

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Львівський національний університет імені Івана Франка**  
**Біологічний факультет**  
**Кафедра біофізики та біоінформатики**

**Затверджено**

на засіданні кафедри

біофізики та біоінформатики

біологічного факультету

Львівського національного університету імені

Івана Франка

(протокол № 13 від 11.02 2025 р.)

Завідувач кафедри, *AM*

д.б.н., проф. А.М. Бабський

**Силабус з навчальної дисципліни**

«Аналіз даних в біотехнології»,

що викладається в межах ОПП Біотехнології та біоінженерія  
другого (магістерського) рівня вищої освіти для здобувачів з  
спеціальності 162 - Біотехнології та біоінженерія

<b>Назва дисципліни</b>	Аналіз даних в біотехнології
<b>Адреса викладання дисципліни</b>	вул. Грушевського 4, 79005 Львів
<b>Факультет та кафедра, за якою закріплена дисципліна</b>	Біологічний факультет, кафедра біофізики та біоінформатики
<b>Галузь знань, шифр та назва спеціальності</b>	16 Хімічна та біоінженерія, 162 - Біотехнології та біоінженерія
<b>Викладачі дисципліни</b>	Тарновська Антоніна Володимирівна, кандидат біол. наук, доцент, доцент кафедри біофізики та біоінформатики;
<b>Контактна інформація викладачів</b>	antonina.tarnovska@lnu.edu.ua <a href="https://bioweb.lnu.edu.ua/employee/tarnovska-a-v">https://bioweb.lnu.edu.ua/employee/tarnovska-a-v</a>
<b>Консультації з питань навчання по дисципліні відбуваються</b>	Згідно розкладу на кафедрі (вул. Грушевського, 4, ауд 323). Також проводяться он-лайн консультації на платформах Teams, Zoom
<b>Сторінка дисципліни</b>	
<b>Інформація про дисципліну</b>	Дисципліна «Аналіз даних в біотехнології» є дисципліною вільного вибору студентів з спеціальності G 21 - Біотехнології та біоінженерія, яка викладається в 3 семестрі в обсязі 4 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
<b>Коротка анотація дисципліни</b>	Ця дисципліна є ключовою для магістрів, що працюють із великими обсягами біологічних даних. Вона надає студентам необхідні теоретичні знання та практичні навички для ефективного збору, обробки, аналізу та інтерпретації різноманітних даних, отриманих у біотехнологічних дослідженнях. Курс охоплює статистичні методи, методи машинного навчання та візуалізації, що застосовуються до геномних, протеомних, метаболомічних та інших типів біологічних даних. Особлива увага приділяється практичному застосуванню програмних інструментів та мов програмування для вирішення реальних завдань у біотехнології.
<b>Мета та цілі дисципліни</b>	Метою вивчення дисципліни є розвиток у здобувачів вищої освіти комплексних компетенцій для самостійного та ефективного проведення аналізу великих біологічних даних, їх інтерпретації та представлення результатів з використанням сучасних обчислювальних методів та програмного забезпечення. Цілями є формування фахових компетентностей магістра біотехнологій та біоінженерії.
<b>Література для вивчення дисципліни</b>	<b>Основна:</b> 1. Forthofer, R. N., Lee, E. S., & Hernandez, M. (2006). <i>Biostatistics: A Guide to the Design, Analysis, and Presentation of Research Results</i> . John Wiley & Sons. 2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). <i>An Introduction to Statistical Learning with Applications in R</i> . Springer. (Доступна онлайн). 3. Grus, J. (2019). <i>Data Science from Scratch: First Principles with Python</i> . O'Reilly Media. 4. Selected online documentation for Python libraries (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn) or R packages (Tidyverse, ggplot2). <b>Додаткова:</b> 1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). <i>The Elements of</i>

