Jan BARWICKI[1], Marek HRYNIEWICZ, Anna GRZYBEK

DOI: 10.15199/180.2020.1.3

[1] Institute of Technology and Life Sciences, Warsaw Branch,
Department of Rural Technical Infrastructure Systems
E-mail of the corresponding author: jbarwicki@.gmail.com

Jan Barwicki ORCID: 0000-0002-5437-5284
Marek Hryniewicz ORCID: 0000-0001-9926-699X ResearcherID: N-1547-2016 Scopus Author ID: 57210414113

YIELD FORECASTING USING ARTIFICIAL INTELLIGENCE

PROGNOZOWANIE PLONÓW PRZY UŻYCIU SZTUCZNEJ INTELIGENCJI

***Summary**: The article reviews and analyzes literature for application of artificial intelligence in forecasting of crop yield. Yield forecasting models were based on neural networks, fuzzy logic or hybrid solutions. When designing new yield forecasting models, analyzes of the main factors of components that are important for yield forecasting should be performed. This is to eliminate unnecessary or negligible factors for forecasting. It is also important to review the databases that will be used for forecasting. The data with unusual numerical results that differ significantly from reality should be deleted. This will improve the quality of the databases and, as a result, will give better forecasting results. In more complex cases, it would be recommended to create hybrid solutions combining neural networks and fuzzy logic to combine the advantages of both solutions.*

***Keywords**: artificial intelligence, hybrid solutions, fuzzy logic, neural networks, yield*

***Streszczenie**: W artykule wykonano przegląd i analizę literatury dla zastosowań sztucznej inteligencji przy prognozowaniu plonów. Modele prognozowania plonów były oparte o sieci neuronowe, logikę rozmytą lub rozwiązania hybrydowe. Przy projektowaniu nowych modeli prognozowania plonów należy przeprowadzić analizy głównych składowych czynników, które są istotne dla prognozowania plonu. Ma to na celu eliminację czynników zbędnych lub mało znaczących dla prognozowania. Istotne jest również dokonanie przeglądu baz danych, które zostaną wykorzystane do prognozowania.*

***Słowa kluczowe**: sztuczna inteligencja, rozwiązania hybrydowe, logika rozmyta, sieci neuronowe, wydajność*

Introduction

Jayram and Marad [2012] found that accurate forecasting of crop yield is of increasing importance in the developed and the developing countries and everywhere where agricultural production is carried out. Reliable forecasts are expected due to the cost-effectiveness of the agricultural production and a high involvement of mechanical equipment. Sawasawa [2003] showed that knowing the size of crop yield before plants harvesting is important for decision-makers and politicians, especially in the regions with major climate changes and a capricious weather. It allows them to make a decision about buying cereals in case of their shortage or selling in case of their excess. This is related to food security, the risk of which can also be assessed using fuzzy logic belonging to one of the methods used in artificial intelligence [Kadir and Inni, 2013]. Plant production can take place in the field or in greenhouses and plastic tunnels. Field production is closely related to the weather [Baruth and Inni, 2008]. In greenhouses, on the other hand, production is rather independent on the weather conditions [Qaddoum i inni, 2013]. Yields in plant production depend on several overlapping factors that usually change relative to each other at the same time, in a non-linear way [Boniecki & Niżewski, 2010]. This makes it difficult to predict yields using traditional methods. Boniecki and Niżewski [2010] proposed the use of procedures based on the artificial intelligence methods. Artificial intelligence appeared as a new field of knowledge along with the development of science. According to Kwatera [2016], artificial intelligence is a relatively new interdisciplinary field of science, a subject of great expectations and lively debates. In a theoretical sense, it combines the issues in the field of computer science, psychology, anthropology, mathematics, electronics, neurophysiology and philosophy. This serves to solve problems based on natural cognitive processes of man [Bartman 2017].

