CYBERNETICS
and COMPUTER
TECHNOLOGIES

UDC 519.67 DOI:10.34229/2707-451X.22.1.3

O. ZOZULYA, V. DOMRACHEV, V. TRETYNYK

APPLICATION OF ARTIFICIAL NEURAL NETWORK TECHNOLOGY FOR PREDICTION OF SUNFLOWER HARVEST LOSSES

This paper describes the usefulness of using neural networks for predicting the probable losses of the farmer's crops of sunflower on the basis of the analysis of the vegetation index distribution on the field.

***Keywords**: sunflower, machine learning, artificial neural networks, forecast model.*

Introduction. The current stage of economic development is characterized by digitalization. Digital technologies in crop production occupy leading positions in agrocybernetics [1].

The digitalization of society has brought to the fore new methods of studying development processes, among which a significant role is played by deep learning and its most successful methods such as artificial neural networks [2–8].

Artificial neural networks (ANNs) have gained popularity an effective tool for offering solutions to a wide variety of different case studies of biological and agricultural background. Their effectiveness emanates from their ability to model complex relationships between observation data from sensors and predicted variables without relying on assumptions about the model structure hence they can predict the real nature of the nonlinear relation between input and output data [3]. Yield prediction is a major challenge in precision agriculture, closely associated to the adoption of best management practices, crop pricing and security. Various techniques and methodologies have been developed to predict crop yield in agriculture.

Yield forecasting requires control of many parameters, including Moisture Content pH, Soil Organic Matter, Total Nitrogen and Organic Carbon, which complicates the forecasting process [3].

The purpose of this paper is to find out and substantiate the possibility of predicting the probable loss of the sunflower crop by the farmer based on the analysis of the distribution of the vegetation index in the field.

Our hypothesis is that the distribution of the vegetation index significantly affects the percentage of losses, of course, with additional parameters.

Let us consider the use of ANN (artificial neural networks) to predict yield loss based on the analysis of changes of vegetation index of sunflower. Very often there is a need to accelerate the ripening of sunflowers by desiccation (desiccation - pre-drying of plants in order to accelerate ripening and facilitate harvesting). Especially when there is uneven plant development in the field.

This is often due to weather conditions during sowing, when there is a lack of moisture in the area of seed germination. In areas with drier soils, the plants often sprout later after the rains.

Or there is a difference in soil temperature, which in early spring depends on the terrain. Somewhere it warms up faster, somewhere slower. As a result, the seedlings appear in places with better heating earlier, and in others later. To balance the ripening process and reduce losses during harvest, desiccation of crops is carried out. And here the farmers face a dilemma: to do desiccation and spend extra money or not to do and save, because the possible losses are not worth the extra investment in the technological operation. In order to estimate the losses in different regions of Ukraine, data on losses from under the combine were collected in areas with different levels of vegetation index [1]. As data for the analysis measurements of indicators of set of fields with sunflower are chosen (fig. 1).

