

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
Факультет біології, географії та екології
Кафедра біології людини та імунології

ВИКОРИСТАННЯ БІОЛОГІЧНОГО МОДЕЛЮВАННЯ ДЛЯ
З'ЯСУВАННЯ РОЛІ ГЕНІВ У ФУНКЦІОНУВАННІ ОРГАНІЗМІВ

Кваліфікаційна робота (проект)
на здобуття ступеня вищої освіти «магістр»

Виконала: здобувачка 2 курсу 211М групи
Спеціальності 091 Біологія
Освітньої програми Біологія
Шевченко Ярослава Миколаївна

Керівник к. б. наук, доцентка, Шкуропат А. В.
Рецензент д. б. наук, професор, Чернозуб А. А.

ЗМІСТ

ВСТУП	4
РОЗДІЛ 1. Особливості та здобутки біологічного моделювання	6
1.1. Загальні особливості моделювання біологічних систем.....	6
1.2. Математична біологія та математичне моделювання.....	8
1.3. Біоінформатика.....	10
1.4. Наукова візуалізація.....	12
1.5. Моделювання у молекулярній біології, генетиці та філогенії...	13
1.6. Моделювання у біофізиці, біомеханіці.....	18
1.7. Моделювання у біомедицині.....	20
1.8. Моделювання у екології.....	21
1.9. Моделювання у мікробіології.....	26
РОЗДІЛ 2. Інструменти біологічного моделювання та аналіз можливостей їх використання у науковій роботі	28
2.1 Бази даних.....	28
2.2 Генетичні карти.....	30
2.3 Штрих-кодування ДНК.....	31
2.4 Омів-технології.....	32
2.4.1 Транскриптоміка.....	32
2.4.2 Геноміка.....	32
2.4.3 Епігеноміка.....	33
2.5 Мультимодальна інтеграція даних за допомогою програмного забезпечення MildInt.....	33
2.6 Математична теорія біологічних угруповань В. Вольтерра.....	34
2.7 Філогенетичні дерева.....	35
2.8 Байєсовська мережа.....	36
РОЗДІЛ 3. Прикладне застосування біологічних моделей	37
2.9 Створення Байєсовської мережі факторів ризику захворіти COVID-19.....	37

2.10 Аналіз можливостей використання інструментів біологічного моделювання у науковій роботі.....	39
ВИСНОВКИ.....	42
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	43

ВСТУП

Актуальність. Модель – це відображення будь-якого предмета, явища у вигляді зменшеної або збільшеної копії, схеми, аналітичного опису [9].

Для біологічних наук моделювання є важливим методом дослідження біологічних систем, оскільки дає доступ до описування, відтворення складних процесів, вивчення їх властивостей і зв'язків. У сучасному світі з'явилася можливість працювати з великими масивами даних та зібралось багато інструментів моделювання: від формул, теорій, схем до складних автоматизованих систем. Важливим є розуміння принципів їх роботи, доцільності по відношенню до галузей науки та задач дослідження, які вони ставлять перед собою, та можливостей їх використання.

Мета: дослідити особливості біологічного моделювання у різних галузях біологічних наук.

Завдання:

1. Зробити огляд особливостей та здобутків біологічного моделювання у різних галузях біологічних наук.
2. Зробити огляд різних інструментів біологічного моделювання.
3. Використати один із інструментів, Байєсівську мережу, для створення моделі факторі ризику захворіти на COVID-19.
4. Проаналізувати інструменти біологічного моделювання на можливості їх практичного використання у наукових дослідженнях.

Об'єктом дослідження кваліфікаційної роботи є біологічне моделювання. Предметом є інструменти моделювання та можливості їх використання у наукових дослідженнях.

Методи дослідження:

1. Огляд навчальної та наукової літератури;

2. Огляд інтернет-ресурсів;
3. Створення схеми у PowerPoint.

Наукова новизна: аналіз можливостей використання інструментів біологічного моделювання на основі наукових публікацій широкого часового проміжку - від минулого століття до 2020-2021 років.

По темі кваліфікаційної роботи була написана *наукова стаття* «Використання моделей та моделювання у біологічних науках» та подана до публікації в електронному альманасі «Магістерські студії» 2022-го року.

РОЗДІЛ 1. Особливості та здобутки біологічного моделювання

1.1 Загальні особливості моделювання біологічних систем

Дамо кілька визначень, що розкривають спосіб модельного розчленовування (декомпозиції) реального досліджуваного об'єкта.

Визначення 1. *Системою* будемо називати «сукупність деяких цілком певних універсальних складових одиниць – елементів, які перебувають у певних співвідношеннях і зв'язках між собою, завдяки чому вони і складають деяку певну неподільність, унітарність, цілісність. Елементи системи об'єднані спільним функціональним середовищем (а для соціальних і економічних систем – ще й спільною метою їх функціонування), у рамках якої елементи під дією системних взаємозв'язків частково втрачають свої індивідуальні властивості й здобувають спеціалізацію».

Визначення 2. *Функціональне середовище системи* – це «характерна для системи сукупність правил і параметрів (часто сформульованих у вигляді законів або алгоритмів), за якими здійснюється взаємодія (взаємообмін) між елементами системи й функціонування (розвиток) системи в цілому. Інакше кажучи, функціональне середовище системи – це сукупність зв'язків між елементами системи».

Визначення 3. *Компоненти системи* – «множина відносно однорідних елементів, які об'єднані спільними функціями при забезпеченні виконання спільних задач розвитку системи».

Визначення 4. *Елементи системи* це «умовно неподільна, самостійно функціонуюча частина системи. Підкреслимо, що виділення елементів (розбивка системи на елементи) – це операція, у певному сенсі слова, суб'єктивна. І хоча вона найчастіше повністю визначає успіх або

невдачу всього дослідження, вона надзвичайно важко піддається регламентації. Як правило, таке розчленовування системи здійснюється відповідно до якихось апріорних уявлень дослідника. І, звичайно, виділення елементів істотно залежить від постановки задачі, яка стоїть перед дослідником».

Структура системи – це «сукупність «ключових» елементів, які перебувають між собою у "сильних" зв'язках, по яким забезпечується обмін енергією, масою та інформацією між елементами системи, і які визначають як функціонування системи в цілому, так і способи її взаємодії із зовнішнім середовищем. Такі елементи, що "задають структуру", є свого роду "унікальними", виділеними, – але виділеними не по своїй індивідуальній (наприклад, для соціальних або економічних систем – особистісній) специфіці, а по своєму місцю та своїй ролі у функціонуванні всієї системи».

Визначення 5. *Границя системи* – це «сукупність зв'язаних між собою елементів, які – узяті у своїй сукупності – дозволяють зробити поділ на "внутрішнє" (наприклад, функціональне середовище системи) і "зовнішнє" середовища для розглянутої системи. Цікаво, що саме через такі "прикордонні" елементи – а, точніше, "місця", які вони займають, і відбувається весь обмін масою, енергією та інформацією між системою і її оточенням».

Визначення 6. «Система, в якій наявний хоч один елемент, який має біологічну природу, називаються *біологічними системами*».

Визначення 7. «Системи, елементи яких включають людей та технічні об'єкти різного походження, називаються *ергатициними системами*» [21].

Таким чином, щоб задати систему, ми повинні описати (задати) наступне:

а) універсальні складові одиниці – (функціональні) елементи системи;

- б) зв'язки, які існують між цими елементами;
- в) спеціально виділити структуру системи (як сукупність «специфічних місць», потрапляючи в які елементи здобувають «особливу вагу й значення», а також систему зв'язків між такими «виділеними» місцями);
- г) сукупність «прикордонних» (що належать до границі) елементів (скоріше навіть – не їх самих, а тих «місць», тих положень елементів у системі, знаходження в яких надає цим елементам здатність «відмежовувати» внутрішність системи від навколишнього середовища).

Основним методом дослідження біологічних систем є моделювання. По способу свого здійснення моделювання можна розділити на:

– *Моделювання функції системи* – з надією, що, тим самим, вдасться адекватно описати майбутню поведінку самого досліджуваного об'єкта. Наприклад, моделювання функції «активність» або «переробка інформації» для оператора в ергатичній системі.

– *Моделювання властиво самого об'єкта дослідження* – з надією, що при цьому вдасться отримати адекватний опис також і його функції. Наприклад, моделювання ухвалення рішення людиною з наступним застосуванням отриманих результатів при описі інтер'єру «управління технічною системою» (наприклад, медичним діагностичним комплексом).

Основними методами моделювання біологічних систем є аналіз і синтез. По суті, це просто класичний розподіл етапів рішення задач [11].

1.2 Математична біологія та математичне моделювання

Математична біологія (біоматематика) – це наука, яка використовує математичні розрахунки для опису біологічних систем.

Вона використовується у широкому спектрі наук про живе, оперуючи чисельним аналізом у дослідженні наукових проблем. Як результат, створюються математичні моделі біологічних процесів. Біоматематика також характерна тим, що застосовує одну й ту ж математичну концепцію до біологічних проблем різного походження. Завдяки математичним моделям є можливим вивчення різноманітних гіпотез. Встановлюється ланцюг взаємозв'язків між біологічними змінними. Потім, використовуючи експерименти, виміри та спостереження у природі, параметри отримують свої визначення у рівняннях. Змінні – це невідомі. Вони обчислюються на основі моделі, піддаються аналізу. Прості математичні моделі використовуються для інтерпретації та оцінки параметрів. Наприклад, в екології та епідемології. Дослідження у математичній біології націлені не лише на біологічні питаннями по відношенню до математиків. Також вони формують поєднання математичних понять, методів і підходів з експериментальними дослідженнями. Таким чином досягається суттєве покращення розуміння у біологічних науках. Проводиться опрацювання гіпотетичного механізму у вигляді математичної моделі. Формується певне математичне припущення. Потім пропонують експерименти, що біологічно підтверджують чи спростовують модель. Під час конкретної побудови математичної моделі вона являє собою об'єкт аналітичного та комп'ютеризованого дослідження, які далі надають значення її розв'язкам. Найбільший фокус треба мати на дослідженнях, що ґрунтуються на конкретних показниках (даних) та на уважному порівнянні передбачень й спостережень. Тоді математична теорія відіграє роль у розв'язанні практичних задач біології [8].

