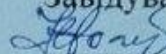


НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Факультет біотехнології і біотехніки
Кафедра біоенергетики, біоінформатики та екобіотехнології

До захисту допущено:

Завідувач кафедри

 Наталія ГОЛУБ

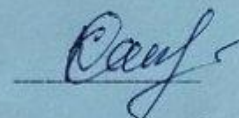
«03» травня 2025 р.

Дипломна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою «Біотехнології»
спеціальності 162 «Біотехнології та біоінженерія»
на тему: «Машинне навчання для передбачення форми магнітних
наночастинок у магнітотаксисних бактеріях»

Виконав:

студент IV курсу, групи ББ-12

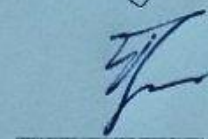
Хахно Кирил Юрійович



Керівник:

Проф. каф. біоенергетики, біоінформатики та екобіотехнології

Горобець Світлана Василівна

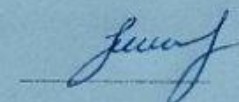


Рецензент:

Старший викладач каф. промислової

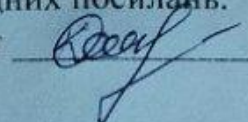
біотехнології та біофармації, к.т.н.

Тітова Лариса Олександрівна



Засвідчую, що у цій дипломній роботі немає
запозичень з праць інших авторів без
відповідних посилань.

Студент



Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет біотехнології і біотехніки
Кафедра біоенергетики, біоінформатики та екобіотехнології

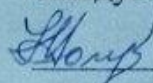
Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 162 «Біотехнології та біоінженерія»

Освітньо-професійна програма «Біотехнології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

 Наталія ГОЛУБ

«21» квітня 2025 р.

ЗАВДАННЯ

на дипломну роботу студенту

Хахно Кирилу Юрійовичу

1. Тема роботи «Машинне навчання для передбачення форми магнітних наночастинок у магнітотаксисних бактеріях», керівник роботи Горобець Світлана Василівна, д.т.н., професор, затверджені наказом по університету від «29» травня 2025 р. № 1838-с
2. Термін подання студентом роботи 03 червня 2025 року
3. Вихідні дані до роботи: дані про геноми та білки для 53 штамів магнітотаксисних бактерій та інформація з літератури про їх форму біогенних магнітних наночастинок.
4. Зміст роботи: літературний огляд за темою дослідження, методи збору додаткових дескрипторів для тренування моделей машинного навчання, опис роботи моделей дерева рішень та випадкового лісу, розробка алгоритму на основі двох моделей машинного навчання для передбачення форми біогенних магнітних

наночастинок магнітотаксисних бактерій, пошук недосліджених штамів магнітотаксисних бактерій та передбачення їх форми біогенних магнітних наночастинок, аналіз отриманих даних.

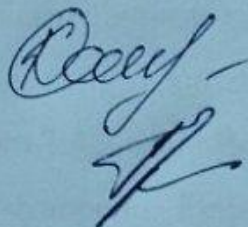
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): Презентація (А4).

6. Дата видачі завдання _____ 21 квітня 2025 року _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Огляд літературних джерел та наукових публікацій	28.04.2025	Виконано
2	Розробка програми та алгоритму, отримання результатів	12.05.2025	Виконано
3	Аналіз отриманих результатів	27.05.2025	Виконано
4	Огляд нормативних документів з охорони праці	29.05.2025	Виконано
5	Оформлення роботи	01.06.2025	Виконано
7	Подання дипломної роботи	03.06.2025	Виконано

Студент



Кирил ХАХНО

Керівник

Світлана ГОРОБЕЦЬ

РЕФЕРАТ

Дипломна робота містить 86 сторінок, 16 рисунків, 3 таблиці, 3 додатки, 38 джерел літератури.

Мета роботи: побудова моделі для передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій на основі даних про наявність генів *matA*, *matB*, *matE*, *matI*, *matK*, *matM*, *matP*, *matQ*, *matH*, *matF*, *matS*, *matT*, *matC*, *matD*, *matG*, *matR*, *matL*, *matO*, *matN*, *matX*, *matZ*, *matY*, *matJ*, *matU*, *matV*, *matW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічних властивостей кодованих цими генами білків методами машинного навчання.

Об'єкт дослідження: геноми та протеоми магнітотаксисних бактерій.

Предмет дослідження: проблема взаємозв'язку форми біогенних магнітних наночастинок магнітотаксисних бактерій з наявністю генів *matA*, *matB*, *matE*, *matI*, *matK*, *matM*, *matP*, *matQ*, *matH*, *matF*, *matS*, *matT*, *matC*, *matD*, *matG*, *matR*, *matL*, *matO*, *matN*, *matX*, *matZ*, *matY*, *matJ*, *matU*, *matV*, *matW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічними властивостями кодованих цими генами білків.

Матеріали і методи дослідження: експериментальні дані про форму 53 біогенних магнітних наночастинок магнітотаксисних бактерій, дані про геноми та протеоми магнітотаксисних бактерій бази даних GenBank та Protein NCBI. Методи машинного навчання дерево рішень та випадковий ліс, методи бібліотек Biopython 1.84, pandas 2.2.2 та scikit-learn 1.4.2 мови Python 3.12.3.

Результати роботи: створено Python-скрипт для передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій одного з 4 класів (призматична, кубо-октаедрична, форма кулі чи вигнутої кулі). Знайдено дані про протеоми 6 штамів в базі даних GenBank та Protein NCBI, форма БМН для яких експериментально не досліджена, показано, що ці 4 штами є потенційними продуцентами призматичних біогенних магнітних наночастинок.

Наукова новизна:

- Вперше встановлено взаємозв'язок форми біогенних магнітних наночастинок магнітотаксисних бактерій з даними про наявність генів *mamA*, *mamB*, *mamE*, *mamI*, *mamK*, *mamM*, *mamP*, *mamQ*, *mamH*, *mamF*, *mamS*, *mamT*, *mamC*, *mamD*, *mamG*, *mamR*, *mamL*, *mamO*, *mamN*, *mamX*, *mamZ*, *mamY*, *mamJ*, *mamU*, *mamV*, *mamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічними властивостями кодованих цими генами білків.
- Вперше передбачено форму біогенних магнітних наночастинок для наступних магнітотаксисних бактерій: *Alpha proteobacterium* LM-2, *Alpha proteobacterium* LEMS, *Alpha proteobacterium* KR-1 та *Alpha proteobacterium* CB-1, *Candidatus Magnetominusculus xianensis* HCH-1 та *Magnetospirillum* sp. SS-4.

Ключові слова: БІОГЕННІ МАГНІТНІ НАНОЧАСТИНКИ, МАГНІТОТАКСИСНІ БАКТЕРІЇ, МАШИННЕ НАВЧАННЯ, ДЕРЕРО РІШЕНЬ, ВИПАДКОВИЙ ЛІС.

ABSTRACT

This thesis contains 86 pages, 16 figures, 3 tables, 3 appendixes, 38 references to literary sources.

Purpose of the study: building a model to predict the shape of biogenic magnetic nanoparticles of magnetotaxis bacteria based on data on the presence of the genes *mamA*, *mamB*, *mamE*, *mamI*, *mamK*, *mamM*, *mamP*, *mamQ*, *mamH*, *mamF*, *mamS*, *mamT*, *mamC*, *mamD*, *mamG*, *mamR*, *mamL*, *mamO*, *mamN*, *mamX*, *mamZ*, *mamY*, *mamJ*, *mamU*, *mamV*, *mamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, as well as numerous genes of the *mad* (from *mad1* to *mad31*) and *man* (from *man1* to *man6*) groups in their genomes, as well as the physicochemical properties of the proteins encoded by these genes using machine learning methods.

Object of research: genomes and proteomes of magnetotaxis bacteria.

Subject of the research: the problem of the relationship between the shape of biogenic magnetic nanoparticles of magnetotaxis bacteria with the presence of the genes *mamA*, *mamB*, *mamE*, *mamI*, *mamK*, *mamM*, *mamP*, *mamQ*, *mamH*, *mamF*, *mamS*, *mamT*, *mamC*, *mamD*, *mamG*, *mamR*, *mamL*, *mamO*, *mamN*, *mamX*, *mamZ*, *mamY*, *mamJ*, *mamU*, *mamV*, *mamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, as well as numerous genes of the *mad* (from *mad1* to *mad31*) and *man* (from *man1* to *man6*) groups in their genomes, as well as the physicochemical properties of the proteins encoded by these genes.

Materials and methods: experimental data on the form of 53 biogenic magnetic nanoparticles of magnetotaxis bacteria, data on genomes and proteomes of magnetotaxis bacteria from the GenBank and Protein NCBI databases. Machine learning methods decision tree and random forest, methods of Biopython 1.84, pandas 2.2.2 and scikit-learn 1.4.2 libraries for Python 3.12.3.

Results: the Python script was created to predict the shape of biogenic magnetic nanoparticles of magnetotaxis bacteria of one of 4 classes (prismatic, cubo-octahedral, spherical or curved spherical). Data on proteomes of 6 strains were found in the GenBank and Protein NCBI databases, for which the shape of BMN has not been experimentally

studied, it was shown that these 4 strains are potential producers of prismatic biogenic magnetic nanoparticles.

Original:

- For the first time, the relationship between the shape of biogenic magnetic nanoparticles of magnetotaxis bacteria and data on the presence of the genes *mamA*, *mamB*, *mamE*, *mamI*, *mamK*, *mamM*, *mamP*, *mamQ*, *mamH*, *mamF*, *mamS*, *mamT*, *mamC*, *mamD*, *mamG*, *mamR*, *mamL*, *mamO*, *mamN*, *mamX*, *mamZ*, *mamY*, *mamJ*, *mamU*, *mamV*, *mamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, as well as numerous genes of the *mad* (from *mad1* to *mad31*) and *man* (from *man1* to *man6*) groups in their genomes, as well as the physicochemical properties of the proteins encoded by these genes.
- For the first time, the shape of biogenic magnetic nanoparticles has been predicted for the following magnetotaxis bacteria: *Alpha proteobacterium* LM-2, *Alpha proteobacterium* LEMS, *Alpha proteobacterium* KR-1, *Alpha proteobacterium* CB-1, *Candidatus Magnetominusculus xianensis* HCH-1 and *Magnetospirillum sp.* SS-4.

Key words: BIOGENIC MAGNETIC NANOPARTICLES, MAGNETOTACTIC BACTERIA, MACHINE LEARNING, DECISION TREE, RANDOM FOREST.

ЗМІСТ

Перелік умовних скорочень	10
Вступ.....	11
РОЗДІЛ 1 Літературний огляд	15
1.1 Магнітогенетика.....	15
1.1.1 Визначення та принципи магнітогенетики.....	15
1.1.2 Компоненти магнітогенетичних систем	16
1.1.3 Сучасні застосування та перспективи магнітогенетики	18
1.2 Види магнітних актуаторів для задач магнітогенетики	21
1.2.1 Типи магнітних матеріалів.....	21
1.2.2 Хімічний синтез магнітних наночастинок.....	22
1.2.3 Біогенні магнітні наночастинок як альтернативне джерело	22
1.3 Магнітотаксисні бактерії.....	23
1.3.1 Загальна характеристика магнітотаксисних бактерій	23
1.3.2 Біогенні магнітні наночастинок.....	24
1.3.3 Виклики культивування магнітотаксисних бактерій	25
1.3.4 Досягнення у методах культивування МТБ	26
1.4 Вплив форми магнітних наночастинок на їх властивості.....	26
1.4.1 Вплив форми на фізичні властивості магнітних наночастинок	26
1.4.2 Дослідження форми та морфології магнітних наночастинок.....	28
1.5 Генетичний апарат синтезу біогенних магнітних наночастинок	29
1.5.1 Магнітосомний острівець.....	29
1.5.2 Основні групи генів у складі магнітосомного острівця та їх функції	30
РОЗДІЛ 2 Матеріали та методи дослідження.....	34
2.1 Матеріали дослідження	34
2.2 Методи дослідження.....	40
2.2.1 Дерево рішень.....	40
2.2.2 Випадковий ліс	42
2.2.3 Параметри тренування моделей	43
2.2.5 Зовнішня валідація	46

РОЗДІЛ 3 Результати та їх обговорення.....	48
3.1 Тренування моделей машинного навчання	48
3.1.1 Результати тренування моделей	48
3.1.2 Результати перехресної перевірки моделей	51
3.2 Результати зовнішньої валідації	52
3.3 Передбачення форми для недосліджених штамів	57
РОЗДІЛ 4 Охорона праці та довкілля.....	60
4.1 Загальні положення.....	60
4.2 Вимоги безпеки перед початком роботи	60
4.3 Вимоги безпеки під час роботи	61
4.4 Вимоги охорони праці при користуванні хімічним посудом	62
4.5 Вимоги охорони праці при користуванні персональним комп'ютером.....	62
4.6 Вимоги безпеки після закінчення роботи.....	63
4.7 Вимоги безпеки в аварійних ситуаціях.....	63
Висновки	66
Перелік посилань.....	67
Додаток А.....	71
Додаток Б.....	76
Додаток В	80

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

БМН – біогенні магнітні наночастинки

МТБ – магнітотаксисні бактерії

МНЧ – магнітні наночастинки

АСМ – атомно-силова мікроскопія

ТЕМ – трансмісійна електронна мікроскопія

МСМ – магнітна силова мікроскопія

ЗІЗ – засоби індивідуального захисту

ВСТУП

Останніми роками спостерігається стрімкий розвиток міждисциплінарних досліджень на межі біології та магнетизму, що призвело до появи інноваційної галузі – магнітогенетики [1]. Ця сфера відкриває унікальні перспективи для неінвазивного керування клітинними процесами за допомогою магнітних полів. На відміну від традиційних методів впливу, таких як оптогенетика або хемогенетика, магнітогенетика використовує магнітні поля, які легко проникають крізь біологічні тканини, пропонуючи глибокий та нешкідливий вплив. В основі магнітогенетичного підходу лежить використання магніточутливих елементів, здатних трансформувати зовнішній магнітний сигнал у біологічну відповідь клітини [1].

Особливу увагу в контексті магнітогенетики привертають біогенні магнітні наночастинки (БМН), зокрема магнітосоми, що виробляються магнітотаксисними бактеріями (МТБ). Ці природні органели, що складаються з кристалів магнетиту або грейгіту [1], оточених мембраною, є ідеальними кандидатами для розробки магнітогенетичних інструментів завдяки своїм винятковим магнітним властивостям, біосумісності та здатності до біоконтрольованого утворення. Ефективність та функціональність цих наночастинок у біологічних застосуваннях значною мірою залежить від їхніх фізичних характеристик, включаючи розмір, кристалічну структуру, але особливо – від форми.

Розуміння та контроль форми БМН є **актуальними** для точного налаштування їхніх магнітних властивостей та, як наслідок, їхньої ефективності у магнітогенетиці. Однак традиційні методи визначення форми БМН, такі як електронна мікроскопія, є надзвичайно трудомісткими, потребують значних ресурсів та не дозволяють проводити швидкий скринінг великої кількості зразків. У цьому контексті, застосування сучасних обчислювальних підходів, зокрема методів машинного навчання, пропонує перспективний шлях для розробки швидких, ефективних та неруйнівних методів передбачення форми БМН на основі інших, легше вимірюваних характеристик, таких як генетичні ознаки та біохімічні властивості пов'язаних білків. Розробка таких підходів може значно оптимізувати

процес відбору та модифікації магнітних наночастинок для специфічних магнітогенетичних застосувань, а також поглибити фундаментальне розуміння процесів біомінералізації у МТБ [1-3].

Метою цієї роботи є побудова моделі для передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій на основі даних про наявність генів *matA*, *matB*, *matE*, *matI*, *matK*, *matM*, *matP*, *matQ*, *matH*, *matF*, *matS*, *matT*, *matC*, *matD*, *matG*, *matR*, *matL*, *matO*, *matN*, *matX*, *matZ*, *matY*, *matJ*, *matU*, *matV*, *matW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічних властивостей кодованих цими генами білків методами машинного навчання.

Для досягнення мети було поставлено такі **завдання**:

- виконати пошук наукової літератури за темою роботи;
- сформулювати та підготувати набір даних, що включає генетичну інформацію та властивості білків магнітосомного острівця;
- розробити алгоритм для передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій;
- реалізувати програму передбачення форми біогенних магнітних наночастинок, використовуючи алгоритми машинного навчання та мову програмування Python;
- оцінити точність та стабільність розроблених моделей за допомогою перехресної перевірки;
- провести тестування моделей на незалежних експериментальних даних;
- описати основні заходи щодо охорони праці під час роботи із персональним комп'ютером та у мікробіологічній лабораторії.

Об'єктом дослідження є геноми та протеоми магнітотаксисних бактерій.

Предметом дослідження є проблема взаємозв'язку форми біогенних магнітних наночастинок магнітотаксисних бактерій з наявністю генів *matA*, *matB*, *matE*, *matI*, *matK*, *matM*, *matP*, *matQ*, *matH*, *matF*, *matS*, *matT*, *matC*, *matD*, *matG*, *matR*, *matL*, *matO*, *matN*, *matX*, *matZ*, *matY*, *matJ*, *matU*,

tamV, *tamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічними властивостями кодованих цими генами білків.

Наукова новизна.

- Вперше встановлено взаємозв'язок форми біогенних магнітних наночастинок магнітотаксисних бактерій з даними про наявність білків MamL та MamO в їх протеомах, а також фізико-хімічними властивостями білків.
- Вперше передбачено форму біогенних магнітних наночастинок для наступних магнітотаксисних бактерій: *Alpha proteobacterium* LM-2, *Alpha proteobacterium* LEMS, *Alpha proteobacterium* KR-1 та *Alpha proteobacterium* CB-1, *Candidatus Magnetominusculus xianensis* HCN-1 та *Magnetospirillum* sp. SS-4.

Практична цінність.

- Розроблений алгоритм передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій може бути застосований в галузях, де широко використовуються магнітні наночастинок: лікування раку, електроніка, адресна доставка ліків та інші для пошуку нових продуцентів БМН з генетично контрольованою формою.
- Встановлення взаємозв'язку форми біогенних магнітних наночастинок магнітотаксисних бактерій з даними про наявність генів *tamA*, *tamB*, *tamE*, *tamI*, *tamK*, *tamM*, *tamP*, *tamQ*, *tamH*, *tamF*, *tamS*, *tamT*, *tamC*, *tamD*, *tamG*, *tamR*, *tamL*, *tamO*, *tamN*, *tamX*, *tamZ*, *tamY*, *tamJ*, *tamU*, *tamV*, *tamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічними властивостями кодованих цими генами білків сприятиме більш повному розумінню генетичного механізму контролю форми біогенних магнітних наночастинок, що важливо в генній інженерії.

