

Article

Use of Machine Learning Methods for Predicting Amount of Bioethanol Obtained from Lignocellulosic Biomass with the Use of Ionic Liquids for Pretreatment

Małgorzata Smuga-Kogut ¹, Tomasz Kogut ^{2,*}, Roksana Markiewicz ³ and Adam Słowik ⁴

¹ Department of Agrobiotechnology, Faculty of Mechanical Engineering, Koszalin University of Technology, Raclawicka 15-17, 75-620 Koszalin, Poland; malgorzata.smuga-kogut@tu.koszalin.pl

² Department of Geoinformatic, Koszalin University of Technology, Sniadeckich 2, 75-453 Koszalin, Poland

³ NanoBioMedical Centre, Adam Mickiewicz University in Poznań, Wszechnicy Piastowskiej 3, 61-614 Poznań, Poland; roksana.markiewicz@amu.edu.pl

⁴ Department of Computer Engineering, Koszalin University of Technology, Sniadeckich 2, 75-453 Koszalin, Poland; adam.slowik@tu.koszalin.pl

* Correspondence: tomasz.kogut@tu.koszalin.pl; Tel.: +48-943-486-720

Abstract: The study objective was to model and predict the bioethanol production process from lignocellulosic biomass based on an example of empirical study results. Two types of algorithms were used in machine learning: artificial neural network (ANN) and random forest algorithm (RF). Data for the model included results of studying bioethanol production with the use of ionic liquids (ILs) and different enzymatic preparations from the following biomass types: buckwheat straw and biomass from four wastelands, including a mixture of various plants: stems of giant miscanthus, common nettle, goldenrod, common broom, fireweed, and hay (a mix of grasses). The input variables consisted of different ionic liquids (imidazolium and ammonium), enzymatic preparations, enzyme doses, time and temperature of pretreatment, and type of yeast for alcoholic fermentation. The output value was the bioethanol concentration. The multilayer perceptron (MLP) was used in the artificial neural networks. Two model types were created; the training dataset comprised 120 vectors (14 elements for Model 1 and 11 elements for Model 2). Assessment of the optimum random forest was carried out using the same division of experimental points (two random datasets, containing 2/3 for training and 1/3 for testing) and the same criteria used for the artificial neural network models. Data for mugwort and hemp were used for validation. In both models, the coefficient of determination for neural networks was <0.9 , while for RF it oscillated around 0.95. Considering the fairly large spread of the determination coefficient, two hybrid models were generated. The use of the hybrid approach in creating models describing the present bioethanol production process resulted in an increase in the fit of the model to $R^2 = 0.961$. The hybrid model can be used for the initial classification of plants without the necessity to perform lengthy and expensive research related to IL-based pretreatment and further hydrolysis; only their lignocellulosic composition results are needed.

Keywords: hemp; mugwort; bioethanol; machine learning; enzymatic hydrolysis



Citation: Smuga-Kogut, M.; Kogut, T.; Markiewicz, R.; Słowik, A. Use of Machine Learning Methods for Predicting Amount of Bioethanol Obtained from Lignocellulosic Biomass with the Use of Ionic Liquids for Pretreatment. *Energies* **2021**, *14*, 243. <https://doi.org/10.3390/en14010243>

Received: 16 November 2020

Accepted: 4 January 2021

Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The production of bioethanol is a current topic raised by scientists, technologists, and representatives of fuel companies in the European Union who are working on satisfying the percentage share of this biocomponent in conventional fuels. The policy of the European Union countries is aimed at the search for low-emission technology, which is sustainable and aimed at reducing CO₂ production at relatively low cost of production lines. Taking this into account, research targeting the development of bioethanol preparation from the available lignocellulosic biomass, using the existing production lines or innovative pretreatment methods, is an important topic. The preparation of bioethanol is a complex procedure, the success of which hinges primarily on the use of raw material

that is inexpensive, available in every region of Europe, and has effective and cost-efficient pretreatment, which will result in a high yield of fermenting sugars.

Numerous examples of the production of fermenting sugars from lignocellulose can be found in the literature. A large portion of these publications contain data on the impact of pretreatment on enzymatic hydrolysis of the biomass. The most common include treatments using ammonia fiber expansion (AFEX), steam explosion, ionic liquids (ILs) and other methods [1–5]. There are numerous examples of conversion of rye, corn, rice, wheat straw, poplar, willow, and coniferous tree chips into bioethanol based on the method using pretreatment with ionic liquids and enzymatic hydrolysis [6–10]. However, comparison of the majority of the data has proven to be difficult due to various amounts of substrates used for the process or other pretreatment parameters. Therefore, the use of literature data to design the bioethanol production process at a greater scale is very complicated and problematic. In this study, an attempt was made to create a model that would allow for the estimation or prediction of bioethanol concentration from various lignocellulosic raw materials. Such a model would allow the classification of biomass for bioethanol production on the basis of its chemical composition and facilitate choice of a suitable ionic liquid and enzyme preparation used for pretreatment. Such an approach would save time and costs in laboratory research, which is now a way to search for ideal production methods and an efficient source of lignocellulosic biomass.

Mathematical and statistical models provide essential information for understanding, analyzing, and predicting biological processes, and are necessary to optimize key parameters to improve process performance. Modeling and optimization of biofuel manufacturing processes will contribute to a better understanding of the process expenditures in order to obtain the optimum efficiency. The main purpose of modeling is to optimize the operations involved in their production in order to achieve efficiency improvement [11–13].

Artificial intelligence tools have appeared as a promising method for modeling and optimizing bioprocesses. In the last decade, artificial neural networks (ANNs) were applied in multidimensional, nonlinear research and development of bioprocesses. They have been found effective in developing bioprocess models devoid of prior information on the kinetics and metabolic flows occurring in cells and cell surroundings [14]. Furthermore, ANN are completely based on data, without prior knowledge on the events regulating the process [15]. The appeal of ANN as a modeling tool derives from their unique functions of processing information that is assigned primarily—linearity, high parallelism, and error and noise acceptance—as well as their capability to learn and generalize. ANNs have gained much attention from significant soft computing tools that are not only limited to data processing and analysis, but can also be used to solve problems in multifaceted and nonlinear processes [16].

For these reasons, the aim of this study was to collect the current data on the production of bioethanol from lignocellulosic biomass using ionic liquids and to create a model that can be used to predict ethanol concentration from lignocellulosic raw materials. For this purpose, machine learning (ML) methods were used. Increased interest in the use of ML procedures has been observed, e.g., artificial neural networks (ANNs) and random forests (RFs) in the context of bioethanol production from biomass [17–20]. With the use of these algorithms, it was assumed that on the basis of the results of laboratory tests obtained, it will be possible to predict the concentration of bioethanol in various biomass species based on their lignocellulosic composition. This would aid the classification of biomass type to the suitable ionic liquid or type of cellulolytic enzyme. This preliminary estimation in the case of laboratory studies is important because in this way the costs associated with the synthesis or purchase of ionic liquid and enzymes can be kept to the minimum. It would be possible to perform more rapid verification and grouping of plant types that are found in the given region. Such model could also inspire classification of other pretreatments, and ML data could be applied to a wider extent in scaling-up processes. In order to reduce the costs of the process, plants can be grouped in terms of similarities in their cellulose

content, affinity for ionic liquids, and their influence on the exposure of cellulose fibers or the selection of cellulolytic enzymes for the appropriate type of biomass.

In the present study, the prognostic model was verified by empirical studies of bioethanol production from hemp (*Cannabis sativa* L.) and mugwort (*Artemisia vulgaris* L.). The stems of hemp contain high amounts of cellulose (up to 80%), with lignin content of about 15–20%. In addition, it is a plant that is increasingly commonly grown for the production of oil, fiber, essential oils, etc. The use of both plants properly fit in the criteria of sustainable development, especially considering their biological and agrotechnical properties, which makes them economically and environmentally favorable. The properties exhibited by this biomass source make them admirable for the development of multiprocessor systems by gradually separating the biomass into several useful components. This trait provides hemp with an advantage over other industrial crops, as they are usually used for extraction of one component [21,22]. As far as energy consumption is concerned, it is necessary to emphasize that the mean yield value of green hemp plants was 14.5 t/ha [23] (by means of dry weight). It is possible to obtain about 10.5 t/ha of raw material, which can be potentially used for energy purposes [24]. Parts of the stems that were considered waste or could be used for the production of solid fuel (pellets) were used to produce bioethanol. For contrast, mugwort was used, which is a plant commonly viewed as a weed, growing in arable land as well as boundary strips and agricultural wastelands. This plant possesses approximately 1–1.5 m stems with 35% cellulose content, 25% lignin, and approximately 20% hemicellulose. Mugwort is a native species to Poland. It was introduced to North America, where it has spread and is treated as an invasive species. The mugwort species found in Europe are typically weed and ruderal plants. Common mugwort is the most widespread species. It is characterized by high growth force and good regenerative abilities. New plants can emerge from even finely cut rhizome fragments [25,26].

