

ДОДАТОК А

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Ім'я користувача:
Нечволод Вадим Юрійович каф. ПІ

ID перевірки:
1016362898

Дата перевірки:
15.06.2024 12:04:14 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
15.06.2024 16:17:25 EEST

ID користувача:
94949

Назва документа: 2024_М_ПІ_ІПЗм-22-1_Сергієнко_О_С_скорочений

Кількість сторінок: 36 Кількість слів: 6005 Кількість символів: 45453 Розмір файлу: 2.31 MB ID файлу: 1016168191

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

5%
Схожість

Найбільша схожість: 2.7% з Інтернет-джерелом (<https://simple.wikipedia.org/wiki?curid=122283>)

4.51% Джерела з Інтернету 176

Сторінка 38

0.63% Джерела з Бібліотеки 9

Сторінка 38

0% Цитат

Не знайдено жодних цитат

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи 5

Підозріле форматування 9 сторінок

ДОДАТОК Б

Слайди презентації



Дослідження методів програмної реалізації оцінки властивостей складних структур

Виконала:
ст. гр. ПЗМ-22-1 Сергієнко О.С.

Керівник:
доц. каф. ПІ Туруга О.П.



Актуальність дослідження

01 **Зростаюча потреба у нових лікарських засобах**

В умовах глобальних пандемій, таких як COVID-19, виникає гостра необхідність у швидкій та ефективній розробці нових медикаментів.

02 **Зменшення витрат і часу на дослідження**

Ефективні методи прогнозування допомагають значно скоротити витрати та час, необхідний для розробки нових лікарських препаратів, з урахунок зменшення кількості необхідних експериментів і оптимізації процесу дослідження.

