КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

УДК 004.9

В.М. ГРУША

НОРМАЛІЗАЦІЯ ТА ЗМЕНШЕННЯ РОЗМІРНОСТІ ДАНИХ ХЛОРОФІЛ-ФЛУОРОМЕТРІВ

Вступ. Один із способів спостереження за станом рослин є метод вимірювання індукції флуоресценції хлорофілу (ІФХ), що полягає в освітлені листа рослини у синьому спектрі світла з подальшою реєстрацією випромінювання хлорофілу в червоному спектрі світла. В результаті отримується так звана крива індукції флуоресценції хлорофілу (ІФХ). В даний час ІФХ вимірюються спеціальними портативними приладами хлорофіл-флуорометрами. Ряд таких приладів розроблено в Інституті кібернетики імені В.М. Глушкова. Зокрема, розроблено сімейство приладів «Флоратест», на заміну яких у даний час розроблено мережу бездротових сенсорів ІФХ [1]. Аналіз літератури та дослідження, проведені в Інституті, свідчать, що діагностика стану рослин на базі кривих ІФХ ускладнена через значну залежність стану рослин від факторів навколишнього середовища та природну варіацію, що притаманна біологічним об'єктам [2]. В зв'язку з цим, в останні роки все ширше застосовуються методи машинного навчання і зокрема нейронні мережі (НМ) [3], що дозволяють враховувати приховані залежності у виміряних даних. При обробці даних за допомогою нейронних мереж найчастішими задачами при підготовці вхідних векторів НМ є нормалізація та зменшення розмірності цих даних. Нормалізація – це приведення значень у певний діапазон, переважно застосовуються інтервали [0,1], [–1,1] або ж близькі до них. Це дозволяє привести вимірювання різних величин до однієї шкали та усунути деякі відмінності у даних, виміряних різними флуорометрами

V. Grusha

CHLOROPHYLL FLUOROMETER DATA NORMALIZATION AND DIMENSIONALITY REDUCTION

The paper describes few approaches to normalization and dimensionality reduction of chlorophyll fluorometer data.
Key words: chlorophyll fluorometer, chlorophyll fluorescence induction, neural networks.

Описаны несколько подходов к нормализации данных хлорофилл-флуорометров и уменьшение их размерности.
Ключевые слова: хлорофилл-флуорометр, индукция флуоресценции хлорофилла, нейронные сети.

Описано декілька підходів до нормалізації даних хлорофіл-флуорометрів та зменшення їхньої розмірності.
Ключові слова: хлорофіл-флуорометр, індукція флуоресценції хлорофілу, нейронні мережі.

(невідкаліброваними або ж флуорометрами з дещо відмінними характеристиками освітлення та діапазоном вимірювання). Зменшення розмірності даних дозволяє зменшити кількість входів НМ, нейронів у прихованих шарах та, відповідно, скоротити час навчання нейронної мережі і зменшити вимоги до обчислювальної продуктивності.

Дана стаття містить: 1) вивчення можливості покращення класифікації рослин обприсканих різними дозами гербіциду за формою ІФХ, застосовуючи різні способи нормалізації вхідного вектора нейронної мережі; 2) дослідження кількох способів зменшення розмірності вимірювань ІФХ на прикладі задачі класифікації виду рослин, використовуючи НМ; 3) визначення водного дефіциту шляхом апроксимації даних ІФХ та параметрів навколишнього середовища НМ.

ІФХ. Крива ІФХ та найчастіше використовувані параметри для її аналізу зображені на рис. 1.