Апробація. Методи, що були використані в цій дипломній роботі, були апробовані у таких публікаціях:

- Горобець С. В., Горобець О. Ю., Хахно К. Ю. Мова python для інженерних та наукових задач : підручник для здобувачів ступеня бакалавра за спеціальністю 162 «Біотехнології та біоінженерія», магістра за спеціальністю за спеціальністю 162 «Біотехнології та біоінженерія», бакалавра за спеціальністю 104 «Фізика та астрономія» та магістра за спеціальністю 104 «Фізика та астрономія». Київ : КПІ ім. Ігоря Сікорського, 2024. 277 с. URL: <https://ela.kpi.ua/handle/123456789/67188>.
- Хахно К. Ю, Горобець С. В. Побудова дерева рішень для передбачення форми біогенних магнітних наночастинок у магнітотаксисних бактеріях / «Біотехнологія XXI століття»: матеріали XVIII Міжнар. науково-практ. конф., м. Київ, 17 трав. 2024 р. Київ, КПІ ім. Ігоря Сікорського, 2024. – С. 283–286.
- Хахно К. Ю, Горобець С. В. Передбачення форми біогенних магнітних наночастинок у магнітотаксисних бактеріях за допомогою дерев рішень та випадкових лісів / «Біотехнологія XXI століття»: матеріали XIX Міжнар. науково-практ. конф., м. Київ, 16 трав. 2025 р. Київ, КПІ ім. Ігоря Сікорського, 2025. – С. 299–302.

РОЗДІЛ 1 ЛІТЕРАТУРНИЙ ОГЛЯД

1.1 Магнітогенетика

1.1.1 Визначення та принципи магнітогенетики

Магнітогенетика – це нова міждисциплінарна галузь, що знаходиться на перетині фізики, хімії, матеріалознавства, інженерії та біології, яка досліджує можливості використання магнітних полів та магнітних матеріалів (здебільшого у вигляді наночастинок) для дистанційного контролю біологічних процесів у живих системах. На відміну від оптогенетики, яка використовує світло для активації біологічних функцій, магнітогенетика оперує магнітними полями, що мають значно кращу проникність у біологічні тканини, дозволяючи впливати на глибоко розташовані клітини та органи неінвазивним або мінімально інвазивним шляхом [1]. Ця перевага робить магнітогенетику перспективним інструментом для фундаментальних досліджень та потенційних терапевтичних застосувань.

Основний принцип магнітогенетики полягає у використанні магнітних наночастинок (МНЧ) як трансд'юсерів, які здатні перетворювати енергію зовнішнього магнітного поля на біологічно значущий сигнал. Цей сигнал потім використовується для активації або інгібування специфічних клітинних функцій. Взаємодія між зовнішнім магнітним полем та МНЧ може відбуватися різними шляхами, залежно від характеристик поля (статичне, змінне, пульсуюче, однорідне, градієнтне), властивостей самих МНЧ (розмір, форма, склад, магнітні властивості) та їхнього розташування відносно біологічної мішені [1].

Розвиток методів неінвазивного контролю біологічних процесів на клітинному та тканинному рівнях є критично важливим для фундаментальних досліджень та розробки нових терапевтичних підходів. Традиційні методи стимуляції, такі як електрична або оптична, мають певні обмеження, пов'язані з інвазивністю у випадку електричної стимуляції або обмеженою глибиною проникнення в біологічні тканини у випадку оптогенетики. У відповідь на ці виклики виникла та активно розвивається нова галузь біоінженерії – магнітогенетика. Цей напрямок інтегрує принципи генетики або молекулярної біології та фізики магнетизму, дозволяючи маніпулювати функціями клітин за

допомогою зовнішніх магнітних полів. Ключовою перевагою магнітних полів є їхня здатність проникати глибоко в біологічні тканини практично без ослаблення, що відкриває шлях до створення неінвазивних біоінтерфейсів [1, 2]. Магнітогенетичні системи, як правило, складаються з трьох основних компонентів: зовнішнього джерела магнітного поля, магнітних наночастинок або інших магнітних структур, доставлених до біологічної мішені, та біологічної мішені, що відповідає на дію МНЧ, такої як іонний канал, білок, що змінює конформацію, або генетичний елемент [1]. Ефективність та специфічність магнітогенетичного контролю значною мірою залежать від властивостей використаних магнітних наночастинок та методів їх зв'язку з біологічними мішенями.

1.1.2 Компоненти магнітогенетичних систем

Для ефективного здійснення магнітогенетичного контролю над біологічними процесами необхідна злагоджена взаємодія кількох ключових компонентів, які спільно забезпечують перетворення зовнішнього магнітного сигналу у специфічну біологічну відповідь. Як детально описано в оглядових роботах з магнітогенетики [1, 3], ці компоненти функціонують як система «тригер-трансдуктор-ефектор».

Першим ініціюючим компонентом є зовнішнє магнітне поле. Це первинний тригер системи, який передає енергію або силу дистанційно, без фізичного контакту з біологічною тканиною. Різні типи магнітних полів можуть застосовуватися для індукування специфічних фізичних ефектів у магнітних актуаторах, зв'язаних з клітинними мішенями, як це наочно продемонстровано на рисунку 1. Так, градієнтне постійне поле може створювати силу притягання або відштовхування, спричиняючи рух або натяг магнітного актуатора. Однорідне постійне поле може призводити до дипольних взаємодій між сусідніми магнітними актуаторами, спричиняючи їх агрегацію або зміну конформації зв'язаних білків. Обертове однорідне постійне поле здатне генерувати крутний момент, змушуючи магнітні актуатори обертатися. Змінне магнітне поле є найбільш універсальним і може індукувати генерацію тепла в магнітних актуаторах через процес релаксації, що призводить до локального підвищення температури (магнітна гіпертермія).

Також існують повідомлення про те, що деякі типи МНЧ під дією змінного магнітного поля можуть генерувати активні форми кисню (ROS) [4]. Перевага магнітних полів як тригера полягає у їхній високій проникності через біологічні тканини порівняно зі світлом, що дозволяє маніпулювати клітинами у глибоко розташованих ділянках організму. Практична реалізація цього компонента вимагає використання спеціалізованих котушок або магнітних джерел для генерації полів з точно заданими параметрами (частотою, амплітудою, градієнтом), що відповідають бажаному механізму активації.

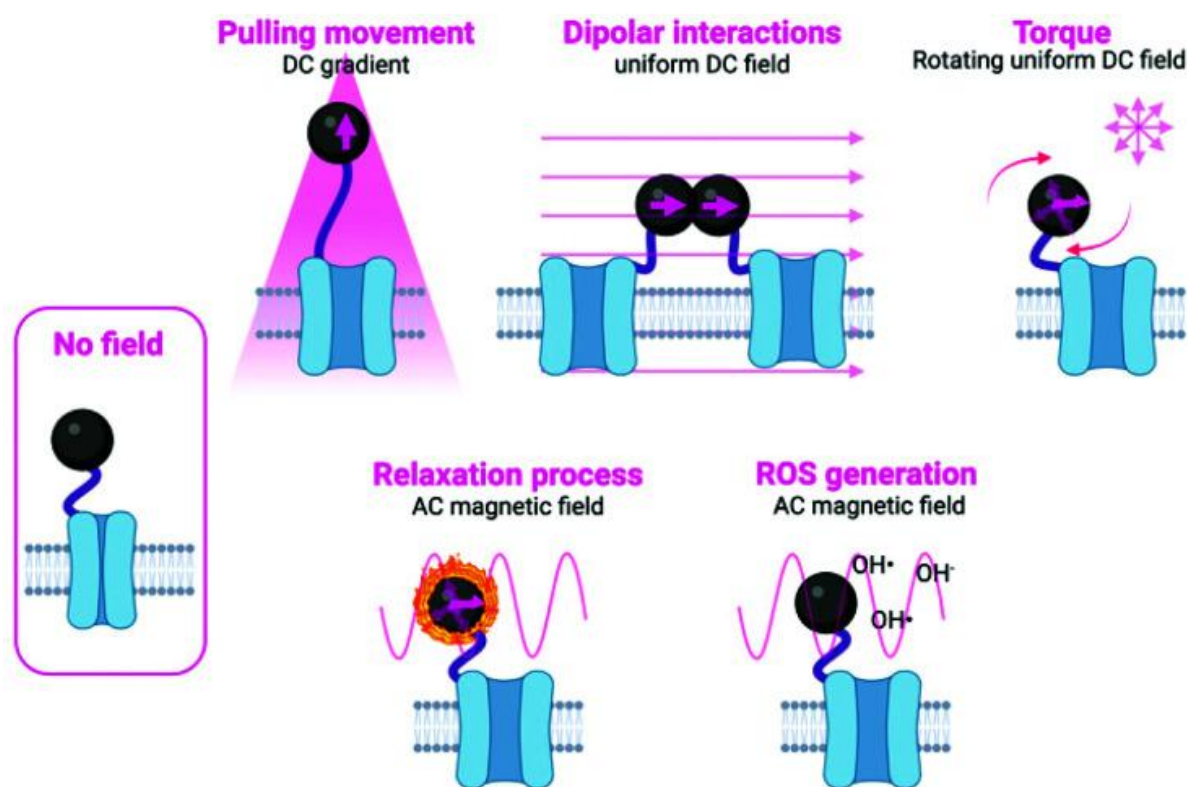


Рисунок 1 – Механізми впливу на магнітні актуатори [1]

Другий компонент системи – магнітні актуатори. Ці елементи є посередниками, які перетворюють енергію або силу зовнішнього магнітного поля у фізичний сигнал, що може вплинути на біологічну систему. Найчастіше в ролі магнітних актуаторів виступають магнітні наночастинки, переважно на основі оксидів заліза (Fe_3O_4 , $\gamma\text{-Fe}_2\text{O}_3$) через їхню відносно низьку токсичність та добре вивчені магнітні властивості (суперпарамагнетизм або феромагнетизм). Під дією зовнішнього магнітного поля, як описано вище, МНЧ можуть генерувати локальне тепло, відчувати значний крутний момент та механічні сили. Також як актуатори

можуть використовуватись генетично інженерні білки, що містять магнітні домени або здатні асоціюватись з ендogenous залізом, змінюючи свою конформацію або локалізацію під дією поля [4].

Третім компонентом системи є клітинні мішені. Це біологічні елементи у клітинах або тканинах, які здатні реагувати на фізичний сигнал, що генерується магнітними актуаторами, і, як наслідок, змінювати свою функцію або ініціювати специфічні клітинні процеси. Ці мішені є ефекторами біологічної відповіді. Залежно від типу актуатора та механізму трансдукції, клітинними мішенями можуть бути: термочутливі іонні канали (наприклад, TRP-канали [5]), які активуються локальним підвищенням температури, спричиненим МНЧ; механочутливі іонні канали (наприклад, Piezo1 [6]), які відкриваються у відповідь на механічні сили або деформацію мембрани, індуковану рухом або обертанням актуаторів; або рецептори та сигнальні білки, чия активність може бути модульована шляхом кон'югації з магнітними актуаторами та застосування механічного впливу або репозиціонування. Таким чином, клітинні мішені переводять фізичний вплив від актуаторів у специфічну біологічну відповідь, завершуючи магнітогенетичний цикл контролю.

1.1.3 Сучасні застосування та перспективи магнітогенетики

Однією з найбільш перспективних сфер застосування є нейронауки. Тут магнітогенетика надає можливість дистанційно та контрольовано модулювати активність нейронів [7] – збуджувати або пригнічувати їхню електричну активність. Це дозволяє дослідникам вивчати складні нейронні ланцюги, розшифровувати механізми, що лежать в основі поведінки, та моделювати патологічні стани, пов'язані з дисфункцією нейронів. Глибоке проникнення магнітного поля робить цей метод особливо цінним для маніпуляцій з нейронами у глибоко розташованих структурах мозку без необхідності інвазивного введення оптоволокон, як у випадку з оптогенетикою.

Магнітогенетика також має значні перспективи у протипухлинній терапії. Один з добре вивчених підходів – це магнітна гіпертермія [8], де магнітні

наночастинки накопичуються в пухлинній тканині та під дією змінного магнітного поля генерують тепло, що призводить до загибелі ракових клітин або робить їх більш чутливими до хіміо- чи радіотерапії. Крім того, МНЧ використовуються як носії для адресної доставки [3] хіміотерапевтичних препаратів або інших терапевтичних агентів до пухлини, часто з можливістю контрольованого вивільнення під впливом магнітного поля.

Для фундаментальних досліджень сигнальних шляхів магнітогенетика надає можливість точно та локалізовано активувати специфічні сигнальні каскади всередині клітин [3]. Прив'язуючи магнітні актуатори до ключових компонентів сигнальних шляхів або використовуючи канали, що ініціюють сигналізацію у відповідь на магнітний стимул, дослідники можуть вивчати динаміку та функціональне значення цих шляхів. Цей метод дозволяє детально аналізувати клітинну відповідь на фізичні стимули, такі як механічний натяг або локальне підвищення температури.

Попри значний прогрес, досягнутий за останні роки, магнітогенетика як відносно молода галузь все ще стикається з низкою суттєвих викликів, подолання яких є критично важливим для її подальшого розвитку та потенційного клінічного застосування. Ці виклики охоплюють різні аспекти – від властивостей самих магнітних елементів до складності контролю та питань безпеки.

Одним з першочергових завдань є покращення ефективності та специфічності магнітних актуаторів [1], насамперед магнітних наночастинок. Незважаючи на успіхи у синтезі, часто існує необхідність підвищення їхньої чутливості до магнітного поля, щоб досягти бажаного біологічного ефекту за меншої напруженості або частоти поля, що є безпечнішим для організму. Це вимагає вдосконалення методів синтезу для отримання МНЧ з оптимальними магнітними властивостями (наприклад, високою питомою втратою потужності для гіпертермії або значним магнітним дипольним моментом для механічної стимуляції), контрольованим розміром та формою. Крім того, ключовим є забезпечення специфічного взаємодії МНЧ з цільовими клітинними мішенями (іонними каналами, рецепторами тощо) та мінімізація нецільового впливу. Це часто

досягається шляхом функціоналізації поверхні МНЧ специфічними лігандами або антитілами.

Значним викликом є зниження токсичності та забезпечення біосумісності компонентів магнітогенетичних систем [1]. Хоча оксиди заліза вважаються відносно біосумісними, їх довготривала присутність в організмі, біорозподіл та шляхи виведення потребують ретельного вивчення. Розмір, форма, поверхневий заряд та покриття МНЧ можуть впливати на їх взаємодію з біологічними системами, схильність до агрегації, клітинне поглинання та імунну відповідь [9]. Необхідні подальші дослідження для розробки МНЧ з покращеними профілями безпеки, можливо, шляхом використання біорозкладних покриттів або дослідження альтернативних біогенних магнітних матеріалів, таких як магнітосоми з магнітотаксисних бактерій, які мають природну біосумісність та стабільність [10].

Розробка складніших систем контролю є ще одним важливим напрямком. Більшість сучасних магнітогенетичних систем забезпечують лише просту активацію або інактивацію. Для тонкої модуляції складних біологічних процесів (наприклад, патернів активності нейронів або каскадів сигналізації) потрібні системи, що дозволяють здійснювати просторово- та часово-точний контроль з можливістю регулювання інтенсивності стимулу. Це вимагає розробки більш складних магнітних актуаторів, які реагують на різні параметри поля, а також вдосконалення систем генерації магнітних полів для створення локалізованих полів або полів зі складними часовими профілями [11].

Зрештою, головним викликом є поширення технології на клінічну практику. Це комплексне завдання, що включає не лише вирішення вищезазначених проблем безпеки та ефективності, але й проходження суворих доклінічних та клінічних випробувань. Необхідно продемонструвати відтворюваність результатів, масштабованість виробництва компонентів, стандартизацію процедур та відповідність регуляторним вимогам. Клінічна трансляція магнітогенетики потребує інтеграції з існуючими діагностичними та терапевтичними підходами та вирішення логістичних питань, пов'язаних з введенням магнітних елементів та застосуванням зовнішнього магнітного поля у клінічних умовах.

1.2 Види магнітних актуаторів для задач магнітогенетики

1.2.1 Типи магнітних матеріалів

Основа магнітних наночастинок, що використовуються або досліджуються для біомедичних застосувань, складають матеріали з відповідними магнітними властивостями. До найбільш поширених належать оксиди заліза, зокрема магнетит (Fe_3O_4) та магеміт ($\gamma\text{-Fe}_2\text{O}_3$) [12]. Ці матеріали є привабливими завдяки відносно низькій токсичності, біосумісності та можливості отримання в нанорозмірі з бажаними магнітними характеристиками. Інші матеріали включають грейгіт (Fe_3S_4), який за своєю структурою схожий на магнетит, а також різноманітні ферити (наприклад, кобальтові або нікелеві ферити) та сплави металів (такі як FePt , FeCo). Останні можуть мати сильніші магнітні властивості, проте їх застосування часто обмежене потенційною токсичністю або схильністю до окислення [12, 13]. Ключовими магнітними властивостями, релевантними для біомедицини, є суперпарамагнетизм та феромагнетизм.

Суперпарамагнетизм – це властивість, при якій наночастинка має магнітний момент, але у відсутності зовнішнього магнітного поля він швидко змінює напрямок під дією теплової енергії, в результаті чого сумарна намагніченість зразка дорівнює нулю [13]. Це надзвичайно важливо для біомедичних застосувань, оскільки суперпарамагнітні МНЧ не агрегують під дією залишкової намагніченості після зняття зовнішнього поля, що забезпечує їх стабільність у біологічних рідинах. Феромагнетизм проявляється у більших частинок або при нижчих температурах і характеризується наявністю залишкової намагніченості після зняття поля, що може призводити до агрегації, але забезпечує сильнішу реакцію на статичні поля [12, 13].

Вибір матеріалу та його розміру визначає, чи буде частинка суперпарамагнітною чи феромагнітною, впливаючи на її придатність для конкретного застосування (наприклад, гіпертермія, механічна стимуляція, контрастування).

1.2.2 Хімічний синтез магнітних наночастинок

На сьогоднішній день хімічний синтез є найбільш поширеним методом отримання магнітних наночастинок для дослідницьких та комерційних цілей. Існує широкий спектр хімічних підходів, що дозволяють контролювати різні аспекти синтезу. Серед основних методів можна виділити: співосадження (coprecipitation) – простий, масштабований, як правило, водний метод, що передбачає осадження іонів металів з розчину; термічний розклад (thermal decomposition) – метод, що часто проводиться в органічних розчинниках при високих температурах з використанням металоорганічних прекурсорів, відомий своєю здатністю отримувати МНЧ з високою монодисперсністю та кристалічністю [12, 14]; гідротермальний та сольвотермальний синтез – проводяться у закритих посудинах при високих температурах і тисках, дозволяючи контролювати фазовий склад та морфологію частинок; та мікроемульсійні методи, де реакція відбувається у обмежених просторах міцел, що сприяє отриманню частинок заданого розміру [12]. Ці методи дають можливість контролювати розмір, форму, склад та поверхню синтетичних МНЧ шляхом варіювання таких параметрів, як температура, рН, концентрація прекурсорів, тип розчинника та використання поверхнево-активних речовин або стабілізаторів [13]. Однак хімічний синтез має і певні недоліки. Багато методів потребують високотемпературних або високотискових умов, використання токсичних або вибухонебезпечних реагентів [12]. Часто кінцевий продукт є гетерогенним за розміром та формою, особливо у випадку співосадження, що вимагає подальших етапів сортування. Також МНЧ, отримані хімічним шляхом, зазвичай потребують подальшої функціоналізації поверхні (наприклад, покриття полімерами, діоксидом кремнію або ліпідами) для покращення біосумісності, запобігання агрегації, забезпечення стабільності в біологічних середовищах та приєднання специфічних лігандів для таргетування [12].

