

УДК 544.165:004.85:547-326+547.596+547.466:615.212

Є.К. Пестерев ^{а, б}, Л.М. Огніченко ^б, А.Г. Артеменко ^б, В.Є. Кузьмін ^б

УДОСКОНАЛЕННЯ МЕТОДУ НАЙБЛИЖЧИХ СУСІДІВ ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ МОЛЕКУЛ. АНАЛГЕТИЧНА АКТИВНІСТЬ ДЕЯКИХ ТЕРПЕНОЇДІВ

^а Одеський національний університет імені І.І. Мечникова, м. Одеса, Україна^б Фізико-хімічний інститут ім. О.В. Богатського НАН України, м. Одеса, Україна

Розроблено модифікований метод k найближчих сусідів (k -NN) при $k=2$ для оцінювання структурної подібності/відмінності молекул з метою прогнозування значень досліджуваної властивості X молекули m_x на основі відомих значень відповідних властивостей молекул $m_1, m_2, \dots, m_k$, розташованих у багатовимірному дескрипторному просторі. Запропоновано при розрахунку значення властивості X враховувати не лише відстані від точки m_x до точок $m_1, m_2, \dots, m_k$, а й кути між цією точкою і парами її найближчих сусідів. Для порівняння класичного та модифікованого методів 2-NN використано вибірку з 19 похідних терпеноїдів. Як досліджувану властивість розглядали аналгетичну активність сполук у чотирьох тестах: алілізотіоціанатному, формаліновому, капсаїциновому та тесті «гарячої пластини». Розрахунок структурних дескрипторів здійснювали на 2D рівні моделювання молекулярної структури в рамках симплексного підходу, при цьому для диференціації вершин симплексів використовували не лише мітки, що відображають природу атомів, а й інші атомні характеристики, зокрема ван-дер-ваальсові взаємодії, електронегативність, ліпофільність тощо. Порівняння статистичних характеристик моделей показало, що модифікований підхід забезпечує зменшення середньоквадратичної похибки на 15–17% порівняно з класичним 2-NN у всіх досліджених тестах, що підтверджує важливість урахування не лише відстаней до найближчих точок, а й їхнього взаємного розташування у просторі.

Ключові слова: k -NN, QSAR, симплексне представлення молекулярної структури, терпеноїди, аналгетична активність.

DOI: 10.32434/0321-4095-2026-167-4-82-89

Вступ

За останні роки значного розвитку набули дослідження у галузі кількісних співвідношень «структура–активність/властивість» (QSAR/QSPR) органічних сполук. Це насамперед зумовлено тим, що встановлення таких залежностей дає змогу заощаджувати час і ресурси при створенні нових матеріалів, реагентів, препаратів тощо, а також

сприяє глибшому розумінню механізмів їх функціонування. Використання комп'ютерних технологій, зокрема, відкриває можливість на основі встановлених закономірностей здійснювати не лише попередній відбір (скринінг), а й молекулярний дизайн органічних сполук із заданим комплексом корисних властивостей.

Питання підвищення прогнозуючої здатності

© Є.К. Пестерев, Л.М. Огніченко, А.Г. Артеменко, В.Є. Кузьмін, 2026



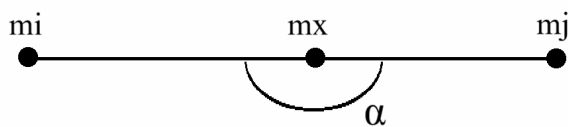
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

E.K. Pesterev, L.M. Ognichenko, A.G. Artemenko, V.E. Kuz'min

методів машинного навчання завжди залишається актуальним. Одним із популярних підходів у машинному навчанні є метод k найближчих сусідів (k -NN – k Nearest Neighbours) [1], який використовують для прогнозування значень досліджуваної властивості об'єкта (молекули) на основі відомих значень відповідних властивостей схожих, тобто найближчих до цього об'єкта, k об'єктів (молекул). Метод k -NN вирізняється концептуальною простотою, що забезпечує високу швидкість обчислень, особливо за помірно великої кількості об'єктів. Метод базується на припущенні, що близьким об'єктам у дескрипторному просторі відповідають подібні значення цільової властивості. Для оцінки подібності молекул аналізують їх розподіл у структурному (дескрипторному) просторі. При цьому значення структурних дескрипторів визначають координати точок, що відповідають молекулам, у просторі, осями якого є відповідні дескриптори. Для оцінки близькості між точками використовують різні метрики, зокрема евклідову. Класичний метод k -NN враховує відстані та прогнозує невідоме значення властивості X для точки m_x , використовуючи значення цієї властивості для найближчих до неї точок $m_1, m_2, \dots, m_k$.