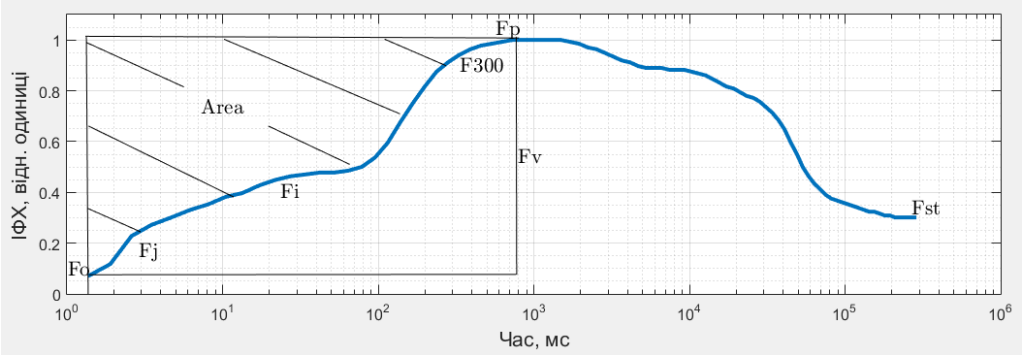


РИС. 1. Крива ІФХ (логарифмічна шкала)

Найбільш застосовувані параметри ІФХ [4]: F_0 – початковий рівень флуоресценції; F_j – флуоресценція в точці J (приблизний час 2 мс); F_I – флуоресценція в точці I (приблизний час 30 мс); F_t – флуоресценція в час t; F_M або F_p – максимальний рівень флуоресценції хлорофілу; t_{F_M} – час досягнення максимуму флуоресценції (мс); t_{F_0} – час вимірювання початкового рівня флуоресценції (мс); F_{st} – стаціонарний рівень флуоресценції; $F_v = F_m - F_0$ – максимальна варіабельна флуоресценція; $k_1 = F_m / F_{st}$ – розрахунковий параметр, який характеризує ефективність темнової стадії фотосинтезу; F_v / F_m – розрахунковий па-

раметр, який характеризує ефективність фотосистеми II; $Rfd = (Fm - Fst)/Fst$ – так званий «індекс життєдіяльності рослини»; $k_2 = (F_m - F_{st})/F_m$ – коефіцієнт ІФХ, який змінюється під дією стресових чинників; $V_t = (F_t - F_0)/(F_m - F_0)$ – відносна змінна флуоресценція в час t; $V_j = (F_j - F_0)/(F_m - F_0)$ – відносна змінна флуоресценція (відносно точки J); $V_I = (F_I - F_0)/(F_M - F_0)$ – відносна змінна флуоресценція (відносно точки I); $V_{st} = (F_{st} - F_0)/(F_M - F_0)$ – відносна змінна флуоресценція (відносно стаціонарного рівня); $Area = \int_{t_{F_0}}^{t_{F_m}} (F_m - F_t)\,dt$ – площа над кривою ІФХ між F_0 та F_m.

Нормалізація даних. Виміряні дані флуоресценції утворюють матрицю *X*, в якій кожен стовпець складається з 89 значень ІФХ, що утворюють виміряну криву ІФХ. x_{ik} – елемент матриці в *i*-му рядку, *k*-му стовпчику. $\tilde{x}_{ik}$ – нормалізоване значення флуоресценції в *i*-му рядку, *k*-го стовпчика. $\bar{x}_k$ – середнє значення *k*-го стовпця, δ_{x_k} – стандартне відхилення значень ІФХ в *k*-му стовпці. $F_{m,k}$ – максимальне значення в *k*-му стовпчику, $F_{0,k}$ – мінімальне значення флуоресценції в *k*-му стовпчику, $x_{\max ik}$ – максимальне значення матриці *X*.

