УДК 519.22:631.5:633.1
DOI: https://doi.org/10.17721/AIT.2025.2.06

Олексій ШОЛОХОВ, канд. фіз.-мат. наук
ORCID ID: 0000-0002-8676-3724
e-mail: gyroalex@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

Денис ПРОСЯНКІН, асп.
ORCID ID: 0009-0000-4402-6921
e-mail: dipruman@gmail.com
Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, Україна

МЕТОД ПОПЕРЕДНЬОГО ОБРОБЛЕННЯ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР У СИСТЕМІ ПРОГРАМУВАННЯ ВРОЖАЇВ

Вступ. ***Розроблено метод оброблення метеорологічних показників та агрономічних даних, у тому числі результатів польових дослідів. Метод призначений для застосування у системі програмування врожаїв, зокрема, для прогнозного моделювання та розроблення сценаріїв агрохімічних заходів в умовах кліматичних змін. Особливістю системи програмування є те, що вищеназваний метод дозволяє інтегрувати різні дані, отримані з різних джерел, і створювати набори інформативних показників, які відображають умови розвитку культур, що забезпечує більш повне подання стану агроекосистем і підвищує якість отримуваних прогнозів. Метою роботи є розроблення методу оброблення даних у системі програмування врожаїв, який забезпечує врахування як часових залежностей, так і ієрархічності структури даних, адаптований до використання у системі програмування врожаїв.***

Методи. ***У процесі дослідження використано методи аналізу та синтезу, системного аналізу, експертного оцінювання, багатокритеріального аналізу, статистичні, аналізу часових рядів, байєсівський підхід, машинне навчання.***

Результати. ***Розроблений метод забезпечує можливість адаптивного моделювання врожайності зернових культур, ураховуючи особливості певної агрокліматичної зони. Це дозволило врахувати як часові залежності, так і опрацювати структуру даних "рослина → ділянка → господарство → регіон". Ураховані лагові зв'язки між урожайністю поточного року та факторами попереднього періоду, що є особливо важливим під час використання багаторічних агрометеорологічних спостережень, оскільки дозволяє простежувати вплив кліматичних змін на певній території.***

Висновки. ***Ефективність розробленого методу перевірена на даних польових дослідів. Розроблені математичні моделі забезпечили покращення якості прогнозів урожайності зернових культур майже на 10 %. Використання методики інтелектуального аналізу даних дозволило більш точно визначити вплив метеорологічних чинників на розвиток рослин на різних етапах агроценозу. Побудовані сценарії урожайності сільськогосподарських культур забезпечують поглиблений аналіз факторів, які вливають на розвиток рослин на різних етапах вегетації, що підвищує якість прогнозів за різних варіантів впливу керованих факторів та некерованих умов і, відповідно, підвищення урожайності й біологізацію процесів виробництва зернових культур. Розроблені метод, моделі та сценарії призначені для інтеграції в систему програмування врожаю, але можуть бути застосовані й окремо. Їхнє використання створить умови для оптимізації агротехнологій, підвищивши ефективність виробництва зернових.***

Ключові слова: ***оброблення даних, байєсівські мережі, агрометеорологічні дані, інтеграція даних, машинне навчання, програмування урожайності, сценарії.***

Вступ

Традиційно Україна є одним зі світових лідерів у виробництві продукції сільського господарства та зернових зокрема. Саме рослинницька галузь, навіть в умовах сучасних викликів, демонструє сталий розвиток, забезпечує продовольчу безпеку країни та продовжує нарощувати експортний потенціал. Передумови стійкості галузі забезпечені потужним науково-технологічним потенціалом аграрної сфери, створеним завдяки багаторічним дослідженням учених за активної участі фахівців-практиків. Сьогодні, як особливо актуальні, визначені задачі забезпечення високих і сталих врожаїв сільськогосподарських культур в умовах деградації земель, зміни клімату, зростаючого забруднення довкілля, в тому числі й внаслідок військового конфлікту. Вирішенню цих задач присвячені роботи багатьох вітчизняних і закордонних фахівців, зокрема такі, у яких розглядаються питання удосконалення планування, прогнозування та програмування врожаїв. Програмування урожаїв – розроблення програми одержання запланованої урожайності сільськогосподарських культур через оптимізацію співвідношення регульованих і нерегульованих чинників – в умовах певної агротехнології за використання наявних ресурсів стає дедалі актуальнішим з огляду на необхідність забезпечення високої якості сільськогосподарської продукції, біологізацію рослинництва тощо. На сьогодні інформатизація та комп'ютеризація аграрної сфери більше орієнтовані на розв'язання задач моніторингу вирощування польових культур, і меншою мірою – на керування процесами та формування врожаю. Проблемою існуючих розробок є те, що вони орієнтовані на дослідження поточного стану посівів, а не на отримання оптимального варіанта технології, за якої реальна господарська урожайність наблизиться до дійсно можливої урожайності, а дійсно можлива урожайність – до потенційної. Для забезпечення реалізації цієї задачі потрібні підходи, які передбачають комплексне використання технології інтелектуального аналізу даних, апарату математичного моделювання, штучного інтелекту для урахування ґрунтово-кліматичних факторів, диференціації агротехнічних і меліоративних прийомів, оптимального використання матеріальних, технічних та трудових ресурсів господарства з метою запланування врожайності й забезпечення рівня її отримання. Усі вони поєднують досягнення аграрної науки з комп'ютерним моделюванням та сучасними інформаційно-комунікаційними технологіями. Однак повністю адаптованих до використання існуючих наборів сільськогосподарських даних, у тому числі результатів польових дослідів, на сьогодні немає. Інформаційні технології такого типу на даний момент практично відсутні. Мало уваги приділено й розробленню аналітичного

© Шолохов Олексій, Просянкін Денис, 2025

ISSN 2788-6603

апарату для дослідження оптимального кількісного співвідношення регульованих факторів і не регульованих умов, які в системі технологічного процесу забезпечать одержання запланованої урожайності при найбільш економному витрачанні наявних ресурсів та за раціонального природокористування.