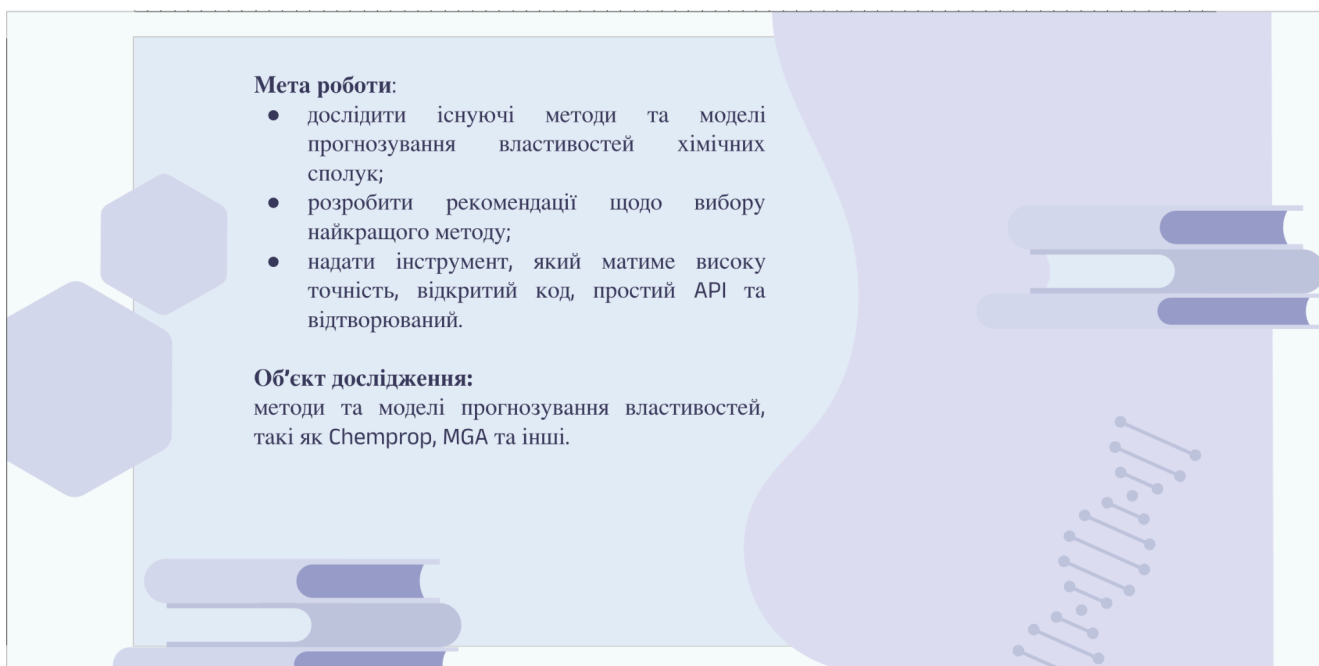
03 **Поліпшення безпеки**

Дослідження методів прогнозування ADME/Т властивостей сприяє створенню більш безпечних та ефективних лікарських засобів, зменшуючи ризик побічних ефектів та покращуючи терапевтичну ефективність препаратів.

Мета роботи:

- дослідити існуючі методи та моделі прогнозування властивостей хімічних сполук;
- розробити рекомендації щодо вибору найкращого методу;
- надати інструмент, який матиме високу точність, відкритий код, простий API та відтворюваний.

Об'єкт дослідження:
методи та моделі прогнозування властивостей, такі як Chemprop, MGA та інші.



Постановка задачі

Збір даних

Актуальні хімічні дані для передбачення метаболізму необхідно зібрати з таких баз даних як PubChem, ChEMBL



Експериментальні дослідження

Визначити ефективність моделей та порівняти їх точність, розробити рекомендації по покращенню

Пайплайн для обробки

Визначити правила для навчальних та прогнозованих даних, вирішити проблему сумісності та стандартизації даних з різних джерел