	<p><i>Statistical Learning: Data Mining, Inference, and Prediction</i>. Springer. (Доступна онлайн).</p> <p>2. Zar, J. H. (2010). <i>Biostatistical Analysis</i>. Pearson Education.</p> <p>3. Курси та tutoriали на онлайн-платформах (Coursera, edX, DataCamp, Kaggle) з аналізу даних, машинного навчання та біоінформатики.</p> <p>4. Наукові статті та огляди з актуальних методів аналізу даних у біотехнології (наприклад, публікації в журналах <i>Bioinformatics</i>, <i>Nature Biotechnology</i>).</p> <p>Інформаційні ресурси  Навчальні платформи (Moodle, Google Classroom тощо).  Репозиторії даних (NCBI GEO, TCGA, PDB).  Відкриті освітні ресурси та відеолекції (Coursera, edX, Khan Academy – курси з баз даних та біоінформатики).</p>
<b>Тривалість курсу</b>	один семестр
<b>Обсяг курсу</b>	120 годин. З них 32 години лекцій, 16 годин практичних занять та 72 годин самостійної роботи.
<b>Очікувані результати навчання</b>	<p>Після завершення цього курсу студент буде:</p> <p><b>знати</b> основні концепції та методи статистичного аналізу даних, принципи роботи алгоритмів машинного навчання та їх застосування у біотехнології, різні типи візуалізації даних та їх ефективне використання, методи очищення та попередньої обробки біологічних даних.</p> <p><b>вміти</b> збирати, очищати та попередньо обробляти біологічні дані з різних джерел застосовувати відповідні статистичні тести для перевірки гіпотез, будувати та оцінювати моделі машинного навчання для класифікації, кластеризації та прогнозування, візуалізувати складні біологічні дані для їх інтерпретації, проводити комплексний аналіз геномних, протеомних, транскриптомних та інших біологічних даних.</p> <p><b>демонструвати</b> здатність до самостійного вирішення аналітичних завдань у біотехнології, навички критичного мислення при інтерпретації результатів аналізу, здатність до презентації складних аналітичних результатів у зрозумілій формі, готовність до безперервного навчання та освоєння нових інструментів аналізу даних.</p>
<b>Ключові слова</b>	біоінформаційні бази даних.
<b>Формат курсу</b>	Очний
	Проведення лекцій, практичних робіт та консультації для кращого розуміння тем
<b>Теми</b>	<i>Наведено у табл. 1</i>
<b>Підсумковий контроль, форма</b>	Залік у кінці семестру.
<b>Пререквізити</b>	Для вивчення курсу студенти потребують базових знань з дисципліни генетика, біоінформатика
<b>Навчальні методи та техніки, які будуть використовуватися під час викладання курсу</b>	Лекції, презентація (ілюстрація, демонстрація), розповіді, пояснення, розв'язування ситуативних задач, дискусія.
<b>Необхідне обладнання</b>	Персональний комп'ютер, загальнонавчальні комп'ютерні програми і операційні системи, проектор.
<b>Критерії оцінювання</b>	Оцінювання проводиться за 100-бальною шкалою. Бали нарахо-

<p>(окремо для кожного виду навчальної діяльності)</p>	<p>вуються за наступним співвідношенням:</p> <p>Поточний контроль:</p> <ul style="list-style-type: none"> <li>• (за результатами представлення і захисту презентацій, участі в дискусії на практичних заняттях): 40 % семестрової оцінки; максимальна кількість балів – 40; виконання презентацій та їх представлення – максимум 5 балів за 1 презентацію (4 презентації по 5 балів кожна), участь у дискусії – максимум 20 балів за семестр.</li> <li>• контрольні заміри (модулі): 60 % семестрової оцінки; максимальна кількість балів – 60.</li> </ul> <p>Модуль здобувач отримує на підставі тестування в системі Moodle (2 модулі по 30 балів кожен) – максимальна кількість балів - 60.</p> <p><b>Академічна доброчесність:</b> очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману. <b>Відвідання занять</b> є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і практичні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися усіх строків визначених для виконання усіх видів письмових робіт, передбачених курсом. <b>Література.</b> Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p><b>Політика виставлення балів.</b> Враховуються бали набрані на поточному тестуванні, самостійній роботі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття.; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін. Жодні форми порушення академічної доброчесності не толеруються.</p>
<p><b>Питання до іспиту (замірів знань)</b></p>	<ol style="list-style-type: none"> <li>1. Вступ до аналізу даних в біотехнології.</li> <li>2. Значення даних у сучасній біотехнології.</li> <li>3. Життєвий цикл даних: збір, зберігання, обробка, аналіз, інтерпретація, представлення.</li> <li>4. Типи біологічних даних (нуклеотидні, протеїнові, експресійні, метаболомні, клінічні).</li> <li>5. Попередня обробка та очищення даних.</li> <li>6. Імпорт та експорт даних.</li> <li>7. Обробка пропущених значень, викидів.</li> <li>8. Нормалізація та стандартизація даних.</li> </ol>