Тому тема цього дослідження, присвяченого розв'язанню задач розроблення інструментарію для попередньої обробки даних, їхнього аналізу, підготовки до побудови прогнозів, сценаріїв, планів у системі програмування урожайності, є актуальною та має практичну цінність.

Використання розробки сприятиме підвищенню якості управлінських рішень у сільськогосподарському виробництві, а саме щодо підвищення продуктивності агроценозів, екологізації рослинництва, переведення його на контрольовану основу.

Аналіз останніх досліджень і публікацій

Забезпечення ефективності аграрного сектору в контексті кліматичних змін, погіршення екологічної ситуації, раціонального використання ресурсного потенціалу, зменшення дефіциту продовольства є однією з найактуальніших як вітчизняних, так і світових проблем. Одним зі шляхів її вирішення є підвищення продуктивності рослинництва за рахунок програмування врожаїв сільськогосподарських культур. Програмування врожаїв передбачає розроблення програми раціонального використання чинників формування урожаю під час вегетації з урахуванням оптимального кількісного співвідношення регульованих чинників та нерегульованих умов, які в системі технологічного процесу забезпечують одержання запланованої урожайності за раціонального використання ресурсів.

Jabed, M. A. та Azmi Murad, M. A. (Jabed, & Azmi Murad, 2024), підкреслюючи важливість проблеми інформаційно-аналітичного забезпечення сільськогосподарського виробництва, зазначають її недостатню опрацьованість. На їхню думку, розробка точних та інтерпретованих моделей прогнозування виробництва сільськогосподарських культур залишається складною задачею через багатофакторні та різноспрямовані впливи. Різні методи, включаючи польові дослідження, статистичні моделі, моделі росту сільськогосподарських культур і дистанційне зондування вже традиційно використовуються для прогнозування, планування та програмування врожайності сільськогосподарських культур. Ці методи потребують великих наборів даних, оскільки необхідно врахувати й особливості виробничого циклу в сільському господарстві, і складні залежності факторів сільськогосподарського виробництва. Проблемою існуючих розробок є те, що вони орієнтовані на дослідження поточного стану посівів, а не на отримання оптимального варіанта технології, за якої реальна господарська урожайність наблизиться до дійсно можливої, а дійсно можлива урожайність – до потенційної. Вони зосереджені більше на вирішенні задач моніторингу вирощування польових культур, і меншою мірою – на коригуванні процесів, формуванні врожаю та якості сільськогосподарської продукції. Мало уваги приділено й використанню інтелектуального аналізу даних, математичних моделей для дослідження оптимального кількісного співвідношення регульованих факторів і не регульованих умов, які в системі технологічного процесу забезпечать одержання запланованої урожайності при найбільш економному витрачанні наявних ресурсів.

Для реалізації даної задачі потрібні методики, які забезпечують оброблення агроекологічних даних, результатів польових дослідів, даних дистанційного зондування тощо. Тобто такі, які б дозволяли врахувати вплив ґрунтово-кліматичних факторів, диференціації агротехнічних і меліоративних прийомів, оптимального використання матеріальних, технічних та трудових ресурсів господарства з метою запланування врожайності та забезпечення рівня її отримання.

Вирішенню вищенаведених проблем присвячені роботи багатьох вітчизняних і закордонних фахівців. Зокрема, у роботах (Kussul et al., 2025; Jabed, & Azmi Murad, 2024; Ashfaq et al., 2025; Shahhosseini et al., 2021; Wang et al., 2018; Yan et al., 2025) представлено комплексний огляд підходів машинного та глибокого навчання, сформовано переваги й недоліки вказаних підходів, підкреслено перспективність їхнього застосування у задачах прогнозування урожайності сільськогосподарських культур.

Інтеграція взаємодій генотип – середовище – менеджмент (GxExM) у прогностичні моделі зазначена як пріоритетний напрям сучасної агрономічної науки (Cooper et al., 2021; Mahmood, Ahmed, & Trethowan, 2022). Адже розуміння цих складних взаємодій і взаємовпливів критично важливе для оптимізації сільськогосподарського виробництва та скорочення розриву між потенційною й фактичною врожайністю.

Складність вирішення задачі програмування урожайності сільськогосподарських культур стимулює розвиток індустрії інформаційно-аналітичного забезпечення сільськогосподарського виробництва: від електронних таблиць до потужних спеціалізованих аналітичних платформ.

Потужні системи управління аграрним виробництвом, такі як EOS Data Analytics і Cropwise Operations, не зважаючи на значний прогрес у моніторингу посівів, залишаються обмеженими в можливостях програмування та прогнозування урожайності через відсутність адекватних математичних моделей, здатних описувати складну, нелінійну та нестаціонарну природу біологічних процесів у рослинництві (Бідюк та ін., 2017).

Як зауважують (Shahhosseini et al., 2021; Corbeels et al., 2018), є необхідність переходу до гібридних архітектур, що поєднують фізичну достовірність процесно-орієнтованих моделей (WOFOST, DSSAT, APSIM) із гнучкістю машинного навчання.

Однак багаторічні агрометеорологічні спостереження та результати польових дослідів поки залишаються поза увагою при розробленні прогнозних моделей (Біляєва, 2015), що є критичним у контексті зростаючої кліматичної мінливості та посилення її впливу на продуктивність сільськогосподарських культур (Zhao et al., 2017), а також погіршення екологічної ситуації.