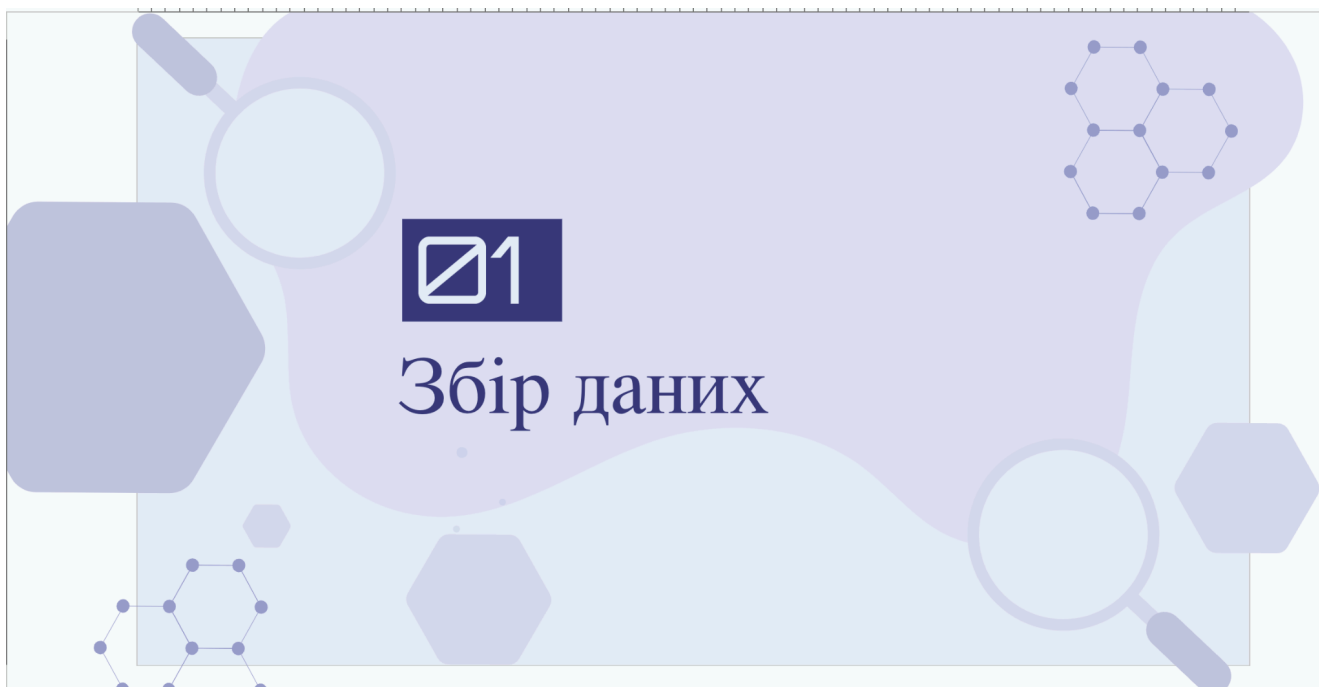


Розробка API

Розробка простого API для прогнозування ADMET властивостей з подальшою можливістю інтеграції

Огляд поточних рішень

Chemprop, MGA та інші



01

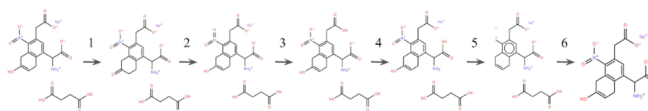
Збір даних

Дані з бази даних PubChem AID 1851

Метаболізм - це складна біохімічна система, відповідальна за управління енергетичними та молекулярними процесами в живих організмах.

Назва	Опис	Обробка	Кількість	Позитивні	Негативні
CYP1A2-Inhibitor	Це ймовірність бути інгібітором цитохрому CYP1A2, який каталізує метаболізм ліків. Інгібування CYP1A2 може бути як бажаною, так і не бажаною властивістю в залежності від цілей розробки ліків	Завантаження Фільтрування за «panel name» у p450-cyp1a2 Видалення недостовірних записів у «PUBCHEM_ACTIVITY_OUTCOME» Фільтрування TRUE за «Inhibition Observed» Встановлення Activity = 1, якщо «PUBCHEM_ACTIVITY_OUTCOME» є Active, і 0 у іншому випадку	13239	5997	7242

Пайплайн для обробки структур



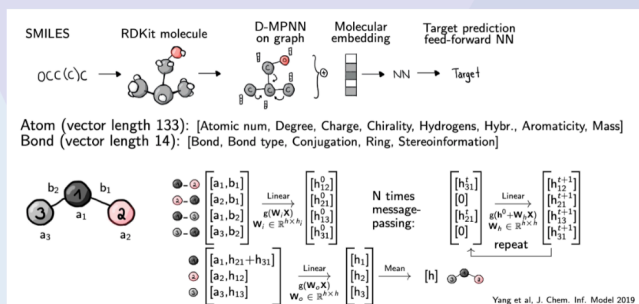
- 1 Tautomerization
- 2 Chemotype standardization
- 3 Neutralization
- 4 Normalization
- 5 Aromatization
- 6 Main component extraction

Перший етап обробки даних забезпечує сумісність та стандартизацію даних з різних джерел для моделювання. Ми видаляємо дані, що не мають відношення до структури та властивостей/активності та додаємо додаткові дані, такі як тип джерела, одиниці виміру та молекулярні мішені.