Найбільш дієвим засобом у застосуванні кількісної методології біологічного дослідження є математичне моделювання. Процес

проникнення математичних ідей та методів у біологію зробив його можливим.

Процедура моделювання – це імітація структур та функцій з реального світу. Математична та електронна моделі являють собою аналоги об'єкта, який вивчається. Цей аналог працює згідно з чітко визначеними математичними законами [2].

Математичні моделі полегшують розуміння складних взаємозв'язків між різними компонентами, такими як ДНК, білки, ферменти, сигнальні молекули та інші складові у біологічних систем. Як результат, це дає змогу краще зрозуміти систему в цілому та передбачити її поведінку, якщо вона буде перебувати у зміненому стані (наприклад, хвороба). Математичне моделювання дало доступ до більш глибокого розуміння багатьох складних біологічних процесів. Такими процесами є, наприклад, кінетика ферментів, метаболізм, генне регулювання, електрофізіологія [22].

Математична модель при використанні експериментальних або клінічних даних може бути використана для створення та оцінки гіпотез. Таких як постановка питання «що буде, якщо», проведення експериментів, керування майбутніми експериментами. Це дасть змогу підтвердити їх валідність. Такі моделі можуть бути використані для виявлення механізмів, які провокують зміни динамічних систем. Математична модель ніколи не замінить реальні експерименти. Тим не менш, вона взаємодіє з експериментами, що допомагає заощадити час та ресурси, визначити умови експерименту, які навряд чи є сприятливими для отримання результатів та завдяки принципам оптимізації визначити успішні експерименти [26].

1.3 Біоінформатика

З 50-х років минулого століття проводяться дослідження в області молекулярної біології. Вони націлені на отримання нуклеотидних послідовностей різних організмів, включаючи людину. Постало питання необхідності оброблювати експериментальні дані. Це й призвело до появи *біоінформатики*. Вона є однією з найбільш сучасних напрямків молекулярної біології.

З точки зору біоінформатики молекули нуклеїнових кислот та білків представляють собою текст, що складається з обмеженої кількості алфавітних знаків (4 для нуклеїнових кислот та 20 для білків). Задачею є аналіз змісту цього тексту. Можна виділити три основних напрямки біоінформатики:

- створення баз даних, які дозволяють зберігати велику кількість класифікованих біологічних даних і керувати ними;
- розробка алгоритмів і методів статистичного аналізу для визначення співвідношень між тими чи іншими даними;
- використання сучасних інструментів (наприклад, бази даних та програмні забезпечення) для аналізування різних типів біологічної інформації. Такою інформацією можуть бути послідовності ДНК, РНК та білків, білкові структури, процеси експресії генів і біохімічних реакцій.

Крім того, варто відзначити, що просторова структура макромолекул ДНК, РНК і білків є визначальною у формуванні механізмів їх взаємодії з іншими атомами та молекулами. Визначити тривимірну структуру біологічних макромолекул за допомогою електронної мікроскопії, рентгеноструктурного аналізу та методу ядерного магнітного резонансу є складним процесом. Доволі часто це взагалі неможливо. Це спрямувало розвиток методів біоінформатики, що націлені на передбачення структури молекули. Таким чином, визначення

точної просторової структури біомолекули є однією з найбільш актуальних задач біоінформатики [4].

Біоінформатика є важливою у роботі клінічних лабораторій, що генерують, аналізують дані молекулярних та генетичних тестів. А потім зберігають їх та займаються інтерпретацією[39]. Таким чином, епігенетика відіграє важливу роль у регуляції великої кількості важливих клітинних процесів. Порушення регуляції на епігенетичному рівні призводить до серйозних патологій. Наприклад, неврологічні розлади, які є причиною нейродегенерації та інтелектуальної недостатності. Зараз існує значний технологічний прогрес для процесу епігенетичного профілювання. Разом з тим, існує потреба в систематичному розумінні епігенетичного формування клітинних схем при патогенезі захворювань. Розробка більш точних обчислювальних підходів для аналізу складних епігенетичних профілів важлива для розшифровки механізмів, які лежать в основі клітинного розвитку та складних взаємодій. Вони впливають на модифікацію хроматину та метилювання ДНК, що являє собою контроль експресії генів [13].

1.4 Наукова візуалізація

Біологічні науки мають та постійно генерують величезні набори даних. Головним чином, завдяки останнім технологічним досягненням. Через їхню складність, такі набори даних потребують ефективних методів візуалізації. Тому що інакше можуть бути незрозумілими [41].

Наукова візуалізація — це наука, що займається візуалізацією наукових явищ та процесів. Вона є міждисциплінарною [20].

Технологія покращила візуальне представлення даних. Адже так можна швидко створювати складну візуальну інформацію. Також

можливо широко та просто поширювати її за допомогою цифрових засобів (порівняно з папером і чорнилом, наприклад). Для наукової літератури також візуальна інформація стала більш значима. Цифрові представлення є частим явищем у наукових публікаціях. Також зараз більше журналів стало вимагати графічні додатки [30].

Десять принципів (за Stephen R. Midway), які допоможуть покращити візуальне представлення інформації:

1. Перед створенням візуалізації варто визначити пріоритети інформації, уявити її. А потім спроектувати;
2. Використовувати підходяще програмне забезпечення;
3. Використовувати доречну геометрію та відображення даних;
4. Використання кольору є потужним інструментом візуалізації у більшій частині випадків;
5. Додавання невизначеності (наприклад, такі додаткові зображення, як панель помилок та затінена область);
6. Візуально виділити відмінності;
7. Всі візуальні зображення моделі мають бути пояснені в підписі до малюнка або мати посиланням до іншого місця в документі. Тоді читач може знайти повну інформацію стосовно того, що представляє собою візуальний образ моделі;
8. Наявність детальних підписів, які чітко пояснюють все, що є на малюнку;
9. Знати що таке інфографіка. Діаграми, наприклад, зосереджуються на представленні даних у вигляді моделей. Інфографіка ж зазвичай включає елементи діаграм;
10. Показати візуальну модель іншим людям з проханням прокоментувати її. Це допоможе дізнатися наскільки ефективно дане візуальне зображення передає потрібну інформацію [34].

1.5 Моделювання у молекулярній біології, генетиці та філогенії

У 1951 році К. Л. Полінг разом та Р. Б. Корі розробили уявлення про структуру поліпептидного ланцюга у білках. Вони першими висунули гіпотезу, що вона має спіральну будову. Таким чином, була описана альфа-спіраль [40].

1953 рік став знаковим для молекулярної біології та генетики. Він ознаменував початок нової епохи у зв'язку з відкриттям подвійної спіралі ДНК Джеймсом Уотсоном та Френсісом Кріком [45, 3].

У 1957 році Ф. Крік окреслив основи «Центральної догми» у молекулярній біології. Вона являє собою взаємне відношення ДНК, РНК і білків між собою [16]. Бичачий інсулін – це перший розшифрований білок у 1956 році. Як виявилось, він складається із 51 залишку амінокислот [44]. У наступне десятиліттям відбулася розшифровка першої нуклеїнової послідовності тРНК дріжджів, що складається із 77 основ [28].

У 1965 році Маргарет Дейхофф та співробітники Національного Фонду біомедичних досліджень у Вашингтоні систематизували усі наявні дані про амінокислотні послідовності. Результатом стало створення першої біоінформаційної бази даних – атласу білкових послідовностей та їхніх структур. Вона містила опис із 65 послідовностей. Через декілька років також були створені перші біологічні бази даних білків і нуклеотидних послідовностей. У 1982 році був створений банк даних нуклеотидних послідовностей – GenBank у США. Це база послідовностей ДНК. Також була створена EMBL у Європі. Вона являє собою молекулярно-біологічну бібліотеку [3].

Важливим у молекулярній біології є відкриттям зворотньої транскрипції ДНК [42], а також клонуванням [37]. Це відбулося у 70-х роках 20 століття. Відбулося зростанням ринкового попиту на

секвенування послідовностей та цілих геномів. У зв'язку з цим, було сконструйовано автоматичний секвенатор Л. Худ та Т. Хункапиллер у 1987 році. Оскільки обробка та аналіз такої інформації є практично неможливою традиційними методами статистики.

У кінці 80-х років та почали активно розвиватися бази даних біологічного та біоінформаційного спрямування. Це було пов'язано з великим накопиченням інформаційних даних[3].

За означенням, молекулярне моделювання біологічних систем — це опис біологічних об'єктів та явищ на рівні молекул. У цьому допомагають фізико-математичні моделі та чисельні алгоритми, котрі, як правило, вимагають використання потужних комп'ютерів. Ця мультидисциплінарна галузь науки знаходиться на перетині молекулярної і статистичної фізики, біофізики, молекулярної біології, біохімії, комп'ютерних наук. Вона відіграє важливу роль для розвитку молекулярної біології та біофізики. Також це практично використовується у фармакології та біомедичних дослідженнях.

