УДК 004.89:633.854.78 DOI: 10.36710/IOC-2025-39-01

ПЕРСПЕКТИВИ ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ СИСТЕМАТИЗАЦІЇ СЕЛЕКЦІЙНО-НАСІННИЦЬКОГО МАТЕРІАЛУ СОНЯШНИКА

Е. Б. Алієв[1], **О. Ю. Алієва**[2]

[1] *Дніпровський державний аграрно-економічний університет*
[2] *Інститут олійних культур Національної академії аграрних наук України*

Розроблено метод класифікації насіння соняшнику (Helianthus annuus L.) за морфологічними та колірними ознаками із використанням цифрової обробки зображень і штучних нейронних мереж. Модель із двома прихованими шарами забезпечила високу точність класифікації (R^2 = 0,98, RASE < 0,0003, Accuracy = 1,0), що підтверджено показниками F1-score та MCC. Отримані результати свідчать про значний потенціал нейронних мереж для систематизації селекційно-насінницького матеріалу олійних культур. Використання нейромережевих підходів дає змогу автоматизувати і підвищити точність оцінювання фенотипічних ознак, прискорити аналіз великих колекцій зразків та створювати цифрові бази даних генофонду. Впровадження штучного інтелекту в селекційно-насінницьку практику відкриває нові можливості для підвищення ефективності процесів добору, оцінки й прогнозування продуктивності олійних культур.

Ключові слова: штучні нейронні мережі, класифікація насіння, соняшник, морфологічні ознаки, фенотипування, цифрові технології.

Вступ. Сучасний розвиток селекції олійних культур характеризується зростанням обсягів експериментальних даних, отриманих у процесі польових, лабораторних і генетичних досліджень (Makhova et al. 2023; Kryvosheyeva et al. 2009; Katal et al. 2013). Обробка таких даних традиційними статистичними методами стає дедалі складнішою через їх багатовимірність, неоднорідність та наявність прихованих нелінійних залежностей (Najafabadi 2021; Hesami et al. 2022; Zhang et al. 2021). У зв'язку з цим актуальним напрямом розвитку є впровадження інтелектуальних систем аналізу даних, зокрема нейронних мереж, здатних до самонавчання, виявлення закономірностей і прогнозування результатів селекційних процесів (Libbrecht et al. 2015; Monino et al. 2006).

Проте ефективність використання нейронних мереж значною мірою залежить від якості підготовки вхідних даних. Більшість результатів наукових досліджень у селекції насіння представлені у вигляді таблиць, текстових описів або графічних матеріалів, що не відповідають вимогам машинної обробки (Kim et al. 2020; Niazian et al. 2020). Тому виникає необхідність розроблення методики трансформації цих даних у формалізований, структурований формат, придатний для навчання та тестування моделей штучного інтелекту.

Розроблення такої методики забезпечить можливість автоматизованої обробки великих масивів наукових даних, підвищить точність і швидкість селекційного аналізу, а також створить передумови для формування єдиної цифрової бази знань у галузі олійних культур. Це, у свою чергу, сприятиме

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

*Corresponding author

підвищенню ефективності селекційного процесу, скороченню термінів створення нових сортів та адаптації методів штучного інтелекту до аграрних досліджень.

Helianthus annuus L. (соняшник культурний) є однією з провідних олійних культур світу, що належить до родини Asteraceae. Його вирощують у багатьох країнах, зокрема в Україні (Vedmedeva 2023; Aliiev et al. 2024), яка входить до числа світових лідерів із виробництва насіння соняшнику та соняшникової олії. Завдяки високій адаптивній здатності до кліматичних умов, потужній кореневій системі, великій листковій поверхні та високій продуктивності, соняшник є стратегічною культурою для аграрного сектору. Проте селекційний процес ускладнюється значною морфологічною подібністю між сортами та гібридами, а також варіабельністю ознак, зумовленою умовами вирощування (Vedmedeva et al. 2022).