Objective, scope and methodology of work

The aim of the study was to review the artificial intelligence methods used to forecast crop yields. The scope of the work was to specify the methods and factors used in these methods that were utilized to forecast yields, and to prepare a brief description of the basics of the developed crop forecasting models. Literature analysis was adopted as the methodology of work.

Literature analysis

Niedbała et al. [2005] analyzed the assumptions for modelling sugar beet crop using artificial neural networks. Based on the literature information, own research and the conclusions from the regression analysis, they selected the following factors important for the needs of neuronal forecasting: soil bonitation class, level of organic and mineral fertilization expressed in the pure NPK component, sowing date and seed sowing standard, final plant density, date of beet harvest, temperature, sunshine and precipitation. Factors causing a decrease in the quality and quantity of the crop (diseases and pests) were omitted because it was assumed that their appearance was earlier detected and appropriate preventive measures were taken. The authors created, learned and tested various types of networks to select an artificial neural network with the right topology. MLP multi-layer perceptron networks were chosen for use due to lower error values. The designed network correctly predicted the sugar beet yield based on the input data provided. The authors conclude that the artificial neural networks can be an efficient tool for forecasting the effects of agricultural production not only in the case of sugar beet.

Boniecki and Niżewski [2010] designed, implemented and tested an IT system based on artificial neural network technologies that allowed forecasting the yield and starch content in potato tubers. This system forecasted the results on the basis of the following input data: average annual temperature, average annual amount of precipitation, potato variety, and average rainfall in the agrotechnical period and average temperature in the agrotechnical period. The authors used the Neural Networks package implemented in the Statistica v 7.1 IT system. The best yield prediction results were obtained by generating a RBF (Radial Basis Function) type topology with 5 input variables, 23 hidden neurons and one output neuron. In the case of starch forecasting, the best results were also obtained through the use of RBF networks. However, in this case, it turned out that the average rainfall of the agrotechnical period was variable with a negligible low weight for forecasting. Thus, input variables were reduced to four. 90 neurons were contained in the hidden neuron layer and one output neuron was used. The authors developed an IT system which they called 'Ziemniak'. This system contained previously described neural networks. The system made short-term forecasts of yield and starch content in potato tubers in an efficient and convenient way. The fact that the best results were obtained through the use of RBF type neural networks pointed to the non-linear nature of the relationship between input and output factors.

Qaddoum et al. [2013] used evolving fuzzy neural networks to predict the yield of tomatoes grown in greenhouses. This type of network was decided because they have the following advantages: the ability to model non-linear relationships within systems, resistance to imprecise, incomplete and uncertain input data. The advantages of this type of network have been revealed in the applications in various fields of technology related to forecasting, control, optimization and pattern recognition [Moraes & Machado, 2005]. As input variables, the temperature inside the greenhouse, the amount of CO_2 in the atmosphere, the water vapour pressure deficit in the air and the intensity of solar radiation and the history of previous crops were used. They adopted tomato yield per square meter as the output variable. The authors compared the results of the use of evolving fuzzy neural networks to predict the yield of tomatoes grown in the greenhouses for the use of a regular neural network and found that evolving fuzzy neural networks had a lower forecasting error than ordinary neural networks and required the last lower computing power

An average of 90% accuracy in predicting weekly fluctuations in tomato yields was achieved. In addition, the combination of the application of fuzzy logic in neural networks allowed reducing the sensitivity of the results for the adopted model of tomato yielding from imprecise input data. Its important feature was the inclusion of fuzzy logic inference rules (if-then rules) for each of the input variables. The authors stated that the network they developed and fully trained could be replaced by a set of if-then principles. An improvement of this system could be proposed by Frausto i Pieters [2004] by the use of neural networks to model the temperature inside the greenhouse as a function of external air temperature, air humidity, sunlight and cloudiness. For research they chose a network with one hidden layer of neurons. They stated that the network they proposed gave good results for long periods of operation, but they did not provide a statistical evaluation of their results.