Об'єктами дослідження у молекулярному моделюванні біологічних систем є біологічні молекули. Це можуть бути власне білки, нуклеїнові кислоти, ліпіди. Також їх надмолекулярні компоненти — біологічні мембрани або білкові комплекси. На додаток, можуть входити природні або штучні наноструктури, які містять у своєму складі біологічні молекули.

Можна зазначити, що біологічні системи на рівні молекулярних структур мають особливу специфіку. По-перше, у них поєднані ознаки твердих тіл, рідин, колоїдів, полімерів та рідких кристалів. При цьому вони не зводяться до жодного типу цих речовин. По-друге, вони представляють собою вкрай складну просторову будову та мають складну динамічну поведінку в широкому для молекул такого розміру діапазоні часів — від фемтосекунд до цілих годин. По-третє, для

функціонування таких систем характерна дуже висока чутливістю до мінімальних змін на атомарному рівні, наприклад, зміна декількох атомів може привести до мутацій та мати критичний вплив на здоров'я цілого організму. Таким чином, для дослідження біологічних систем на молекулярному рівні потрібне поєднання з декількох методів, котрі працюють у різних часових, а також просторових масштабах. Одним таким методом є молекулярне *комп'ютерне моделювання*.

Напрямок молекулярного комп'ютерного моделювання цих систем розвивається більше 30 років. Сьогодні воно є достатньо розгалуженою областю науки. З різноманітних методів, котрі використовуються цією галуззю, можна виділити три основні різновиди: структурне моделювання, динамічне моделювання і моделювання взаємодій між молекулами.

Структурне моделювання — це визначення структури молекул різних розмірів у просторі. Від малих молекул до макромолекул. Завдання цього моделювання – це розрахунок геометрії молекули, електронної структури та хімічних реакцій (хімічними методами), побудова структурних моделей для потреб кристалографії, ядерного магнітного резонансу та кріоелектронної мікроскопії, передбачення структури і гомологічне моделювання білків та нуклеїнових кислот.

Динамічне моделювання — це представлення у вигляді моделей рухів біологічних молекул. У цьому моделюванні використовують найбільш часто методи квантової динаміки, атомістичної класичної молекулярної динаміки, грубозернистої класичної молекулярної динаміки, блочної та доменної динаміки білків та континуальної динаміки супрамолекулярних комплексів.

До *моделювання міжмолекулярних взаємодій* належать методи молекулярного докінгу (жорсткий, гнучкий, по конформаційному ансамблю тощо) та віртуального скринінгу.

Стосовно молекулярної динаміки, часові та просторові масштаби, які доступні цьому методу, майже ідеально відповідають завданням з дослідження біологічних молекул. Зокрема, класична молекулярна динаміка, яка ґрунтується на класичних рівняннях Ньютона, має справу з системами, розміром до сотень нанометрів, на часових масштабах порядку 1 мкс. Таким чином, подібні системи можуть містити близько мільйона атомів. Грубозерниста молекулярна динаміка досліджує системи, в яких декілька атомів об'єднані в одну частину. Цей метод дозволяє працювати з системами розміром порядку мікрометрів на часових масштабах у сотні мікросекунд, котрі містять до 10^8 атомів. В якості порівняння: квантова хімія, що основана на рівняннях Шредінгера, на сьогоднішній день може оперувати тільки близько сотнями атомів у часовому масштабі порядку пікосекунд [6].

З того часу як почався процес картування та секвенування геному людини, нові технології значно розвинулися. Тепер ми можемо отримати велику кількість молекулярних вимірювань у тканинах та клітинах. Такі технології застосовуються для біологічних систем з метою отримання зображень з високою роздільною здатністю. Наукові галузі, що вимірюють біологічні молекули таким високопродуктивним способом, називаються «омікою» (omics). До категорій омік можна віднести такі напрямки досліджень:

- протеоміка – це глобальний аналіз білків;
- транскриптоміка – аналіз РНК;
- геноміка – аналіз генів;
- метаболоміка – аналіз метаболітів;
- ліпідоміка – аналіз ліпідів;
- епігеноміка – це аналіз метильованої ДНК або модифікованих білків гістонів у хромосомах [19].

Вперше в 1913 - 1915 роках Т. Морган з колегами показали на основі експериментів, що опираючись на явища зчеплення генів та кросинговера є можливість побудови спеціальних карт – генетичних карт хромосом [36]. Якщо гени розташовані близько один до одного в хромосомі, то є велика вірогідність що вони успадкуються разом. Завдяки закономірностям успадкування, можливо визначити взаємне розташування генів. Ще недавно вченими використовувалися методи рекомбінантної ДНК (рДНК). Вони використовувалися для встановлення фактичного фізичного розташування генів у хромосомах. Протягом багатьох років для виведення генетичних карт використовувалися різноманітні методи. Нині найбільш точним та ефективним підходом до побудови генних карт є секвенування генома (або його частини), щоб потім використати різні комп'ютерні програми для аналізу певної послідовності та ідентифікації генів у цій послідовності. Цей підхід дає можливість створення доволі точної карти кожного гену та його субкомпонентів [23].

Генетичні карти успішно використані дослідниками для пошуку гена, який відповідає за доволі рідкісні одногенні спадкові захворювання. Наприклад, муковісцидоз та м'язова дистрофія Дюшенна. Генетичні карти корисні для орієнтировки вчених на гени, котрі підвищують ризик більш поширених захворювань. Наприклад, астма, серцеві захворювання, діабет, рак і психічні розлади [21]. На основі ранніх робіт, котрі базувалися на секвенуванні ДНК у мікробних спільнотах з використанням гена 5S рРНК, були розроблені методики штрих-кодування ДНК [25].

В еволюційній біології з часів Чарльза Дарвіна використовуються філогенетичні дерева. *Філогенетичне дерево* – своєрідна діаграма, що відображає лінії еволюційного походження у різних видів, організмів та генів від спільного предка [14].

В сучасні часи з'явилися також філогенетичні сітки, які є узагальненими філогенетичними деревами. Такі сітки дозволяють створювати більш складні еволюційні історії родів. Вони включають процеси, які подібні до гібридизації [27].

1.6 Моделювання у біофізиці та біомеханіці

Моделювання є основним методом біофізики. Воно використовується на всіх рівнях вивчення живих систем, зокрема молекулярної біофізики, біофізики мембран, біофізики клітини та органів. Також використовується для біофізики складних систем.

Основні етапи моделювання у біофізиці:

- 1) збір первинної інформації;
- 2) визначення завдання;
- 3) аргументація основних допущень;
- 4) створення та дослідження моделі;
- 5) перевірка адекватності моделі по відношенню до реального об'єкта.

Таким чином відбувається узгодження моделі та реального об'єкта для дослідження.

В біомеханіці математичне моделювання використовують для опису поведінки досліджуваного матеріалу або явища.

Існують певні принципи для побудови біофізичних та біомеханічних моделей. У зв'язку з вивченням механічних властивостей біологічних тканин та систем, необхідно абстрагуватися від того, що не належить до механічних явищ. Завдяки цьому абстрагуванню буде досягнута найбільша простота моделі. Наприклад, розгляд руху рідини по трубі як чисто механічне явище. Тоді можна не брати до уваги теплопередавання, теплове випромінювання, хімічні реакції та інші

параметри. Що враховується у процесі створення таких моделей? Те, що під впливом механічних природних або штучних дій у біологічних тканинах, органах і системах формується механічний рух. У зв'язку з цим виникає напруження та його наслідки, такі як деформації та поширення хвиль. Фізіологічна відповідь, що виникає через ці чинники залежить від механічних особливостей біологічних тканин та рідин. Знання зміни реакції та властивості в тканинах і органах є важливо для профілактики та захисту організму при використанні штучних органів і тканин, а також для глибшого розуміння фізіології в нормі та патології органів і тканин.

Однією з цілей моделювання тканин організмів є дослідження можливостей протезування. Труднощі полягають в складній анізотропній структурі біологічних тканин. Ця структура є залежною від функцій призначених природою для цих тканин [5].

1.7 Моделювання у біомедицині

За останнє століття експериментальні методи в медицині зіштовхуються із рядом обмежень. У зв'язку з цим виявилось, що проведення окремих досліджень неможливе без моделювання. Це викликано такими факторами:

- втручанням в біологічні системи є ризик призвести до неможливості встановлення причин змін, котрі виникають при цьому;
- певні теоретично обґрунтовані експерименти майже неможливо здійснити через недостатній рівень розвитку експериментальної техніки;
- ряд експериментів необхідно проводити на людях. А це є жорстким відхиленням від морально-етичних та правових норм [38, 1].

У біомедичних дослідженнях є дуже корисними методи візуалізації. Вони дають можливість виявити патологічні механізми. І потім це можна використовувати для розробки та оцінки нових діагностичних і лікувальних підходів [47].

У міру того, як стають доступними великі обсяги неоднорідних біомедичних даних, було розроблено численні методи для інтеграції таких наборів даних, щоб отримати додаткові знання з кількох джерел. Останнім часом такий підхід більш глибокого аналізу показав багатообіцяючі результати в різних областях досліджень. Наприклад, такий підход був використаний для дослідження хвороби Альцгеймера. [31].

Також була підкреслена важливість мультимодальної інтеграції даних для розуміння молекулярної еволюції станів злоякісних клітин протягом життєвого циклу раку. Широке поширення мутацій-драйверів та епігенетичних змін у тканинах, що виглядають нормально, спонукає до переоцінки того, як визначається ініціація раку; у пізніх стадіях раку – вивчення ролі клонального відбору, епігеномної адаптації, резистентності до терапії ін. Покращення оцінки ризику раку та прискорення терапевтичних відкриттів вимагають надійних, всебічних та інтегрованих наборів даних з використанням сучасних комп'ютерних технологій, таких як створення 3D моделей пухлин протягом усього життєвого циклу раку [43].

