

УДК 004.8

Pavlo Krysenko, Doctor of Philosophy, Research Assistant
ORCID ID: <https://orcid.org/0000-0002-5612-9474> **e-mail:** p.krysenko@gmail.com

Alim Bektimirov, lead engineer
ORCID ID: <https://orcid.org/0009-0007-8572-7952> **e-mail:** alim.bektimirov@gmail.com

Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

THE PROBLEM OF IC50 PREDICTION FOR LIGAND-PROTEIN PAIRS USING TRANSFORMER ARCHITECTURE UNDER LIMITED RESOURCES

***Annotation.** The article discusses approaches to optimizing the training of models for predicting the half-maximal inhibitory concentration (IC50) of ligand-protein pairs under limited computational resources. A method of smart bucketing of data by protein length with a dynamic selection of the number of groups to improve randomization is proposed. To solve the problem of the quadratic complexity of the Transformer architecture, a convolution layer was used to compress the input data. Based on 4 conducted experiments, the relationship between the degree of sequence compression and the obtained root mean square error (RMSE) for lgIC50 was analyzed.*

***Key words:** deep learning, ligand, protein, bioinformatics, convolutional neural networks, data compression.*

П. Крисенко, А. Бектіміров

Інститут телекомунікацій і глобального інформаційного простору НАН України, м. Київ, Україна

ПРОБЛЕМАТИКА ПРОГНОЗУВАННЯ ПОКАЗНИКА IC50 ДЛЯ ПАР ЛІГАНД-БІЛОК ЗА ДОПОМОГОЮ АРХІТЕКТУРИ TRANSFORMER В УМОВАХ ОБМЕЖЕНИХ РЕСУРСІВ

***Анотація.** У статті розглядаються підходи до оптимізації навчання моделей для прогнозування напівмаксимальної інгібіторної концентрації (IC50) пар ліганд-білок в умовах обмежених обчислювальних ресурсів. Запропоновано метод розумного групування (bucketing) даних за довжиною білка з динамічним підбором кількості груп для покращення рандомізації. Для вирішення проблеми квадратичної складності архітектури Transformer використано шар згортки для компресії вхідних даних. На основі 4 проведених експериментів проаналізовано співвідношення між ступенем стиснення послідовності та отриманим коренем з середньоквадратичної похибки (RMSE) для lgIC50.*

***Ключові слова:** глибоке навчання, ліганд, білок, біоінформатика, згорткова нейронна мережа, компресія даних.*

<https://doi.org/10.32347/2411-4049.2026.2.287-292>

© П. Крисенко, А. Бектіміров, 2026

Вступ

Прогнозування спорідненості зв'язування між лігандом та білком-мішенню є ключовим етапом у сучасному комп'ютерному дизайні ліків [1]. Однією з найважливіших кількісних метрик цієї взаємодії є напівмаксимальна інгібіторна концентрація (IC50) [2]. З розвитком методів глибокого навчання, зокрема архітектур на основі механізмів уваги, точність прогнозування значно зросла. Проте використання класичного Transformer стикається з серйозною перешкодою – квадратичною обчислювальною складністю відносно довжини вхідної послідовності. Білкові ланцюги можуть бути надзвичайно довгими, що призводить до експоненційного зростання споживання пам'яті та унеможливорює ефективне навчання моделей в умовах обмежених апаратних ресурсів. Метою цієї роботи є дослідження методів оптимізації пайплайну навчання для зменшення обчислювального навантаження без значної втрати точності прогнозування.

Результати дослідження

Для подолання жорстких обмежень обчислювальних ресурсів під час дослідження було розроблено та імплементовано комплекс оптимізаційних підходів: на рівні попередньої обробки даних та на рівні самої архітектури нейронної мережі.

При формуванні пакетів даних для навчання моделей на послідовностях різної довжини традиційно використовується додавання нульових токенів (padding) до довжини найдовшого елемента в батчі. У випадку білків це призводить до того, що відеокарта витрачає ліву частку обчислень на обробку «порожніх» токенів. Прямолінійне сортування [4] всього датасету за довжиною частково вирішує цю проблему, але порушує базове правило стохастичного градієнтного спуску щодо випадковості вибірки, що веде до перенавчання моделі.

