*Article*

Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models

Dragana Rajković [1], Ana Marjanović Jeromela [1,*], Lato Pezo [2], Biljana Lončar [3], Federica Zanetti [4,*], Andrea Monti [4] and Ankica Kondić Špika [1]

[1] Institute of Field and Vegetable Crops, 21000 Novi Sad, Serbia; dragana.rajkovic@ifvcns.ns.ac.rs (D.R.); ankica.spika@ifvcns.ns.ac.rs (A.K.Š.)

[2] Institute of General and Physical Chemistry, 11000 Belgrade, Serbia; latopezo@yahoo.co.uk

[3] Faculty of Technology, University of Novi Sad, 21000 Novi Sad, Serbia; biljanacurcicc@gmail.com

[4] Department of Agricultural and Food Sciences (DISTAL), Università di Bologna—Alma Mater Studiorum, 40126 Bologna, Italy; a.monti@unibo.it

* Correspondence: ana.jeromela@ifvcns.ns.ac.rs (A.M.J.); federica.zanetti5@unibo.it (F.Z.);
  Tel.: +381-21-4898-111 (A.M.J.); +39-051-2096655 (F.Z.)

Abstract: As one of the greatest agricultural challenges, yield prediction is an important issue for producers, stakeholders, and the global trade market. Most of the variation in yield is attributed to environmental factors such as climate conditions, soil type and cultivation practices. Artificial neural networks (ANNs) and random forest regression (RFR) are machine learning tools that are used unambiguously for crop yield prediction. There is limited research regarding the application of these mathematical models for the prediction of rapeseed yield and quality. A four-year study (2015–2018) was carried out in the Republic of Serbia with 40 winter rapeseed genotypes. The field trial was designed as a randomized complete block design in three replications. ANN, based on the Broyden–Fletcher–Goldfarb–Shanno iterative algorithm, and RFR models were used for prediction of seed yield, oil and protein yield, oil and protein content, and 1000 seed weight, based on the year of production and genotype. The best production year for rapeseed cultivation was 2016, when the highest seed and oil yield were achieved, 2994 kg/ha and 1402 kg/ha, respectively. The RFR model showed better prediction capabilities compared to the ANN model (the r^2 values for prediction of output variables were 0.944, 0.935, 0.912, 0.886, 0.936 and 0.900, for oil and protein content, seed yield, 1000 seed weight, oil and protein yield, respectively).

Keywords: rapeseed; yield; oil content; mathematical modelling; artificial neural network; random forest

1. Introduction

High and stable yield and oil content are the most important traits in rapeseed (*Brassica napus* L.) breeding programs. According to [1], in the last five years the world average rapeseed yield was about 2.1 t/ha. Rapeseed seed yield and quality vary depending on location, cultivar and their mutual interaction [2,3]. Seed yield is mainly affected by environmental variation such as climatic factors (temperature, precipitation, length of photoperiod, abiotic stresses), soil type, and cultivation practice (density and time of sowing, fertilization). Due to the abovementioned factors, seed yield prediction is an exceedingly challenging task.

Early yield prediction especially comes to focus in years when extreme weather events unfavourably influence crop yield. Being able to forecast low yield leaves space to make on-time warning and develop a strategy to maintain a stable food supply chain. It is forecasted that in the near future precipitation levels will rise in northern Europe, which is among others expected to reflect on higher seed yield [4]. On the other hand, southern Europe will suffer from high temperatures accompanied by drought, which both adversely

affect yield [5]. Ref. [6] tested several regression models and concluded that in all of the models an increase in precipitation during autumn and winter was negatively correlated with the yield of winter rapeseed, whereas a temperature rise during flowering had a positive effect. However, different authors claimed that higher temperatures negatively affect rapeseed yield [7–9]. Differences in the observed temperature effects may have risen because of different growing conditions in locations where the trials were set up. Namely, the trials of [6] were conducted in Denmark, where the climate was cooler with temperate springs. Hence, it is possible that measured temperatures did not surpass critical values as in [7–9], when they reflected negatively on seed yield. Seed and silique forming and development are the most important phenological phases that affect yield, which is mainly determined before ripening [10].

Yield prediction is an important part of the precision agriculture concept. Knowledge of weather and plant conditions may assist farmers, big producers, output buyers and suppliers in the early prediction of crop yield by providing them with valuable information regarding return and expected financial benefit. Data gained via remote sensing imaging with unmanned aerial vehicles (UAVs) are not only valuable for monitoring crop conditions, especially changes in crop nitrogen concentration, disease occurrence, flowering time and pod ripening [11–13], but can also help in estimating final yield. In the study of [14], remote sensing of vegetation indices with UAV during flowering was used to estimate rapeseed yield before harvest. They tested various vegetation indices, where the most accurate had an estimation error under 13%. Machine learning models are handy for different tasks, especially when considering living systems in which linear regression models often disregard complex interactions between variables. In [15], an enhanced vegetation index, solar-induced chlorophyll fluorescence, climate and different combinations of the mentioned variables were used as input data to compare the performance of different models. Non-linear models, such as random forest regression (RFR) and neural networks, outperformed linear models mostly because relations between examined variables were non-linear. Ref. [16] emphasized the efficacy of RFR in staple crop yield prediction. An RFR model that relies on near-infrared vegetation reflectance during several growth stages was successfully used to forecast rapeseed yield [17]. Apart from regression analysis, cutting-edge statistical models that use artificial neural network (ANN) models can be incorporated into yield predictions [18]. In addition, machine learning models are capable of establishing patterns and correlations among data [19]. Still, they do not reveal the actual cause of a relationship. This is why the dataset for a model of interest needs to go through a training phase first.

Lately, ANN was used for the estimation of crop yield and quality [18,20]. ANN models are considered to have higher accuracy in comparison with regression models [21]. The number of hidden nodes influences the precision of yield prediction in terms that models with fewer nodes than the starting number of nodes are better [21]. Ref. [22] reported that machine learning models perceive seed yield as a function of input variables, such as genotypes and environments. Ref. [23] developed an ANN with weather, soil and management data as input and predicted maize yield with 80% accuracy. Ref. [24] predicted genotypic effects of rapeseed lines and hybrids with the aid of SNP markers. Since the correlation of genetic prediction with phenotype for yield-related traits produced similar values, such as estimated heritability, it was highlighted that this approach could be used to predict best-performing genotypes [24].

ANN is used as an additional tool to assist in the seed classification of rapeseed varieties [25,26]. In [18], input consisted of quantitative (precipitation, temperatures, applied fertilizers) and qualitative data (fertilizer type, liming, tillage type, sowing date and previous forecrop). Ref. [18] tested three models that differed in terms of predictive dates for plant development stages to foresee rapeseed yield.

To the best of the authors' knowledge, this is the first study to use the ANN model to predict yield-related traits as well as oil and protein content in rapeseed. The objective of this study was to investigate the possibility of predicting oil and protein content, seed yield,

oil and protein yield, and 1000 seed weight, based on the year of production and genotype, using artificial neural network (ANN) and random forest regression (RFR) models.

2. Materials and Methods

2.1. Plant Material and Trial Design

Six traits related to yield and seed quality were surveyed (i.e., oil and protein content (OC, PC), seed yield (SY), oil and protein yield (OY, PY), and thousand seed weight (TSW)) on 40 winter-type rapeseed genotypes (Table 1) during four consecutive years (2015–2018).

Table 1. Winter rapeseed genotypes tested during 2015–2018.

