

Голові разової спеціалізованої вченої ради  
Національного університету  
«Львівська політехніка»  
д.т.н., професору Наталії ШАХОВСЬКІЙ

## **ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА**

доктора технічних наук, професора Гнатушенка Володимира Володимировича  
на дисертаційну роботу Гурбича Олександра Вікторовича  
«Методи та засоби аналізу хімічних сполук засобами штучного інтелекту»,  
подану на здобуття наукового ступеня доктора філософії за  
спеціальністю 122 Комп'ютерні науки (12 – Інформаційні технології)

### **Актуальність теми дослідження.**

Розробка нових лікарських речовин — це тривалий ітеративний процес, який включає кілька етапів, включаючи відкриття, обчислювальні випробування, доклінічні та клінічні дослідження, перевірку та схвалення регуляторних органів. Щоб вивести на ринок новий препарат, може знадобитися до 15 років і буде коштувати мільярди доларів. На жаль, більшість молекул-кандидатів відхиляються регуляторними організаціями через такі проблеми, як токсичність і низька селективність, що робить процес ще дорожчим. Насправді лише 12% нових молекул, які проходять клінічні випробування, врешті-решт схвалюються FDA. З цієї причини, менше ніж 5% рідкісних захворювань мають доступні ліки. Однак сучасна наука лише почала досліджувати множину потенційних лікарських речовин. І зараз хімічна та фармацевтична промисловість переживають пік трансформації, при якій методи машинного навчання можуть підвищити ефективність розробки нових лікарських молекул. Але обмеженість існуючих теоретико-практичних засад застосування методів машинного навчання при аналізі хімічних сполук потребує розвитку нових підходів до цієї проблематики і визначає актуальність дисертаційної роботи Олександра Гурбича, що полягає у розробці методів та засобів аналізу хімічних сполук засобами штучного інтелекту з подальшим поєднанням створених елементів у вигляді інформаційної системи повного циклу дизайну ліків.

**Ступінь обґрунтованості наукових положень, висновків та рекомендацій.** Основні теоретичні положення дисертації одержані шляхом коректного застосування методів опорних векторів, градієнтного бустингу, машинного навчання та математичного моделювання. Прогнозування виходу

продукту хімічної реакції виконувалось деревами рішень, нейронними мережами прямого поширення та моделями-трансформерами. Аналіз змісту розділів, використаної інструментарію та способів його застосування дозволяє зробити висновок про належну обґрунтованість отриманих автором наукових результатів. Наукові положення, висновки та рекомендації, сформульовані у дисертації, обґрунтовано теоретичним аналізом і підтверджено результатами практичного використання.

Достовірність результатів підтверджується також комп'ютерною реалізацією розроблених алгоритмів з можливістю кількісного та візуального контролю її результатів, тестовими розрахунками з прогнозованими результатами та впровадженнями роботи, підтвердженими відповідними документами.

Все це свідчить про високий ступінь достовірності та обґрунтованості результатів дисертації О.В. Гурбича.

### **Структура, обсяг роботи.**

Дисертація загальним обсягом 251 сторінка складається із вступу, п'яти розділів, висновків, списку використаних джерел із 292 найменувань та додатків, у які винесені лістинги програмних модулів та документи, що підтверджують впровадження результатів роботи. Основний текст дисертації займає 127 сторінок. Оформлення дисертації відповідає усім необхідним вимогам.

### **Характеристика роботи, новизна розроблених наукових положень.**

У вступі автором подано загальну характеристику дисертації, визначено актуальність теми, сформульовано мету дослідження, окреслено коло наукових та прикладних задач, розв'язання яких забезпечує реалізацію мети роботи, показана наукова новизна та практична цінність роботи. Наведено відомості про апробацію та публікації результатів досліджень.

У першому розділі дисертантом проведений критичний аналіз літературних джерел та існуючих методів та підходів машинного навчання до прогнозування молекулярної спорідненості, фактичного виходу продукту хімічної реакції та інформаційної розробки молекулярних структур. За результатами огляду охарактеризовано недоліки існуючих досліджень в галузі застосування інформаційних технологій для розробки ліків та сформовано завдання дисертаційної роботи. Результатом цього розділу можна вважати сформовані автором функціональні вимоги до платформи для розробки нових лікарських речовин.

У другому розділі автором запропоновано метод мета-навчання для прогнозування молекулярної спорідненості між рецептором (велика біомолекула) та лігандами (малі органічні молекули). З практичних результатів

розділу слід відзначити те, що запропоноване дисертантом поєднання моделей методом мета-стекингу збільшує Recall класифікації на 34,9 % і підвищує загальну точність та статистичну достовірність результатів (R2) на 21 %, що дозволяє виключити або компенсувати помилки, допущені іншими моделями ансамблю.