Another solution would be the use of neural networks proposed by Singh and Tiwari [2017] for one-day prediction of average daily temperature and average daily relative humidity in the greenhouse based on input variables such as registered maximum temperature and minimum and relative humidity in the greenhouse, as well as average wind speed and insolation outside the greenhouse. As a result of research, an optimal network was built having 6 neurons in the input layer, 4 neurons in the hidden layer and 2 output neurons. The authors conducted a statistical analysis of the results and found that for the optimal network between the actual measured temperature and the predicted temperature, the mean square error was 0.711 °C and the mean absolute error was 0.558 °C. Between the measured and predicted humidity, the mean square error was 2.514% and the mean absolute error was 1.976%.

Jayram and Marad [2012] used a fuzzy logic system to forecast sorghum yield. They considered combinations of various plant characteristics that affect yield such as days for 50% flowering, the percentage of insect damage, plant height, inflorescence length, inflorescence weight, and number of ovules. There was found a very high correlation identified as a logarithmic relationship between the height of the crop and the length of the inflorescences with the mean square error value of 1.39. About 1000 data from reality were available. About 70%

of this data was used to train the system. The remaining 30% of the data was used to test the yield forecast. As a result of simulations and tests, the system was finally selected as input to the system: days for 50% flowering, height of inflorescences, weight of inflorescences, and the number of ovules. Very good forecasts made by the system were found.

Stathakis et al. [2010] created an adaptive inference system based on neural networks and fuzzy logic for forecasting wheat yield. As input data, some of the parameters of the crop growth simulation model used in the WOFOST system [Supit et al., 1994] were selected, such as soil moisture, the amount of biomass above the ground and the amount of biomass contained in the roots. In addition, information obtained from remote sensing in the form of a normalized differential vegetation indicator (NDVI English abbreviation) was used. NDVI values were collected every 10 days and were divided into two periods to create two different time profiles. The first period covered the data from the time of sowing to harvest (until August 1). The second period covered the data from the time of sowing to the point of heading (until June 1). This was done to investigate the differences between early crop prediction (when heading) and late forecast (during crop ripening). For comparison, an ordinary neural network was created that did not use fuzzy logic.

The ordinary neural network turned out to be completely unstable in its behaviour. In contrast, the neural network in conjunction with solutions derived from fuzzy logic gave quite good results. Overall forecasting accuracy averaged 74% for both the first and second periods. The authors suggested improving their system by taking into account the results of modelling dry matter production and meteorological factors such as temperature, amount of precipitation, snow depth, etc. The system can be improved by using genetic algorithms to select the optimal number of input parameters and their values.

Guo and Xue [2012] proposed forecasting wheat yield in the state of Queensland, Australia by testing and training the neural network. They stated that the yield depends on many factors such as the area of cultivation, water consumption, rainfall and temperature variability, seed quality, terrain topography, soil quality, emergence of diseases and pests, etc. It turned out that some of these factors are irrelevant to the time factor. Efforts have been made to use statistics to identify relationships between these factors and to use neural networks to map the effects of some of these factors on the crop yields. The complexity of the problem required the creation of a multi-criteria hybrid system that would take into account all or most of these factors in order to achieve a satisfactory level of monitoring crop growth and yield forecasting. First, historical data in the state of Queensland were statistically examined. As a result of this analysis, the following factors were determined that had the greatest impact on wheat yield: crop area, rainfall and temperature.

Secondly, unusual results for designated factors were removed to achieve better results of future forecasts. Thirdly, a neural network was created, the operation of which was based on data from previously determined factors. The authors showed that their neural network with 100 nodes predicted wheat yield with an average absolute error of 2.06%, a maximum error of 9.64% and a standard deviation of 2.96%. It was stated that at the beginning a statistical analysis (multifactorial regression) should be carried out to identify the most important factors affecting the yield. This analysis should also help detect unusual data values to remove them from the data set and identify trends in the data changes.

Only then you can build and train a neural network. In the future, the neural network can be expanded by taking into account other factors.