1.2.3 Біогенні магнітні наночастинок як альтернативне джерело

На відміну від хімічного синтезу, природа пропонує альтернативне джерело магнітних наночастинок у вигляді біогенних структур, синтезованих певними

живими організмами. Це явище, відоме як біомінералізація, включає контрольоване утворення неорганічних кристалів клітинами. Найбільш відомим прикладом біогенних магнітних наночастинок є магнітосоми, що синтезуються групою мікроорганізмів, відомих як магнітотаксисні бактерії [2]. На відміну від хімічно синтезованих МНЧ, магнітосоми утворюються всередині клітини під суворим генетичним контролем, що призводить до їх унікальних властивостей [2]. Переваги біогенних магнітосом над хімічно синтезованими МНЧ для біомедичних застосувань є значними і будуть детальніше розглянуті у наступному розділі. Вони включають високу кристалічність, хімічну чистоту, надзвичайно вузький розподіл за розміром, однорідну та часто видоспецифічну форму, природну стабільність та біосумісність завдяки мембранному оточенню, а також оптимальні магнітні властивості [15]. Ці характеристики роблять магнітосоми привабливою альтернативою хімічно синтезованим МНЧ для багатьох біомедичних цілей, включаючи магнітогенетику.

1.3 Магнітотаксисні бактерії

1.3.1 Загальна характеристика магнітотаксисних бактерій

Магнітотаксисні бактерії – це унікальна група філогенетично різноманітних прокаріотичних мікроорганізмів, які мають виняткову здатність синтезувати внутрішньоклітинні, оточені мембраною кристали магнітних мінералів [16]. Ці кристали, відомі як магнітосоми, дозволяють бактеріям орієнтуватися та рухатися вздовж ліній геомагнітного поля Землі – феномен, що отримав назву магнітотаксис [16]. Екологічне значення цієї поведінки полягає в тому, що вона допомагає МТБ ефективно знаходити та підтримувати оптимальне положення в стратифікованих водних середовищах (наприклад, озерах чи болотах). Більшість МТБ є мікроаерофільними (потребують низьких концентрацій кисню) або анаеробними, і магнітотаксис, по суті, спрямовує їх до ділянок з низьким вмістом кисню, розташованих нижче насиченої киснем поверхні [16, 17]. Різноманітність МТБ вражає. Вони демонструють широку морфологічну різноманітність, існуючи у формах сферичних клітин (коки), вигнутих паличок (вібріони), спіралеподібних

клітин (спірили) та прямих паличок [16]. Ще більш значущою є їхня філогенетична різноманітність: МТБ знаходяться серед представників різних таксономічних груп бактерій, включаючи *Pseudomonadota*, а також *Desulfobacterota* та *Nitrospirota* [18]. Ця широка філогенетична поширеність свідчить про можливий горизонтальний перенос генів, відповідальних за синтез магнітосом [18]. Крім того, МТБ відрізняються за типом магнітного матеріалу, який вони синтезують: більшість синтезують магнетит (Fe_3O_4), мінерал із залізом в ступені окиснення +2 та +3, тоді як деякі штами синтезують грейгіт (Fe_3S_4), сірковмісний магнітний мінерал [18].

1.3.2 Біогенні магнітні наночастинки

Центральним елементом магнітотаксису є магнітосоми. Це високоорганізовані внутрішньоклітинні органели, які складаються з магнітного кристала, оточеного тонкою ліпідно-білковою мембраною, відомою як мембрана магнітосоми [16]. Ця мембрана є не просто оболонкою, а активним компонентом, що відіграє ключову роль у біомінералізації, забезпечуючи контроль над транспортом іонів заліза, нуклеацією та ростом кристала [16, 17]. Особливістю магнітосом у більшості видів МТБ є їхнє розташування у клітині. Вони зазвичай утворюють ланцюжки вздовж цитоскелетних елементів, таких як білок MamK, що є гомологом актину [19]. Така лінійна організація магнітосом призводить до того, що сумарні магнітні моменти окремих частинок додаються, створюючи сильний загальний магнітний дипольний момент клітини, який дозволяє їй орієнтуватися у слабкому геомагнітному полі [16].

Біогенні магнітосоми пропонують низку значних переваг, що роблять їх привабливою альтернативою хімічно синтезованим магнітним наночастинкам для різних біомедичних застосувань, включаючи магнітогенетику та адресну доставку препаратів [1, 15]. По-перше, їхній синтез відбувається під надзвичайно точним біологічним та генетичним контролем, що призводить до отримання МНЧ з високою кристалічністю та хімічною чистотою, на відміну від часто дефектних або гетерогенних кристалів, отриманих хімічним шляхом [15]. По-друге, бактерії синтезують магнітосоми з надзвичайно вузьким розподілом за розміром – розмір

частинок є характерним для кожного виду бактерії та варіює в дуже малих межах, що важко досягти хімічними методами без додаткових етапів відбору [15]. По-третє, магнітосоми одного виду мають однорідну та контрольовану форму (наприклад, призматична, кубо-октаедрична, форма кулі чи вигнутої кулі) [18], що також є результатом генетично контрольованої біомінералізації і суттєво впливає на їхні магнітні властивості та взаємодію з біологічними системами. Крім того, природна біосумісність є важливою перевагою; мембрана, що оточує магнітосому, не тільки забезпечує її стабільність у водних середовищах та запобігає агрегації, але й містить білки, які можуть бути легко модифіковані або використані для функціоналізації поверхні з метою спрямування до певних клітин або приєднання біомолекул [15]. Важливо також, що синтез магнітосом відбувається за м'яких фізіологічних умов (кімнатна температура, атмосферний тиск), що є більш екологічною та менш ресурсомісткою альтернативою багатьом високотемпературним або високотисковим хімічним методам [16]. Нарешті, завдяки високій кристалічній якості та унікальній організації у ланцюжки, магнітосоми мають оптимальні магнітні властивості (наприклад, високу намагніченість насичення, контрольовані коерцитивність та температуру блокування), що робить їх високоефективними для багатьох біомедичних застосувань [20].

1.3.3 Виклики культивування магнітотаксисних бактерій

Не дивлячись на значні переваги біогенних магнітосом, їх отримання у достатній кількості для широких досліджень та потенційного застосування стикається з певними викликами, пов'язаними з культивуванням самих магнітотаксисних бактерій [16]. Більшість штамів МТБ є вибагливими до умов росту. Вони часто потребують специфічних анаеробних або мікроаерофільних умов, суворого контролю рівня кисню, оскільки надлишок кисню може бути токсичним або інгібувати синтез магнітосом [21]. Крім того, вони потребують певних джерел вуглецю, азоту, мікроелементів та іноді специфічних факторів росту, які можуть бути відсутні у стандартних лабораторних середовищах [21].

МТБ також часто чутливі до світла, особливо анаеробні штами. Ще одним обмеженням є відносно низька швидкість росту багатьох штамів порівняно з широко використовуваними лабораторними бактеріями, такими як *Escherichia coli* [21]. Варіабельність виходу магнітосом є поширеною проблемою, оскільки кількість та якість синтезованих частинок значною мірою залежать від точного складу середовища, умов культивування та, особливо, від доступності заліза в певній формі та концентрації [21].

1.3.4 Досягнення у методах культивування МТБ

Незважаючи на виклики, значний прогрес був досягнутий у розробці методів для ефективного лабораторного культивування МТБ та оптимізації синтезу магнітосом [21]. Дослідники розробили оптимізовані поживні середовища, які враховують специфічні метаболічні потреби різних штамів МТБ, забезпечуючи належні джерела енергії, вуглецю, азоту та мікроелементів у потрібних концентраціях [21]. Використання біореакторів дозволило здійснювати точний контроль ключових параметрів культивування, таких як температура, рН, рівень розчиненого кисню (або його відсутність), а також забезпечувати рівномірне перемішування, що є критично важливим для гомогенізації середовища та підтримки специфічного газового складу [21]. Були розроблені специфічні стратегії контролю доступності заліза, оскільки його концентрація та хімічна форма безпосередньо впливають на кількість, розмір, форму та магнітні властивості синтезованих магнітосом [16, 21]. Також постійно ведуться роботи з ізоляції та дослідження нових штамів МТБ з різними характеристиками магнітосом та, що важливо, легших для культивування в лабораторних умовах [17, 18].

1.4 Вплив форми магнітних наночастинок на їх властивості

1.4.1 Вплив форми на фізичні властивості магнітних наночастинок

Форма магнітних наночастинок має прямий вплив на їхні магнітні властивості через явище магнітної анізотропії форми. Навіть якщо кристал матеріалу ізотропний за своєю кристалічною структурою, відхилення від

сферичної форми (наприклад, призми, куби, пластини) призводить до появи переважного напрямку легкого намагнічування вздовж довгої осі або певної площини. Це пов'язано з тим, що намагнічування вздовж такого напрямку мінімізує енергію розмагнічуючого поля [13].

Вплив анізотропії форми на магнітні властивості є суттєвим. Вона впливає на такі параметри, як коерцитивність (здатність чинити опір розмагнічуванню), залишкова намагніченість (намагніченість, що залишається після зняття зовнішнього поля) та температура блокування (для суперпарамагнітних частинок) [13]. Частинки з вираженою анізотропією форми, як правило, мають вищу коерцитивність та намагніченість порівняно зі сферичними частинками того ж об'єму та матеріалу. Це означає, що вони перемагнічуються в більших полях.

Взаємодія з зовнішнім магнітним полем також залежить від форми. Частинки з вираженою анізотропією форми ефективніше орієнтуються вздовж напрямку зовнішнього поля [13, 20]. Цей обертальний момент може бути використаний для механічної стимуляції клітинних мішеней у магнітогенетиці, спричиняючи фізичне переміщення, обертання або деформацію актуатора, пов'язаного з мішенню [20]. Крім того, форма може впливати на дисперсію та стабільність МНЧ у розчинах. Анізотропні частинки можуть мати іншу схильність до агрегації або потребувати іншого типу поверхневої модифікації для стабілізації [12].

Форма магнітних наночастинок відіграє не менш важливу роль у їхній взаємодії з біологічними системами. Одним з найбільш вивчених аспектів є поглинання клітинами (ендоцитоз). Численні дослідження показали, що форма МНЧ суттєво впливає на ефективність та механізм їхнього клітинного поглинання. Наприклад, призми, куби або диски можуть поглинатися клітинами інакше, ніж сферичні частинки того ж об'єму чи площі поверхні. Часто виявляється, що для поглинання анізотропних частинок потрібне більше часу або енергії, а їх орієнтація відносно клітинної мембрани під час взаємодії є критично важливою [22]. Це має пряме значення для магнітогенетики та доставки препаратів, оскільки визначає, скільки магнітних актуаторів потрапить всередину цільових клітин.

Розподіл у тканинах та фармакокінетика МНЧ в організмі також можуть залежати від їхньої форми. Форма може впливати на гідродинамічну поведінку частинок у кровотоці, їхню взаємодію з компонентами крові (наприклад, білками плазми), проникнення через біологічні бар'єри та накопичення у органах [22].

Нарешті, форма МНЧ може впливати на їхню токсичність та імунну відповідь. Різні форми можуть по-різному взаємодіяти з клітинними органелами, лізосомами, білками або імунними клітинами, потенційно викликаючи різну ступінь цитотоксичності, запалення або імунної відповіді [23].

1.4.2 Дослідження форми та морфології магнітних наночастинок

Для розуміння та контролю властивостей МНЧ, зокрема магнітосом, вкрай важливим є їх детальне структурне та морфологічне дослідження. Існує кілька ключових методів мікроскопії, які дозволяють візуалізувати форму на нанорозмірному рівні.

Трансмисійна електронна мікроскопія (ТЕМ) є ключовим та найбільш широко використовуваним методом для візуалізації форми, розміру, кристалічної структури та внутрішніх дефектів окремих магнітних наночастинок [12, 23]. ТЕМ використовує пучок електронів, що проходить через тонкий зразок, надаючи двовимірне проєкційне зображення з високою роздільною здатністю (до ангстрем). Це дозволяє безпосередньо спостерігати морфологію частинок, вимірювати їхні лінійні розміри та отримувати статистичну інформацію про розподіл за розміром та формою у популяції.

Скануюча електронна мікроскопія (СЕМ) також використовує електронний пучок, але візуалізує поверхню зразка, аналізуючи вторинні або відбиті електрони. СЕМ корисна для вивчення поверхні частинок, їхнього розподілу та агрегатів, надаючи тривимірне уявлення про морфологію зразка в цілому [12]. Однак, порівняно з ТЕМ, СЕМ зазвичай має меншу роздільну здатність для візуалізації форми окремих наночастинок, особливо якщо вони знаходяться близько одна до одної.

Атомно-силовою мікроскопією (АСМ) використовується надзвичайно тонкий зонд, що сканує поверхню зразка, вимірюючи сили взаємодії між зондом і поверхнею. АСМ може надавати високоякісну топографічну інформацію про форму та розмір частинок, особливо якщо вони розташовані на плоскій підкладці, даючи уявлення про їхню висоту та латеральні розміри [24]. Крім того, АСМ дозволяє досліджувати механічні властивості матеріалів на нанорівні.

Магнітна силова мікроскопія (МСМ) є варіантом скануючої зондової мікроскопії, де накінчивник зонда виготовлений із феромагнітного матеріалу, покритий магнітним матеріалом. МСМ сканує поверхню зразка і вимірює магнітну взаємодію між магнітним зондом та магнітним полем зразка. Важливо зазначити, що МСМ досліджує не саму геометричну форму частинки безпосередньо, а розподіл магнітного поля над її поверхнею або магнітні взаємодії [25]. Однак, магнітне зображення, отримане за допомогою МСМ, відображає розподіл намагніченості, який залежить від форми частинки, її розміру, матеріалу та магнітного стану (наприклад, кількості магнітних доменів). Таким чином, МСМ є корисним інструментом для розуміння магнітної поведінки частинок, яка опосередковано пов'язана з їхньою формою, а також для вивчення магнітних доменів або взаємодій між частинками [25].

1.5 Генетичний апарат синтезу біогенних магнітних наночастинок

1.5.1 Магнітосомний острівець

Ключовим генетичним елементом, відповідальним за синтез магнітосом у МТБ, є так званий магнітосомний острівець. Він визначається як кластер генів, що кодують більшість білків, необхідних для всіх етапів утворення магнітосом: від захоплення заліза та формування внутрішньоклітинної везикули до нуклеації, росту кристала та його організації у клітині [26, 27]. Локалізація магнітосомного острівця в геномі МТБ часто вказує на його особливе походження. Дослідження геномів різних видів МТБ показали, що магнітосомний острівець нерідко є ділянкою, яка має ознаки горизонтального переносу генів [16, 18]. Це означає, що цей кластер генів міг бути набутий МТБ від інших організмів у процесі еволюції,

що пояснює присутність складного магнітосомального апарату у філогенетично віддалених таксономічних групах бактерій, які в іншому випадку не пов'язані між собою [18]. Таке походження підкреслює функціональну цілісність магнітосомного острівця як набору генів, оптимізованого для реалізації процесу біомінералізації.

1.5.2 Основні групи генів у складі магнітосомного острівця та їх функції

Магнітосомний острівець містить значну кількість генів (від ~20 до >50 залежно від виду МТБ), які часто організовані у вигляді декількох оперонів (наприклад, MamAB, MamXY, Mms6, FeoAB1). Ці гени кодують білки, що виконують різноманітні функції на різних етапах магнітосомогенезу як показано у табл. 1 та на рис. 2 [19].

Таблиця 1 – Функції білків, пов'язаних із магнітосомогенезом [19]

	Інвагінація мембрани	Об'єднання білків	Транспорт заліза	Окисл./ віднов. заліза	Нуклеація кристалу	Ріст кристалу	Збірка ланцюга
MamB	+		+				
MamM	+		+		+		
MamL	+						
MamI	+						
MamQ	+						
MamY	+						+
MamU	+						
Mms16	+						
MamE	+			+	+		
MamC	+						
MamF	+						
MamA		+					
MamH			+				
MamZ			+	+			
MamT			+	+			
MamP				+			
MamX				+			
MamO					+		
Mms5						+	
Mms6						+	
MamR						+	
MamD						+	
MamJ							+
MamK							+

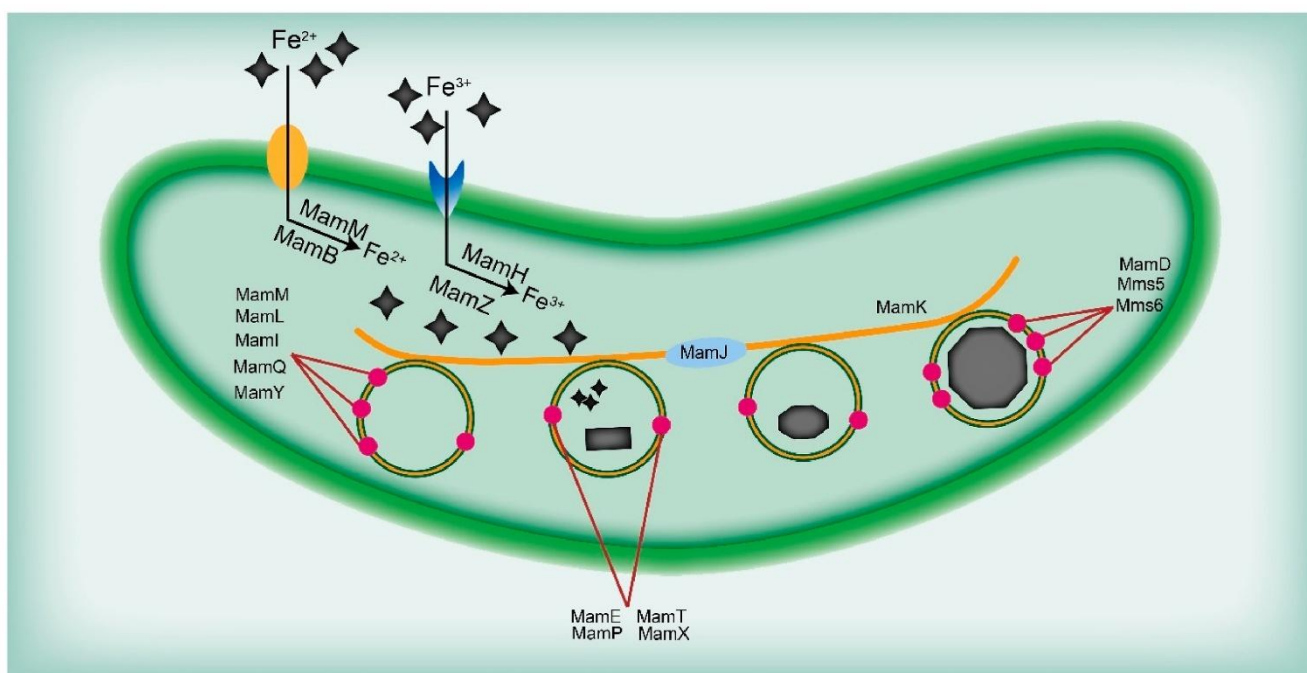


Рисунок 2 – Графічне представлення ролі білків, що пов'язані із утворенням магнітосом [26]

На початкових етапах формування магнітосом цитоплазматична мембрана бактеріальної клітини впинається всередину, утворюючи везикули, які стануть компартментами для росту магнітних кристалів [26]. Цей процес вимагає злагодженої роботи багатьох білків, більшість з яких є компонентами майбутньої мембрани магнітосоми. Згідно з табл. 1 та рис. 2, до цієї стадії залучені численні білки: MamB, MamM, MamL, MamI, MamQ, MamY, MamU, Mms16, MamE, MamC, MamF. Зокрема, MamB та MamM є основними інтегральними мембранними білками, що відіграють структурну роль у формуванні та стабілізації магнітосомної везикули [19]. Інші білки, як MamL, MamI, MamQ, MamY та білки з кластера Mms6, також локалізуються на мембрані і, ймовірно, беруть участь у регуляції процесу її інвагінації та початкового утворення везикул [19, 26].

