomass is 18.23 MJ kg $^{-1}$, while that of wood biomass is 18.89 MJ kg $^{-1}$. Voća et al. (2021) report the mean values of FC (9.78%), VM (81.88%), and ash (1.91%) for Miscanthus biomass. The authors also give the mean value of *HHV* (17.78 MJ kg $^{-1}$).

The average percentage of C in agricultural and wood biomass is 47.61% and 48.13%, respectively. The values for H and N are higher on average in agricultural biomass (5.52% and 1.27%), while they are lower in wood biomass (5.37% and 0.70%). The share of S in agricultural biomass is 0.26%, while it is slightly higher in wood biomass (0.33%). The average value of O in wood biomass is 47.75%, while in agricultural biomass it is 39.87%. The *HHV* has an average higher value for forest biomass (18.89 MJ kg $^{-1}$), as in the other observed analyses, while the *HHV* value for agricultural biomass is 18.73 MJ kg $^{-1}$. The ultimate and calorimetric analysis of various biomass fuels were studied by Ismaila, Abdullahi, and Garba (2013). Authors reported that C content ranged from 35.92% to 75.98%, S content from 0.01% to 0.11%, H content from 2.95% to 6.10%, N content from 0.70% to 3.50%, O content from 14.50% to 56.78%, and *HHV* from 11.03 MJ kg^{-1} to 28.92 MJ kg^{-1} based on the Boie equation.

Table 2 displays the developed ANN models, considering inputs from ultimate, proximate, and structural analyses. The ANN model created using input values of C, H, N, S, and O demonstrated superior performance with a network structure comprising 5 artificial neurons in the input layer, 11 in the hidden layer, and 1 in the output layer. This model exhibited strong *HHV* prediction capabilities, boasting a coefficient of determination of 0.87 for training and 0.93 for testing, while the training error was 0.49 and the testing error was 0.54. Noushabadi et al. (2021) conducted a study aimed at estimating the HHV of different types of biomass fuels based on the input variables of elemental analysis. After evaluation for the developed ANN model, it showed high performance in modeling ($R^2 = 0.92$ and RMSE = 1.08), while Xing et al. (2019) reported the values of R^2 (0.90) and RMSE (3.55).

Furthermore, the ANN model based on proximate analysis inputs (FC and VM) utilized a 2-12-1 structure, yielding outstanding results. The network demonstrated high efficiency in predicting *HHV*, evident from the R^2 value of 0.96 for both training and testing, and low errors of 0.19 and 0.23, respectively. Veza et al. (2022) conducted studies to accurately estimate HHV biomass using input data from proximate analysis. After statistical analysis, authors determined a high coefficient of determination ($R^2 = 0.94$) and a low error level (RMSE = 0.99), while Güleç et al. (2022) gives $R^2 = 0.82–0.87$.

Lastly, using inputs from structural analysis (Cel, Lig, and Hem), an ANN structure of 3-9-1 was developed. This model displayed strong predictive ability, with coefficients of determination of 0.92 for training and 0.86 for testing. The model exhibited minimal errors for both training (0.12) and testing (0.16), underscoring its accuracy in predicting the initial *HHV* value. Maksimuk et al. (2021) evaluated 30 equations for predicting HHV, presented two with less than 3% error and suggested using ultimate analysis for better accuracy, as predicting HHV from structure analysis dataset alone is more prone to bigger errors. The ANN (artificial neural network) model consistently demonstrated the highest accuracy across all three types of

analyses – ultimate, proximal, and structural – with an R^2 value of ≥ 0.9 for each analysis. The SVM (support vector machine) and SOP models also exhibited commendable accuracy, boasting R^2 values ranging from 0.74 to 0.95. In contrast, the RFR (random forest regression) model generally displayed lower accuracy compared to the other models. Notably, all models performed most accurately when estimating *HHV* using the proximate analysis dataset.

Analyzing the ultimate analysis inputs, the model showed significant impacts on *HHV* output based on the following variables: C (35.51%), H (−18.71%), N (31.88%), S (10.09%), and O (−3.81%). The sensitivity analysis depicted in Figure 2(a) revealed that an increase in C, N, and S, coupled with reduced H and O variables, primarily influenced the rise in *HHV* output.