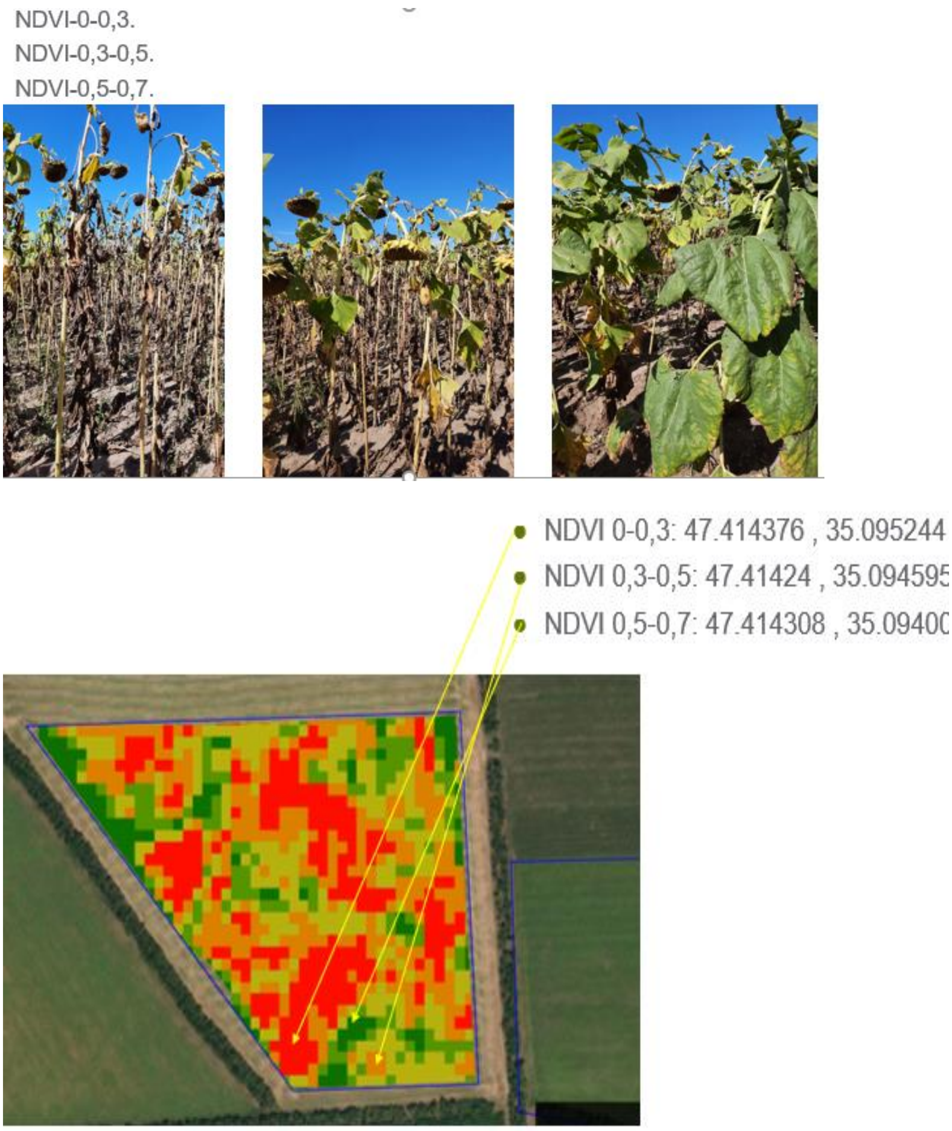

FIG. 1. The level of sunflower ripening in one field in areas with different levels of vegetation index (NDVI), Zaporozhye region, 2021

To determine the losses in areas with different vegetation indices, special frames were laid out, from which the lost and shed sunflower seeds were removed after the combine was passed, weighed, and the losses per hectare were counted (fig. 2).

FIG. 2. The process of determining losses during harvesting, Zaporizhzhia region, 2021

Such investigation has been conducted in more than ten fields in different regions, on different backgrounds of vegetation indices and potential sunflower yields. There were also differences in grain moisture at the beginning of harvest. Processing by the method of normal correlation in this case did not give a clear answer regarding the construction of the formula for the dependence of parameters. What is the formula for the dependence of losses on the difference in the value of the vegetation index? If we take each individual experiment, there was an almost linear dependence of losses on the value of the delta indices in different areas. But it is difficult to calculate the exact formula of this dependence for the whole data set from all plots, because there are many different input data: different values of indices, different background of total yield, seed moisture, combine brands, different hybrids and more. Data for analysis are given in the table below (table 1).

Results

The influence of parameters that characterize the harvest on its losses is, but a clear regression relationship can not be built. Therefore, the technology of artificial neural networks is used to build the model. The model is formed in the form of an algorithm at the input of which input parameters are given (value of vegetation index at the beginning of the study, change of index value during the study period, seed moisture in the accounting area, percentage of study area from field area), at the output we get the percentage of possible crop losses (fig. 3). The algorithm is automatically translated into a program in the C ++ programming language (or another programming language), which allows in practice to model the farmer's possible crop losses depending on his actions in relation to growing crops.

