

**МІНІСТЕРСТВО ОХОРОНИ ЗДОРОВ'Я
НАЦІОНАЛЬНИЙ МЕДИЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ О. О. БОГОМОЛЬЦЯ**

**ФАРМАЦЕВТИЧНИЙ ФАКУЛЬТЕТ
КАФЕДРА АНАЛІТИЧНОЇ, ФІЗИЧНОЇ ТА КОЛОЇДНОЇ ХІМІЇ**

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

**на тему «ПОРІВНЯННЯ МЕТОДІВ ПРОГНОЗУВАННЯ
ЛІПОФІЛЬНОСТІ МОЛЕКУЛ ЛІКАРСЬКИХ ЗАСОБІВ»**

Виконала: здобувачка вищої освіти 5 курсу
групи 98Ф2А

напряму підготовки 22 «Охорона здоров'я»
спеціальності 226 «Фармація, промислова
фармація»

освітньої програми «Фармація»

Дорош Анна Сергіївна

Керівник: кандидат хімічних наук, доцентка
Пушкарьова Я.М.

Рецензент: кандидат біологічних наук, доцентка
Кузнецова О.В.

КИЇВ-2024

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	4
ВСТУП	5
ОСНОВНА ЧАСТИНА.....	8
1. ЛІТЕРАТУРНИЙ ОГЛЯД	8
1.1. Значення ліпофільності на стадії розробки потенційних лікарських речовин.....	8
1.2. Методи експериментального визначення ліпофільності	9
<i>Висновки до розділу 1</i>	12
2. ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА	13
2.1. Масиви даних для дослідження.....	13
2.2. Опис штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі.....	18
<i>Висновки до розділу 2</i>	20
3. РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ.....	21
3.1. Прогностичні моделі з літературних джерел, що використовувалися для порівняння.....	21
3.2. Результати прогнозування ліпофільності із застосуванням мережі прямого поширення сигналу та каскадної мережі.....	22
3.2.1. Прогнозування ліпофільності із застосуванням 6 дескрипторів (енергія нульової точки; теплова енергія; ентальпія; вільна енергія; енергія активації; теплоємність).....	22
3.2.2. Прогнозування ліпофільності із застосуванням 7 дескрипторів (ентропія; енергія Гартрі-Фока; стеричний фактор; найвища зайнята молекулярна орбітальна енергія; найнижча незайнята молекулярна орбітальна енергія; молекулярна рефракція; коефіцієнт розподілу)	26
3.2.3. Прогнозування ліпофільності із застосуванням 2 дескрипторів (коефіцієнту розподілу та енергії нульової точки)	30
<i>Висновки до розділу 3</i>	35
ВИСНОВКИ.....	36

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	37
ДОДАТОК.....	40
SUMMARY	41

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

$\log P$	ліпофільність
MSE	середньоквадратична похибка
R	коефіцієнт кореляції

ВСТУП

Актуальність теми. Ліпофільність обов'язково враховується та оцінюється на етапі моделювання нових лікарських засобів, оскільки є однією із складових успішного застосування досліджуваної молекули як потенційного лікарського препарату. Існує багато експериментальних підходів щодо оцінки ліпофільності (метод струшування колби, метод вортекс рідинно-рідинної мікроекстракції, метод повільного перемішування, метод аспірації/ін'єкції водяної пробки, метод на основі потоку, метод на основі наноабсорбенту, метод на основі діалізу), але вони, зазвичай, трудомісткі, довготривалі та мають свої певні обмеження.

На початковому етапі досліджень із виявлення нових можливих лікарських засобів необхідно перевірити велику кількість хімічних молекул, щоб виявити сполуки з необхідною біологічною активністю. Цей процес займає багато часу через високий рівень невдач і збільшує вартість розробки ліків.

Через значну важливість параметра ліпофільності та його роль у розумінні фармакокінетичних властивостей препаратів-кандидатів виникла потреба в точних та простих розрахункових моделях (*in silico*) для його прогнозування. Таким чином, представлену роботу можна трактувати як актуальну та корисну.

Мета і завдання дослідження. Мета дослідження – порівняти ефективність різних підходів щодо прогнозування ліпофільності молекул лікарських засобів.

Досягнення поставленої мети зумовлює вирішення наступних завдань:

- 1) визначити оптимальний набір фізико-хімічних параметрів для прогнозування ліпофільності молекул лікарських засобів;
- 2) визначити оптимальну архітектуру штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі, що забезпечують високу точність прогнозування ліпофільності молекул лікарських засобів на

рівні інших методів прогнозування, знайдених в літературних джерелах;

3) оцінити прогностичну можливість запропонованих алгоритмів прогнозування ліпофільності молекул лікарських засобів.

Предмет дослідження: параметри архітектури штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі та вплив різних наборів фізико-хімічних параметрів на ефективність їх навчання та прогностичну силу.

Об'єкт дослідження: фізико-хімічні властивості молекул лікарських засобів.

Методи дослідження. Штучна нейронна мережа прямого поширення сигналу та каскадна нейронна мережа для прогнозування значення ліпофільності молекул лікарських засобів. Реалізацію вказаних алгоритмів штучних нейронних мереж виконано із застосуванням програмного комплексу Matlab R2023b (trial individual license 11937601).

Новизна та значення одержаних результатів. Встановлено, що алгоритми штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі є досить ефективними процедурами прогнозування ліпофільності молекул лікарських засобів. Оптимальна кількість нейронів прихованого шару для вказаних типів штучних нейронних мереж знаходиться в межах від чотирьох до дев'яти. Показано, що для ефективного та точного прогнозування ліпофільності достатнім є набір лише з двох дескрипторів: коефіцієнту розподілу та енергії нульової точки. Отримані результати дозволяють віддати перевагу каскадній нейронній мережі, що характеризується дещо точнішими результатами прогнозування, ніж нейронній мережі прямого поширення сигналу.

Алгоритми штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі є конкурентоспроможними по відношенню до методології QSAR.

Результати представленої роботи вносять певний вклад у розвиток розрахункових методів прогнозування ліпофільності молекул лікарських

засобів.

Апробація результатів дослідження. Результати цієї роботи були представлені на I International Scientific and Theoretical Conference “Scientific review of the actual events, achievements and problems”, 01.12.2023, Berlin, Federal Republic of Germany (сертифікат учасника IST № 23/1201-155).

Публікації. Оpubліковано статтю: Дорош, А., & Пушкарьова, Я. (2023). ПРОГНОЗУВАННЯ ЛІПОФІЛЬНОСТІ МОЛЕКУЛ ЛІКАРСЬКИХ ЗАСОБІВ ЗА ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ. *Collection of scientific papers «SCIENTIA»*, (December 1, 2023; Berlin, Germany), 304-309.

Структура роботи: 43 сторінки, 3 розділи, 28 використаних джерел.

ОСНОВНА ЧАСТИНА

1. ЛІТЕРАТУРНИЙ ОГЛЯД

1.1. Значення ліпофільності на стадії розробки потенційних лікарських речовин

Відкриття ліків – це складний і важкий процес. Щоб визначити, чи буде запропонована молекула успішним лікарським засобом, необхідно оцінити її ефективність і можливість лікування певних захворювань. На початковому етапі експериментальних досліджень із виявлення нових можливих лікарських засобів необхідно перевірити велику кількість хімічних молекул, щоб виявити сполуки, які демонструють певну хімічну активність. Цей процес займає багато часу через високий рівень невдач і збільшує вартість розробки ліків. Зі збільшенням відсотка невдач під час доклінічних і клінічних експериментів збільшується кількість досліджень щодо відкриття нових ліків. Сьогодні такі обчислювальні методи, як віртуальний скринінг, широко використовуються для зменшення вартості та часу виконання таких досліджень. Віртуальний скринінг полягає у перевірці (скринінгу) великих бібліотек малих молекул, і лише потім експериментально перевіряються сполуки, які, за прогнозами, будуть добре поводитися в біологічній системі [1–4].

Ліпофільність важлива для прогнозування успіху молекули як лікарського препарату, оскільки є показником метаболізму, фармакокінетики, фармакодинаміки та молекулярної токсичності. Ліпофільність виражається через $\log P$, де P – коефіцієнт розподілу молекули між *n*-октанолом і водою (двома розчинниками, які не змішуються). Висока ліпофільність ($\log P > 5$) органічної молекули вказує на підвищену здатність молекули долати гематоенцефалічний бар'єр. У той же час висока ліпофільність може вказувати на сполуки з низькою розчинністю і поганим всмоктуванням, що не є бажаним для будь-якого препарату. Якщо значення

занадто низьке, сполуки демонструють низьку проникність. Встановлено, що оптимальний діапазон ліпофільності сполук, які можуть бути успішними як ліки, має значення $\log P$ від 1 до 3 [1].

Коефіцієнт розподілу октанол/вода є мірою гідрофобності та гідрофільності речовини. Гідрофобні взаємодії мають вирішальне значення в багатьох галузях хімії, включаючи взаємодії ферментів і лігандів, взаємодії ліків і рецепторів, транспортування ліків до активного центру, збирання ліпідів у біомембранах, агрегацію поверхнево-активних речовин, коагуляцію та миючі властивості тощо. Гідрофобне «зв'язування» насправді взагалі не є утворенням зв'язків, а скоріше тенденцією гідрофобних молекул або гідрофобних частин молекул уникати води, оскільки вони не можуть легко пристосуватися до високовпорядкованої структури води з водневими зв'язками. Гідрофобна взаємодія є термодинамічно сприятливою через збільшення ентропії молекул води, яка супроводжує асоціацію неполярних молекул, які видавлюють воду [5, 6].

1.2. Методи експериментального визначення ліпофільності

Пряме визначення ліпофільності вимагає кількісного визначення концентрації сполуки, присутньої у неводній та водній фазах. Прямий метод, відомий як метод струшування колби, є золотим стандартом для визначення $\log P$ [7]. Прямі методи зазвичай точні при вимірюванні значень $\log P$ в діапазоні від -2 до 4, але вони трудомісткі і зазвичай вимагають відносно великих кількостей чистих сполук.