Головна ідея вдосконалення методу k -NN полягає в такому. Ми вважаємо, що оцінка лише відстаней від досліджуваної точки (молекули) у структурному просторі до референтних точок (сусідів) є недостатньою. Важливо також враховувати взаємне розташування цих точок, що може істотно підвищити точність прогнозу.

Так, наприклад, для методу k -NN, де $k=2$, ідеальною є ситуація, коли досліджувана точка m_x та референтні точки (m_i, m_j) лежать на одній прямій:



кут $\alpha(m_i, m_x, m_j) = 180^\circ$.

Зрозуміло, що референтні точки необхідно шукати серед найближчих сусідів ($d(m_i, m_x), d(m_j, m_x) \rightarrow \min$). Потрібно встановити розумне порогове значення (cutoff value) $d(m_i, m_x), d(m_j, m_x) \leq d(cv)$. У наших моделях використовується значення $d(cv)$, яке є меншим за середню відстань між точками робочої вибірки.

Для гострого кута неминучі помилки розрахунку (див. нижче). Граничне значення кута в наших моделях становить $\alpha \geq 120^\circ$.

Розвиваючи далі запропоновану методологію, можна зробити висновок, що для $k=3$ іде-

альною є ситуація, коли досліджувана точка знаходиться всередині трикутника референтних точок; для $k=4$ – всередині тетраедра тощо.

У цій роботі ми обмежилися моделями з $k=2$.

Для дослідження ефективності і апробації удосконаленого 2-NN підходу було використано вибірку терпеноїдів з експериментальними даними щодо їхньої аналгетичної активності [2]. Проблема розроблення лікарських препаратів комплексної дії, тобто сполук, здатних одночасно взаємодіяти з різними типами рецепторів, посідає чільне місце серед пріоритетних напрямів розвитку сучасної медицини та фармакології. Відомо, що терпеноїди та їх похідні можуть взаємодіяти з каналами транзитного рецепторного потенціалу (TRP-іонними каналами), рецепторами γ -аміномасляної кислоти (GABA), ацетилхолінергетичною, натрієвими каналами, системами вторинних месенджерів і низкою ферментів, що може зумовлювати їхню аналгетичну активність [3–5].

Таким чином, предметом цього дослідження було розроблення, реалізація та апробація модифікованого методу 2-NN на вибірці терпеноїдів та їх естерів.

Матеріали і методи

Найпоширенішим розширенням методу k найближчих сусідів є використання зважування при усередненні, за якого внесок кожного сусіда залежить від відстані до досліджуваної сполуки, для якої виконується прогнозування. Як вагову функцію використовують функцію, обернено пропорційну відстані.

Значення властивості P нової молекули m_x розраховується як середньозважена величина на основі значень властивості найближчих референтних (властивість відома) молекул m_i (де $i=1\dots k$) та відповідних їхніх відстаней (d) в дескрипторному просторі до молекули m_x . Формула має вигляд:

$$P(m_x) = \frac{\sum_{i=1}^k \frac{P(m_i)}{d(m_x, m_i)}}{\sum_{i=1}^k \frac{1}{d(m_x, m_i)}}, \quad (1)$$

де $P(m_x)$ – прогнозоване значення властивості нової молекули; $P(m_i)$ – відоме значення властивості i -ї референтної молекули; $d(m_x, m_i)$ – відстань між новою та референтною молекулами у дескрипторному просторі; k – кількість найближчих сусідів.

У межах класичного 2-NN підходу враховується одна пара з найближчих референтних точок (молекул), незалежно від кута між цією па-

рою та точкою (молекулою) m_x . Унаслідок цього можлива ситуація, коли m_x розташована в зовсім іншому діапазоні значень властивостей порівняно з обраною парою найближчих точок (m_1 і m_2), а кут, утворений цією парою точок і m_x , є гострим (рис. 1а).

У межах запропонованого модифікованого підходу така пара точок у подібній ситуації не використовується для розрахунку властивості m_x . Натомість здійснюється пошук іншої пари найближчих точок (m_3 і m_4), для яких кут між ними та точкою m_x є максимально наближеним до 180° (рис. 1б).