Формуючи систему програмування урожайності важливо враховувати особливості вхідних даних. Адже агрономічні дані "прив'язані" до місця знаходження ділянок, на яких вирощуються певні культури. Тому мають ієрархічну структуру (рослина → ділянка → господарство → регіон). Для урахування такої ієрархічності у системі оброблення даних доцільно використовувати ієрархічний байєсівський висновок (Jensen, 2001). Перевагами такого підходу є те, що він дозволяє враховувати варіативність на різних рівнях, захоплюючи GxExM ефекти. Як зазначають (Valleggi, Scutari, & Stefanini, 2023; Salakpi et al., 2022; Rivière et al., 2015), використання ієрархічних байєсівських мереж є перспективним інструментом аналізу даних у системі програмування урожайності. Саме байєсівські підходи (Spirtes, Glymour,

& Scheines, 1993) продемонстрували свою ефективність у моделюванні складних сільськогосподарських систем з урахуванням невизначеності різних типів (Cho et al., 2023; Kocian et al., 2020). Крім того, використання ієрархічних байєсівських моделей дозволяє ефективно інтегрувати дані з різних рівнів їх організації та враховувати просторову неоднорідність в умовах обмеженої інформації (Salakpi et al., 2022; Rivière et al., 2015; Cho et al., 2023). Це особливо актуально для України, де накопичені довготривалі агрометеорологічні дані з різних регіонів (Kussul, Drozd, & Yailymova, 2023; Концеба та ін., 2021; Kussul et al., 2025; Бідюк та ін., 2017), результати польових дослідів, моніторингу посівів у фермерських господарствах, які можуть бути використані для побудови прогностичних моделей. Але ці дані належать різним розпорядникам і на сьогодні використовуються досить обмежено.

Тому метою статті є опис розробленої методики попереднього оброблення агрометеорологічних даних, даних урожайності та результатів польових дослідів, призначеної для використання у системі програмування урожайності зернових культур.

Для досягнення мети були такі задачі:

- проаналізовано наявні інформаційні технології оброблення даних у системах програмування урожайності сільськогосподарських культур;
- досліджено особливості застосування інтелектуального аналізу даних, математичного моделювання, штучного інтелекту для оброблення даних для використання їх для прогнозування врожаїв;
- розроблено методику виявлення найбільш значущих нерегульованих і регульованих чинників для прогнозування урожайності сільськогосподарських культур у системі програмування врожаїв в умовах неповних різнорідних даних.

Інформаційну основу дослідження становлять часові ряди агрометеорологічних даних (за 1965–2005 рр.), дані власних польових дослідів та часові ряди урожайності сільськогосподарських культур, накопичені дослідниками кафедри агрохімії та землеробства Уманського національного університету за період 1965–2005 рр.

Методи

Методологічна основа дослідження спрямована на формування комплексного підходу до попереднього оброблення даних, підготовки їх до подальшого використання у системі програмування урожайності зернових культур. Запропонована методика поєднує статистичні, машинні та ймовірнісно-графічні підходи, що дозволяє враховувати просторову неоднорідність, часові лаги й невизначеність у даних. Основну увагу приділено етапам інтеграції різнорідних даних, отриманих з різних джерел інформації (Data Fusion), формуванню похідних ознак (Feature Engineering) тощо.

Пропоновану методику можна подати у вигляді послідовності кроків (табл. 1).

Таблиця 1

Методика інтеграції агрометеорологічних даних

Крок	Дії
1	Створення набору метеорологічних даних (meteo_df)
2	Агрегування даних за роками (year) та місяцями (imonth)
3	Розрахунок похідних показників: GDD (сумарні активні температури) і SMD (дефіцит вологи)
4	STL-декомпозицію NDVI для вилучення сезонної складової
5	Нормалізація основних ознак (z-score)
6	Виведення перших 12 записів для перевірки

Метеорологічні показники (розглянуто дані за період 1965–2005 рр.): значення середньомісячних температур (°C), кількість опадів (мм), відносна вологість повітря (%) отримані із системи моніторингу досліджуваної агроекосистеми. Агрономічні дані (1965–2005 рр.): урожайність озимої пшениці, ячменю, кукурудзи (ц/га) за 10 варіантів удобрення та без добрив (добрива $N_{45}P_{45}K_{45}$, $N_{90}P_{90}K_{90}$, $N_{135}P_{135}K_{135}$; гній 4,5/9/13,5 т/га з додатковим внесенням NPK). Обсяг вибірки: 40 років × 10 варіантів = 400 спостережень на культуру (результати польових дослідів)..

На етапі попередньої підготовки дані було агреговано за вегетаційними періодами культур, зокрема для озимих: березень – липень; для ярих – квітень – серпень. Виділено ключові метеофактори: T_{sum} – сума середньодобових температур; P_{May} – опади у травні; RH_{avg} – середня вологість повітря. Застосовано алгоритми заповнення пропущених даних та Isolation Forest для видалення викидів (Trofymchuk et al., 2019).

Для покращення репрезентативності факторів побудовано часові ряди найбільш значущих чинників (derived features), що описують динаміку агрометеорологічних процесів:

1. Сума ефективних температур Growing Degree Days (GDD):

$$GDD = \sum_{i=1}^{n} \max(T_{mean,i} - T_{base}, 0),$$

де $T_{mean,i}$ – середньодобова температура повітря в день i (°C), T_{base} – базова (порогова) температура росту: для озимих культур (пшениця, жито) – 10° C; кукурудзи – 10° C; рису – 8° C; для соняшнику – 6° C, n – кількість днів вегетаційного періоду, $\max(x, 0)$ – функція, що повертає максимальне значення між x та 0.

Цей показник характеризує сумарну теплову енергію, необхідну для проходження фенологічних фаз.

ISSN 2788-6603

2. Дефіцит вологи в ґрунті – Soil Moisture Deficit (SMD):

$$SMD = P - ET_p,$$

де P – кількість опадів (мм), ET_p – потенційна евапотранспірація (розрахована за формулою Пенмана-Монтейта) (мм).

Показник відображає водний баланс ґрунту. Негативне значення (-) вказує на дефіцит вологи (посуха), позитивне значення (+) – на надлишок вологи, нульове значення (0) – на збалансований водний режим.