Метод рідинної хроматографії зазвичай використовується для кількісного визначення речовини в кожній фазі двофазної системи. Метод рідинної хроматографії забезпечує ширший діапазон застосування, ніж газова хроматографія, і забезпечує нижчу межу виявлення, ніж ультрафіолетова та видима спектроскопія. Межа виявлення особливо актуальна для високоліпофільних сполук, оскільки сполуки зі значеннями $\log P$ більше 4 часто обмежені мінімальною межею виявлення аналіту у водній фазі [8].

1) Метод струшування колби

Метод струшування колби використовує *n*-октанол (гідрофобний) і воду (гідрофільну) як двофазну рідинну систему. Ця класична процедура складається з розчинення зразка, струшування до досягнення рівноваги та вимірювання концентрації сполуки в кожній фазі двофазної системи [7].

Метод струшування колби є прямим вимірюванням коефіцієнта розподілу, і це його основна перевага. Однак ця процедура має кілька недоліків [9–11]:

- метод вимагає контролю над великою кількістю експериментальних параметрів (різні умови експерименту зазвичай дають різні значення $\log P$ для того самого аналіту);
- метод дуже трудомісткий, оскільки розподіл повинен досягти рівноваги;
- час, необхідний для досягнення рівноважних концентрацій, дуже різниться (від 1 до 24 годин);
- швидкість розподілу залежить від ліпофільності розчиненої речовини та ступеня струшування;
- метод споживає велику кількість розчинника;
- метод не підходить для сполук, що розкладаються, для поверхнево-активних матеріалів та для погано розчинних сполук.

2) Метод повільного перемішування [12]

У методі повільного перемішування *n*-октанол і водну фазу змішують при повільному перемішуванні (замість струшування). Діючи таким чином, можна запобігти утворенню емульсії. Отримані значення за допомогою методу струшування колби та методу повільного перемішування дуже схожі для сполук, що демонструють $\log P < 4,5$, але вони відрізняються для сполук із вищим $\log P$, що можна пояснити утворенням емульсії октанолу під час процедури струшування колби. Як недолік, цей метод вимагає тривалого перемішування, до 2-3 днів, для досягнення рівноваги.

3) Метод аспірації/ін'єкції водяної пробки [13]

Метод аспірації/ін'єкції за допомогою водяної пробки є ще однією модифікацією методу струшування колби, яка особливо підходить для високоліпофільних сполук. У процедурі струшування колби важко вручну відокремити водну фазу від фази н-октанолу без забруднення водної фази через високу в'язкість н-октанолу. Щоб уникнути цієї проблеми, перед збором водної фази в голку аспірують невеликий об'єм води, щоб зменшити потенційне забруднення, оскільки вода в голці має відштовхувати забруднюючу фазу н-октанолу.

4) Метод на основі потоку [14, 15]

Інжекційний аналіз потоку можна використати для встановлення варіанту методу струшування колби, який дозволяє стандартизувати вимірювання. У цьому методі досліджувана речовина розчиняється або в н-октанолі, або у водній фазі. Потім речовина впорскується в потік капілярної трубки з відповідною фазою, і дві незмішувані фази безперервно перекачуються до точки змішування в системі. Речовина розподіляється між двома фазами, і після встановлення рівноваги частина водного потоку відокремлюється, і визначається концентрація аналіту. Велика площа поверхні та коротка відстань призводять до швидкого транспортування аналіту з однієї фази в іншу, що дозволяє йому швидко досягти рівноваги.

5) Метод вортекс рідинно-рідинної мікроекстракції [10]

Метод вортекс рідинно-рідинної мікроекстракції – це метод мікроекстракції, який у поєднанні з рідинною хроматографією спрямований на підвищення ефективності вимірювань $\log P$. У методі вортекс рідинно-рідинної мікроекстракції мікрооб'єми н-октанолу диспергують у водній фазі за допомогою вихрового перемішування. Механічна напруга, викликана вихровим перемішуванням, розбиває фазу н-октанолу на кілька менших крапель, утворюючи емульсію. Дрібні мікрокраплі утворюються для збільшення площі міжфазного контакту та зменшення товщини застійної водної плівки, яка зазвичай присутня на межі розділу н-октанол/вода. Отже, час, необхідний для досягнення рівноваги, різко скорочується (умови

рівноваги досягаються за 2 хвилини вихрового перемішування).

6) Метод на основі наноабсорбенту [16, 17]

Наночастинки з абсорбуючими властивостями можна використовувати для визначення коефіцієнта розподілу. Один із цих методів заснований на використанні інкапсульованих у пористому кремнеземі магнітних наночастинок, які попередньо завантажені відомою кількістю н-октанолу. Потім ці наночастинки диспергують у водну фазу, що містить досліджувану сполуку. Невеликий розмір крапель н-октанолу та ефективне змішування, яке забезпечують наночастинки, збільшують площу міжфазного контакту між двома фазами, що значно скорочує час, необхідний для досягнення рівноваги. Магнітні властивості наночастинок дозволяють легко розділяти фази. Коефіцієнт розподілу визначається шляхом вимірювання концентрації цільового аналіту у водній фазі до та після розподілу.

7) Метод на основі діалізу [18]

Для визначення ізотропної ліпофільності можна використовувати діалізні мішки. У цьому методі діалізний мішок заповнюють розчином досліджуваної сполуки, розчиненої в н-октанолі, насиченому водою або буфером. Потім діалізний мішок занурюють у водну фазу, насичену н-октанолом, і застосовують ультразвукову обробку, щоб скоротити час урівноваження. Після обробки ультразвуком фази легко розділяються та аналізуються за допомогою рідинної хроматографії.

Висновки до розділу 1

1. Параметр ліпофільності надзвичайно важливий для розуміння та прогнозування фармакокінетичних властивостей сполук-кандидатів у ефективні лікарські засоби.

2. Експериментальні методи визначення ліпофільності, зазвичай, довготривалі, кожен зі своїми недоліками та перевагами.

2. ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА

2.1. Масиви даних для дослідження

У даній роботі для порівняння методів прогнозування ліпофільності молекул лікарських засобів використали нижченаведені масиви даних [19]:

1) 34 лікарські сполуки, що характеризуються такими фізичними та фізико-хімічними параметрами (таблиця 2.1):

- енергія нульової точки (Гартрі);
- теплова енергія (Гартрі);
- ентальпія (Гартрі);
- вільна енергія (Гартрі);
- енергія активації (ккал/моль);
- теплоємність (кал/(моль-К));

2) ті самі 34 лікарські сполуки, що характеризуються такими фізико-хімічними та електронними параметрами (таблиця 2.2):

- ентропія (кал/(моль-К));
- енергія Гартрі-Фока (Гартрі);
- стеричний фактор;
- найвища зайнята молекулярна орбітальна енергія (а.о.);
- найнижча незайнята молекулярна орбітальна енергія (а.о.);
- молекулярна рефракція;
- коефіцієнт розподілу.

Таблиці 2.1 та 2.2 також містять експериментальні значення ліпофільності ($\log P$) наведених лікарських сполук.

Таблиця 2.1. Лікарські сполуки та значення їх деяких дескрипторів та ліпофільності [19]

№	Лікарська сполука	Log P	Енергія нульової точки (Гартрі)	Теплова енергія (Гартрі)	Ентальпія (Гартрі)	Вільна енергія (Гартрі)	Енергія активації (ккал/моль)	Теплоємність (кал/(моль·К))
1	Абакавір	0,72	0,32838	0,34716	0,34811	0,27737	217,848	70,565
2	Ацикловір	-1,76	0,23925	0,25410	0,25504	0,19589	159,448	54,410
3	Аденозин	-1,46	0,25538	0,27237	0,27332	0,20836	170,915	62,614
4	Альбуцид	0,02	0,33358	0,35226	0,35320	0,28447	221,046	68,569
5	Атропін	1,53	0,38692	0,40565	0,40660	0,33719	254,550	72,015
6	Азациитидин	-1,99	0,23007	0,24585	0,24679	0,18601	154,270	57,956
7	Карбідоба	-0,19	0,24834	0,26441	0,26536	0,20412	165,920	59,183
8	Цитарабін	-2,3	0,24291	0,25859	0,25954	0,19869	162,269	58,527
9	Децитабін	-1,93	0,22512	0,23991	0,24086	0,18207	150,547	54,131
10	Десвенлафаксин	2,26	0,39517	0,41415	0,41510	0,34679	259,884	72,647
11	Добутамін	2,49	0,38899	0,41093	0,41187	0,33115	257,860	81,035
12	Дифілін	-1,12	0,25841	0,27588	0,27683	0,21060	173,119	62,450
13	Флоксуридин	-1,2	0,21981	0,23464	0,23558	0,17618	147,238	55,063
14	Ганцикловір	-2,07	0,24891	0,26584	0,26679	0,20226	166,818	61,578
15	Гоматропін	1,57	0,35820	0,37548	0,37642	0,31073	235,617	67,254
16	Гідроксихлорохін	3,54	0,42694	0,45075	0,45169	0,36738	282,849	86,496
17	Ізоетарин	1,13	0,33339	0,35212	0,35306	0,28513	220,957	68,544
18	Ізопротеренол	0,25	0,27601	0,29223	0,29318	0,23010	183,379	58,751
19	Ізоксупрін	2,58	0,38944	0,41082	0,41177	0,33470	257,794	80,003
20	Ламівудин	-1,02	0,20450	0,21852	0,21946	0,16227	137,123	51,660
21	Левобунолол	2,86	0,40632	0,42739	0,42833	0,35275	268,191	78,991
22	Метипранолол	2,67	0,42742	0,45283	0,45377	0,36751	284,155	89,840