Після формування везикули або одночасно з цим відбувається цілеспрямований транспорт та інтеграція специфічних білків магнітосомного острівця до мембрани везикули [19]. Центральну роль у цьому процесі відіграє білок MamA, який згідно з таблицею, відповідає за об'єднання білків. MamA є периферичним мембранним білком, що утворює олігомерний каркас на

цитоплазматичній стороні магнітосомної мембрани. Цей каркас слугує платформою для взаємодії з іншими білками, забезпечуючи їх правильну локалізацію та функціонування всередині або на поверхні магнітосоми [19].

Ефективне накопичення заліза всередині магнітосомної везикули є критично важливим для біомінералізації. На рисунку 2 показано транспорт як Fe^{2+} , так і Fe^{3+} . Таблиця 1 вказує на залучення білків MamB, MamM, MamH, MamZ, MamT до транспорту заліза. MamB та MamM, крім структурної ролі, також є передбачуваними транспортерами іонів заліза, які можуть переносити Fe^{2+} через магнітосомну мембрану, використовуючи протонний градієнт як рушійну силу [19]. Білки MamH та MamZ пов'язані з транспортом та перетвореннями Fe^{3+} [26]. MamT, як показано на схемі разом із MamP та MamX, також ідентифікований як потенційний Fe^{2+} транспортер до везикули, працюючи, можливо, разом з іншими білками для забезпечення достатнього потоку заліза для росту кристала [19].

Для синтезу магнетиту (Fe_3O_4), який містить як Fe^{2+} , так і Fe^{3+} , необхідний точний контроль окисно-відновного стану заліза всередині магнітосомної везикули. Табл. 1 та рис. 2 вказують на участь білків MamE, MamZ, MamT, MamP, MamX у цих процесах. MamE є ключовим білком з функцією ферооксидази, який каталізує окислення Fe^{2+} до Fe^{3+} , що є необхідним кроком для формування магнетиту [19]. MamP є мульти-мідною оксидазою, також залученою до процесів окислення заліза [19]. MamZ може діяти як фериредуктаза, перетворюючи Fe^{3+} назад у Fe^{2+} [19]. MamT і MamX, як зазначено, також пов'язані з цим етапом, можливо, впливаючи на доступність іонів заліза в потрібному стані окиснення [19, 26]. Цей комплекс білків забезпечує створення необхідних хімічних умов для біомінералізації магнетиту.

Процес утворення магнітосоми починається з нуклеації – формування початкових зародків кристала всередині везикули. Згідно з таблицею 1, білки MamM, MamE та MamO залучені до цього етапу. MamM та MamE можуть відігравати роль у створенні сприятливого мікросередовища для нуклеації або бути безпосередньо задіяними в ініціації формування кристалічної фази [19]. MamO є особливо важливим білком, який, як вважають, відіграє центральну роль у контролі

нуклеації та визначенні кількості кристалів, що формуються в кожній везикулі, а також впливає на їх початкову морфологію [19].

Після нуклеації кристалічні зародки починають зростати, приєднуючи іони заліза. Цей процес, особливо у випадку магнетиту, є висококонтрольованим білками, які взаємодіють з поверхнею кристала, визначаючи швидкість росту різних граней і, як наслідок, кінцеву форму частинки [16, 19]. Таблиця 1 вказує на участь білків Mms5, Mms6, MamR, MamD у рості кристала. Білки оперону Mms6 (переважно Mms5, Mms6) є білками магнітосомної мембрани, які специфічно зв'язуються з певними кристалічними гранями магнетиту, контролюючи їх ріст і таким чином визначаючи форму кристала (наприклад, сприяючи формуванню призматичних або кубо-октаедричних форм) [19].

На фінальному етапі формування магнітосом вони організовуються у лінійні структури – ланцюжки, які забезпечують сильний магнітний дипольний момент клітини [16, 19]. Згідно з табл. 1 та рис. 2, за цей процес відповідають білки MamY, MamJ та MamK. MamK є центральним компонентом системи, формуючи актин-подібний філамент, що простягається вздовж цитоплазми бактерії [19, 26]. MamJ є білком-адаптором, який зв'язує магнітосому везикулу з філаментом MamK, таким чином закріплюючи магнітосому вздовж цього каркасу і формуючи ланцюжок [26]. Білок MamY, також розташований на магнітосомній мембрані, ймовірно, теж бере участь у закріпленні магнітосом на філаменті MamK [26]. Таким чином, ці білки забезпечують правильне просторове розташування магнітосом всередині клітини.

РОЗДІЛ 2 МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

2.1 Матеріали дослідження

Для дослідження можливості передбачення форми біогенних магнітних наночастинок у магнітотаксисних бактеріях за допомогою методів машинного навчання було сформовано набір даних та застосовано відповідні алгоритми класифікації (Додаток А та Додаток Б).

Первинний набір даних був отриманий зі статті [18], яка містить інформацію про геномний склад 47 штамів МТБ та відповідні, експериментально визначені форми синтезованих ними біогенних магнітних наночастинок. Набір даних включав ознаки, що характеризують наявність або відсутність широкого спектра генів, асоційованих з процесами біомінералізації та формування магнітосом. До цих генів належать:

1. група генів *mam*:

- *mamA*,
- *mamB*,
- *mamE*,
- *mamI*,
- *mamK*,
- *mamM*,
- *mamP*,
- *mamQ*,
- *mamH*,
- *mamF*,
- *mamS*,
- *mamT*,
- *mamC*,
- *mamD*,
- *mamG*,
- *mamR*,
- *mamL*,
- *mamO*,
- *mamN*,
- *mamX*,
- *mamZ*,
- *mamY*,
- *mamJ*,
- *mamU*,
- *mamV*,
- *mamW*;

2. група генів *mms*:

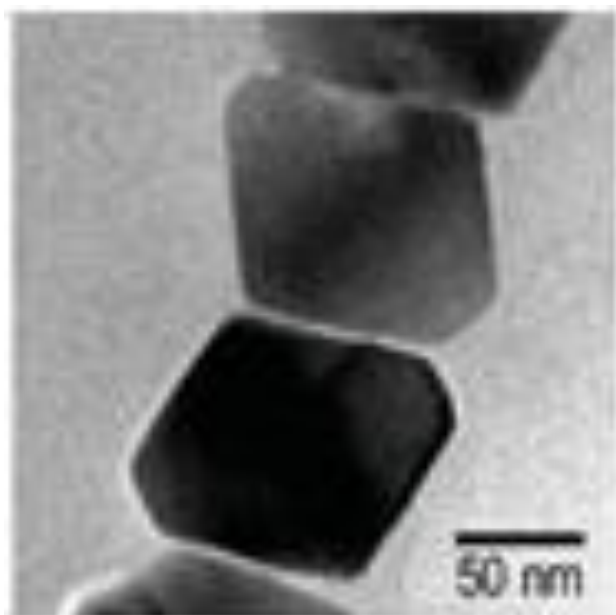
- *mms5*,
- *mms6*,
- *mms6-L*,
- *mms36*,
- *mms48*;

3. група генів *mad* (від *mad1* до *mad31*);

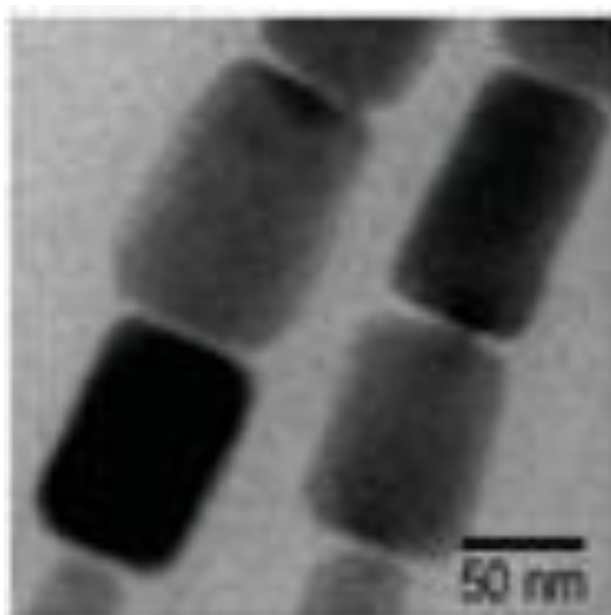
4. група генів *man* (від *man1* до *man6*).

Для кожного штаму МТБ у наборі даних ці геномні ознаки слугували вхідними змінними. Цільовою змінною для передбачення була форма БМН –

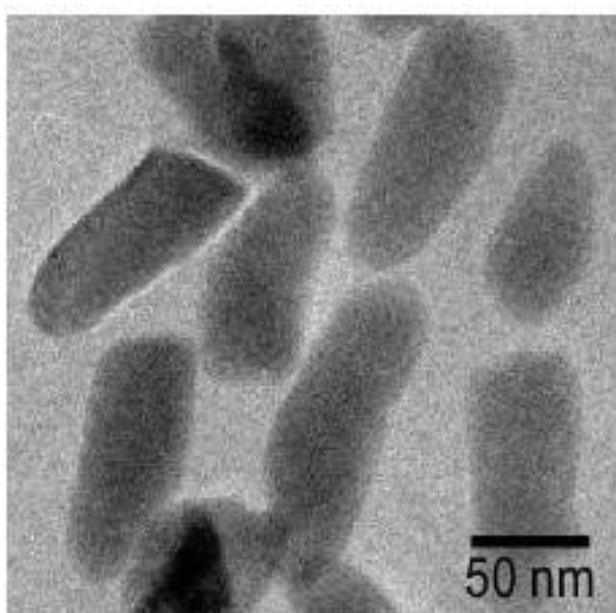
категоріальна змінна, що включає класи Bullet (куля), Curved bullet (вигнута куля), (Cubo)-octahedron (кубічна чи октаедрична) та Prism (призматична).



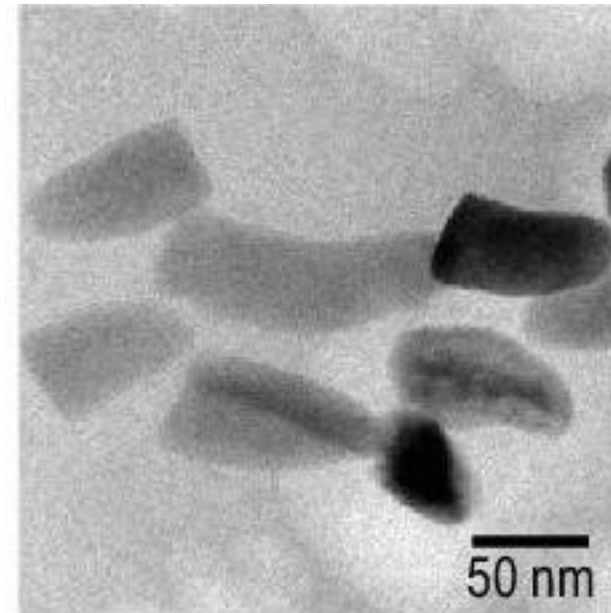
Кубічно-октаедрична форма



Призматична форма



Форма кулі



Форма вигнутої кулі

Рисунок 3 – Приклади форми біогенних магнітних наночастинок [18]

На додаток до геномних ознак, для підвищення інформативності даних було проведено пошук послідовностей генів *tamO* та *tamL* у базі даних GenBank, що підтримується Національним центром біотехнологічної інформації (NCBI). Оскільки білки, що кодуються цими генами, відіграють важливу роль у формуванні

магнітосом та наявні в усіх представників типу *Pseudomonadota*, для якого притаманні кубічно-октаедричні чи призматичні БМН, було вирішено включити їхні фізико-хімічні властивості як додаткові ознаки. Розрахунок цих властивостей здійснювався за допомогою модуля ProtParam, що входить до складу бібліотеки Biopython для мови програмування Python. Модуль ProtParam аналізує амінокислотну послідовність білка та надає низку інших характеристик. Зокрема, для даного дослідження були використані:

- Молекулярна вага – розраховується як сума атомних мас усіх атомів, що входять до складу амінокислотних залишків, з яких складається білок, за вирахуванням маси молекул води, які відщеплюються при утворенні пептидних зв'язків. ProtParam використовує стандартні середні маси амінокислотних залишків для цього розрахунку.
- Ароматичність – визначається як відносна кількість ароматичних амінокислот (фенілаланін, триптофан, тирозин) у складі білка. Цей показник може корелювати зі стабільністю білка. ProtParam обчислює його як відношення кількості ароматичних залишків до загальної кількості амінокислот у послідовності.
- Ізоелектрична точка (pI) – це значення рН, при якому сумарний електричний заряд білкової молекули дорівнює нулю. Розрахунок pI є ітеративним процесом, який враховує кислотно-основні властивості всіх іонізованих груп у білку – N-кінцевої та C-кінцевої груп, а також бічних ланцюгів іонізованих амінокислот (аспартат, глутамат, лізин, аргінін, гістидин, тирозин, цистеїн). ProtParam використовує набір відомих значень константи кислотної дисоціації pK_a для цих груп і застосовує ітеративний алгоритм для знаходження такого значення рН, при якому сума позитивних зарядів (від основних груп) дорівнює сумі негативних зарядів (від кислотних груп).

Окрім стандартних фізико-хімічних властивостей, для оцінки ступеня подібності амінокислотних послідовностей білків MamO та MamL між різними штамми МТБ та групами форм БМН, було введено показник відносної ваги вирівнювання. Цей показник базується на попарному вирівнюванні

послідовностей. Для його розрахунку використовувалася функціональність попарного вирівнювання (pairwise alignment) з бібліотеки Biopython, зокрема об'єкт PairwiseAligner.

Процедура розрахунку відносної ваги вирівнювання була реалізована за допомогою двох функцій. Перша функція, calculate_alignment_weight, приймала дві амінокислотні послідовності: референсну (reference_seq) та цільову (target_seq). Вона створювала об'єкт PairwiseAligner і виконувала попарне вирівнювання цільової послідовності відносно референсної. За замовчуванням, PairwiseAligner прагне знайти оптимальне вирівнювання, максимізуючи суму балів за відповідності (matches), невідповідності (mismatches) та пропуски (gaps) згідно з вбудованою або заданою ваговою матрицею. У цьому випадку, використовувалась стандартна матриця ваг, де рахується кількість збігів амінокислот:

- оцінка за співпадіння: 1;
- оцінка за неспівпадіння: 0;
- режим: глобальне вирівнювання.

Отримувався результат вирівнювання, і з нього вибиралося перше (або єдине) найкраще вирівнювання (best_alignment). Значення ваги вирівнювання (match_count) отримувалося з атрибуту .score цього об'єкта, що за стандартних налаштувань відповідало кількості ідентичних амінокислот у вирівняних позиціях. Для нормалізації цього значення та отримання показника, який відображає схожість відносно довжини референса, match_count ділився на довжину референсної послідовності (len(reference_seq)). Отримане дробове значення і становило відносну вагу вирівнювання.

Друга функція, get_alignment_weight_list, автоматизувала розрахунок цього показника для груп білків, що належать до різних класів форм БМН (одна група білків від штамів з кубічно-октаедричною формою, інша – з призматичною). Як глобальна референсна послідовність для всіх подальших вирівнювань була обрана перша послідовність з першої вхідної групи (funct_prot[0]), тобто це були послідовності білків *Paramagnetospirillum magneticum* АМВ-1. Далі, у циклах, для кожної послідовності в обох вхідних списках (funct_prot та funct_prot1),

викликалася функція `calculate_alignment_weight`, використовуючи фіксовану референсну послідовність. Результати зберігались у двох окремих списках, по одному для кожної групи. Важливо відзначити, що відносні ваги вирівнювання розраховувалися для всіх послідовностей, крім самої референсної послідовності з першої групи (що відображено у масиві ваг вирівнювання `alignment_weights[1:]`). Це дозволило отримати числові ознаки, що кількісно описують схожість кожного білка у вибірці з обраним референсом, що потенційно може бути пов'язано з відмінностями у структурі та функції білка, а отже – у формі БМН.

Код для розрахунку відносної ваги вирівнювання мав вигляд:

```
def calculate_alignment_weight(reference_seq: str, target_seq: str) -> float:
```

```
    aligner = PairwiseAligner()
```

```
    best_alignment = aligner.align(reference_seq, target_seq)[0]
```

```
    match_count = best_alignment.score
```

```
    alignment_length = len(reference_seq)
```

```
    return match_count / alignment_length
```

```
def get_alignment_weight_list(func_prot: list, func_prot1: list) -> tuple[list, list]:
```

```
    reference_sequence = func_prot[0]
```

```
    alignment_weights = []
```

```
    for seq in func_prot:
```

```
        weight = calculate_alignment_weight(reference_sequence, seq)
```

```
        alignment_weights.append(weight)
```

```
    alignment_weights1 = []
```

```
    for seq in func_prot1:
```

```
        weight = calculate_alignment_weight(reference_sequence, seq)
```

```
        alignment_weights1.append(weight)
```

```
    return alignment_weights[1:], alignment_weights1
```

Таким чином, кожен зразок у фінальному наборі даних був представлений бінарним вектором наявності генів та числовими значеннями відносної ваги вирівнювання, молекулярної ваги, ароматичності та ізоелектричної точки білків MamO і MamL, а цільовим класом була відповідна форма БМН. Ці дані було записано у Excel-таблицю (таблиця 2) за допомогою пакету pandas для мови програмування Python.