The production of bioethanol from lignocellulose could become profitable if the costs associated with the production of biomass are very low. Each geographic region is distinguished by a great variety of flora. These can be found in many literature reports stating that the success of the bioethanol production process is influenced by the biomass composition: proportions of cellulose, hemicellulose, and lignin; and the type of fiber arrangement, and thus also the plant species and the degree of its maturity. Taking all these factors into account, it becomes very difficult to test all plant species that could potentially form a good source for the production of bioethanol. Therefore, it is easier to use energy crops for the production of bioethanol. A mathematical model based on experimental data can be a helpful tool in determining the suitability of a plant species for processing into bioethanol. The production of bioethanol is one of the biotechnological processes the complexity of which is high and difficult to present via ready-made algorithms. Therefore, in such situations it is perfect to use ML, including e.g., ANN or RF, as a prediction tool for future biomass samples. This publication attempts to create a model based on experimental data of the bioethanol production process from biomass on a laboratory scale and validate this model based on the results of the experiment—fermentation of hemp and mugwort stems.

2. Materials and Methods

2.1. Raw Materials

Two plant species were used for model validation: stems of hemp, cultivar Finola (plant collection of the Department of Agrobiotechnology, Koszalin University of Technology) and common mugwort stems collected from agricultural wastelands. Stems of mugwort were taken from two wastelands and hemp samples from two extreme cultivation sites within 30 km from Koszalin (N 54°11'26.11", E 16°10'53.77"), and then the bioethanol production process was carried out for each sample separately in triplicate. The mean values of three replicates for four biomass samples are presented in Table 1. The biomass was collected in late autumn (second half of October 2019). The stems were dried and ground, then the material was pretreated using different ionic liquids.

Table 1. Chemical composition of biomass and glucose content after enzymatic hydrolysis and ethanol content, following alcoholic fermentation of samples of material used for model validation.

Material Name	Composition			Pretreatment	Enzymatic Hydrolysis	Alcoholic Fermentation
	Cellulose [%]	Hemicellulose [%]	Lignin [%]	Ionic Liquid	Glucose [g/L]	Ethanol [g/L]
Hemp (<i>Cannabis sativa</i> L.)	55.18	20.42	15.78	[BMIM][OAc]	11.54	8.33
	55.51	17.54	17.8	[BMIM][OAc]	12.27	9.93
	58.37	15.7	18.22	[CHDMA-C4][OAc]	9.16	5.17
	58.98	15.3	11.5	[CHDMA-C4][OAc]	9.85	5.81
	61.79	18.32	10.39	[CHDMA-C6][OAc]	8.56	6.01
	60.11	17.12	12.34	[CHDMA-C6][OAc]	8.49	5.63
	30.9	9.93	13.8	[EMIM][DEP]	5.78	4.78
	32.5	10.1	12.7	[EMIM][DEP]	6.86	4.08
	46.4	15.12	16.26	[EMIM][OAc]	11.32	7.97
	48.25	16.5	15.78	[EMIM][OAc]	11.32	8.28
	62.22	17.72	19.98	untreated	3.49	3.28
	61.51	16.27	11.56	untreated	3.38	2.60
Common mugwort (<i>Artemisia vulgaris</i> L.)	46.42	21.05	18.78	[BMIM][OAc]	2.59	1.65
	41.37	21.93	20.44	[EMIM][DEP]	0.26	0.22
	42.78	19.9	16.94	[EMIM][OAc]	3.15	1.50
	46.42	21.05	18.78	[BMIM][OAc]	4.80	2.37
	42.78	19.9	16.94	[EMIM][OAc]	4.86	2.21
	45.1	13.79	20.38	untreated	1.81	0.95
	43.27	11.9	18.76	untreated	2.22	1.12
	41.9	17.72	24.3	[CHDMA-C4][OAc]	2.43	1.42
	42.19	21.03	24.42	[CHDMA-C4][OAc]	2.43	1.53
	42.4	18.14	23.99	[CHDMA-C6][OAc]	1.87	1.29
	41.37	21.93	20.44	[EMIM][DEP]	2.49	1.03
	41.37	21.93	20.44	[EMIM][DEP]	0.60	0.37
	46.42	21.05	18.78	[BMIM][OAc]	3.18	1.56
	42.78	19.9	16.94	[EMIM][OAc]	2.97	1.67

2.2. Ionic Liquids (ILs)

For the pretreatment of cellulose-rich material, five ILs from imidazolium and ammonium groups were chosen. Three were commercially available (Iolitech GmbH, Germany) imidazolium ILs: 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]), 1-butyl-3-methylimidazolium acetate ([BMIM][OAc]), and 1-ethyl-3-methylimidazolium diethyl phosphate ([EMIM][DEP]). The remaining two, belonging to the group of ammonium ILs, were synthesized for the study: butyl(cyclohexyl)dimethylammonium acetate ([CHDMA-C4][OAc]) and (cyclohexyl) hexyldimethylammonium acetate ([CHDMA-C6][OAc]).

2.3. Synthesis of [CHDMA-C4][OAc] and [CHDMA-C6][OAc] Ionic Liquids

Synthesis of [CHDMA-C4][OAc] and [CHDMA-C6][OAc], and its bromide precursors was conducted according to a previous work, where apart from the synthetic route, full characteristics and cellulose dissolution were also described [27]. In that way, appropriate bromides were prepared in a quaternization reaction between cyclohexyldimethylamine and an appropriate 1-bromoalkane with 10% extension, in acetonitrile, at room temperature for 72 h. The crude product was obtained by evaporation of the reaction background in a rotary evaporator. After the ethyl acetate addition, the prepared bromides were filtered

and dried at 40 °C under reduced pressure for 48 h (vacuum dryer). The product was then kept in a desiccator to avoid moisture uptake from the environment. The second step of the ILs preparation was the anion exchange reaction between quaternary bromide and acetic acid. An appropriate bromide (with hexyl or butyl substituent) was dissolved in methanol. To this solution, a previously prepared stoichiometric amount of KOH dissolved in methanol was added. The solutions were stirred for 3 h at ambient temperature. Partially precipitated side product (KBr) was filtrated and an appropriate (stoichiometric) amount of acetic acid was added to the reaction mixture. Solutions were further stirred at the same conditions for 1 h and the reaction background was evaporated in a rotary evaporator (40 °C). The crude product was purified with the addition of anhydrous acetone. The remaining potassium bromide was filtered, and solvent once again removed (rotary evaporator) to give the final product. The ILs obtained ([CHDMA-C4][OAc] and [CHDMA-C6][OAc] were finally dried for 48 h under reduced pressure in a vacuum dryer.

2.4. Pretreatment, Enzymatic Hydrolysis, and Alcoholic Fermentation

Imidazolium ILs, namely [EMIM][OAc], [BMIM][OAc], [EMIM][DEP], [CHDMA-C4][OAc], and [CHDMA-C6][OAc] were used for biomass purification. To achieve this, solutions of appropriate ground material (5 g) and a specific IL (50 mL) were prepared, which were further subjected to homogenization (2 min) and incubation at 120 °C for 2 h. Samples were afterwards left to cool to room temperature. In the next step, the cellulose fibers were isolated via thorough rinsing of the prepared mixture with deionized water. This was repeated at least three times, to the point of total IL removal. The solid fraction obtained was further dissolved in a 50 mM acetate buffer with a pH equal to 5.0 (100 mL). Enzymatic hydrolysis was then performed on the pretreated and nontreated lignocellulosic biomass.