23	Мідодрин	-0,32	0,30600	0,32443	0,32538	0,25729	203,585	67,361
24	Піридоксин	-1,9	0,18992	0,20219	0,20313	0,15134	126,875	43,813
25	Ризедронова кислота	-2,94	0,20154	0,21959	0,22053	0,15486	137,794	65,868
26	Ритодрін	1,61	0,36055	0,37988	0,38083	0,30928	238,379	73,810
27	Ставудин	-0,91	0,22577	0,24064	0,24159	0,17852	151,006	53,326
28	Тербугалін	0,48	0,30482	0,32204	0,32299	0,25716	202,085	63,950
29	Тригексифенідил	5,06	0,48781	0,50820	0,50914	0,43689	318,898	80,508
30	Тропікамід	1,16	0,34962	0,36833	0,36928	0,29971	231,132	70,344
31	Відарабін	-1,46	0,25544	0,27234	0,27328	0,20893	170,895	62,586
32	Воринонат	0,86	0,33606	0,35572	0,35666	0,27938	223,216	69,985
33	Залцитабін	-1,51	0,23336	0,24698	0,24793	0,19143	154,985	50,659
34	Золедронова кислота	-2,28	0,18294	0,20023	0,20117	0,13721	125,644	62,757

Таблиця 2.2. Лікарські сполуки та значення їх деяких дескрипторів та ліпофільності [19]

№	Лікарська сполука	Log P	Ентропія (кал/(моль·К))	Енергія Гартрі-Фока (Гартрі)	Стеричний фактор	Найвища зайнята молекулярна орбітальна енергія (а.о.)	Найнижча незайнята молекулярна орбітальна енергія (а.о.)	Молекулярна рефракція	Коефіцієнт розподілу
1	Абакавір	0,72	148,877	0,15097	40,67	-0,30382	0,00319	7,8973	0,8057
2	Ацикловір	-1,76	124,491	-0,08041	15,51	-0,31690	0,00994	5,4966	-2,1354
3	Аденозин	-1,46	136,720	-0,14181	38,25	-0,32661	-0,00929	6,2955	-2,1577
4	Альбуцид	0,02	144,658	-0,22051	-0,12	-0,33796	0,00395	6,7632	0,0614
5	Атропін	1,53	146,082	-0,17317	32,90	-0,34748	0,00358	8,1462	1,2992
6	Азацитидин	-1,99	127,931	-0,25863	12,92	-0,36931	-0,01385	5,3911	-2,1981
7	Карбідоба	-0,19	128,890	-0,22215	-25,99	-0,32092	0,01247	5,7762	-0,4448
8	Цитарабін	-2,3	128,064	-0,2743	15,34	-0,33873	-0,00018	5,6022	-2,1951
9	Децитабін	-1,93	123,727	-0,18278	20,16	-0,36249	-0,00838	5,2380	-1,9012
10	Десвенлафаксин	2,26	143,772	-0,14428	17,43	-0,32439	0,01927	7,8241	2,6830
11	Добутамін	2,49	169,890	-0,1674	1,32	-0,32084	0,00837	8,8106	2,4330
12	Дифілін	-1,12	139,395	-0,15661	26,39	-0,33688	-0,02165	6,2242	-1,2861
13	Флоксуридин	-1,2	125,014	-0,36648	22,05	-0,36340	-0,02766	5,3090	-1,4048
14	Ганцикловір	-2,07	135,806	-0,14534	22,87	-0,31957	-0,01469	6,0691	-2,5448
15	Гоматропін	1,57	138,272	-0,16281	27,30	-0,34750	0,00392	7,6824	1,4274
16	Гідроксихлорохін	3,54	177,453	-0,02964	24,86	-0,31718	-0,01819	9,7216	4,1159
17	Ізоетарин	1,13	142,972	-0,20758	3,23	-0,32652	0,00245	6,7632	0,9914
18	Ізопротеренол	0,25	132,761	-0,20465	-7,38	-0,32934	0,00299	5,8356	0,1534
19	Ізоксупрін	2,58	162,193	-0,14145	8,84	-0,32683	0,00385	8,8106	2,6150
20	Ламівудин	-1,02	120,367	-0,09953	9,64	-0,32222	-0,01616	5,6385	-1,4624
21	Левобунолол	2,86	159,088	-0,19797	17,93	-0,33775	-0,01415	8,3236	2,2623

22	Метипранолол	2,67	181,563	-0,26987	19,15	-0,33847	-0,00504	8,6541	2,5454
23	Мідодрин	-0,32	143,306	-0,2255	-5,84	-0,32555	-0,00065	6,7038	-0,4248
24	Піридоксин	-1,9	109,014	-0,19064	4,10	-0,34360	-0,00701	4,3282	-0,3450
25	Ризедронова кислота	-2,94	138,228	-0,66112	28,74	-0,37478	-0,01503	5,7520	-2,6224
26	Ритодрін	1,61	150,580	-0,15598	-1,74	-0,32844	0,01275	8,3468	1,6514
27	Ставудин	-0,91	132,740	-0,1791	7,40	-0,35086	-0,00871	5,5788	-0,4875
28	Тербуталін	0,48	138,542	-0,20335	0,30	-0,33354	0,01039	6,2994	0,4824
29	Тригексифенідил	5,06	152,058	-0,1029	22,65	-0,33388	0,01885	9,3488	5,1510
30	Тропікамід	1,16	146,405	-0,05002	4,58	-0,35844	0,00185	8,3290	1,1806
31	Відарабін	-1,46	135,449	-0,14293	41,99	-0,31980	-0,00317	6,2955	-2,1577
32	Воринонат	0,86	162,648	-0,14613	-8,32	-0,32383	0,01037	7,3609	0,9890
33	Залцитабін	-1,51	118,919	-0,14032	15,58	-0,34090	-0,00424	5,2960	-1,2469
34	Золедронова кислота	-2,28	134,613	-0,61319	41,98	-0,34492	0,00261	5,1815	-3,0656

2.2. Опис штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі

Штучні нейронні мережі прямого поширення сигналу, як і будь-яка інша штучна нейронна мережа, складається з шарів нейронів. Сигнали (тобто вхідні вектори чи вхідна інформація) надходять на вхідний рівень нейронів. Потім вони надходять та обробляються нейронами прихованого шару. Виходи (кінцевий результат) надають нейрони вихідного шару. Таким чином, сигнали рухаються тільки в прямому напрямку. Кожен наступний шар має з'єднання з попереднім шаром [20–22].

Каскадна нейрона мережа є різновидом мережі прямого поширення сигналу, але з однією відмінністю. Каскадна мережа має додаткові підключення від входу до кожного рівня та від кожного рівня до всіх наступних рівнів [20–22].

Для обробки будь-якої інформації і для вирішення різноманітних завдань можна використовувати мережу прямого поширення сигналу та каскадну мережу з одним прихованим шаром і достатньою кількістю нейронів у прихованому шарі [23, 24].

Мережа прямого поширення сигналу та каскадна мережа, як правило, мають один прихований шар нейронів із сигмоїдальною функцією активації, за яким слідує вихідний рівень нейронів із лінійною функцією передачі:

$$f_{\text{linear}} = n, \quad (2.1)$$

$$f_{\text{tangent sigmoid}} = \frac{e^n - e^{-n}}{e^n + e^{-n}}, \quad (2.2)$$

$$n = wp + b, \quad (2.3)$$

де p – вхідний вектор, w – ваговий вектор нейрона, b — зміщення нейрона.

Застосування нелінійних функцій активації в архітектурі штучної

нейронної мережі дозволяє вивчити нелінійні та лінійні зв'язки введення-виведення.

Для правильного застосування штучної нейронної мережі необхідно визначити такі параметри її архітектури: функції активації на кожному рівні, метод навчання та кількість нейронів у кожному шарі. Вхідні сигнали, помножені на вагові параметри, підсумовуються та пропускаються через передатну функцію для створення виходу для нейронів.

Початкові ваги були призначені за методом Нгуена-Відроу. Цей метод швидкий і в поєднанні з методом навчання Левенберга-Марквардта забезпечує швидку збіжність і точність алгоритмів нейронних мереж в цілому. Крім того, алгоритм Нгуена-Відроу найкраще підходить для використання з сигмоїдальною та лінійною функціями активації [25, 26]. У [27] було показано, що для прямих і каскадних нейронних мереж найкращі результати були досягнуті за допомогою методу навчання Левенберга-Марквардта та сигмоїдальної / лінійної функції активації, відповідно, для прихованого і вихідного шарів. Для підвищення ефективності навчання нейронних мереж, зазвичай, як додатковий метод застосовують алгоритм градієнтного спуску.

Середня квадратична похибка була використана як функція похибки, яку потрібно мінімізувати під час навчання штучних нейронних мереж [26]:

$$MSE = \frac{1}{M} \sum_{i=1}^M (y_i - t_i)^2, \quad (2.4)$$

де M – кількість сполук у навчальній вибірці, y_i та t_i – прогнозоване значення досліджуваного параметру і його реальне (експериментальне) значення.

Таблиця 2.3 містить інформацію щодо архітектури реалізованих у даній роботі штучних нейронних мереж. Реалізацію алгоритмів нейронної мережі прямого поширення сигналу та каскадної мережі виконано із застосуванням програмного комплексу Matlab R2023b (trial individual license 11937601) [28].

Таблиця 2.3. Параметри навчання прямої і каскадної нейронних мереж

<i>Тип параметру</i>	<i>Опис</i>
Тип навчання	Навчання «із вчителем»
Метод навчання	Левенберга-Марквардта, алгоритм градієнтної оптимізації
Мінімізована функція похибки	Середня квадратична похибка
Кількість прихованих нейронів	$n-m$
Функції активації	Сигмоїдальна / лінійна
Метод ініціалізації	Алгоритм Нгуєна-Відроу

Висновки до розділу 2

1. Якщо досліджувані масиви є великими, то застосування штучних нейронних мереж до їх обробки є необхідним та прийнятним.