3. Індекс стабільності погодних умов – Weather Stability Index (WSI):

$$WSI = \frac{\sigma_T}{\bar{T}} \times 100\%,$$

де σ_T – стандартне відхилення температури (°C), $\bar{T}$ – середня температура за період (°C).

Показник характеризує варіабельність (мінливість) температурного режиму упродовж вегетаційного періоду. Відображає ступінь коливань температури. Низькі значення WSI – стабільні температурні умови, високі значення WSI – різкі коливання температур, що можуть викликати стрес у рослин.

4. Коефіцієнт кумулятивних опадів – Cumulative Precipitation Ratio (CPR):

$$CPR = \frac{P_{first_half}}{P_{total}},$$

де P_{first_half} – сума опадів за першу половину вегетаційного періоду (мм), P_{total} – загальна сума опадів за весь вегетаційний період (мм).

Цей показник відображає відношення опадів за першу половину вегетаційного сезону до загальних опадів – індикатор своєчасності вологозабезпечення у критичні фази розвитку: якщо CPR > 0,6 – опади зміщені на ранній період (сприятливо), 0,4 ≤ CPR ≤ 0,6 – рівномірний розподіл опадів, CPR< 0,4 – опади зміщені на пізній період (ризик для врожайності)

Обробка сезонності й нормалізація виконувалась таким чином. Для усунення міжрічної та міжрегіональної варіабельності здійснено сезонне згладжування часових рядів методом STL-декомпозиції (Seasonal-Trend decomposition using Loess), що дозволяє виділити трендову, сезонну й залишкову компоненти.

Після усунення сезонної складової проведено нормалізацію за z-score:

$$X' = \frac{X - \mu_X}{\sigma_X},$$

де μ_X та σ_X – середнє значення та стандартне відхилення відповідної ознаки за весь період спостережень.

Таке перетворення забезпечує зіставність різних показників.

У пропонованій методиці (табл. 1) використано підхід, який поєднує Data Fusion та Feature Engineering, що дозволяє інтегрувати різнорідні дані, отримані з різних джерел, і створювати набори інформативних показників, які відображають умови розвитку культур. Перевагою такої інтеграції є забезпечення більш повного представлення стану агроекосистем.

Проведене оброблення сезонності та нормалізація даних дозволили суттєво підвищити якість вхідних ознак, використовуваних для моделювання урожайності зернових культур. Застосовування STL-декомпозиції забезпечило вилучення сезонних коливань і виділення стійкої трендової компоненти, що є критично важливим для аналізу довгих часових рядів. Використання z-score нормалізації уніфікувало масштаби показників, зробивши їх порівнюваними між різними роками та регіонами. У результаті отримано стабільні та збалансовані дані, придатні для побудови моделей і прогнозів.

Виходячи з особливостей оброблюваних наборів даних, інструментом для дослідження причинно-наслідкових зв'язків та побудови прогнозів обрано байєсівські моделі.

Байєсівські моделі, зокрема мережі Байєса, як зазначають П.І. Бидюк та ін. (Бідюк та ін., 2017), добре враховують особливості агрометеорологічних даних, що забезпечує підвищення якості прогнозування.

Загальне рівняння байєсівської моделі урожайності має вигляд:

$$P(Y_t|Y_{t-1}, X_t) = \sum_i P(Y_t|X_t, Z_i) \cdot P(Z_i|Y_{t-1}),$$

де, Y_t – урожайність у році t, Y_{t-1} – урожайність попереднього року, X_t – вектор агрометеорологічних факторів (опади, температура, вологість тощо), Z_i – приховані (латентні) стани, що відображають стійкість агроекосистеми.

Спочатку було побудовано ієрархічну байєсівську модель (Hierarchical Bayesian Model), а для моделювання часових залежностей (між роками) було розширено її структуру до динамічної байєсівської моделі (Dynamic Bayesian Model, (Kocian et al., 2020)).

Для моделювання методами класифікації Y – цільова змінна врожайності – спочатку була перетворена в ординарну змінну, що має три стани: низька врожайність ($Y = 1$), середня врожайність ($Y = 2$), висока врожайність ($Y = 3$).

Оскільки цільова змінна має три категорії, то була використана мультиноміальна логістична регресія, при чому клас "низька врожайність" був обраний як базовий. У загальному ця модель має такий вигляд:

$$\log\frac{P(Y = k|x)}{P(Y = 1|x)} = \beta_{k0} + \beta_{kT} \cdot T + \beta_{kR} \cdot R + \beta_{kH} \cdot H + \sum_i^m \beta_{kF_i} \cdot F_i,$$

де β_{k0} – вільний член моделі, T – середньомісячна температура, R – кількість опадів, H – відносна вологість повітря, F_i – даммі-змінні типу добрива, при чому i змінюється від 1 до m (кількість видів добрив).

На рис. 1 наведено спрощену топологію побудованої мережі Байєса у програмному забезпеченні Genie 2.0 (www.bayesfusion.com).