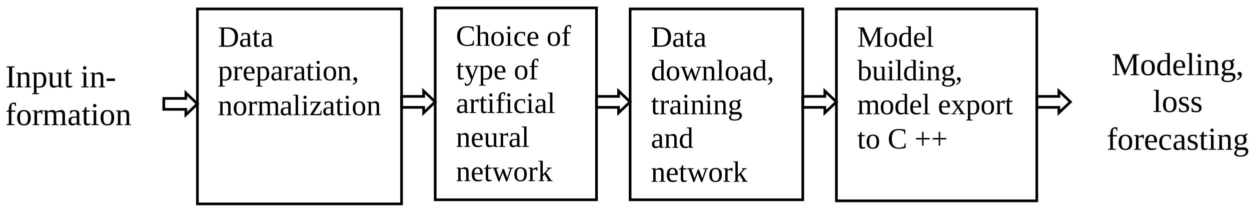

FIG. 3. General scheme of the method of forecasting crop losses

The input data are presented in the form of a table: the value of the vegetation index at the beginning of the investigation, the change in the index for the investigation period, loss (kg) per 1 ha, seed moisture in the accounting area, percentage of yield losses field area.

TABLE 1

Vegetation index at the beginning of the investigation	Change the value of the index during the investigation period	Crop losses, kg/ha	Seed moisture in the accounting area, %	Crop losses % ha	Percentage of study area from field area, %
1	2	3	4	5	6
0.16	0.09	40.80	12.1	1.48	13.7
0.25	0.03	60.4	12.7	2.20	79.7
0.28	0.06	56.6	12.2	2.06	6
0.13	0.06	132	9.5	3.27	10
0.13	0.06	91.7	10.8	2.27	28.67
0.19	0.03	76.5	10.5	1.89	54.94
0.22	0.12	297	13.1	7.35	16.31
0.13	0.06	227.3	9.3	6.89	11.1
0.13	0.06	125.4	7.5	3.80	9.2
0.19	0.03	99	8.8	3.00	51.81
0.22	0.12	171.6	10.7	5.20	24.25
0.19	0.06	57.2	8.2	1.91	52.82
0.25	0.03	77	8.7	2.57	29.58
0.28	0.15	96.8	9.9	3.23	11.01
0.07	0.12	30.3	9.1	0.81	35.63
0.19	0.06	132	9.4	3.52	38.02
0.25	0.09	189.8	9.7	5.06	23.88
0.22	0.03	63.3	5.5	3.00	3.7
0.25	0.09	2.8	6.5	0.13	84.1
0.34	0.09	5.5	8	0.26	12.2
0.22	0.09	93.5	7.1	3.60	24.88
0.31	0.06	71.5	8	2.75	50
0.37	0.12	33	9.4	1.27	25.12
0.31	0.09	44	10.1	1.42	13.22
0.25	0.06	44	8	1.42	37.48
0.19	0.06	165	7.3	5.32	49.3
0.2	0.1	24.2	6.9	0.83	43.94
0.3	0.1	11	6.4	0.38	33.67
0.4	0.1	37.4	4.2	1.28	21.42
0.3	0.1	116.6	10.1	3.62	20.29
0.2	0.1	105.6	9.7	3.28	51.25
0.1	0.1	173.8	8.1	5.40	26.55
0.3	0.1	123.2	10.9	3.83	7.39

Continuation of Table 1

1	2	3	4	5	6
0.2	0.1	125.4	9.4	3.89	70.1
0.1	0.1	202.4	8.1	6.29	19.33
0.19	0.09	151.8	10.1	4.40	15.72
0.16	0.03	136.4	8.4	3.95	39.07
0.1	0.06	112.2	7.7	3.25	43.45
0.19	0.09	121	9.7	4.13	8.95
0.16	0.03	112.8	8.3	3.85	63.3
0.1	0.06	99	8.2	3.38	27.4
0.43	0.09	330	18	10.31	8.9
0.37	0.06	151.3	15	4.73	53.03
0.25	0.12	82.5	10	2.58	38
0.5	0.1	185.9	16	6.00	7
0.4	0.1	110	15	3.55	57
0.2	0.2	55	10.5	1.77	35
0.5	0.2	103.4	14.9	3.34	8.03
0.3	0.2	50.6	12.2	1.63	84.34
0.01	0.29	35.2	10.3	1.14	6.16
0.7	0.2	41.8	13.7	1.59	42.96
0.5	0.2	46.2	10.3	1.76	55.67
0.3	0.2	41.8	9.1	1.59	1.37
0.28	0.03	45.3	4.2	1.44	7
0.37	0.03	82.7	10.8	2.64	85.9
0.4	0.03	101	10.9	3.22	7.1
0.31	0.03	63.4	4.9	1.62	14.2
0.37	0.03	66.9	7.7	1.66	78
0.43	0.03	87.1	13.2	2.11	7.2