2. Для прогнозування ліпофільності лікарських сполук обрано дві штучні нейронні мережі, що одночасно є однаковими з точки зору поширення сигналу з одного шару на інший та різні з точки зору зв'язків між шарами нейронів.

3. РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

3.1. Прогностичні моделі з літературних джерел, що використовувалися для порівняння

Результати прогнозування ліпофільності з використанням штучних нейронних мереж порівнювалися із результатами, що наведені в літературі [19] з метою оцінки їх якості та ефективності:

$$\log P = 0,939 + 8,343 \cdot \text{вільна енергія} - 0,009 \cdot \text{теплоємність} - 0,006 \cdot \text{ентропія} - 1,082 \cdot \text{енергія Гартрі-Фока} - 0,002 \cdot \text{стеричний фактор} + 9,957 \cdot \text{найвища зайнята молекулярна орбітальна енергія} - 5,811 \cdot \text{найнижча незайнята молекулярна орбітальна енергія} + 0,241 \cdot \text{молекулярна рефракція} + 5,597 \cdot \text{коефіцієнт розподілу} \quad (3.1)$$

$$\log P = -2,695 + 0,615 \cdot \text{коефіцієнт розподілу} + 8,687 \cdot \text{теплова енергія} \quad (3.2)$$

$$\log P = -2,365 + 4,621 \cdot \text{вільна енергія} + 0,028 \cdot \text{теплоємність} - 0,015 \cdot \text{ентропія} - 52,41 \cdot \text{енергія Гартрі-Фока} - 0,001 \cdot \text{стеричний фактор} + 0,859 \cdot \text{найвища зайнята молекулярна орбітальна енергія} + 1,647 \cdot \text{найнижча незайнята молекулярна орбітальна енергія} + 0,178 \cdot \text{молекулярна рефракція} + 0,512 \cdot \text{коефіцієнт розподілу} \quad (3.3)$$

$$\log P = -2,328 + 0,605 \cdot \text{коефіцієнт розподілу} + 8,042 \cdot \text{вільна енергія} \quad (3.4)$$

Якість отриманих результатів в роботі [19] оцінювали за допомогою коефіцієнта кореляції між експериментальним значенням ліпофільності та прогнозованим значенням ліпофільності: для рівняння (3.1) $R = 0,990$; для рівняння (3.2) $R = 0,988$; для рівняння (3.3) $R = 0,989$; для рівняння (3.4) $R = 0,988$.

3.2. Результати прогнозування ліпофільності із застосуванням мережі прямого поширення сигналу та каскадної мережі

Для кожного нижче описаного набору дескрипторів знаходилося оптимальне число прихованих нейронів та оцінювалася прогностична сила нейронних мереж. Оптимальне число прихованих нейронів оцінювалося за допомогою навчання нейронних мереж на усіх 34 лікарських сполуках та оцінкою середньоквадратичної похибки. Тобто як якісно нейронні мережі «вивчили» подану інформацію та відтворюють її. Для оцінки прогностичної сили 34 досліджувані лікарські сполуки випадковим чином були поділені на навчальну вибірку (75%) та тестову вибірку (25%).

Склад тестової вибірки: Ацикловір, Азацитидин, Десвенлафаксин, Ганцикловір, Ізопротеренол, Метипранолол, Ставудин, Відарабін.

3.2.1. Прогнозування ліпофільності із застосуванням 6 дескрипторів (енергія нульової точки; теплова енергія; ентальпія; вільна енергія; енергія активації; теплоємність)

Встановлено, що оптимальна кількість нейронів прихованого шару для нейронної мережі прямого поширення сигналу складає 9, а для каскадної нейронної мережі – 7. Відповідні залежності наведено на рисунках 3.1 та 3.2.



Рисунок 3.1. Залежність середньоквадратичної похибки від кількості

прихованих нейронів для нейронної мережі прямого поширення сигналу



Рисунок 3.2. Залежність середньоквадратичної похибки від кількості прихованих нейронів для каскадної нейронної мережі

Рисунки 3.3 та 3.4 представляють кореляцію між прогнозованим значенням ліпофільності та експериментальним значенням ліпофільності у випадку навчання нейронних мереж із встановленим оптимальним значенням прихованих нейронів: для нейронної мережі прямого поширення сигналу $R = 0,965$ та для каскадної мережі $R = 0,962$.

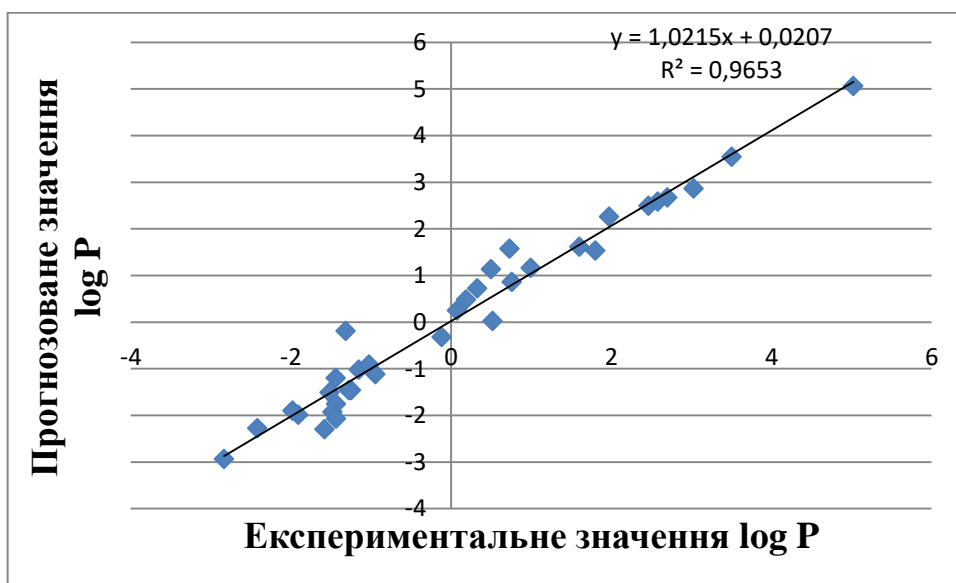


Рисунок 3.3. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу

для усіх 34 лікарських сполук

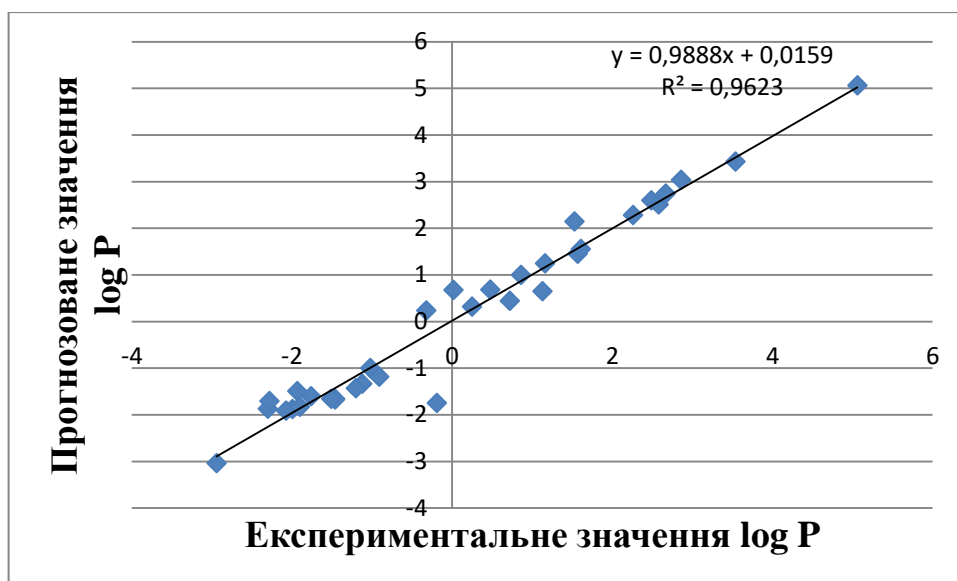


Рисунок 3.4. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для усіх 34 лікарських сполук

Нагадаємо, що у результаті застосування математичного рівняння QSAR (3.1), що застосовує той самий набір дескрипторів, $R = 0,990$ [19].

При оцінці прогностичної сили запропонованих моделей отримані такі результати: для нейронної мережі прямого поширення сигналу $MSE = 0,165$ та $R = 0,958$ (рисунок 3.5); для каскадної нейронної мережі $MSE = 0,478$ та $R = 0,874$ (рисунок 3.6). Отже, можна зробити висновок, що нейронна мережа прямого поширення сигналу має більшу прогностичну силу, ніж каскадна нейронна мережа.

