

Development of new nonlinear mathematical models for modelling the higher heating value of biomass

DANI
DOKTORATA
BIOTEHNIČKOG
PODRUČJA

12. i 13. rujna 2024.



Ivan Brandić

University of Zagreb, Faculty of Agriculture

Biomass

- Renewable energy source
- Types of biomass
- Multiple application
- Energy potential
- Research and optimization





Biomass characteristics

- Ultimate analysis
- Proximate analysis
- Structural analysis
- Calorimetric analysis

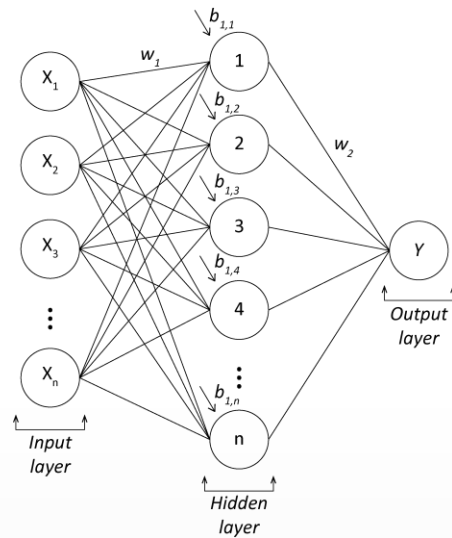
Features of machine learning regression models

- Continuous Modeling
- Pattern Recognition
- Multidimensional Inputs
- Generalization Ability
- Model Adjustment
- Regularization

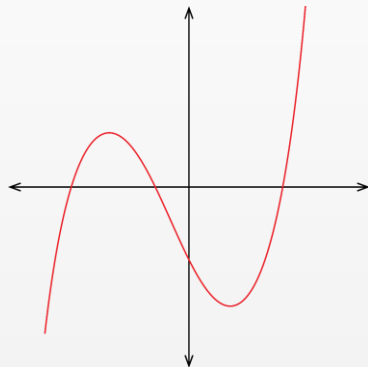


Selection of nonlinear and machine learning models

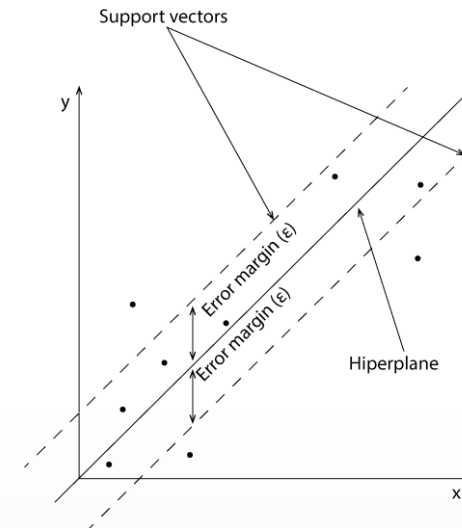
- Artificial neural networks (ANN)



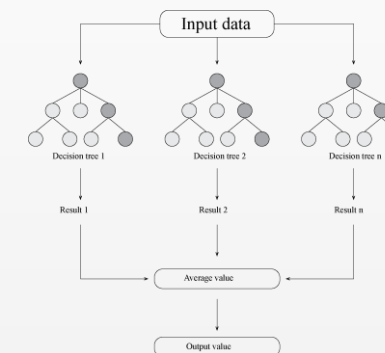
- High order polynomial (HOP)



- Support vector machine (SVM)



- Random forest regression (RFR)