Furthermore, the most effective proximate analysis model was constructed using two input variables: FC (−44.88%) and VM (−55.12%), yielding high accuracy in predicting *HHV*. In the ANN model based on structural analysis inputs, the greatest impact on the highest *HHV* output was observed with an increase in Cel (33.65%) and Lig (32.01%), along with a decrease in Hem (−34.34%). These findings underline the superior performance of the ANN model across diverse analyses, as evidenced by its highest R^2 values and lowest RMSE and MBE values.

Conclusions

Machine learning models, including non-linear ones such as SOP, SVM, RFR, and ANN, have proven effective in modeling the higher heating value (*HHV*) of biomass using input parameters derived from ultimate, proximate, and structural analyses. In this study, a total of 180 models were developed: 124 based on ultimate analysis, 28 on proximate analysis, and 28 on structural analysis. The models were generated using a combination method without repeating the elements (variables) from characteristic biomass analyses. The best-performing SOP, SVM, RFR, and ANN models for each dataset are presented. Among all the models developed, ANNs emerged as the most effective in modeling *HHV* for agricultural and forest biomass datasets. For instance, the coefficient of determination for ANN model 5-11-1 (based on inputs C, H, N, S, and O) was $R^2 = 0.90$, while ANN model 2-12-1 (inputs FC and VM) achieved $R^2 = 0.96$. Additionally, the model based on inputs Cel, Lig, and Hem attained an R^2 value of 0.91. Furthermore, models incorporating second-order polynomials ($R^2 = 0.95$), SVM ($R^2 = 0.95$), and RFR ($R^2 = 0.90$) developed based on proximate analysis inputs demonstrated the lowest errors in *HHV* modeling. This study extends existing knowledge by providing a detailed approach to predicting the HHV of biomass using different inputs in ML models, improving accuracy and model optimization. While this study demonstrates the effectiveness of ML models in predicting the HHV of biomass, future research must address both the challenges of scalability across different biomass types and practical applications, as well as the difficulties of data collection and model generalization, to develop robust, universally applicable algorithms for comprehensive biomass calorific value estimation.

Acknowledgements

The publication was supported by the Croatian Science Foundation, under project No. IP-2018-01-7472 "Sludge management via energy crops production" and within the project "Young Researchers' Career Development Project —Training of Doctoral Students", co-financed by the European Union, under the OP "Efficient Human Resources 2014–2020" from the ESF funds.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Ivan Brandić – Research Assistant (Ph.D. Candidate) at the Department of Sustainable Technologies and Renewable Energy, Faculty of Agriculture, University of Zagreb. Research interests encompass mathematical modeling in renewable energy sources and biomass

Neven Voća – Full Professor at the Department of Sustainable Technologies and Renewable Energy, Faculty of Agriculture, University of Zagreb. Research areas include biomass, renewable energy sources, and waste management.

Jerko Gunjača - Full Professor at the Department of Plant Breeding, Genetics, and Biometrics, Faculty of Agriculture, University of Zagreb. Specializes in Quantitative Genetics, Genotype by Environment Interaction, Molecular Data Analysis, Genetic Similarity and Diversity, and Association Mapping.

Biljana Lončar – Senior Research Associate at the Faculty of Technology, University of Novi Sad, Novi Sad. Focuses on Food Engineering and Chemical Engineering.

Nikola Bilandžija - Associate Professor in the Department of Mechanization and Autonomous Systems in Agriculture, Faculty of Agriculture, University of Zagreb. Research interests include Agricultural Biomass and Energy Crops, Energy Consumption and Potential in Agriculture, Energy Crop Cultivation Engineering, and Horticultural Production Engineering.

Anamarija Peter - Assistant at the Department of Sustainable Technologies and Renewable Energy, Faculty of Agriculture, University of Zagreb. Research focuses on Renewable Energy Sources, Biomass and Biofuels, Waste Management, Biological Diversity, Invasive Plant Species, and Wild Plant Species.

Jona Šurić - Research Assistant (Ph.D. Candidate) at the Department of Sustainable Technologies and Renewable Energy, Faculty of Agriculture, University of Zagreb. Research interests are in Renewable Energy Sources, Biomass and Biofuels, and Waste Management.

Lato Pezo – Researcher at the University of Belgrade Institute of General and Physical Chemistry. Research areas include Industrial Design, Biochemistry, and Chemical Kinetics.