02

Огляд поточних рішень

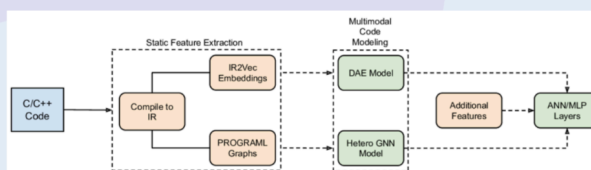
Chemprop-RDKit6

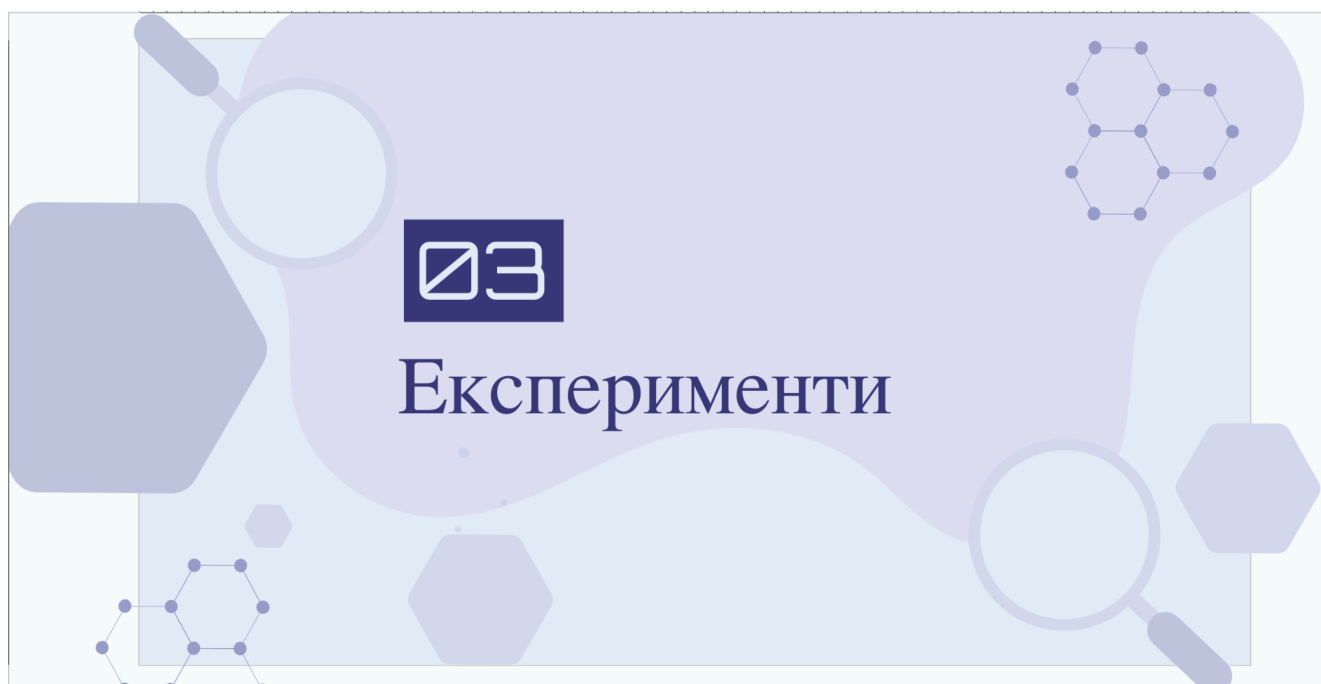


- використовує графову нейронну мережу;
- використовує пакет хемінформатики RDKit для обчислення різноманітних фізико-хімічних ознак для кожного атома та зв'язку у молекулі





Фреймворк MGA

Переваги	Недоліки
Ефективне одночасне вирішення регресійних і кваліфікаційних завдань	Потреба у великій кількості даних для навчання ефективної моделі
Здатність враховувати структурні особливості молекул за допомогою графової уваги	Складність налаштування гіперпараметрів для досягнення оптимальної продуктивності
Можливість налаштування для різних завдань прогнозування ADMET	





Chemprop-RDKit6

Назва	Особливість	Чутливість	Точність	Збалансова на точність	ROC AUC
CYP1A2-Inhibitor	0.899	0.956	0.925	0.928	
CYP3A4-Inhibitor	0.861	0.949	0.896	0.905	
CYP3A4-Substrate	0.562	0.857	0.776	0.709	
CYP2C19-Inhibitor	0.887	0.890	0.888	0.889	
CYP2C9-Inhibitor	0.899	0.880	0.893	0.890	

API та інтеграція

Screenshot of a web application interface for Chemprop-RDKit6, showing a REST client and a data visualization dashboard.