Hypotheses and research goals

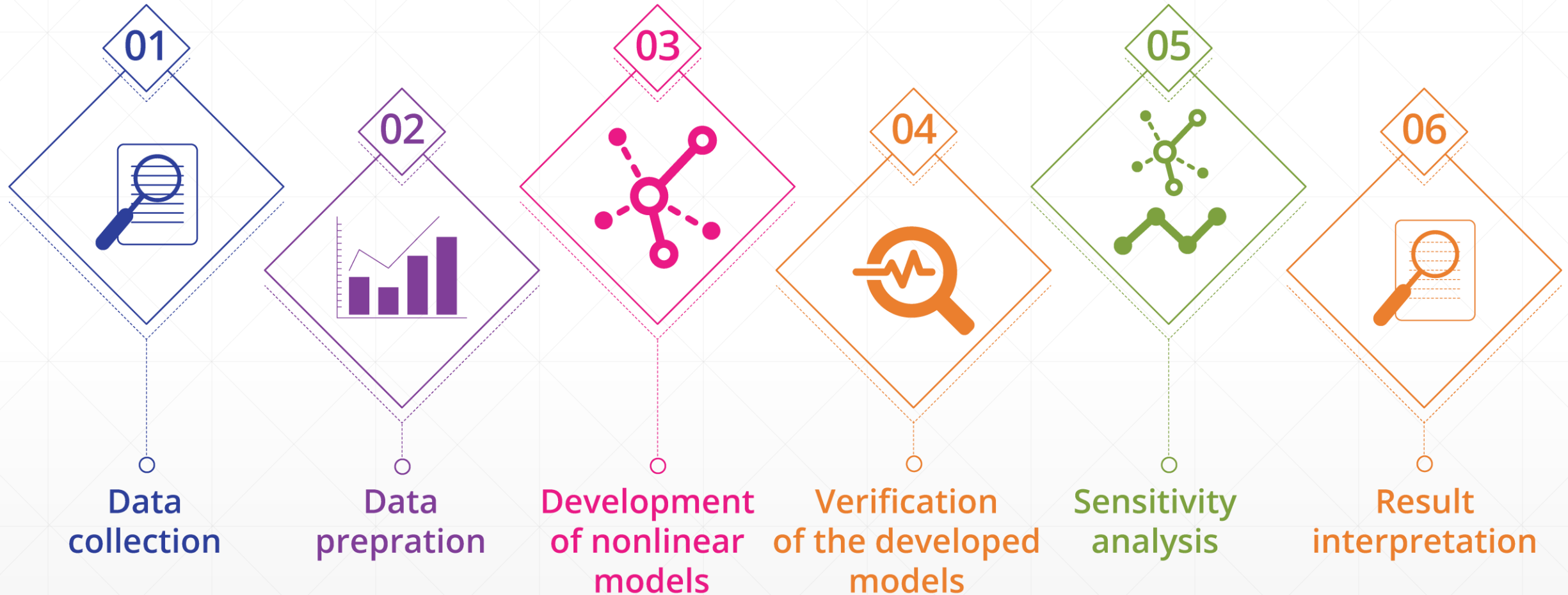
Hypothesis:

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

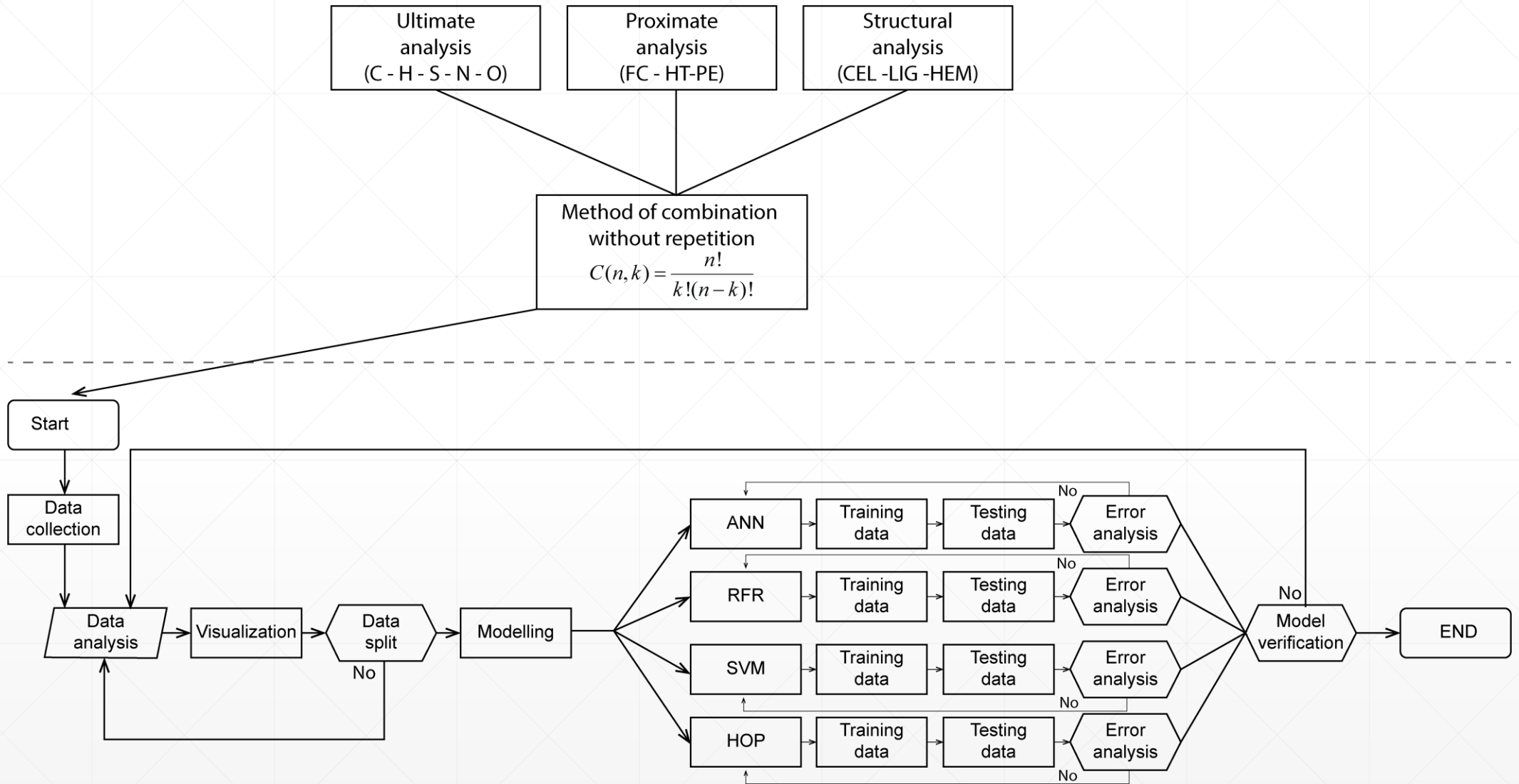
Objectives:

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

Materials and methods



Flowchart of modeling



Model verification

Chi-squared test (**X²**):

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

Sum of squared error (**SSE**):

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

Where:

$x_{pre,i}$ Output value

$x_{exp,i}$ Input value

Root mean squared error (**RMSE**):

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

Average absolute relative deviation (**AARD**):

$$AARD = \frac{100}{n} \sum_{i=1}^n \frac{|x_i^{predicted} - x_i^{experimental}|}{x_i^{exp}}$$

Mean percentage error (**MPE**):

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

Coefficient of determination (**R²**):

$$R^2 = 1 - \frac{\sum_{i=1}^n [x_i^{predicted} - x_i^{experimental}]^2}{\sum_{i=1}^n [x_i^{predicted} - x^m]^2}, x_m = \frac{\sum_{i=1}^n x_i^{experimental}}{n}$$

In addition. **residual analysis** is also calculated:

- Skewness (**Skew**)
- Kurtosis (**Kurt**)
- Standard deviation (**SD**)
- Variance (**Var**)

Sensitivity analysis based on ANN model (Yoon's method)

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

Where:

W is the weight coefficient of the ANN model.

i is the input variable.

j is the output variable.

k is the artificial neuron of the hidden layer.

n is the number of artificial neurons in the hidden layer.

m is the number of input variables.