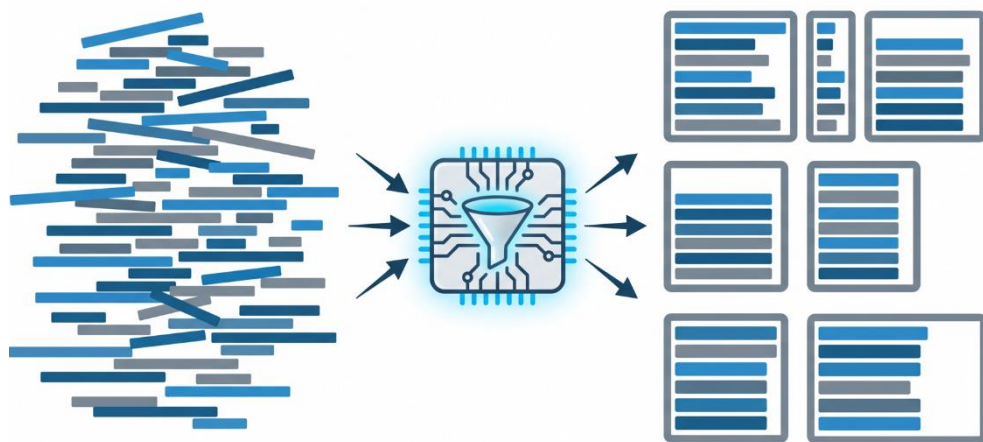


Рис. 1. Схематичний вигляд алгоритму smart bucketing

У даній роботі був застосований алгоритм розумного групування (smart bucketing). Він динамічно аналізує розподіл довжин білків у навчальній вибірці та самостійно розбиває дані на оптимальну кількість груп таким чином, щоб усередині кожного кошика довжини послідовностей були максимально близькими. При формуванні батчу алгоритм випадковим чином обирає кошик, а потім дістає звідти необхідну кількість семплів. Це радикально економить пам'ять, зберігаючи необхідний рівень рандомізації між ітераціями.

Основою запропонованої моделі є адаптована архітектура типу Енкодер-Декодер (Encoder-Decoder) [3], яка найкраще підходить для задач парної взаємодії, якою і є зв'язування ліганду з білком-мішенню, що зображена на рис. 2. Замість базової конкатенації ознак обох молекул, архітектура обробляє їх паралельно з подальшим структурним злиттям.

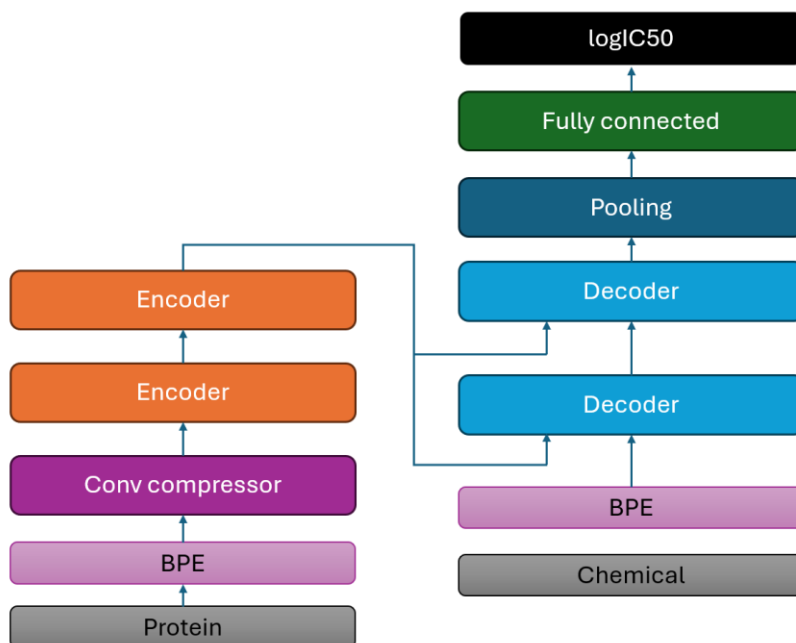


Рис. 2. Схематичне зображення архітектури

Енкодер білка (Protein Encoder): саме на цьому етапі класичний механізм самоуваги (Self-Attention) стикається з проблемою квадратичної складності. Щоб Transformer не «страждав» від надмірної довжини вхідних даних, перед блоками уваги було інтегровано модуль компресії, який імплементований за допомогою одновимірної згортки (1D Convolution), який зображений на рис. 3. Цей шар діє як локальний екстрактор ознак, виявляючи біологічні мотиви (наприклад, альфа-спіралі) та здійснюючи просторову компресію вхідних послідовностей (ембедингів). Завдяки згортці з визначеним кроком (stride), фізична довжина послідовності зменшується кратно, що експоненційно знижує навантаження на наступні шари Transformer Encoder, які формують глибоке контекстне розуміння структури білка.