Інтегрована мультимодальна візуалізація серця та легенів відіграє центральну роль у різних клінічних дослідженнях і є важливою для діагностики, визначення ризику та лікування пацієнтів із COVID-19 [12].

Також використання математичних методів та автоматизованих комп'ютерних систем у медицині та фармації суттєво розширило здійснення розробок методик контролю якості лікарських засобів,

фармацевтичних субстанцій, лікарської рослинної сировини і допоміжних речовин із використанням фізичних, фізико-хімічних та хімічних методів контролю [1].

1.8 Моделювання у екології

Кінцева мета вивчення штучних і природних екосистем – це найбільш раціональне їх використання для потреб людини, яке передбачає оптимальне управління екосистемами. Вирішення завдання оптимального управління неможливе без побудови математичної моделі об'єкта управління, оскільки метод "проб і помилок" стосовно природних екосистем не придатний через їх унікальність і неприпустимість викликати в них необоротні зміни. Якщо модель досить точно імітує дійсність, вона являє собою необмежені можливості для експериментування, в неї можна вводити нові чинники і збурення, для того, щоб з'ясувати їх вплив на систему.

До необхідності побудови математичних моделей популяцій, співтовариств і екосистем можна підійти не лише з позицій оптимального управління, але й виходячи з вимог кількісного опису зв'язків і функціонування цих складних систем далеко не у всіх деталях, доступних безпосередньому спостереженню.

Процеси енерго- і масообміну в біогеоценозах відбуваються за участі різних рослинних і тваринних організмів під впливом чинників навколишнього середовища, елементи якого безпосередньо залучаються до перетворень біомаси ланок екологічних систем. Кожна популяція, оскільки вона існує на обмеженій території і використовує для свого існування обмежену кількість речовини, входить у деяке співтовариство

популяцій тварин і рослин, створюючих певну структуру харчових (трофічних) зв'язків і метаболізму, що оточують її. Разом з використовуваною неживою речовиною таке співтовариство і становить біогеоценоз. При дослідженні складних біогеоценозів (екологічних систем) плідними виявилися методи загальносистемного підходу. Один з них – виділення з екологічної системи взаємодіючих складових структурних елементів, таких, як види, що належать до різних трофічних рівнів або вікових і статевих внутрішньовидових груп. Інший важливий елемент системного підходу – встановлення характеру процесів, в яких бере участь кожен елемент (процеси розмноження і росту, взаємодії між елементами типу конкуренції, хижацтва і т. д.). До екологічних систем також може застосовуватися принцип ізоморфізму, що дозволяє описувати схожими математичними рівняннями системи, різні за своєю природою, але однакові за структурою і типом взаємодії між їх складовими елементами. У даному випадку є велика схожість систем рівнянь хімічної кінетики і міжпопуляційної динаміки.

Складність екосистем посилюється мінливістю самих живих організмів, яка може виявлятися і при взаємодії організмів один з одним, і в реакції організмів на умови навколишнього середовища. Ця реакція може виражатися у зміні швидкості росту, відтворення і в різній здатності до виживання в умовах, що сильно відрізняються. До цього додаються зміни, що відбуваються незалежно від таких факторів середовища, як клімат і характер місцепроживання. Тому дослідження і регулювання

екологічних процесів і екологічних систем є винятково важливим завданням. Важливим чинником, що ускладнює дослідження, є тривалість екологічних процесів. Наприклад, дослідження у галузі землеробства і садівництва пов'язані з визначенням врожайності, а урожай збирають раз на рік, через це один цикл експерименту займає 1

рік і більше. Щоб знайти оптимальну кількість добрив і провести інші можливі заходи щодо окультурення, може знадобитися декілька років, особливо коли необхідно розглядати взаємозв'язок між експериментальними результатами і погодою. У лісоводстві залежно від тривалості колообігу урожаїв деревини найтриваліший експеримент займає 25 років, а довготривалі експерименти можуть тривати від 40 до 120 років. Аналогічні масштаби часу необхідні і для проведення досліджень за керуванням природними ресурсами. Для опису динаміки чисельності видів у складних екосистемах і вирішення питань оптимального керування природними і штучними співтовариствами потрібно застосування комп'ютерів, особливо коли мова йде про кількісну відповідність моделі і реальності. Проте в основі суворого математичного опису екосистем лежать якісь елементарні моделі, що описують "елементарні" взаємодії в екосистемі і піддаються якісному аналізу. Методи якісного дослідження дозволяють визначити такі важливі характеристики моделей екосистем, як їх стійкість і наявність коливань чисельності видів.

Із проблемою стійкості екосистем доводиться все частіше стикатися при вирішенні питань експлуатації природних популяцій і угруповань, оцінці меж забруднення середовища, урахування наслідків, здійснення тих або інших природогосподарських заходів. При цьому кількісні оцінки стійкості або "стабільності" екосистем можуть бути отримані лише шляхом побудови математичних моделей і формалізації самого терміна "стійкість". Дійсно, інтуїтивно зрозуміло, що біогеоценоз, екосистема, біологічне суспільство, що існують у більш-менш незмінному вигляді достатній час, мають деяку внутрішню здатність протистояти збурювальним чинникам, що впливають на неї з боку зовнішнього

середовища (серед них і антропогенний чинник). При цьому стійкою, як правило, вважається система, незмінна на різних рівнях. Передбачається незмінним у часі географічний регіон або ландшафт, що включає систему, що вивчається і що визначає умови проживання рослин і тварин. Основними процесами, що визначають динаміку регіону, будуть не зміни чисельності окремих видів, що населяють його, а глобальні біогеохімічні цикли. У стійкій системі передбачається збереження числа видів у даному біологічному співтоваристві. Вважають також, що співтовариство стабільно стійке, якщо чисельності складових її популяцій не піддаються різким відхиленням від деяких середніх значень. Це визначення близьке до термодинамічного поняття стійкості такої системи, в якій мала ймовірність великих флуктуації, здатних відвести систему від рівноважного стану або навіть зруйнувати її.

Як правило, при математичному моделюванні біогеоценозу завдання полягає в тому, щоб отримати обґрунтований прогноз кінетики ценозу, знаючи його склад і співвідношення його компонентів. При цьому будуть зроблені різні початкові припущення і переслідуються відповідні цілі при вивченні моделей, які один з піонерів математичної екології А. А. Ляпунов (1972) сформував так:

А. Біологічні характеристики компонентів вважаються незмінними, так само, як і взаємодії між ними. Ценоз вважається однорідним у просторі. Вивчаються зміни в часі кількості (біомаси) компонентів ценозу.

Б. При збереженні гіпотези однорідності ценозу вводиться припущення про закономірну зміну системи відносин між компонентами. Це може відповідати або закономірній зміні зовнішніх умов (наприклад, сезонній), або заданому характеру еволюції форм, що створюють цей ценоз. При цьому, як і раніше, вивчається кінетика кількості компонентів.

Апаратом для вивчення цих двох класів завдань служать системи звичайних диференціальних і диференціально-різнісних рівнянь з постійними (А) і змінними (Б) коефіцієнтами.

В. Об'єкти, що становлять ценоз, вважаються за різнорідні за своїми властивостями і здатні до дії відбору. Передбачається, що еволюція форм визначається умовами існування ценозу. У цих умовах вивчається, з одного боку, кінетика кількості компонентів, з іншого – дрейф характеристик популяцій, що створюють цей ценоз. При вирішенні таких завдань істотно використовується апарат теорії ймовірності. До цієї групи належать багато завдань генетики популяції.

Г. Нарешті, можлива відмова від територіальної однорідності ценозу та урахування залежності усереднених концентрацій від координат. Тут виникають питання, пов'язані із просторовим перерозподілом складових ценозу, як живих, так і неживих. Наприклад, чисельність (біомаса) видів може змінюватися зі зміною глибини водоймища. Для опису таких систем необхідно залучення апарату диференціальних рівнянь у часткових похідних.

На початку ХХ ст. з'явилися перші моделі взаємодії видів, наприклад, модель А. Лотки (1925), наведена ним у книзі "Елементи фізичної біології". Проте засновником сучасної математичної екології насправді вважається італійський математик Віте Вольтерра, який розробив математичну теорію біологічних угруповань, апаратом якої були системи диференціальних і інтегродиференціальних рівнянь [10].

1.9 Моделювання у мікробіології

Прогностична біологія є наступним великим розділом у синтетичній та системній біології, особливо для мікроорганізмів.

Завдання, які колись здавалися нездійсненними, все частіше реалізуються, наприклад, розробка та реалізація складних синтетичних генних ланцюгів, які виконують складні функції сприйняття та активації, а також збирання багатовидових бактеріальних спільнот зі специфічним, заздалегідь визначеним складом. Ці досягнення стали можливими завдяки інтеграції різноманітних знань у біології, фізиці та інженерії, що призвело до появи

кількісного розуміння біологічного дизайну. Оскільки набори даних постійно розширюються, стають доступними, їхня потенційна корисність у перетворенні теорії на практику залишається міцно вкоріненою в основних кількісних принципах, які керують біологічними системами [32].