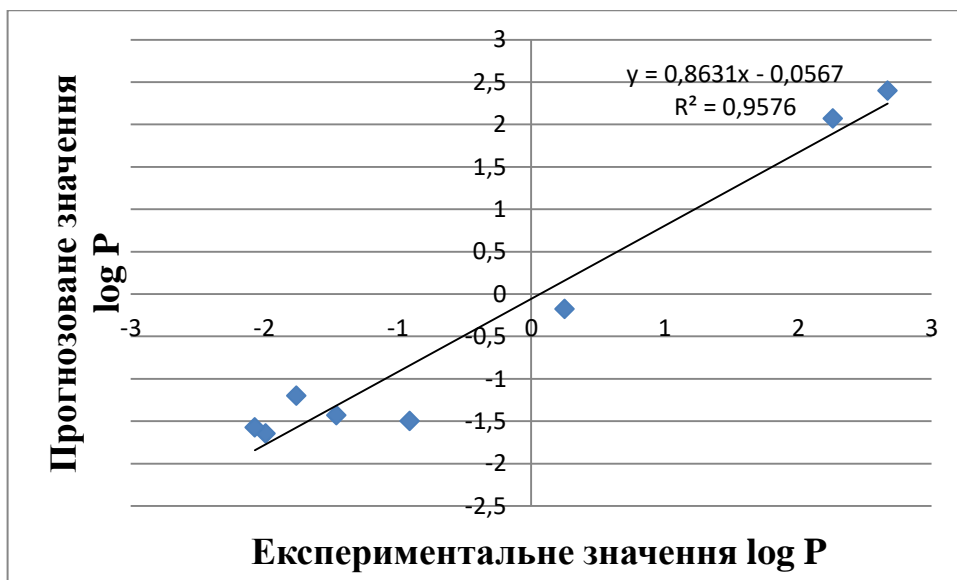


Рисунок 3.5. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу для сполук тестової вибірки

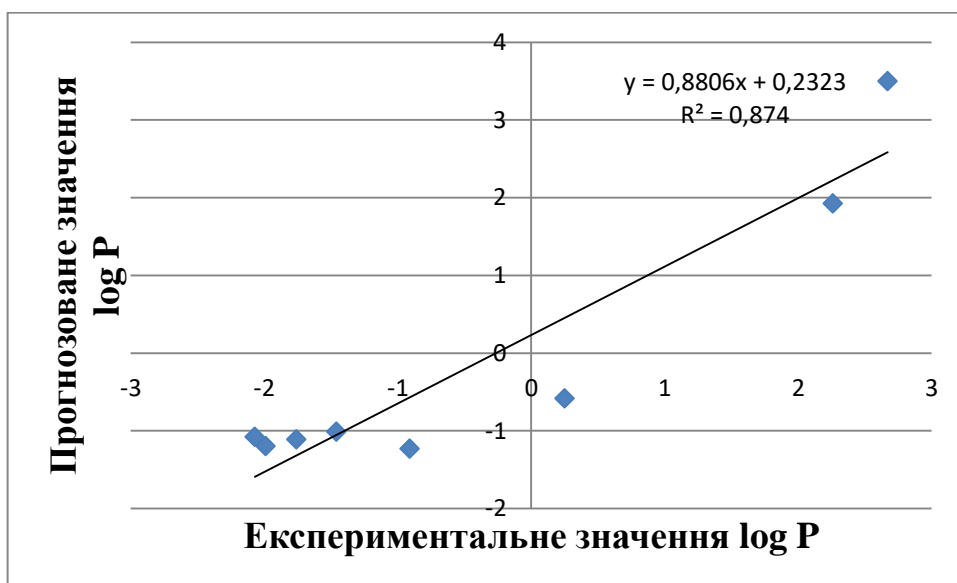


Рисунок 3.6. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для сполук тестової вибірки

3.2.2. Прогнозування ліпофільності із застосуванням 7 дескрипторів (ентропія; енергія Гартрі-Фока; стеричний фактор; найвища зайнята молекулярна орбітальна енергія; найнижча незайнята молекулярна орбітальна енергія; молекулярна рефракція; коефіцієнт розподілу)

Встановлено, що оптимальна кількість нейронів прихованого шару для нейронної мережі прямого поширення сигналу складає 4, а для каскадної нейронної мережі – 7. Відповідні залежності наведено на рисунках 3.7 та 3.8.

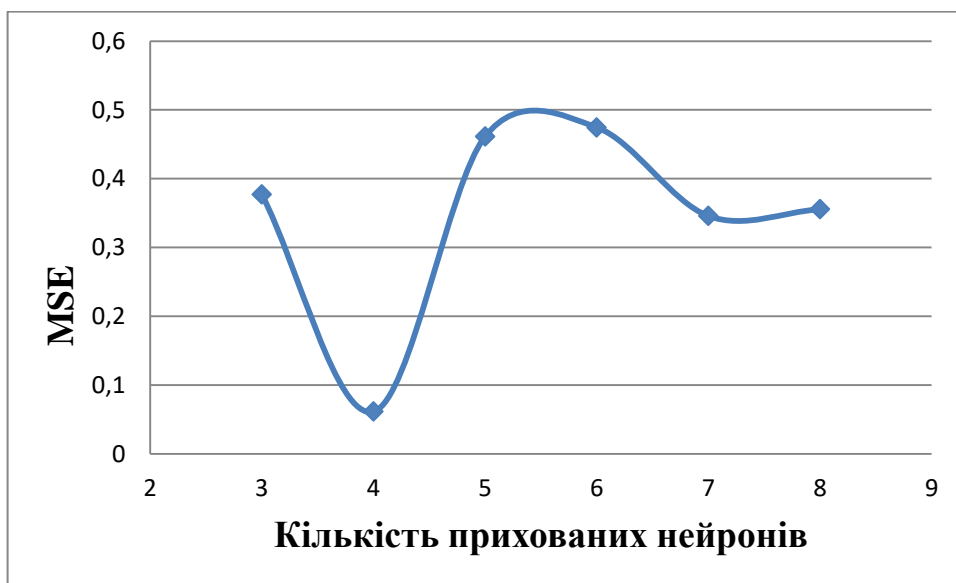


Рисунок 3.7. Залежність середньоквадратичної похибки від кількості прихованих нейронів для нейронної мережі прямого поширення сигналу



Рисунок 3.8. Залежність середньоквадратичної похибки від кількості прихованих нейронів для каскадної нейронної мережі

Рисунки 3.9 та 3.10 представляють кореляцію між прогнозованим значенням ліпофільності та експериментальним значенням ліпофільності у випадку навчання нейронних мереж із встановленим оптимальним значенням прихованих нейронів: для нейронної мережі прямого поширення сигналу $R = 0,985$ та для каскадної мережі $R = 0,980$.

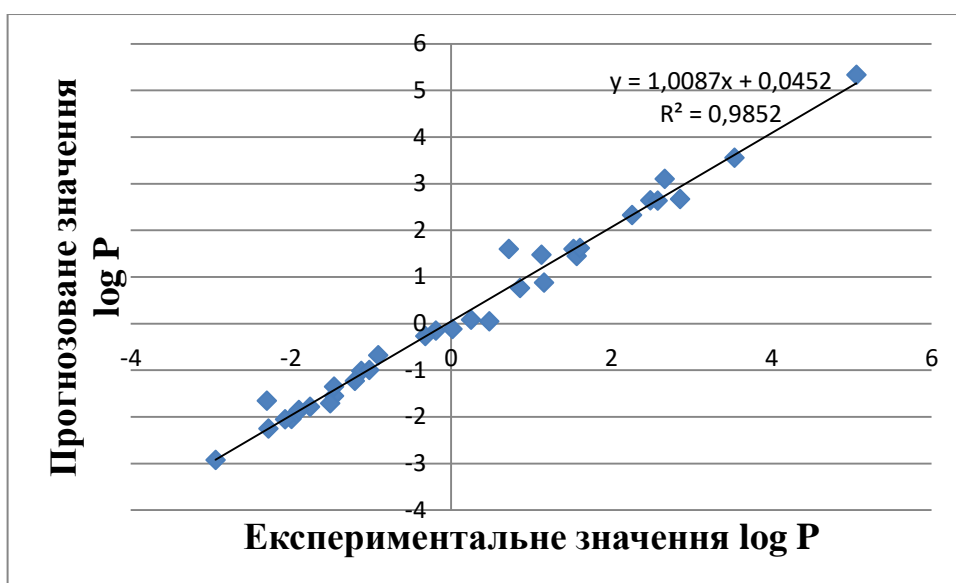


Рисунок 3.9. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу для усіх 34 лікарських сполук

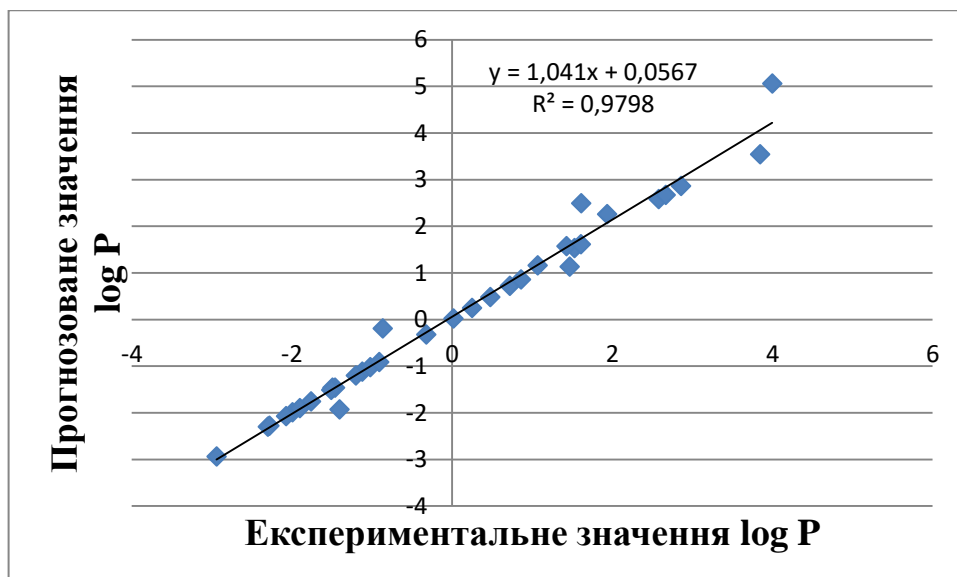


Рисунок 3.10. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для усіх 34 лікарських сполук

Нагадаємо, що у результаті застосування математичного рівняння QSAR (3.3), що застосовує той самий набір дескрипторів, $R = 0,989$ [19].

При оцінці прогностичної сили запропонованих моделей отримані такі результати: для нейронної мережі прямого поширення сигналу $MSE = 0,108$ та $R = 0,983$ (рисунок 3.11); для каскадної нейронної мережі $MSE = 0,175$ та $R = 0,982$ (рисунок 3.12).

Необхідно відмітити, що при прогнозуванні ліпофільності із застосуванням 7 дескрипторів у порівнянні із результатами прогнозування ліпофільності із застосуванням 6 дескрипторів, середньоквадратична похибка зменшується, а коефіцієнт кореляції збільшується. Отже, набір із 7 дескрипторів є більш інформаційним та репрезентативним для прогнозування ліпофільності.