Косинус кута α у n -мірному просторі розраховується через відношення скалярного добутку векторів між парою точок (m_A , m_B) та m_x до добутку відстаней до відповідних точок:

$$\cos \alpha = \frac{\sum_{i=1}^n ((m_x - m_A)_i \times (m_x - m_B)_i) / (d(m_x, m_A) \times d(m_x, m_B))}{(2)}$$

Слід зазначити, що, як правило, для точок, які розташовані на периферії структурного простору (далі «периферійні точки»), можлива ситуація, коли неможливо знайти пару з тупим кутом. У такому випадку для даних точок (молекул) прогноз здійснюється за класичним підходом.

Раніше нами було проведено QSAR-дослідження анагетичної активності 19 похідних терпеноїдів [2] (табл. 1) із використанням симплексного підходу до опису молекулярної структури. У даному дослідженні зазначену вибірку використано для апробації модифікованого 2-NN підходу та його порівняння з класичним 2-NN підходом. Як досліджувані властивості розглядалися час реакції у тестах з алізіотіо-

ціанатом (АІТЦ-тест), формаліном (Ф-тест), капсаїцином (К-тест) та латентний час у тесті «гарячої пластини» (Н-тест) (у секундах).

Для розрахунку структурних дескрипторів було використано симплексний метод представлення молекулярної структури, який добре зарекомендував себе при розв'язанні низки різноманітних QSAR/QSPR завдань [6]. Важливою особливістю симплексного підходу є те, що при ідентифікації вершин симплексів використовують не лише мітки, що відображають природу атомів, а й різні вагові параметри (властивості) атомів, зокрема, електронегативність (E_n), електронну поляризованість (R_f), характеристики ван-дер-ваальсових взаємодій (ВДВ) універсального силового поля [7,8] (параметри притягнення ($Attr$) та відштовхування (Rep)), атомний заряд (Chg), ліпофільність (Lip).

Для атомних характеристик, які мають реальні значення, діапазон усіх можливих значень попередньо розбивається на певну кількість дискретних груп. Мітка атома визначається з точки зору його належності до певної групи. При розрахунку симплексних дескрипторів були використані такі діапазони поділу значень атомних властивостей на інтервали:

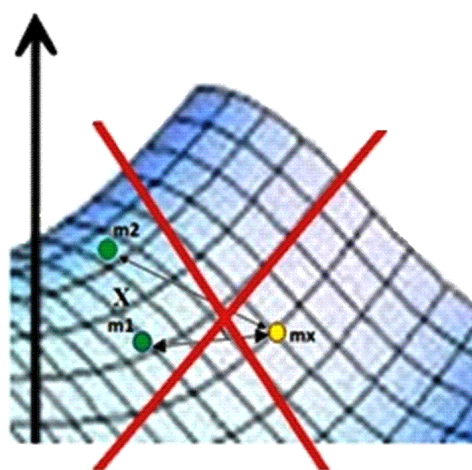
E_n : $A < 2,19 \leq B < 2,5 \leq C < 3,0 \leq D$;

R_f : $A < 1,5 \leq B < 3,0 \leq C < 8,0 \leq D$;

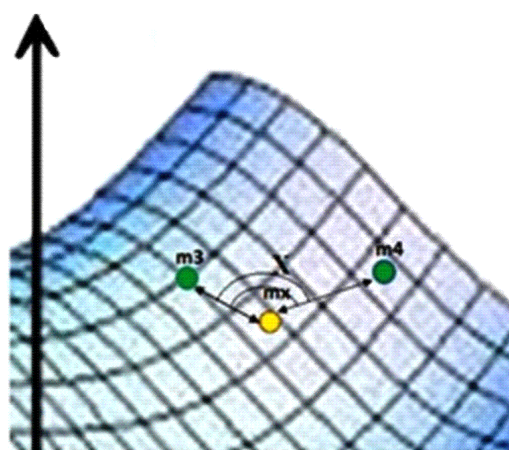
$Attr$:

$A < 50 \leq B < 100 \leq C < 250 \leq D < 400 \leq E < 650 \leq F < 2000 \leq G$;

Rep : $A < 20000 \leq B < 32000 \leq C < 50000 \leq D < 100000$;



а



б

Рис. 1. Розраховане X та реальне (●) розташування молекули m_x на модельній поверхні властивості