Загалом, моделі, розроблені в прогностичній мікробіології, спрямовані на кількісну оцінку впливу внутрішніх, зовнішніх факторів обробки на результативну мікробну проліферацію в харчових продуктах або харчових модельних системах, наприклад, буферній системі. Ці моделі покладаються на можливість інтерполяції результативної мікробної проліферації для комбінацій, які не тільки спочатку досліджуються, але також включені в діапазон планування експерименту. Таким чином, прогностичну мікробіологію можна розглядати як потужний інструмент для дослідження та стислого узагальнення впливу різноманітних умов (у межах рецептури та обробки їжі) на мікробну екологію.

Історична перспектива моделювання розробок у галузі прогностичної мікробіології була представлена Мафартом у 2005 році. Відповідно до цього, перші розробки відносяться до 1920-х років, коли термостійкість мікроорганізмів описувалась або рівнянням Арреніуса, або моделлю Бігелоу. Проте принципи та цілі дисципліни з'явилися значно пізніше, на початку 1990-х років, після чого послідували

розробка та опис мікробних моделей та генерація відповідних баз даних та інших програмних засобів [33].

Ще одним цікавим прикладом застосування моделювання є прикладна мікробіологія. Наприклад, у таких технологічних процесах, як очищення стічних вод, виробництво безпечних харчових продуктів, концентрування певних речовин, що містяться у корисних копалинах за допомогою мікробів, можуть застосовуватися різні методи моделювання: побудова первинних, вторинних та третинних математичних моделей, механістичних або кінетичних моделей, феноменологічних моделей, моделей реактивного транспорту, моделей байєсовської мережі, тощо [18].

РОЗДІЛ 2

Інструменти біологічного моделювання та аналіз можливостей їх використання у науковій роботі

2.1 Бази даних

Кожна розшифрована нуклеотидна або амінокислотна послідовність сама по собі представляє значний інтерес для генної інженерії та біотехнології, але разом з цим, послідовність може слугувати колосальним джерелом інформації при порівнянні її з іншими послідовностями. Це стимулювало розробку добре структурованих баз даних (БД), які могли б працювати з великими об'ємами експериментальних даних, і спеціальних програмних засобів, що допомогли б інтерпретувати результати експериментів. Відомості про відкриття нових послідовностей розміщувалися у провідних журналах, а

потім ці дані заносилися до БД вручну. Однак, коли почалося лавиноподібне зростання обсягів інформації, такий процес став неможливим. Журнали почали вимагати, щоб послідовності розміщувалися у БД самими авторами. Сьогодні, коли секвенування ДНК є достатньо поширеним процесом, який можуть виконувати роботи або студенти на лабораторних роботах, багато послідовностей можуть потрапляти до БД без опублікування у наукових журналах. На даний час існують такі БД, де інформацію може розмістити кожен науковець, і такі, де інформація суворо перевіряється, а відповідальність за її достовірність покладається на власника бази даних [3].

В молекулярній біології та біоінформатиці найбільш популярним та розповсюдженим є FASTA формат. Він виглядає наступним чином: кожен запис в ньому складається з двох рядків. Перший рядок, який починається зі знака >, за ним може знаходитись інформація про базу даних в якій зберігається послідовність. Далі може знаходитись ідентифікатор або код доступу до послідовності. Ідентифікатор зазвичай містить три перших літери родової та дві літери видової таксономічної приналежності організму, для якої була отримана послідовність. Після службової інформації в наступному рядку безпосередньо записана сама послідовність. FASTA формат дозволяє зберігати інформацію для нуклеотидних та амінокислотних послідовностей. Наприклад, послідовність мікротубулярного протеїну людини у FASTA форматі має наступний вигляд:

```
>gi|31563518|ref|NP_852610.1| microtubule-associated proteins 1A/1B light
chain 3A isoform b [Homo sapiens]
MKMRFFSSPCGKAAVDPADRCKEVQQIRDQHPSKIPVPIERYKGEKQL
PVLDKTKFLVPDHVNMSE
LVKIIRRLQLNPTQAFLLVNQHSMVSVSTPIADIYEQEKDEDGFLY
MVYASQETFGFIRENE [7]
```

Однією з найбільш відомих міжнародних організацій серед тих, які створюють інструменти для аналізу інформації та здійснюють контроль за наповненням баз даних біоінформаційного напрямку, є Міжнародна система баз даних нуклеотидних послідовностей («International Nucleotide Sequence Database Collaboration (INSDC)»), яка об'єднує 3 установи, нуклеотидні бази даних, що мають різний набір сервісів:

NCBI – National Center for Biotechnology Information, USA
<http://www.ncbi.nlm.nih.gov/>,

NIG – National Institute of Genetics, Japan <http://www.ddbj.nig.ac.jp/>,

EBI – European Bioinformatics Institute, United Kingdom
<http://www.ebi.ac.uk/>.

До розповсюджених організацій також відносяться «RCSB – Research Collaboratory for Structural Bioinformatics», USA, <http://www.rcsb.org/>, та SIB – Swiss Institute of Bioinformatics, Switzerland <http://www.isb-sib.ch/>.

база даних білкових послідовностей та 3d структур:

– Swissprot (Department of Medical Biochemistry of the University, Switzerland, Geneva, <http://www.expasy.ch/sprot/sprot-top.html>)

база даних нуклеотидних послідовностей:

DDBJ (Center for Information Biology (cib), DnaData Bank of Japan, <http://www.ddbj.nig.ac.jp/>)

EMBL (European Molecular Biology Laboratory, Geyzelberg, Germany, <http://www.ebi.ac.uk/embl/>)

GenBank (National Centre for Biotechnology Information, Los-Alamos, USA, <http://www.ncbi.nlm.nih.gov/>)

Також існують більш спеціалізовані біоінформаційні ресурси:

– PMB (Protein Mutant Database, Japan, http://www.genome.ad.jp/htbin/www_bfind?pmd) – БД призначена для досліджень в області білкової інженерії.

- OMIM (John Hopkins University, USA, <http://www3.ncbi.nlm.nih.gov/omim>) – БД по генах людини та пов'язаними з ними хворобами.
- Genes Kegg (Kioto, Japan, <http://www.genome.ad.jp/kegg/>) – БД по систематичному аналізу функцій генів Складається із шести баз даних – метаболічних шляхів (PATHWAY), генів (GENES), лігандів (LIGAND), дослідних даних по експресії генів (EXPRESSION и BRITE), білків (SSDB) [4].

2.2 Генетичні карти

Генетичні карти застосовуються для встановлення розташування генів у хромосомах. Ранні генні карти використовували аналіз зчеплення [23].

Щоб створити генетичну карту, дослідники збирають зразки крові або тканин у членів сімей, в яких переважає певна хвороба або ознака. Використовуючи різні лабораторні методи, вчені виділяють ДНК з цих зразків і досліджують її на наявність унікальних моделей, які можна побачити лише у членів сім'ї, які мають хворобу або ознаку. Ці характерні закономірності в ДНК називаються маркерами. ДНК-маркери самі по собі не ідентифікують потрібний ген, але вони можуть приблизно показати де знаходиться ген у хромосомі [21].

2.3 Штрих-кодування ДНК

Штрих-код ДНК — це метод ідентифікації певного виду за допомогою короткого фрагмента ДНК з його геному [17]. Маркери, або специфічні короткі послідовності ДНК, зі стандартизованої області

геному можуть створити штрих-код ДНК для ідентифікації видів [25]. Базується штрих-кодування ДНК на тому, що порівнюється окрема послідовність ДНК з довідковою бібліотекою таких ділянок ДНК-послідовностей, і може використовуватися для ідентифікації організму як певного виду, так само, як сканер супермаркету використовує чорні смуги УРС штрих-коду для ідентифікації товару на його складі за його довідковою базою даних [17].

Для того, щоб вдало визначати види за допомогою штрих-кодів ДНК, відбір необхідної інформативних ділянок ДНК має вирішальне значення. Хороший штрих-код ДНК повинен мати низьку внутрішньовидову та високу міжвидову мінливість [46]. Ідентифікація видів, перерахованих у додатках до Конвенції про міжнародну торгівлю видами, що перебувають під загрозою зникнення (CITES), за допомогою штрих-кодування використовується для моніторингу незаконної торгівлі [29].

2.4 Омів-технології

2.4.1 Транскриптоміка

Транскриптом – це повний набір транскриптів РНК з ДНК у клітині або тканині. Транскриптом включає рибосомну РНК(рРНК), інформаційну РНК(мРНК), транспортну РНК (тРНК), мікроРНК (міРНК) та некодуючу РНК (нкРНК). У людини лише від 1,5 до 2 відсотків геному представлено в транскриптомі як гени, що кодують білок. Двома домінуючими технологіями вимірювання транскриптомів є мікрочипи та секвенування РНК (RNAseq). Мікрочипи засновані на олігонуклеотидних зондах, які гібридизуються зі специфічними РНК-

транскриптами. RNAseq — це набагато більш сучасний підхід, який дозволяє здійснювати пряме секвенування РНК без необхідності використання зондів.

2.4.2 Геноміка

Геном — це повна послідовність ДНК у клітині чи організмі. Цей генетичний матеріал можна знайти в ядрі клітини або в інших органелах, таких як мітохондрії. За винятком мутацій і хромосомних перебудов, геном організму залишається практично незмінним з часом. Повну або часткову послідовність ДНК можна аналізувати за допомогою різних експериментальних платформ, включаючи чіпи одонуклеотидного поліморфізму (SNP) і технологію секвенування ДНК. Чіпи SNP – це масиви тисяч олігонуклеотидних зондів, які гібридизуються (або зв’язуються) зі специфічними послідовностями ДНК, у яких, як відомо, зустрічаються варіанти нуклеотидів. За допомогою чіпів SNP можна аналізувати лише відомі варіанти послідовності, а на практиці таким чином аналізують лише загальні варіанти. Геномний аналіз також може виявити вставки та делеції та варіацію кількості копій, що стосується втрати або ампліфікації двох копій кожного гена (один від матері та один від батька в кожному генному локусі).

