

УДК 519.23:633.111.1:528.88

DOI <https://doi.org/10.32851/2226-0099.2020.113.9>

ЕФЕКТИВНІСТЬ МЕТОДІВ СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ У ПРОГНОЗУВАННІ ВРОЖАЇВ ПШЕНИЦІ ОЗИМОЇ НА РЕГІОНАЛЬНОМУ РІВНІ ЗА ДАНИМИ СУПУТНИКОВОГО МОНІТОРИНГУ

Лиховид П.В. – к.с.-г.н., старший науковий співробітник відділу маркетингу, трансферу інновацій та економічних досліджень, Інститут зрошуваного землеробства Національної академії аграрних наук України
Лаверенко С.О. – к.с.-г.н., доцент кафедри землеробства, ДВНЗ «Херсонський державний аграрно-економічний університет»
Лаверенко Н.М. – к.с.-г.н., доцент кафедри землеустрою, геодезії та кадастру, ДВНЗ «Херсонський державний аграрно-економічний університет»

Здійснено аналіз ефективності статистичної обробки даних супутникового моніторингу та врожайності пшениці озимої на регіональному рівні для одержання прогнозованих рівнів урожайності в Херсонській області методами лінійної регресії та штучних нейронних мереж.

Вихідні дані, застосовувані для побудови прогностичних моделей, включають регіональні рівні врожайності пшениці озимої за період 2012-2019 рр. (згідно звітів Державної статистичної служби України) і результати обчислення величин вегетаційних індексів у програмі QGIS 3.10 за супутниковими знімками MODIS Terrain NDVI/EVI з роздільною здатністю 250 м.

Регресійний аналіз виконували за допомогою програми BioStat v7, штучні нейронні мережі будували, навчали та тестували у програмі Tiberius. Обидва методи статистичної обробки даних є прийнятними для одержання надійних прогностичних моделей врожайності пшениці озимої в досліджуваному регіоні за величинами нормалізованого диференційного (NDVI) та поліпшеного (EVI) вегетаційних індексів із 95% рівнем достовірності.

Встановлено, що застосування штучних нейронних мереж із двома типами нейронів (лінійним і нелінійним) істотно поліпшувало точність і надійність прогнозування врожайності зерна пшениці озимої при використанні даних NDVI, про що свідчать величини коефіцієнтів детермінації (95,46% для нейронних мереж і 89,72% для лінійної регресії) та середньої абсолютної похибки (2,63% і 5,70%). Не доведено суттєвої переваги нейронних мереж у прогнозуванні врожайів культури на регіональному рівні за величинами EVI, оскільки величини коефіцієнтів детермінації (88,47% для нейронних мереж і 88,72% для лінійної регресії) та середньої абсолютної похибки (5,29% і 6,52%) за обох методів прогнозування виявилися дуже близькими.

Таким чином, абсолютної переваги штучних нейронних мереж над лінійною регресією для одержання надійних прогнозів регіонального рівня врожайності пшениці озимої не доведено, хоча в конкретних випадках вони забезпечують децю вищий рівень точності. Неоліком методу є неможливість одержання рівняння прогнозу, яке є доступним у разі регресійного моделювання врожайності.

Ключові слова: пшениця озима, лінійна регресія, штучні нейронні мережі, супутниковий моніторинг, нормалізований диференційний вегетаційний індекс, поліпшений вегетаційний індекс.

Lykhovyd P.V., Lavrenko S.O., Lavrenko N.M. Efficiency of the methods of statistical data analysis in forecasting winter wheat yields on the regional scale using the remote sensing data

The analysis of the efficiency of statistical processing of the remote sensing data and winter wheat yields on the regional scale in order to obtain predicted levels of the yields in Kherson oblast by the means of linear regression and artificial neural networks has been performed.

The inputs for the prediction models include regional levels of winter wheat yields for 2012-2019 (with accordance to the reports of the State Statistical Service of Ukraine) and the results of calculations of vegetation indices using QGIS 3.10 software by the spatial imagery of MODIS Terrain NDVI/EVI with a resolution of 250 m. Regression analysis has been performed using

BioStat v7 software, artificial neural networks have been developed, trained and validated using Tiberius software.