В наукових публікаціях, присвячених ІФХ та методам машинного навчання застосовують різноманітні способи нормалізації як то ділення на F_m або F_0 [5], мінімаксна нормалізація тощо. Розглянемо деякі з найбільш застосовуваних способів нормалізації:

- десяткове масштабування:

$$\tilde{x}_{ik} = \frac{x_{ik}}{10^3}; \tag{1}$$

- мінімаксна нормалізація в межах [0,1]

$$\tilde{x}_{ik} = \frac{x_{ik} - F_{0,k}}{F_{m,k} - F_{0,k}}; \tag{2}$$

- мінімаксна нормалізація в межах [–1,1]

$$\tilde{x}_{ik} = 2\frac{x_{ik} - F_{0,k}}{F_{m,k} - F_{0,k}} - 1; \tag{3}$$

- централізація значення навколо середнього $\overline{x}_k$ (*z*-оцінка).

$$\tilde{x}_{ik} = \frac{x_{ik} - \overline{x}_k}{\delta_{x_k}}; \tag{4}$$

-нормалізація вектора (перетворення заданого вектора у колінеарний з одиничною довжиною)

$$\tilde{x}_{ik} = \frac{x_{ik}}{\sqrt{\sum x_k^2}}; \tag{5}$$

- нормалізація на основі максимуму кривої (Fm)

$$\tilde{x}_{ik} = \frac{x_{ik}}{F_{m,k}}; \tag{6}$$

- нормалізація на основі максимального значення матриці X

$$\tilde{x}_{ik} = \frac{x_{ik}}{x_{\max ik}}; \tag{7}$$

- нормалізація на основі мінімального значення кривої ІФХ (F_0)

$$\tilde{x}_{ik} = \frac{x_{ik}}{F_{0,k}}; \tag{8}$$

- нормалізація на основі максимуму та мінімуму кривої (Fm, F_0)

$$\tilde{x}_{ik} = \frac{x_{ik}}{F_{m,k} - F_{0,k}}. \tag{9}$$

Для тестування методів нормалізації були використані дані експерименту з розпізнавання впливу гербіциду [2], виміряні на сьомий день після обприскування. Рослини були поділені на три групи: контрольну групу, дві групи з різними дозами оброблення рослин гербіцидом «Раундап» (гліфосат). Для тестування використовувалася двошарова нейронна мережа з 89 входами, 25 нейронами в прихованому шарі та трьома у вихідному шарі. У прихованому шарі використовувалася сигмоїдна функція активації, а у вихідному – софтмакс функція.

В табл. 1 наведено максимальну точність розпізнавання в % (A_m) та середню точність розпізнавання ($\overline{A}$), що отримані після 100 повторних навчань і тестувань нейронної мережі.

Точність розпізнавання обчислювалась за формулою

$$A = \frac{N_p}{N_T} * 100,$$

де N_p – кількість правильно класифікованих кривих, N_T – загальна кількість кривих.

Як видно з таблиці, найкращі результати дають мінімаксний спосіб нормалізації у межах [–1,1] – формула (3) та *z*-оцінка – формула (4).

ТАБЛИЦЯ 1. Результати тестування методів нормалізації

Спосіб нормалізації	A_m,%	$\overline{A}$,%
Без нормалізації	71,4	56
Формула 1	66.7	54.4
Формула 2	76,2	61,3
Формула 3	80,9	62,0
Формула 4	80,9	62,9
Формула 5	71,4	57,5
Формула 6	76,2	64,3
Формула 7	76,2	57,0
Формула 8	71,4	57,8
Формула 9	66,7	54,2

Зменшення розмірності даних вимірювань хлорофіл-флуорометрів. Можна використати ряд способів зменшення кількості входів нейронної мережі: 1) застосування геометричних параметрів ІФХ (як F_0, F_m, F_s і т. п.), що є поширеним підходом при роботі з кривими ІФХ; 2) вибір значень ІФХ за нелінійною шкалою; 3) використання коефіцієнтів апроксимуючих поліномів; 4) застосування методу головних компонент.