Таблиця 1

Структури та анагетична активність терпеноїдів

Структурний остов	№	R	тест			
			АІПЦ	Ф	К	Н
			Реакційний час, с			Латентний час, с
	1	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$23 \pm 4,3$	$31 \pm 8,3$	$12 \pm 1,2$	$19 \pm 3,8$
	2	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$15 \pm 0,9$	$29 \pm 6,9$	$14 \pm 0,8$	$21 \pm 2,5$
	3	$-\text{OH}$	$35 \pm 2,2$	$38 \pm 3,8$	$23 \pm 6,4$	$19 \pm 1,9$
	4	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$20 \pm 5,8$	$52 \pm 6,6$	$20 \pm 2,6$	$21 \pm 0,9$
	5	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$13 \pm 0,9$	$47 \pm 1,3$	$16 \pm 3,2$	$47 \pm 1,3$
	6	$-\text{OH}$	$25 \pm 3,8$	$50 \pm 5,2$	$35 \pm 4,1$	$15 \pm 0,3$
	7	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$3 \pm 0,3$	$30 \pm 7,4$	$4 \pm 0,9$	$30 \pm 2,6$
	8	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$16 \pm 4,6$	$19 \pm 3,7$	$11 \pm 1,5$	$29 \pm 6,5$
	9	$-\text{OH}$	$7 \pm 1,2$	$24 \pm 5,6$	$11 \pm 2,7$	$24 \pm 3,9$
	10	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$22 \pm 3,1$	$32 \pm 4,8$	$17 \pm 2,2$	$26 \pm 2,9$
	11	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$25 \pm 1,2$	$30 \pm 5,6$	$18 \pm 0,9$	$32 \pm 3,8$
	12	$-\text{OH}$	$30 \pm 2,2$	$40 \pm 3,3$	$20 \pm 3,4$	$21 \pm 3,1$
	13	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$25 \pm 1,5$	$47 \pm 9,8$	$24 \pm 3,3$	$17 \pm 3,7$
	14	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$30 \pm 8,3$	$35 \pm 1,9$	$20 \pm 3,6$	$23 \pm 2,8$
	15	$-\text{OH}$	$21 \pm 2,8$	$54 \pm 6,8$	$23 \pm 2,7$	$20 \pm 0,5$
	16	$-\text{OOC}-(\text{CH}_2)_3-\text{NH}_2$	$23 \pm 3,0$	$23 \pm 0,8$	$15 \pm 3,0$	$46 \pm 2,2$
	17	$-\text{OOC}-\text{CH}_2-\text{NH}_2$	$20 \pm 1,2$	$19 \pm 1,5$	$15 \pm 1,3$	$49 \pm 0,9$
	18	$-\text{OH}$	$8 \pm 3,5$	$26 \pm 4,7$	$12 \pm 5,0$	$27 \pm 2,8$
	19	—	$48 \pm 2,0$	$36 \pm 2,4$	$29 \pm 6,6$	$18 \pm 0,9$

Chg:

$A < -0,16 \leq B < -0,08 \leq C < 0,00 \leq D < 0,08 \leq E < 0,16 \leq F < 0,23 \leq G$;

Lip:

$A < -1,53 \leq B < -1,00 \leq C < -0,47 \leq D < 0,05 \leq E < 0,58 \leq F < 1,11 \leq G$.

Крім того, всі атоми, що відповідають вершинам симплексів, були класифіковані за мітками на три групи: D – донори потенційного Н-зв'язку, A – акцептори потенційного Н-зв'язку та I – індіферентні атоми.

Як дескриптори також використовували інтегральні параметри молекули, такі як молекулярна маса, молекулярна рефракція, сумарна електронегативність, ліпофільність. Для кожної молекули разом з симплексами (чотирьохатомними фрагментами) розраховували також фрагменти меншого розміру – двійки та трійки.

Розрахунок симплексних дескрипторів здійснювали на 2D рівні моделювання молекулярної структури. Таким чином, для кожної досліджуваної сполуки було розраховано 2051 структурний дескриптор.

Після обчислення структурних дескрипторів було сформовано чотири окремі вибірки – по одній для кожного тесту. На попередньому етапі здійснювали попередню обробку дескрипторів для кожної вибірки: вилучали константні змінні та дескриптори з попарною кореляцією вище порогового значення ($R=0,95$). Додатково застосовано процедуру тренд-вектору [9] для формування початкових наборів дескрипторів, на основі яких були побудовані PLS моделі з двома латентними змінними (A1, A2). Отримані латентні змінні надалі використовувалися як дескриптори відповідних сполук у 2-NN методах.