Number	Genotype	Type of Material	Origin	Registration Year
1	NS-H-R-1	hybrid	Serbia	n.r. *
2	NS-H-R-2	hybrid	Serbia	n.r.
3	NS-H-R-3	hybrid	Serbia	n.r.
4	Banaćanka	cultivar	Serbia	1998
5	Slavica	cultivar	Serbia	2003
6	Valeska tamna	line	Sweden	n.r.
7	Valeska svetla	line	Sweden	n.r.
8	Zlatna	cultivar	Serbia	2008
9	NS-L-74	line	Serbia	n.r.
10	Branka	cultivar	Serbia	2007
11	Express	cultivar	Germany	1993
12	NS-L-7	line	Serbia	n.r.
13	Nevena	cultivar	Serbia	2008
14	Valeska	cultivar	Sweden	EU **
15	Ilia	cultivar	Serbia	2011
16	Kata	cultivar	Serbia	2006
17	Nena	cultivar	Serbia	2005
18	NS-L-31	line	Serbia	n.r.
19	NS-L-126	line	Serbia	n.r.
20	NS-L-33	line	Serbia	n.r.
21	NS-L-128	line	Serbia	n.r.
22	Svetlana	cultivar	Serbia	2016
23	Jasna	cultivar	Serbia	2009
24	NS-L-101	line	Serbia	n.r.
25	Zorica	cultivar	Serbia	2010
26	NS-L-102	line	Serbia	n.r.
27	NS-L-134	line	Serbia	n.r.
28	NS-L-32	line	Serbia	n.r.
29	NS-L-136	line	Serbia	n.r.
30	NS-L-137	line	Serbia	n.r.
31	NS-L-138	line	Serbia	n.r.
32	NS-L-251	line	Serbia	n.r.
33	NS-L-210	line	Serbia	n.r.
34	NS-L-44	line	Serbia	n.r.
35	NS-L-45	line	Serbia	n.r.
36	NS-L-46	line	Serbia	n.r.
37	NS-L-47	line	Serbia	n.r.
38	Jelena	cultivar	Serbia	2011
39	Forward	line	Serbia	n.r.
40	Maidan	line	Serbia	n.r.

* n.r. non-registered; ** EU registered within the European Union.

The trial was set up as a randomized complete block design with three replicates at Rimski šančevi, Serbia (45°19′53.7″ N 19°50′12.6″ E). The size of the experimental plot was sized 4×1.5 m with 55–65 plants/m^2 at the harvest. Sowing and harvesting were carried out in the optimum times, which are in September and June in each season. Prior to sowing, the soil received an adequate amount of NPK 15:15:15 (nitrogen, phosphorus, potassium) fertilizer (250–450 kg/ha), with respect to soil analysis results. Standard production technol-

ogy was applied during plant cultivation. The yield was surveyed on each plot. Thousand seed weight was calculated by counting subsamples of 200 seeds per plot per replicate. Oil content in dry seeds was determined using Newport 4000 NMR and is represented as % of dry matter (d.m.). Nitrogen content was determined by the Dumas combustion method (EN ISO 16634-1:2008) and expressed in % of dry matter. Nitrogen content in % was multiplied with a conversion factor 6.25 to gain the overall protein content. Oil and protein yield in kg/ha were obtained by multiplying seed yield by seed oil and protein content, respectively.

Meteorological data (average daily temperature and precipitation) were collected from the meteorological station "Rimski šančevi" of the Republic hydrometeorological service of Serbia, which is located near the experimental field.

The colour plot diagram for mean genotypic values of the rapeseed samples was calculated and plotted using R software v.4.0.3 (64-bit version). The corrplot instruction was applied, with the "circle" method enabled, as a graphical tool to represent the correlation between the mean genotypic values of the observed samples.

Two different machine learning algorithms were employed to foresee the oil and protein content, seed yield, oil and protein yield, and 1000 seed weight based on the year of production and genotype, including ANN and RFR. These two machine learning methods are broadly utilized and proved to be effective [27,28].

2.2. ANN Modeling

The artificial neural network model is inspired by the structure and function of the neural network of human brain. An ANN consists of three input layers in addition to hidden and output layers. The nodes of such a network are interconnected and pass on information in the same way as neurons do in a brain. Our ANN model was built using data from Table 1. The inputs were the year of production and genotype. A multi-layer perceptron model (MLP) scheme, which consisted of three layers, was used for modelling two artificial neural network models (ANN) for the prediction of oil and protein content, seed yield, oil and protein yield, and 1000 seed weight, based on the year of production and genotype. According to the literature, the ANN models were proven as quite capable of approximating non-linear functions [29–32]. This is important for the study of living organisms where many relations between the examined variables are complex and non-linear. An important advantage of the ANN model is its ability to derive previously unseen relationships. Before the calculation, both input and output data were normalized (according to the min–max normalization scheme) to improve the behaviour of the ANN. During this iterative process, input data were repeatedly presented to the network [33,34]. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm was used as an iterative method for solving unconstrained non-linear optimization during the ANN modelling.

The experimental database for the ANN was randomly divided into training, cross-validation, and testing datasets (with 60%, 20%, and 20% of experimental data, respectively). The training dataset was used for the learning cycle of the ANN and also for the evaluation of the optimal number of neurons in the hidden layer and the weight coefficient of each neuron in the network. A series of different topologies were used, in which the number of hidden neurons varied from 5 to 10, and the training process of the network was run 100,000 times with random initial values of weights and biases. The optimization process was performed based on validation of error minimization. It was assumed that successful training was achieved when the learning and cross-validation curves approached zero.

Coefficients associated with the hidden layer (weights and biases) were grouped in matrices W_1 and B_1, respectively. Similarly, coefficients associated with the output layer were grouped in matrices W_2 and B_2. It is possible to represent the neural network by using matrix notation (Y is the matrix of the output variables, f_1 and f_2 are transfer functions in the hidden and output layers, respectively, and X is the matrix of input variables) [35]:

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

Weight coefficients (elements of matrices W_1 and W_2) were determined during the ANN learning cycle. They were updated using optimization procedures to minimize the error between the network and experimental outputs [33,36,37], according to the sum of squares (SOS) and BFGS algorithms, used to speed up and stabilize convergence [38]. The coefficients of determination were used as parameters to check the performance of the obtained ANN model.

The collected data for ANN modelling were processed statistically using the software package StatSoft Statistica, ver. 10.0, Palo Alto, CA, USA.

2.3. RFR Modeling

The random forest model (RF) is a broadly employed machine learning algorithm that is constructed upon decision trees to predict outputs according to prediction variables [39]. The RF model can be utilized for classification or regression purposes. The random forest regression method is used for the mean prediction of individual trees, in consistence with decision trees developed according to the training dataset [40]. Both ANN and RFR as machine learning models have limitations regarding interpretation, which is very important, particularly in life sciences. Still, they offer valuable insights not only into yield assessment, but also into seed quality parameters, such as oil and protein content. In addition, the RF model can reveal the importance of features. The RFR models were constructed upon the data presented in Table 1. Similarly to the ANN model, the inputs for the RFR models were the year of production and genotype. During random forest regression model calculation for the prediction of seed yield, oil and protein yield, 1000 seed weight, and oil and protein content, based on the year of production and genotype, a large set of decision trees was constructed and each tree was built according to the specific bootstrap sample within a training dataset [41]. In this study, the bootstrap function was employed to randomly split the dataset into homogeneous subsets, namely training and test subsets, which explained 60% and 40% of the entire data, respectively [42]. New sub-samples were selected from the input sample dataset and multiple trees were added to the RFR structure to fit the obtained sub-samples. During the training cycle, the RFR model averaged the results of the created trees in order to minimize the error of prediction [28]. During the RFR calculation, the number of trees parameter was set to 100, 200, 300, 400, 500, and 10,000, while the random test data proportion was set to 40% and the sample proportion was 50%.

The building of the RFR models was performed using StatSoft Statistica, ver. 10.0, Palo Alto, CA, USA.