ORCID

Ivan Brandić http://orcid.org/0000-0003-4135-8757
Neven Voća http://orcid.org/0000-0003-1016-3260
Jerko Gunjača http://orcid.org/0000-0003-3999-4519
Biljana Lončar http://orcid.org/0000-0003-2994-6871
Anamarija Peter http://orcid.org/0000-0002-2817-1130
Jona Šurić http://orcid.org/0000-0003-2003-1894
Lato Pezo http://orcid.org/0000-0002-0704-3084

Credit author statement

Ivan Brandić: Conceptualization, Methodology, Investigation, Data curation, Visualization, Writing- Original draft preparation, Writing- Reviewing and Editing; **Neven Voća**: Conceptualization, Supervision, Validation; **Jerko Gunjača**: Conceptualization, Methodology, Investigation, Writing- Original draft preparation; **Biljana Lončar**: Conceptualization, Methodology, Investigation, Writing- Original draft preparation; **Nikola Bilandžija**: Conceptualization, Methodology, Investigation, Writing- Original draft preparation; **Anamarija Peter**: Writing- Reviewing and Editing; **Jona Šurić**: Conceptualization, Methodology, Investigation, Writing- Original draft preparation; **Lato Pezo**: Conceptualization, Writing- Original draft preparation, Writing- Reviewing and Editing;

References

Ang, K. L. M., and J. K. P. Seng. 2021. Big data and machine learning with hyperspectral information in agriculture. *Institute of Electrical and Electronics Engineers Access* 9:36699–718. doi:10.1109/ACCESS.2021.3051196.

Arsenović, M., L. Pezo, S. Stanković, and Z. Radojević. 2015. Factor space differentiation of brick clays according to mineral content: Prediction of final brick product quality. *Applied Clay Science* 115:108–14. doi:10.1016/j.clay.2015.07.030.

Bakshaev, A., and R. Rudzkis. 2017. Goodness-of-fit tests based on the empirical characteristic function. *Lithuanian Mathematical Journal* 57 (2):155–70. doi:10.1007/s10986-017-9350-7.

Benos, L., A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis. 2021. Machine learning in agriculture: A comprehensive updated review. *Sensors* 21 (11):1–55. doi:10.3390/s21113758.

Bilandzija, N., N. Voca, B. Jelcic, V. Jurisic, A. Matin, M. Grubor, and T. Kricka. 2018. Evaluation of Croatian agricultural solid biomass energy potential. *Sustainable Energy Reviews* 93:225–30. doi:10.1016/j.rser.2018.05.040.

Brandić, I., L. Pezo, N. Bilandžija, A. Peter, J. Šurić, and N. Voća. 2022. Artificial neural network as a tool for estimation of the higher heating value of miscanthus based on ultimate analysis. *Mathematics* 10 (20):1–12. doi:10.3390/math10203732.

Çakman, G., S. Gheni, and S. Ceylan. 2021. Prediction of higher heating value of biochars using proximate analysis by artificial neural network. *Biomass Conversion and Biorefinery* 1–9. doi:10.1007/s13399-021-01358-4.

Callejón-Ferre, A. J., J. Carreño-Sánchez, F. J. Suárez-Medina, J. Pérez-Alonso, and B. Velázquez-Martí. 2014. Prediction models for higher heating value based on the structural analysis of the biomass of plant remains from the greenhouses of Almería (Spain). *Fuel* 116:377–87. doi:10.1016/j.fuel.2013.08.023.

Cholewa, T., C. A. Balaras, J. Kurnitski, A. Siuta-Olcha, E. Dascalaki, R. Kosonen, C. Lungu, M. Todorovic, I. Nastase, C. Jolas, M. Cakan 2022. Energy efficient renovation of existing buildings for HVAC professionals, (Rehva guidebook; No. 32). Federation of European Heating, Ventilation and Air Conditioning Associations (REHVA).

Collard, F. X., and J. Blin. 2014. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renewable and Sustainable Energy Reviews* 38:594–608. doi:10.1016/j.rser.2014.06.013.

Cvetkov-Iliev, A., A. Allauzen, and G. Varoquaux. 2022. Analytics on non-normalized data sources: More learning, rather than more cleaning. *Institute of Electrical and Electronics Engineers Access* 10:42420–31. doi:10.1109/ACCESS.2022.3168013.

Dai, Z., Z. Chen, A. Selmi, K. Jermsittiparsert, N. M. Denić, and Z. Nešić. 2021. Machine learning prediction of higher heating value of biomass. *Biomass Conversion and Biorefinery* 13 (5):1–9. doi:10.1007/s13399-021-01273-8.

Darvishan, A., H. Bakhshi, M. Madadkhani, M. Mir, and A. Bemani. 2018. 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: Recovery, Uti-Lization and Environmental Effects* 40 (24):2960–66. doi:10.1080/15567036.2018.1514437.