Для оцінювання якості моделей одним із найпоширеніших статистичних показників є коефіцієнт детермінації R^2 , який показує яку частку загальної дисперсії (варіації) даних пояснює модель:

$$R^2 = \frac{SS - RSS}{SS},$$

де RSS (Residual Sum of Squares) – сума квадратів

залишків: $RSS = \sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{obs}})^2$; y_i^{pred} і y_i^{obs} –

значення властивості, що розраховане за допомогою побудованої моделі, і спостережуване значення властивості для i -об'єкту навчальної вибірки відповідно; N – кількість об'єктів навчальної вибірки; SS (Sum of Squares) – загальна сума квадратів відхилень значень залежної змінної від

її середнього значення: $SS = \sum_{i=1}^N (y_i^{\text{obs}} - \bar{y})^2$; $\bar{y}$ –

середнє значення властивості для вибірки з N спо-

$$\text{лук: } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i^{\text{obs}}.$$

Також важливим статистичним показником є середньоквадратична похибка (RMSE), яка обчислюється за формулою:

$$RMSE = \sqrt{\frac{RSS}{N}}.$$

Результати та їх обговорення

У цій роботі для алілізотіоціанатного, формалінового, капсаїцинового тестів та тесту «гарячої пластини» проведено аналіз результатів прогнозу аналгетичної активності для вибірки терпеноїдів із використанням класичного 2-NN підходу та розробленого модифікованого підходу. У табл. 2 наведено статистичні параметри побудованих 2-NN моделей для кожного з тестів.

Порівняння статистичних характеристик розроблених моделей демонструє найкращий результат для модифікованого підходу. Зокрема середньоквадратична похибка для модифікованого підходу є приблизно на 15% меншою, ніж для класичного.

Для кожного тесту на рис. 2 наведено розподіл молекул вибірки в дескрипторному просторі латентних змінних. Точки, позначені

Таблиця 2

Статистичні характеристики 2-NN-моделей

Тест	Класичний підхід		Модифікований підхід		Покращення
	R^2	RMSE	R^2	RMSE	$y\%$ RMSE%
АІТЦ-тест	0,74	5,6	0,82	4,8	15
Формаліновий тест	0,91	3,4	0,94	2,8	16
Тест «гарячої пластини»	0,82	4,3	0,88	3,6	17
Капсаїциновий тест	0,72	3,7	0,80	3,2	15

чорним кольором, є периферійними точками — для них не знайдено жодної пари найближчих сусідів, що відповідають заданим умовам. Для алілізотіюанатного, формалінового тестів і тесту «гарячої пластини» виявлено чотири периферійні точки (рис. 2а,б,г), для капсаїцинового — три (рис. 2в). При оцінюванні якості отриманих моделей значення прогнозованої властивості для периферійних точок розраховували за класичним методом.

Діаграми абсолютних похибок прогнозу класичного та модифікованого 2-NN підходів для досліджуваних тестів наведено на рис. 3.

Таким чином, як видно з табл. 2 та рис. 3, модифікований підхід забезпечує кращі результати прогнозу анальгетичної активності похідних терпеноїдів для всіх досліджених тестів. Інтерпретація моделей у цій роботі не проводилася, оскільки вона була виконана раніше [2], а головним завданням цієї роботи є проведення по-

рівняльного аналізу класичного та модифікованого 2-NN підходів.

Висновки

Порівняння статистичних характеристик прогнозу анальгетичної активності за класичним та модифікованим підходами найближчих сусідів 2-NN у всіх випадках продемонструвало перевагу розробленого модифікованого підходу, що проявляється також у зменшенні середньоквадратичної похибки на 15–17%. Таким чином, отримані результати підтверджують доцільність урахування не лише найближчих точок за відстанню, але й їхнього взаємного розташування у структурному просторі. Розроблений підхід може бути надалі застосований до більших вибірок або інших класів сполук.