For the enzymatic hydrolysis of hemp stems, Cellic CTec2 (Sigma-Merck, Darmstadt, Germany) was used at the amount of 20 FPU/g of cellulose. Samples were incubated at 50 °C for 72 h. On the other hand, the mugwort samples were hydrolyzed using the following enzymatic preparations: Cellic CTec2, cellulase from *Aspergillus* sp., cellulase from *T. reesei* (Sigma-Merck, Darmstadt, Germany). The incubation of biomass fractions mixed with *Aspergillus* sp. and *T. reesei* cellulases was performed for 72 h at 47 °C.

Before performing the alcoholic fermentation, the hydrolysate solutions were purified by means of filtration to get rid of any residual lignocellulose. The pH of the fermentation broth was kept constant at 5.0 for each sampling point. The pH control was performed by the adding a solution of H₂SO₄ (10 wt.%) or NaOH (20 wt.%). Freeze-dried distiller's yeast *Saccharomyces cerevisiae* type II (purchased from Sigma-Aldrich) (5%, w/v) were used to initiate ethanol fermentation. This was afterwards allowed to proceed in anaerobic conditions for four days. After fermentation, the samples were further analyzed to establish the ethanol concentrations.

Control samples were hemp and mugwort stalks not pretreated with ionic liquids. Samples were dissolved in an acetate buffer, according to the protocol described for IL pretreated samples. The material was characterized to establish the cellulose, hemicellulose, and lignin content. Glucose content after the process of enzymatic hydrolysis and alcohol concentration after the fermentation process were also determined.

2.5. Analytical Techniques

An Ankom A200 fiber analyzer was used to determine the amounts of lignin/cellulose/hemicellulose in all biomass samples (with the use of filter bag encapsulation). Fiber test results were determined as neutral detergent fibers (NDF) with the use of Van Soest method, and acidic detergent fiber (ADF) and acidic detergent lignin (ADL) according to the standard [28]. The difference between the ADF and ADL fractional share was the cellulose content, while the difference between NDF and ADF fractional share was the hemicellulose content. High performance liquid chromatography (HPLC) was used to determine the amounts of glucose and ethanol (Merck Hitachi, Darmstadt, Germany). For that purpose, the prepared samples were, in the first step, subjected to centrifugation (10 min, 4000 × g,

4 °C) with the use of a Multifuge 3SR (Darmstadt, Germany) and filtered in the second step using membrane filters with a pore diameter of 0.22 µm (Millex-GS, Millipore, Burlington, MA, USA). An Aminex HPX-87P column (Bio-Rad, Hercules, CA, USA) was used for the separation of glucose and ethanol with a 5 mM solution of H₂SO₄ (mobile phase) at a flow rate of 0.6 mL/min at 30 °C. The detection of glucose and ethanol was performed with a refractive index detector (Model L-7490, Merck Hitachi, Darmstadt, Germany).

3. Experimental Strategy and Overview of Proposed Machine Learning Methods

3.1. Materials—Data for the Model

To create the model, empirical data from experiments on the bioethanol preparation from the following types of lignocellulosic biomass were used: buckwheat straw, biomass from four wastelands, including a mixture of various plants: stems of giant miscanthus, common nettle, goldenrod, common broom, fireweed, and hay. The production process of bioethanol from the abovementioned types of biomass was carried out on the basis of an identical production scheme, which included disintegration of the raw material, pretreatment with the use of IL, enzymatic hydrolysis with the use of five enzymatic preparations and alcoholic fermentation with the use of *Saccharomyces cerevisiae* type II or *Saccharomyces cerevisiae* Ethanol Red yeast. The cellulose, hemicellulose and lignin amounts were determined in each material. After enzymatic hydrolysis, glucose content was determined in the samples, whereas after alcoholic fermentation, ethyl alcohol content was determined. A total of 120 experiments were conducted, on which basis two model types were created. Model validation was carried out on the basis of the concentrations of bioethanol obtained from hemp and mugwort. In summary, 26 experiments were performed and the results obtained were used to validate the model. Biomass samples—mugwort and hemp—were pretreated with the use of various ILs and enzymatic hydrolysis with the use of enzyme preparations. Both in the native material and after pretreatment with ILs, determinations were made for the content of cellulose, hemicellulose, and lignin.

In Model 1 (Figure 1), the following input data were determined: biomass composition, including the content of cellulose (%), hemicellulose (%), lignin (%), and types of ILs used for pretreatment expressed in amount [mL]: [BMIM][OAc], [CHDMA-C4][OAc], [CHDMA-C6][OAc], [EMIM][DEP], [EMIM][OAc], EMIM[Cl], as well as types of enzymatic preparations expressed in amounts (g/L) added in the process of enzymatic hydrolysis and glucose content (g/L) tested after this process.

In Model 2 (Figure 1), the input data were arranged in a different manner and they consisted of the following variables: biomass composition (content of cellulose (%), hemicellulose (%) and lignin (%)), types of ionic liquids ([BMIM][OAc], [CHDMA-C4][OAc], [CHDMA-C6][OAc], [EMIM][DEP], [EMIM][OAc], [EMIM][Cl]), their amounts [mL], and time of purifying material in ionic liquids [min]. Input data of the enzymatic hydrolysis process included the possibility of additions of combinations of two enzymes at the same time—expressed as addition of enzyme 1 and their amount and addition of enzyme 2 and their amount [µL]. The last variable in this model is the content of glucose tested after 72 h of enzymatic hydrolysis. This was necessary because two enzyme preparations simultaneously were used in certain processes to hydrolyze them in order to increase the content of simple sugars in the fermented solutions. In both models, the starting variable was the concentration of bioethanol (g/L) tested after 96 h of alcoholic fermentation. To create models, Matlab for RF and Keras and Tensorflow library in Python for ANN were used.

Model 1

Input features

Composition			Pretreatment				Hydrolize						
U ₁ -Cellulose [%]	U ₂ -Hemicellulose [%]	U ₃ -Lignin [%]	U ₄ -[BMIM][OAc]	U ₅ -[CHDMA-C4][OAc]	U ₆ -[EMIM][DEP]	U ₇ -[EMIM][OAc]	U ₈ -[EMIM]Cl	U ₉ -Aspergillus	U ₁₀ -Cellic Ctec 2	U ₁₁ -T. reesei	U ₁₂ -Viscozyme	U ₁₃ -Xylanaza	U ₁₄ -Glucose

Output

Ethanol

Model 2

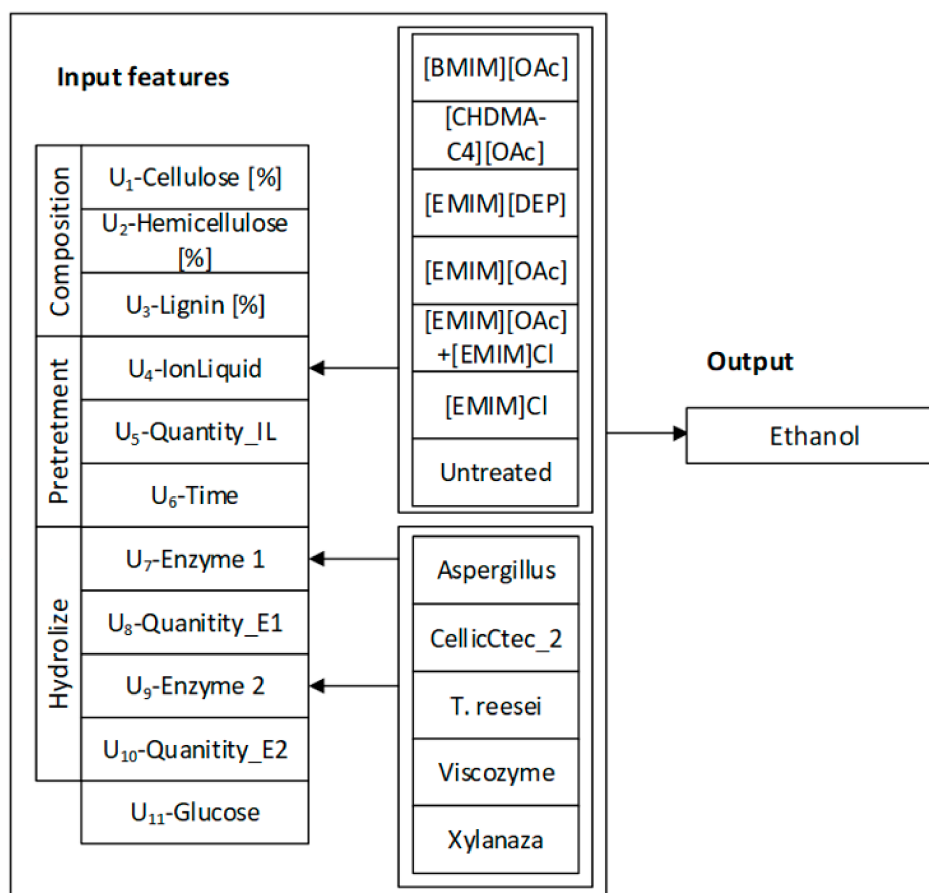


Figure 1. Two models used for the artificial neural network (ANN) and random forest algorithm (RF).