2.4.3 Епігеноміка

Епігеном складається з оборотних хімічних модифікацій ДНК або гістонів, які зв’язують ДНК, і викликають зміни в експресії генів, не змінюючи послідовність азотистих основ. Епігеномні модифікації можуть відбуватися специфічно для тканин, у відповідь на фактори

навколишнього середовища або в результаті розвитку хворобливих станів і можуть зберігатися протягом поколінь. Епігеном може істотно відрізнятися між різними типами клітин в межах одного організму. Біохімічно епігенетичні зміни, які вимірюються при високій пропускній здатності, належать до двох категорій: метилювання залишків цитозину ДНК і різноманітні види модифікацій специфічних гістонових білків у хромосомах (гістонові позначки). Редагування РНК є ще одним механізмом епігенетичних змін у експресії генів, що вимірюється в основному за допомогою транскриптомних методів [19].

2.5 Мультиmodalна інтеграція даних за допомогою програмного забезпечення MildInt

Мультиmodalна інтеграція даних вимагає досвіду для побудови глибокого масиву мультиmodalних даних. Для цієї мети був розроблений спеціальний пакет Python (MildInt), який у подальшому, на основі масиву даних, здатен задавати класифікації.

Для отримання реальних даних в якості пілотного дослідження використовували набори клінічних і мультиmodalних нейровізуалізаційних даних при хворобі Альцгеймера для прогнозування прогресування захворювання. MildInt скомбінував різні форми числових даних, включаючи часові ряди, та дані не часових рядів для вилучення додаткових функцій із мультиmodalного набору даних [31].

2.6 Математична теорія біологічних угруповань В. Вольтерра

Системи, які вивчалися Вольтерра, складаються із декількох видів. В окремих випадках розглядається також запас їжі, який використовують деякі з видів. Про компоненти системи формулюються такі припущення:

1. Їжа або ϵ в необмеженій кількості, або її кількість у часі жорстко регламентована;
2. Особини кожного виду вмирають так, що за одиницю часу гине стала частина існуючих особин;
3. Хижі види поїдають жертв, причому за одиницю часу кількість жертв, яких з'їдають, завжди пропорційна імовірності зустрічі особин цих двох видів, тобто добутку кількості хижаків на кількість жертв;
4. Якщо ϵ їжа в обмеженій кількості і декілька видів, які здатні її споживати, частка їжі, споживана кожним видом за одиницю часу, пропорційна кількості особин цього виду, взятого з деяким коефіцієнтом, що залежить від виду (моделі міжвидової конкуренції);
5. Якщо вид харчується їжею, наявною в необмеженій кількості, приріст чисельності виду за одиницю часу пропорційний чисельності виду;
6. Якщо вид харчується їжею, наявною в обмеженій кількості, його розмноження регулюється швидкістю споживання їжі, тобто за одиницю часу пропорційний чисельності виду;
6. Якщо вид харчується їжею, наявною в обмеженій кількості, його розмноження регулюється швидкістю споживання їжі, тобто за одиницю часу приріст пропорційний кількості з'їденої їжі.

Перелічені гіпотези дозволяють описувати складні біоценози за допомогою систем звичайних диференціальних рівнянь, у правих частинах яких є суми лінійних і білінійних членів. Такими рівняннями описуються і системи хімічних реакцій [10].

2.7 Філогенетичні дерева

Філогенетичні дерева є фрактальними - можна знайти однакову картину, незалежно від того, чи розглядаємо ми лінії, які нещодавно розійшлися, або глибокі розколи в дереві життя. Як результат, структура дерева надзвичайно корисна для відстеження біологічного різноманіття на всіх рівнях.

Більшість філогенетичних дерев мають коріння, що означає, що одна гілка (яка зазвичай не позначена) відповідає спільному предку всіх видів, включених до дерева.

Мітки на «кінчиках» філогенезу можуть відповідати окремим організмам, видам або наборам видів, якщо кожна верхівка становить окрему гілку на дереві життя. У певних контекстах такі кінці можуть відповідати окремим генам. Точки розгалуження всередині дерева, які відповідають виведеним подіям видоутворення, називаються вузлами. Кожен вузол представляє останнього спільного предка двох ліній, що походять від цього вузла. Внутрішні гілки або міжвузля з'єднують два вузли, тоді як зовнішні гілки з'єднують кінчик і вузол.

Клада - це частина філогенезу, яка включає родовід предків і всіх нащадків цього предка [14].

2.8 Байєсовська мережа

Моделі байєсовської мережі є інструментом для ймовірнісного та причинно-наслідкового моделювання. Байєсовські мережі визначаються як спрямовані ациклічні графи, де причинно-наслідкові зв'язки та пов'язана з ними невизначеність кількісно визначені в таблицях умовної ймовірності. Байєсовські мережі за своєю суттю включають невизначеність і можуть інтегрувати різноманітні типи інформації, включаючи експертне виявлення [35].

РОЗДІЛ 3

Прикладне застосування біологічних моделей

3.1 Створення Байєсовської мережі факторів ризику захворіти COVID-19

Була взята інформація із сайту www.cdc.gov (Centers for Disease Control and Prevention) щодо факторів ризику захворіти на COVID-19.

Фактори ризику захворіти на COVID-19

- 1) Рак.
- 2) Послаблення організму через лікування раку.
- 3) Хронічні захворювання нирок.
- 4) Хронічні захворювання печінки.
- 5) Хронічні захворювання легень.
- 6) Кістозний фіброз.
- 7) Деменція.
- 8) Хвороба Альцгеймера.
- 9) Діабет (1-го та 2-го типу).
- 10) Інвалідність.
- 11) Серцеві захворювання на COVID-19.
- 12) Наявність ВІЛ (вірусу імунодефіциту людини).
- 13) Ослаблена імунна система.
- 14) Умови психічного здоров'я.
- 15) Надмірна вага.
- 16) Ожиріння.
- 17) Вагітність.
- 18) Серповидно-клітинна анемія або таласемія.

- 19) Куріння, теперішнє чи колишнє.
- 20) Трансплантація твердих органів або стовбурових клітин крові.
- 21) Інсульт або цереброваскулярні захворювання.
- 22) Розлади вживання психоактивних речовин.
- 23) Наявність розладу вживання психоактивних речовин.
- 24) Туберкульоз [15].

Створені окремо графи факторів та графа ризику захворіти (Рис. 2.1). Всі фактори направлені стрілками до ризику. Деякі графи факторів пов'язані між собою взаємними стрілками, що означає їх взаємозв'язок. Наприклад, рак та ослаблення імунної системи через його лікування, зайва вага та ожиріння.

Створена Байєсовська мережа є доволі простою, тому що включає тільки загальні дані, не має процентного співвідношення (не було вказано на сайті). Але навіть без процентного співвідношення зрозуміло, що якщо людина має більше ніж один фактор ризику (наприклад, ожиріння та хронічне захворювання органів дихання), то ймовірність захворіти на COVID-19 зростає. І чим більше цих факторів, тим більшим є ризик та можливість подальших ускладнень, або навіть летального випадку.

Виходячи з вищезазначеного, Байєсовську мережу можна використовувати для аналізу та візуалізації загальних даних, але для більш повного дослідження потрібен більш ретельний опис рівнів ризику факторів з приблизним процентним співвідношенням.