Ідентифікація сортів і гібридів соняшнику за морфологічними ознаками часто є складним завданням через подібність таких характеристик, як форма сім'янки, забарвлення оболонки, наявність опушення, маса 1000 насінин, товщина оболонки та геометрія насінної поверхні (Aliiev et al. 2024). Багато з цих ознак проявляються лише на певних етапах розвитку рослини або мають перехресні значення, що ускладнює достовірну диференціацію матеріалу (Aliiev 2023; Aliiev et al. 2024). У випадках міжсортової або міжлінійної гібридизації виникають форми з проміжними характеристиками, що ще більше ускладнює ідентифікацію та класифікацію насіння. Точна ідентифікація генотипів має важливе значення для селекційного добору, контролю сортової чистоти та оптимізації насінництва.

Штучні нейронні мережі (ШНМ) – це потужні інструменти інтелектуального аналізу даних, створені за аналогією з біологічними нейронними системами (Mahesh 2020). Вони здатні до розпізнавання складних закономірностей, класифікації та прогнозування на основі великих обсягів вхідних даних. У рослинництві ШНМ активно застосовуються для розпізнавання видів, аналізу морфологічних параметрів насіння, оцінки стану посівів та прогнозування урожайності. Для цього моделі навчаються на основі експериментальних даних, що включають виміряні морфологічні або спектральні ознаки (Chandra et al. 2020). Під час навчання нейронна мережа поступово коригує свої внутрішні параметри для мінімізації різниці між прогнозованими та фактичними результатами, формуючи здатність узагальнювати знання.

У даному дослідженні розглядається можливість використання морфологічних ознак насіння соняшнику для побудови моделі класифікації за допомогою штучних нейронних мереж. Було сформовано вибірку насіння з колекції Інституту олійних культур НААН. Морфологічні параметри насіння – маса, довжина, ширина, товщина, індекс форми, колір та текстура поверхні – використовуються як вхідні дані для моделі ШНМ.

Метою дослідження є визначення інформативних ознак насіння соняшнику, що найбільше впливають на точність класифікації, оцінка можливості їх використання для автоматичного розпізнавання сортів та розроблення методики трансформації результатів наукових досліджень у формат, придатний для обробки нейронними мережами. Це дозволить удосконалити цифрові інструменти в селекції, підвищити ефективність аналізу даних і створити основу для автоматизованої системи ідентифікації сортів олійних культур.

© Е. Б. Алієв, О. Ю. Алієва

Матеріали і методи досліджень. У дослідженні використовували насіння соняшнику (Helianthus annuus L.) з колекції Інституту олійних культур НААН, вирощене в умовах Південного Степу України. До складу вибірки увійшли сортозразки різного генетичного походження: InK2830, КГ111, SL2966, НА298, APS10, I3K1070, Л7242, In18917, LD723, AH70029RF, LD722, СЛ2354, КГ13, In7034, I2K670, НАР7, In18906, RHA273, APS04, КГ113, LD4, InK85, InK1276, СЛ1218, I2K20031, M19, I2K87, КГ15, Л259524, InK1124, НА300Б, In18906, I2K87, InK1276. Зразки відрізнялися між собою за маркерними морфологічними ознаками, передусім за забарвленням насіння, що дозволило сформувати репрезентативну колекцію для подальшого морфологічного та цифрового аналізу.

Зображення насіння соняшнику отримували за допомогою фотокамери ELP-USBFHD01M-BFV. Оптичне збільшення становило 300×, роздільна здатність знімків – 1280×1024 пікселі. Для забезпечення стабільних умов освітлення використовували комбіноване галогенне підсвічування зверху та знизу. Перед фотографуванням насіння розміщували на предметному склі у горизонтальному положенні та фіксували покривним склом, щоб уникнути зміщення під час зйомки.

Морфометричні параметри насіння визначали за допомогою програмного середовища ImageJ. Для обробки зображень використовували плагіни Seed Analysis та Color Threshold.

Аналіз включав такі основні етапи (рис. 1):

1. Попередня обробка зображень – видалення фону та зменшення шумів за допомогою фільтра Gaussian blur із радіусом 2 пікселі.

2. Сегментація насіння – автоматичне виявлення контурів окремих насінин на основі їх форми та кольору.

3. Вимірювання параметрів – за допомогою функції Measure отримували довжину, ширину, площу проєкції.

4. Оцінка кольору – плагін Color Threshold визначав інтенсивність каналів RGB (червоного, зеленого та синього), що використовувалось для подальшої класифікації за відтінками оболонки.