3.2. Methods of Machine Learning Used to Predict Bioethanol Content

3.2.1. Artificial Neural Networks

To carry out the ethanol content predicting process, multilayer perceptron (MLP) artificial neural networks were used, with the architecture shown in Figure 2 (for Model 1) and in Figure 3 (for Model 2).

For both model types, the training dataset comprised 120 vectors (14 elements for Model 1 and 11 elements for Model 2). Tables 1 and 2 present the designations of inputs

and outputs of the artificial neural network from Figure 1 (Model 1) and Figure 2 (Model 2), respectively.

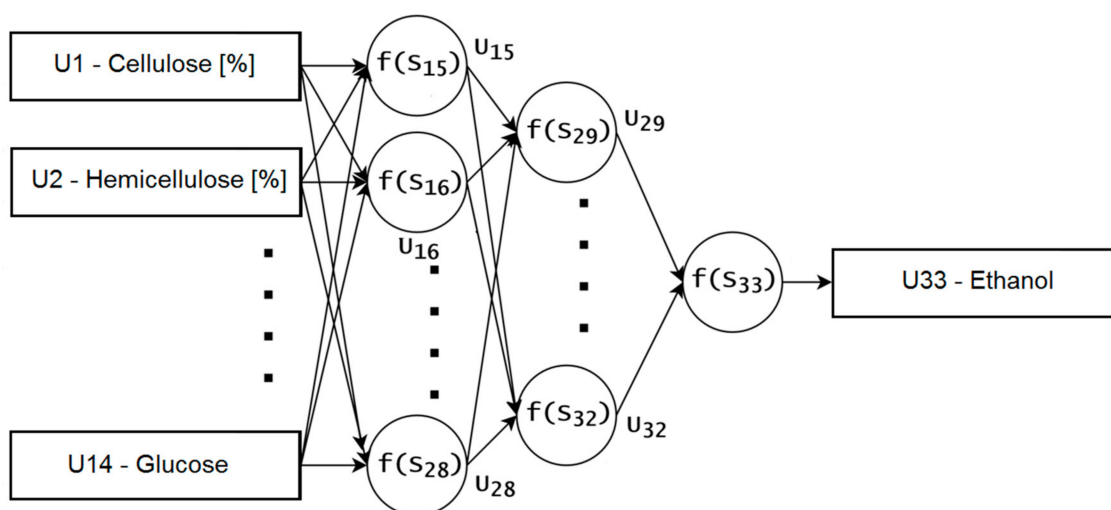


Figure 2. Multilayer perceptron (MLP) artificial neural network for Model 1 implementation.

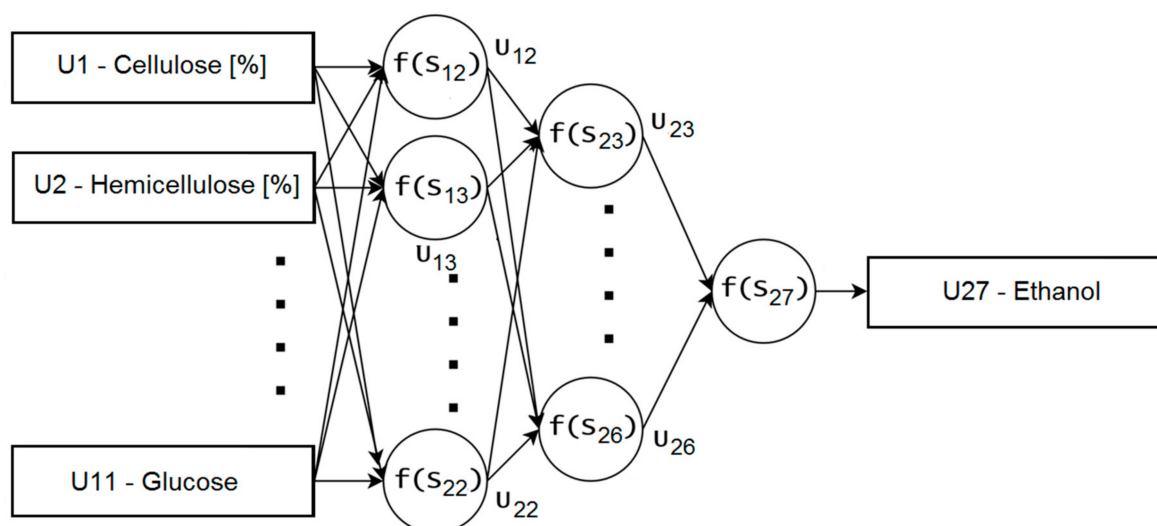


Figure 3. MLP artificial neural network for Model 2 implementation.

Table 2. Validation of the models for bioethanol production from *Cannabis sativa* L. and *Artemisia vulgaris* L.

	Model 1		Model 2		Hybrid 1		Hybrid 2	
	ANN	RF	ANN	RF	Model 1	Model 2	Model 1	Model 2
R ²	0.78	0.94	0.88	0.96	0.95	0.96	0.96	0.96
RMSE hemp	2.08	1.20	1.55	0.99	1.25	1.04	1.25	0.82
RMSE mugwort	0.59	0.55	0.28	0.46	0.32	0.36	0.19	0.33
ME hemp	0.59	0.77	0.73	0.71	0.59	0.67	0.84	0.16
ME mugwort	0.40	−0.47	−0.08	−0.43	−0.26	−0.31	−0.10	−0.24
MAE hemp	1.96	0.96	1.38	0.78	1.02	0.92	1.05	0.74
MAE mugwort	0.41	0.47	0.24	0.43	0.26	0.31	0.14	0.29
est_mean hemp	5.50	5.31	5.36	5.38	5.50	5.42	5.25	5.93
est_mean mugwort	1.16	2.04	1.64	1.99	1.82	1.87	1.66	1.81
est_median hemp	5.40	5.03	5.40	5.03	4.69	4.95	4.65	5.38
est_median mugwort	1.24	1.86	1.65	1.92	1.67	1.80	1.56	1.68

The error back-propagation algorithm with a constant value of the learning coefficient $\alpha = 0.01$ was adopted as the training algorithm. Before starting the training process, each of the training sets was subjected to the normalization process according to the relationship:

$$U_i^* = \frac{U_i - \text{mean}_i}{\text{stdev}_i} \quad (1)$$

where $i \in \{1, \dots, 14\}$ (for Model 1) or $i \in \{1, \dots, 11\}$ (for Model 2), U_i^* —value of input of neural network after normalization, mean_i —mean value for all training data for i -th input parameter (attribute), and stdev_i —standard deviation value for i -th input attribute.

The mean_i mean value and stdev_i standard deviation values determined were saved to be used for normalization of input data during the process of predicting ethanol value on the test set. The test set comprised 17 vectors (14 elements for Model 1 and 11 elements for Model 2).

After the training data normalization process was carried out, both models were trained with the corresponding training data. The training data comprised two phases. The forward propagation phase consists of randomly selected training vector values of weighed sums S_k and values of $f(S_k)$ outputs for all neurons determined, according to the relationship:

$$S_k = \sum_{t=m}^n w_{k,t} * U_t + w_{k,0} \quad (2)$$

where: $k \in \{15, \dots, 33\}$ (for Model 1) or $k \in \{12, \dots, 27\}$ (for Model 2). In addition, for Model 1 $m = 1$ and $n = 14$ for input layer neurons ($k \in \{15, \dots, 28\}$), $m = 15$ and $n = 28$ for intermediate layer neurons ($k \in \{29, \dots, 32\}$), and $m = 29$ and $n = 32$ for output layer neuron ($k = 33$); for Model 2 $m = 1$ and $n = 11$ for input layer neurons ($k \in \{12, \dots, 22\}$), $m = 12$ and $n = 22$ for intermediate layer neurons ($k \in \{23, \dots, 26\}$), and $m = 23$ and $n = 26$ for output layer neuron ($k = 27$); $w_{k,t}$ —value of the weight connecting U_k neuron with neuron or input U_t , and $w_{k,0}$ —value of the so-called threshold value for the U_k neuron.