To construct a model based on an artificial neural network, we use classification algorithms that provide better statistics [2–8]. To do this, in the input table, divide the percentage of yield loss by the corresponding number of so-called deciles (in descriptive statistics, the decile is any of the nine values that divide the sorted data into ten equal parts, so that each part is 1/10 samples or aggregates) to which the neural network will include the results of losses (fig.4). Thus, we move from determining the absolute value of the percentage of crop losses to determining the appropriate class (decile), which simplifies the procedure for establishing weights in an artificial neural network and provides an appropriate (reasonable) statistical justification for the determination algorithm.

Before data can be entered into the network, it must be prepared in some way. It is also important that the source data can be interpreted correctly.

The variety of neural network models and the many parameters that need to be set (network size, learning algorithm parameters, etc.) can be problematic. Therefore, it is better to consider an ensemble of models and choose the best one.

There are automated neural network systems that can automatically search for the appropriate network architecture of any complexity and compare network performance.

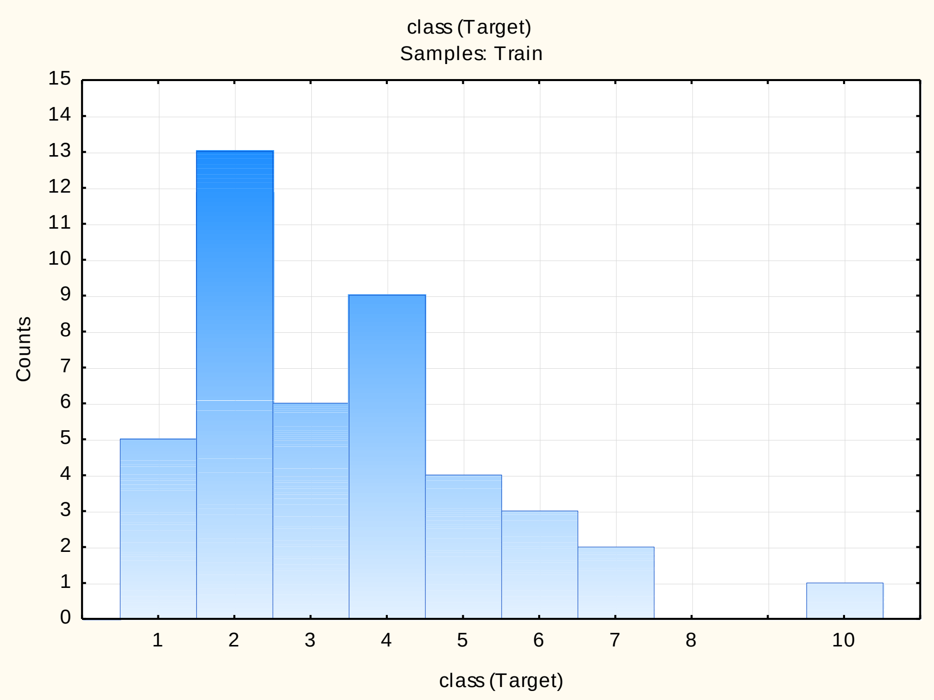

FIG. 4. Histogram of the distribution of the percentage of sunflower yield losses by deciles

Investigations have shown that the best statistical approximation is the model of ANN, which provides predictive values (table 2).

TABLE 2

case	class target	class - model Output
1	**2**	**3**
1	2	5
3	2	2
4	4	2
5	3	2
8	7	2
9	4	7
10	3	3
11	5	4
12	2	4
13	3	3
14	3	2
17	5	4
19	1	1
20	1	2
22	3	2
23	1	1
24	2	4

Continuation of Table 2

1	2	3
25	2	2
26	6	6
29	1	1
30	4	4
32	6	7
33	4	4
34	4	4
35	7	7
36	5	4
37	4	2
39	4	5
42	10	10
43	5	5
44	3	4
45	6	4
46	4	4
47	2	1
48	4	4
49	2	2
50	1	1
51	2	2
52	2	2
53	2	1
54	2	2
57	2	2
58	2	2

Inconsistent deciles of values are marked in red.
The values of statistics in this case are given in the Table 3.

