

УДК 004.4'2: 631.526.3

Стариченко Є. М., канд. екон. наук, завідувач відділу науково-технічної інформації**Мажуга К. М.**, заступник завідувача відділу науково-технічної інформації**Орленко Н. С.**, канд. екон. наук, доцент, старший науковий співробітник відділу науково-технічної інформації

Український інститут експертизи сортів рослин

E-mail: n.s.orlenko@gmail.com

ІДЕНТИФІКАЦІЇ СХОЖИХ СОРТІВ РОСЛИН ІЗ ЗАСТОСУВАННЯМ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ ТА МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ

Ідентифікація сортів сільськогосподарських культур є важливою складовою у процесі селекції, реєстрації, виробництва й торгівлі насінням та посівним матеріалом. Традиційний підхід до ідентифікації нових сортів рослин передбачає спостереження і опис морфологічних ознак вегетативних і генеративних органів рослини під час польових та лабораторних досліджень, що проводяться за методиками кваліфікаційної експертизи на відмінність, однорідність та стабільність (ВОС). Складовими елементами, які використовують процесі ідентифікації сорту, є:

1. Методична база проведення кваліфікаційної експертизи.

2. Інформаційна база результатів кваліфікаційної експертизи.

3. Математико-статистичні методи та інструментальні засоби, які застосовують під час оброблення результатних даних кваліфікаційної експертизи на ВОС.

Міжнародною базою методичного забезпечення ВОС є технічні документи UPOV, зокрема TGP/7/3. У відповідності до цього документа розробляються національні адаптовані методичні вказівки та рекомендації. Це забезпечує систему контролю якості методології проведення ідентифікації сорту рослини.

Важливою складовою процесу ідентифікації нового сорту є порівняння його зі схожими сортами які становлять загальновідому колекцію сортів (еталонну колекцію). Зауважимо, що суть терміну «загальновідомий» розкрито у в документі TGP / 3 UPOV.

Інформаційну базу для ідентифікації сорту складають міжнародні та національні бази даних сортів, які підтримують ряд країн та організацій. Ці дані доступні для широкого кола користувачів. База даних сортів рослин UPOV (<https://www.upov.int/genie/index.xhtml>) є найбільшою базою у світі та відображає дані у режимі онлайн. В базі даних Українського інституту експертизи сортів рослин зберігаються дані щодо 46 391 сорту рослин 664 ботанічних таксонів. Перелік сортів можна переглянути за допомогою он-лайн інформаційно-пошукової системи «Реєстр сортів», яка доступна за посиланням <http://service.ukragroexpert.com.ua>.

Математико-статистичний апарат, який служить цілям пошуку схожих сортів відносить-

ся до методів інтелектуального аналізу даних. Більш традиційним інструментальним засобом класифікації даних є метод ієрархічної агломератної кластеризації. Застосування кластерного аналізу під час дослідження морфологічних характеристик сортів рослин дозволяє зменшити розмірність вибірки даних, що у свою чергу сприяє більш точній ідентифікації нових сортів. Кластерний аналіз, зазвичай, проводять у такій послідовності: формують вибірки об'єктів для кластеризації (набір морфологічних ознак у випадку експертизи ВОС); визначають множину змінних, за якими будуть оцінюватися об'єкти у вибірці.

Інший метод класифікації – це машинне навчання. Його ще називають “здатністю програмного алгоритму робити висновки, на підставі певного набору даних”. Головною особливістю цього метода є те, що комп'ютерна програма, яка реалізує цей метод, вчиться з власного досвіду. У контексті кваліфікаційної експертизи сортів рослин на ВОС таким набором даних є сукупність морфологічних ознак сортів рослин й складається з таких категорій даних:

– сорти рослин, що були внесені до Державного реєстру сортів рослин, придатних для поширення в Україні. Дані цієї категорії можуть використовуватися як навчальна вибірка.

– нові сорти, на які подано заявки на внесення до Реєстру.

Одним з алгоритмів машинного навчання є алгоритм *K*-найближчих сусідів (KNN алгоритм), який передбачає, що вже існує певна кількість об'єктів з точною класифікацією. В контексті кваліфікаційної експертизи – це набір схожих загальновідомих сортів рослин, які вже було занесено до Державного реєстру сортів рослин, придатних для поширення в Україні.

Враховуючи великомасштабний характер ботанічної морфометрії автоматизація пошуку схожих сортів рослин є необхідністю. Автоматизована класифікація сортів рослин за їх морфологічними ознаками безперечно знижує трудомісткість процесу пошуку схожих сортів й сприяє більш точній класифікації нового сорту. Однак, потрібно розуміти, що рослини є надзвичайно різноманітні за формою, розміром і кольором, тому жоден метод не є панацеєю для рішення задачі ідентифікації сортів рослин.