Вищезгадані методи мінімізації даних досліджено на задачі класифікації різних видів рослин. Зібрано набір кривих для 6 видів рослин: сої, лободи, фікуса еластіка, молочая ребристого, фікуса Бенджаміна, цинії. Проводилося вимірювання кривих 5 хв та 10 с. ІФХ рослин вимірювались у тіні, темнова адаптація становила 5 хв. В результаті зібрано 176 кривих. Для класифікації використовувалася мережа з прямим поширенням сигналів. Нейронна мережа містить 89 входів, прихований шар – 25 нейронів, вихідний шар – 6 нейронів. В прихованому шарі використовується сигмоїдна функція активації, а на вихідному – нормалізована експоненційна функція (softmax function). В роботі [6] показано, що криві розроблених сенсорів ІФХ можна використовувати для таксономічного розрізнення рослин, 5-хвилинні криві дещо краще підходять для даної задачі. В даній статті пропонується дослідити можливість мінімізації входів нейронної мережі.

Степенева шкала. Оскільки зміна ІФХ за часом нелінійна, значення ІФХ у приладах сімейства «Флоратест» вимірюються за нелінійною степеневою шкалою. Таким чином криві ІФХ у розроблених сенсорах дискретизовано з викори-

станням 90 точок. Для того, щоб визначити моменти вимірювання, в які прилад повинен виміряти значення ІФХ, спочатку реальну тривалість вимірювання переводять у нелінійне значення за наступними формулами:

$$\tau = t^{1/8}, \tag{10}$$

де t – реальний час в мс; τ – нелінійний «час».

Моменти відліку інтенсивності флуоресценції хлорофілу в такому нелінійному часі беруться так:

$$\tau_1 = 1;\ \tau_2 = 1 + \Delta\tau\ ;\tau_2 = 1 + 2 * \Delta\tau\ ;\ldots;\ \tau_i = 1 + \Delta\tau * (i-1), \tag{11}$$

$$\Delta\tau = \frac{(\tau_N - 1)}{N}, \tag{12}$$

де N – кількість необхідних відліків.

Відповідні моменти справжнього часу для розрахованих відліків визначають за формулою

$$t_i = \tau_i^8. \tag{13}$$

Маючи виміряні значення для 89 точок (перше значення вимірювання ІФХ було вирішено відкинути), використовуючи формули 10 – 13 та інтерполяцію можна представити криві меншою кількістю значень від 2 до 88. Результати тестування нейронної мережі в залежності від кількості взятих значень ІФХ показано на рис. 2.

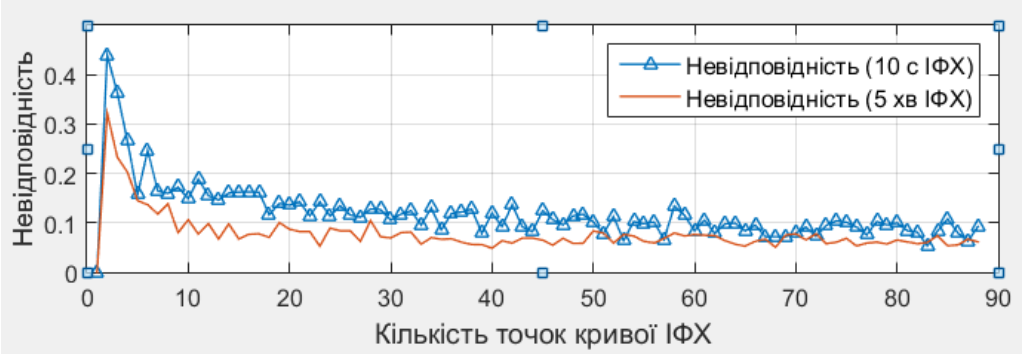


РИС. 2. Частка хибних розпізнавань у залежності від кількості точок на кривій ІФХ, взятих за степеневою шкалою

Таким чином, за усередненими значеннями очевидно, що для отримання якості розпізнавання більше 90 % для 5-хвилинної кривої ІФХ достатньо 9 точок, водночас як 10-секундна крива для задачі розпізнавання виду рослин потребує більше 30 точок.