Для визначення внеску окремих морфометричних і колірних ознак насіння соняшника у класифікаційні результати було застосовано номінальну логістичну модель. Для оцінки впливу кожного параметра використовували показники log-worth та p-value. Значення log-worth визначається як $-\log_{10}$(p-value); при цьому значення >2 вважається статистично значущим на рівні 0,01. На основі цих показників було виконано покроковий регресійний аналіз для відбору найінформативніших змінних, які згодом використовувалися при побудові штучної нейронної мережі (ШНМ).

Для класифікації насіння соняшника за фенотипічними ознаками (наприклад, колір оболонки, маса, форма, розмір) дані було розділено на три підмножини:

– 50% – для навчання моделі (коригування ваг і зміщень);

– 30% – для валідації (контроль якості навчання і запобігання перенавчанню);

– 20% – для тестування (оцінювання остаточної точності моделі).

© Е. Б. Алієв, О. Ю. Алієва

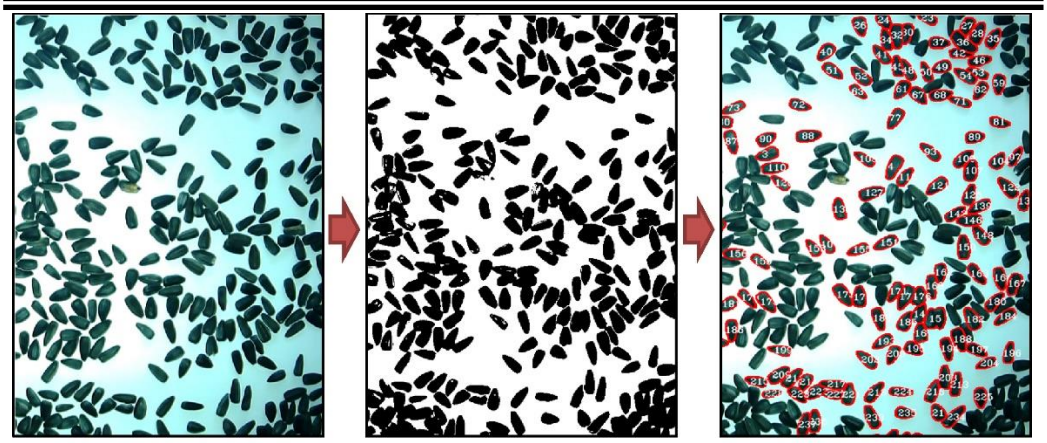

Рис. 1. **Етапи обробки зображення**

Навчання ШНМ виконували за алгоритмом стохастичної зворотної поширюваності помилки. Кількість епох навчання встановлено на рівні 600, що забезпечило стабільну збіжність без втрати узагальнюючої здатності.

Вхідний шар містив сім нейронів, що відповідали кількості вхідних ознак (маса 1000 насінин M_{1000}, довжина L, ширина B, товщина T, колір R, G, B). Вихідний шар складався з двох нейронів, які відповідали трьом класам. У прихованій частині мережі застосовували два шари, кожен із комбінацією трьох гіперболічних тангенсів і трьох Гаусових функцій активації (рис. 2). Згідно досліджень (Aliiev et al. 2025) вихідний шар складався з трьох нейронів, що відображали відповідні групи насіння (рис. 2).