REST Client (POST):

```

url: http://127.0.0.1:5000/api/upload?models=Caco2
body:
  1: smiles
  2: CCC
  
```

Data Visualization Dashboard:

- smiles:** 10 categories, Most diverse structures, 98
- piechart:** CYP1A2-inhibitor, CYP2C19-inhi
- Metabolism:**
 - CYP1A2-inhibitor: 0.98
 - CYP2C19-inhibitor: 0.18
 - CYP2C9-inhibitor: 0.03
 - CYP2C9-Substrate: 0.24
 - CYP2D6-inhibitor: 0.68
 - CYP2D6-Substrate: 0.41
- Absorption:**
 - Pgp-Substrate: 0.82
 - Caco2: -4.58
 - Lipophilicity: 2.84
 - Solubility: -3.51
- Excretion:**
 - CL-Hepa: 44.45
 - CL-Micro: 32.94
 - Half Life: 8.58
- Distribution:**
 - PPBR: 96.24
 - VDss: 4.52

ДОДАТОК В

Апробація результатів роботи

УДК 004.85

DOI: <https://doi.org/10.30837/IYF.IIS.2024.442>

ДОСЛІДЖЕННЯ МЕТОДІВ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ОЦІНКИ ВЛАСТИВОСТЕЙ СКЛАДНИХ СТРУКТУР

Сергієнко О. С.

Науковий керівник – к.т.н., доцент Турута О. П.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: oleksandra_serhiienko@nure.ua

This work is dedicated to validating QSAR models, which are pivotal for predicting molecular activity based on structural attributes, with a particular focus on ADMET. It meticulously examines the performance of two prominent methodologies: chemprop, employing graph neural networks for biological activity prediction, and a SVM classification model utilizing molecular fingerprints. Diverse mathematical approaches are rigorously assessed, employing metrics such as balanced accuracy (BA) and area under the ROC curve (AUC). The ultimate objective of this research is to enhance the efficiency of evaluating molecular properties, offering promising implications in the realm of drug discovery.

Актуальність дослідження ADMET властивостей складних структур у галузі медицини та фармацевтики пояснюється наявністю постійної потреби у нових лікарських засобах, особливо для подолання пандемій, таких як COVID-19. Використання сучасних комп'ютерних методів аналізу даних дозволяє ефективно оцінювати властивості молекул, що значно спрощує процес досліджень, зменшує ризики невдалих експериментів та дозволяє оперативно реагувати на зміни в галузі громадського здоров'я. Такі дослідження є ключовими для розвитку сучасної фармацевтичної науки та сприяють створенню більш ефективних та безпечних лікарських препаратів.

Метод (Q)SAR відіграє важливу роль у сучасній медичній хімії та фармацевтиці, особливо в контексті пошуку нових лікарських засобів. Цей метод відображає залежність між характеристиками молекули x та її активністю або властивістю y . Для оцінки активності молекули за її характеристиками використовуються статистичні оцінки та методи машинного навчання. Вибір характеристик молекули x для створення (Q)SAR моделі є важливим завданням. Один з підходів – це представлення молекули у вигляді вектора чисел – дескрипторів, що описують її хімічну структуру та властивості. Інші підходи можуть використовувати 3D дескриптори або графові нейронні мережі для аналізу молекулярної структури.

Для побудови QSAR моделей використовуються різні математичні підходи, такі як метод опорних векторів, логістична регресія, випадковий ліс та штучні нейронні мережі. Ці методи відрізняються за кількістю

використовуваних змінних, інтерпретованістю та прогностичною здатністю.

У даному дослідженні використовувалися дані з бази даних Tox21, яка містить інформацію про токсичність хімічних сполук. Дані Tox21 представляють собою масиви даних, які є не збалансованими та містять різні співвідношення активних та неактивних сполук.