	<ol style="list-style-type: none"> <li>9. Злиття та трансформація даних.</li> <li>10. Описова статистика.</li> <li>11. Міри центральної тенденції (середнє, медіана, мода).</li> <li>12. Міри розсіювання (дисперсія, стандартне відхилення, квартилі).</li> <li>13. Розподіли даних (нормальний розподіл).</li> <li>14. Основи статистичного висновку.</li> <li>15. Тестування гіпотез (нульова та альтернативна гіпотези).</li> <li>16. Помилки типу I та типу II.</li> <li>17. Статистична значущість (p-значення).</li> <li>18. Параметричні та непараметричні тести (t-критерій, ANOVA, критерій <math>\chi^2</math>-квадрат, критерій Манна-Вітні).</li> <li>19. Типи машинного навчання: навчання з учителем (supervised), без учителя (unsupervised), з підкріпленням (reinforcement).</li> <li>20. Основні поняття: ознаки (features), мітки (labels), модель, навчання, передбачення.</li> <li>21. Оцінка моделей (метрики).</li> <li>22. Класифікація та регресія.</li> <li>23. Алгоритми класифікації: логістична регресія, SVM (метод опорних векторів), дерева рішень, випадковий ліс.</li> <li>24. Алгоритми регресії: лінійна регресія.</li> <li>25. Застосування в біотехнології (прогнозування активності сполук, діагностика захворювань).</li> <li>26. Кластеризація та зниження розмірності.</li> <li>27. Алгоритми кластеризації: k-means, ієрархічна кластеризація, DBSCAN.</li> <li>28. Зниження розмірності: PCA (метод головних компонент), t-SNE, UMAP.</li> <li>29. Застосування (виявлення підтипів клітин, групування генів).</li> <li>30. Візуалізація даних.</li> <li>31. Принципи ефективною візуалізації.</li> <li>32. Типи графіків (гістограми, діаграми розсіювання, box plot, heatmaps, volcano plots, PCA plots).</li> <li>33. Використання бібліотек Matplotlib, Seaborn (Python) / ggplot2 (R).</li> <li>34. Аналіз високопродуктивних біологічних даних.</li> <li>35. Аналіз даних секвенування нового покоління (RNA-Seq, scRNA-Seq, ChIP-Seq).</li> <li>36. Аналіз протеомних та метаболомних даних.</li> <li>37. Робота з оміксними даними.</li> <li>38. Інтеграція даних та системна біологія.</li> <li>39. Мульти-оміксний аналіз.</li> <li>40. Побудова мереж взаємодій (ген-ген, білок-білок). Системна інтерпретація результатів.</li> </ol>
<b>Опитування</b>	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Таблиця 1

Схема курсу «Аналіз даних в біотехнології»