Relative importance ; **input**  **output**

- Contribution of each input variable to the overall variability of the output variable
- Identifies key variables that have the greatest impact on the model
- Assesses how sensitive the model is to changes in input variables

Results – Biomass characteristics

Dataset	Biomass type	Agricultural biomass	Wood biomass	Statistical significance
Ultimate analysis	C (%)	47.61 ± 7.76	48.13 ± 9.35	n.s.
	H (%)	5.52± 1.47	5.37 ± 1.50	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.
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.
Structural analysis	Cellulose (%)	44.70 ± 13.40	44.04 ± 7.63	n.s.
	Lignin (%)	15.08 ± 7.87	27.64 ± 7.04	*
	Hemicellulose (%)	23.21 ± 5.81	27.25 ± 4.41	*
	HHV (MJ kg ⁻¹)	17.67 ± 1.66	19.68 ± 0.62	*

C – carbon; H – Hydrogen; N – nitrogen; S – sulfur; O – Oxygen; HHV – higher heating value ;FC – Fixed carbon; VM – Volatile matter;
 Statistical significance: * p<0.01; ** p<0.05; ns – not significant

Results – Model error






Dataset	Model	Inputs	χ^2	RMSE	MBE	MPE	SSE	AARD	R ²	Skew	Kurt	SD	Var
Ultimate analysis	ANN	CHNSO	1.02	1.01	0.06	4.21	251.19	196.44	0.9	-0.46	1.23	1.01	1.01
	HOP	CHNO	1.77	1.33	0.1	5.08	437.11	237.51	0.82	0.94	8.1	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 analysis	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
	HOP	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	8.8	1907.5	505.41	0.9	0.85	9.09	2.56	6.55
Structural analysis	ANN	Cel Lig Hem	0.26	0.51	0	2.3	74.81	193.72	0.91	-1.09	3.34	0.51	0.26
	HOP	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	-0.8	2.16	1.15	1.32
	RFR	Lig Hem	0.54	0.73	0	3.19	153.37	275.97	0.82	-0.46	2.91	0.73	0.54

C – carbon; H – Hydrogen; N – nitrogen; S – sulfur; O – Oxygen; HHV – higher heating value ;FC – Fixed carbon; VM – Volatile matter; ANN – Artificial neural networks; HOP – High order polynomials; SVM – Support vector machine; RFR – Random forest regression.







Summary of the performance of the most significant models analyzed in 5 published scientific papers

Input datasets↓	Model ↓	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
Ultimate analysis	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	ANN	-	-	-	-	R ² =0.96
	SVM	-	-	-	-	R ² =0.95
	RFR	-	-	-	-	R ² =0.90
	HOP	-	-	-	-	R ² =0.95
Structural analysis	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






