

УДК 631.15

DOI: <https://doi.org/10.17721/3041-2323.2024.82-88>

Владислав ГНАТІЄНКО, асп.
ORCID ID: 0009-0000-2678-5158
e-mail: hnatienko.vladyslav@knu.ua
Київський національний університет
імені Тараса Шевченка, Київ, Україна

Віталій СНИТЮК, д-р техн. наук, проф.
ORCID ID: 0000-0002-9954-8767
e-mail: snytyuk@knu.ua
Київський національний університет
імені Тараса Шевченка, Київ, Україна

КОНЦЕПТУАЛЬНІ ЕЛЕМЕНТИ ТЕХНОЛОГІЇ РОЗПОДІЛЕНОГО ПРОГНОЗУВАННЯ ВРОЖАЙНОСТІ СІЛЬСЬКОГОСПОДАРСЬКИХ КУЛЬТУР

Основна мета дослідження – підвищити точність прогнозування врожайності за допомогою розроблення методів оброблення даних і нейромережових технологій. Розроблено технологію прогнозування врожайності, що охоплює моделі розпізнавання образів для аналізу супутникових знімків і підвищення якості даних у навчальних наборах, методи оброблення даних і глибокі нейронні мережі у поєднанні з іншими моделями штучного інтелекту. На основі цієї технології проаналізовано ефективність і доцільність застосування деяких агротехнічних заходів, що забезпечить підтримку прийняття рішень у виробництві сільськогосподарської продукції.

Ключові слова: сільськогосподарська продукція, розподілене прогнозування, точність прогнозування.

Вступ

Цифрова агрономія знаходиться на етапі активного розвитку, власники господарств усе частіше охоплюють стратегії застосування онлайн-менеджменту, що дозволяє дистанційно моніторити та контролювати польові роботи. Фахівці застосовують штучний інтелект, проводять дослідження для поглиблення знань і розроблення ефективних технологій цифрової агрономії. У сучасних дослідженнях методів прогнозування врожайності досяга-

© Гнатієнко Владислав, Снитюк Віталій, 2024

ється середньоквадратична помилка (RMSE) на рівні 10–15 % середніх значень урожайності. Більшість моделей використовують для прогнозування сумарного врожаю поля, без можливості виконання розподіленого прогнозування. Ті моделі, за допомогою яких будують детальні карти прогнозованої врожайності, загалом протестовані на замалих вибірках, через що неможливо достовірно оцінити їхню ефективність.

Нині актуальною задачею є розроблення технології прогнозування з деталізацією до окремих ділянок поля. У процесі дослідження проаналізовано особливості даних, зв'язки та ступені впливу на врожайність різних агрономічних показників. Наукова цінність полягає в тому, що інформація про особливості агрономічних даних дозволить спростити й оптимізувати майбутні дослідження, надавши розуміння, які дані і чому ефективно застосовувати, а деталізація прогнозів до окремих ділянок відкриє нові напрями майбутніх досліджень. З практичного погляду прогнозування дасть можливість планувати бюджет, аналізувати ризики та вживати завчасних заходів.

Очікувані результати дослідження забезпечать відкритий доступ до інноваційних рішень, що сприятиме розвитку цифрової агрономії.

Результати

Ступінь розробленості теми. Нині проводять багато досліджень, метою яких є підвищення точності прогнозування врожайності різноманітних культур. Використовують широкий спектр джерел інформації, таких як генетичні дані рослин, дані про умови навколишнього середовища, супутникові знімки тощо. Для оброблення й аналізу цих даних застосовують різні моделі – від класичних статистичних моделей до сучасних глибоких нейронних мереж.

У дослідженні (Al-Gaadi et al., 2016) для прогнозування врожаю використано дані про генотип рослин, погодні умови та показники ґрунту. Для прогнозування застосували глибоку нейромережу й отримали RMSE на рівні 11–12 % середнього рівня врожаю. Проте розглядають лише прогнозування сумарного врожаю цілого поля без деталізації щодо окремих його ділянок, що є обмеженням для можливих застосувань такого підходу. Такий самий недолік маємо в дослідженні (Amankulova et al., n. d.), в якому використовувалась модель Random forest. Показники RMSE становлять 11.9 %, 16.7, 13.9 і 5.8 % середніх значень для

пшениці, кукурудзи у США, картоплі і кукурудзи в Північно-Східному прибережному регіоні США, відповідно. Автори зазначають, що вибрана модель схильна до перенавчання, через що можуть виникнути складності в узагальненні. Також проблемою є її низька надійність: модель ефективна в середньому, що дозволяє аналізувати загальні особливості великих даних, проте виявлено високу ймовірність отримати значні похибки у процесі виконання окремих прогнозів.