TABLE 3

	Vegetation index at the beginning of the investigation	**Change the value of the index during the investigation period**	**Seed moisture in the accounting area, %**
Standard deviation (Validation)	0.091016	0.033780	2.53000

Neural network learning outcomes can be applied to new datasets (for forecasting) in several ways: you can save trained networks and then apply them to a new dataset (for forecasting, classification or forecasting); you can use a code generator to automatically generate program code in a programming language (C ++, C #, Java, etc.).

An example of a crop loss forecasting program (in the C ++ programming language) is shown below in fig. 5 (based on Table 1).

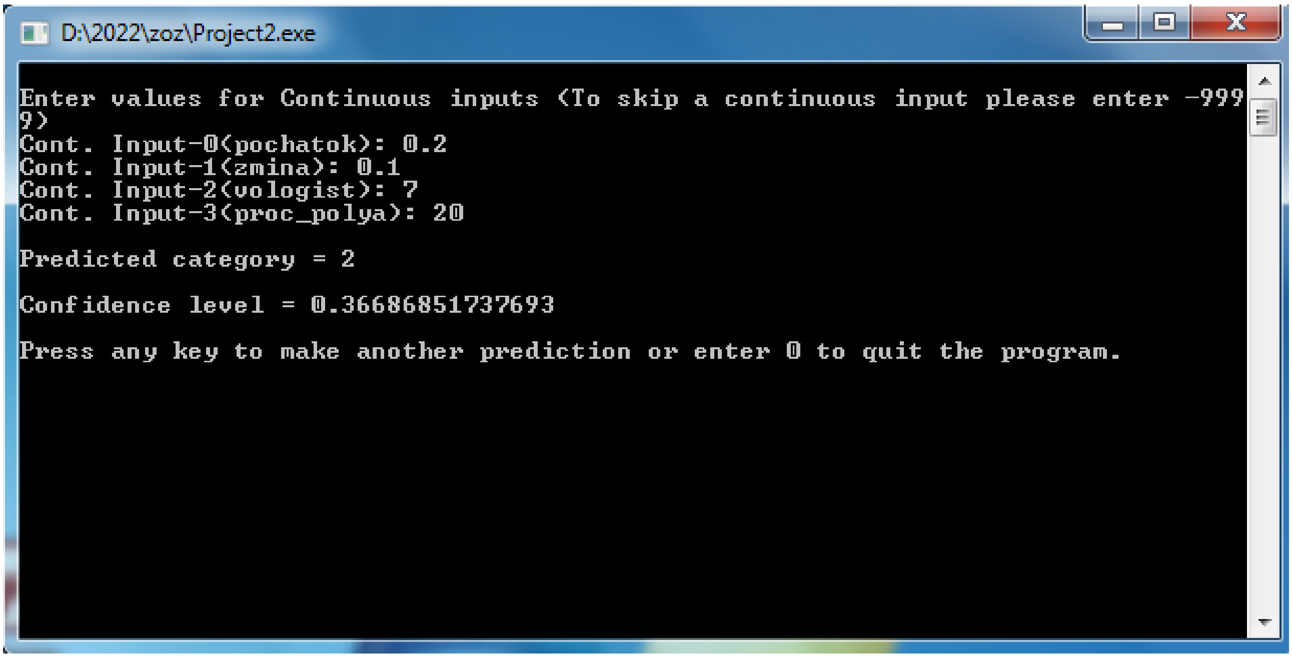

FIG. 5. Screenshot of the program in the C ++ programming language, which predicts the loss of sunflower harvest

Variable pochatok corresponds to the value of the vegetation index at the beginning of the investigation, zmina is a change the value of the index during the investigation period, vologist is a seed moisture in the accounting area, proc_polya is a percentage of study area from field area, Predicted category is a decile to which the program attributed crop losses.

The paper raises questions for further research:

- data preparation: how many classes to divide the initial values of crop losses,
- how losses are affected by various factors influencing the vegetation process, such as desiccation,
- what additional factors need to be included to improve the loss forecast,
- whether it is possible to extend this technique to other crops, such as corn.

The future investigations will be devoted to these questions.

Conclusion

The paper substantiates the possibility of forecasting the probable loss of sunflower harvest by the farmer on the basis of the analysis of the distribution of the vegetation index in the field and constructs a forecasting method (forecast model).