2.4. The Accuracy of the Model

The numerical verification of the developed models was tested using the coefficient of determination (r^2), reduced chi-square (χ^2), mean bias error (*MBE*), root mean square error (*RMSE*), and mean percentage error (*MPE*). MBE and *RMSE* have the same unit-like variable. These commonly used parameters can be calculated as follows [43]:

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

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

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

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

$$SSE = \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2, \tag{6}$$

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

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

3. Results

3.1. Yield-Related Components

Seed and oil yield, as well as oil content, had the highest values in 2016. That year was favourable for rapeseed growing and over half of the examined genotypes yielded more than 2950 kg/ha. According to four-year mean values, Jelena and NS-L-32 belong to the same group with the highest seed yield as determined by Duncan post hoc test (Table 2). On the other hand, NS-L-45 exhibited the lowest yield. The top two genotypes with the highest oil yield were the same as for seed yield, namely Jelena and NS-L-32. NS-L-45 had the lowest oil yield. Protein yield varied among years, whereby in 2016 and 2018 their average values differed by only 19.44 kg/ha. The average protein yield was 444.93 kg/ha. The highest protein yield was noted for NS-L-32 and NS-L-136 and the lowest for Kata and NS-L-45. The mean genotypic value for 1000 seed weight in the period 2015–2018 was 4.28 g. NS-L-44 had the highest and Express the lowest 1000 seed weight. The mean seed oil content ranged between 41.57% and 46.85%, with a grand mean of 44.41%. The highest yearly average of 46.83% was recorded in 2016 and the lowest, 41.56%, in 2015. Protein content was the highest (22.57%) in 2018 and the lowest (18.24%) in 2016. Valeska svetla had the highest seed protein content (23%), which was 2% more than the grand mean for all other genotypes.

3.2. Correlation Analysis

Statistically significant correlations ($p \leq 0.05$) were found for all analysed traits. During 2015–2018, oil content was in a strong negative correlation with protein content (Figure 1). The size and the circle's colour depend on the correlation coefficients; if the colour is blue, a positive correlation was achieved, and on the contrary the red colour represents the negative correlation. Additionally, the circle's size is increased with the absolute value of the correlation coefficient. The highest positive correlations were found between seed and oil yield ($r = 0.995$), seed and protein yield ($r = 0.943$), and oil and protein yield ($r = 0.921$). Oil content was positively correlated with seed yield, 1000 seed weight, and oil and protein yield, whereas correlation with all traits except 1000 seed weight was strong. Interestingly, only 1000 seed weight was weakly correlated with all traits, negatively with protein content and positively with the other traits.

3.3. ANN Model

The acquired optimal neural network model showed good generalization capabilities for the experimental data, and could accurately predict the oil and protein content, seed yield, oil and protein yield, and 1000 seed weight, based on the year of production and genotype. The number of neurons for the ANN model was eight (network MLP 2-8-6) to obtain the highest values of r^2 (during the training cycle r^2 for output variables were: 0.742; 0.757; 0.853; 0.705; 0.872 and 0.807, for oil and protein content, seed yield, 1000 seed weight, oil and protein yield, respectively); see Table 3.

Table 2. Mean genotypic values for six analysed traits during four-year period.

| Genotype	OC (% d.m.)	PC (% d.m.)	SY (kg/ha)	OY (kg/ha)	PY (kg/ha)	TSW (g)
NS-H-R-1	44.16 [h]	20.99 [n]	2006.00 [ef]	915.12 [fghi]	421.59 [fgh]	4.26 [jk]
NS-H-R-2	44.58 [k]	21.04 [o]	2154.25 [lm]	978.96 [mno]	442.81 [klmn]	4.54 [st]
NS-H-R-3	43.78 [f]	20.73 [k]	2446.07 [u]	1084.90 [v]	495.46 [s]	4.29 [kl]
Banaćanka	44.15 [h]	21.01 [no]	2044.96 [fghi]	918.23 [ghi]	423.93 [fgh]	4.49 [rs]
Slavica	44.30 [i]	20.58 [i]	2198.08 [lmn]	979.17 [mno]	453.29 [no]	4.10 [efg]
Valeska tamna	41.57 [a]	22.98 [ć]	1999.24 [ef]	833.85 [b]	456.51 [o]	4.74 [u]
Valeska svetla	42.67 [c]	23.08 [č]	2064.17 [ghij]	892.14 [efg]	471.10 [pq]	4.58 [t]
Zlatna	45.59 [s]	20.62 [j]	2074.53 [hij]	944.51 [jk]	422.48 [fgh]	4.08 [def]
NS-L-74	45.25 [op]	20.06 [c]	2017.42 [ef]	927.32 [hij]	404.87 [cd]	4.40 [pq]
Branka	45.31 [pq]	19.76 [b]	2027.92 [efgh]	935.99 [ijk]	392.77 [b]	4.13 [fgh]
Express	45.18 [no]	20.14 [d]	2085.11 [ijk]	952.32 [kl]	414.03 [de]	3.91 [a]
NS-L-7	46.55 [t]	20.62 [j]	2287.5 [qr]	1075.03 [uv]	467.49 [p]	4.44 [qr]
Nevena	43.55 [e]	21.55 [t]	2236.33 [mno]	998.13 [opqr]	474.58 [pqr]	4.38 [opq]
Valesca	43.31 [d]	22.05 [x]	2021.06 [efg]	889.76 [ef]	439.72 [jklm]	4.25 [jk]
Ilia	45.35 [q]	20.56 [i]	2213.56 [mn]	1013.71 [qr]	446.58 [lmno]	4.13 [fgh]
Kata	45.57 [s]	20.83 [l]	1838.67 [b]	858.55 [cd]	371.19 [a]	4.14 [fgh]
Nena	45.15 [mn]	20.88 [m]	2109.67 [jkl]	969.56 [lmn]	436.80 [ijkl]	4.30 [klm]
NS-L-31	45.09 [m]	21.23 [pq]	1987.58 [de]	903.78 [efgh]	422.75 [fgh]	4.01 [c]
NS-L-126	44.83 [l]	21.41 [s]	1947.08 [d]	878.91 [de]	409.28 [cde]	4.24 [jk]
NS-L-33	45.35 [q]	19.64 [a]	2294.92 [r]	1052.50 [tu]	448.68 [mno]	3.98 [bc]
NS-L-128	44.16 [h]	22.13 [y]	2012.5 [ef]	901.88 [efgh]	433.57 [hijk]	4.14 [fgh]
Svetlana	43.54 [e]	21.96 [w]	2190.08 [lm]	958.09 [klm]	479.64 [qr]	4.14 [fgh]
Jasna	45.45 [r]	20.35 [g]	2123.51 [kl]	986.34 [nop]	425.25 [fgh]	4.33 [lmno]
NS-L-101	43.72 [f]	21.54 [t]	2017.42 [efg]	893.089 [efg]	426.03 [fghi]	4.37 [nop]
Zorica	44.8 [l]	20.74 [k]	2349.5 [s]	1061.10 [tuv]	484.13 [r]	4.57 [t]
NS-L-102	46.85 [u]	19.78 [b]	2085.92 [ijk]	990.96 [nopq]	404.01 [bcd]	4.16 [ghi]
NS-L-134	45.26 [op]	20.22 [e]	2190.65 [lm]	1000.91 [opqr]	440.87 [jklm]	4.37 [nop]
NS-L-32	44.34 [ij]	21.32 [r]	2555.00 [w]	1139.47 [w]	536.93 [t]	4.31 [klmn]
NS-L-136	44.32 [ij]	21.05 [o]	2511.02 [v]	1128.19 [w]	527.48 [t]	4.02 [cd]
NS-L-137	43.54 [e]	20.45 [h]	2357.33 [s]	1043.67 [st]	473.13 [pqr]	4.18 [hi]
NS-L-138	43.89 [g]	20.04 [c]	2216.08 [mn]	988.25 [nopq]	430.04 [ghij]	3.92 [ab]
NS-L-251	45.48 [r]	20.30 [f]	2242.58 [nop]	1022.38 [rs]	451.46 [mno]	4.37 [lmno]
NS-L-210	43.56 [e]	21.61 [u]	1894.07 [c]	841.37 [bc]	401.61 [bc]	4.40 [pq]
NS-L-44	43.90 [g]	21.64 [u]	2368.58 [st]	1056.99 [tu]	501.59 [s]	5.02 [v]
NS-L-45	43.75 [f]	21.81 [v]	1743.5 [a]	776.75 [a]	371.95 [a]	4.48 [rs]
NS-L-46	42.11 [b]	22.40 [z]	1808.34 [b]	773.54 [a]	399.90 [bc]	4.16 [ghi]
NS-L-47	44.41 [j]	21.01 [no]	2410.75 [tu]	1084.32 [v]	497.78 [s]	4.49 [rs]
Jelena	45.32 [pq]	20.08 [c]	2566.58 [w]	1162.62 [x]	501.73 [s]	4.04 [cde]
Forward	43.52 [e]	21.22 [p]	2010.84 [ef]	893.89 [efg]	419.14 [efg]	4.31 [klmn]
Maidan	43.34 [d]	21.27 [q]	2273.86 [pqr]	1006.60 [pqr]	475.04 [pqr]	4.21 [ij]
Mean	44.41	21.02	2149.36	967.82	444.93	4.28
Minimum	39.24	16.66	533.33	209.60	113.17	3.30
Maximum	50.20	24.67	3766.67	1745.85	693.61	5.77

Values in the same column followed with the same letter are not significantly different at the $p \leq 0.05$ level according to Duncan's post-hoc test. OC—oil content; d.m.—dry matter; PC—protein content; SY—seed yield; OY—oil yield; PY—protein yield; TSW—1000 seed weight.