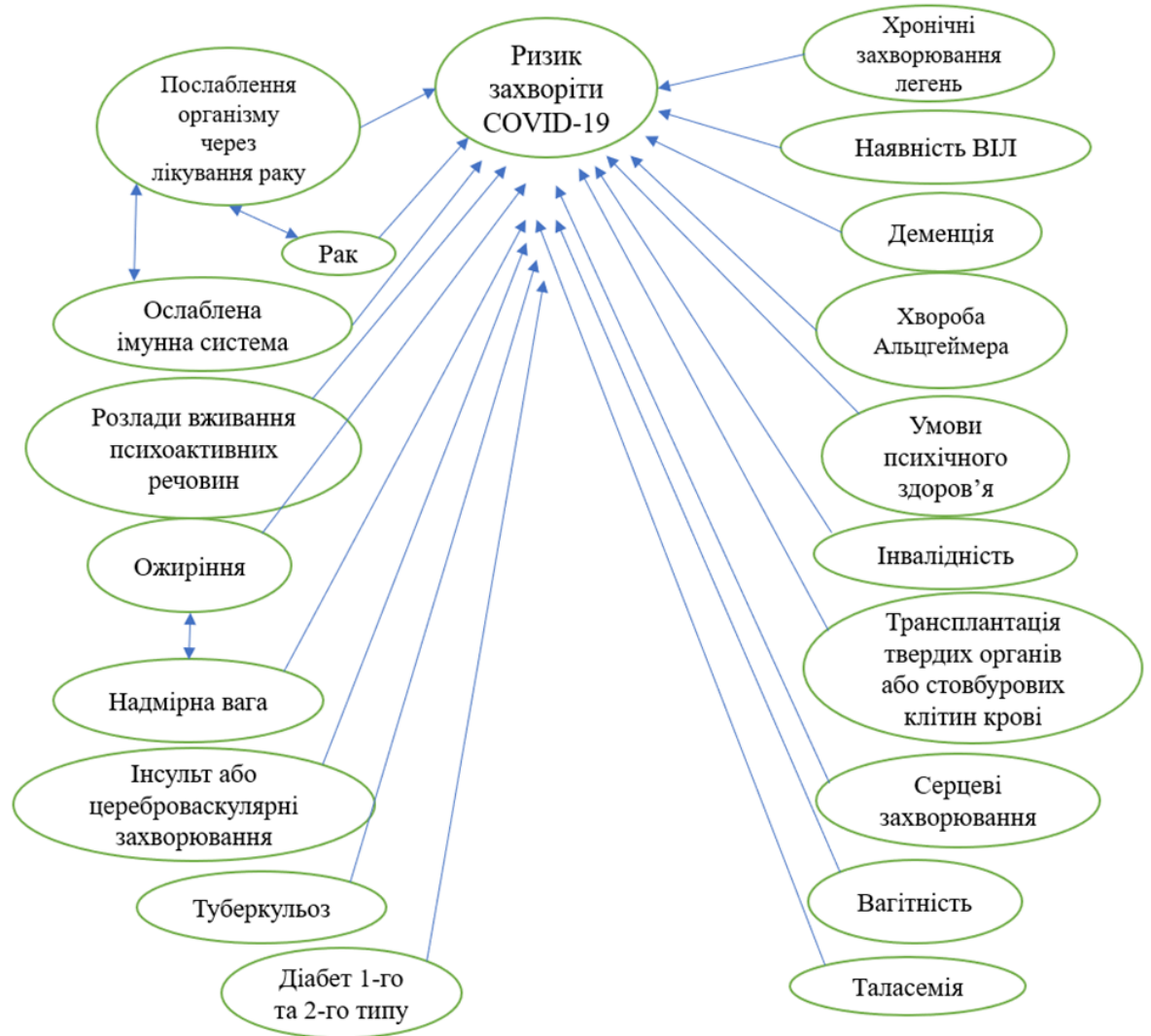


Рис. 3.1 Байєсовська мережа факторів ризику захворіти COVID-19

3.2 Аналіз можливостей використання інструментів біологічного моделювання у науковій роботі

Бази даних є загальнодоступними через мережу Інтернет. За їхньою допомогою можливо знайти інформацію про молекулярні дані біохімічних речовин, таких як, наприклад, нуклеотидні послідовності, білки, їхні 3D зображення. Існують загальні та спеціалізовані бази даних. Наприклад, про генні хвороби. Для практичної роботи з базами даних дослідник має оволодіти практичними методиками роботи.

Такими є, наприклад, попарне вирівнювання та множинне вирівнювання нуклеотидних послідовностей. Використовувати бази даних можливо для наукової роботи на різних рівнях складності - від навчальних практик до складних молекулярних досліджень. Більш складна наукова робота потребує більш поглибленого знання методик біоінформатики.

Для створення *генетичних карт* потрібні спеціальні лабораторні умови та матеріали, чітка задача (наприклад, визначення локуса гена відповідального за певну хворобу). Цей інструмент моделювання є більш доступним у рамках складних спеціалізованих генетичних досліджень.

Для *штрих-кодування ДНК* теж потрібні лабораторні умови та матеріали. Цей інструмент є більш спеціалізованим для вивчення видової різноманітності та охорони природи.

Омік-технології є одними із найновіших інструментів. Вони потребують лабораторних умов та сучасних високотехнологічних матеріалів, таких як мікрочипи, оптичні апарати та ін.

Мультимодальна інтеграція даних за допомогою програмного забезпечення MildInt новий інструмент, який тільки починають ґрунтовно використовувати. Він є складним, оскільки потрібно добре володіти комп'ютерними технологіями для обробітку великого масиву даних. Використовується, в основному, для прогнозування перебігу хвороб та інших патологічних станів організму.

Математична теорія біологічних угруповань В. Вольтерра існує для системних досліджень у галузі екології видів. Можна використовувати на різних рівнях складності досліджень, за допомогою комп'ютерних технологій або без них.

Філогенетичні дерева можна використовувати за допомогою комп'ютерних технологій або без них, для різних рівнів складності досліджень. Більш складні та об'ємні філогенетичні дерева зручніше

створювати за допомогою комп'ютерних технологій. Можна також використовувати у якості навчальних практичних робіт.

Байєсівську мережу можливо створювати для будь-якої теми дослідження, для будь-якого рівня складності досліджень, за допомогою комп'ютерних технологій та без них. Для більш складного та глибокого дослідження важливо мати відсоткові співвідношення вірогідностей - це дає змогу використовувати цей інструмент для прогнозування біологічних явищ.

Таким чином, серед інструментів біологічного моделювання можливо знайти інструменти біологічного моделювання для будь-якої складності наукових досліджень та для будь-якої галузі біологічних наук.

ВИСНОВКИ

1. Біологічне моделювання досягло значного прогресу протягом XIX-XXI століть. Значним поштовхом до його розвитку стала молекулярна біологія та генетика. У створенні біологічних моделей задіяні такі міждисциплінарні галузі, як біологічна математика, наукова візуалізація, біоінформатика. Можливості біологічного моделювання значно розширились в останні роки завдяки технологічному прогресу.

2. Існує значна кількість інструментів біологічного моделювання у різних галузях біологічних наук та для різних рівнів складності досліджень.

3. Байєсовська мережа дала змогу створити загальну графічну модель ризиків, які можуть призвести до захворювання на COVID-19. Але для більш повного аналізу рівнів ризиків потрібне процентне співвідношення.

4. Існує достатня кількість інструментів моделювання, які можна використовувати для різних галузей біологічних знань та різної складності досліджень. Для більш складних та поглиблених досліджень необхідно добре володіти комп'ютерними технологіями та методиками біоінформатики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Болух В. А. Сучасні математичні моделі в термодинаміці біологічних процесів. Медична Освіта. Житомирський базовий медичний коледж. 2018. № 4. с. 127-131.
2. Гвоздяк В.М. Особливості математичного моделювання органічних систем. Мультиверсум. Філософський альманах. К.: Центр духовної культури, 2006. № 56.
3. Горобець С. В., Горобець О. Ю., Дем'яненко І. В. Біоінформатика. Практикум: навч. посіб. для студ. спеціальності 162 «Біотехнологія та біоінженерії». К.: КПІ ім. Ігоря Сікорського, 2020. 87 с.
4. Горобець С. В., Горобець О. Ю., Хоменко Т. А. Основи біоінформатики: підручник для студентів напряму підготовки 6.051401 «Промислова біотехнологія» факультету біотехнології і біотехніки. К.: НТУУ «КПІ», 2010. 506 с.
5. Дослідження в біофізиці та біомеханіці: підручник/ В. С. Антонюк та ін. Київ: НТУУ «КПІ», 2012. 346 с.
6. Єсилевський С. О. Молекулярне моделювання біологічних систем: стан та перспективи. Стенограма наукової доповіді на засіданні Президії НАН України. Вісн. НАН України, 2018. № 6. с. 43-49.
7. Кисляк С. В., Настенко Є.А. Основи молекулярної біології та біоінформатики: комп'ютерний практикум: навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології спеціалізації «Інформаційні технології в біології та медицині». К.: КПІ ім. Ігоря Сікорського, 2018. 95 с.

8. Рябуха Т. В. Математична біологія. Енциклопедія сучасної України. URL: https://esu.com.ua/search_articles.php?id=66927 (дата звернення: 20.11.2022).
9. Савустьяненко Т. Л., Савустьяненко А. В. Інновації на уроках біології. Харків: ВГ «Основа», 2007. 190 с.
10. Теорія систем в екології: підручник / Ю. Г. Масікевич та ін. Суми : СумДУ, 2015. 330 с.
11. Шиян А. А. Основи моделювання біологічних та ергатичних систем: навчальний посібник. Вінниця: ВНТУ, 2008. 131 с.
12. Agricola E. Heart and Lung Multimodality Imaging in COVID-19/ E. Agricola, A. Beneduce, A. Esposito et al.// JACC Cardiovasc Imaging, 2020. Vol. 13. N. 8. P. 1792 – 1808.
13. Angarica V. E. Bioinformatics Tools for Genome-Wide Epigenetic Research/ V. E. Angarica, A. D. Sol// Advances in Experimental Medicine and Biology, 2017. Vol. 978. – P. 489–512.
14. Baum D. Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups/D. Baum// Nature Education, 2008. Vol. 1. N. 1. P. 190.
15. COVID-19. Medical Conditions. Centers for Disease Control and Prevention URL: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html> (дата звернення: 20.11.2022).
16. Crick F. (1955). On degenerate templates and the adaptor hypothesis (Lecture)/ F. Crick // Francis Crick (1916-2004): archives. 1955.
17. DNA Barcoding. International Barcode of Life. URL: <https://ibol.org/about/dna-barcoding/> (дата звернення: 20.11.2022).
18. Esser D. Modeling microbial growth and dynamics/ D. S Esser, J. H J Leveau, K. M Meyer// Applied Microbiology and Biotechnology, 2015. Vol. 99. P. 8831–8846.