Таблиця 2 – Основні дані, що зібрані для навчання моделей машинного навчання

Штам	MACG	MASGP	mamL	mamO	mad	man	Ізоелектрич на точка MamL	Ізоелектрич на точка MamO	Відн. вага вирівнювання MamL	Відн. вага вирівнювання MamO	Молекулярна вага MamL	Молекулярна вага MamO	Ароматичність MamL	Ароматичність MamO	Форма БМН
AMB-1	1	1	1	1	0	0	12,00	6,68	1	1	8 360,16	66 342,56	0,06	0,07	(Cubo)-octahedron
XM-1	1	1	1	1	0	0	12,00	6,72	0,91	0,88	8 520,44	65 497,73	0,08	0,07	(Cubo)-octahedron
MS-1	1	1	1	1	0	0	10,03	7,28	0,50	0,49	9 093,99	66 757,17	0,07	0,08	(Cubo)-octahedron
ME-1	1	1	1	1	0	0	5,60	9,64	0,40	0,49	7 011,06	70 448,49	0,15	0,08	(Cubo)-octahedron
SO-1	1	1	1	1	0	0	12,00	8,08	0,90	0,59	8 330,18	40 038,60	0,06	0,06	(Cubo)-octahedron
SP-1	1	1	1	1	0	0	4,97	9,55	0,50	0,49	11 121,88	71 273,14	0,14	0,09	(Cubo)-octahedron
BB-1	1	1	1	1	0	0	12,00	6,68	1	1,00	8 360,16	66 282,51	0,06	0,07	(Cubo)-octahedron
MSR-1	1	1	1	1	0	0	6,54	9,87	0,45	0,51	8 248,74	70 163,17	0,15	0,09	(Cubo)-octahedron
LBB-42	1	1	1	1	0	0	12,00	6,50	0,97	1,00	8 364,10	66 268,44	0,06	0,07	(Cubo)-octahedron
YQV-1	1	1	1	1	0	0	12,00	6,46	0,86	0,88	8 585,54	65 379,53	0,09	0,07	(Cubo)-octahedron
WYHS-3	1	1	1	1	0	0	12,00	6,68	1	1,00	8 360,16	66 272,47	0,06	0,07	(Cubo)-octahedron
MO-1	1	1	1	1	0	0	12,00	6,12	0,83	0,91	8 411,17	60 421,32	0,06	0,07	(Cubo)-octahedron
MC-1	1	1	1	1	0	0	6,53	9,26	0,44	0,49	8 173,68	70 738,53	0,14	0,07	(Cubo)-octahedron
IT-1	1	1	1	1	0	0	12,00	6,50	1	1,00	8 360,16	66 298,46	0,06	0,07	(Cubo)-octahedron
BW-2	1	1	1	1	0	0	11,72	6,46	0,51	0,62	8 576,39	73 270,21	0,13	0,08	(Cubo)-octahedron
MV-1	1	1	1	1	0	0	5,54	9,65	0,45	0,50	7 623,00	79 037,91	0,13	0,08	Prism
PR-1	1	1	1	1	0	0	5,05	9,45	0,47	0,49	8 100,67	69 405,00	0,13	0,10	Prism
SH-1	1	1	1	1	0	0	9,95	9,39	0,46	0,33	8 341,97	43 112,17	0,09	0,06	Prism
QH-2	1	1	1	1	0	0	10,61	8,40	0,53	0,68	9 648,58	67 284,41	0,11	0,08	Prism
LM-1	1	1	1	1	0	0	10,87	5,90	0,47	0,57	8 576,28	74 205,80	0,09	0,07	Prism
WYHS-4	1	1	1	1	0	0	10,24	9,72	0,46	0,51	8 327,98	72 307,22	0,09	0,08	Prism
DMHC-1	1	1	1	1	0	0	9,78	7,28	0,41	0,52	8 547,55	69 339,09	0,12	0,09	Prism
DMHC-6	1	1	1	1	0	0	10,90	6,44	0,50	0,58	9 184,99	71 507,18	0,08	0,09	Prism
THC-1	1	1	1	1	0	0	9,78	7,28	0,41	0,52	8 547,55	69 339,09	0,12	0,09	Prism
YQC-9	1	1	1	1	0	0	11,33	9,25	0,49	0,49	9 087,82	68 904,69	0,07	0,09	Prism
UR-1	1	1	1	1	0	0	11,33	6,63	0,49	0,51	9 087,82	67 226,23	0,07	0,07	Prism
YQC-3	1	1	1	1	0	0	6,53	9,75	0,46	0,49	7 947,38	67 598,62	0,15	0,08	Prism
YQC-5	1	1	1	1	0	0	11,35	9,55	0,41	0,50	8 033,70	71 112,32	0,08	0,09	Prism
XQGC-1	1	1	1	1	0	0	10,92	8,39	0,50	0,56	8 959,87	68 731,24	0,07	0,08	Prism
MYC-9	1	1	1	1	0	0	10,70	10,30	0,47	0,27	8 370,02	28 001,20	0,09	0,10	Prism
DMHC-8	1	1	1	1	0	0	10,70	9,90	0,47	0,46	8 370,02	56 076,15	0,09	0,07	Prism
SHHR-1	1	1	1	1	0	0	6,71	9,62	0,42	0,49	8 069,61	69 990,12	0,11	0,09	Prism
SS-5	1	1	1	1	0	0	11,72	9,32	0,49	0,51	9 938,87	71 086,62	0,04	0,09	Prism

Продовження таблиці 2 – Основні дані, що зібрані для навчання моделей машинного навчання

Штам	MACG	MACGP	mamL	mamO	mad	map	Ізоелектрична точка MamL	Ізоелектрична точка MamO	Відн. вага вирівнювання MamL	Відн. вага вирівнювання MamO	Молекулярна вага MamL	Молекулярна вага MamO	Ароматичність MamL	Ароматичність MamO	Форма БМН
RS-1	1	0	1	0	1	0									Bullet
FSS-1	1	0	1	0	1	0									Bullet
HK-1	1	0	1	0	1	0									Bullet
BW-1	1	0	1	0	1	0									Bullet
TM-1	1	0	0	1	1	1									Curved bullet
MYR-1	1	0	0	1	1	1									Curved bullet
XYR	1	0	0	1	1	1									Curved bullet
YQR-1	1	0	0	1	1	1									Curved bullet
LBB01	1	0	0	1	1	1									Curved bullet
MYC-10	1	0	0	1	1	1									Curved bullet
LBB02	1	0	0	1	1	1									Curved bullet
XYC	1	0	0	1	1	1									Curved bullet
WYHC-5	1	0	0	1	1	1									Curved bullet
CS-04	1	0	0	0	1	0									Curved bullet

У цій таблиці гени *mamA*, *mamB*, *mamE*, *mamI*, *mamK*, *mamM*, *mamP* та *mamQ* об'єднані у групу MACG (magnetosome core genes), оскільки наявні у всіх магнітотаксисних бактерій незалежно від їх форми біогенних магнітних наночастинок. Гени *mamH*, *mamF*, *mamS* та *mamT* притаманні лише типу *Pseudomonadota* об'єднані у групу MACGP [18].

2.2 Методи дослідження

Для розв'язання задачі класифікації форми БМН були використані два алгоритми машинного навчання: дерево рішень (Decision Tree) та випадковий ліс (Random Forest), реалізовані за допомогою бібліотеки scikit-learn в Python [28].

2.2.1 Дерево рішень

Дерево рішень є одним із фундаментальних алгоритмів машинного навчання, що використовується як для задач класифікації, так і для регресії [29]. Його структура нагадує блок-схему, де кожен внутрішній вузол представляє перевірку певного атрибута (ознаки), кожен внутрішній вузол (гілка) – результат цієї перевірки, а кожен листовий вузол (термінальний вузол) – рішення (клас або

числове значення). Значною перевагою дерева рішень є можливість представлення його внутрішньої будови (рисунок 3) та відтворення процедури пошуку рішення, що рідко зустрічається у складних моделях машинного навчання.

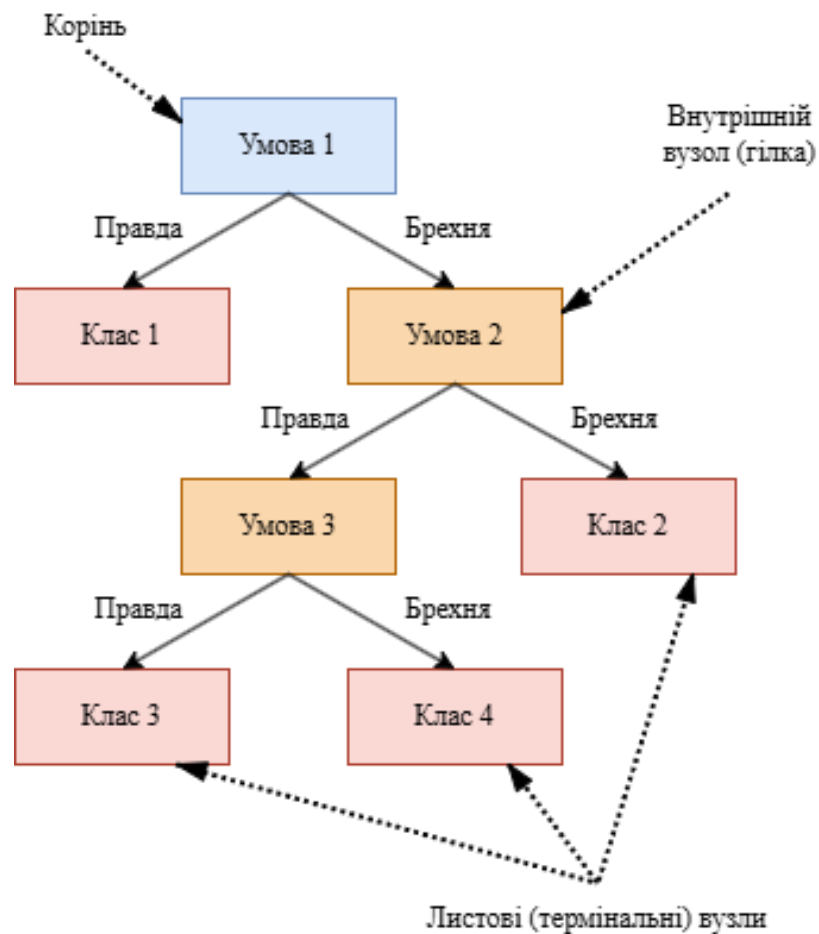


Рисунок 3 – Графічне представлення дерева рішень

Процес побудови дерева полягає у рекурсивному розбитті навчальної вибірки на підмножини на основі значень дескрипторів. Ключовим аспектом є вибір атрибута для розбиття на кожному кроці. У scikit-learn для цього використовується індекс Джині. Індекс Джині – це метрика, що використовується в алгоритмах побудови дерев рішень для оцінки ступеня неоднорідності класів у вибірці даних, тобто у вузлі дерева [29]:

$$Gini = 1 - \sum_{i=1}^c p_{(i|t)}^2, \quad (1)$$

де $p(i|t)$ – відносна частота класу i у вузлі t ; c – кількість класів.

Алгоритм прагне мінімізувати цей індекс, обираючи атрибут, що дає найбільше зменшення невизначеності.

2.2.2 Випадковий ліс

Метод випадкового лісу є ансамблевим алгоритмом машинного навчання, який значно покращує результати в порівнянні з поодинокими деревами рішень. Його суть полягає у створенні сотень незалежних дерев рішень під час тренування. Кожне таке дерево навчається на випадковій підвибірці вихідних даних, отриманій методом бутстрепа (вибірка з поверненням). Додаткова випадковість вноситься на етапі побудови вузлів кожного дерева: замість того, щоб розглядати всі доступні ознаки для знаходження оптимального розділення, алгоритм обирає лише випадкову підмножину ознак. Цей подвійний підхід до внесення випадковості – як на рівні даних, так і на рівні ознак – є ключовим для зменшення кореляції між окремими деревами, що, в свою чергу, підвищує загальну точність та стійкість моделі до перенавчання [29].

В бібліотеці scikit-learn для мови програмування Python випадковий ліс реалізований у класі RandomForestClassifier для задач класифікації. При створенні моделі одним з найважливіших параметрів є кількість дерев рішень у випадковому лісі. Збільшення цього числа зазвичай покращує якість моделі, але до певної межі, після якої зростання стає незначним, а обчислювальні витрати збільшуються [29].

У задачах класифікації кожне дерево в лісі «голосує» за певний клас, і остаточним прогнозом стає клас, що набрав найбільшу кількість голосів, як це візуалізовано на рисунку 4.

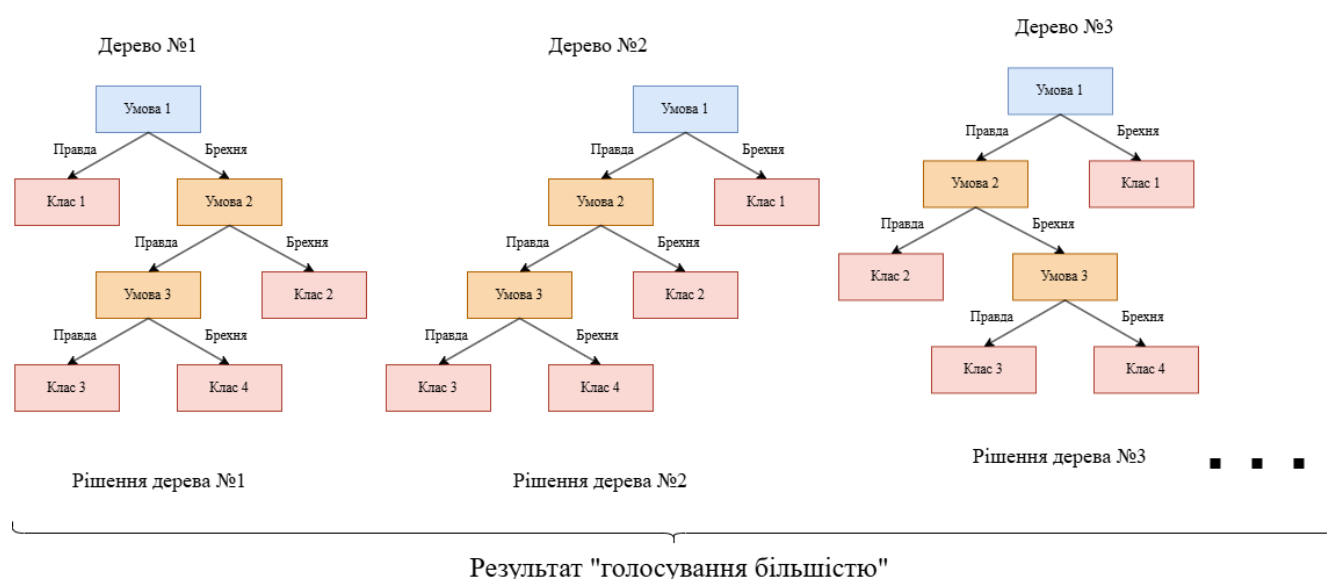


Рисунок 4 – Графічне представлення логіки роботи алгоритму випадкового лісу

Перевагами випадкового лісу є його висока точність, стійкість до перенавчання, здатність обробляти великі обсяги даних з великою кількістю ознак та можливість оцінки важливості цих ознак. Він також добре працює з різними типами даних і не вимагає ретельного масштабування ознак. Однак, модель може бути складнішою для інтерпретації порівняно з одним деревом рішень, а навчання великої кількості дерев може бути обчислювально складним. Попри це, випадковий ліс залишається одним з найпопулярніших та ефективних методів у сучасному машинному навчанні.

2.2.3 Параметри тренування моделей

Спочатку вхідні дані, що склалися з матриці дескрипторів X та цільового вектора y , були розділені на тренувальний та тестувальний набори. Для цього використовувалася функція `train_test_split` з такими ключовими параметрами:

```
X_train, X_test, y_train, y_test = train_test_split(  
    X,  
    y,  
    test_size=0.2,  
    random_state=seed  
)
```

Параметр `test_size=0.2` означав, що 20% від усіх даних було виділено для тестувального набору, а решта 80% склали тренувальний набір. Параметр `random_state=seed` був використаний для забезпечення відтворюваності самого процесу розділення даних. У даному випадку `random_state=0`.

Після підготовки даних було створено об'єкти класів `DecisionTreeClassifier` та `RandomForestClassifier`. Для забезпечення відтворюваності результатів цих моделей при їх створенні було явно встановлено параметр `seed` аналогічно попередньому етапу поділу даних на два набори.

Це гарантує, що навіть при наявності стохастичних елементів у процесі побудови дерева (наприклад, при виборі ознак, якщо декілька з них дають однаковий приріст інформації), результат буде однаковим при кожному запуску.

Всі інші гіперпараметри `DecisionTreeClassifier` та `RandomForestClassifier`, такі як критерій розбиття (`criterion`), максимальна глибина дерева (`max_depth`) або мінімальна кількість зразків для розбиття вузла (`min_samples_split`), були залишені зі своїми значеннями за замовчуванням.

Після ініціалізації з вказаними параметрами обидві моделі, `DecisionTreeClassifier` та `RandomForestClassifier`, навчалися виключно на тренувальному наборі даних (`X_train`, `y_train`), який, як зазначалося, складав 80% від початкового набору даних і містив інформацію про 37 штамів.

2.2.4 Визначення точності моделей

Для отримання більш об'єктивної та узагальненої оцінки продуктивності розроблених класифікаційних моделей, а також для оцінки їхньої стійкості до різних розподілів даних, було застосовано процедуру перехресної перевірки. Зокрема, було обрано метод `ShuffleSplit`, із бібліотеки `scikit-learn` мови програмування Python.

Вибір методу `ShuffleSplit` обумовлений його гнучкістю. На відміну від класичної K-Fold перехресної перевірки, `ShuffleSplit` випадковим чином перемішує дані перед кожним розділенням та дозволяє незалежно контролювати як кількість таких розділень, так і частку даних, що відводиться під тестовий набір. Цей підхід є особливо корисним при роботі з великими наборами даних або коли потрібно провести значну кількість ітерацій для оцінки стабільності [30].

У рамках кожного окремого проходу процедури `ShuffleSplit`, дані були випадковим чином перемішані, а потім поділені на навчальну та тестову вибірки (рисунки 5). Було виконано 3 таких розділення. Для кожного розділення, 20% від загального обсягу даних використовувалися як тестовий набір, а решта 80% – як навчальний. Така конфігурація дозволила отримати три незалежні оцінки продуктивності для кожної моделі на різних підмножинах даних [30].