The potential of the ANN model to predict yield and quality components is presented by scatter plots (Figure 2). The distribution pattern of predicted values differed in comparison with the scatter plots obtained by the RFR model (Figure 3).

Table 3. Artificial neural network model summary (performance and errors), for training, testing, and validation cycles.

Net. Name	Performance			Error			Train. Alg.	Error Func.	Hidden Act.	Output Act.
	Train.	Test.	Valid.	Train.	Test.	Valid.				
MLP 2-8-6	0.788	0.782	0.784	49.749	50.173	52.481	BFGS 95	SOS	Logistic	Logistic

Performance term represents the coefficients of determination, while the error term indicates a lack of data for the ANN model. Net.—network; Train.—training; Test.—testing; Valid.—validation; alg.—algorithm; func.—function; act.—activation
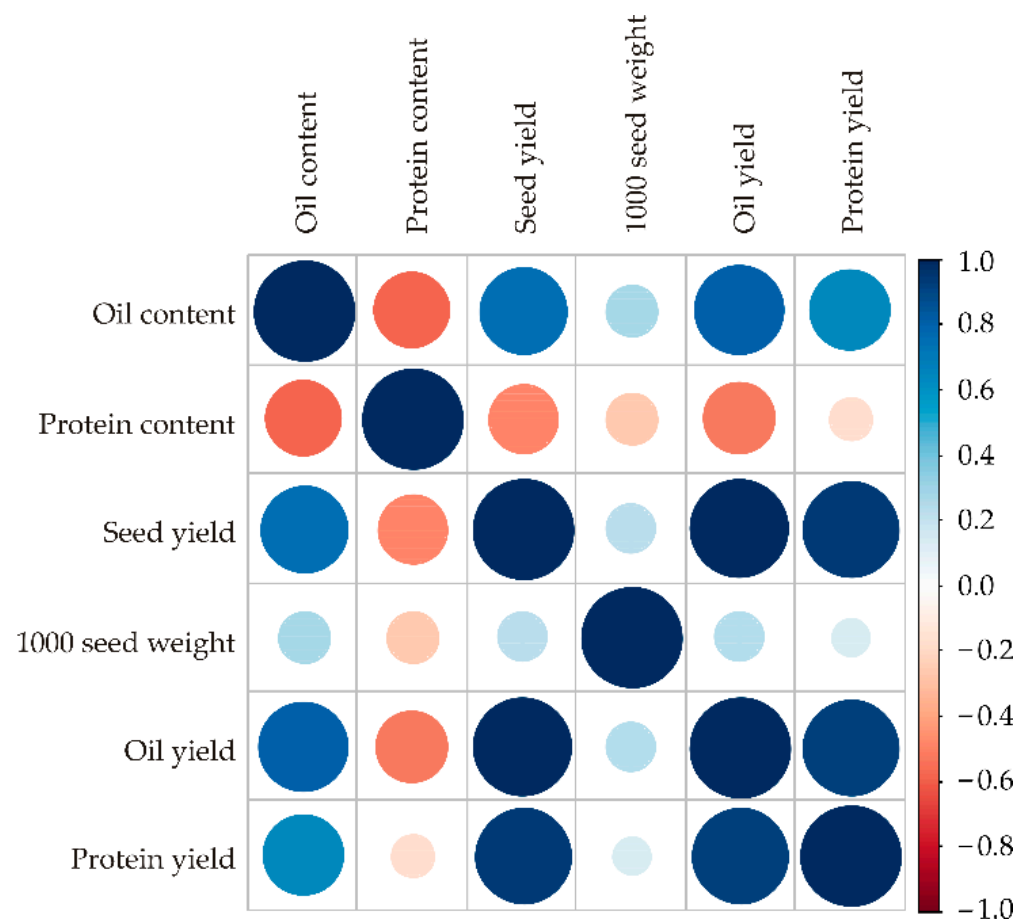
Figure 1. Colour correlation graph between genotypic values for six analysed traits during four-year period. Numerical data represent the coefficients of correlations.

The obtained ANN model for the prediction of output variables was built upon 78 weights-bias coefficients due to the high nonlinearity of the observed system [44,45].

The goodness of fit between experimental measurements and model-calculated outputs, represented as ANN performance, is shown in Table 4. For seed yield, RMSE was 303.25 kg/ha, which accounts for 14.11% of the overall observed yield mean. The RMSE for oil yield and oil content was 139.88 kg/ha and 1.25%, respectively, which accounts for 14.45% and 2.81% of the overall observed oil yield and content, respectively.

The ANN model predicted experimental variables reasonably well for a broad range of the process variables. For the ANN model, the predicted values were very close to the measured values in most cases, in terms of r^2 values.
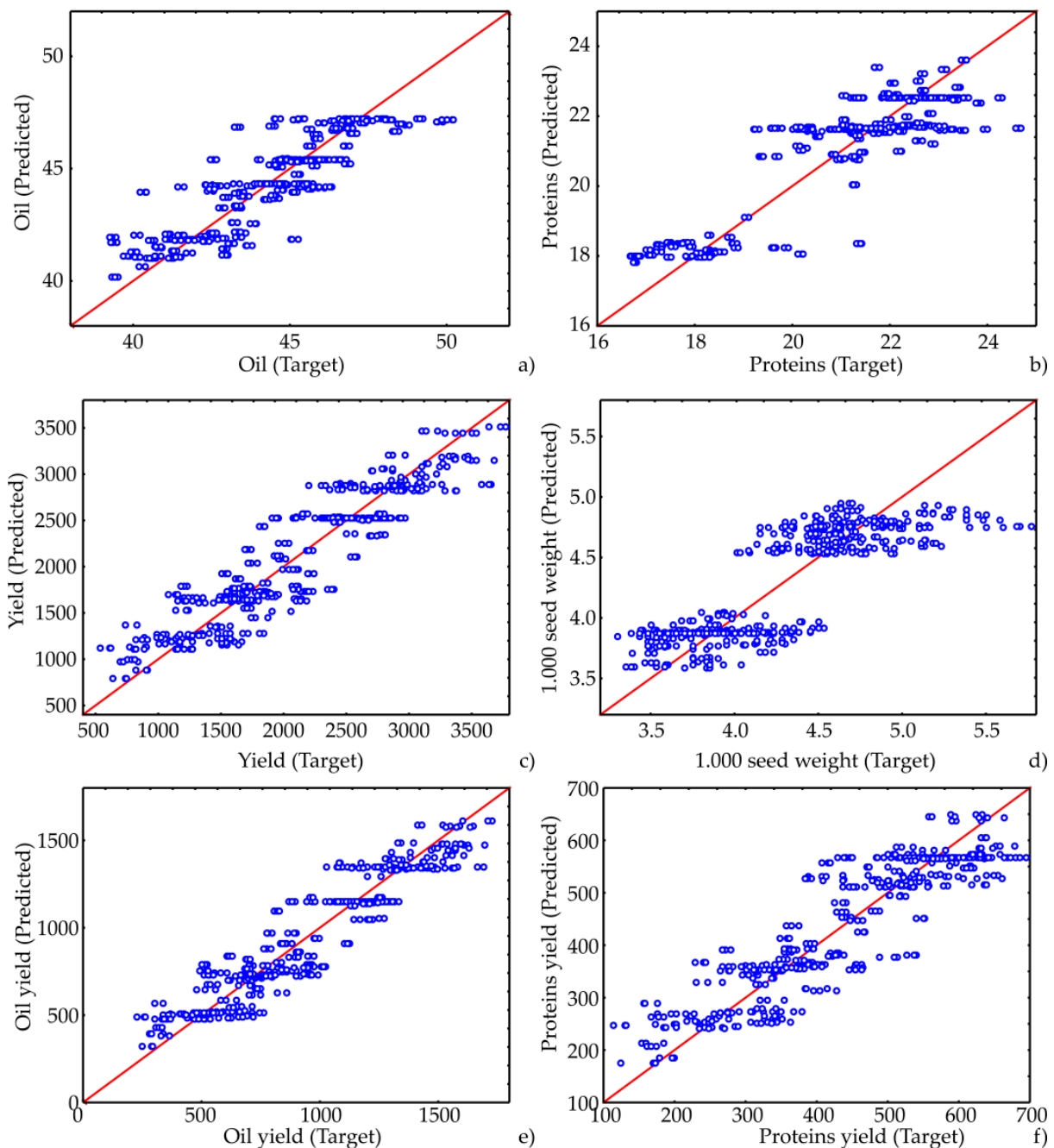
Figure 2. Comparison between experimentally obtained and ANN model predicted values of (**a**) oil and (**b**) protein content, (**c**) yield, (**d**) 1000 seed weight, (**e**) oil yield, and (**f**) protein yield.