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

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


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

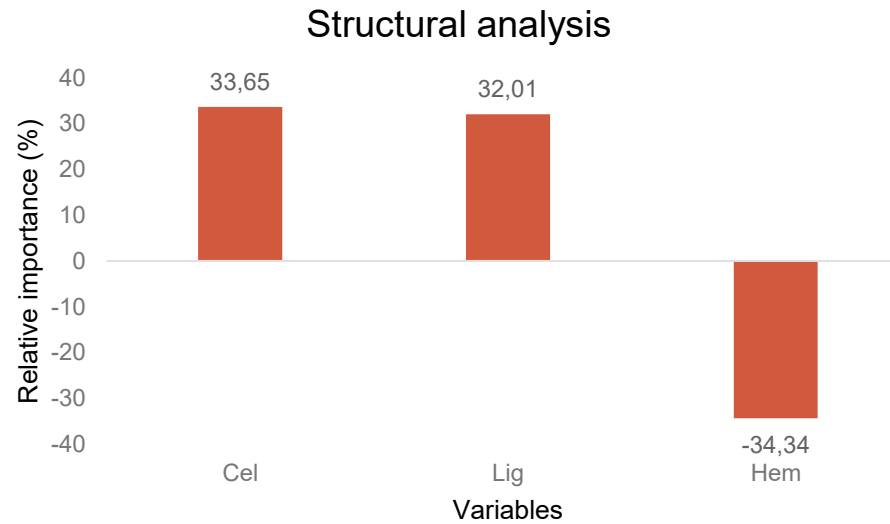
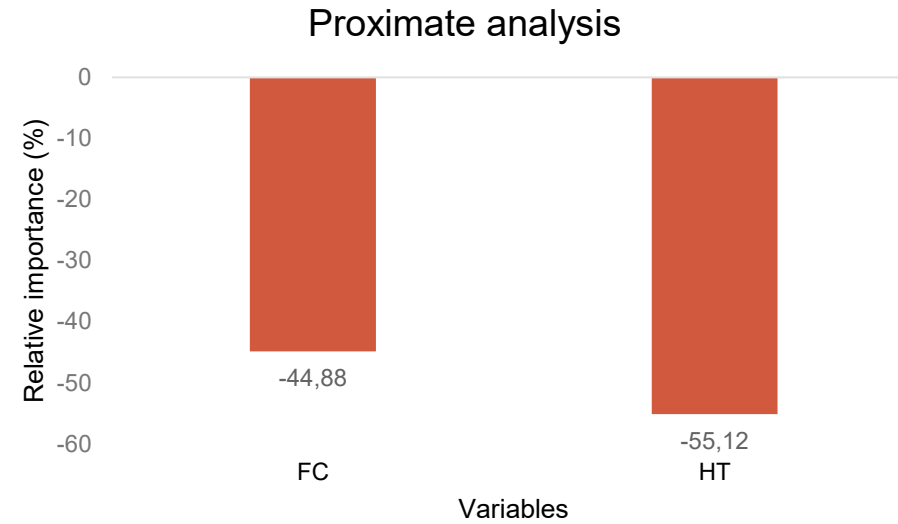
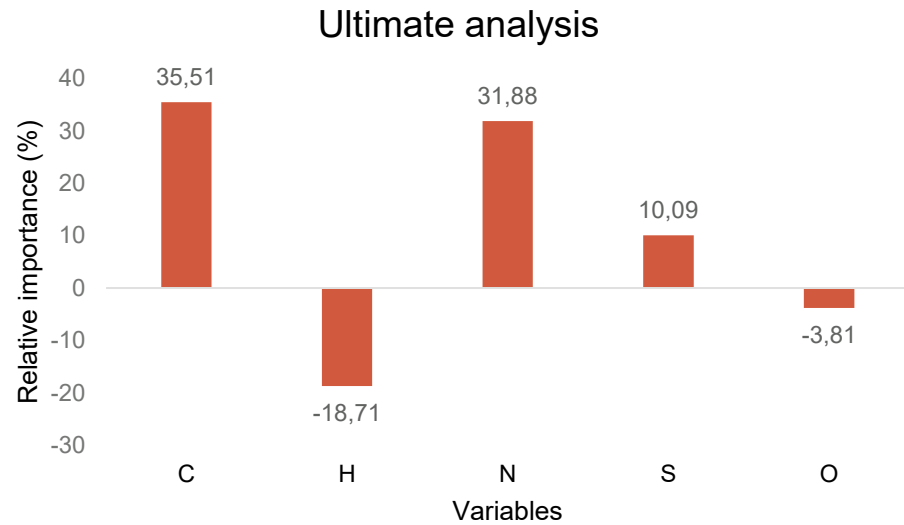
ASSESSING THE PROPERTIES OF *MISCANTHUS X GIGANTEUS* UNDER VARYING LEVELS OF ASH FERTILIZATION TREATMENT AND REGRESSION NEURAL NETWORK INSIGHT INTO CALORIFIC VALUE

Ivan BRANDIĆ¹, Lato PEZO^{2,}, Neven VOĆA¹, Josip LETO¹, Jona ŠURIC¹, Anamarija PETER¹, Nikola BILANDŽIJA¹*

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

Ivan Brandić , Neven Voća , Jerko Gunjača , Biljana Lončar , Nikola Bilandžija ^d, Anamarija Peter , Jona Šurić , and Lato Pezo 

Sensitivity analysis



Conclusion(s)

- Models based on data from proximate analysis achieved the lowest error in modeling HHV of biomass ($R^2=0.96$) compared to models based on datasets from ultimate and structural analysis.
 - ANN models have the lowest error in modeling HHV compared to SVM, RFR, and HOP regardless of the applied dataset.
-

Conclusion(s)

- Using the **method of combination without repetition**, the most suitable models were determined for each dataset and each developed type of model:
 - **Ultimate analysis:** The most suitable model was developed **with all input variables** (C, H, N, S, and O).
 - **Proximate analysis:** The most suitable model was developed with 2 input variables (**FC and HT**).
 - **Structural analysis:** The most suitable model was developed **with all input variables** (cel, hem, and lig).
-

Thank you for your attention!



ibrandic@agr.hr

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.