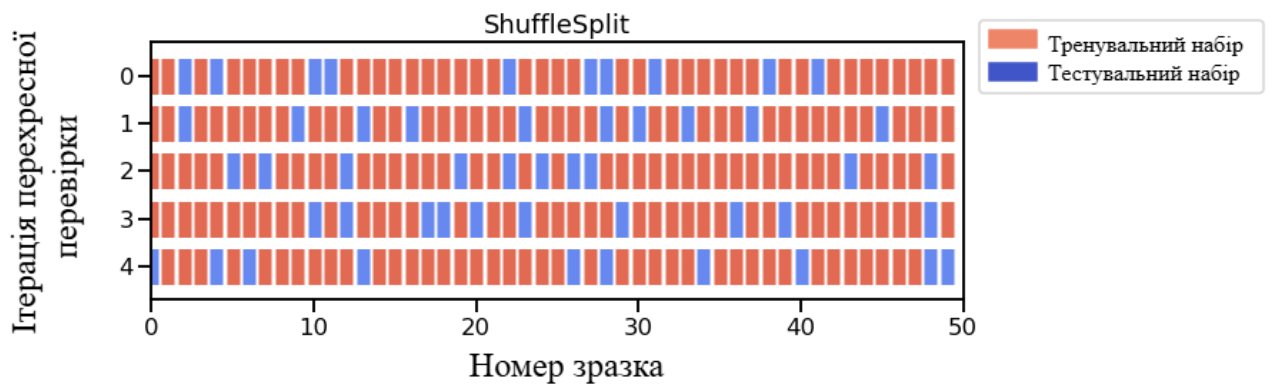


Рисунок 5 – Графічне представлення методу ShuffleSplit із бібліотеки scikit-learn мови програмування Python [30]

Ключовим аспектом дослідження стабільності було повторення всієї процедури перехресної перевірки 100 разів. Кожне з цих 100 повторень ініціювалося з різним початковим значенням `seed` для генератора псевдовипадкових чисел. Це означає, що на кожній з 100 ітерацій дані перемішувалися та поділялися інакше, ніж на попередніх ітераціях. Такий підхід дозволив дослідити, наскільки сильно змінюється оцінка продуктивності моделі залежно від конкретного випадкового поділу даних на навчальну та тестову вибірки. Фактично, для кожної моделі було отримано 300 окремих оцінок точності (100 ітерацій по 3 розділення на ітерацію).

Як метрика оцінки ефективності класифікаційних моделей використовувалася точність (accuracy). Ця метрика визначає частку правильно класифікованих екземплярів у тестовому наборі. Точність є інтуїтивно зрозумілою метрикою і широко застосовується для задач класифікації, особливо якщо класи є відносно збалансованими.

Для кількісного узагальнення та аналізу отриманих результатів по 100 ітераціях перехресної перевірки (тобто, по 300 оцінках точності для кожної моделі), були розраховані:

- Середні значення (`mean`) отриманих оцінок точності. Це дає уявлення про типову або очікувану продуктивність моделі.

- Стандартне відхилення (standard deviation) отриманих оцінок точності. Цей показник є мірою розсіювання оцінок. Менше стандартне відхилення свідчить про більш стабільну роботу моделі, яка менш чутлива до конкретного вибору навчальних та тестових даних, отриманих після випадкового перетасування та поділу.

Для наочного представлення та порівняння розподілу отриманих 300 оцінок точності для кожної з оцінюваних моделей була виконана візуалізація результатів. Розподіл оцінок точності візуалізовано за допомогою діаграм, побудованих з використанням бібліотеки `seaborn` мови програмування Python. Це дозволило наочно порівняти не тільки середню продуктивність, але й варіативність та стабільність результатів для обох алгоритмів.

2.2.5 Зовнішня валідація

Для перевірки узагальнюючої здатності натренованих моделей на даних, які повністю були відсутні у вихідному наборі, було проведено пошук інформації про штами МТБ, не включені до початкової вибірки зі статті [18]. В результаті було обрано такі штами:

1. *Magnetococcus sp.* PR-3
2. *Candidatus Magnetominusculus xianensis* HCH-1
3. *Candidatus Desulfarcum epimagneticum* CR-1
4. *Candidatus Deltaproteobacteria* FH-1
5. *Magnetospirillum sp.* SS-4
6. *Magnetospirillum sp.* UT-4
7. *Magnetospirillum sp.* LM-5
8. *Gamma proteobacteria bacterium* CCP3-SC1AL1
9. *Alpha proteobacterium* LM-2
10. *Alpha proteobacterium* LEMS
11. *Alpha proteobacterium* KR-1
12. *Alpha proteobacterium* CB-1

Для автоматизації процесу передбачення форми біогенних магнітних наночастинок для нових штамів було написано Python скрипт, який збирає всі необхідні дескриптори для математичних моделей лише за назвою штаму чи будь-яким іншим унікальним ідентифікатором. Його схему подано на рисунку 6, а сам код подано у додатку В.

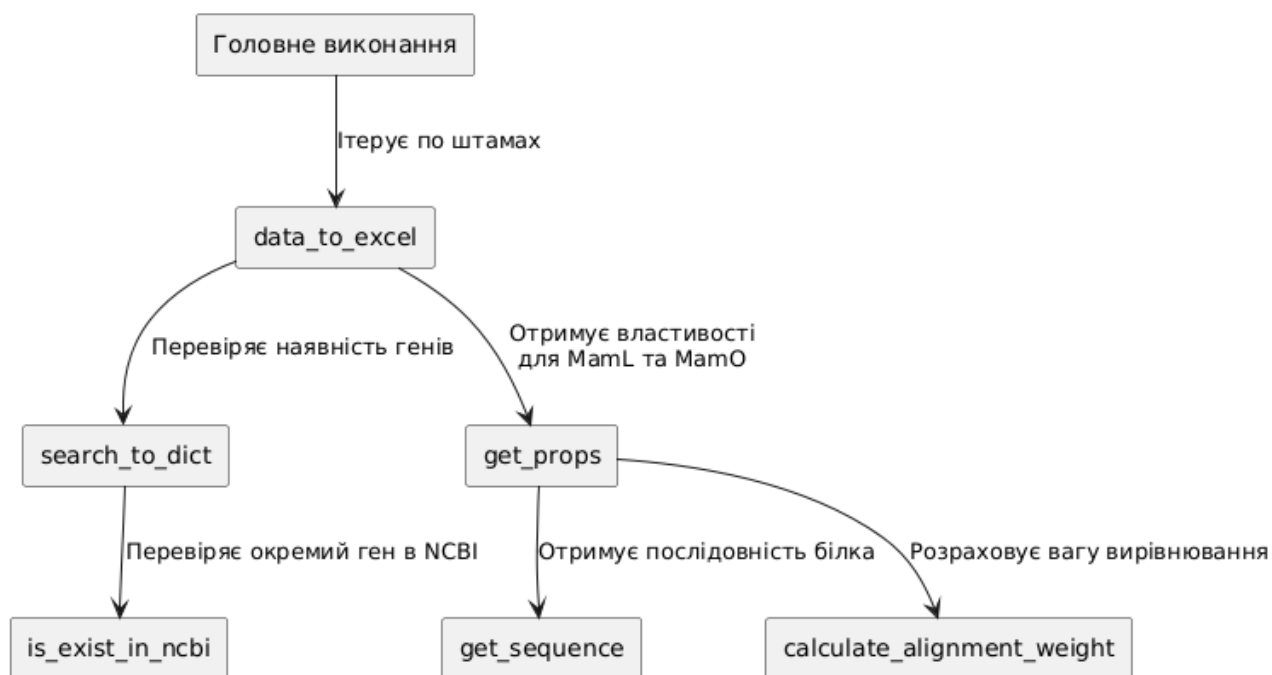


Рисунок 6 – Схема коду для автоматизованого отримання дескрипторів

Для цих штамів були зібрані відповідні вхідні дані (наявність генів та властивості білків MamO та MamL, якщо доступно), і натреновані моделі були використані для передбачення форми їхніх БМН. Отримані передбачення порівнювалися з літературними даними щодо форми біогенних магнітних наночастинок цих штамів.

РОЗДІЛ 3 РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

3.1 Тренування моделей машинного навчання

3.1.1 Результати тренування моделей

Як було вказано у розділі «Матеріали і методи», для тренування моделі дерева рішень використовувались лише 80% даних, тобто дані про 37 штамів. Код написаний для цього поданий у додатку А. Результат побудови дерева рішень візуалізовано на рисунку 7.

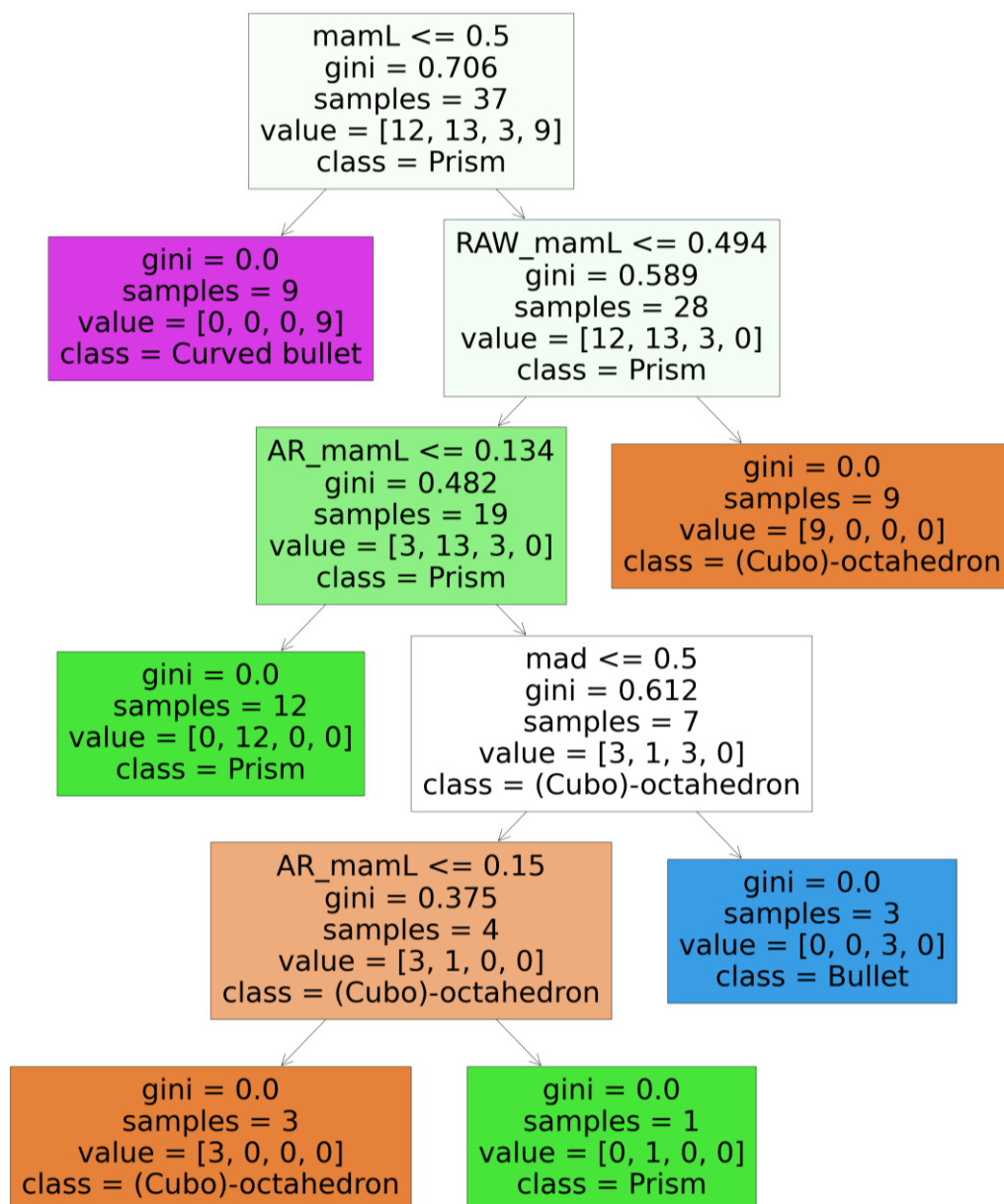


Рисунок 7 – Графічне представлення отриманого дерева рішень

Помітно, що ця модель не має визначених дескрипторів для оптимального розділення МТБ на класи відповідно до форми їх магнітних наночастинок. Модель

розділяє класи БМН у формі кулі та вигнутої кулі за унікальними для них властивостями. Для форми вигнутої кулі таким показником є відсутність гена, що кодує білок MamL. Оскільки після цього етапу зразки з класу вигнутої кулі відкинуті з набору даних, то для виділення класу МТБ, які мають БМН у формі кулі, використовується ознака наявності білків групи Mad.

Класи октаедричних та призматичних БМН неможливо розділити за однією ознакою, але саме вони найбільш важливі для досліджень магнітогенетики. Призматичні мають найбільший дипольний момент, якщо діяти вздовж їх довшої осі, що дозволяє більш точно контролювати клітинні мішені, що зв'язані із такими магнітними актуаторами.

На відміну від дерева рішень, модель випадкового лісу неможливо візуалізувати як єдину цілісну структуру, оскільки вона являє собою ансамбль, що складається з численних (часто сотень або тисяч) індивідуальних дерев рішень. У нашому випадку використовувався набір із 100 дерев рішень, оскільки всі розрахунки виконувались на персональну комп'ютері з обмеженими ресурсами. Кожне з цих дерев будується на основі випадкової підвибірки навчальних даних та з урахуванням випадкової підмножини ознак на кожному етапі розбиття вузла. Така стохастичність у побудові призводить до значної варіативності між окремими деревами в ансамблі.

Кінцевий результат роботи моделі (класифікація або регресія) формується шляхом агрегації прогнозів від усіх цих різнорідних дерев – наприклад, шляхом голосування більшістю для задач класифікації або усередненням для задач регресії. Відповідно, спроба представити всю сукупність цих дерев та механізм їхньої взаємодії у вигляді єдиної, легко інтерпретованої графічної структури, аналогічної візуалізації окремого дерева рішень, є практично нездійсненним завданням через надмірну складність та багатовимірність такого представлення.

Хоча окремі дерева з лісу можуть бути візуалізовані (рисунок 8), це не відображатиме колективну логіку прийняття рішень усього ансамблю.

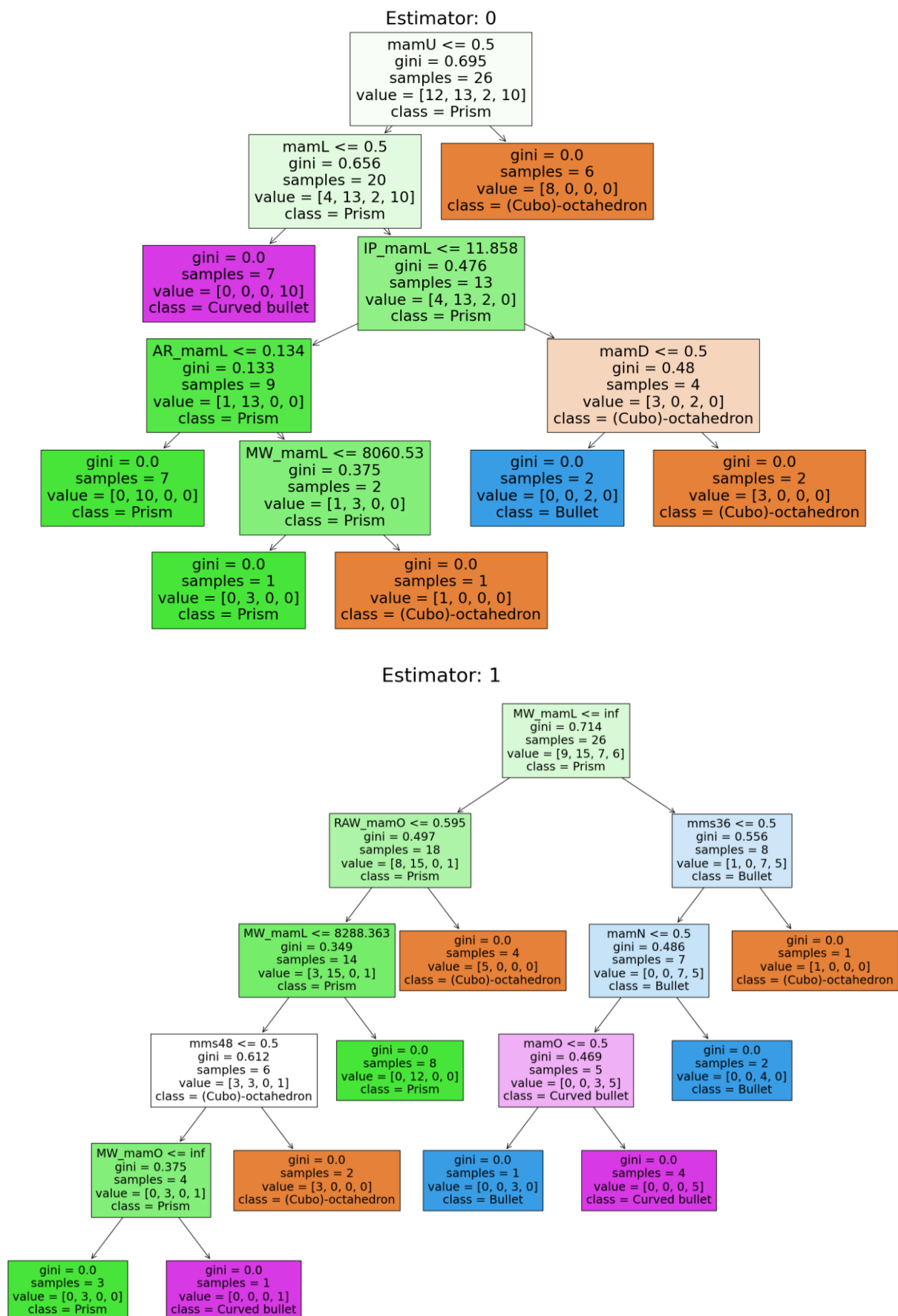


Рисунок 8 – Графічне представлення перших двох дерев у ансамблі випадкового лісу

3.1.2 Результати перехресної перевірки моделей

Результати, отримані за допомогою методології перехресної валідації (крос-валізації), засвідчили значний потенціал обох використаних алгоритмів машинного навчання – дерева рішень та випадкового лісу – у контексті предиктивного моделювання морфологічних характеристик біогенних магнітних наночастинок. Зокрема, ансамблева модель випадкового лісу продемонструвала усереднену точність класифікації на рівні 90% при стандартному відхиленні, що не перевищувало 9%. Натомість, архітектура дерева рішень, хоча й забезпечила прийнятний рівень класифікаційної точності, характеризувалася дещо нижчим середнім значенням – 82%, та суттєво вищим показником стандартного відхилення, що склав 11%. Такі відмінності у значеннях стандартного відхилення вказують на вищу робастність та генералізаційну здатність моделі випадкового лісу, її меншу схильність до перенавчання та вищу стійкість до флуктуацій у складі навчальних вибірок, що використовуються на етапах перехресної перевірки, порівняно з одиночною, більш чутливою до варіацій даних, структурою дерева рішень.