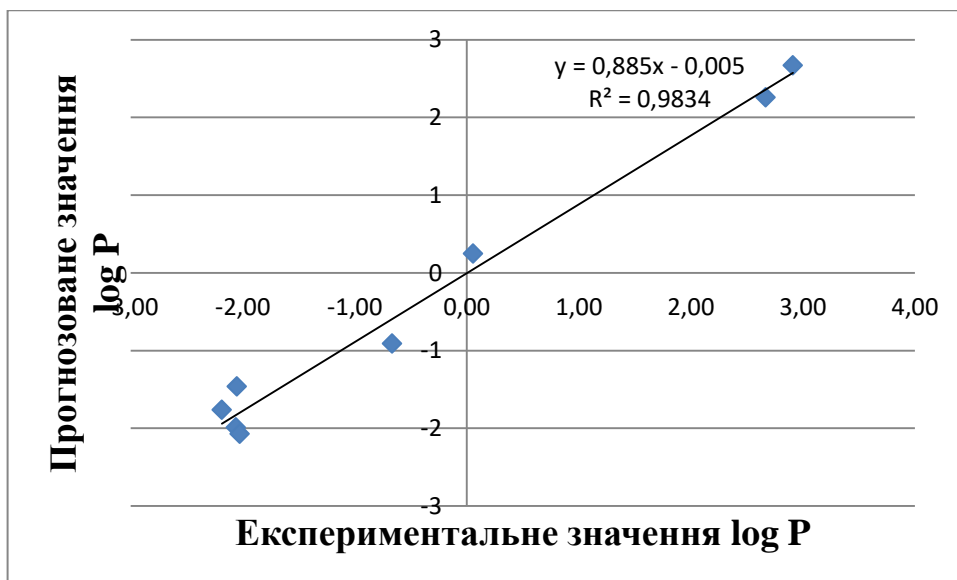


Рисунок 3.11. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу для сполук тестової вибірки

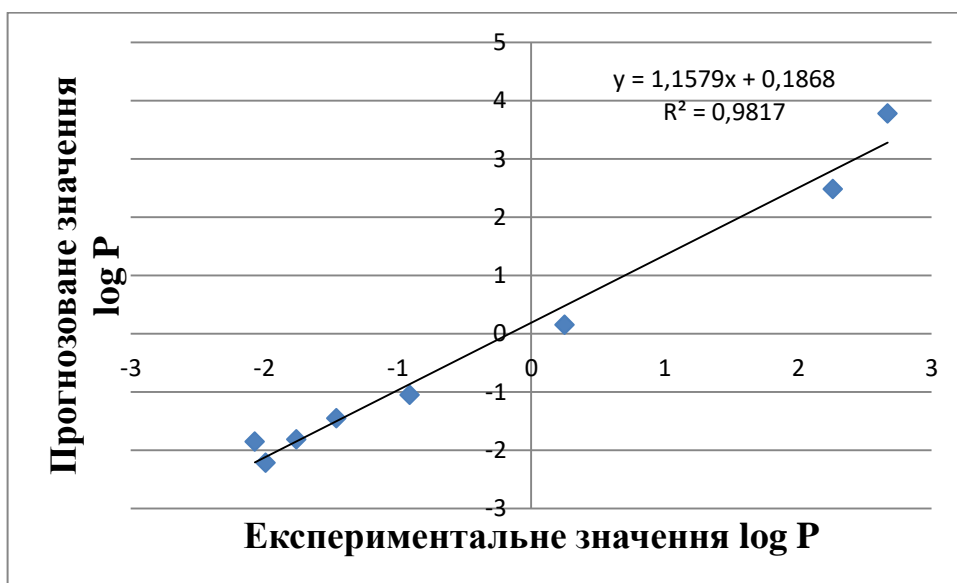


Рисунок 3.12. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для сполук тестової вибірки

3.2.3. Прогнозування ліпофільності із застосуванням 2 дескрипторів (коефіцієнту розподілу та енергії нульової точки)

Наступним кроком нашої роботи стало дослідження скорочення кількості дескрипторів для прогнозування ліпофільності лікарських сполук та визначення найбільш інформативних з них.

Таблиця 3.1 містить результати оцінки кореляції між експериментальним значенням ліпофільності та кожним дескриптором, що представлений у таблицях 2.1 та 2.2.

Таблиця 3.1. Значення коефіцієнта кореляції та рівняння функції лінійної регресії між експериментальним значенням ліпофільності та фізико-хімічними параметрами

Фізико-хімічний параметр	Рівняння лінійної регресії	Коефіцієнт кореляції
Енергія нульової точки	$y = 0,0389x + 0,2966$	0,9223
Теплова енергія	$y = 0,04x + 0,3143$	0,9198
Ентальпія	$y = 0,04x + 0,3152$	0,9197
Вільна енергія	$y = 0,0369x + 0,2485$	0,9228
Енергія активації	$y = 25,081x + 197,21$	0,9198
Теплоємність	$y = 4,4477x + 65,222$	0,6879
Ентропія	$y = 6,3556x + 140,36$	0,5959
Енергія Гартрі-Фока	$y = 0,0294x - 0,1952$	0,1709
Стеричний фактор	$y = -1,089x + 14,692$	0,0187
Найвища зайнята молекулярна орбітальна енергія	$y = 0,0027x - 0,336$	0,1053
Найнижча незайнята молекулярна орбітальна енергія	$y = 0,0025x - 0,002$	0,1827
Молекулярна рефракція	$y = 0,6548x + 6,7116$	0,8400
Коефіцієнт розподілу	$y = 1,0226x - 0,0518$	0,9582

Відповідно до наведених у таблиці 3.1 даних можна зробити наступні висновки:

1) такі фізико-хімічні параметри як енергія Гартрі-Фока, стеричний фактор, найвища зайнята молекулярна орбітальна енергія, найнижча

незайнята молекулярна орбітальна енергія, ентропія та теплоємність є малоінформативними та не репрезентативними для ефективного прогнозування ліпофільності лікарських сполук;

2) такі фізико-хімічні параметри як енергія нульової точки, теплова енергія, ентальпія, вільна енергія, енергія активації та коефіцієнт розподілу є найбільш інформативними та репрезентативними для ефективного прогнозування ліпофільності лікарських сполук.

Прогнозування ліпофільності лікарських сполук із застосуванням пари коефіцієнту розподілу та теплової енергії досліджено у роботі [19], коефіцієнт кореляції між прогнозованими та експериментальними значеннями ліпофільності становить 0,988. Також у вказаній роботі досліджено прогнозування ліпофільності лікарських сполук із застосуванням пари коефіцієнту розподілу та вільної енергії, коефіцієнт кореляції між прогнозованими та експериментальними значеннями ліпофільності теж становить 0,988.

Ми вирішили дослідити ефективність прогнозування ліпофільності лікарських сполук із застосуванням пари коефіцієнту розподілу та енергії нульової точки.

Встановлено, що оптимальна кількість нейронів прихованого шару для нейронної мережі прямого поширення сигналу складає 5, а для каскадної нейронної мережі – 6. Відповідні залежності наведено на рисунках 3.13 та 3.14.



Рисунок 3.13. Залежність середньоквадратичної похибки від кількості прихованих нейронів для нейронної мережі прямого поширення сигнал

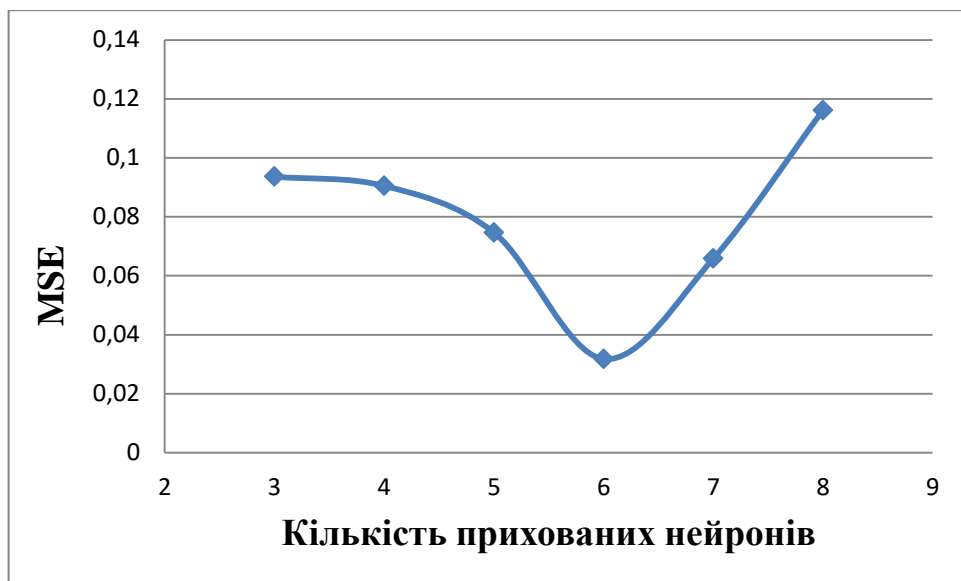


Рисунок 3.14. Залежність середньоквадратичної похибки від кількості прихованих нейронів для каскадної нейронної мережі

Рисунки 3.15 та 3.16 представляють кореляцію між прогнозованим значенням ліпофільності та експериментальним значенням ліпофільності у випадку навчання нейронних мереж із встановленим оптимальним значенням прихованих нейронів: для нейронної мережі прямого поширення сигналу $R = 0,988$ та для каскадної мережі $R = 0,992$. Отримані результати по точності не

поступаються результатам, що наведені в літературі, а результати, що отримані з використанням каскадної нейронної мережі – перевищують їх.