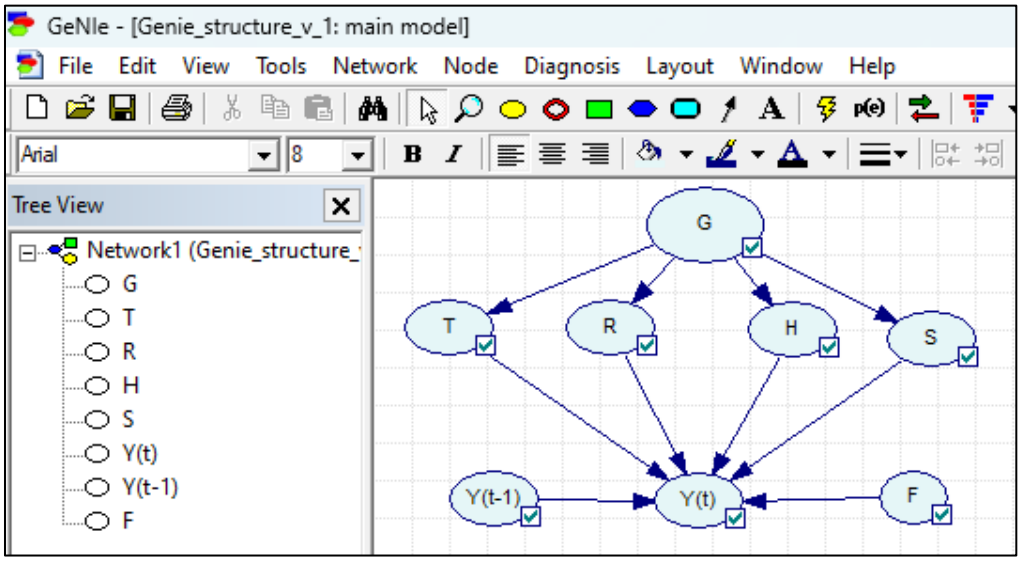

Рис. 1. Структура побудованої мережі Байєса

Для побудови мережі застосовано додаткові фактори аналізу:

- інерційний: $Y(t-1)$ – врожайність сільськогосподарської культури у попередньому році;
- агроекологічні фактори: S – якість ґрунту (приймає значення – низька, середня, висока), F – тип добрива;
- кліматичні зміни: T, R, H (середньомісячна температура, кількість опадів, відносна вологість повітря);
- просторовий чинник: G – просторовий фактор (приймає значення агрокліматичної зони).

Дискретизація змінних аналізу здійснювалася за такими правилами:

- температура: низька, якщо $T < 18$ C, нормальна – при $18 \leq T \leq 25$ C, висока – при $T > 25$ C;
- опади: малі, якщо $R < 50$ мм, середні – при $50 \leq R \leq 100$ мм, великі – при $R > 100$ мм;
- вологість: низька, якщо $H < 50$ %, норма – при $50\ \% \leq H \leq 70$ %, висока – при $H > 70$ %;
- якість ґрунту: низька, якщо гумус < 2 %, середня – при $2 - 4$ %, висока > 4 %.

Параметри часових залежностей оцінювались за Expectation-Maximization (EM) алгоритмом для часових графічних моделей із регуляризацією Лапласа. Це дало змогу врахувати інерційні ефекти клімату (напр., вплив накопиченої вологи або дефіциту опадів попереднього сезону на урожайність поточного року).

Використання динамічної структури дозволило підвищити стійкість прогнозів за різких кліматичних змін; зменшити середньоквадратичну похибку на 6–8 % порівняно з ієрархічною байєсівською мережею; відтворити причинно-часові зв'язки між параметрами (приміром, опади, у момент часу $t-1 \rightarrow$ температура, у момент часу $t \rightarrow$ урожайність, у момент часу t).

Відповідно до пропонованої методики послідовно здійснюється підготовка середовища, формування синтетичних даних з часовими лагами та визначення структури моделі. Далі дані перетворюються у формат MultiIndex, після чого модель "навчається" методом Maximum Likelihood Estimation. Завершальним кроком є аналіз отриманих параметрів умовних розподілів і виконання запиту для прогнозування урожайності поточного року на основі показників попереднього періоду. У табл. 2 подано методику навчання моделі.

Таблиця 2

Методика навчання моделі

Номер етапу	Опис
1	Встановлення бібліотеки Python pgmpy
2	Імпорт модулів
3	Створення еталонного датафрейму (10 років, з лагами t-1)
4	Визначення структури динамічної байєсівської мережі
5	Перетворення даних у формат MultiIndex (для pgmpy)
6	Навчання моделі методом Maximum Likelihood
7	Виведення параметрів умовних розподілів (CPDs)
8	Виконання запиту. Наприклад, запит (Y_t) прогнозу урожайності сільськогосподарської культури Y у рік t за заданих опадів і температурі попереднього року

ISSN 2788-6603

Розширення моделі до динамічної байєсівської мережі (DBN) підтвердило доцільність урахування часових лагів у прогнозуванні урожайності зернових культур. Отримані результати демонструють, що включення факторів попереднього року дозволяє краще відтворювати інерційні ефекти клімату та стану ґрунту, зокрема накопичення вологи чи дефіцит опадів. Застосування динамічної мережі Байєса забезпечило підвищення точності прогнозу – зменшення середньоквадратичної похибки на 6–8 % порівняно з базовою ієрархічною мережею Байєса. Крім того, модель відтворює причинно-часові взаємозв'язки між агрометеорологічними показниками, що робить її придатною для інтеграції у системи підтримки аграрних рішень і довгострокового моніторингу кліматичних ризиків.

Результати

Для досягнення поставленої мети, використовуючи пропоновану методику попереднього оброблення даних, розроблено адекватні моделі з помилкою класифікації MISC 18–22 %, що перевершує традиційні підходи.

Дані включали метеорологічні показники (температура, опади, вологість) та агрономічні дані (урожайність озимої пшениці, ячменю, кукурудзи за 10 варіантів удобрення) за 1965–2005 рр. Структура мережі Байєса побудована PC-алгоритмом, параметри оцінено методом максимальної правдоподібності зі згладжуванням Лапласа. Модель ураховує ієрархічну структуру даних (рослина → ділянка → господарство → регіон).

У результаті застосування побудованої мережі Байєса виявлено такі ключові зв'язки для озимої пшениці: $P_{May} \rightarrow Y$: сильний позитивний вплив ($\beta \approx +0{,}42$); $T_{sum} \rightarrow Y$: нелінійний зв'язок з оптимумом при 210–230°C; $N_{90} \rightarrow Y$: ефект насичення.

