

ЦИФРОВА ТРАНСФОРМАЦІЯ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ В СИСТЕМІ ШІ В ТЕХНОЛОГІЧНИХ СХЕМАХ СЕЛЕКЦІЇ ГОРОХУ ПОСІВНОГО

Чернуський В.В., кандидат с.-г. наук, с.н.с.

Інститут біоенергетичних культур та цукрових буряків, Київ, Україна

Вступ. В умовах воєнного часу високобілкова культура гороху набуває стратегічного значення у вирішенні питання продовольчої безпеки України. Разом з тим врожайність гороху в Україні залишається вкрай низькою і за даними Державної служби статистики у 2024 році склала 2,19 т/га при валовому зборі 416 тис. т з площі 189 тис. га. У зв'язку з кардинальними змінами клімату одним із перспективних шляхів вирішення даної проблеми є одночасне підвищення продуктивності та адаптивності новостворених сортів.

Класична базова проблема селекції більшості сільськогосподарських культур це оптимізація взаємозв'язку формування параметрів цінних продуктивно – адаптивних ознак на еко – градієнтних ландшафтах онтогенетичного розвитку фенотипу. Дані зв'язки як правило негативні, нелінійні, вимагають постійного підлаштування під мінливі умови зовнішнього середовища. Тому відповідно до кардинальних змін клімату необхідно застосування сучасних методів математико – статистичного аналізу у вигляді диференціюючих геометричних поверхонь відгуку для виявлення синергетично – емерджентних балансів продуктивних і адаптивних ознак популяцій, мітки яких знаходяться у максимальній червоній зоні еко – градієнтних ландшафтів.

А також методів машинного навчання (ML), щоб ефективно аналізувати цей величезний обсяг складних даних. Використання великих даних у поєднанні з ML у селекції рослин має потенціал для революції в цій галузі та підвищення продовольчої безпеки [1].

Це пов'язано також з тим, що глобальна сфера даних розширюється в геометричній прогресії, і за прогнозами, досягне 180 зетабайт у 2025 році, тоді як очікується, що поточні технології не будуть масштабуватися з такою ж швидкістю. Вихід - модульний і цілісний підхід, який поєднує глибокі нейронні мережі, навчені на змодельованих даних, коди виправлення помилок на основі тензорного продукту та механізм запасу надійності об'єднаний в єдиний узгоджений конвеєр [2].

Методика досліджень. Для верифікованого аналізу експериментальних даних застосовано метод побудови математичних моделей складних процесів на засадах генетичних алгоритмів. Це евристичний алгоритм пошуку, що використовується для рішення задач оптимізації і моделювання шляхом послідовного підбору, комбінування і варіації шуканих параметрів з використанням механізмів, що нагадують біологічну еволюцію. У порівнянні з багаторядним алгоритмом групового урахування аргументів даний метод дає можливість синтезувати математичні моделі будь-якої складності без

попереднього вибору числа рядів селекції [3]. Використано платформу RYTHIA 8.1.[4].

Результати досліджень. Класична одновимірна та багатоваріантна статистика - найпоширеніші методи, що використовуються для аналізу даних у селекції рослин. Оцінка генетичного різноманіття, класифікація генотипів рослин, аналіз компонентів врожайності, аналіз стабільності врожаю, оцінка біотичних та абіотичних стресів, прогнозування батьківських комбінацій у програмах гібридної селекції виконуються переважно класичними статистичними методами. Незважаючи на успішне застосування, ці класичні статистичні методи мають недостатню ефективність в аналізі даних, отриманих під час досліджень складних процесів онтогенезу рослин, оскільки генотип, середовище та їх взаємодія ($G \times E$) призводять до недетермінованого та нелінійного характеру характеристик рослин. Масштабний потік даних, включаючи феноміку, метаболоміку, геноміку, необхідно проаналізувати для ефективної інтерпретації результатів, на які впливає $G \times E$. Нелінійні непараметричні методи машинного аналізу у вигляді нейромереж: часткова найменша квадратна регресія, випадковий ліс та опорні векторні машини-одні з найбільш актуальних моделей машинного навчання, які широко застосовуються для аналізу нелінійних та складних даних у класичній селекції рослин [5].

В наших дослідженнях використано поєднання методів аналітичної геометрії у вигляді поверхонь відгуку та нейромереж у вигляді платформ генетичних алгоритмів. Зокрема в системі поверхонь відгуку досліджена диференціація зразків у темпоральній системі умов зовнішнього середовища різних років сортовипробування та диференціація зразків у ієрархічній системі добору контрольний розсадник – конкурсне сортовипробування за різними попередниками (рис. 1, рис. 2).