Третій розділ роботи Гурбича О.В. присвячений створенню методу прогнозування виходу продукту хімічної реакції, що виконується за допомогою нової архітектури графової нейронної мережі. Моделі були натреновані та оцінені на двох публічних наборах даних із одного класу реакцій та одному наборі даних від компанії Enamine із десяти типів реакцій. Дослідження доповнене аналізом даних, результатів і помилок, а також впливу просторових факторів, побічних реакцій, ефективності виділення та очищення на результати прогнозування. Спрощення повнозв'язної мережі оцінки відсотку виходу продукту хімічних реакцій дозволило автору покращити значення функції втрат (RMSE) на 3.26 % та коефіцієнту детермінації на 15.43 % із 70.79 % залишкових ваг.

У четвертому розділі автором запропоновано метод, що поєднує кілька глибоких нейронних мереж для створення молекулярних структур із заданими властивостями. Генерування доповнюється виправленням хімічної будови молекул за допомогою рекурентної нейронної мережі з механізмом уваги. Вагомим результатом розділу є підтверджені автором при проведенні серії комп'ютерних симуляцій стабільність та передбачена розчинність згенерованих молекул методами молекулярної динаміки.

У п'ятому розділі дисертантом розроблено за формулюванням автора «архітектурну діаграму інформаційної системи» для створення лікарських речовин із бажаними біологічними та фізико-хімічними властивостями. Інформаційна система поєднує методи, описані дисертантом у попередніх розділах. Сукупність результатів, пов'язаних з комп'ютерною реалізацією запропонованих алгоритмів, дає підстави трактувати їх як цілісний програмно-методичний комплекс створення лікарських речовин із бажаними біологічними та фізико-хімічними властивостями, важливою характеристикою якого є безпосередня інтеграція із постачальниками органічних речовин.

У додатках наведені лістинги розроблених комп'ютерних програм та документи впровадження результатів дисертації.

**Наукову новизну дисертаційної роботи представляють такі основні та вагомні результати:**

- Вперше розроблено метод прогнозування молекулярної спорідненості, що поєднує послідовні ансамблі моделей машинного навчання першого рівня та узагальнені лінійні моделі другого рівня для покращеної точності.

- Вдосконалено архітектуру графової нейронної мережі для передбачення виходу продукту хімічної реакції шляхом додавання інформації про учасників хімічного перетворення, а також молекулярні дескриптори.
- Вперше розроблено метод генерування молекулярних структур, що дозволяє контролювати одну або більше властивостей генерованих молекул, а також має послідовний модуль виправлення помилок хімічної будови. Модель виправлення хімічних помилок підвищує вихід коректних молекулярних структур на 20 %, вихід унікальних молекул - на 67 % порівняно із моделлю-генератором без модулю виправлення помилок.

### **Прикладна цінність дисертації.**

Практичне значення одержаних результатів підтверджується впровадженнями в ТЗОВ "СофтСерв-Індустрія", й у комерційному проекті HTG Molecular Diagnostics. Результати роботи також впроваджені у освітньому процесі Національного університету «Львівська політехніка» при викладанні освітнього компонента «Нейромережеві технології та їх застосування». Також автором розроблено діаграму мікросервісної архітектури інформаційної системи для розрахункового синтезу молекул із заданими властивостями.

### **Зв'язок роботи з науковими програмами.**

Дисертація Гурбича О.В. виконана на кафедрі систем штучного інтелекту Національного університету «Львівська політехніка». Дисертаційне дослідження виконано у рамках тематики науково-дослідної роботи НУ «Львівська політехніка» «Інформаційна технологія формування психофізичного портрету в умовах стресових ситуацій» (№ держ. реєстр. 0119U002257). Це підтверджено актом впровадження (стор. 160 дисертації). Результати роботи були також використані для створення комп'ютерних систем ранньої розробки лікарських речовин у ТОВ «СофтСерв-Індустрія» та у комерційній розробці програмного забезпечення HTG Molecular Diagnostics.

### **Рекомендації щодо впровадження результатів дисертації.**

Коло практичних застосувань результатів роботи, на наш погляд, не обмежується розглянутими у ній впровадженнями. Так, наприклад, запропоновану автором роботи у другому розділі графову нейронну мережу можна взяти за основу та використовувати для передбачення «спорідненості» спільнот у соціальних мережах, а генератор молекулярних структур (Розділ 4) після відповідного дотренування та заміни словника можна використовувати для генерації будь-яких коротких текстових даних.

## **Оформлення дисертації, дотримання вимог академічної доброчесності та повнота викладу наукових положень та результатів в опублікованих працях**

Дисертаційна робота Гурбича О.В. має логічну структуру. Висновки та подальші кроки логічно витікають із результатів попередніх розділів та огляду літератури. У кінці кожного розділу є проміжні висновки, що підсумовують результати автора. Стиль викладення матеріалів дисертаційного дослідження та наукових положень доказовий і в цілому забезпечує доступність їх сприйняття.

У дисертаційній роботі чітко прослідковується авторський стиль подачі матеріалу та формування висновків. Аналіз структури та змісту дисертаційної роботи та наукових праць, що опубліковані автором, дозволяє стверджувати, що усі наукові та практичні результати отримані ним особисто і повною мірою опубліковані та апробовані. У дисертації не виявлено текстових запозичень і використання наукових результатів інших науковців без посилань на відповідні джерела. Підстав для сумнівів у науковій доброчесності здобувача під час детального ознайомлення з дисертацією не виявлено.