Activation function for input layer neurons $\{U_{15}, \dots, U_{28}\}$ (for Model 1) $\{U_{12}, \dots, U_{22}\}$ (for Model 2) and intermediate layer $\{U_{29}, \dots, U_{32}\}$ (for Model 1) and $\{U_{23}, \dots, U_{26}\}$ (Model 2) is described with the following relationship:

$$U_k = f(S_k) = \frac{1}{1 + e^{-S_k}} \quad (3)$$

Function of activation for the output layer neuron U_{33} (for Model 1) and U_{27} (for Model 2) is described with relationship:

$$U_k = f(S_k) = S_k \quad (4)$$

After determining baseline values for all neurons, the forward propagation phase is complete and the signal backward propagation phase begins. This phase consists of determining the values of derivatives for all neurons according to the following relationships.

For neurons of the input layer and intermediate layer:

$$U'_k = f'(S_k) = U_k(1 - U_k) \quad (5)$$

For output layer neuron:

$$U'_k = f'(S_k) = 1 \quad (6)$$

In the backward propagation phase, the so-called delta_k coefficients are also determined for each k -th neuron with the following the relationships:

For output layer:

$$\text{delta}_k = (C_k - U_k) * f'(S_k) \quad (7)$$

For intermediate layer:

$$\delta_k = \delta_t * w_{t,k} * f'(S_k) \quad (8)$$

For input layer:

$$\delta_k = \left(\sum_{t=x}^y w_{t,k} * \delta_t \right) * f'(S_k) \quad (9)$$

where: $x = 29$ and $y = 32$ (for Model 1) and $x = 23$ and $y = 26$ (for Model 2).

After determining δ_k values for all neurons, the backward propagation phase is completed. Then the process of modification of all weight values begins according to the relationship:

$$w_{k,t}^* = w_{k,t} + ro * \delta_k * U_t \quad (10)$$

where $w_{k,t}^*$ is the new value of $w_{k,t}$ weight providing a signal to the U_k neuron from neuron output/input U_t .

After completing the modification of the weights, the first iteration of the training algorithm ends. Then another training vector is randomly selected and the whole process (forward propagation phase, backward propagation phase, and weights update) is repeated until training is completed. The training lasted for 2000 iterations for each neural network described.

3.2.2. Random Forest Algorithm

Random forest is a nonparametric ML algorithm derived from the classification and regression tree. Characteristics of RF include resistance to noise, simplicity of tuning, and capacity to deal with high-dimensional nonlinear problems [29–32]. In this work, RF was used with an RF library in Matlab software and applied to describe the pretreatment and enzymatic hydrolysis. To ensure good predictive performance, the RF was assessed for 11 of the RF samples, similar to the ANN. The model whose RMSE (root mean square error) was the median of all errors was further assessed.

4. Results and Discussion

Mugwort is an example of biomass obtained without the need for cultivation and fertilization, with an average cellulose content of 45%, hemicellulose 13.8%, and lignin 20.4%. For comparison, the conversion was also carried out on hemp stalks, which have recently become very popular for functional reasons. Finola hemp stalks had an average cellulose content of 62%, hemicellulose 17%, and lignin 19%. In this study, for the production of bioethanol, ground plant stalks were used and the process was carried out by performing a pretreatment with the use of various ionic liquids and various enzyme preparations. Glucose content in mugwort samples depended on the type of pretreatment and enzyme preparation used. Table 1 presents glucose contents obtained after 72 h enzymatic hydrolysis, bioethanol content, and chemical composition of hemp and mugwort.

After the enzymatic hydrolysis of mugwort, the highest content of glucose was obtained in the samples where imidazolium ionic liquids ([EMIM][OAc] and [BMIM][OAc]), and Cellic CTec2 for enzymatic hydrolysis were used. A similar relationship was observed in the samples of hemp for reducing sugars, but the results were significantly higher as compared with common mugwort. The content of simple sugars after enzymatic hydrolysis with the use of Cellic CTec2 amounted to 12.27 g/L for material purified with [BMIM][OAc] and 11.32 g/L for biomass purified with [EMIM][OAc]. For comparison, in the sample of native hemp hydrolyzed with Cellic CTec2, 3.2 g/L glucose was obtained after 72 h.

In the experiments with the use of machine learning, including ANN and RF methods for the estimation of bioethanol concentration, results of experiments concerning the processing of hay, agricultural wastelands, and selected energy crops were utilized. The RF method exhibits different advantages than ANN. Each tree represents the learning process and each tree can select traits and samples at random [33]. The final prediction is obtained

by averaging the predictions concerning the trees. This enables efficient avoidance of excessive matching and the effect of single samples [34]. On the other hand, ANN is characterized by singular correlation or learning process. Furthermore, many earlier studies show that the RF may give better predictions for the same problem [35,36].

Experimental data concerning bioethanol production from hemp and mugwort stems were used for the validation of the ANN model. The raw material is characterized by high cellulose content, low lignin content, and better structural properties after processing with ionic liquids; i.e., the material is more porous and there is more area free and available for cellulolytic enzymes; thus enzymatic hydrolysis is facilitated and more efficient. The situation is completely different when a raw material such as common mugwort is used, as its cellulose content is lower by 50% and it contains considerably higher amounts of lignin and hemicellulose. In addition, after dissolution in ionic liquids, common mugwort stems are not deprived of lignin with the same efficiency as for hemp stems. High amounts of lignin remaining in biomass samples directed for enzymatic hydrolysis may be linked to a poorer course of the process, because lignin is an enzyme inhibitor [37]. In this case, the use of ML methods perfectly reflects the processes of pretreatment and enzymatic hydrolysis processes.

The type of biomass, as well as the contribution of cellulose in the composition of plants, has direct influence on the content of simple sugars, including glucose after enzymatic hydrolysis. Thus, the selection of biomass rich in cellulose, as was the case for hemp stems, should be linked to more efficient ethanol production (Figure 4). For Model 1, the use of Cellic CTec2 in enzymatic hydrolysis was most important because, regardless of the type of biomass used, i.e., whether it was weed, woody plants, or energy crops, high concentrations of glucose were obtained when the enzyme was used for the hydrolysis. The pretreatment of biomass and type of ionic liquid applied were also significant for Model 1. More favorable results were obtained in the case of imidazolium liquids, and the most important was the use of [EMIM][OAc] and [BMIM][OAc], for both energy crops—hemp, as well as woody weed with higher lignin content. The content of lignin in the biomass composition is another significant factor affecting the reduction of bioethanol production efficiency. In Model 2, this variable is in the third place, whereas second place is taken by the use of ionic liquids as pretreatment type. Dependable variables that affect the described model are also E1 and E2 enzymes, that is, *T. reesei* and Cellic CTec2 and their amounts, appropriately selected to the content of cellulose after pretreatment.

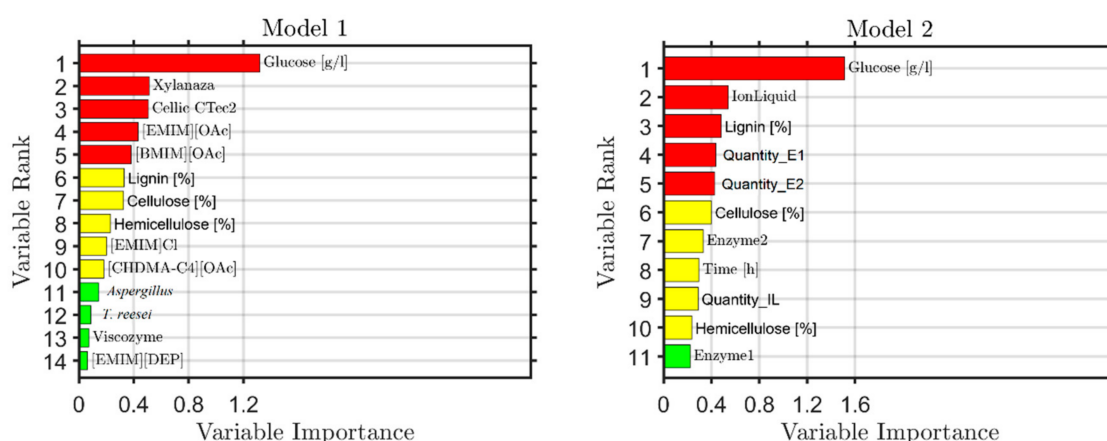


Figure 4. Relative importance of inputs for bioethanol estimation—random forest.