Ключовим етапом прогнозування IC50 є моделювання фізико-хімічної взаємодії між двома молекулами. Для цього використовується декодер з механізмом перехресної уваги (Cross-Attention) [6]. У цій конфігурації стиснуті

та контекстуалізовані репрезентації білка виступають як ключі та значення (Keys, Values), а репрезентації ліганду — як запити (Queries).

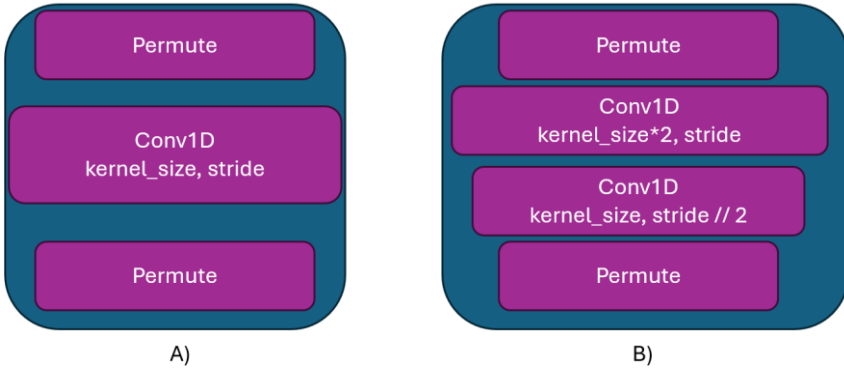


Рис. 3. Модулі A) small conv compression та B) medium conv compression

Це дозволяє моделі динамічно «фокусувати» увагу на тих ділянках білка (наприклад, активних центрах або сайтах зв'язування), які мають найбільшу афінність до конкретних функціональних груп ліганду.

Після блоку Cross-Attention отримані вектори ознак агрегуються і подаються на багатошаровий перцептрон (MLP), який безпосередньо регресує значення логарифму напівмаксимальної інгібіторної концентрації (lgIC50).

Для оновлення ваг моделі було обрано алгоритм оптимізації AdamW [7]. На відміну від стандартного Adam, AdamW відокремлює зменшення ваг (weight decay) від кроку оптимізації. У поєднанні з архітектурою Transformer це критично важливо, оскільки забезпечує значно кращу генералізаційну здатність моделі та запобігає її швидкому перенавчанню на складних патернах ліганд-білкових взаємодій.

Для перевірки гіпотез було проведено серію з 4 експериментів. На рис. 4 зображено значення лосс функції для тренувальної та тестувальної вибірок для кожного з експериментів в залежності від епохи. Головною змінною виступало співвідношення компресії вхідної послідовності (зміна параметру stride (S) у згортковому шарі білкового енкодера). Оцінювався вплив цього коефіцієнта на фінальну метрику RMSE для lgIC50.

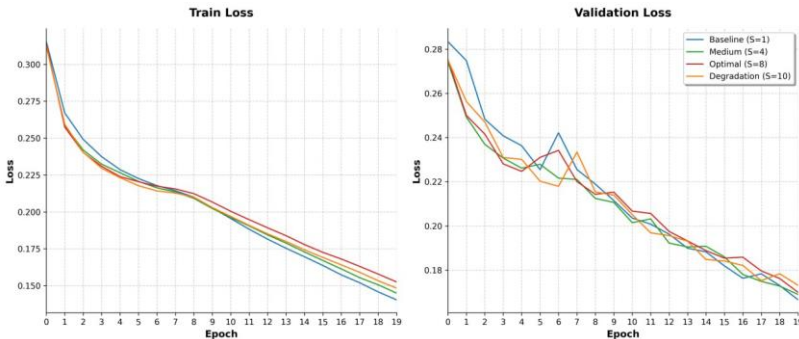


Рис. 4. Графік лосс функцій для 4 експериментів на тренувальній та тестувальній вибірках

Після отримання найкращої моделі для кожного з експериментів було підраховано корінь з середньоквадратичної похибки для тестувальної вибірки. Результати показані в таблиці. 1.