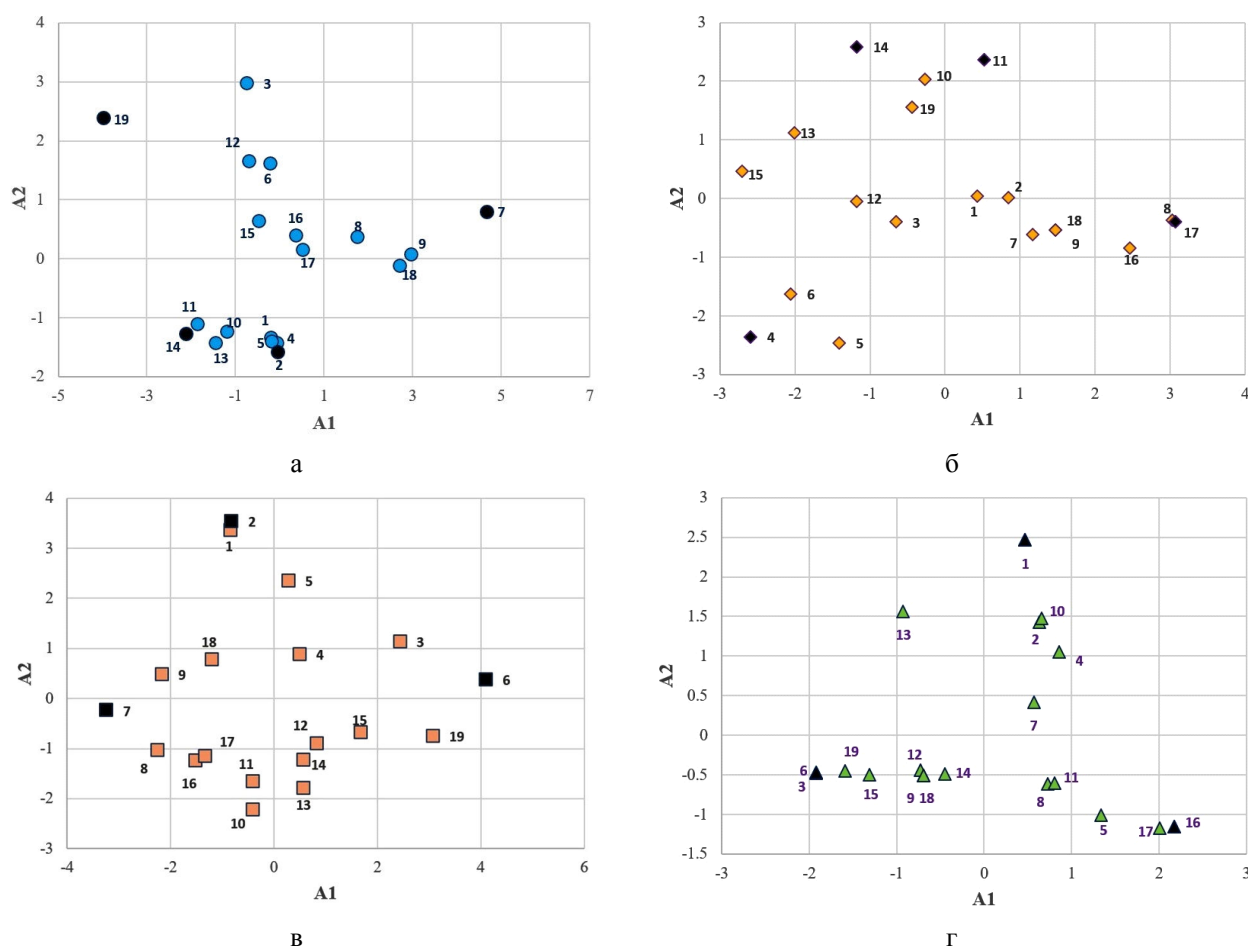


Рис. 2. Розподіл терпеноїдів в дескрипторному просторі латентних змінних (A1, A2) для тестів з:
а — алілізотіюанатом; б — формаліном; в — капсаїцином; г — тест «гарячої пластини»