Тиждень	Тема занять (перелік питань)	Форма діяльності та обсяг годин	Додаткова література / ресурс для виконання завдань (за потреби)	Термін виконання
<b>Змістовний модуль 1: Основи статистичного аналізу даних та програмування для біотехнології</b>				
1	<b>Вступ до аналізу даних в біотехнології.</b> Значення даних у сучасній біотехнології. Життєвий цикл даних: збір, зберігання, обробка, аналіз, інтерпретація, представлення. Типи біологічних даних (нуклеотидні, протеїнові, експресійні, метаболомні, клінічні).	Лекції – 2 год, практична робота – 2 год, самостійна робота – 20 год		1 тиждень
2	<b>Попередня обробка та очищення даних.</b> Імпорт та експорт даних. Обробка пропущених значень, викидів. Нормалізація та стандартизація даних. Злиття та трансформація даних.	Лекції – 4 год, самостійна робота – 10 год		2 тиждень
3	<b>Описова статистика.</b> Міри центральної тенденції (середнє, медіана, мода). Міри розсіювання (дисперсія, стандартне відхилення, квартилі). Розподіли даних (нормальний розподіл).	Лекції – 4 год, практична робота – 4 год, самостійна робота – 10 год		2 тиждень
4	<b>Основи статистичного висновку.</b> Тестування гіпотез (нульова та альтернативна гіпотези). Помилки типу I та типу II. Статистична значущість (p-значення).	Лекції – 2 год, самостійна робота – 10 год		1 тиждень
5	<b>Параметричні та непараметричні тести</b> (t-критерій, ANOVA, критерій хі-квадрат, критерій Манна-Вітні).	Лекції – 2 год, практична робота – 2 год, самостійна робота – 10 год		2 тиждень
<b>Змістовний модуль 2: Методи машинного навчання та спеціалізований аналіз біологічних даних</b>				
6	<b>Вступ до машинного навчання.</b> Типи машинного навчання: навчання з учителем (supervised), без учителя (unsupervised), з підкріпленням (reinforcement). Основні поняття: ознаки (features), мітки (labels), модель, навчання, передбачення. Оцінка моделей (метрики).	Лекції – 4 год, самостійна робота – 10 год		2 тиждень
7	<b>Класифікація та регресія.</b> Алгоритми класифікації: логістична регресія, SVM (метод опорних	Лекції – 2 год, практична робота – 2 год,		1 тиждень

	векторів), дерева рішень, випадковий ліс. Алгоритми регресії: лінійна регресія. Застосування в біотехнології (прогнозування активності сполук, діагностика захворювань).	самостійна робота – 10 год		
8	<b>Кластеризація та зниження розмірності.</b> Алгоритми кластеризації: k-means, ієрархічна кластеризація, DBSCAN. Зниження розмірності: PCA (метод головних компонент), t-SNE, UMAP. Застосування (виявлення підтипів клітин, групування генів).	Лекції – 2 год, самостійна робота – 10 год		1 тиждень
9	<b>Візуалізація даних.</b> Принципи ефективної візуалізації. Типи графіків (гістограми, діаграми розсіювання, box plot, heatmaps, volcano plots, PCA plots). Використання бібліотек Matplotlib, Seaborn (Python) / ggplot2 (R).	Лекції – 2 год, практична робота – 2 год, самостійна робота – 10 год		1 тиждень
10	<b>Аналіз високопродуктивних біологічних даних.</b> Аналіз даних секвенування нового покоління (RNA-Seq, scRNA-Seq, ChIP-Seq). Аналіз протеомних та метаболомних даних. Робота з оміксними даними.	Лекції – 4 год, практична робота – 2 год, самостійна робота – 10 год		2 тиждень
11	<b>Інтеграція даних та системна біологія.</b> Мульти-оміксний аналіз. Побудова мереж взаємодій (ген-ген, білок-білок). Системна інтерпретація результатів.	Лекції – 4 год, практична робота – 2 год, самостійна робота – 10 год		1 тиждень

Автор

Антоніна ТАРНОВСЬКА

"Погоджено"

Голова методичної ради  
біологічного факультету

Віталій ГОНЧАРЕНКО

" 10 " 2025 р.

Гарант ОПІ

Богдан Остап

" 10 " 2025 р.