Both methods of statistical data processing are acceptable for obtaining reliable prediction models of winter wheat yields in the studied region by the values of normalized difference (NDVI) and enhanced (EVI) vegetation indices at the probability level of 95%. It was determined that application of artificial neural networks with two types of neurons (linear and non-linear) improved significantly the accuracy and reliability of the prediction of winter wheat grain yields in case of NDVI data use, that is testified by the coefficients of determination (95.46% for a neural network and 89.72% for a linear regression, respectively) and the mean absolute percentage error (2.63% and 5.70%, respectively).

At the same time, there was no evidence for the advantage of neural networks over linear regression in the forecasting the crop yields on the regional level using the EVI values because the coefficient of determination (88.47% for a neural network and 88.72% for a linear regression, respectively) and the mean absolute percentage error (5.29% and 6.52%, respectively) were almost equal at the both prediction methods.

Therefore, there is no evidence for the absolute benefit of artificial neural networks over linear regression for getting reliable predictions of the regional levels of winter wheat yields, although in some cases they provide somewhat higher level of accuracy. The drawback of the method is impossibility of getting the formula for prediction, while this option is available at the regression modeling.

Key words: winter wheat, linear regression, artificial neural networks, remote sensing, normalized difference vegetation index, enhanced vegetation index.

Постановка проблеми. Раннє прогнозування врожаїв сільськогосподарських культур є важливим завданням сучасної аграрної науки. Широке впровадження інформаційних технологій у землеробство та рослинництво зумовлюють попит у агровиробників на інтелектуальну продукцію, що здатна підвищити ефективність систем точного землеробства та забезпечити максимально високу віддачу від них.

Одним із таких інформаційних продуктів є надання прогностичних даних щодо врожайності. Крім того, важливо розробляти прогностичні моделі як для окремих аграрних масивів, так і для великих за своїми масштабами регіонів (районів, областей, країн). Це має велике значення для забезпечення продовольчої безпеки, формування аграрної політики та імпортно-експортної стратегії в аграрному секторі, прийняття раціональних управлінських рішень тощо.

Нині існує декілька шляхів вирішення зазначеної проблеми. Втім, прогностичні моделі для великих за обсягом територій можливі здебільшого лише завдяки доступу до даних супутникового моніторингу територій, на основі яких розраховують індекси, за якими можна опосередковано характеризувати та оцінювати стан рослинних угруповань. Наприклад, це нормалізований диференційний вегетаційний індекс (далі – NDVI), а також менш поширений, але в окремих випадках більш точний, поліпшений вегетаційний індекс (далі – EVI), який враховує спотворення супутникових знімків через хмарність і наявність у повітрі аерозолів [1]. Хоча це твердження є суперечливим, але інколи NDVI демонструє кращі результати в оцінці вегетативного покриву [2]. Саме на аналізі вегетаційних індексів, а інколи додатково метеорологічних і ґрунтових умов, базується створення масштабних моделей прогнозування врожаїв різних сільськогосподарських культур на великих земельних масивах.

Варто враховувати той факт, що результати прогнозування здебільшого залежатимуть і від технології математичної обробки вхідних даних. Натепер найпопулярнішими методами є різні варіації регресійного аналізу, а також штучні нейронні мережі [3; 4]. Варто визначитися, які саме методики роботи з даними забезпечують найкращу віддачу саме в аграрній сфері.

Аналіз останніх досліджень і публікацій. Лінійна регресія – один із найдавніших і найбільш поширених методів математичного моделювання, який знайшов широке застосування і в найрізноманітніших моделях прогнозування врожаїв сільськогосподарських культур [5]. Головними перевагами регресійного моделювання є його відносна доступність, простота, а також можливість ручного налаштування параметрів моделі, отримання на виході математичного рівняння прогнозу [6]. Головними недоліками є недосконалість лінійних алгоритмів, які найчастіше застосовуються під час моделювання, що позначається на точності прогнозів [7].