Використання коефіцієнтів апроксимуючого полінома. Криву ІФХ можна замінити поліномом певного степеня вигляду:

$$Q(t) = k_0 + k_2 t + k_3 t^2 + ... + k_n t^n,$$

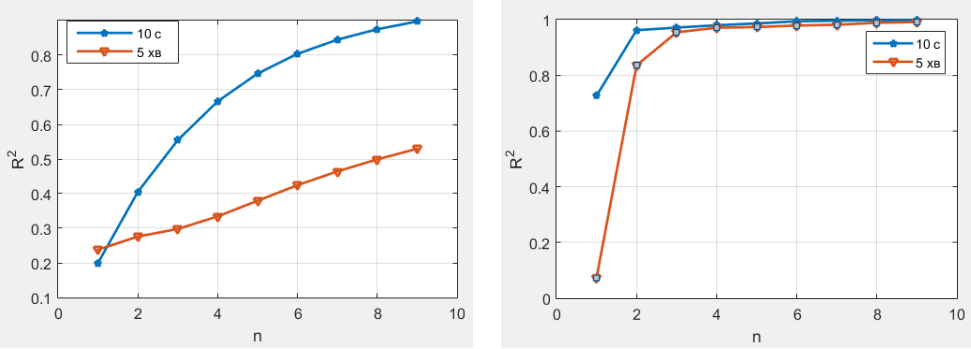
де k_i – коефіцієнти полінома, $i = 1..n$, n – степінь полінома, T – час вимірювання значення ІФХ.

Коефіцієнти полінома будуть індивідуальні для кожної виміряної кривої ІФХ. Підбір коефіцієнтів полінома здійснюється методом найменших квадратів. Більшість сучасних програмних пакетів для математичних обчислень дозволяють швидко обчислити коефіцієнти полінома до дев'ятого степеня.

На рис. 3, а показано залежність коефіцієнта детермітації від степеня полінома, що були отримані з використанням однієї з кривих вибірки. Як видно з рисунка, 10-секундні криві ІФХ апроксимовано поліномом 9 степеня з коефіцієнтом детермінації 0,9. 5-хвилинні криві апроксимуються гірше. Покращити апроксимацію 5-хвилинних кривих можна за допомогою кускової апроксимації, використавши декілька поліномів, проте це збільшить кількість коефіцієнтів, що подаватимуться на вхід нейронної мережі. Ще одним виходом є замінити часові значення порядковими значеннями відліків $x \in Z$, $x = 1..89$ в нелінійній шкалі:

$$Q(x) = k_0 + k_2 x + k_3 x^2 + ... + k_n x^n.$$

В такому разі апроксимація проходить значно краще (рис. 3, б).



а б

РИС. 3. Залежність коефіцієнта детермінації від степеня полінома: а – з використанням реального часу; б – з використанням порядкових номерів дискретних значень ІФХ в нелінійній шкалі

Метод головних компонент – один з найбільш застосовуваних методів зменшення розмірності даних в машинному навчанні, дозволяє здійснити мінімізацію даних шляхом переходу до нових змінних головних компонент, PC (principal components) таких, що будуть враховувати мінливість даних (дисперсію). В результаті вимірювання ІФХ, що складається з 89 дискретних значень, можна замінити меншою кількість головних компонент. На рис. 4 показано виміряні дані за трьома головними компонентами. Очевидно, що дані добре групуються уже при використанні трьох компонент.