The use of xylanase as an additional enzyme in enzymatic hydrolysis resulted in increased content of simple sugars by decomposing hemicellulose, which is linked to cellulose. Enzymatic digestion of hemicellulose resulted in exposing cellulose fibers, which were then digested by cellulase. Therefore, the use of xylanase in such cases directly

contributed to the increase of the content of glucose in the samples. Considering the costs of the process, the use of an additional enzyme (xylanase) is not valid.

Ahmadian-Moghadam et al. [38] examined the influence of the initial concentration of the substrate (molasses), live yeast cells, and dead yeast cells as the input parameters of the process on the production of bioethanol via *Saccharomyces cerevisiae*. An R^2 value equal to 0.93 was obtained, which shows that the model was suitable for pattern recognition into data and these patterns precisely predicted ethanol efficiency. In the latest research conducted by Betiku and Taiwo [39], the influence of breadfruit hydrolysate concentration, hydraulic retention time, and pH on the production of bioethanol was assessed with ANN and response surface methodology (RSM). The ANN had an absolute mean deviation between the predicted and observed value of 0.09%, compared to 1.67% after RSM [39]. These results further confirm the precision of ANN modeling in comparison with other techniques, such as RSM.

ANN and RF algorithms have a random training start point, thus they were repeated 11 times to ensure higher reliability of the results obtained. The following analyses present results of the iteration whose error was a median of error from 11 replications. In the learning process, the R^2 determination coefficient for Model 1 was 0.92 for ANN and 0.93 for RF. In Model 2, the R^2 coefficient increased to near 1 for ANN, whereas for RF it remained at the same level (Figure 5).

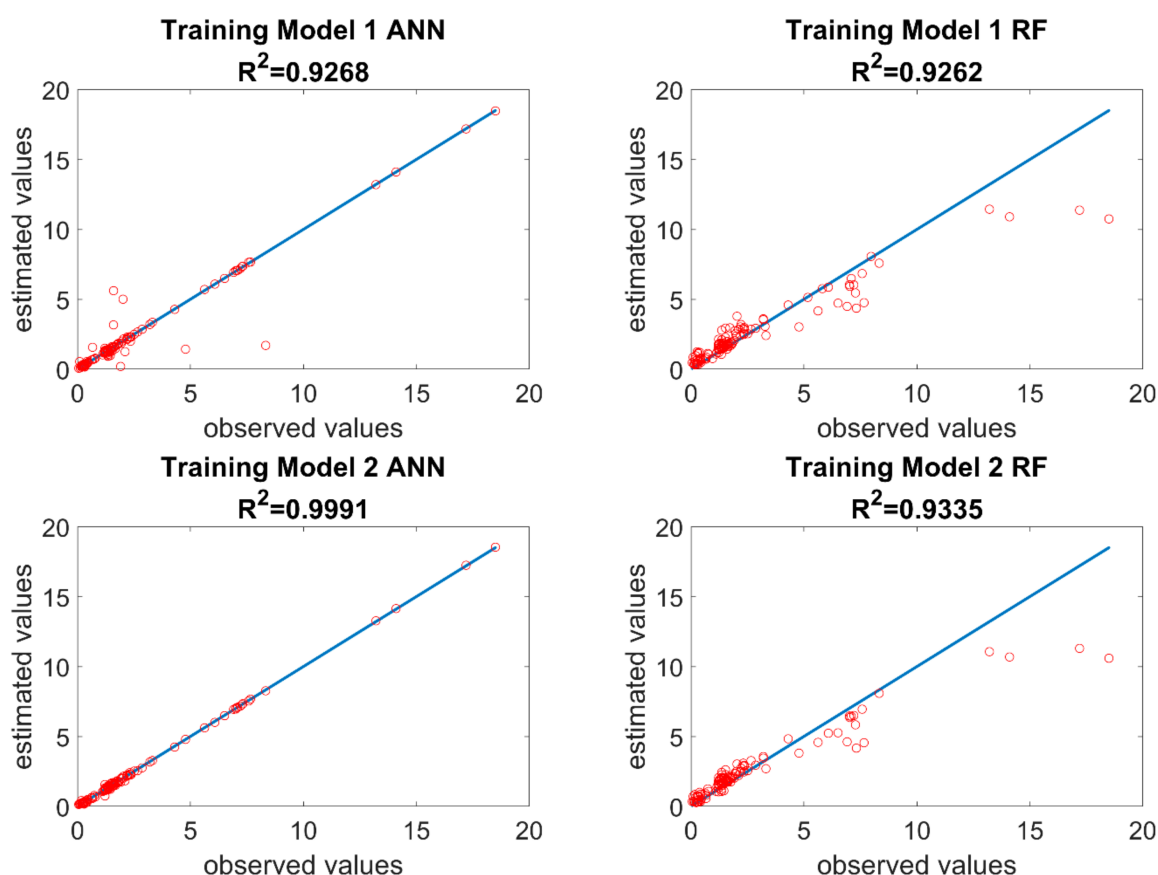


Figure 5. Training models for artificial neural network (ANN) and random forest (RF).

Data for mugwort and hemp were used for validation. In both models, the determination coefficient for neural networks >0.9 , whereas for RF it oscillated around 0.95. In the RF training models, there were four wasteland samples whose observed values significantly differ from the estimated values. The ethanol content of these samples was significantly different from the others and due to the principle of operation of the RF algorithm, those samples could not be included. Considering the rather wide dispersion of

the determination coefficient, two hybrid models were executed. The first hybrid model (Hybrid 1) consisted of assuming the median from the set of data estimated from 22 replications. The second hybrid model (Hybrid 2) assumed determination of a linear function describing the variables from the entire training set of 22 replications (11 ANN and 11 RF). Subsequently, median from the validation set was calculated, on the basis of which new estimated values were calculated. The last step of the process was to calculate the value closest to estimated values from the set of 22 values. During the calculation of these models, 70% of points in Model 1 from ANN were selected and 75% of points in Model 2 from ANN were selected. When both hybrid models were applied, a clear increase of determination coefficient can be observed with regards to ANN, and a considerable decrease of RMSE (root mean square error), ME (mean error), and MAE (mean absolute error) for mugwort and hemp. Furthermore, mean values (est_mean) and medians (est_median) of estimates were calculated for each model and algorithm.

Table 2 above describes the sum parameters concerning the presented models of bioethanol production from hemp and common mugwort. The R^2 determination coefficient depended on the type of applied model. The hybrid approach in the creation of models explaining this process of bioethanol production resulted in increased match of the model to $R^2 = 0.961$ for Hybrid 2. In the original calculations, R^2 reached about 0.96 match only for Model 2–RF. Precision of Model 1–Hybrid 2 for the prediction of ethanol production process from biomass is satisfactory and higher than the ANN and RF models. RMSE values for the RF algorithm in each case of validation sample analysis, that is, hemp and mugwort biomass, were lower, and the model was better matched than for ANN. Moreover, differences in R^2 and RMSE relative to the analyzed material can be observed. RMSE was lower for common mugwort samples, the results of which were predicted with the use of ANN. A reverse situation occurred for validation of the model utilizing hemp samples. In this case, lower RMSE with better match of the model ($R^2 = 0.961$) was obtained for Model 2–RF. Considering that the differences were significant and did not provide a clear answer, it was decided to use a hybrid model, which vastly improved the effects of predicting bioethanol concentrations from lignocellulosic materials and provided a better match of the hybrid model to experimental results–validation, presented in Figure 6. The pink color in the Figure 6 was used to mark experimental results of bioethanol concentration obtained from common mugwort stems, and green was used to mark bioethanol concentrations obtained from cannabis stems. The blue asterisk refers to RF values and the red asterisk to ANN values.