Штучні нейронні мережі є сучаснішим і досконалішим інструментом роботи з даними. Застосування нейронних мереж у моделюванні та прогнозуванні природних процесів і продуктивності сільськогосподарських культур часто є більш ефективним, але головним недоліком є неможливість одержати прогностичне рівняння, тобто дослідник не матиме гадки про те, яким чином нейронна мережа дійшла до таких результатів [8].

Враховуючи той факт, що як регресійні, так і нейромоделі мають свої переваги та недоліки, варто провести низку досліджень із порівняння їх функціональної ефективності для вирішення конкретних завдань агромоделювання.

Постановка завдання. Метою дослідження було порівняння методів лінійної регресії та штучних нейронних мереж щодо їх ефективності в прогнозуванні врожаїв на регіональному рівні для Херсонської області на прикладі пшениці озимої з використанням даних супутникового моніторингу. Для цього попередньо було обраховано величини вегетаційних індексів MODIS NDVI та EVI за супутниковими знімками з роздільною здатністю 250 м, наданими Університетом природних ресурсів та природничих наук (Австрія) в програмному забезпеченні QGIS 3.10 (із використанням функцій програмного пакету «Растровий аналіз»).

Супутникові знімки Херсонської області за період 2012-2019 рр. попередньо було обрізано в програмі QGIS 3.10 за маскою рослинного покриву регіону, оскільки це підвищує точність розрахунків і дозволяє вилучити об'єкти, які не мають відношення до рослинності (будівлі, споруди, піщані масиви, водойми тощо). Прогнозування врожаїв пшениці озимої в регіоні за величинами вегетаційних індексів виконували для періоду травня (30-45 днів до планованого збирання врожаю зерна культури).

Регресійні моделі врожайності зерна пшениці озимої залежно від вегетаційних індексів було розроблено в програмному пакеті BioStat v7 [9]. Для побудови штучних нейронних мереж було використано програмне середовище Tiberius, в якому було побудовано, навчено (протягом 10 000 епох із темпом навчання 0,70), протестовано прогностичні моделі [10].

Порівняння точності методів прогнозування врожайності пшениці озимої базувалося на обчисленні коефіцієнтів детермінації та середньої абсолютної похибки у відсотках [11]. Вищі значення коефіцієнту детермінації та нижчі значення похибки асоціюють із вищою точністю моделі [12].

Виклад основного матеріалу дослідження. Статистична обробка врожайних даних і даних щодо величини NDVI та EVI по Херсонській області за період 2012-2019 рр. дозволила побудувати прості лінійні моделі прогнозування врожаю зерна пшениці озимої на регіональному рівні за даними попередніх періодів (табл. 1).

Схематичне зображення побудованої нейромоделі з двома типами нейронів (лінійним і нелінійним) наведено на рис. 1.

Моделі врожайності зерна пшениці озимої в Херсонській області, побудовані методами лінійного регресійного аналізу та штучних нейронних мереж, дозво-

Таблиця 1

Регресійна статистика відношень між урожайністю зерна пшениці озимої в Херсонській області та величинами регіональних NDVI та EVI

Критерії	NDVI	EVI
Коефіцієнт кореляції (R)	0.9472	0.9426
Середня квадратична похибка (MSE)	0.0676	0.0732
Коефіцієнт детермінації (R ²)	0.8972	0.8886
Коригований R ²	0.8800	0.8700
Прогнозований R ²	0.8376	0.8285
Середня абсолютна похибка (MAPE), %	5.7010	6.5170
Вільний член регресійного рівняння	-8.0022	-4.2962
Аргумент	21.0469	23.7742
Модель	$Y = -8,0022 + 21,0469X$	$Y = -4,2962 + 23,7742X$

лили оцінити кожен із застосованих методів (табл. 2, 3). Так, за результатами апроксимації моделей «врожайність зерна пшениці озимої у зв'язку з NDVI» та «врожайність зерна пшениці озимої у зв'язку з EVI» було встановлено, що штучні нейронні мережі мають помітну перевагу (5,74% за величиною коефіцієнта детермінації та 3,07% за величиною абсолютної середньої похибки) над лінійною регресією лише у першому випадку, в той час як за застосування у якості вхідного параметра величини EVI точність моделей практично не відрізняється за статистичною достовірністю.