Важливим є те, що вплив опадів у травні на урожайність варіюється залежно від регіону (різниця до 35 %), що не можна виявити звичайною мережею Байєса. Помилка класифікації становить 19,3 %; а MSE – 0,032.

На основі моделі побудовано сценарії (табл. 3):

Таблиця 3

Сценарії впливу агрометеорологічних факторів на урожайність

Сценарій	ΔP_{May}	T_{sum}, C °	ΔY, %	Регіон
Базовий	0	220	0	Усі
Оптимізований	+15	225	+13,5	Північ
Оптимізований	+20	222	+16,2	Південь
Стресовий	-20	250	-28,1	Усі

Сценарний аналіз демонструє значущість взаємодій GxExM, які визначають варіабельність відповіді культур на агрометеорологічні умови (Cooper et al., 2021; Mahmood, Ahmed, & Trethowan, 2022).

Проведено порівняння з іншими методами (табл. 4).

Таблиця 4

Порівняння результатів прогнозування урожайності із застосуванням різних методів

Метод	RMSE	MAE	Помилка класифікації, %	Переваги	Недоліки
мережа Байєса	0,042	0,031	23,1	Простота, інтерпретованість	Не враховує ієрархію
Logistic Regression	0,052	0,037	22,3	Швидкість	Не враховує ієрархію
Random Forest	0,035	0,025	18,7	Висока точність	"Чорна скринька"
Запропонований метод	**0,032**	**0,023**	**19,3**	**Фізична достовірність, ієрархія**	**Складне навчання**

Результати підтверджують переваги ієрархічного байєсівського підходу, особливо для роботи з багаторівневими даними та врахування просторової неоднорідності (Rivière et al., 2015; Cho et al., 2023). Порівняння з методами машинного навчання показує конкурентоспроможність запропонованого підходу, з перевагою в інтерпретованості результатів та можливості інтеграції експертних знань.

Виявлено ключові фактори, що впливають на урожайність озимої пшениці: опади у травні для озимої пшениці ($\beta \approx +0{,}42$), температурна сума з оптимумом 210–230° C, ефект насичення азотного живлення. Побудовано оптимальні сценарії: збільшення опадів у травні на 15 % за температури червня +18° C сприяє зростанню урожайності озимої пшениці на 12–14 %, але оптимальні значення різняться за регіонами (Північ: +13,5 %, Південь: +16,2 %).

Дискусія і висновки

В роботі запропоновано метод прогнозування урожайності зернових культур на основі ієрархічних байєсівських мереж, що дозволяє враховувати особливості структури бази даних польових дослідів і метеорологічні дані та складні GxExM-ефекти. Побудовано адекватні моделі з помилкою класифікації 18–22 % та MSE 0,031–0,038, що перевершує традиційні підходи.

Виявлено ключові фактори що впливають на підвищення урожайності сільськогосподарських культур, а саме – опади у травні – для озимої пшениці, внесення азотних добрив – для кукурудзи (з урахуванням особливостей агрокліматичної зони). Найбільший приріст урожайності досягається за комбінації помірних опадів у травні та помірної температури в червні, але оптимальні значення різняться за регіонами.

Подібні результати отримані в дослідженнях із застосуванням глибинного навчання та трансферного навчання для прогнозування врожайності (Wang et al., 2018; Yan et al., 2025), проте байєсівський підхід забезпечує краще моделювання невизначеності та інтерпретацію причинно-наслідкових зв'язків.

Запропонований метод може бути інтегрованим в інформаційні технології програмування урожаю й системи підтримки прийняття рішень у сільському господарстві, забезпечуючи не лише прогнозування, а й оптимізацію агротехнологій з метою підвищення ефективності виробництва.

Перспективи подальших досліджень містять розширення бази даних до інших культур і регіонів, інтеграцію супутникових даних та розробку мобільних додатків для практичного використання в господарствах.

Внесок авторів: Олексій Шолохов – написання – перегляд і редагування; Денис Просянкін – методологія, концептуалізація, формальний аналіз, програмне забезпечення, валідація даних, написання – оригінальна чернетка.

Джерела фінансування. Це дослідження не отримало жодного гранту від фінансової установи в державному, комерційному або некомерційному секторах.

Використання інструментів штучного інтелекту. Автори заявляють, що під час підготовки цього рукопису, проведення аналізу даних та написання тексту не використовувалися інструменти генеративного штучного інтелекту або інші ШІ-технології. Робота є повністю оригінальним продуктом інтелектуальної діяльності авторів.

Список використаних джерел

Бідюк, П. І., Терентьєв, О. М., Просянкіна-Жарова, Т. І., & Ефендієв, В. В. (2017). Прогнозне моделювання нелінійних нестаціонарних процесів у рослинництві з використанням інструментів SAS Enterprise Miner. *Наукові вісті НТУУ КПІ*, 1, 24–36. https://doi.org/10.20535/1810-0546.2017.1.87423

Біляєва, І.М. (2015) Наукове обґрунтування та практичне використання агрометеорологічних методів прогнозування врожайності польових культур в умовах зрошення. *Таврійський науковий вісник*, 94, 8–15. https://www.tnv-agro.ksauniv.ks.ua/archives/94_2015/4.pdf

Концеба, С., Ліщук, Р., Скуртол, С., Родащук, Г., & Васильченко, І. (2021). Прогнозування обсягів виробництва сільськогосподарської продукції за допомогою методів інтелектуального аналізу даних. *Системні дослідження та інформаційні технології*, 1, 87–97. https://doi.org/10.20535/SRIT.2308-8893.2021.1.07

Ashfaq, M., Khan, I., Afzal, R. F., Shah, D., Ali, S., & Tahir, M. (2025). Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques. *Scientific Reports*, 15*, 18093*. https://doi.org/10.1038/s41598-025-02700-w