Kumar et al. [2010] applied fuzzy logic to time series to predict wheat production in the experimental fields of the University of Agriculture and Technology in Pantnagar, India. Time series were created based on available yield data from 1981-2002. Three models based on fuzzy logic were created. The first of these was based on Chen's arithmetic [1996]. The second one was built using modified medium weights. The third of them was a time-unchanging model. Crop forecasts were made using all three models. It turned out that they gave very similar results with an average error of around 11%. The mean square error was 138.458, 135.105 and 140.901 for these models, respectively.

Aggarwal et al. [2017] used a fuzzy logic system to forecast rice production yield. This system was based on the analysis of fuzzy time series of historical data on rice yields and forecasting future yield on this basis. They proposed dividing the range of predicted yields into 7 equal intervals and compared the results of forecasting with divisions into 9 and 11 equal intervals proposed by other authors. The system operation was assessed on the basis of the average forecasting percentage error and mean square error, which were respectively: 3.45% and 22.737.5 for 7 ranges, 3.75% and 28.088,6 for 9 ranges and 3.13% and 23.478, 1 for 11 compartments. The system with 7 compartments gave the best results.

Kandala and Prajneshu [2002] proposed a method based on fuzzy regression using remotely collected data. The input data came from photographs of crops taken in the visible and near infrared range. Based on these calculations, a normalized differential vegetation index (NDVI abbreviation) and vegetation index (RVI abbreviation) were calculated. According to the results of ANSARI et al [1999], these indicators are closely correlated and cause the problem of collinearity of these indicators. This results in an increase in standard errors for test results using these indicators. The methods used so far using principal component analysis, dorsal regression, etc., did not give satisfactory results and were not reliable. Only using the fuzzy regression method allowed to obtain satisfactory results. However, the authors did not provide errors or any statistical evaluation of the results obtained.

Summary and conclusions

Advances in the development of artificial intelligence have made it possible to solve the problems of combining non-linear factors with a high degree of uncertainty and with low-quality data into working models for forecasting future yields. This was done successfully for all kinds of plant species. These models used information from one to several factors. Researchers have proved that it was possible to forecast yields correctly based only on data on the history of yields collected over a dozen or so decades. Crop forecasting was also carried out using more factors. There was a limitation of data availability.

To develop yield forecasting models, solutions based only on neural networks or fuzzy logic were used, as well as hybrid solutions combining these two issues. In the case of designing new models for yield forecasting, it would be advisable to carry out an analysis of the main components of factors that are important for yield forecasting in order to eliminate redundant or insignificant factors for forecasting. It would also be important to review the databases that will be used for forecasting. The data with unusual numerical results, which differ significantly from reality, should be removed. This will improve the quality of the databases and, as a result, will give better forecasting results. In more complex cases, it would be recommended to create hybrid solutions combining neural networks and fuzzy logic to achieve the synergy of the advantages of both solutions.

References

[1] Aggarwal S., Sokhal J., Garg B. 2017. Forecasting Production Values using Fuzzy Logic Interval based Partitioning in Different Intervals. International Journal of Advanced Computer Science and Applications, Vol. 8(5), ISSN 2156-5570, s. 292-299.

[2] Ansari M., Mahley R., Sidhu S. 1999. Cotton yield prediction through spectral parameters. Journal of the Indian Society of Remote Sensing. Vol 27, ISSN 0255-660X, s. 185-192.

[3] Bartman J. 2017. Sztuczna inteligencja. Wprowadzenie. Artificial intelligence. Introduction

[4] [online]. Rzeszów. [Dostęp 30.07.2017]. Dostępny w Internecie. http://www.neurosoft.edu.pl/media/pdf/jbartman/sztuczna_inteligencja/SI_1_wprow.pdf

[5] Baruth B., Royer A., Klisch A., Genovese G. 2008. The use of remote sensing within the MARS crop yield monitoring system of European Commission. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B8, Beijing, s. 935-940.