3.4. RFR Model

The acquired optimal random forest models showed good prediction capabilities for the experimental data, and could be used to adequately foresee the oil and protein content, seed yield, oil and protein yield, and 1000 seed weight, based on the year of production and genotype. The number of trees for the RFR models were 1000, 590, 590, 1000, 200, and 1000, respectively for oil and protein content, seed yield, 1000 seed weight, oil and protein yield, to obtain the highest values of r^2 (during the training cycle, r^2 values for the output variables were 0.944, 0.935, 0.912, 0.886, 0.936, and 0.900, respectively); see Table 5.
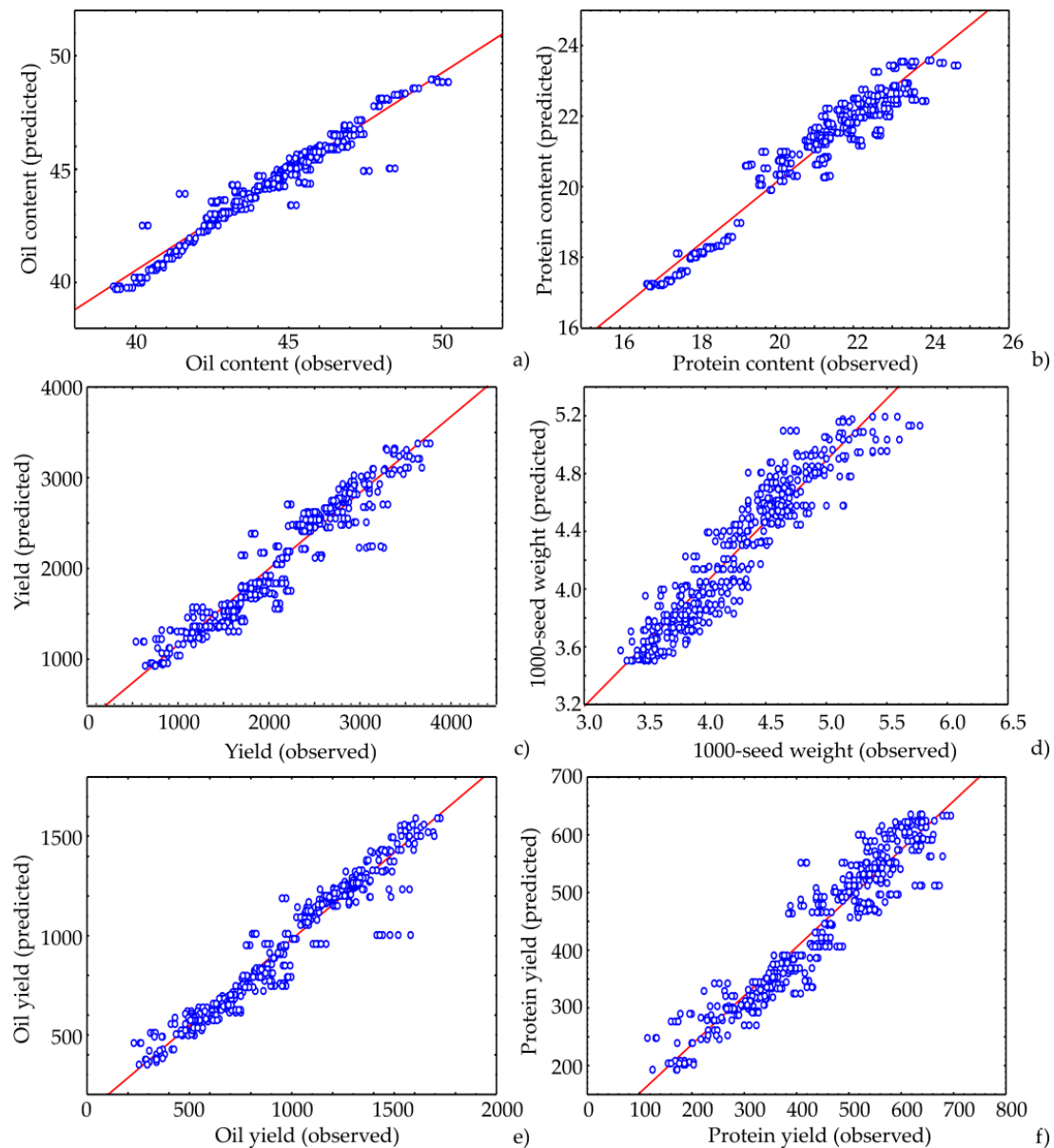
Figure 3. Comparison between experimentally obtained and RFR model predicted values of (**a**) oil and (**b**) protein content, (**c**) yield, (**d**) 1000 seed weight, (**e**) oil yield, and (**f**) protein yield.

Table 4. The goodness of fit tests for the developed ANN model.

Output Variable	χ^2	RMSE	MBE	MPE	SSE	AARD	r^2
Oil content	1.615	1.247	−0.101	2.188	247.112	164.245	0.715
Protein content	1.090	1.024	0.001	3.719	167.920	198.625	0.726
Seed yield	9.6×10^4	303.249	−18.881	14.340	1.5×10^7	5.6×10^4	0.830
1000 seed weight	0.087	0.290	−0.002	5.348	13.454	61.797	0.693
Oil yield	2.0×10^4	139.881	−6.242	15.026	3.1×10^6	2.5×10^4	0.851
Protein yield	4.2×10^3	63.798	−3.269	13.918	6.5×10^5	1.3×10^4	0.785

χ^2—reduced chi-square; RMSE—root mean square error; MBE—mean bias error; MPE—mean percentage error; SSE—sum of squared errors; AARD—absolute average relative deviation; r^2—coefficient of determination.

Table 5. The goodness of fit tests for the developed RFR model.