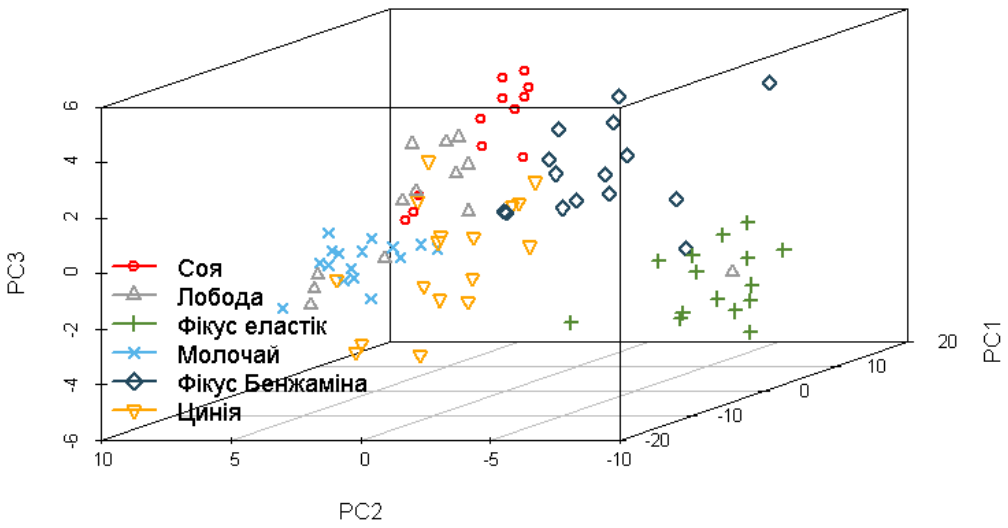


РИС. 4. Результат візуалізації за трьома головними компонентами виміряних даних ІФХ 6-ти видів рослин

На рис. 5 показано результат тестування нейронної мережі за різною кількістю взятих компонент. Очевидно, що найкращі результати отримуються при подачі на вхід нейронної мережі від 7 до 12 компонент.

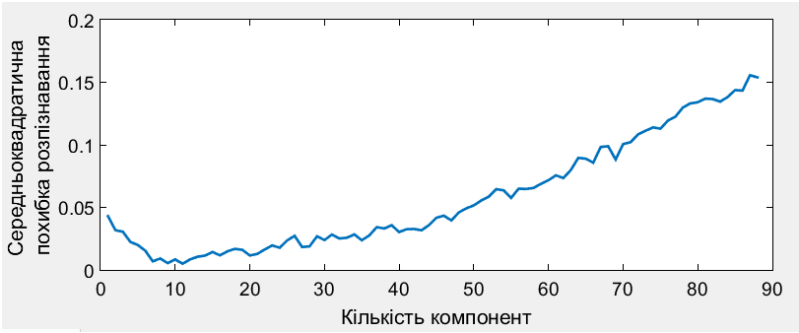


РИС. 5. Залежність середньоквадратичної похибки розпізнавання від кількості взятих компонент (для 5-хвилинних кривих)

В табл. 2 наведено результати тестування розглянутих способів мінімізації в результаті 100 повторних тестувань, де Pmin- мінімальна середньоквадратична похибка, Pmean – усереднена середньоквадратична похибка, Pmax – максимальна середньоквадратична похибка, Cmin – мінімальна частка кількості помилко-

вих розпізнавань, Cmean – середня частка кількості помилкових розпізнавань, Cmax – максимальна частка кількості помилкових розпізнавань.

Найкращий результат демонструє нейронна мережа, вхідним вектором якої є 89 нормалізованих значень ІФХ. Метод головних компонент забезпечує найкращу мінімізацію даних, дещо гірший результат дає подальше взяття значень на кривій за нелінійною шкалою та використанні 13 розрахованих параметрів ІФХ. Щоправда, при усіх трьох способах відбулась невелика втрата середньої точності розпізнавання (2 – 3 %). Використання коефіцієнтів полінома у вхідному векторі нейронної мережі значно погіршує результат.