Cho, W., Lambert, D. M., Brorsen, B. W., Maples, C. H., Fornah, A., & Raun, W. R. (2023). A Bayesian approach for analyzing crop yield response data with limited treatments. *Agrosystems, Geosciences & Environment*, 6*, e20358*. https://doi.org/10.1002/agg2.20358

Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang T., & Hammer, G. L. (2021). Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *Theoretical and Applied Genetics*, 134, 1625–1644. https://doi.org/10.1007/s00122-021-03812-3

Corbeels, M., Berre, D., Rusinamhodzi, L., & Lopez-Ridaura, S. (2018). Can we use crop modelling for identifying climate change adaptation options? *Agricultural and Forest Meteorology*, 256*,* 46–52. https://doi.org/10.1016/j.agrformet.2018.02.026

Jabed, M. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10, e40836. https://doi.org/10.1016/j.heliyon.2024.e40836

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer. https://doi.org/10.1007/978-1-4757-3502-4

Kocian, A., Massa, D., Cannazzaro, S., Incrocci, L., Di Lonardo, S., Milazzo, P., & Chessa, S. (2020). Dynamic Bayesian network for crop growth prediction in greenhouses. *Computers and Electronics in Agriculture*, 169, 105167. https://doi.org/10.1016/j.compag.2019.105167

Kussul, N., Drozd, S., & Yailymova, H. (2023). Forecast of yield of major crops in Ukraine in war conditions 2022 based on MODIS and Sentinel-2 satellite data. In *11th International Conference on Advanced Information and Communication Technology (ICAIIT)* (pp. 89–95). IEEE. http://dx.doi.org/10.25673/101923

Kussul, N., Shelestov, A., Yailymov, B., Yailymova, H., Lemoine, G., & Deininger, K. (2025). Assessment of war-induced agricultural land use changes in Ukraine using machine learning applied to Sentinel satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 140*,* 104551. https://doi.org/10.1016/j.jag.2025.104551

Mahmood, T., Ahmed, T., & Trethowan, R. (2022). Genotype x Environment x Management (GEM) reciprocity and crop productivity. *Frontiers in Agronomy*, *4, 800365*. https://doi.org/10.3389/fagro.2022.800365

Rivière, P., Dawson, J. C., Goldringer, I., & David, O. (2015). Hierarchical Bayesian modeling for flexible experiments in decentralized participatory plant breeding. *Crop Science*, *55,* 1053–1067. https://doi.org/10.2135/cropsci2014.07.0497

Salakpi, E. E., Hurley, P. D., Muthoka, J. M., Bowell, A., Oliver, S., & Rowhani, P. (2022). A dynamic hierarchical Bayesian approach for forecasting vegetation condition. *Natural Hazards and Earth System Sciences*, 22*,* 2725–2749. https://doi.org/10.5194/nhess-22-2725-2022

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11, 1606. https://doi.org/10.1038/s41598-020-80820-1

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction and Search*. Springer. https://doi.org/10.1007/978-1-4612-2748-9

Trofymchuk, O., Bidiuk, P., Prosyankina-Zharova, T., & Terentiev, O. (2019). *Decision support systems for modelling, forecasting and risk estimation*. Beau Bassin: LAP Lambert Academic Publishing.

Valleggi, L., Scutari, M., & Stefanini, F. M. (2023). *Learning Bayesian networks with heterogeneous agronomic data sets via mixed-effect models and hierarchical clustering.* arXiv. https://doi.org/10.48550/arXiv.2308.06399

Wang, A. X., Tran, C., Desai, N., Lobell, D., & Ermon, S. (2018). Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18) (Article 50, pp. 1–5). Association for Computing Machinery.* https://doi.org/10.1145/3209811.3212707

Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). *Crop yield time-series data prediction based on multiple hybrid machine learning models*. arXiv. https://doi.org/10.48550/arXiv.2502.10405

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., … & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35), 9326–9331. https://doi.org/10.1073/pnas.1701762114

References

Ashfaq, M., Khan, I., Afzal, R. F., Shah, D., Ali, S., & Tahir, M. (2025). Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques. *Scientific Reports*, 15*,* 18093. https://doi.org/10.1038/s41598-025-02700-w

Bidjuk, P. I., Terent'jev, O. M., Prosjankina-Zharova, T. I., & Efendijev, V. V. (2017). Prognostic modeling of nonlinear non-stationary processes in crop production using SAS Enterprise Miner tools. *Scientific News of NTUU KPI*, 1, 24–36 [in Ukrainian]. https://doi.org/10.20535/1810-0546.2017.1.87423

Bilyaeva, I.M. (2015). Scientific justification and practical use of agrometeorological methods for predicting the yield of field crops under irrigation conditions. *Taurida Scientific Herald*, 94, 8–15. https://www.tnv-agro.ksauniv.ks.ua/archives/94_2015/4.pdf

Cho, W., Lambert, D. M., Brorsen, B. W., Maples, C. H., Fornah, A., & Raun, W. R. (2023). A Bayesian approach for analyzing crop yield response data with limited treatments. *Agrosystems, Geosciences & Environment*, 6*, e20358*. https://doi.org/10.1002/agg2.20358

Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang T., & Hammer, G. L. (2021). Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity *Theoretical and Applied Genetics*, 134, 1625–1644. https://doi.org/10.1007/s00122-021-03812-3

ISSN 2788-6603

Corbeels, M., Berre, D., Rusinamhodzi, L., & Lopez-Ridaura, S. (2018). Can we use crop modelling for identifying climate change adaptation options? *Agricultural and Forest Meteorology*, 256, 46–52. https://doi.org/10.1016/j.agrformet.2018.02.026

Jabed, M. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10, e40836. https://doi.org/10.1016/j.heliyon.2024.e40836

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer. https://doi.org/10.1007/978-1-4757-3502-4

Kocian, A., Massa, D., Cannazzaro, S., Incrocci, L., Di Lonardo, S., Milazzo, P., & Chessa, S. (2020). Dynamic Bayesian network for crop growth prediction in greenhouses. *Computers and Electronics in Agriculture*, 169, 105167. https://doi.org/10.1016/j.compag.2019.105167