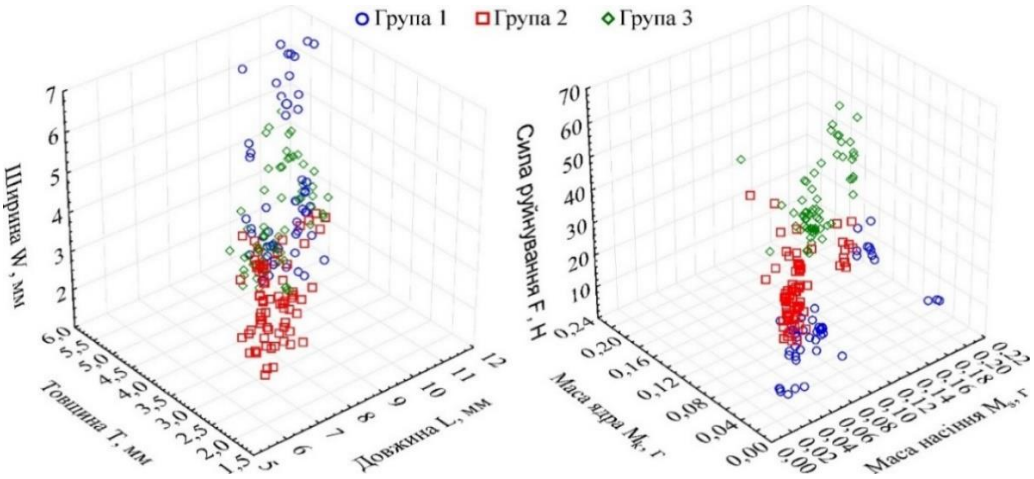

Рис. 2. **Результати групування даних**

Для оцінки ефективності моделі розраховували такі показники:

– Generalized R^2, Entropy R^2, Maximum Likelihood – для оцінки узгодженості моделі;

– Misclassification Rate – частота неправильних класифікацій;

© Е. Б. Алієв, О. Ю. Алієва

– Sensitivity (TP/(TP+FN)) – частка правильно визначених позитивних зразків;

– Specificity (TN/(TN+FP)) – частка правильно визначених негативних зразків;

– Accuracy ((TP+TN)/(TP+TN+FP+FN)) – загальна точність моделі;

– F1-score (2TP/(2TP+FP+FN)) – баланс точності та повноти;

– Matthews Correlation Coefficient (MCC) – комплексна метрика, що враховує всі чотири типи результатів (TP, TN, FP, FN) і варіює від –1 (повна невідповідність) до +1 (повна відповідність).

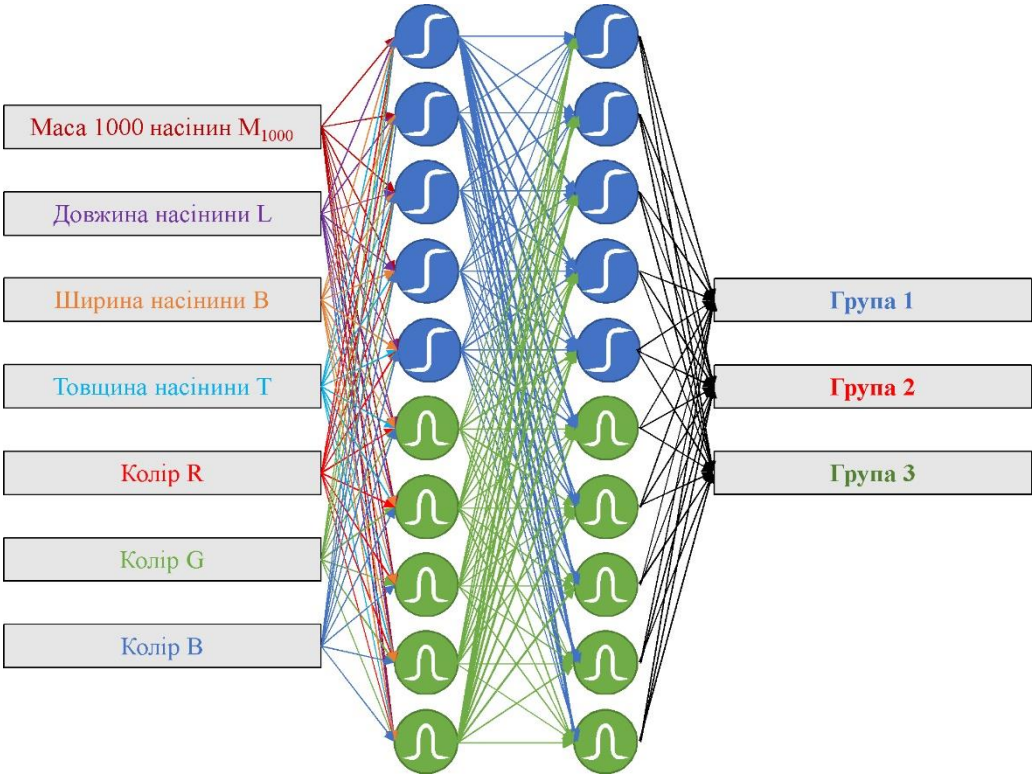

Рис. 3. **Схематичне зображення структури нейронної мережі для класифікації насіння соняшника за морфологічними ознаками**

Для аналізу кореляцій між ознаками насіння та просторовими даними застосовано метод головних компонент (PCA). Визначено оптимальні поєднання головних компонент на основі scree plot та власних значень, після чого для візуалізації й інтерпретації було обрано перші дві компоненти (PC1 і PC2).