Output Variable	χ^2	RMSE	MBE	MPE	SSE	AARD	r^2
Oil content	0.344	0.581	0.041	0.749	161.021	185.343	0.944
Protein content	0.256	0.501	0.002	1.768	120.560	307.080	0.935
Seed yield	5.3×10^4	227.082	26.850	9.080	2.4×10^7	1.8×10^5	0.912
1000 seed weight	0.033	0.179	0.003	3.169	15.318	66.079	0.886
Oil yield	9.4×10^3	95.937	13.992	7.889	4.3×10^6	8.1×10^4	0.936
Protein yield	2.0×10^3	44.431	1.704	9.006	9.5×10^5	3.6×10^4	0.900

The RFR model showed much better prediction characteristics for oil and protein content, seed yield, oil and protein yield, and 1000 seed weight based on the year of production and genotype in comparison with the ANN model (Tables 4 and 5).

The RFR and ANN models had an insignificant lack of fit tests, which means the models satisfactorily predicted output variables.

4. Discussion

In this research, 40 rapeseed genotypes were analysed during four years for the assessment of yield and quality components. NS-L-45 was the lowest yielding genotype, probably due to low performance in 2015, which was below 1000 kg/ha. In this study the following pattern was observed: genotypes that had the highest (Jelena, NS-L-32) and the lowest yield (NS-L-45) also had the highest and lowest oil yield. Ref. [10] reported that at the same environments (locations) and years, the highest seed and oil yield were recorded. Since they evaluated only three genotypes, our pattern cannot be extrapolated and discussed for comparison of genotype performance. In 2016, rapeseed had the highest yearly mean value for oil content (46.83%). Throughout May of the same year, during flowering and at the beginning of the seed filling phase, precipitation was higher than the long-term average (64.6 mm for the period 1964–2014). Although precipitation is one of the main factors that positively influence oil content in rapeseed [46,47], it should be kept in mind that it is not the only factor influencing oil content, since in years with high precipitation during seed filling oil content may not be as high as expected [48]. This is in line with our study, as in May 2015 precipitation was three times higher than the long-term average and the oil content was lowest in that year. Considering that oil and protein content are negatively correlated traits [49–51], 2016, the year with the lowest average protein content, was advantageous in terms of oil content. Still, not all analysed genotypes with high oil content had a lower share of proteins, e.g., in 2015 NS-L-7 had 2% higher oil content and 0.71% higher protein than average, and in 2017 NS-H-R-2 had both higher oil and protein content relative to average year values for that year. These genotypes are regarded as good resources for further breeding towards high oil and protein content, because their protein content does not sink abruptly with increasing oil content, as in the case of other genotypes.

Most traits that are used for rapeseed breeding are polygenic and represent the result of the interaction of several components. Knowledge regarding trait correlations is important for success in breeding. Due to the low heritability of yield, indirect selection appears to be the best breeding solution. The strength of the correlation between two analysed traits may differ in different agroecological growing conditions. A strong negative correlation between oil and protein content was previously reported [48,52,53]. An increase in oil content in the seed can arise whether at the expense of decreasing protein content, or by reduction of other seed components [49]. Oil content is under the control of a large number of genes that have additive and non-additive effects, whereas the environment has an impact only on additive components of genetic variance [54]. Refs. [55,56] also reported a high positive correlation between oil content and seed yield. However, unlike our results, they did not find a significant correlation between oil content and oil yield. In relation to a high positive correlation between seed yield on the one hand, and oil and

protein yield on the other, it can be concluded that with higher seed yield oil and protein yield tends to increase, as can be realized from Figure 1. The analysis of 20 rapeseed traits with path analysis revealed 1000 seed weight to be the most important trait that influences yield [57]. A positive relationship between 1000 seed weight and seed yield was also reported by [58,59]. Results from [60] are contrasting in terms of claiming a negative correlation between these traits. Observed differences probably occurred because of the stronger influence of environmental (weather) variables on yield and 1000 seed weight over the analysed years. It can be hypothesized that in years with adverse weather conditions, rapeseed plants will decrease the number of seeds per silique, but seed size may increase, thus resulting in higher seed weight.

Similar to our dataset, which consisted of temperature, precipitation, and cultivar data, on the list of variables [19] that were most often used for prediction of crop yield using different machine learning models, temperature was positioned first, rainfall third, and crop information (e.g., cultivar, crop density) fourth. The process of seed filling is generally susceptible to environmental conditions, especially temperature and precipitation. Thus, we assume that in that period information regarding weather conditions is more important than crop information for yield prediction.

The use of classic statistical procedures for the analysis of both dependent and independent variables is not as efficient as the use of machine learning models. Machine learning models make it possible to predict a larger number of variables. Non-linear machine learning models for the evaluation of yield-related traits enable deciphering non-linear relations among dependent and independent variables [61]. Prediction models can be efficiently used for rapeseed and other crop yield prediction, offering the possibility for early yield assessment, thus enhancing farmers in the decision-making process toward optimum production. ANN and RFR, among other machine learning models, cope well with the analysis of complex data. The developed mathematical models provided an efficient insight in the prediction of oil and protein content, seed yield, 1000 seed weight, oil and protein yield, and the influences of production parameters, such as year of production and crop genotype, on the abovementioned traits. With the aid of these models, it is easier to predict the effects of different weather circumstances, or of the selected cultivar on yield and quality, as well as to choose which cultivar will have the best performance in a certain environment. Knowledge regarding cultivars' reaction to specific environmental conditions is valuable for the estimation of their final performance. Additionally, information on weather, such as temperatures and precipitations, is accessible to farmers, and on the other hand, cultivar/hybrid recommendation and production technology can be obtained from agricultural extension service. All of this should be incorporated into applicable models and used for yield forecasting.

ANN was successfully used for the prediction of oil content in other species, such as sesame (*Sesamum indicum* L.) and ajowan (*Carum copticum* L.) [62,63]. The best fit of predicted to measured traits in our ANN model was observed for oil yield (r^2 = 0.851). Negative MBE values occurred for all traits except protein content. This indicates that the predicted values were smaller than the observed ones. SOS values obtained with the ANN model were of the same order of magnitude as experimental errors for output variables reported in the literature [33,37].

A high r^2 is indicative that the variation was accounted for and that the data fitted the proposed RFR model satisfactorily [64,65]. The RMSE for seed yield was 227.08 kg/ha, which represents 10.56% of the overall observed yield mean. When comparing this result with the RMSE for the ANN model, it can be noticed that the RFR model offered a more acceptable RMSE. This finding also goes in favour of using RFR for rapeseed yield prediction.

5. Conclusions

The best performances of the analysed rapeseed genotypes (e.g., highest seed and oil yield) were achieved in 2016. The highest positive correlation was found between