Das, N., P. K. Jena, D. Padhi, M. Kumar Mohanty, and G. Sahoo. 2021. A comprehensive review of characterization, pretreatment, and its applications on different lignocellulosic biomass for bioethanol production. *Biomass Conversion and Biorefinery* 13 (2):1–25. doi:10.1007/s13399-021-01294-3.

Fernández-Navarro, F., M. Carbonero-Ruz, D. Becerra Alonso, M. Torres-Jiménez 2016. IEEE Transactions on neural networks and learning systems 1 global sensitivity estimates for neural network classifiers, USA, 1–13.

García Nieto, P. J., E. García-Gonzalo, F. Sánchez Lasheras, J. P. Paredes-Sánchez, and P. Riesgo Fernández. 2019. Forecast of the higher heating value in biomass torrefaction by means of machine learning techniques. *Journal of Computational and Applied Mathematics* 357:284–301. doi:10.1016/j.cam.2019.03.009.

Ge, Z., Z. Song, S. X. Ding, and B. Huang. 2017. Data mining and analytics in the process industry: The role of machine learning. *Institute of Electrical and Electronics Engineers Access* 5:20590–616. doi:10.1109/ACCESS.2017.2756872.

Gironès, V. C., S. Moret, E. Peduzzi, M. Nasato, and F. Maréchal. 2017. Optimal use of biomass in large scale energy systems. *Insights for Energy Policy Energy* 137:789–97. doi:10.1016/j.energy.2017.05.027.

Güleç, F., D. Pekaslan, O. Williams, and E. Lester. 2022. Predictability of higher heating value of biomass feedstocks via proximate and ultimate analyses – A comprehensive study of artificial neural network applications. *Fuel* 320 (January):123944. doi:10.1016/j.fuel.2022.123944.

Ighalo, J. O., A. G. Adeniyi, and G. Marques. 2020. Application 349 of artificial neural networks in predicting biomass higher heating value: An early appraisal. *Energy Sources Part A: Recovery, Utilization, and Environmental Effects* 1–8. https://www.tandfonline.com/loi/ueso20.

Ismaila, A., I. Abdullahi, and N. N. Garba. 2013. Investigation on biomass briquettes as energy source in relation to their calorific values and measurement of their total carbon and elemental contents for efficient biofuel utilization. *Pelagia Research Library Advances in Applied Science Research* 4 (2016):303–09.

Khayyaty, Z., I. F. Ilyasz, A. Jindal, S. Madden, M. Ouzzani, P. Papotti, J. A. Quiané-Ruiz, N. Tang, S. Yin 2015. BigDansing: A system for big data cleansing. Proceedings of the ACM SIGMOD International Conference on Management of Data, 1215–30. 10.1145/2723372.2747646

Kollo, T., and D. Von Rosen. 2005. Advanced multivariate statistics with matrices. *Advanced Multivariate Statistics with Matrices*. doi:10.1007/1-4020-3419-9.

Kołodziej, J., K. Pudełko, and J. Mańkowski. 2023. Energy and biomass yield of industrial hemp (Cannabis sativa L.) as influenced by seeding rate and harvest time in Polish agro-climatic conditions. *Journal of Natural Fibers* 20 (1). doi:10.1080/15440478.2022.2159609.

Li, T. Y., H. Xiang, Y. Yang, J. Wang, and G. Yildiz. 2021. Prediction of char production from slow pyrolysis of lignocellulosic biomass using multiple nonlinear regression and artificial neural network. *Journal of Analytical and Applied Pyrolysis* 159:105286. doi:10.1016/j.jaap.2021.105286.

Li, Y., C. Zou, M. Berecibar, E. Nanini-Maury, J. C. W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar. 2018. Random forest regression for online capacity estimation of lithium-ion batteries. *Applied Energy* 232:197–210. doi:10.1016/j.apenergy.2018.09.182.

Maksimuk, Y., Z. Antonava, V. Krouk, A. Korsakova, and V. Kursevich. 2021. Prediction of higher heating value (HHV) based on the structural composition for biomass. *Fuel* 299 (March):120860. doi:10.1016/j.fuel.2021.120860.

Mansor, A. M., J. S. Lim, F. N. Ani, H. Hashim, and W. S. Ho. 2019. Characteristics of cellulose, hemicellulose and lignin of MD2 pineapple biomass. *Chemical Engineering Transaction* 72:79–84. doi:10.3303/CET1972014.