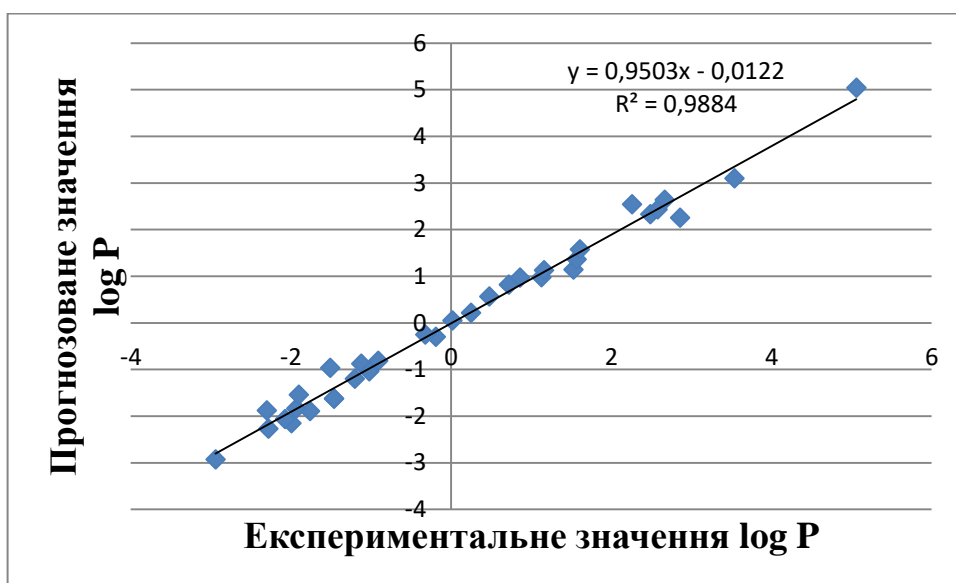


Рисунок 3.15. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу для усіх 34 лікарських сполук

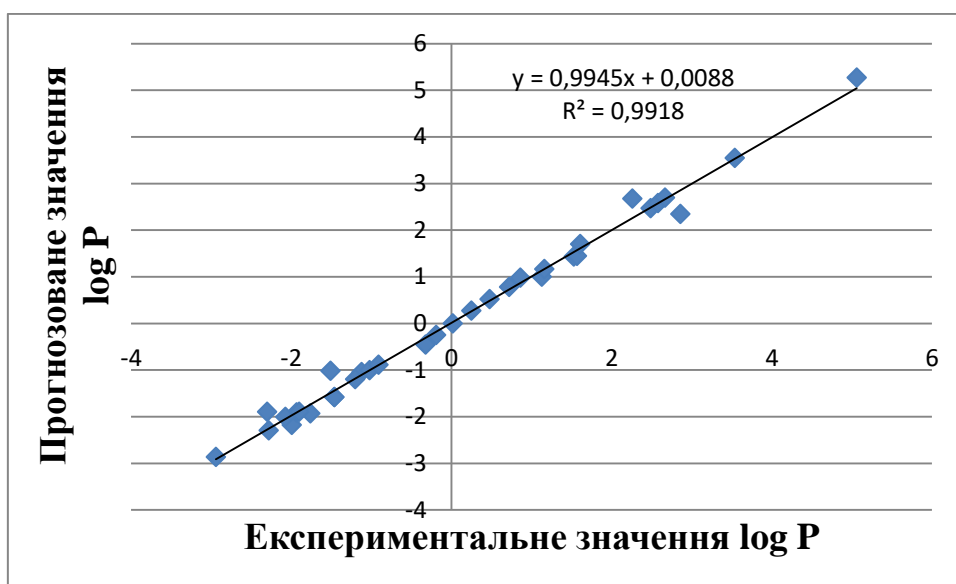


Рисунок 3.16. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для усіх 34 лікарських сполук

Нагадаємо, що у результаті застосувань математичних рівнянь QSAR (3.2) та (3.4), що теж базуються на застосуванні лише 2 дескрипторів, $R = 0,988$ та $R = 0,988$, відповідно. Отже, запропонована нами модель, характеризується більшою точністю.

При оцінці прогностичної сили запропонованих моделей отримані такі результати: для нейронної мережі прямого поширення сигналу $MSE = 0,143$ та $R = 0,988$ (рисунок 3.17); для каскадної нейронної мережі $MSE = 0,177$ та $R = 0,991$ (рисунок 3.18). Отже, можна зробити висновок, що обидві нейронні мережі мають сильну прогностичну силу.

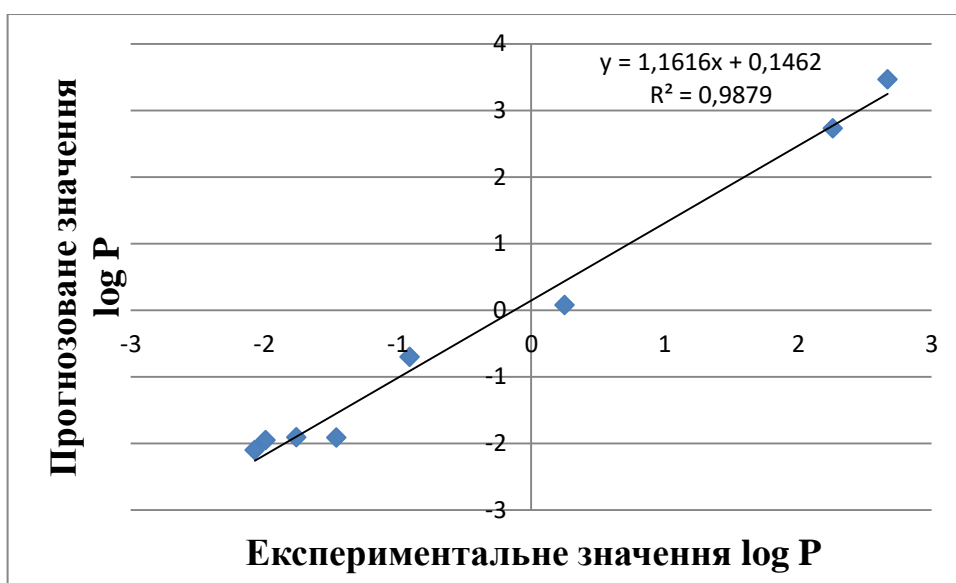


Рисунок 3.17. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для нейронної мережі прямого поширення сигналу для сполук тестової вибірки

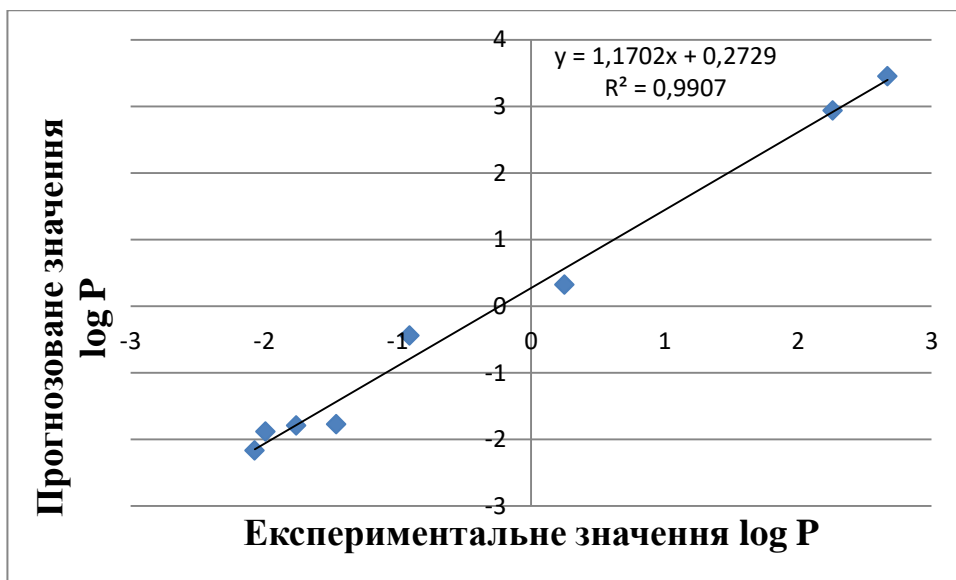


Рисунок 3.17. Кореляція між прогнозованими та експериментальними значеннями ліпофільності для каскадної нейронної мережі для сполук тестової вибірки

Висновки до розділу 3

1. Оптимальна кількість нейронів прихованого шару штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі для прогнозування ліпофільності молекул лікарських засобів на основі різних наборів фізико-хімічних параметрів не перевищує 9.

2. Проведене дослідження прогнозування ліпофільності із застосуванням різних наборів фізико-хімічних дескрипторів та математичних моделей дозволяють зробити висновок, що більш точні результати отримані із застосуванням каскадної нейронної мережі та лише двох дескрипторів (коефіцієнту розподілу та енергії нульової точки).

3. Алгоритми штучних нейронних мереж є конкурентоспроможними відповідно до методології QSAR.

ВИСНОВКИ

1. Поєднання параметрів коефіцієнту розподілу та енергії нульової точки є достатнім та ефективним для прогнозування ліпофільності молекул лікарських засобів.

2. Алгоритми штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі характеризуються високою точністю прогнозування ліпофільності молекул лікарських засобів із застосуванням різних наборів фізико-хімічних параметрів.