In the case of modeling such complex processes as bioethanol production from different lignocellulosic raw materials and taking into account numerous initial samples, the use of only one algorithm type results in difficulties. Due to the concerns that the use of ANN would result in flattening or not using all process conditions and mechanisms, scientists often refer to the comparison with such algorithms as random forest, adaptive neuro–fuzzy inference system, and support vector machine [35,36,40]. The application of a hybrid approach to the discussed issues aimed for a more comprehensive inclusion of the mechanisms that are not yet discovered in bioethanol production, or have not yet been classified as of key importance on bioethanol concentration. In this study, the hybrid model is well matched to the process presented, and it further includes a very wide spectrum of lignocellulosic biomass, not including raw materials due to, e.g., an excessive amount of lignin. This may largely contribute to the expansion of knowledge in the field of bioethanol production from mixed types of lignocellulosic biomass with different chemical compositions, and acceleration of the selection of pretreatment type based only on several input variables.

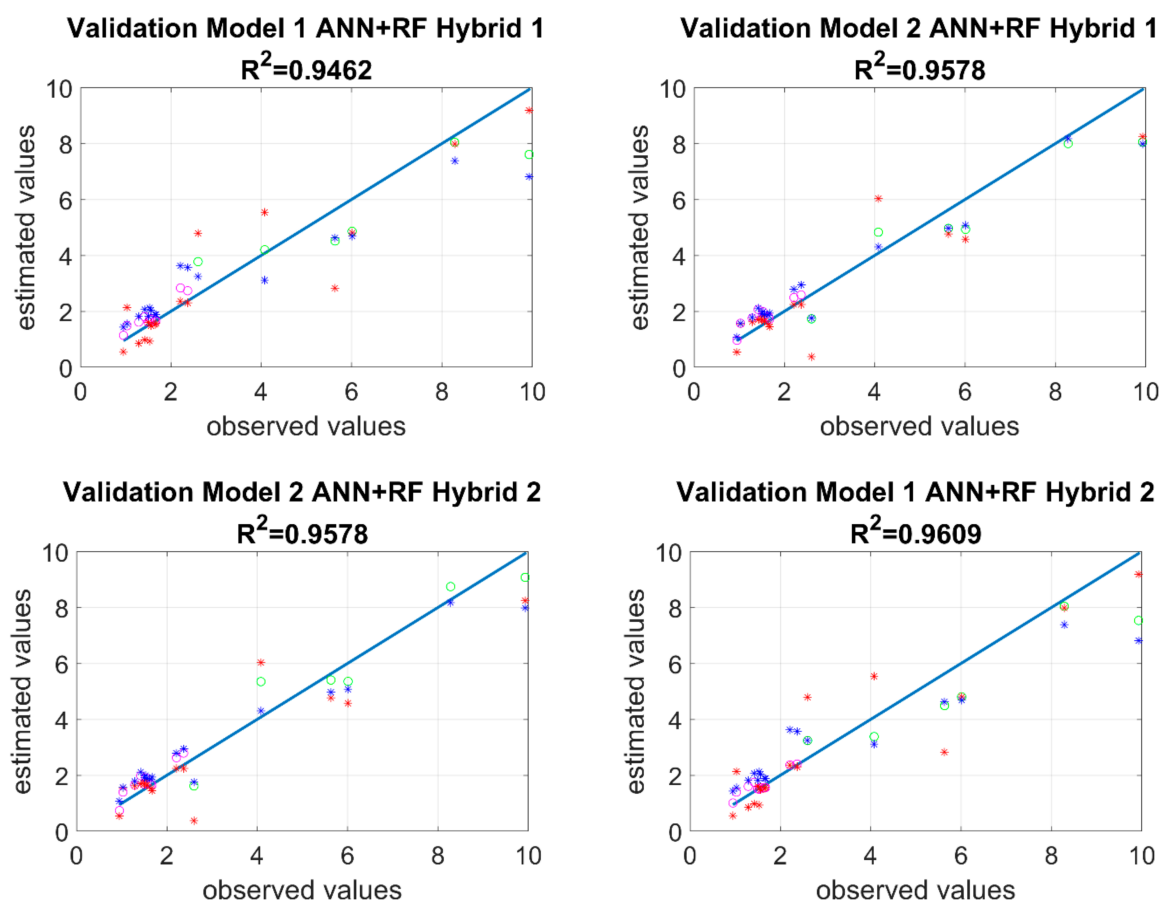


Figure 6. Measured values relative to approximated values for the Hybrid 1 and Hybrid 2 models.

5. Summary

The use of machine learning methods, i.e., ANN and RF, for the prediction biotechnological processes outcomes, in our case bioethanol production, even at a laboratory scale is a very good first step to understand the production mechanism, before going to a large scale. Results of this study suggest that ML is a good tool to predict the final concentration of ethanol obtained in a multistage process of hydrolysis and fermentation of lignocellulosic biomass. Data for this model includes results of bioethanol production with the use of ILs and different enzymatic preparations from the following biomass types: buckwheat straw and biomass from four wastelands, including a mixture of various plants—stems of giant miscanthus, common nettle, goldenrod, common broom, fireweed, and hay. The results obtained for each of the models applied are in a very good agreement with the experimental results. For the process, two extreme biomass cases (hemp and mugwort) were used and the simulations determined the final ethanol value with high likelihood. Importantly, the ANN model alone qualifies the biomass as a good source of bioethanol, mainly on the basis of the cellulose content (as in the case of hemp). The RF, on the other hand, also takes into consideration other variables, such as lignin content. Therefore, the hybrid model proposed is more adequate and takes into consideration other constituents and the level of their changes during the pretreatment process. The hybrid model can be successfully used for the preliminary classification of plants on the basis of the results of their lignocellulosic composition, which means that the selection of an appropriate biomass source can be carried out without long-term and often expensive research. ML is a perfect tool for these types of processes, which can be developed by means of continuous network training. The quality of this study indicates that further research results on the production

of bioethanol from lignocellulose can be used for extending and continuously increasing the verification of the hybrid model.