[6] Boniecki P., Niżewski P. 2010. Modelowanie neuronowe w rozwiązywaniu wybranych problemów predykcyjnych Inżynierii Rolniczej. Neural modeling in solving selected predictive problems of Agricultural Engineering.Journal of Research and Applications in Agricultural Engineering. Vol. 55(1), ISSN 1732-1719, s. 16-19.

[7] Chen S. 1996. Forecasting enrolments based on fuzzy time series. Fuzzy Sets and Systems, vol. 81, ISSN 0165-0114, str. 311- 319.

[8] Frausto H., Pieters J. 2004. Modelling greenhouse temperature using system identification by means of neural networks. Neurocomputing, vol. 56, ISSN 0925-2312.

[9] Guo W., Xue H. 2012. An incorporative statistic and neural approach for crop yield modelling and forecasting. Neural Computing and Applications. Vol. 21(1), ISSN 1433-3058, s. 109-117.

[10]. Jayram M., Marad N. 2012. Fuzzy Inference Systems for Crop Yield Prediction. Journal of Intelligent Systems, no 21. ISSN 2191026X, s. 363-372.

[11] Kadir M., Hines E., Qaddoum K., Collier R., Dowler E., Grant W., Leeson M., Iliescu D., Subramanian A., Richards K., Merali Y., Napier R. 2013. Food security risk level assessment: a fuzzy logic-based approach. Applied Artificial Intelligence, Volume 27 (Number 1), ISSN 1087-6545, s. 50-61.

[12] Kandala V., Prajneshu. 2002. Fuzzy Regression Methodology for Crop Yield Forecasting Using Remotely Sensed Data. Journal of the Indian Society of Remote Sensing. Vol. 30, No. 4, ISSN 0974-3006, s. 191-195.

[13] Kwater T. 2016. Sztuczna inteligencja. Definicja. Artificial intelligence. Introduction [online]. Rzeszów. [Dostęp 30.06.2017]. Dostępny w Internecie: http://www.neurosoft.edu.pl/media/pdf/tkwater/nowoczesne_techniki_informatyczne/szint.pdf

[14] Moraes R., Machado L. 2005 Evaluation System Based on EfuNN for On-line Training Evaluation in Virtual Reality. Lecture Notes in Computer Science. Berlin, v. 3773, ISSN 0302-9743, s. 778-785.

[15] Niedbała G., Przybył J., Boniecki P., Sęk T. 2005. Analiza założeń dla modelowania plonu buraka cukrowego z wykorzystaniem sztucznych sieci neuronowych. Analysis of assumptions for modeling sugar beet crop using artificial neural networks Inżynieria Rolnicza. Vol. 2, ISSN 2083-1587, s. 123-130.

[16] Qaddoum K., Hines E., Iliescu D. 2013. Yield Prediction for Tomato Greenhouse Using EFuNN. ISRN Artificial Intelligence. Vol. 2013, ISSN 2356-7872, s. 1-9.

[17] Sawasawa H. 2003. Crop yield estimation: Integrating RS, GIS and management factors. Praca magisterska wykonana na wydziale Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente[online]. Enschede, Holandia. [Dostęp 30.06.2017]. Dostępny w Internecie: https://pdfs.semanticscholar.org/6ad3/7c4754e7ef2b3770c92fb74ce098543bfabf.pdf

[18] Singh V., Tiwari K. 2017. Prediction of greenhouse micro-climate using artificial neural networks. Applied ecology and environmental research, vol. 15(1), ISSN 1785 0037, s. 767-778.

[19] Stathakis D., Savin I., Negre T. 2010. Neuro-fuzzy modelling for crop yield prediction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 34, part XXX, ISSN 0924-2716, str. 4.

[20] Supit I., Hooijer A., van Diepen C. 1994 System description of the WOFOST 6.0 crop growth simulation model. JRC, Brussel-Luxembourg, str. 125.

Article reviewed

Received: 30.01.2020/Accepted: 10.02.2020