Kontseba, S., Lishchuk, R., Skurtol, S., Rodashchuk, H., & Vasylchenko, I. (2021). Forecasting of agricultural production volumes using methods of data mining. *Systemni doslidzhennia ta informatsiini tekhnolohii*, 1, 87–97 [in Ukrainian]. https://doi.org/10.20535/SRIT.2308-8893.2021.1.07

Kussul, N., Drozd, S., & Yailymova, H. (2023). Forecast of yield of major crops in Ukraine in war conditions 2022 based on MODIS and Sentinel-2 satellite data. In *11th International Conference on Advanced Information and Communication Technology (ICAIIT)* (pp. 89–95). IEEE. http://dx.doi.org/10.25673/101923

Kussul, N., Shelestov, A., Yailymov, B., Yailymova, H., Lemoine, G., Deininger, K. (2025). Assessment of war-induced agricultural land use changes in Ukraine using machine learning applied to Sentinel satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 140, 104551. https://doi.org/10.1016/j.jag.2025.104551

Mahmood, T., Ahmed, T., & Trethowan, R. (2022). Genotype x Environment x Management (GEM) reciprocity and crop productivity. *Frontiers in Agronomy*, 4, 800365. https://doi.org/10.3389/fagro.2022.800365

Rivière, P., Dawson, J. C., Goldringer, I., & David, O. (2015). Hierarchical Bayesian modeling for flexible experiments in decentralized participatory plant breeding. *Crop Science*, *55,* 1053–1067. https://doi.org/10.2135/cropsci2014.07.0497

Salakpi, E. E., Hurley, P. D., Muthoka, J. M., Bowell, A., Oliver, S., & Rowhani, P. (2022). A dynamic hierarchical Bayesian approach for forecasting vegetation condition. *Natural Hazards and Earth System Sciences*, 22, 2725–2749. https://doi.org/10.5194/nhess-22-2725-2022

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11, *1606*. https://doi.org/10.1038/s41598-020-80820-1

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction and Search*. Springer. https://doi.org/10.1007/978-1-4612-2748-9

Trofymchuk, O., Bidiuk, P., Prosyankina-Zharova, T., & Terentiev, O. (2019). *Decision support systems for modelling, forecasting and risk estimation*. Beau Bassin: LAP Lambert Academic Publishing.

Valleggi, L., Scutari, M., & Stefanini, F. M. (2023). *Learning Bayesian networks with heterogeneous agronomic data sets via mixed-effect models and hierarchical clustering.* arXiv. https://doi.org/10.48550/arXiv.2308.06399

Wang, A. X., Tran, C., Desai, N., Lobell, D., & Ermon, S. (2018). Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18) (Article 50, pp. 1–5). Association for Computing Machinery.* https://doi.org/10.1145/3209811.3212707

Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). *Crop yield time-series data prediction based on multiple hybrid machine learning models*. arXiv. https://doi.org/10.48550/arXiv.2502.10405

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., … & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35), 9326–9331. https://doi.org/10.1073/pnas.1701762114

Отримано редакцією журналу / Received: 10.11.25
Прорецензовано / Revised: 18.12.25
Схвалено до друку / Accepted: 29.12.25
Опубліковано / Published: 06.05.26

Oleksiy SHOLOKHOV, PhD (Phys. & Math.)
ORCID ID: 0000-0002-8676-3724
e-mail: gyroalex@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Denis PROSYANKIN, PhD Student
ORCID ID: 0009-0000-4402-6921
e-mail: dipruman@gmail.com
Institute of Telecommunications and Global Information Space, NAS of Ukraine, Kyiv, Ukraine

DATA PRE-PROCESSING METHOD FOR FORECASTING GRAIN YIELD IN A HARVEST PROGRAMMING SYSTEM

Background. ***A method for processing meteorological indicators and agronomic data, including the results of field experiments, has been developed, intended for use in the crop programming system, in particular, for predictive modeling and development of scenarios of agrochemical measures in conditions of climate change. Its feature is that it allows integrating heterogeneous data obtained from various sources and creating sets of informative indicators that reflect the conditions of crop development, which provides a more complete representation of the state of agroecosystems and improves the quality of the obtained forecasts. The aim of the work is to develop a method for processing data in the crop programming system, which ensures the consideration of both time dependencies and the hierarchy of the data structure, adapted for use in the crop programming system.***

Methods. ***The research used methods of analysis and synthesis, system analysis, expert evaluation, multi-criteria analysis, statistical, time series analysis, Bayesian approach, machine learning.***

Results. ***The developed method provides the possibility of adaptive modeling of grain crop yields, taking into account the characteristics of a certain agroclimatic zone. This allowed us to take into account both time dependencies and to process the data structure "plant → plot → farm → region". Lag relationships between the yield of the current year and the factors of the previous period were taken into account, which is especially important when using long-term agrometeorological observations, since it allows us to trace the impact of climate change on a certain territory.***

Conclusions. ***The effectiveness of the developed method was tested on field experiment data. The developed mathematical models improved the quality of grain crop yield forecasts by almost 10%. The use of the method of intelligent data analysis made it possible to more accurately determine the impact of meteorological factors on plant development at different stages of the agrocenosis. The constructed crop yield scenarios provide an in-depth analysis of factors that influence plant development at different stages of vegetation, which improves the quality of forecasts for different options for the influence of controlled factors and uncontrolled conditions and, accordingly, increases yield and biologization of grain crop production processes. The developed method, models and scenarios are intended for integration into the yield programming system, but can also be used separately. Their use will create conditions for optimizing agricultural technologies, increasing the efficiency of grain production.***

Keywords: ***data processing, Bayesian networks, agrometeorological data, data integration, machine learning, yield programming, scenarios.***

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