Author Contributions: Conceptualization, M.S.-K. and T.K.; methodology, M.S.-K., T.K., R.M., and A.S.; software and validation, T.K. and A.S.; formal analysis, M.S.-K., T.K., R.M., and A.S.; investigation, M.S.-K., T.K., R.M., and A.S.; writing—review and editing, M.S.-K., T.K., and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Elgharbawy, A.A.; Alam, M.Z.; Moniruzzaman, M.; Goto, M. Ionic liquid pretreatment as emerging approaches for enhanced enzymatic hydrolysis of lignocellulosic biomass. *Biochem. Eng. J.* **2016**, *109*, 252–267. [\[CrossRef\]](#)
2. Baral, N.R.; Shah, A. Comparative techno-economic analysis of steam explosion, dilute sulfuric acid, ammonia fiber explosion and biological pretreatments of corn stover. *Bioresour. Technol.* **2017**, *232*, 331–343. [\[CrossRef\]](#)
3. Jönsson, L.J.; Alriksson, B.; Nilvebrant, N.-O. Bioconversion of lignocellulose: Inhibitors and detoxification. *Biotechnol. Biofuels* **2013**, *6*, 16. [\[CrossRef\]](#)
4. Konda, N.; Shi, J.; Singh, S.; Blanch, H.W.; Simmons, B.A.; Klein-Marcuschamer, D. Understanding cost drivers and economic potential of two variants of ionic liquid pretreatment for cellulosic biofuel production. *Biotechnol. Biofuels* **2014**, *7*, 86. [\[CrossRef\]](#) [\[PubMed\]](#)
5. He, L.; Wang, C.; Shi, H.; Zhou, W.; Zhang, Q.; Chen, X. Combination of steam explosion pretreatment and anaerobic alkalization treatment to improve enzymatic hydrolysis of *Hippophae rhamnoides*. *Bioresour. Technol.* **2019**, *289*, 121693. [\[CrossRef\]](#) [\[PubMed\]](#)
6. da Silva, S.P.M.; da Lopes, A.M.C.; Roseiro, L.B.; Bogel-Lukasik, R. Novel pre-treatment and fractionation method for lignocellulosic biomass using ionic liquids. *RSC Adv.* **2013**, *3*, 16040–16050. [\[CrossRef\]](#)
7. Momayez, F.; Karimi, K.; Karimi, S.; Horváth, I.S. Efficient hydrolysis and ethanol production from rice straw by pretreatment with organic acids and effluent of biogas plant. *RSC Adv.* **2017**, *7*, 50537–50545. [\[CrossRef\]](#)
8. da Costa Lopes, A.M.; João, K.G.; Bogel-Lukasik, E.; Roseiro, L.B.; Bogel-Lukasik, R. Pretreatment and Fractionation of Wheat Straw Using Various Ionic Liquids. *J. Agric. Food Chem.* **2013**, *61*, 7874–7882. [\[CrossRef\]](#)
9. Rosenau, T.; Potthast, A.; Sixta, H.; Kosma, P. The chemistry of side reactions and byproduct formation in the system NMMO/cellulose (Lyocell process). *Prog. Polym. Sci.* **2001**, *26*, 1763–1837. [\[CrossRef\]](#)
10. Goshadrour, A.; Karimi, K.; Taherzadeh, M. Ethanol and biogas production from birch by NMMO pretreatment. *Biomass Bioenergy* **2013**, *49*, 95–101. [\[CrossRef\]](#)
11. Sewsynker-Sukai, Y.; Faloye, F.; Kana, E.B.G. Artificial neural networks: An efficient tool for modelling and optimization of biofuel production (a mini review). *Biotechnol. Biotechnol. Equip.* **2017**, *31*, 221–235. [\[CrossRef\]](#)
12. Fischer, J.; Lopes, V.S.; Cardoso, S.L.; Coutinho Filho, U.; Cardoso, V.L. Machine learning techniques applied to lignocellulosic ethanol in simultaneous hydrolysis and fermentation. *Braz. J. Chem. Eng.* **2017**, *34*, 53–63. [\[CrossRef\]](#)
13. Bohdal, T.; Charun, H.; Kruzel, M.; Sikora, M. An investigation of heat transfer coefficient during refrigerants condensation in vertical pipe minichannels. *E3S Web Conf.* **2018**, *70*, 02001. [\[CrossRef\]](#)
14. Gueguim Kana, E.B.; Oloke, J.K.; Lateef, A.; Adesiyun, M.O. Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm. *Renew. Energy* **2012**, *46*, 276–281. [\[CrossRef\]](#)
15. Shi, Y.; Gai, G.; Zhao, X.; Zhu, J.; Zhang, P. Back Propagation Neural Network (BPNN) Simulation Model and Influence of Operational Parameters on Hydrogen Bio-Production through Integrative Biological Reactor (IBR) Treating Wastewater. In Proceedings of the 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, 18–20 June 2010; pp. 1–4.
16. Almeida, J.S. Predictive non-linear modeling of complex data by artificial neural networks. *Curr. Opin. Biotechnol.* **2002**, *13*, 72–76. [\[CrossRef\]](#)
17. Vani, S.; Sukumaran, R.K.; Savithri, S. Prediction of sugar yields during hydrolysis of lignocellulosic biomass using artificial neural network modeling. *Bioresour. Technol.* **2015**, *188*, 128–135. [\[CrossRef\]](#)
18. Das, S.; Bhattacharya, A.; Haldar, S.; Ganguly, A.; Gu, S.; Ting, Y.P.; Chatterjee, P.K. Optimization of enzymatic saccharification of water hyacinth biomass for bio-ethanol: Comparison between artificial neural network and response surface methodology. *Sustain. Mater. Technol.* **2015**, *3*, 17–28. [\[CrossRef\]](#)
19. Giordano, P.C.; Beccaria, A.J.; Goicoechea, H.C.; Olivieri, A.C. Optimization of the hydrolysis of lignocellulosic residues by using radial basis functions modeling and particle swarm optimization. *Biochem. Eng. J.* **2013**, *80*, 1–9. [\[CrossRef\]](#)
20. Gitifar, V.; Eslamloueyan, R.; Sarshar, M. Experimental study and neural network modeling of sugarcane bagasse pretreatment with H₂SO₄ and O₃ for cellulosic material conversion to sugar. *Bioresour. Technol.* **2013**, *148*, 47–52. [\[CrossRef\]](#)
21. Struik, P.; Amaducci, S.; Bullard, M.J.; Stutterheim, N.; Venturi, G.; Cromack, H.T.H. Agronomy of fibre hemp (*Cannabis sativa* L.). *Ind. Crops Prod.* **2000**, *11*, 107–118. [\[CrossRef\]](#)
22. Burczyk, H. Oilseed hemp (*Cannabis sativa* L. var. *oleifera*) grown for seeds, oil and biogas. *Probl. Inż. Rol.* **2016**, *24*, 109–116.

23. Schluttenhofer, C.; Yuan, L. Challenges towards Revitalizing Hemp: A Multifaceted Crop. *Trends Plant Sci.* **2017**, *22*, 917–929. [[CrossRef](#)] [[PubMed](#)]
24. Kraszkiewicz, A.; Kachel, M.; Parafiniuk, S.; Zajac, G.; Niedziółka, I.; Sprawka, M. Assessment of the Possibility of Using Hemp Biomass (*Cannabis Sativa* L.) for Energy Purposes: A Case Study. *Appl. Sci.* **2019**, *9*, 4437. [[CrossRef](#)]
25. Holst, N.; Rasmussen, I.; Bastiaans, L. Field weed population dynamics: A review of model approaches and applications. *Weed Res.* **2007**, *47*, 1–14. [[CrossRef](#)]
26. Jezierska-Domaradzka, A.; Kuźniewski, E. Ruderal plants within segetal communities in south-west part of Poland. *Zesz. Przyr. OTPN* **2000**, *34*, 5–11.
27. Pernak, J.; Kordala, R.; Markiewicz, B.; Walkiewicz, F.; Popławski, M.; Fabiańska, A.; Jankowski, S.; Łożyński, M. Synthesis and properties of ammonium ionic liquids with cyclohexyl substituent and dissolution of cellulose. *RSC Adv.* **2012**, *2*, 8429–8438. [[CrossRef](#)]
28. Van Soest, P.J.; Robertson, J.B.; Lewis, B.A. Methods for Dietary Fiber, Neutral Detergent Fiber, and Nonstarch Polysaccharides in Relation to Animal Nutrition. *J. Dairy Sci.* **1991**, *74*, 3583–3597. [[CrossRef](#)]
29. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: London, UK, 1983.
30. Breiman, L. Random Forests. *Mach Learn* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Seyedhosseini, M.; Tasdizen, T. Disjunctive normal random forests. *Pattern Recognit.* **2015**, *48*, 976–983. [[CrossRef](#)]
32. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *2*, 18–22.
33. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; IEEE Computer Society Press: Montreal, QC, Canada, 1995; Volume 1, pp. 278–282.
34. Hu, S.; Liu, H.; Zhao, W.; Shi, T.; Hu, Z.; Li, Q.; Wu, G. Comparison of Machine Learning Techniques in Inferring Phytoplankton Size Classes. *Remote Sens.* **2018**, *10*, 191. [[CrossRef](#)]
35. You, H.; Ma, Z.; Tang, Y.; Wang, Y.; Yan, J.; Ni, M.; Cen, K.; Huang, Q. Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. *Waste Manag.* **2017**, *68*, 186–197. [[CrossRef](#)] [[PubMed](#)]
36. Xing, J.; Wang, H.; Luo, K.; Wang, S.; Bai, Y.; Fan, J. 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* **2019**, *136*, 104–114. [[CrossRef](#)]
37. Kellock, M.; Maaheimo, H.; Marjamaa, K.; Rahikainen, J.; Zhang, H.; Holopainen-Mantila, U.; Ralph, J.; Tamminen, T.; Felby, C.; Kruus, K. Effect of hydrothermal pretreatment severity on lignin inhibition in enzymatic hydrolysis. *Bioresour. Technol.* **2019**, *280*, 303–312. [[CrossRef](#)]
38. Ahmadian-Moghadam, H.; Elegado, F. Prediction of Ethanol Concentration in Biofuel Production Using Artificial Neural Networks. *Am. J. Model. Optim.* **2013**, *1*, 31–35. [[CrossRef](#)]
39. Betiku, E.; Taiwo, A.E. Modeling and optimization of bioethanol production from breadfruit starch hydrolyzate vis-à-vis response surface methodology and artificial neural network. *Renew. Energy* **2015**, *74*, 87–94. [[CrossRef](#)]
40. Grahovac, J.; Jokić, A.; Dodić, J.; Vučurović, D.; Dodić, S. Modelling and prediction of bioethanol production from intermediates and byproduct of sugar beet processing using neural networks. *Renew. Energy* **2016**, *85*, 953–958. [[CrossRef](#)]