References

1. Zozulya O.L., Mykhalska L.M., Kovel O.L., Schwartau V.V. Digital technologies in crop production. Kyiv, 2020. 72 p. (in Ukrainian)
2. Snityuk V.E., Miroshnik O.N. Modeling and forecasting of processes in the real estate market. Cherkasy, 2014. 332 p. (in Russian)
3. Pantazi X., Moshou D., Bochtis D. Intelligent Data Mining and Fusion Systems in Agriculture. Academic Press Elsevier Inc., London, 2020. 319 p. https://doi.org/10.1016/B978-0-12-814391-9.00001-7
4. Haikin S. Neural networks Dialectics, 2020. 1104 p. (in Russian)
5. Tadeusiewicz R., Chaki R., Chaki N. Exploring Neural Networks with C #. New York CRC Press, 2015. 296 p. https://doi.org/10.1201/b17332

6. Sugomori Y., Kaluza B., Soares F.M., Souza A.M.F. Deep Learning: Practical Neural Networks with Java. Packt Publishing, 2017. 721 p.
7. Livshin I. Artificial Neural Networks with Java: Tools for Building Neural Network Applications. APress. 2019. 575 p. https://doi.org/10.1007/978-1-4842-4421-0
8. Virender S. C ++ Neural Networks and Fuzzy Logic. Amazon Digital Services, Inc., 2015. 447 p.

Received 13.12.2021

Zozulya Alexander,
PhD, Technical Support Lead, Marketing Department TOV Syngenta,
alexandr.zozulya@syngenta.com

Domrachev Volodymyr,
PhD, Associate Professor, Taras Shevchenko National University of Kyiv,
https://orcid.org/0000-0002-0624-460X
mipt@ukr.net

Tretynyk Violeta,
PhD, Associate Professor, The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".
https://orcid.org/0000-0002-3538-8207
viola.tret@gmail.com

УДК 519.67

О.Л. Зозуля [1], В.М. Домрачев [2], В.В. Третиник [3*]

Застосування технології штучних нейронних мереж для прогнозування втрат врожаю соняшника

[1] *ТОВ «Сингента»*
[2] *Київський національний університет імені Тараса Шевченка*
[3] *НТУУ «КПІ ім. Ігоря Сікорського»*
[*] *Листування:* viola.tret@gmail.com

Вступ. Штучні нейронні мережі (ШНМ) завоювали популярність як ефективний інструмент пропонування рішень для широкого спектру різноманітних біологічних досліджень та сільськогосподарської сфери. Їхня ефективність випливає із здібностей моделювати складні зв'язки між даними спостережень із датчиків і прогнозованими змінними, не покладаючись на припущення щодо структури моделі. Отже вони можуть передбачити реальну природу нелінійного відношення між вхідними та вихідними даними. Прогнозування врожайності є серйозною задачею в землеробстві пов'язаною з прийняттям найкращих методів управління, ціноутворення та безпеки.

Для прогнозування врожайності сільськогосподарських культур розроблено різноманітні методи та методики у сільському господарстві. Прогнозування врожайності вимагає контролю за багатьма параметрами, включаючи pH вмісту вологи, органічний склад ґрунту, загальний азот та органічний вуглець, що ускладнює процес прогнозування.

Мета. З'ясувати та обґрунтувати можливість прогнозувати імовірні втрати фермером врожаю соняшника на основі аналізу розподілу індексу вегетації на полі.

Результати. Вплив параметрів, що характеризують урожай, на його втрати є, але чітку регресійну залежність побудувати неможливо. Тому для побудови моделі використовується технологія штучних нейронних мереж. Модель формується у вигляді алгоритму, на вході якого задаються вхідні параметри (значення вегетаційного індексу на початку дослідження, зміна значення індексу за період дослідження, вологість насіння на спостережуваній ділянці, відсоток спостережуваної площі від площі поля), на виході отримуємо відсоток можливих втрат врожаю. Алгоритм автоматично транслюється в програму мовою програмування С++ (або іншою мовою програмування), що дозволяє на практиці моделювати можливі втрати врожаю фермера залежно від його дій щодо вирощування сільськогосподарських культур.

Ключові слова: соняшник, машинне навчання, штучні нейронні мережі, прогнозна модель.