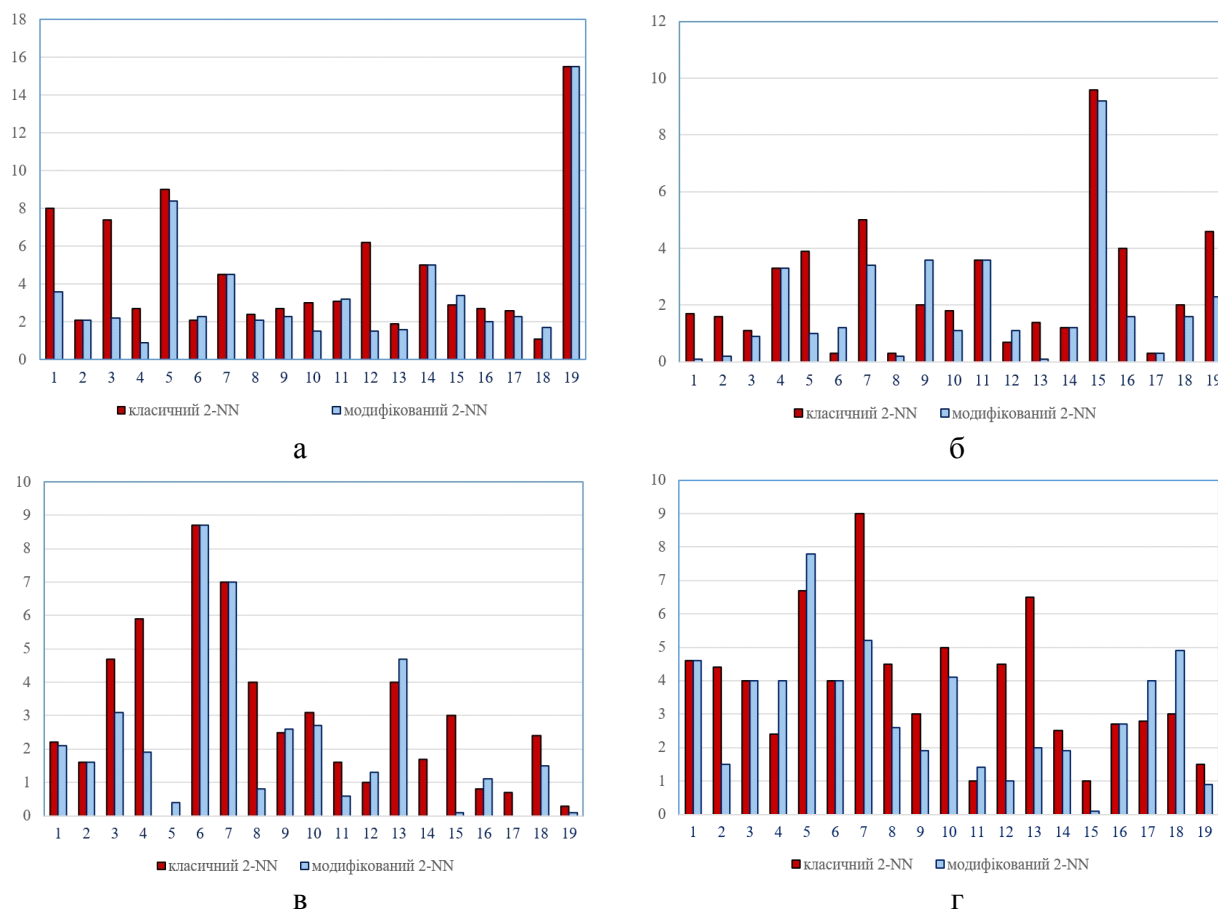


Рис. 3. Абсолютні похибки прогнозу анальгетичної активності терпеноїдів для тестів з:
а – алілізотіоціанатом; б – формаліном; в – капсаїцином; г – тест «гарячої пластини»

СПИСОК ЛІТЕРАТУРИ

1. Cover T., Hart P. Nearest neighbor pattern classification // IEEE Trans. Inf. Theory. – 1967. – Vol.13. – P.21-27.
2. QSAR models for analgesic activity prediction of terpenes and their derivatives / Nesterkina M., Ognichenko L., Shyrykalova A., Kravchenko I., Kuz'min V. // Struct. Chem. – 2020. – Vol.31. – P.947-954.
3. Plant terpenoids as the promising source of cholinesterase inhibitors for Anti-AD therapy / Lai Shi Min S., Liew S.Y., Chear N.J.Y., Goh B.H., Tan W.N., Khaw K.Y. // Biology. – 2022. – Vol.11. – No. 2. – Art. No. 307.
4. Repellent activity of monoterpene esters with neurotransmitter amino acids against yellow fever mosquito, *Aedes aegypti* / Nesterkina M., Bernier U.R., Tabanca N., Kravchenko I. // Open Chem. – 2018. – Vol.16. – P.95-98.
5. Russo E.B. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects // Br. J. Pharmacol. – 2011. – Vol.163. – P.1344-1364.
6. Simplex representation of molecular structure as universal QSAR/QSPR tool / Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., et al. // Struct. Chem. – 2021. – Vol.32. – No. 4. – P.1365-1392.
7. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations / Rappe A.K., Casewit C.J., Colwell K.S., Goddard III W.A., Skiff W.M. // J. Am. Chem. Soc. – 1992. – Vol.114. – No. 25. – P.10024-10035.
8. Quantitative structure-property relationship analysis of critical properties of individual organic compounds / Pesterev E., Ognichenko L., Artemenko A., Kichova M., Muratov N., Kuz'min V. // Proceedings of Odessa Polytechnic University. – 2025. – Vol.1(71). – P.220-234.
9. Carhart R.E., Smith D.H., Venkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and applications // J. Chem. Inf. Comput. Sci. – 1985. – Vol.25. – P.64-73.