Візуалізація результатів за допомогою стовпчастих діаграм на рисунку 9 підтвердила варіативність оцінок точності для обох моделей за різних випадкових розділень даних, особливо для дерева рішень.

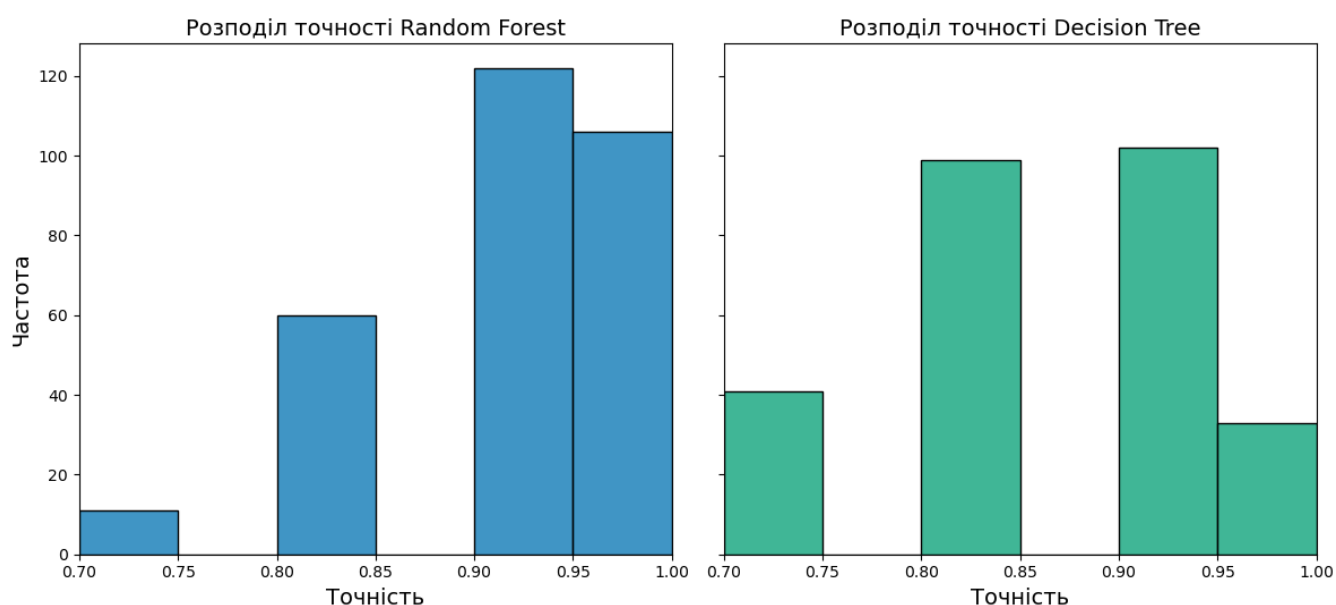


Рисунок 9 – Порівняння розподілу точності моделей випадкового лісу та дерева рішень за результатами перехресної перевірки

Проте, загальний розподіл оцінок для випадкового лісу був ближчим до 100% точності, що свідчить про його кращу узагальнювальну здатність.

3.2 Результати зовнішньої валідації

Перевірка на нових даних також показала перевагу методу Random Forest. Для штаму PR-3 дерево рішень визначило форму БМН як октаедричну (рисунок 10), тоді як випадковий ліс – як призматичну. За літературними даними [31] *Magnetococcus sp.* PR-3 має саме призматичну форму біогенних магнітних наночастинок, що підтверджує вищу точність передбачення методом випадкового лісу для цього конкретного випадку.

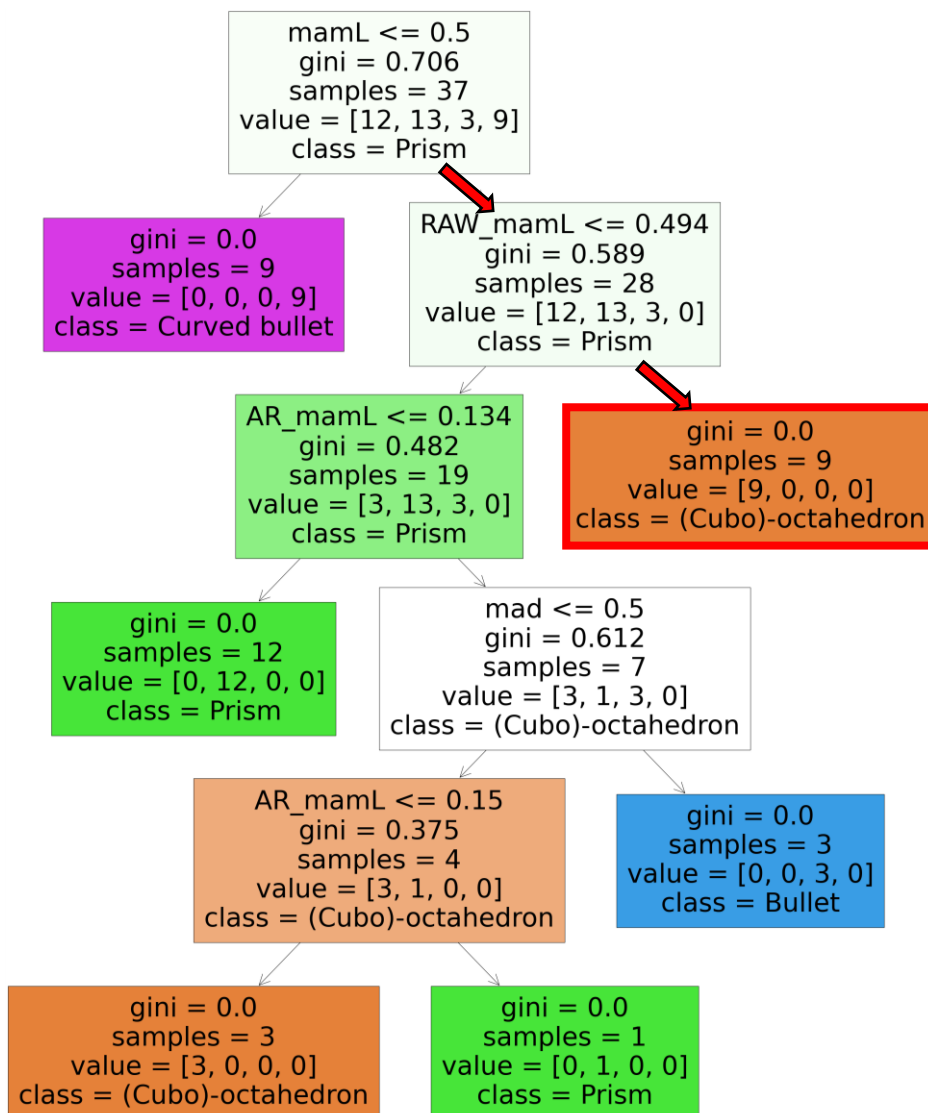


Рисунок 10 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Magnetococcus sp.* PR-3 за допомогою моделі дерева рішення

На відміну від PR-3, для штаму *Candidatus Desulfarcum epimagneticum* CR-1, який за літературними даними має невизначену призматичну форму, близьку до форми кулі [32], обидві моделі – дерево рішень та випадковий ліс – передбачили форму БМН як форму кулі (рисунок 11). Це передбачення, ймовірно, пов'язано з тим, що цей штам має типові ознаки для МТБ, що характеризуються формою кулі, які використовуються моделями для класифікації, зокрема, наявність генів, що кодують Mad білки та білок MamL.

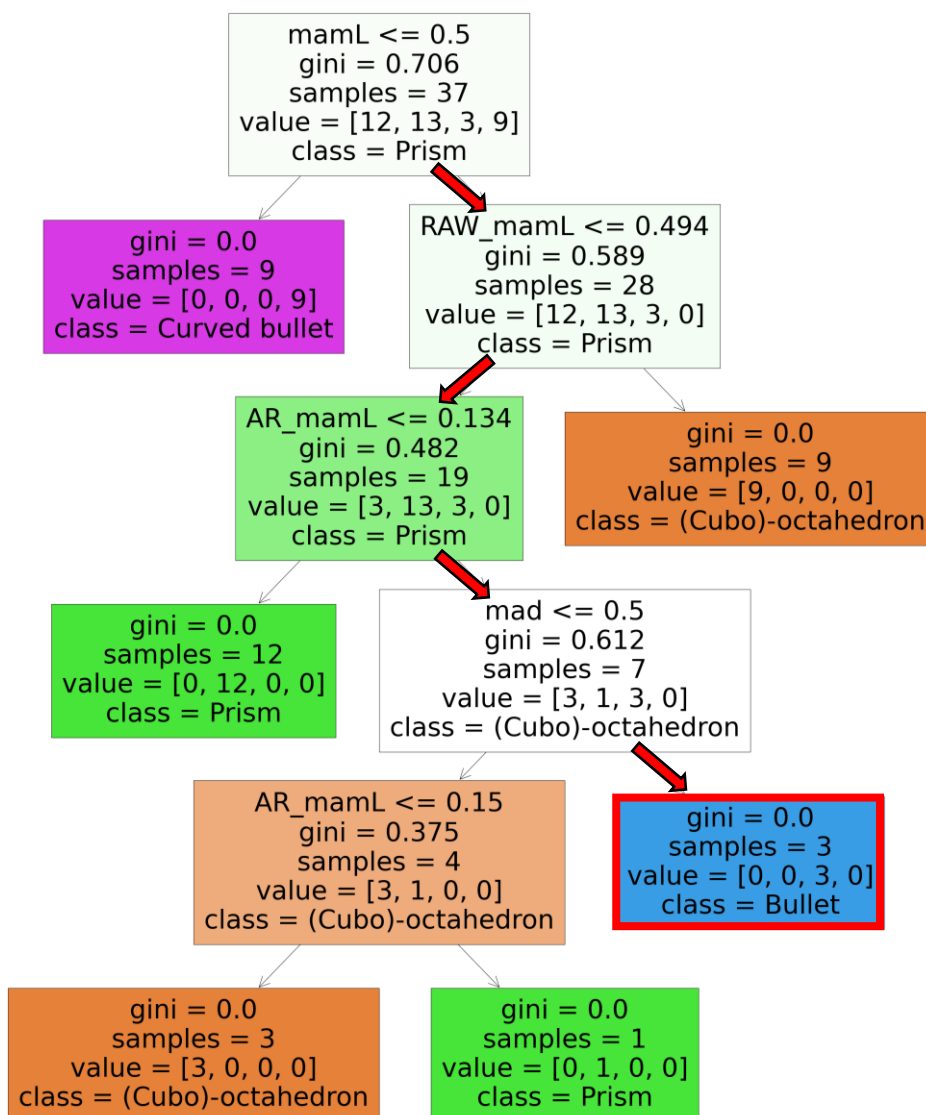


Рисунок 11 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Candidatus Desulfarcum epimagneticum* CR-1 за допомогою моделі дерева рішення

Цікавий випадок спостерігається для штаму *Candidatus Deltaproteobacteria* FH-1, для якого літературні дані вказують на форму БМН у вигляді кулі [33]. У цьому випадку дерево рішень видало помилкове передбачення – призматична форма (рисунок 12). Це могло статися через специфічну структуру дерева (рисунок 7), де критерій поділу за ароматичністю білка MamL був оцінений раніше, ніж більш сильний предиктор форми кулі – наявність Mad-білків, що призвело до неправильного розгалуження. Натомість модель випадкового лісу, інтегруючи рішення багатьох дерев, змогла подолати цю особливість окремого дерева і успішно передбачити правильну форму БМН – кулю.

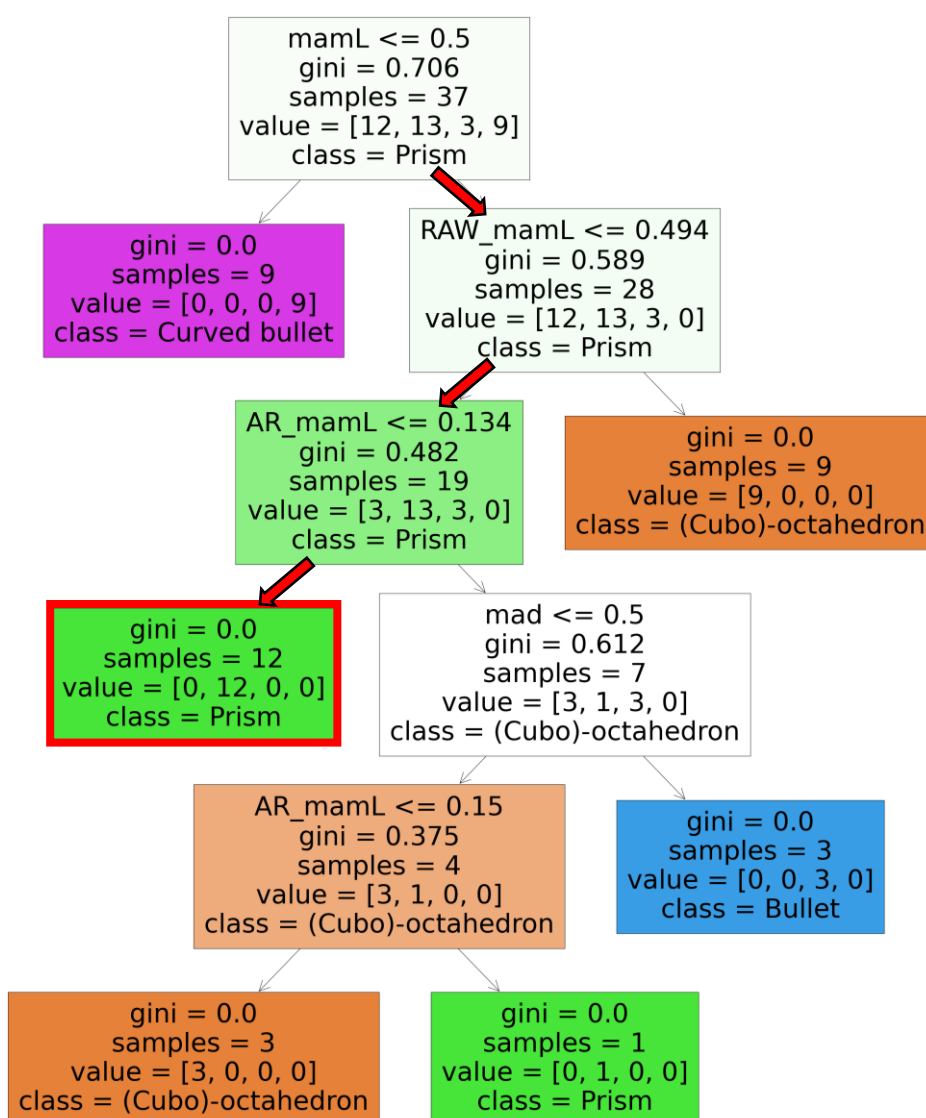


Рисунок 12 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Candidatus Deltaproteobacteria* FH-1 за допомогою моделі дерева рішення

Наступними розглянутими штамами були *Magnetospirillum sp.* UT-4 та *Magnetospirillum sp.* LM-5. Для штаму UT-4 літературні джерела [31] підтверджують наявність БМН у кубічно-октаедричній формі, і обидві розроблені моделі успішно передбачили цю форму. Аналогічна ситуація спостерігається і для штаму LM-5, для якого літературні дані [34] також вказують на кубічно-октаедричну форму БМН, і обидві моделі класифікації дали правильне передбачення (рисунок 13). Це свідчить про адекватність обраних ознак для класифікації даної форми.

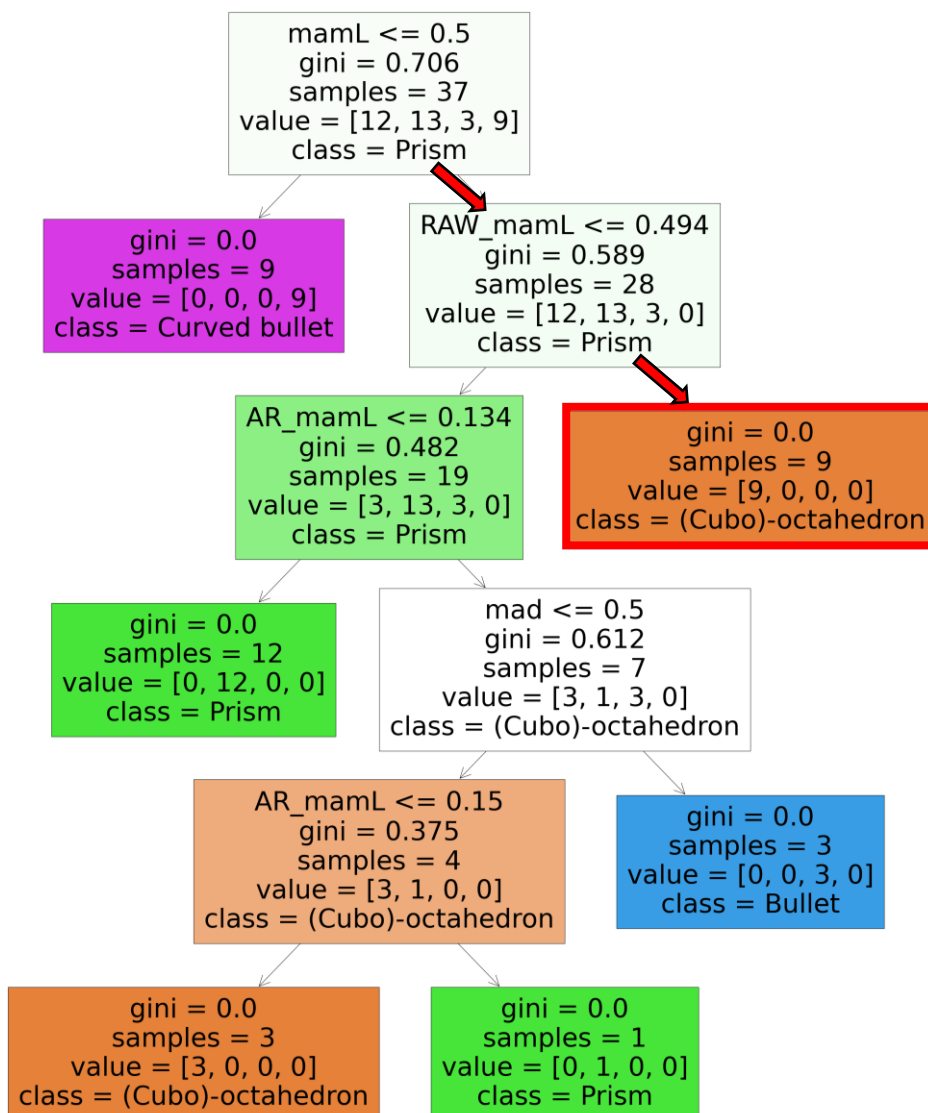


Рисунок 13 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Magnetospirillum sp.* UT-4 та *Magnetospirillum sp.* LM-5 за допомогою моделі дерева рішення

Ще один цікавий випадок спостерігається для штаму *Gammaproteobacteria bacterium* CCP3-SC1AL1, у якого за літературними даними біогенні магнітні наночастинки мають призматичну форму [35]. Для цього штаму обидві натреновані моделі – як дерево рішень (рисунок 14), так і випадковий ліс – правильно передбачили форму БМН, ідентифікувавши її як призматичну.