У ході експерименту використовувалися різні моделі, включаючи моделі на основі глибокого навчання, такі як chemprop, і класичні класифікаційні моделі на основі молекулярних фінгерпринтів.

Модель chemprop є методом для передбачення біологічної активності молекул на основі їх хімічної структури. Вона використовує графові нейронні мережі для аналізу молекулярної структури та передбачення їх властивостей.

Класифікаційна модель на основі молекулярних відбитків є методом, заснованим на представленні молекули у вигляді бінарного вектора, де кожен елемент відповідає наявності або відсутності певного хімічного фрагмента. Цей метод використовується для передбачення біологічної активності молекул на основі їх структури.

Для порівняння і валідації моделей використовувалися різні критерії оцінки точності, включаючи збалансовану точність та площу під ROC-кривою при крос-валідації. Ці критерії були обрані для оцінки прогностичної здатності моделей на перехресному контролі.

Збалансована точність при крос-валідації визначається як середнє арифметичне точності для кожного класу:

$$BA = \frac{1}{n_{classes}} \sum_{i=1}^{n_{classes}} \frac{TP_i}{TP_i + FN_i},$$

де:

TP_i – це кількість вірно передбачених позитивних прикладів для класу i ,

FN_i – це кількість хибно відкинутих прикладів для класу i ,

Площа під ROC-кривою при крос-валідації обчислюється як інтеграл від ROC-кривої:

$$AUC = \int_0^1 TPR(FPR^{-1}(t)) dt,$$

де:

TPR – чутливість,

FPR – специфічність,

$FPR^{-1}(t)$ – обернена функція до функції, що повертає значення FPR при заданому порозі t .

Результати порівняння за наведеними вище критеріями наведені у таблиці.

Таблиця 1 – Результати валідації моделей

Модель	Збалансована точність (BA)	Площа під ROC-кривою (AUC)
Chemprop	0.85	0.92
Класифікаційна модель (SVM)	0.78	0.86

Ці результати свідчать про вищу прогностичну здатність моделі chemprop у порівнянні з класифікаційною моделлю на основі молекулярних відбитків. Для подальшого розвитку дослідження, рекомендується акцентувати увагу на оптимізації параметрів та розширенні набору даних для навчання моделі chemprop. Також, варто розглянути можливість поєднання chemprop з іншими методами машинного навчання для досягнення ще кращих результатів.

Список використаних джерел:

1. In silico Prediction of Chemical Ames Mutagenicity / Congying Xu, Feixiong Cheng, Lei Chen та ін. // J Cheminform. 2012. URL: <https://pubs.acs.org/doi/abs/10.1021/ci300400a> (дата звернення: 20.01.2024).
2. Computational Models for Human and Animal Hepatotoxicity with a Global Application Scope / Denis Mulliner, Friedemann Schmidt, Manuela Stolte та ін. // Chem. Res. Toxicol. 2016. URL: <https://pubs.acs.org/doi/10.1021/acs.chemrestox.5b00465> (дата звернення: 13.12.2023).
3. Turuta O. Collection of questionnaire results, received by using the visual analog scale method, for its further processing in the medical web application / O. Turuta, Y. Daniil // ScienceRise. 2017. URL: https://www.researchgate.net/publication/317612987_Collection_of_questionnaire_results_received_by_using_the_visual_analog_scale_method_for_its_further_processing_in_the_medical_web_application (дата звернення: 20.03.2024).
4. Intelligent information system of heterogeneous medical data analysis / A.Yerokhin, O. Turuta, A. Babii, A. Nechyporenko. 2017. URL: https://www.researchgate.net/publication/320968847_Intelligent_information_system_of_heterogeneous_medical_data_analysis (дата звернення: 23.03.2024).

ДОДАТОК Г

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІПЗМ-21-1
(група)

Сергієнко Олександра Сергіївна

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт

(підпис)

Олена ОЛІЙНИК
(прізвище, ініціали)

17.06.2024