Таблиця 1. Результати моделей на тестувальній вибірці

Модель	RMSE
Stride = 1, Kernel size = 3	0.5914
Stride = 4, Kernel size = 9	0.5951
Stride = 8, Kernel size = 17	0.5952
Stride = 10, Kernel size = 21	0.6010

Висновки

Проведена серія з 4 експериментів переконливо доводить, що квадратична обчислювальна складність архітектур на базі механізмів уваги не є нездоланною перешкодою для їх використання в задачах біоінформатики за умов суворо обмежених апаратних ресурсів.

Аналіз отриманих результатів дозволяє зробити декілька ключових висновків щодо запропонованого оптимізаційного пайплайну.

Інтеграція шару 1D-згортки в Енкодер білка виконала подвійну функцію. З одного боку, вона забезпечила критично необхідну просторову компресію послідовності, кратно зменшивши споживання відеопам'яті. З іншого боку, згортка спрацювала як ефективний індуктивний зсув. Вона агрегувала локальні амінокислотні мотиви (елементи вторинної структури білка) у більш інформативні токени ще до того, як Transformer почав розраховувати глобальні залежності. В ході експериментів було визначено оптимальний поріг стиснення (баланс між розміром ядра та кроком згортки), після якого метрика MSE для IgIC50 починає різко деградувати, що свідчить про межу незворотної втрати структурної інформації.

Практична цінність роботи полягає у суттєвому зниженні бар'єра входження для проведення досліджень у галузі *in silico* скринінгу. Запропонована архітектура з компресією вхідного простору не лише дозволяє розгортати високоточні моделі на апаратному забезпеченні споживчого класу, але й створює фундамент для подальшої оптимізації подібних моделей під інференс на пристроях із жорсткими обмеженнями за кількістю параметрів та пам'яттю (edge computing). Подальші дослідження можуть бути спрямовані на застосування розроблених методів компресії до тривимірних (3D) просторових графів молекул.

СПИСОК ЛІТЕРАТУРИ \ REFERENCES

- [1] Zoziuk, M., Krysenko, P., Dovgiy, S., Makarov, V., Yakimenko, Y., & Koroliouk, D. (2026, April 10). Transformer with BPE tokenization for analysis of interactions of chemical substances and proteins. *Computational Methods and Mathematical Modeling in Cyberphysics and Engineering Applications* 2, pp. 289–299. doi:10.1002/9781394454518.ch9
- [2] Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* (Oxford, England), 34(17), i821–i829. doi:10.1093/bioinformatics/bty593

- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [4] Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). Efficient Transformers: A Survey. *ACM Comput. Surv.*, 55(6). doi:10.1145/3530811
- [6] Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C., & Sun, J. (2021). DeepPurpose: a deep learning library for drug-target interaction prediction. *Bioinformatics (Oxford, England)*, 36(22–23), 5545–5547. doi:10.1093/bioinformatics/btaa1005
- [7] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. doi:10.48550/arXiv.1711.05101

Стаття надійшла до редакції 10.02.2026, надійшла після рецензування 23.03.2026, прийнята 16.04.2026

The article was received 10.02.2026, received after revision 23.03.2026, accepted 16.04.2026

Крисенко Павло

доктор філософії, молодший науковий співробітник, Інститут телекомунікацій і глобального інформаційного простору НАН України

Адреса робоча: 03186, м. Київ, Чоколівський бульвар, 13

ORCHID ID: <https://orcid.org/0000-0002-5612-9474> **e-mail:** p.krysenko@gmail.com

Бектіміров Алім

провідний інженер, Інститут телекомунікацій і глобального інформаційного простору НАН України

Адреса робоча: 03186, м. Київ, Чоколівський бульвар, 13

ORCHID ID: <https://orcid.org/0009-0007-8572-7952> **e-mail:** alim.bektimirov@gmail.com