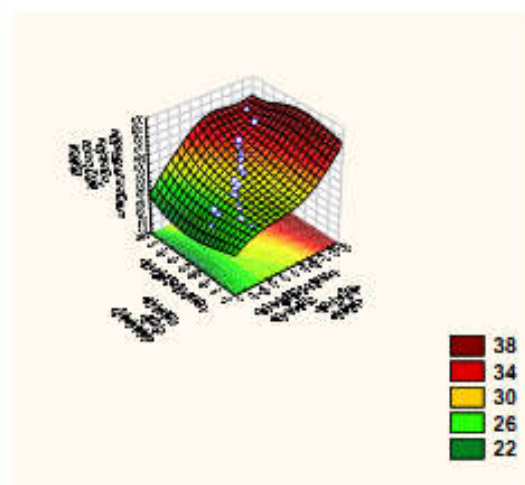


Рис. 1. Диференціація зразків у темпоральній системі умов зовнішнього середовища різних років сортовипробування на аналітичних поверхнях відгуку. Конкурсне сортовипробування 2021 – 2023рр.

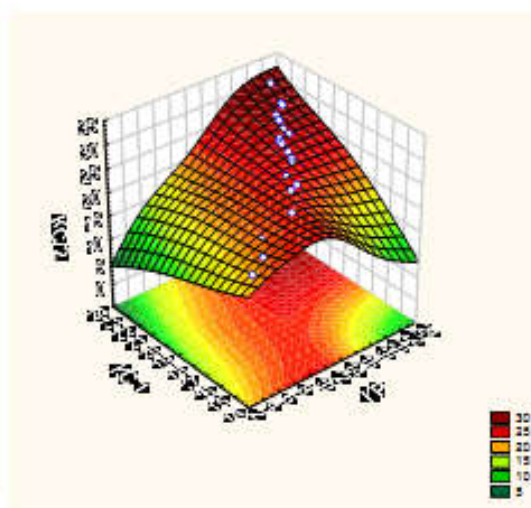


Рис. 2. Диференціація зразків у ієрархічній системі добору контрольний розсадник – конкурсне сортовипробування за різними попередниками, 2022р.

На платформі інтерпретації результатів, на які впливає $G \times E$, доведено значну та достовірну диференціюючу здатність поверхонь відгуку в системах різних умов зовнішнього середовища, різних експериментальних ділянок та різних ієрархічних популяцій добору.

Відпрацьовано методологію добору перспективних зразків вусатого морфо типу гороху з оптимальним поєднанням компонентних ознак на міжпопуляційному рівні шляхом аналізу в системі нейромереж методом генетичних алгоритмів з попередньою нормалізацією параметрів компонентів (рис. 3). За результатами аналізу шляхом оптимізації параметрів компонентних ознак виділено цінні фенотипи за № 7, 10, 12, 16-20.

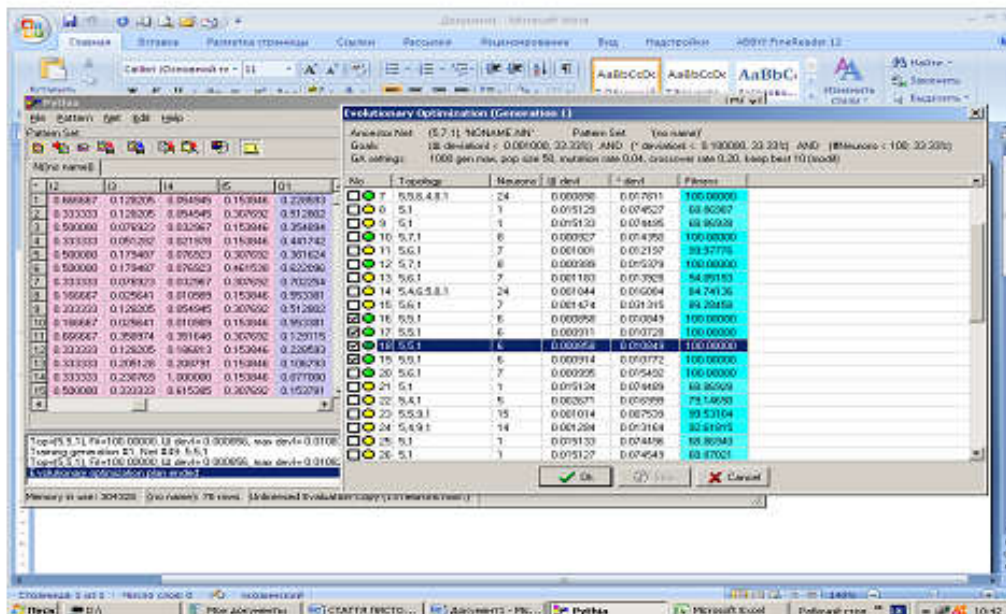


Рис. 3. Аналіз в системі нейромереж III методом генетичних алгоритмів з метою оптимізації параметрів компонентних ознак для виділення цінних фенотипів

Висновки. Доведено значну та достовірну диференціюючу здатність поверхонь відгуку та нейромереж в системах $G \times E$ різних умов зовнішнього середовища, різних експериментальних ділянок та різних ієрархічних популяцій добору.

Література

1. Yoosefzadeh Najafabadi M., Hesami M., Eskandari M. Machine Learning-Assisted Approaches in Modernized Plant Breeding Programs. *Genes*. 2023. 14. 777. <https://doi.org/10.3390/genes14040777>
2. Bar-Lev D., Orr I., Sabary O. et al. Scalable and robust DNA-based storage via coding theory and deep learning. *Nat Mach Intell*. 2025. <https://doi.org/10.1038/s42256-025-01003-z>