19. Evolution of Translational Omics Lessons Learned and the Path Forward./ Edited by C. M. Micheel, S. J. Nass, Gilbert S. Omenn// Washington (DC): National Academies Press (US), 2012.
20. Friendly M. Milestones in the history of thematic cartography, statistical graphics, and data visualization. Published 2008.
21. Genetic Mapping Fact Sheet. National Human Genome Research Institute. URL: <https://www.genome.gov/about-genomics/fact-sheets/Genetic-Mapping-Fact-Sheet> (дата звернення: 20.11.2022).].
22. Glynn P. Mathematical modeling of physiological systems: an essential tool for discovery/ P. Glynn, S. D Unudurthi, T. J. Hund// Life Sciences, 2014. Vol. 111. Issues 1-2. P. 1-5.
23. Green E. D. Gene Mapping. National Human Genome Research Institute. URL: <https://www.genome.gov/genetics-glossary/Gene-Mapping> (дата звернення: 20.11.2022).
24. Hebert P. D. N. Biological identifications through DNA barcodes/ P. D. N. Hebert, A. Cywinska, S. L. Ball et al.// Proceedings of the Royal Society B: Biological Sciences, 2003. – Vol. 270. N. 1512. P. 313 – 321.
25. Hebert P. D. N. Identification of Birds through DNA Barcodes/ P. D. N. Hebert, M. Y. Stoeckle, T. S. Zemplak et al.// PLoS Biology, 2004. Vol. 2. N. 10. e312.
26. Heidbuechel J. P. W. Mathematical Modeling of Oncolytic Virotherapy/ J. P. W. Heidbuechel, D. Abate-Daga, C. E. Engeland et al.// Methods in molecular biology, 2020. Vol. 2058. P. 307 – 320.
27. Hendriksen M. Tree-based unrooted nonbinary phylogenetic networks/ M. Hendriksen// Mathematical Biosciences, 2018. Vol. 3-2. P. 131 – 138.
28. Khorana H. G. Nucleic Acid Synthesis in the Study of the Genetic Code/ H. G. Khorana// Nobel Lecture. December 12, 1968.

29. Lahaye, R. DNA barcoding the floras of biodiversity hotspots/ R. Lahaye, M. van der Bank, D. Bogarin et al.// *Proceedings of the National Academy of Sciences*, 2008. Vol. 105. N. 8. P. 2923 – 2928.
30. Lane S. Graphical abstracts: a taxonomy and critique of an emerging genre/ S. Lane, A. Karatsolis, L. Bui// *Proceedings of the 33rd Annual International Conference on the Design of Communication*, 2015. Article No. 41. P. 1–9.
31. Lee G. MildInt: Deep Learning-Based Multimodal Longitudinal Data Integration Framework/ G. Lee, B. Kang, K. Nho et al.// *Frontiers in Genetics*, 2019. Vol. 10. P. 617.
32. Lopatkin A. J., Collins J. J. Predictive biology: modelling, understanding and harnessing microbial complexity. *Reviews Microbiology*, 2020. Vol. 18. P. 507–520
33. Membré J. M., Valdramidis V. *Modeling in Food Microbiology: from predictive microbiology to exposure assessment*. ISTE Ltd, 2016. 102 p.
34. Midway S. R. *Principles of Effective Data Visualization*/ S. R. Midway// *Patterns*, 2020. Vol. 1. N. 9. P. 100–141.
35. Moe J. Increased Use of Bayesian Network Models Has Improved Environmental Risk Assessments/J. Moe, J. F Carriger, M. Glendell// *Integrated Environmental Assessment and Management*, 2020. Vol. 00, N. 00. P. 1–9.
36. Morgan T. H. et al. *The Mechanism of Mendelian Heredity*. New York: Henry Holt, 1923.
37. Nathans D. Restriction endonucleases in the analysis and restructuring of dna molecules/ D. Nathans, H. O. Smith// *Annual Review of Biochemistry*, 1975. Vol. 44. P. 273 – 293.
38. Olar O.I., Mykytiuk O.Yu., Fediv V.I., Ostafichuk D.I. Methods of mathematical modeling as an important part of knowledge for students of medical and pharmaceutical specialties. *Proceedings of the IX International*

- Scientific and Practical Conference “Scientific Industry of the European Continent 2013”. Praha: Publishing House “Education and Science”, 2013.
39. Oliver G. R. Bioinformatics for clinical next generation sequencing/ G. R. Oliver , S. N. Hart, E. W. Klee// *Clinical Chemistry*, 2015. Vol. 61, N. 1. P. 124–135.
40. Pauling L. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain / L. Pauling L, R. B. Corey, H. R. Branson// *Proceedings National Academy of Sciences*, 1951.Vol. 37, N. 4. P. 205–211.
41. Pavlopoulos G. A. Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future/ G. A. Pavlopoulos, D. Malliarakis, N. Papanikolaou et al.// *GigaScience*, 2015.Vol. 4, Issue 1.
42. Temin H. M. RNA-dependent DNA polymerase in virions of Rous sarcoma virus/ H. M. Temin, S. Mizutani// *Nature*, 1970. Vol. 226. N. 5252. P. 1211– 1213.
43. Ushijima T. Mapping genomic and epigenomic evolution in cancer ecosystems/ T. Ushijima, S. J. Clark, P. Tan// *Science*, 2021. Vol. 373. N. 6562. P. 1474-1479.
44. Vecchio I. The Discovery of Insulin:An Important Milestone in the History of Medicine/ I. Vecchio, C. Tornali, N. L. Bragazzi et al.// *Frontiers in Endocrinology*, 2018. Vol. 9. P. 613
45. Watson J., Crick F. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid / J. Watson J, F. Crick // *Nature*, 1953. Vol. 171, N. 4356. P. 737 – 738
46. Woese C. R. Towards a natural system of organisms: proposalforthe domains Archaea, Bacteria, and Eucarya// *Proceedings ofthe National Academy of Sciences*, 1990. Vol. 87. N. 12. P. 4576–4579.

47. Yamoah G. G. Data Curation for Preclinical and Clinical Multimodal Imaging Studies/ G. G. Yamoah, L. Cao, C. W. Wu et al.// *Molecular Imaging and Biology*, 2019. Vol. 21. N. 6. P.1034–1043.

**КОДЕКС АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ
ЗДОБУВАЧА ВИЩОЇ ОСВІТИ ХЕРСОНЬСЬКОГО
ДЕРЖАВНОГО УНІВЕРСИТЕТУ**

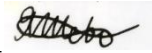
Я, Шевченко Ярослава Миколаївна, учасник(ця) освітнього процесу Херсонського державного університету, **УСВІДОМЛЮЮ**, що академічна доброчесність – це фундаментальна етична цінність усієї академічної спільноти світу.

ЗАЯВЛЯЮ, що у своїй освітній і науковій діяльності **ЗОБОВ'ЯЗУЮСЯ**:

- дотримуватися:
 - вимог законодавства України та внутрішніх нормативних документів університету, зокрема Статуту Університету;
 - принципів та правил академічної доброчесності;
 - нульової толерантності до академічного плагіату;
 - моральних норм та правил етичної поведінки;
 - толерантного ставлення до інших;
 - дотримуватися високого рівня культури спілкування;
- надавати згоду на:
 - безпосередню перевірку курсових, кваліфікаційних робіт тощо на ознаки наявності академічного плагіату за допомогою спеціалізованих програмних продуктів;
 - оброблення, збереження й розміщення кваліфікаційних робіт у відкритому доступі в інституційному репозитарії;
 - використання робіт для перевірки на ознаки наявності академічного плагіату в інших роботах виключно з метою виявлення можливих ознак академічного плагіату;
- самостійно виконувати навчальні завдання, завдання поточного й підсумкового контролю результатів навчання;
 - надавати достовірну інформацію щодо результатів власної навчальної (наукової, творчої) діяльності, використаних методик досліджень та джерел інформації;
 - не використовувати результати досліджень інших авторів без використання покликань на їхню роботу;
 - своєю діяльністю сприяти збереженню та примноженню традицій університету, формуванню його позитивного іміджу;
 - не чинити правопорушень і не сприяти їхньому скоєнню іншими особами;
 - підтримувати атмосферу довіри, взаємної відповідальності та співпраці в освітньому середовищі;
 - поважати честь, гідність та особисту недоторканність особи, незважаючи на її стать, вік, матеріальний стан, соціальне становище, расову належність, релігійні й політичні переконання;
 - не дискримінувати людей на підставі академічного статусу, а також за національною, расовою, статевою чи іншою належністю;
 - відповідально ставитися до своїх обов'язків, вчасно та сумлінно виконувати необхідні навчальні та науково-дослідницькі завдання;
 - запобігати виникненню у своїй діяльності конфлікту інтересів, зокрема не використовувати службових і родинних зв'язків з метою отримання нечесної переваги в навчальній, науковій і трудовій діяльності;
 - не брати участі в будь-якій діяльності, пов'язаній із обманом, нечесністю, списуванням, фабрикацією;
 - не підроблювати документи;
 - не поширювати неправдиву та компрометуючу інформацію про інших здобувачів вищої освіти, викладачів і співробітників;
 - не отримувати і не пропонувати винагород за несправедливе отримання будь-яких переваг або здійснення впливу на зміну отриманої академічної оцінки ;
 - не залякувати й не проявляти агресії та насильства проти інших, сексуальні домагання;
 - не завдавати шкоди матеріальним цінностям, матеріально-технічній базі університету та особистій власності інших студентів та/або працівників;
 - не використовувати без дозволу ректорату (деканату) символіки університету в заходах, не пов'язаних з діяльністю університету;
 - не здійснювати і не заохочувати будь-яких спроб, спрямованих на те, щоб за допомогою нечесних і негідних методів досягати власних корисних цілей;
 - не завдавати загрози власному здоров'ю або безпеці іншим студентам та/або працівникам.

УСВІДОМЛЮЮ, що відповідно до чинного законодавства у разі недотримання Кодексу академічної доброчесності буду нести академічну та/або інші види відповідальності й до мене можуть бути застосовані заходи дисциплінарного характеру за порушення принципів академічної доброчесності.

27.09.2021
(дата)



(підпис)

Ярослава Шевченко
(ім'я, прізвище)