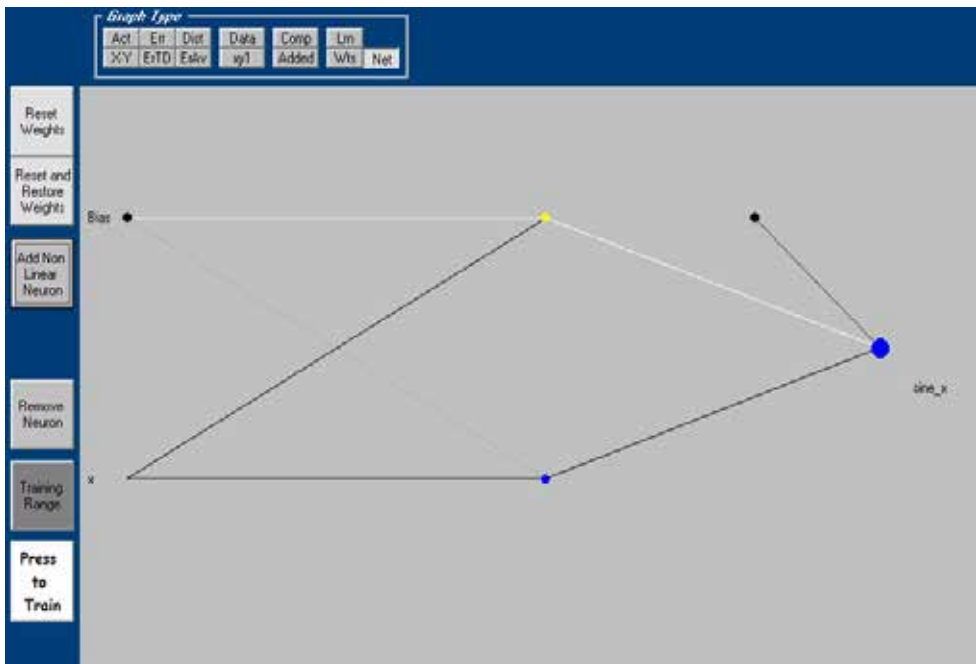


Рис. 1. Схематичне зображення нейронної мережі для прогнозування врожайності пшениці озимої

Таблиця 2

**Урожайність зерна пшениці озимої, прогнозована
за величиною регіонального NDVI по Херсонській області**

Фактична	Прогнозована нейронною мережею	Прогнозована за лінійною регресій- ною моделлю
1,69	1,69	1,68
2,25	2,25	2,52
2,94	2,95	2,94
3,89	3,53	3,57
3,62	3,62	3,78
3,49	3,23	3,15
3,22	3,23	3,15
3,49	3,62	3,78
Коефіцієнт кореляції	0,9770	0,9472
Коефіцієнт детермінації, %	95,46	89,72
Середня абсолютна похибка, %	2,63	5,70

Таблиця 3

**Урожайність зерна пшениці озимої, прогнозована
за величиною регіонального EVI по Херсонській області**

Фактична	Прогнозована нейронною мережею	Прогнозована за лінійною регресій- ною моделлю
1,69	1,68	1,65
2,25	2,26	2,60
2,94	2,26	3,07
3,89	3,62	3,79
3,62	3,62	3,55
3,49	3,23	3,07
3,22	3,23	3,07
3,49	3,62	3,79
Коефіцієнт кореляції	0,9406	0,9419
Коефіцієнт детермінації, %	88,47	88,72
Середня абсолютна похибка, %	5,29	6,52

Мінімальна різниця в коефіцієнті детермінації для EVI-моделі врожаїв зерна пшениці озимої, що склала лише 0,25%, а також незначна перевага штучної нейронної мережі в показнику середньої абсолютної похибки (1,23%), підтверджують факт відсутності переваг нейромоделі над простішою лінійною регресійною моделлю.