Результати досліджень та їхнє обговорення. Номінальний логістичний аналіз показав, що маса 1000 насінин, довжина, ширина і товщина насінини мали найнижчі значення p і найвищі значення log-worth (табл. 1). Це свідчить про те, що саме ці ознаки є найбільш значущими для розрізнення між групами. Ознаки забарвлення (колір R, G, B) були визначені як другорядні за важливістю (усі $p > 0{,}0001$).

© Е. Б. Алієв, О. Ю. Алієва

Таблиця 1

Внесок морфологічних ознак насіння соняшнику у класифікацію за результатами номінального логістичного аналізу

Морфологічний параметр насіння	Log-worth	p-value
Маса 1000 насінин M1000	4,27	0,000054
Довжина насінини L	4,47	0,000034
Ширина насінини B	4,66	0,000022
Товщина насінини T	4,05	0,000089
Колір R	3,55	0,00028
Колір G	3,19	0,00065
Колір B	3,64	0,00023

Отже, на основі результатів номінального логістичного аналізу та покрокового комбінованого аналізу було обрано такі ознаки для подальшого використання як вхідні параметри моделі ШНМ (рис. 2).

Таблиця 2

Ефективність моделей штучних нейронних мереж у класифікації соняшника за морфологічними ознаками насіння

Характеристика	Навчання	Валідація	Тестування
Generalized R^2	0,98	0,95	0,98
Entropy R^2	0,98	0,95	0,98
RASE	0,00008	0,000015	0,00006
Mean Abs Dev	0,000004	0,0003	0,00003
Misclassification Rate	0	0	0
Sum Freq	12520	7526	4331

Ефективність моделі ШНМ оцінювали на етапах навчання, валідації та тестування. Модель продемонструвала відмінні результати, забезпечивши узагальнені та ентропійні значення R^2 на рівні 0,98 і RASE < 0,0003. Частота помилкових класифікацій залишалася стабільно нульовою протягом усіх етапів – навчання, валідації та тестування (табл. 2). Крім того, значення чутливості, специфічності, точності, F1-міри та коефіцієнта Метьюса (MCC) підтверджують високу ефективність моделі (табл. 3). Отримані результати свідчать, що розрізнення між групами соняшника за морфологічними ознаками їх насіння може бути успішно реалізоване за допомогою моделі штучної нейронної мережі.

Таблиця 3

Показники для оцінювання штучних нейронних мереж, отримані з матриці плутанини тестових даних

Модель	Sensitivity	Specificity	Accuracy	F1-score	MCC
A	1	1	1	1	1
B	0,96	0,95	0,95	0,96	0,95
C	0,81	0,80	0,82	0,80	0,81

© Е. Б. Алієв, О. Ю. Алієва

Узагальнена методологія систематизації селекційно-насінницького матеріалу олійних культур на основі нейронних мереж, побудована на представлених експериментальних, морфометричних і фенотипічних даних.

Етап 1. Формування бази вихідних даних. На першому етапі здійснюється комплексний збір інформації про морфометричні, фізико-механічні та оптичні властивості насіння (маса 1000 насінин, довжина, ширина, товщина, колір тощо). Для цього використовуються традиційні методи вимірювання, а також комп'ютерні системи зору на основі OpenCV, TensorFlow і HSV-моделі кольору, що дозволяє автоматично зчитувати параметри з цифрових зображень насіння.

Етап 2. Цифрова обробка та нормалізація даних. Після збирання даних проводиться їх очищення, усунення шумів та нормалізація до уніфікованого діапазону (0–1) для кожної ознаки. Створюються векторизовані набори даних (feature vectors), які описують кожен зразок насіння у багатовимірному просторі ознак. Для підвищення надійності використовується метод Principal Component Analysis (PCA) для виділення головних компонент, що визначають максимальну варіацію у вибірці.