Martino, L., and J. Read. 2021. A joint introduction to gaussian processes and relevance vector machines with connections to Kalman filtering and other kernel smoothers. *Information Fusion* 74:17–38. doi:10.1016/j.inffus.2021.03.002.

Matin, S. S., and S. C. Chelgani. 2016. Estimation of coal gross calorific value based on various analyses by random forest method. *Fuel* 177:274–78. doi:10.1016/j.fuel.2016.03.031.

Maydeu-Olivares, A. 2017. Maximum likelihood estimation of structural equation models for continuous data: standard errors and goodness of fit. *Structural Equation Modeling* 24 (3):383–94. doi:10.1080/10705511.2016.1269606.

Noushabadi, A. S., A. Dashti, F. Ahmadijokani, J. Hu, and A. H. Mohammadi. 2021. Estimation of higher heating values (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation. *Renewable Energy* 179:550–62. doi:10.1016/j.renene.2021.07.003.

Ostertagová, E. 2012. Modelling using polynomial regression. *Procedia Engineering* 48:500–06. doi:10.1016/j.proeng.2012.09.545.

Pezo, L. L., B. L. Ćurčić, V. S. Filipović, M. R. Nićetin, G. B. Koprivica, N. M. Mišljenović, and L. B. Lević. 2013. Artificial neural network model of pork meat cubes osmotic dehydration. *Chemical Industry* 67 (3):465–75. doi:10.2298/HEMIND120529082P.

Qiu, J., Q. Wu, G. Ding, Y. Xu, and S. Feng. 2016. A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*. 2016(1). doi:10.1186/s13634-1.

Ranganayaki, V., and S. N. Deepa. 2019. Linear and non-linear proximal support vector machine classifiers for wind speed prediction. *Cluster Computing* 22 (S1):379–90. doi:10.1007/s10586-018-2005-6.

Schonlau, M., and R. Y. Zou. 2020. The random forest algorithm for statistical learning. *Stata Journal* 20 (1):3–29. doi:10.1177/1536867X20909688.

Scornet, E., G. Biau, and J. P. Vert. 2015. Consistency of random forests. *The Annals of Statistics* 43 (4):1716–41. doi:10.1214/15-AOS1321.

Soponpongpipat, N., D. Sittikul, and U. Sae-Ueng. 2015. Higher heating value prediction of torrefaction char produced from non-woody biomass. *Front Energy* 9 (4):461–71. doi:10.1007/s11708-015-0377-3.

Taki, M., and A. Rohani. 2022. Machine learning models for prediction the Higher Heating Value (HHV) of Municipal Solid Waste (MSW) for waste-to-energy evaluation. *Case Studies in Thermal Engineering* 31:101823. doi:10.1016/j.csite.2022.101823.

Tipping, M. E. 2001. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 1 (3):211–44.

Tursi, A. 2019. A review on biomass: Importance, chemistry, classification, and conversion. *Biofuel Research Journal* 6 (2):962–79. doi:https://doi.org/10.18331/BRJ2019.6.2.3.

Veza, I., H. Irianto Panchal, P. A. Paristiawan, M. Idris, I. M. R. Fattah, N. R. Putra, R. Silambarasan, and R. Silambarasan. 2022. Improved prediction accuracy of biomass heating value using proximate analysis with various ANN training algorithms. *Results in Engineering* 16 (September):100688. doi:10.1016/j.rineng.2022.100688.

Voća, N., J. Leto, T. Karažija, N. Bilandžija, A. Peter, H. Kutnjak, J. Šurić, and M. Poljak. 2021. Energy properties and biomass yield of miscanthus x giganteus fertilized by municipal sewage sludge. *Molecules* 26 (14):4371. doi:10.3390/molecules26144371.

Xiaorui, L., Y. Jiamin, and Y. Longji. 2023. Predicting the high heating value and nitrogen content of torrefied biomass using a support vector machine optimized by a sparrow search algorithm. *RSC Advances* 13 (2):802–07. doi:10.1039/d2ra06869a.

Xing, J., K. Luo, H. Wang, Z. Gao, and J. Fan. 2019. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* 188:116077. doi:10.1016/j.energy.2019.116077.

Yoon, Y., G. Swales, and T. M. Margavio. 2017. A comparison of discriminant analysis versus artificial neural networks. *The Journal of the Operational Research Society* 44 (1):51–60. doi:10.1057/jors.1993.6.