Таким чином, нами не було отримано переконливі докази щодо вищої ефективності застосування штучних нейронних мереж у прогнозуванні врожаїв сільськогосподарських культур за даними вегетаційних індексів. Втім, це питання залишається відкритим, оскільки для його остаточного вирішення необхідною умовою є проведення досліджень за різними алгоритмами навчання нейронних мереж у різних програмних середовищах, за даними урожайності різних сільськогосподарських культур, вирощуваних за різних агрокліматичних умов і техноло-

гій, залучення масивів даних із різним набором вхідних параметрів (малі та великі вибірки) тощо [8].

Висновки і пропозиції. Прогнозування врожаїв на регіональному рівні є важливим та актуальним завданням сучасної аграрної науки. Застосування даних супутникового моніторингу створює можливість для вирішення цього завдання. Втім, проблема обрання методики побудови прогностичної моделі залишається не вирішеною.

Проведена нами оцінка двох поширених методів, що застосовуються для програмування врожаїв сільськогосподарських культур, а саме традиційного (лінійна регресія) та інноваційного (штучні нейронні мережі), показала, що штучні нейронні мережі не завжди мають перевагу над регресійними моделями, а також засвідчила, що ця перевага (в межах 5-7% точності) не може бути визначальною.

Отже, обидва методи є прийнятними для програмування та прогнозування урожайності сільськогосподарських культур на регіональному рівні. Втім, подальше вивчення проблеми з більшими масивами дослідних даних є вкрай необхідним для остаточного вирішення цього питання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Rocha A.V., Shaver G.R. Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology*. 2009. Vol. 149. № 9. P. 1560–1563. DOI: <https://doi.org/10.1016/j.agrformet.2009.03.016>.
2. Feysa G.L., Palao L.K., Nelson A., Gumma M.K., Paliwal A., Win K.T., Nge K.H., Johnson D.E. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Computers and Electronics in Agriculture*. 2020. Vol. 175. P. 105595. DOI: <https://doi.org/10.1016/j.compag.2020.105595>.
3. Wang Y.P., Chang K.W., Chen R.K., Lo J.C., Shen Y. Large-area rice yield forecasting using satellite imageries. *International Journal of Applied Earth Observation and Geoinformation*. 2010. Vol. 12. № 1. P. 27–35. DOI: <https://doi.org/10.1016/j.jag.2009.09.009>.
4. Jiang D., Yang X., Clinton N., Wang N. An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing*. 2004. Vol. 25. № 9. P. 1723–1732. DOI: <https://doi.org/10.1080/0143116031000150068>.
5. Ушкаренко В.О., Лиховид П.В. Регресійна модель урожайності кукурудзи цукрової залежно від агротехнології в зрошуваних умовах Сухого Степу України. *Вісник Уманського національного університету садівництва*. 2016. № 2. С. 31–34.
6. Fox J. Applied regression analysis, linear models, and related methods. Sage Publications, Inc., 1997.
7. Graybill F.A. Theory and application of the linear model. North Scituate, MA : Duxbury press, 1976.
8. Vozhehova R.A., Lykhovyd P.V., Kokovikhin S.V., Biliaieva I.M., Markovska O.Y., Rudik O.L. Artificial neural network and their implementation in agricultural science and practice. Warsaw : Diamond Trading Tour, 2019.
9. Айвазян С.А. Прикладная статистика. Основы эконометрики. Том 2. М. : Юнити-Дана, 2001.
10. Brierley P., Batty B. Data mining with neural networks-an applied example in understanding electricity consumption patterns. *Knowledge discovery and data mining*. 1999. Vol. 12. P. 240–303.
11. De Myttenaere A., Golden B., Le Grand B., Rossi F. Mean absolute percentage error for regression models. *Neurocomputing*. 2016. Vol. 192. P. 38–48. DOI: <https://doi.org/10.1016/j.neucom.2015.12.114>.
12. Nagelkerke N.J.D. A note on a general definition of the coefficient of determination. *Biometrika*. 1991. Vol. 78. № 3. P. 691–692. DOI: <https://doi.org/10.1093/biomet/78.3.691>.