seed and oil yield. In order to forecast yield and quality components, machine learning models were developed based on available genotype and weather data. The current study suggests that RFR and ANN modelling can be successfully exploited for the purpose of rapeseed oil and protein content, seed yield, oil and protein yield, and 1000 seed weight prediction, based on the year of production and genotype. The artificial neural network model showed itself to be adequate for the prediction of output variables. The highest r^2 values were obtained with the RFR model. The mentioned r^2 values justified the use of the developed models in the prediction of the observed parameters. According to the results, the RFR models were more accurate and their results were closer to the experimental data, in comparison with the ANN models. It is assumed that during the phase of seed filling, input data regarding environmental conditions are more valuable for yield prediction. The incorporation of more input data can improve the efficiency of both tested models. The tested models proved their usefulness for yield prediction and suggested the possibility to use them for the prediction of oil and protein content. This study has the potential to direct new ways for promising research related to rapeseed quality prediction, such as fatty acids and glucosinolates contents. In the end, it will encourage and promote research related to the use of machine learning algorithms for yield forecasts.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. FAOSTAT. 2020 Food and Agriculture Organization of the United Nations (FAO). *FAOSTAT Statistical Database.* Available online: http://www.fao.org/faostat/en/#data/QC/visualize (accessed on 13 October 2021).
2. Marjanović Jeromela, A.; Terzić, S.; Jankulovska, M.; Zorić, M.; Kondić-Špika, A.; Jocković, M.; Hristov, N.; Crnobarac, J.; Nagl, N. Dissection of year related climatic variables and their effect on winter rapeseed (*Brassica napus* L.) Development and Yield. *Agronomy* **2019**, *9*, 517. [CrossRef]
3. Spasibionek, S.; Mikołajczyk, K.; Ćwiek–Kupczyńska, H.; Piętka, T.; Krótka, K.; Matuszczak, M.; Nowakowska, J.; Michalski, K.; Bartkowiak-Broda, I. Marker assisted selection of new high oleic and low linolenic winter oilseed rape (*Brassica napus* L.) inbred lines revealing good agricultural value. *PLoS ONE* **2020**, *15*, e0233959. [CrossRef]
4. Olesen, J.E.; Trnka, M.; Kersebaum, K.C.; Skjelvåg, A.O.; Seguin, B.; Peltonen-Sainio, P.; Rossi, F.; Kozyra, J.; Micale, F. Impacts and adaptation of European crop production systems to climate change. *Eur. J. Agron.* **2011**, *34*, 96–112. [CrossRef]
5. Webber, H.; Ewert, F.; Olesen, J.E.; Müller, C.; Fronzek, S.; Ruane, A.C.; Bourgault, M.; Martre, P.; Ababaei, B.; Bindi, M.; et al. Diverging importance of drought stress for maize and winter wheat in Europe. *Nat. Commun.* **2018**, *9*, 4249. [CrossRef]
6. Sharif, B.; Makowski, D.; Plauborg, F.; Olesen, J.E. Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. *Eur. J. Agron.* **2017**, *82*, 11–20. [CrossRef]
7. Aksouh-Harradj, N.M.; Campbell, L.C.; Mailer, R.J. Canola response to high and moderately high temperature stresses during seed maturation. *Can. J. Plant Sci.* **2006**, *86*, 967–980. [CrossRef]
8. Dreccer, M.F.; Fainges, J.; Whish, J.; Ogbonnaya, F.C.; Sadras, V.O. Comparison of sensitive stages of wheat, barley, canola, chickpea and field pea to temperature and water stress across Australia. *Agr. Forest. Meteor.* **2018**, *248*, 275–294. [CrossRef]
9. Brown, J.K.M.; Beeby, R.; Penfield, S. Yield instability of winter oilseed rape modulated by early winter temperature. *Sci. Rep. UK* **2019**, *9*, 6953. [CrossRef]

10. Weymann, W.; Böttcher, U.; Sieling, K.; Kage, H. Effects of weather conditions during different growth phases on yield formation of winter oilseed rape. *Field Crop. Res.* **2015**, *173*, 41–48. [CrossRef]

11. Liu, S.; Li, L.; Gao, W.; Zhang, Y.; Liu, Y.; Wang, S.; Lu, J. Diagnosis of nitrogen status in winter oilseed rape (*Brassica napus* L.) using in-situ hyperspectral data and unmanned aerial vehicle (UAV) multispectral images. *Comput. Electron. Agr.* **2018**, *151*, 185–195. [CrossRef]

12. Kong, W.; Zhang, C.; Huang, W.; Liu, F.; He, Y. Application of Hyperspectral Imaging to Detect *Sclerotinia sclerotiorum* on Oilseed Rape Stems. *Sensors* **2018**, *18*, 123. [CrossRef]

13. Singh, K.D.; Duddu, H.S.N.; Vail, S.; Parkin, I.; Shirtliffe, S.J. UAV-Based Hyperspectral Imaging Technique to Estimate Canola (*Brassica napus* L.) Seedpods Maturity. *Can. J. Remote Sens.* **2021**, *47*, 33–47. [CrossRef]

14. Gong, Y.; Duan, B.; Fang, S.; Zhu, R.; Wu, X.; Ma, Y.; Peng, Y. Remote estimation of rapeseed yield with unmanned aerial vehicle (UAV) imaging and spectral mixture analysis. *Plant Methods* **2018**, *14*, 70. [CrossRef]

15. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Peng, B. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agr. For. Meteorol.* **2019**, *274*, 144–159. [CrossRef]

16. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random forests for global and regional crop yield predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef]

17. Fan, H.; Liu, S.; Li, J.; Li, L.; Dang, L.; Ren, T.; Lu, J. Early prediction of the seed yield in winter oilseed rape based on the near-infrared reflectance of vegetation (NIRv). *Comput. Electron. Agr.* **2021**, *186*, 106166. [CrossRef]

18. Niedbała, G.; Piekutowska, M.; Weres, J.; Korzeniewicz, R.; Witaszek, K.; Adamski, M.; Pilarski, K.; Czechowska-Kosacka, A.; Krysztofiak-Kaniewska, A. Application of artificial neural networks for yield modeling of winter rapeseed based on combined quantitative and qualitative data. *Agronomy* **2019**, *9*, 781. [CrossRef]

19. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agr.* **2020**, *177*, 105709. [CrossRef]

20. Wojciechowski, T.; Niedbala, G.; Czechlowski, M.; Nawrocka, J.R.; Piechnik, L.; Niemann, J. Rapeseed seeds quality classification with usage of VIS-NIR fiber optic probe and artificial neural networks. In Proceedings of the International Conference on Optoelectronics and Image Processing (ICOIP), IEEE, Warsaw, Poland, 10–12 June 2016; pp. 44–48. [CrossRef]

21. Kaul, M.; Hill, R.L.; Walthall, C. Artificial neural networks for corn and soybean yield prediction. *Agr. Syst.* **2005**, *85*, 1–18. [CrossRef]

22. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef]

23. Liu, J.; Goering, C.; Tian, L. A neural network for setting target corn yields. *Trans. ASAE* **2001**, *44*, 705. [CrossRef]

24. Luo, X.; Ding, Y.; Zhang, L.; Yue, Y.; Snyder, J.H.; Ma, C.; Zhu, J. Genomic prediction of genotypic effects with epistasis and environment interactions for yield-related traits of rapeseed (*Brassica napus* L.). *Front. Genet.* **2017**, *8*, 15. [CrossRef]

25. Li, J.; Liao, G.; Ou, Z.; Jin, J. Rapeseed seeds classification by machine vision. In Proceedings of the Workshop on Intelligent Information Technology Application (IITA 2007), IEEE, Zhangjiajie, China, 2–3 December 2007; pp. 222–226. [CrossRef]

26. Qadri, S.; Qadri, S.F.; Razzaq, A.; Rehman, M.U.; Ahmad, N.; Nawaz, S.A.; Saher, N.; Akhtar, N.; Khan, D.M. Classification of canola seed varieties based on multi-feature analysis using computer vision approach. *Int. J. Food Prop.* **2021**, *24*, 493–504. [CrossRef]

27. Wang, J.; Sun, X.; Cheng, Q.; Cuia, Q. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Sci. Total Environ.* **2021**, *762*, 143099. [CrossRef]

28. Yang, S.; Yang, D.; Chen, J.; Santisirisomboon, J.; Lu, W.; Zhao, B. A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. *J. Hydrol.* **2020**, *590*, 125206. [CrossRef]

29. Johnson, D.P.; Stanforth, A.; Lulla, V.; Luber, G. Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Appl. Geogr.* **2012**, *35*, 23–31. [CrossRef]

30. Yun, T.S.; Jeong, Y.J.; Han, T.S.; Youm, K.S. Evaluation of thermal conductivity for thermally insulated concretes. *Energ. Buildings* **2013**, *61*, 125–132. [CrossRef]

31. Kleijnen, J.P.C. Design and Analysis of Simulation Experiments. In *Statistics and Simulation: IWS 8, Vienna, Austria, September 2015*, 1st ed.; Springer Proceedings in Mathematics&Statistics, Pilz, J., Rasch, D., Melas, V.B., Moder, K., Eds.; Springer: Cham, Switzerland, 2018; Volume 231, pp. 3–22. [CrossRef]

32. Pavlić, B.; Pezo, L.; Marić, B.; Peić Tukuljac, L.; Zeković, Z.; Bodroža Solarov, M.; Teslić, N. Supercritical Fluid Extraction of Raspberry Seed Oil: Experiments and Modelling. *J. Supercrit. Fluid.* **2020**, in press. [CrossRef]