Етап 3. Вибір інформативних ознак. На основі статистичних критеріїв (p-value, log-worth, коефіцієнти кореляції, дисперсійний аналіз) здійснюється відбір найбільш значущих морфологічних і оптичних характеристик, що мають найвищу вагу в процесі класифікації. Результати використовуються для формування оптимального набору вхідних параметрів нейронної мережі.

Етап 4. Навчання штучної нейронної мережі. Формується багатошарова нейронна архітектура (Multilayer Perceptron або Convolutional Neural Network), що навчається на навчальній вибірці фенотипових даних. Навчання здійснюється із застосуванням методів backpropagation і cross-validation для уникнення перенавчання. Цільовими класами можуть бути сорти, генотипи або класи за якісними показниками. Оцінка точності проводиться за метриками Accuracy, F1-score, Sensitivity, Specificity, MCC.

Етап 5. Інтерпретація та валідація результатів. Модель проходить перевірку на тестовій вибірці, порівнюються прогнозовані та експериментальні значення. За допомогою heatmap-візуалізацій та градієнтних карт ознак визначаються параметри, що мають найбільший вплив на розпізнавання сортів. Проводиться експертна оцінка і корекція моделі на основі реальних польових та лабораторних спостережень.

Етап 6. Кластеризація та систематизація селекційного матеріалу. На основі нейронної моделі здійснюється кластеризація зразків за подібністю фенотипових характеристик. Формуються групи зразків з близькими властивостями. Це дозволяє структурувати генофонд і виділити цільові групи для подальшої селекції.

Етап 7. Формування рекомендацій для створення нових ліній. За результатами класифікації та кластеризації відбираються зразки з найкращими поєднаннями цільових ознак. На цій основі формується віртуальна селекційна модель, що прогнозує потенційні результати схрещування та відбір перспективних генотипів для польових випробувань.

Етап 8. Інтеграція у систему селекційно-насінницького менеджменту. Остаточний етап передбачає інтеграцію нейромережевої системи у селекційно-насінницькі інформаційні платформи. Це забезпечує автоматизований аналіз фенотипових ознак, формування електронного паспорта кожного зразка,

© Е. Б. Алієв, О. Ю. Алієва

рекомендації для добору батьківських форм і прогнозування властивостей гібридів.

Запропонована методологія поєднує традиційні агрономічні підходи з інтелектуальним аналізом даних на основі штучних нейронних мереж. Це дозволяє значно скоротити час селекційного циклу, підвищити точність оцінки генотипів і створити науково обґрунтовану систему для систематизації та розробки нового селекційно-насінницького матеріалу олійних культур.

Висновки. Розроблено методику класифікації насіння соняшнику (Helianthus annuus L.) за морфологічними та колірними ознаками із застосуванням цифрової обробки зображень і штучних нейронних мереж. Нейронна модель із двома прихованими шарами забезпечила високу точність класифікації (R^2 = 0,98, RASE < 0,0003, Accuracy = 1,0), що підтверджено показниками F1-score і MCC.

Проведені дослідження підтверджують високий потенціал використання штучних нейронних мереж для систематизації селекційно-насінницького матеріалу олійних культур. Застосування нейромережевих моделей дає змогу автоматизувати класифікацію зразків за морфологічними та колірними ознаками, підвищити точність і об’єктивність оцінювання фенотипічних характеристик. Використання таких моделей значно скорочує час аналізу великомасштабних колекцій, сприяє виявленню закономірностей між морфометричними параметрами та генетичними особливостями зразків, а також відкриває можливості для формування цифрових баз даних генофонду.

Отже, впровадження нейронних мереж у селекційно-насінницьку практику є перспективним напрямом розвитку сучасних технологій у аграрній науці, що забезпечує ефективну інтеграцію штучного інтелекту в управління процесами добору, оцінки та прогнозування продуктивності олійних культур.

References

Aliiev EB (2023) The prospects of quantitative phenotyping of oilseed crops. Agrology. 6(3): 49–59 DOI: 10.32819/021109

Aliiev EB, Aliieva OYu (2025) Mathematical model for the systematization of breeding and seed material of oil crops based on neural networks. Scientific and Technical Bulletin of the Institute of Oilseed Crops NAAS. 38: 6–17 DOI: 10.36710/IOC-2025-38-01

Aliiev EB, Vedmedeva KV (2024) Quantitative phenotyping of sunflower genotypes: monograph. Dnipro: LIRA. 204 p.