Надійшла до редакції 11.05.2026

Надійшла після виправлення 26.06.2026

Прийнята до публікації 07.07.2026

Опублікована 10.08.2026

IMPROVING THE NEAREST NEIGHBOR METHOD FOR PREDICTING MOLECULAR PROPERTIES. ANALGESIC ACTIVITY OF SOME TERPENOIDS

E.K. Pesterev^{a, b}, *L.M. Ognichenko*^b, *A.G. Artemenko*^b,
V.E. Kuz'min^{b, *}

^a Odesa I.I. Mechnikov National University, Odesa, Ukraine

^b O.V. Bogatsky Physico-Chemical Institute NAS of Ukraine, Odesa, Ukraine

* e-mail: theorchem@gmail.com

A modified 2-nearest neighbors (2-NN) method has been developed to assess the structural similarity/dissimilarity of molecules in order to predict the value of a studied property X for a molecule *mx* based on known values of the corresponding properties of molecules *m1*, *m2*, ..., *mk* located in a multidimensional descriptor space. It is proposed that, in calculating the value of property X, not only distances from point *mx* to points *m1*, *m2*, ..., *mk* should be taken into account, but also the angles formed by this point and pairs of its nearest neighbors. To compare the classical and modified 2-NN methods, a dataset of 19 terpenoid derivatives was used. The studied property was the analgesic activity of the compounds evaluated in four tests: allyl isothiocyanate, formalin, capsaicin, and the «hot plate» test. Using the simplex approach, the structural descriptors were calculated at the 2D level of molecular structure modeling. Not only labels reflecting the nature of atoms were used to differentiate simplex vertices, but also other atomic characteristics: van der Waals interactions, electronegativity, lipophilicity, etc. Comparison of the statistical characteristics of the models demonstrated that the modified approach provides a 15–17% reduction in root-mean-square error relative to the classical 2-NN method across all tests performed, thereby confirming the importance of accounting not only for the distances to the nearest points, but also for their mutual spatial arrangement.

Keywords: k-NN; QSAR; simplex representation of molecular structure; terpenoids; analgesic activity.

REFERENCES

1. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967; 13(1): 21-27. doi: 10.1109/TIT.1967.1053964.
2. Nesterkina M, Ognichenko L, Shyrykalova A, Kravchenko I, Kuz'min V. QSAR models for analgesic activity prediction of terpenes and their derivatives. *Struct Chem*. 2020; 31: 947-954. doi: 10.1007/s11224-019-01479-7.
3. Lai Shi Min S, Liew SY, Chear NJY, Goh BH, Tan WN, Khaw KY. Plant terpenoids as the promising source of cholinesterase inhibitors for anti-AD therapy. *Biology*. 2022; 11: 307. doi: 10.3390/biology11020307.
4. Nesterkina M, Bernier UR, Tabanca N, Kravchenko I. Repellent activity of monoterpenoid esters with neurotransmitter amino acids against yellow fever mosquito, *Aedes aegypti*. *Open Chem*. 2018; 16: 95-98. doi: 10.1515/chem-2018-0015.
5. Russo EB. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br J Pharmacol*. 2011; 163(7): 1344-1364. doi: 10.1111/j.1476-5381.2011.01238.x.
6. Kuz'min V, Artemenko A, Ognichenko L, Hromov A, Kosinskaya A, Stelmakh S, et al. Simplex representation of molecular structure as universal QSAR/QSPR tool. *Struct Chem*. 2021; 32: 1365-1392. doi: 10.1007/s11224-021-01793-z.
7. Rappe AK, Casewit CJ, Colwell KS, Goddard III WA, Skiff WM. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc*. 1992; 114(25): 10024-10035. doi: 10.1021/ja00051a040.
8. Pesterev E, Ognichenko L, Artemenko A, Kichova M, Muratov N, Kuz'min V. Quantitative structure-property relationship analysis of critical properties of individual organic compounds. *Proceedings of Odessa Polytechnic University*. 2025; 1(71): 220-234. doi: 10.15276/opu.1.71.2025.24.
9. Carhart RE, Smith DH, Venkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J Chem Inf Comput Sci*. 1985; 25: 64-73. doi: 10.1021/C100046A002.