3. Алгоритми штучних нейронних мереж можуть бути використані для експлораторної оцінки параметру ліпофільності на стадії дослідження молекул-претендентів у потенційні лікарські засоби на рівні прогностичних моделей QSAR.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Datta, R., Das, D., & Das, S. (2021). Efficient lipophilicity prediction of molecules employing deep-learning models. *Chemometrics and Intelligent Laboratory Systems*, 213, 104309.
2. Lionta, E., Spyrou, G., K Vassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry*, 14(16), 1923-1938.
3. Ghosh, S., Nie, A., An, J., & Huang, Z. (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Current opinion in chemical biology*, 10(3), 194-202.
4. Walters, W. P. (2018). Virtual chemical libraries: miniperspective. *Journal of medicinal chemistry*, 62(3), 1116-1124.
5. Saaidpour, S. (2014). Prediction of drug lipophilicity using back propagation artificial neural network modeling. *Oriental Journal of Chemistry*, 30(2), 793.
6. Testa, B., Vistoli, G., Pedretti, A., van de Waterbeemd, H., Avdeef, A., Ivanciuc, O., & Muresan, S. (2007). Molecular drug properties: measurement and prediction.
7. OECD. (1995). Partition Coefficient (n-Octanol/Water), Shake Flask Method, No. 107. *OCED*, 1(July), 1-4.
8. Liang, C., & Lian, H. Z. (2015). Recent advances in lipophilicity measurement by reversed-phase high-performance liquid chromatography. *TrAC Trends in Analytical Chemistry*, 68, 28-36.
9. Godard, T., & Grushka, E. (2011). The use of phospholipid modified column for the determination of lipophilic properties in high performance liquid chromatography. *Journal of Chromatography A*, 1218(9), 1211-1218.
10. Román, I. P., Mastromichali, A., Tyrovola, K., Canals, A., & Psillakis, E. (2014). Rapid determination of octanol–water partition coefficient using vortex-assisted liquid–liquid microextraction. *Journal of Chromatography A*, 1330, 1-5.

11. Hartmann, T., & Schmitt, J. (2004). Lipophilicity–beyond octanol/water: a short comparison of modern technologies. *Drug Discovery Today: Technologies*, 1(4), 431-439.
12. De Bruijn, J., Busser, F., Seinen, W., & Hermens, J. (1989). Determination of octanol/water partition coefficients for hydrophobic organic chemicals with the “slow-stirring” method. *Environmental Toxicology and Chemistry: An International Journal*, 8(6), 499-512.
13. Dohta, Y., Yamashita, T., Horiike, S., Nakamura, T., & Fukami, T. (2007). A System for LogD Screening of 96-Well Plates Using a Water-Plug Aspiration/Injection Method Combined with High-Performance Liquid Chromatography– Mass Spectrometry. *Analytical chemistry*, 79(21), 8312-8315.
14. Marine, N. A., Klein, S. A., & Posner, J. D. (2009). Partition coefficient measurements in picoliter drops using a segmented flow microfluidic device. *Analytical chemistry*, 81(4), 1471-1476.
15. Zhang, Y. H. (1996). Methods for determining n-octanol-water partition constants. *TrAC trends in analytical chemistry*, 15(4), 188-196.
16. Gao, X., Yu, C. H., Tam, K. Y., & Tsang, S. C. (2005). New magnetic nano-absorbent for the determination of n-octanol/water partition coefficients. *Journal of pharmaceutical and biomedical analysis*, 38(2), 197-203.
17. Yu, C. H., Tam, K., & Tsang, S. C. (2011). A new high-throughput method utilizing porous silica-based nano-composites for the determination of partition coefficients of drug candidates. *Journal of separation science*, 34(18), 2505-2512.
18. Schröder, W., & Andersson, J. T. (2001). Fast and direct method for measuring 1-octanol–water partition coefficients exemplified for six local anesthetics. *Journal of pharmaceutical sciences*, 90(12), 1948-1954.
19. Ibrahim, A. A., Abd-Alrazzak, A. Y., AbdalRazaq, E. A., Sulliman, E. A., & Shamil, T. (2020). Theoretical Prediction of Lipophilicity for Some Drugs Compounds. *Oriental J. of Chem*, 36(1), 114-119.
20. Beale, M. H., Hagan, M. T., & Demuth, H. B. (2010). Neural network

toolbox. *User's Guide, MathWorks*, 2.

21. Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, 22(5), 717-727.

22. Mutihac, L., & Mutihac, R. (2008). Mining in chemometrics. *Analytica Chimica Acta*, 612(1), 1-18.

23. Scarselli, F., & Tsoi, A. C. (1998). Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural networks*, 11(1), 15-37.

24. Zheng, P., Harrington, P. D. B., & Davis, D. M. (1996). Quantitative analysis of volatile organic compounds using ion mobility spectrometry and cascade correlation neural networks. *Chemometrics and Intelligent Laboratory Systems*, 33(2), 121-132.

25. Nguyen, D., & Widrow, B. (1990, June). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *1990 IJCNN international joint conference on neural networks* (pp. 21-26). IEEE.

26. Balabin, R. M., Safieva, R. Z., & Lomakina, E. I. (2009). Universal technique for optimization of neural network training parameters: gasoline near infrared data example. *Neural Computing and Applications*, 18, 557-565.

27. Pushkarova, Y., & Kholin, Y. (2012). The classification of solvents based on solvatochromic characteristics: the choice of optimal parameters for artificial neural networks. *Open Chemistry*, 10(4), 1318-1327.

28. Paluszek, M., & Thomas, S. (2016). *MATLAB machine learning*. Apress.

ДОДАТОК



CERTIFICATE OF PARTICIPATION

Certificate provides at least a 0.1 ECTS credits to awarded participants for being involved

IST № 23/1201-155

Anna Dorosh

participated in the I International Scientific and Theoretical Conference

Scientific review of the actual events, achievements and problems

01.12.2023 | Berlin, Federal Republic of Germany

The conference is included in the Academic Resource Index ResearchBb catalog and UAR/STET catalog (Certificate № 311 dated June 16, 2023).

President of the International Center of Scientific Research
MARIIA HOLDENBLAT




SECTION 24.

PHARMACY AND PHARMACOTHERAPY

Дорош Анна Сергіївна

здобувач вищої освіти фармацевтичного факультету
Національний медичний університет імені О.О. Богомольця, Україна

Науковий керівник: Пушкарьова Ярослава Миколаївна 

канд. хім. наук, доцент, доцент кафедри аналітичної, фізичної та колоїдної хімії
Національний медичний університет імені О.О. Богомольця, Україна

ПРОГНОЗУВАННЯ ЛІПОФІЛЬНОСТІ МОЛЕКУЛ ЛІКАРСЬКИХ ЗАСОБІВ ЗА ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Анотація. Ліпофільність обов'язково враховується та оцінюється на етапі моделювання нових лікарських засобів, оскільки є однією із складових успішного застосування досліджуваної молекули як потенційного лікарського препарату. Досліджено можливість застосування штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі для прогнозування ліпофільності молекул лікарських засобів за різними наборами молекулярних дескрипторів. Показано, що алгоритми штучних нейронних мереж є ефективними засобами вирішення такого роду задач прогнозування.

Вступ. Відкриття ліків – це складний і важкий процес. Щоб визначити, чи буде запропонована молекула успішним лікарським засобом, необхідно оцінити її ефективність і можливість лікування певних захворювань. На початковому етапі експериментальних досліджень із виявлення нових потенційних лікарських засобів необхідно перевірити велику кількість хімічних молекул, щоб виявити сполуки, які демонструють певну біологічну активність [1–3].

Ліпофільність важлива для прогнозування успіху молекули як лікарського препарату, оскільки є показником метаболізму, фармакокінетики та фармакодинаміки [1]. Експериментальне визначення ліпофільності вимагає кількісного визначення концентрації сполуки, присутньої у неводній та водній фазах. Є різні підходи щодо визначення ліпофільності, але, зазвичай, вони довготривалі та трудомісткі, кожен зі своїми недоліками та перевагами [4, 5].

Мета дослідження – оцінити ефективність алгоритмів штучної нейронної мережі прямого поширення сигналу та каскадної нейронної мережі щодо прогнозування ліпофільності молекул лікарських засобів із застосуванням різних наборів фізико-хімічних параметрів.

Методи дослідження. Реалізацію штучних нейронних мереж виконано із застосуванням програмного комплексу Matlab R2022b.

Штучні нейронні мережі прямого поширення сигналу, як і будь-яка інша штучна нейронна мережа, складається з шарів нейронів. Сигнали (тобто вхідні вектори чи вхідна інформація) надходять на вхідний рівень нейронів. Потім вони надходять та обробляються нейронами прихованого шару. Виходи (кінцевий результат) надають нейрони вихідного

SUMMARY

Anna Dorosh

Topic: “Comparison of methods for prediction the lipophilicity of drug molecules”

Department of analytical, physical and colloid chemistry

Scientific supervisor: Yaroslava Pushkarova

Keywords: artificial neural network, drug design, lipophilicity, multilinear regression.

Introduction. Drug discovery is a complex and difficult process. To determine whether a molecule will be proposed as a successful drug, it is necessary to evaluate its effectiveness and the possibility of treating certain diseases.

Lipophilicity refers to the tendency of a molecule to dissolve in lipids or non-polar solvents, indicating its affinity for fat or lipid-like substances. Lipophilicity is an important property in pharmacology and medicinal chemistry, especially in the context of drug development. It influences the absorption, distribution, metabolism, and excretion (ADME) of a drug in the body.

Lipophilicity is necessarily taken into account and evaluated at the stage of modeling new drugs, as it is one of the components of the successful use of the studied molecule as a potential drug. Artificial neural networks can be used for exploratory estimation of the lipophilicity parameter at the stage of research of candidate molecules for potential drugs.

Materials and methods. Feedforward artificial neural network and cascade neural network for prediction of lipophilicity of drug molecules.

Results. Data set contains 34 drug molecules, which are characterized by different physicochemical properties (thermodynamic parameters, electronic descriptors). It was established the efficiency of artificial neural networks for lipophilicity prediction at different set of descriptors.

The optimal number of neurons of the hidden layer of the feedforward artificial neural network and the cascade neural network for predicting the lipophilicity of drug molecules based on different sets of physicochemical

parameters does not exceed nine.

The conducted study of lipophilicity prediction using different sets of physicochemical descriptors and mathematical models allows us to conclude that more accurate results are obtained using a cascade neural network and only two descriptors (distribution coefficient and zero point energy).

Conclusion. The combination of the parameters of the distribution coefficient and the zero point energy is sufficient and effective for predicting the lipophilicity of drug molecules.

Algorithms of feedforward artificial neural network and cascade neural network are characterized by high accuracy of predicting the lipophilicity of drug molecules using different sets of physicochemical parameters.