Ефективними є методи, в яких використовують супутникові знімки для прогнозування. Наприклад, у дослідженні (Jeong et al., 2016) прогнозується врожай картоплі на основі виявлення залежності між значеннями вегетаційних індексів і значеннями врожайності з відхиленнями на рівні 5–9 % від середніх показників, проте елементи навчального та тестового наборів належали одному полю, через що складно оцінити можливості узагальнення такого підходу. В дослідженні (Khaki, & Wang, 2019) 5 полів формують навчальний набір, а на інших 5 полях відбувається тестування моделі. Результати вказують на те, що найефективніша модель – Random forest, яка працює зі значеннями RMSE від 0,284 до 0,473 т/га, що становить близько 9–14 % від середніх показників урожайності соняшника, але про надійність таких результатів стверджувати складно, з огляду на невеликий розмір тестової вибірки. Як зазначають автори, під час дослідження був вдалий період для збору інформації: вдалось зібрати 16 супутникових знімків у сонячні дні впродовж періоду дозрівання, що створило сприятливі умови для навчання моделі. Зазвичай через постійну наявність хмар вдається зібрати лише від 3 до 7 знімків, що значно ускладнює прогнозування з використанням подібних моделей.

Актуальність. Основними недоліками сучасних технологій прогнозування є недостатня точність і сильна залежність від погодних умов. Також у більшості досліджень зосереджуються на прогнозуванні сумарного врожаю, а ті, в яких виконується розподілене прогнозування, зазвичай мають значно обмежені вибірки в експериментальних верифікаціях, через що неможливо достовірно оцінити ефективність запропонованих підходів. Додаткові проблеми виникають через відсутність знань про доцільність використання різних джерел даних: невідомо, які фактори найбільше впливають на врожайність рослин, через це в дослідженнях можна побачити використання інформації про генотипи рослин,

погодні умови, перепади рельєфу, вміст азоту і фосфору в ґрунті, супутникові знімки тощо. Усі ці дані надходять із різних джерел різними способами та мають великі відмінності у складності й вартості отримання.

З погляду практичного застосування, агрономічні експерти та власники господарств стикаються з тим, що сучасні сервіси прогнозування не забезпечують достатню точність і надійність. Відхилення прогнозованих значень від реальних часто коливається від 3–5 % до 30–40 %, що є закономірним наслідком недостатньої дослідженості цього напрямку в науковому просторі. Це призводить до того, що від прогнозування часто відмовляються, що негативно впливає на розвиток сільськогосподарського виробництва: постачальники технологій аналітики даних і прийняття рішень не отримують достатнього фінансування на їхній розвиток, а власники господарств втрачають потенційний прибуток, який могли б отримати, застосовуючи такі технології.

Наукова новизна роботи полягає в особливій архітектурі системи прогнозування та методах оброблення даних, що розроблятимуться для мінімізації впливу погодних умов, зокрема і наявності великої кількості хмар. В процесі реалізації, з метою оптимального налаштування моделей, буде проаналізовано вплив різних факторів на врожайність і виявлено, які з них є найважливішими, що дозволить спростити майбутні дослідження. Також для підвищення якості навчальних наборів і кращого розуміння даних проводитиметься аналіз внутрішніх просторів моделей і внутрішніх представлень даних, що дасть багато корисної інформації про особливості агрономічних даних, їхні зв'язки та патерни і, зокрема, це дозволить доповнити аналіз впливу факторів на точність моделі. Важливим напрямом є прогнозування з деталізацією окремих ділянок, що відкриває багато можливостей для майбутніх досліджень – оптимізація диференційованого внесення хімічних речовин, аналіз впливу локальних особливостей поля, таких як перепади висоти ділянок над рівнем моря та нерівномірність ґрунтової вологості на відмінності врожайності рослин, а також застосування цих знань для підвищення врожайності (напр., оптимізація поливного режиму, прогнозування появи шкідників і хвороб для своєчасного реагування тощо).

Практична цінність полягає в тому, що розроблена технологія дозволить виконувати прогнозування з високою точністю, що

дасть можливість аналізувати і завчасно реагувати на ризики, та ефективніше планувати бюджет. На її основі будуть проведені експерименти з метою визначення доцільності застосування хімічних оброблень. Крім цього, прогнозування приросту врожаю від застосування продуктів хімічних оброблень може бути використано як маркетинговий інструмент для підвищення продажів. Отже, побудована система сприятиме розвитку сільськогосподарського виробництва за рахунок підвищення якості інформаційно-аналітичних послуг і відкриття нових можливостей.