Chandra AL, Desai SV, Guo W, Balasubramanian VN (2020) Computer vision with deep learning for plant phenotyping in agriculture: A survey. arXiv. 20202006.11391. DOI: 10.48550/arXiv.2006.11391

Hesami M, Alizadeh M, Jones AMP, Torkamaneh D (2022) Machine learning: Its challenges and opportunities in plant system biology. Appl. Microbiol. Biotechnol. 106: 3507–3530 DOI: 10.1007/s00253-022-11963-6

Katal A, Wazid M, Goudar RH (013) Big data: Issues, challenges, tools and good practices; Proceedings of the2013 Sixth International Conference on Contemporary Computing (IC3). Noida. India. 8–10 August: 404–409

Kim KD, Kang Y, Kim C (2020) Application of genomic big data in plant breeding: Past, present, and future. Plants. 9: 1454 p. DOI: 10.3390/plants9111454

© Е. Б. Алієв, О. Ю. Алієва

Kryvosheyeva HV, Ryabchoun VK, Leonova NN, Vedmedeva KV (2009) National sunflower collection as a source for valuable traits in breeding. Scientific and Technical Bulletin of the Institute of Oilseed Crops NAAS. 14: 28–34

Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics. Nat. Rev. Genet.. 6: 321–332 DOI: 10.1038/nrg3920

Mahesh B (2020) Machine learning algorithms-a review. Int. J. Sci. Res. 9: 381–386

Makhova TV, Vedmedev SR, Polyakov OI (2023) Creation of databases of phenotypic signs and adding lines of special recognition. Scientific and Technical Bulletin of the Institute of Oilseed Crops NAAS. 35: 51–62 DOI: 10.36710/IOC-2023-35-05

Monino J-L, Sedkaoui S (2016) Big Data, Open Data and Data Development. Volume 3 John Wiley & Sons. Hoboken. NJ. USA

Najafabadi MY (2021) Using Advanced Proximal Sensing and Genotyping Tools Combined with Bigdata Analysis Methods to Improve Soybean Yield. Ph.D. Thesis. University of Guelph. Guelph. ON. Canada

Niazian M, Niedbała G (2020) Machine learning for plant breeding and biotechnology. Agriculture.10: 436. DOI: 10.3390/agriculture10100436

Vedmedeva K (2023) Marker morphological traits of sunflower: identification, inheritance and effects of gene alleles: monograph. Kyiv: Agrarna nauka. 240

Vedmedeva KV, Bilozub AS, Kobzeva DO (2022) Manifestation of economic symptoms in sesame samples with different seed colors. Scientific and Technical Bulletin of the Institute of Oilseed Crops NAAS. 33: 36–45 DOI: 10.36710/IOC-2022-33-04

Zhang N, Gupta A, Chen Z, Ong Y-S (2021) Evolutionary machine learning with minions: A case study in feature selection. IEEE Trans. Evol. Comput. 26: 130–144 DOI: 10.1109/TEVC.2021.3099289

PROSPECTS FOR USING NEURAL NETWORKS TO SYSTEMATIZE BREEDING AND SEED MATERIAL OF SUNFLOWER

E. B. Aliiev[1], O. Yu. Aliieva[2]

[1] *Dnipro State Agrarian and Economic University*
[2] *Institute of Oilseed Crops of the National Academy of Agrarian Sciences of Ukraine*

A method for classifying sunflower (Helianthus annuus L.) seeds based on morphological and color traits using digital image processing and artificial neural networks has been developed. The model with two hidden layers achieved high classification accuracy (R^2 = 0.98, RASE < 0.0003, Accuracy = 1.0), as confirmed by F1-score and MCC metrics. The obtained results indicate the significant potential of neural networks for systematizing breeding and seed material of oilseed crops. The use of neural network approaches allows automation and improvement of the accuracy of phenotypic trait assessment, accelerates the analysis of large sample collections, and enables the creation of digital gene bank databases. The implementation of artificial intelligence in breeding and seed production practice opens new opportunities for increasing the efficiency of selection, evaluation, and productivity forecasting processes for oilseed crops.

Key words: artificial neural networks, seed classification, sunflower, morphological traits, phenotyping, digital technologies.

© Е. Б. Алієв, О. Ю. Алієва