Основні результати дисертації досить повно висвітлено у 10 роботах автора: 3 статті у наукових фахових виданнях України, 5 статей у наукових виданнях інших держав, а також 2 праці апробаційного характеру – у матеріалах і тезах конференцій та препринтах. Основі положення дисертації повністю викладено в опублікованих працях. Вимоги щодо кількості та якості публікацій виконано.

Тема, зміст та отримані наукові результати роботи відповідають спеціальності 122 «Комп'ютерні науки».

### **Зауваження по роботі.**

1. У першому розділі при проведенні аналізу існуючих методів машинного навчання до прогнозування молекулярних властивостей автору варто було б акцентувати більше уваги на обмеженнях/недоліках підходів інших авторів та використати їх для обґрунтованої постановки завдань дисертаційної роботи.

2. Не повністю розкриті потенційні можливості та обмеження запропонованого у другому розділі методу прогнозування молекулярної спорідненості, основу якого складає мета-навчання. Автор обмежується лише порівнянням повної та спрощеної FFN для передбачення молекулярної спорідненості (табл.2.2).

3. Результати модифікованого автором методу редукції порівнюються лише із повними (нередукованими) моделями. Слід було провести порівняння результатів автора із методами прунінгу нейронних мереж інших авторів та випробувати модифікований метод редукції на публічних наборах даних, які використовувалися іншими дослідниками.

4. Автором не обґрунтовано вибір мікросервісної архітектури інформаційної системи, немає порівняння переваг та недоліків з іншими архітектурними підходами.

5. Поза увагою залишені питання числової стійкості та необхідних обчислювальних ресурсів розроблених автором моделей та методів при їхній комп'ютерній реалізації.

6. Висновки по розділах бажано б було закінчувати фразою: «*Основні наукові результати розділу опубліковані в працях [...]*». Посилання на власні праці автора дало б змогу легко пересвідчитися у виконанні вимоги щодо обов'язкової публікації основних результатів дисертації.

7. Апробація результатів синтезу молекулярних структур із контролем розчинності дисертантом проводилася у симуляційному експерименті. Автору доречно було б визначити реальну розчинність синтезованої нової лікарської речовини у хімічній лабораторії у реальному розчиннику та порівняти із прогнозованою розчинністю.

8. У тексті роботи багато спеціальних термінів, зокрема англомовного походження, які не мають загальновідомих визначень, що ускладнює сприйняття окремих розділів. Перелік умовних позначень містить окремі аббревіатури, які взагалі не згадуються у тексті роботи.

9. Список використаних джерел дисертантом оформлено за різними стилями та з помилками.

10. Наведені у додатку лістинги розроблених програмних модулів майже не супроводжуються коментарями, що утруднює їхню змістову інтерпретацію. В тексті дисертації зустрічаються друкарські помилки та стилістичні вади.

Однак, наведені зауваження мають окремий характер, не знижують високий науковий рівень та практичну цінність дисертаційної роботи і суттєво не впливають на її загальну позитивну оцінку.

### **Загальна оцінка дисертаційної роботи.**

Дисертація Гурбича Олександра Вікторовича на тему «Методи та засоби аналізу хімічних сполук засобами штучного інтелекту» є завершеною науково-дослідницькою працею, що містить нові науково обґрунтовані результати та виконана на високому науковому рівні. Зміст роботи засвідчує про її цілісність, яка включає всі необхідні стадії — від формулювання теми дослідження до впровадження здобутих результатів у практику. Для дисертації характерний тісний логічний зв'язок окремих питань дослідження. У дисертації вирішено актуальну науково-прикладну задачу підвищення ефективності розробки нових лікарських речовин із заданими властивостями з використанням методів та засобів штучного інтелекту.

Одержані висновки та практичні результати представляють наукову новизну та практичну цінність і є значущими для галузі інформаційних технологій. Вважаю, що за актуальністю теми дисертації, обґрунтованістю наукових результатів, їх новизною та практичною цінністю, повнотою викладу матеріалу в наукових публікаціях дисертація Гурбича О.В. відповідає існуючим вимогам відповідного «Порядку присудження ступеня доктора філософії...», а її автор – Гурбич Олександр Вікторович – заслуговує на присудження йому наукового ступеня доктора філософії за спеціальністю 122 Комп'ютерні науки (галузі знань 12 – Інформаційні технології)).

**Офіційний опонент —**

Завідувач кафедри інформаційних технологій  
та комп'ютерної інженерії  
Національного технічного університету  
«Дніпровська політехніка»,  
доктор технічних наук, професор



Володимир ГНАТУШЕНКО

Підпис проф. В.В. Гнатушенка підтверджую

**Вчений секретар**

Національного технічного університету  
«Дніпровська політехніка»



Тайсія КАЛЮЖНА