ТАБЛИЦЯ 2. Результати мінімізації для класифікації рослин

Складові вхідного вектора	Pmin	Pmean	Pmax	Cmin	Cmean	Cmax
$F_1,\ldots,F_{89}$ ненормалізовані значення	2,35e-15	0,0275	0,1552	0	0,0633	0,6364
$F_1,\ldots,F_{89}$, нормалізовані значення	7,54e-16	0,0103	0,0943	0	0,0273	0,8636
Fo, Fj, Fi, Fm, Fs	9,24e-09	0,047	0,2803	0	0,1303	0,8636
Fo, Fj, Fi, Fm, Fs, Fv, k1, k2, Rfd, Area, Vj, Vi, Vs	2,77e-07	0,0276	0,1219	0	0,0597	0,3523
$PC_1,\ldots,PC_7$	6,38e-18	0,0147	0,1741	0	0,0533	0,6818
$F_1,\ldots,F_{10}$ значення ІФХ отримані за степеневою шкалою	1,01e-06	0,0254	0,1844	0	0,0691	0,8636
$k_i(i=1,\ldots,10)$ полінома P(x)	8,42e-06	0,0406	0,211	0	0,1299	0,9205
$k_i(i=1,\ldots,10)$ полінома P(t)	0,01	0,0886	0,2532	0,0568	0,283	0,9205

Визначення водного дефіциту шляхом апроксимації даних нейронною мережею. З метою вивчення зміни параметрів ІФХ при водному дефіциті використано рослини Цинії, які були поділені на три варіанти: з надмірним поливом, з помірним поливом та відсутністю поливу. Протягом двох тижнів реєструвалися зміни ІФХ у дослідних рослин. Зібрані дані включають криві ІФХ (F), вологість ґрунту (Hgr), кислотність ґрунту, температуру ґрунту та температуру повітря (Tair). Оскільки кислотність ґрунту корелює з вологістю ґрунту, а температура ґрунту з вологістю повітря, дані параметри не враховувалися у моделі ней-

ронної мережі. Також з віком рослини ІФХ теж може мінятися, тож у вхідний вектор також включено порядковий номер дня експерименту (D). Усі значення нормувались. Для кривих ІФХ використовувалась формула 4, а Tair та D нормувались шляхом ділення на максимальне значення. Таким чином нейронна мережа здійснила апроксимацію такого вигляду:

$$\varphi(F, Tair, D) = Hgr\ .$$

Для цих цілей використано трьохшарову нейронну мережу з прямим поширенням сигналів з 40 входами на які подавались значення ІФХ (обчислені з початково виміряних кривих за степеневою шкалою, з використанням інтерполяції), з сигмоїдною функцією активації у двох прихованих шарах (25 та 8 нейронів) та лінійною функцією активації на вихідному шарі (1 нейрон). Початково від даних випадковим чином було виділено 75 (із 335 всього) кривих ІФХ для тестування. Решта даних застосовувалися при навчанні із застосуванням п'ятикратної крос-валідації. Найкращий досягнений коефіцієнт кореляції *R* становив 0,78. Один з результатів навчання показано на рис. 6.