Об'єкт дослідження – інформаційно-аналітичний супровід прогнозування та процесів прийняття рішень у виробництві сільськогосподарської продукції.

Предмет дослідження – моделі, методи й інструментальні засоби прогнозування врожайності сільськогосподарських культур.

Мета дослідження – підвищення точності прогнозування й ефективності процесів прийняття рішень шляхом розроблення моделей, методів та інструментальних засобів оптимізації і прогнозування сільськогосподарського виробництва.

Дослідницькі задачі та підходи. Основні задачі спрямовано на підвищення точності прогнозування й експериментальну верифікацію ефективності обраних підходів:

- Розроблення методу виявлення хмар на супутникових знімках і відновлення даних, спотворених цими хмарами.
- Розроблення й аналіз ефективності підходу до прогнозування, заснованого на інтеграції моделей різного спрямування: прогнозування відносних значень із деталізацією та визначення коефіцієнтів для зведення прогнозованих значень до абсолютних на основі загальних показників поля.
- Застосування і налаштування методів автоматизованого вибору параметрів вхідних даних для зменшення обчислювальних витрат.
- Аналіз доцільності використання різних джерел інформації та параметрів опису стану, таких як висота ділянки над рівнем моря, координати ділянки, дата посіву тощо.
- Аналіз доцільності використання даних, що описують стан поля за попередні сезони.
- Застосування моделей для прогнозування на основі часових рядів, зокрема LSTM і трансформерів. Порівняння точності їхніх

прогнозів із точністю попередніх підходів на основі моделей LightGBM та U-Net.

- Аналіз векторного представлення навчених моделей і розв'язання задач: формування репрезентативної навчальної вибірки порівнянням векторних представлень її елементів з представленнями інших доступних даних і забезпечення її максимальної повноти, виявлення аномальних даних і видалення їх із навчальної вибірки.

Дискусія і висновки

Результатом роботи стане технологія розподіленого прогнозування врожайності. Перспективним застосуванням технології є підтримка прийняття рішень в сільськогосподарському виробництві:

- Визначення економічної доцільності застосування десикації. Прогнозування приросту врожайності й обчислення потенційного прибутку з урахуванням вартості хімікатів і витрат на їхнє застосування.

- Налаштування системи для прогнозування на ранніх етапах дозрівання рослин.

- Аналіз комбінацій значень параметрів у наборах даних із метою визначення ефективних умов для вирощення різних типів рослин.

Список використаних джерел

Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A.G., Madugundu, R., Alblewi, B., Assiri, F. (2016). *Prediction of Potato Crop Yield Using Precision Agriculture Techniques*. PLoS One.

Amankulova, K., Farmonov, N., Mukhtorov, U. & Mucsi, L. (n. d.) Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology. *Geocarto Int.*, 38, 1.

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., Kim, S. H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLoS One*, 11(6).

Khaki, S. & Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.*, 10, 621.

References

Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A.G., Madugundu, R., Alblewi, B., Assiri, F. (2016). *Prediction of Potato Crop Yield Using Precision Agriculture Techniques*. PLoS One.

Amankulova, K., Farmonov, N., Mukhtorov, U. & Mucsi, L. (n. d.) Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology. *Geocarto Int.*, 38, 1.

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., Kim, S. H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLoS One*, 11(6).

Khaki, S. & Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.*, 10, 621.

Отримано редакцією журналу / Received: 17.09.24

Прорецензовано / Revised: 27.09.24

Схвалено до друку / Accepted: 01.10.24

Vladyslav HNATIENKO, PhD Student

ORCID ID: 0009-0000-2678-5158

e-mail: hnatiienko.vladyslav@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Vitalii SNYTYUK, DSc (Engin.), Prof.

ORCID ID: 0000-0002-9954-8767

e-mail: snytyuk@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

CONCEPTUAL ELEMENTS OF THE TECHNOLOGY OF DISTRIBUTED FORECASTING OF AGRICULTURAL CROPS YIELD

The main objective of the research is to improve the accuracy of yield forecasting by developing data processing methods and neural network technologies. A yield forecasting technology will be developed, incorporating image recognition models for satellite image analysis and improving data quality in training sets, data processing methods, and deep neural networks in combination with other artificial intelligence models. Additionally, based on this technology, the efficiency and feasibility of applying certain agricultural practices will be studied, providing decision support in agricultural production.

Keywords: *agricultural production, site-specific forecasting, forecasting accuracy.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.