33. Kollo, T.; von Rosen, D. Advanced Multivariate Statistics with Matrices. In *Mathematics and Its Applications*, 1st ed.; Hazewinkel, M., Ed.; Springer: Dordrecht, The Netherlands, 2005; Volume 579, pp. 1–485. [CrossRef]

34. Pezo, 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 dehydration. *Hem. Ind.* **2013**, *67*, 465–475. [CrossRef]

35. Ochoa-Martínez, C.I.; Ayala-Aponte, A.A. Prediction of mass transfer kinetics during osmotic dehydration of apples using neural networks. *LWT–Food Sci. Technol.* **2007**, *40*, 638–645. [CrossRef]

36. Berrueta, L.A.; Alonso-Salces, R.M.; Héberger, K. Supervised pattern recognition in food analysis. *J. Chromatogr. A.* **2007**, *1158*, 196–214. [CrossRef]

37. Doumpos, M.; Zopounidis, C. Preference disaggregation and statistical learning for multicriteria decision support: A review. *Eur. J. Oper. Res.* **2011**, *209*, 203–214. [CrossRef]

38. Taylor, B.J. *Methods and Procedures for the Verification and Validation of Artificial Neural Networks*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–278. [CrossRef]

39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

40. Rasaei, Z.; Bogaert, P. Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. *Geoderma* **2019**, *344*, 50–62. [CrossRef]

41. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agr.* **2018**, *153*, 213–225. [CrossRef]

42. Zhang, L.; Yang, L.; Ma, T.; Shen, F.; Cai, Y.; Zhou, C. A self-training semi supervised machine learning method for predictive mapping of soil classes with limited sample data. *Geoderma* **2021**, *384*, 114809. [CrossRef]

43. 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. Crop Prod.* **2020**, *154*, 112752. [CrossRef]

44. Montgomery, D.C. *Design and Analysis of Experiments*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 1984; pp. 1–556.

45. Chattopadhyay, P.B.; Rangarajan, R. Application of ANN in sketching spatial nonlinearity of unconfined aquifer in agricultural basin. *Agr. Wat. Manag.* **2014**, *133*, 81–91. [CrossRef]

46. Pritchard, F.M.; Eagles, H.A.; Norton, R.M.; Salisbury, P.A.; Nicolas, M. Environmental effects on seed composition of Victorian canola. *Aus. J. Exp. Agr.* **2000**, *40*, 679–685. [CrossRef]

47. Si, P.; Walton, G.H. Determinants of oil concentration and seed yield in canola and Indian mustard in the lower rainfall areas of Western Australia. *Aust. J. Agr. Res.* **2004**, *55*, 367–377. [CrossRef]

48. Tetteh, E.T.; de Koff, J.P.; Pokharel, B.; Link, R.; Robbins, C. Effect of winter canola cultivar on seed yield, oil, and protein content. *Agron. J.* **2019**, *111*, 2811–2820. [CrossRef]

49. Si, P.; Mailer, R.J.; Galwey, N.; Turner, D.W. Influence of genotype and environment on oil and protein concentrations of canola (*Brassica napus* L.) grown across southern Australia. *Aust. J. Agric. Res.* **2003**, *54*, 397–407. [CrossRef]

50. Hu, Z.Y.; Hua, W.; Zhang, L.; Deng, L.B.; Wang, X.F.; Liu, G.H.; Hao, W.J.; Wang, H.Z. Seed structure characteristics to form ultrahigh oil content in rapeseed. *PLoS ONE* **2013**, *8*, e62099. [CrossRef]

51. Gu, J.; Chao, H.; Wang, H.; Li, Y.; Li, D.; Xiang, J.; Gan, J.; Lu, G.; Zhang, X.; Long, Y.; et al. Identification of the relationship between oil body morphology and oil content by microstructure comparison combining with QTL analysis in brassica napus. *Front. Plant Sci.* **2017**, *7*, 1989. [CrossRef] [PubMed]

52. Vujaković, M.; Marjanović Jeromela, A.; Jovičić, D.; Marinković, R. Dependence of rapeseed quality and yield on density, variety and year of production. *Ratarstvo Povrtarstvo* **2015**, *52*, 61–66. [CrossRef]

53. Hammac, W.A.; Maaz, T.M.; Koenig, R.T.; Burke, I.C.; Pan, W.L. Water and temperature stresses impact canola (*Brassica napus* L.) fatty acid, protein, and yield over nitrogen and sulfur. *J. Agric. Food Chem.* **2017**, *65*, 10429–10438. [CrossRef] [PubMed]

54. Guo, Y.; Si, P.; Wang, N.; Wen, J.; Yi, B.; Ma, C.; Tu, J.; Zou, J.; Fu, T.; Shen, J. Genetic effects and genotype × environment interactions govern seed oil content in *Brassica napus* L. *BMC Genet.* **2017**, *18*, 1. [CrossRef]

55. Fikere, M.; Barbulescu, D.M.; Malmberg, M.M.; Maharjan, P.; Salisbury, P.A.; Kant, S.; Panozzo, J.; Norton, S.; Spangenberg, G.C.; Cogan, N.O.I.; et al. Genomic prediction and genetic correlation of agronomic, blackleg disease, and seed quality traits in Canola (*Brassica napus* L.). *Plants* **2020**, *9*, 719. [CrossRef]

56. Stepien, A.; Wojtkowiak, K.; Pietrzak-Fiecko, R. Nutrient content, fat yield and fatty acid profile of winter rapeseed (*Brassica napus* L.) grown under different agricultural production systems. *Chil. J. Agric. Res.* **2017**, *77*, 266–272. [CrossRef]

57. Sabaghnia, N.; Dehghani, H.; Alizadeh, B.; Mohghaddam, M. Interrelationships between seed yield and 20 related traits of 49 canola (*Brassica napus* L.) genotypes in non-stressed and water-stressed environments. *Span. J. Agric. Res.* **2010**, *8*, 356–370. [CrossRef]

58. Ivanovska, S.; Stojkovski, C.; Dimov, Z.; Marjanović Jeromela, A.; Jankulovska, M.; Jankuloski, L. Interrelationship between yield and yield related traits of spring canola (*Brassica napus* L.) genotypes. *Genetika* **2007**, *39*, 325–332. Available online: https://www.dgsgenetika.org.rs/abstrakti/vol39_no3_rad5.pdf (accessed on 1 December 2021). [CrossRef]

59. Verdejo, J.; Calderini, D.F. Plasticity of seed weight in winter and spring rapeseed is higher in a narrow but different window after flowering. *Field Crop. Res.* **2020**, *250*, 107777. [CrossRef]

60. Lu, G.; Zhang, F.; Zheng, P.; Cheng, Y.; Liu, F.I.; Fu, G.; Zhang, X. Relationship among yield components and selection criteria for yield improvement in early rapeseed (*Brassica napus* L.). *Agr. Sci. China* **2011**, *10*, 997–1003. [CrossRef]

61. Niazian, M.; Niedbała, G. Machine learning for plant breeding and biotechnology. *Agriculture* **2020**, *10*, 436. [CrossRef]

62. Abdipour, M.; Ramazani, S.H.R.; Younessi-Hmazekhanlu, M.; Niazian, M. Modeling oil content of sesame (*Sesamum indicum* L.) using artificial neural network and multiple linear regression approaches. *J. Am. Oil Chem. Soc.* **2018**, *95*, 283–297. [CrossRef]

63. Niazian, M.; Sadat-Noori, S.A.; Abdipour, M. Artificial neural network and multiple regression analysis models to predict essential oil content of ajowan (*Carum copticum* L.). *J. Appl. Res. Med. Aromat. Plants* **2018**, *9*, 124–131. [CrossRef]

64. Erbay, Z.; Icier, F. Optimization of hot air drying of olive leaves using response surface methodology. *J. Food Eng.* **2009**, *91*, 533–541. [CrossRef]

65. Turanyi, T.; Tomlin, A.S. *Analysis of Kinetics Reaction Mechanisms*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–363. [CrossRef]