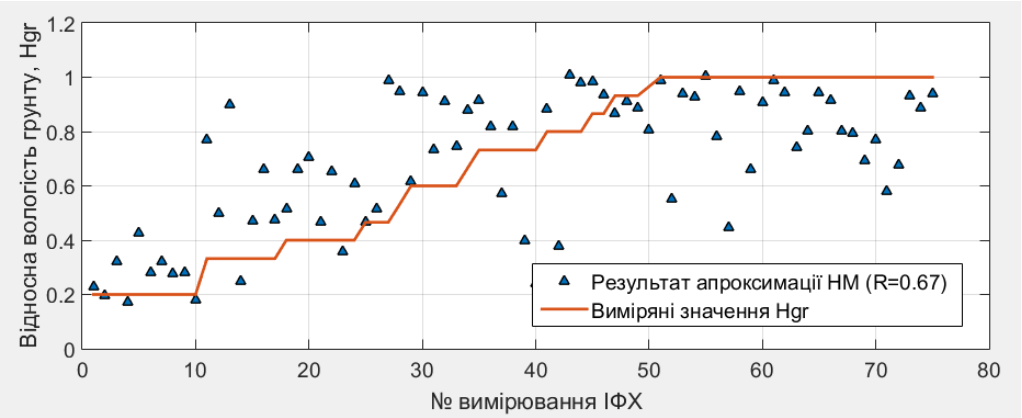


РИС. 6. Результати апроксимації нейронною мережею

Оскільки в даних наявна значна варіація, особливо у рослинних об'єктів з помірною вологістю, нейронна мережа найкраще здійснює апроксимацію даних, які відповідають крайнім значенням вологості ґрунту. Криві ІФХ, виміряні з проміжними значеннями вологості апроксимуються гірше, що можна пояснити тим, що вологість різних листів рослини знижувалась по різному. Крім того, Цинія є посухостійкою рослиною, що також могло відбитися на результатах обробки експериментальних даних та присутні похибки вимірювання вологості ґрунту. Більш кращі результати навчання нейронних мереж досягаються на основі безпосередньо вимірювання відносного вмісту води у листі, як це показано у [7], проте це потребує додаткових засобів та процедур і не усуває неоднорідності вологи у різних листах на рослині. Очевидно, що необхідно здійснювати вимірювання кількох кривих ІФХ, після чого доцільно приймати рішення так

званим «методом голосування», наприклад, коли результат умови Hgr < 0,5 справджується для більшості апроксимованих значень.

Тестування нейронної мережі при різній кількості взятих головних компонент показало, що найменша середньоквадратична похибка досягається при включенні у вхідний вектор 24 – 29 головних компонент.

Висновки. При попередній обробці даних ІФХ найкращі результати демонструє нормалізація шляхом z-оцінки та мінімаксна нормалізація. Найкращий результат при зменшенні розмірності експериментальних даних демонструє метод головних компонент, дещо гірші результати отримані при застосуванні степеневої шкали та використання параметрів ІФХ. Проте усі три згадані способи демонструють незначне зниження середньої точності розпізнавання на 2 – 3 %. НМ демонструє кращі результати апроксимації значень вологості ґрунту на основі ІФХ, які відповідають крайнім значенням вологості.

1. Palagin O., Romanov V., Galelyuka I., Hrusha V., Voronenko O. Wireless smart biosensor for sensor networks in ecological monitoring. Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 21–23 September, 2017. Bucharest, Romania.
2. Груша В.М. Обробка результатів експериментальних досліджень, проведених з використанням портативного флуорометра "Флоратест". *Комп'ютерні засоби, мережі та системи*. 2015. № 14. С. 109 – 116.
3. Kalaji H.M., Schansker G., Brestic M. at al. Frequently asked question about chlorophyll fluorescence, the sequel. Photosynthesis Research. Vol. 132, Issue 1, Springer, 2017. P. 13 – 66.
4. Strasser R.J., Srivastava A., Tsimilli-Michael M. Analysis of the chlorophyll a fluorescence transient ,in: G. Papageorgiou, Govindjee (Eds.), Chlorophyll a Fluorescence: A Signature of Photosynthesis, Advances in Photosynthesis and Respiration, Vol. 19, Kluwer Academic Publishers. 2004. P. 321–362.
5. Xanthoula Eirini Pantazi, Dimitrios Moshou, Dimitrios Kasampalis and Pavlos Tsouvaltzis. Automatic Accessment of Phenotypes in lettuce plants by using Chlorophyll Fluorescence Kinetics and Machine Learning. Proceedings International Conference of Agricultural Engineering. AgEng 2014 Zurich 6-10.07.2014. P. 167–176.
6. Palagin O., Grusha V., Antonova H., Kovyrova O., Lavrentyev V. Application of biosensors for plants monitoring. International Journal "Information theories & applications". Vol. 24, N 2. 2017. P. 115–126.
7. Goltsev V. et al. Drought-induced modification of photosynthetic electron transport in intact leaves: Analysis and use of neural network as a tool for a rapid non-invasive estimation. Biochimica et Biophysica Acta 1817. 2012. P. 1490–1498.

Одержано 25.10.2017