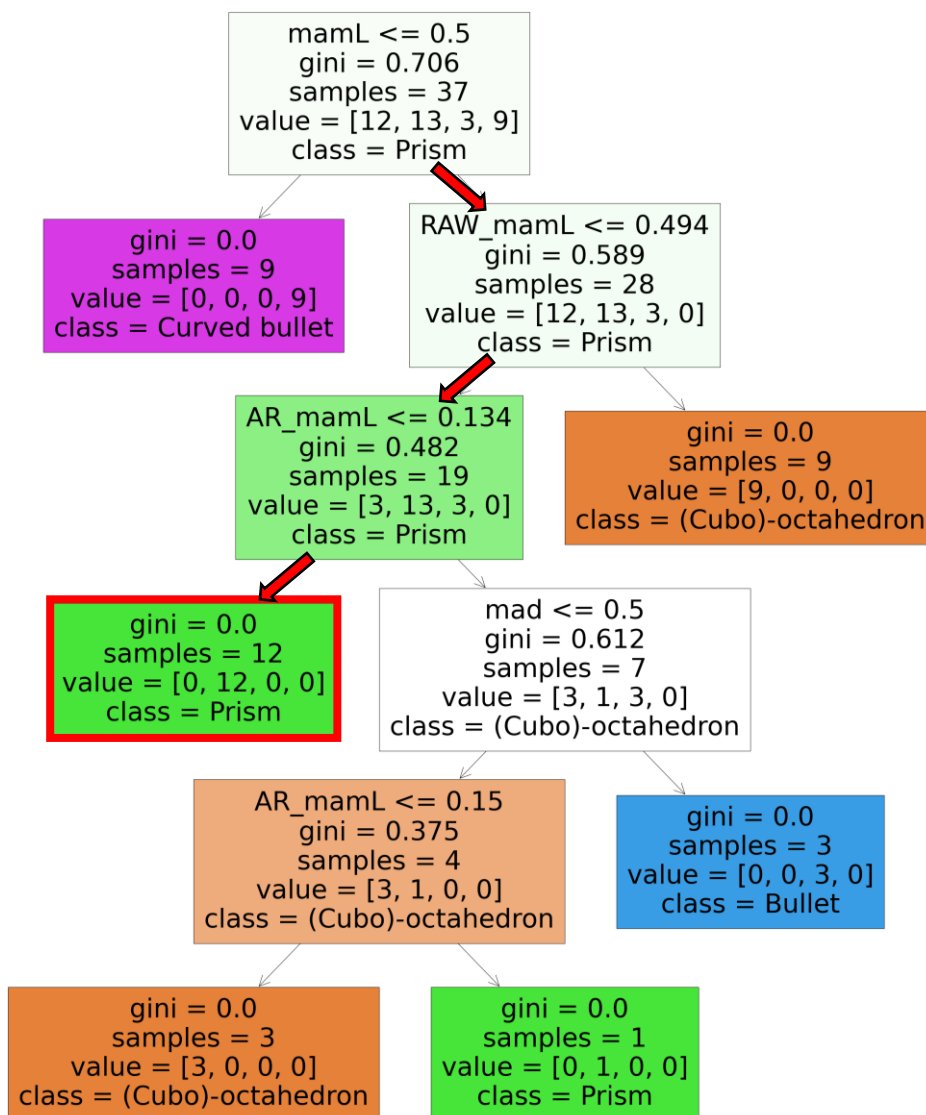


Рисунок 14 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Gammaproteobacteria bacterium* CCP3-SC1AL1 за допомогою моделі дерева рішення

3.3 Передбачення форми для недосліджених штамів

Форму вигнутої кулі БМН штаму *Candidatus Magnetominusculus xianensis* НСН-1 передбачили обидві моделі, дерево рішень та випадковий ліс. Це пояснюється тим, що дескриптор наявності Mad-білків є дуже показовим і явно виділяє цей клас магнітотаксисних бактерій, роблячи передбачення більш очевидним. Також особливим для цього класу є відсутність гену, що кодує білок MamL. Друга особливість використовується деревом рішення на рисунку 15. З літературних даних відомо, що генетично цей штам близький до *Magnetominusculus linsii* LVB02, що має БМН у формі вигнутої кулі також [36].

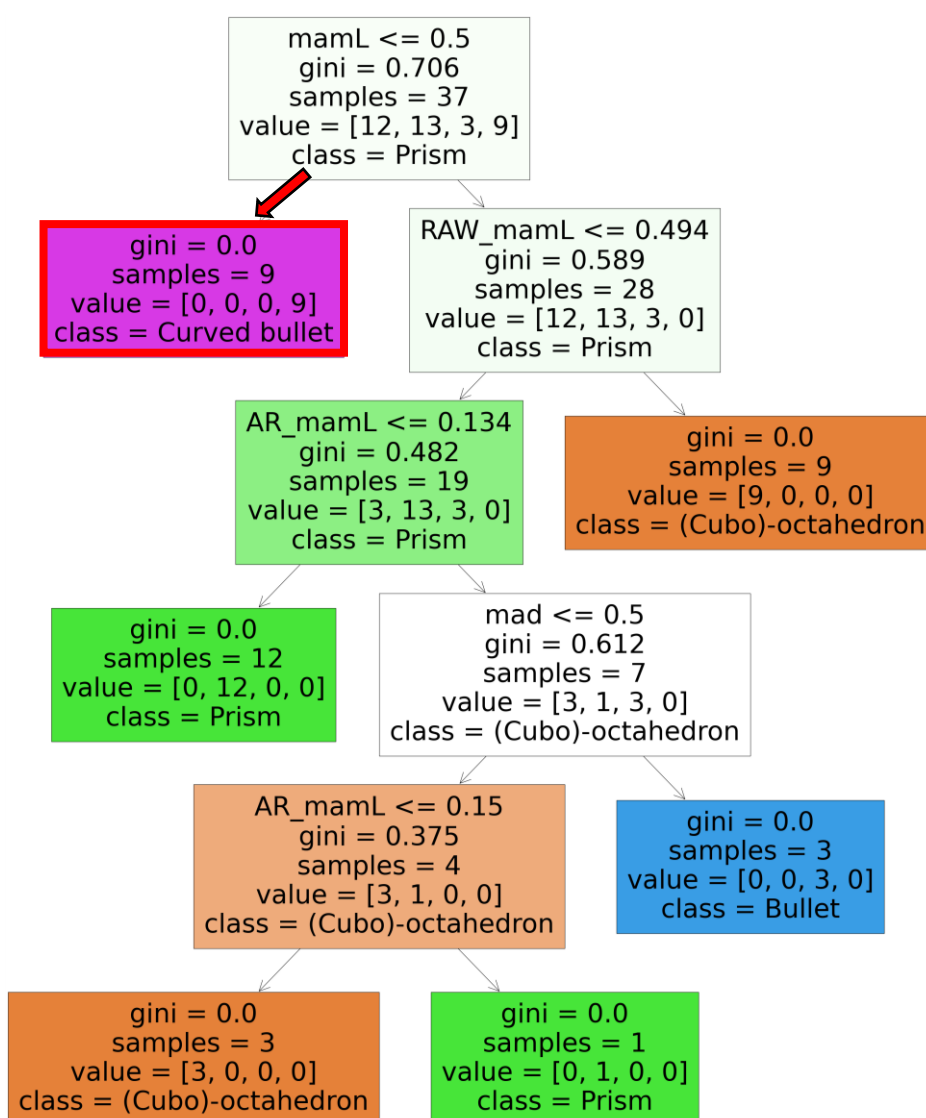


Рисунок 15 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Candidatus Magnetominusculus xianensis* НСН-1 за допомогою моделі дерева рішення

Щодо штаму *Magnetospirillum sp.* SS-4, який відомий своєю генетичною наближеністю до *Magnetospirillum magneticum* AMB-1 [37], літературні дані про форму його БМН наразі відсутні. Застосування обох моделей машинного навчання дозволило передбачити для нього кубічно-октаедричну форму наночастинок (рисунок 16). Враховуючи, що для застосувань у магнітогенетиці часто більш важливими є видовжені призматичні частинки з анізотропними магнітними властивостями, передбачена форма SS-4, ймовірно, вказує на меншу перспективність цього штаму для подальших глибоких досліджень у контексті таких застосувань.

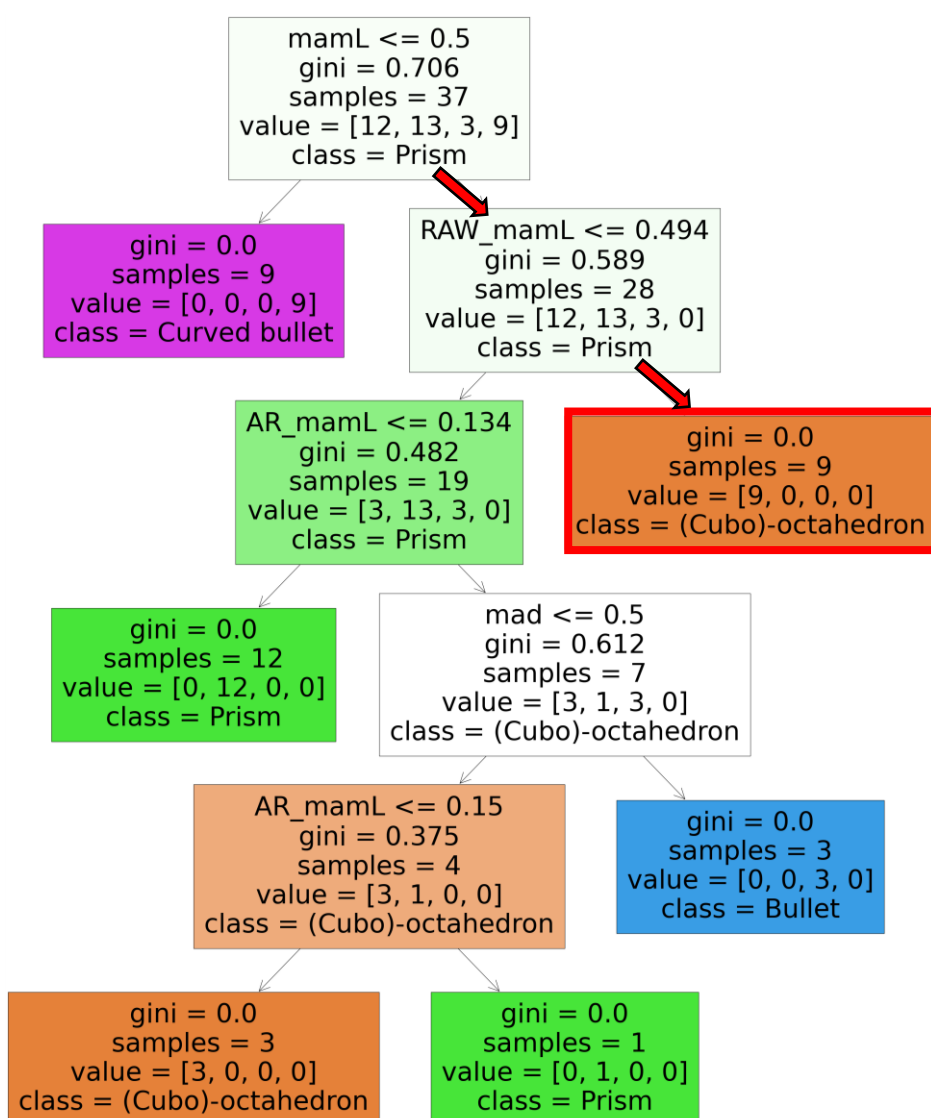


Рисунок 16 – Шлях рішення для передбачення форми біогенних магнітних наночастинок *Magnetospirillum sp.* SS-4 за допомогою моделі дерева рішення

Штами *Alpha proteobacterium* LM-2, *Alpha proteobacterium* LEMS, *Alpha proteobacterium* KR-1 та *Alpha proteobacterium* CB-1 не мають повних розшифрованих геномів (таблиця 3), оскільки були відкриті та досліджені у 2012 році [34], коли ще не було значеного поширення технології секвенування наступного покоління. Встановлено, що всі вказані штамми магнітотаксисних бактерій мають ген, що кодує білок MamO та білки, що специфічні для типу *Pseudomonadota* [34], тобто потенційно мають призматичні чи кубічно-октаедричні біогенні магнітні наночастинки. Ці та подібні штамми потребують більш детального дослідження, а щоб відсіяти серед них ті, що мають точно форму кулі, вигнутої кулі чи кубічно-октаедричну форму, можна використати розроблений у даній роботі алгоритм на основі машинного навчання.

Таблиця 3 – Геномні та протеомні дані, що зібрані для *Alpha proteobacterium* LM-2, *Alpha proteobacterium* LEMS, *Alpha proteobacterium* KR-1 та *Alpha proteobacterium* CB-1

Штам	MACG	MASGP	mamL	mamO	mad	man	Ізоелектрична точка MamL	Ізоелектрична точка MamO	Відн. вага вирівнювання MamL	Відн. вага вирівнювання MamO	Молекулярна вага MamL	Молекулярна вага MamO	Ароматичність MamL	Ароматичність MamO
LM-2	1	1	?	1	0	0	?	6,08	?	0,81	?	60070,16	?	0,07
LEMS	1	1	?	1	0	0	?	6,49	?	0,82	?	60494,73	?	0,07
KR-1	1	1	?	1	0	0	?	6,51	?	0,82	?	60510,77	?	0,07
CB-1	1	1	?	1	0	0	?	6,38	?	0,92	?	61253,43	?	0,07

РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА ДОВКІЛЛЯ

Організація безпечних умов праці в біотехнологічній лабораторії є критично важливою для збереження здоров'я працівників та запобігання аварійним ситуаціям. У даному розділі викладено основні вимоги та правила охорони праці та безпеки в надзвичайних ситуаціях, що ґрунтуються на Інструкції з охорони праці для працюючих в лабораторії Інституту клітинної біології та генетичної інженерії НАН України [38] та загальноприйнятих стандартах лабораторної практики.

4.1 Загальні положення

- Працівники лабораторії повинні неухильно дотримуватись правил внутрішнього трудового розпорядку, посадових інструкцій та інструкцій з охорони праці.
- Не допускається виконання робіт у стані алкогольного, наркотичного чи токсичного сп'яніння, а також у хворобливому стані або стані втоми, що може вплинути на координацію та увагу.
- До самостійної роботи допускаються особи, які пройшли відповідне навчання, інструктажі з охорони праці (первинний, повторний), медичний огляд та не мають протипоказань для виконання робіт.
- Кожен працівник несе особисту відповідальність за дотримання правил безпеки на своєму робочому місці та на території лабораторії.

4.2 Вимоги безпеки перед початком роботи

- Перед початком роботи необхідно одягнути передбачений ЗІЗ (засоби індивідуального захисту): лабораторний халат (застібнутий), змінне закрите взуття. Додаткові ЗІЗ (рукавички, захисні окуляри, маски) надягаються відповідно до специфіки майбутньої роботи.
- Перевірити стан свого робочого місця, прибрати зайві предмети, що можуть заважати роботі.
- Перевірити справність обладнання, приладів, електромереж, вентиляції, освітлення. Про всі виявлені несправності негайно повідомити керівника робіт.

- Переконатись у наявності та придатності необхідних для роботи реактивів, матеріалів, посуду.
- Ознайомитись з протоколом експерименту та переконатись у розумінні послідовності дій та потенційних ризиків.
- Знати розташування засобів пожежогасіння (вогнегасники, пісок), аптечки першої допомоги, аварійного душу та фонтанчика для промивання очей.

4.3 Вимоги безпеки під час роботи

- Використовувати ЗІЗ протягом усього часу виконання робіт, пов'язаних з потенційною небезпекою (робота з хімікатами, біоагентами, гарячими рідинами, УФ-випромінюванням тощо).
- Роботи з леткими, їдкими, токсичними, вибухонебезпечними речовинами, а також роботи, що супроводжуються утворенням парів або пилу, проводити виключно у справній витяжній шафі або ламінарному боксі при увімкненій вентиляції [38].
- Всі ємності з реактивами повинні бути чітко марковані із зазначенням назви речовини, концентрації та дати приготування (за необхідності).
- Не залишати відкритими ємності з леткими або агресивними речовинами.
- Не проводити переливання агресивних рідин на відкритому столі, використовувати для цього піддони.
- Піпетувати рідини тільки за допомогою груші або автоматичних піпеток. Категорично забороняється піпетувати ротом [38].
- Дотримуватись обережності при роботі з гострими предметами (скальпелі, леза, голки). Використовувати спеціальні контейнери для утилізації гострих відходів.
- Не працювати одному з небезпечними речовинами або під час виконання потенційно небезпечних маніпуляцій (наприклад, робота під високим тиском або з високою температурою).
- Підтримувати вільними проходи між робочими столами та до виходів.

- Не відволікатись та не відволікати інших працівників під час виконання відповідальних операцій.

4.4 Вимоги охорони праці при користуванні хімічним посудом

- Перед використанням скляний посуд необхідно ретельно оглянути на відсутність тріщин, сколів або інших пошкоджень. Пошкоджений посуд використовувати заборонено.
- При з'єднанні скляних деталей або вставлянні скляних трубок у гумові пробки чи шланги обов'язково змащувати поверхні водою або гліцерином та захищати руки щільною тканиною [38].
- Нагрівати скляний посуд (колби, пробірки) слід рівномірно, використовуючи азбестову сітку або піщану баню. Не нагрівати посуд з товстими стінками на відкритому вогні.
- Розбите скло негайно прибирати щіткою та совком (не руками) та викидати у спеціально відведені контейнери для уламків скла [38].
- Ретельно мити та висушувати хімічний посуд після використання.

4.5 Вимоги охорони праці при користуванні персональним комп'ютером

- Робоче місце користувача ПК повинно бути організовано відповідно до ергономічних вимог: монітор розташований на рівні очей або трохи нижче, клавіатура та миша – на зручній відстані для уникнення напруження в руках та плечах.
- Крісло повинно мати регульовану висоту та спинку для підтримки правильної постави.
- Освітлення на робочому місці має бути достатнім, але без прямих відблисків на моніторі.
- Необхідно робити регулярні перерви в роботі за комп'ютером (кожні 45-60 хвилин) для розминки, гімнастики для очей та відпочинку.
- Підтримувати чистоту навколо комп'ютера, не допускати скупчення рідин або їжі біля обладнання.

4.6 Вимоги безпеки після закінчення роботи

- Привести у порядок своє робоче місце: прибрати посуд, реактиви, матеріали.
- Відключити всі електроприлади та обладнання (нагрівальні плити, шейкери, центрифуги тощо), які не вимагають постійної роботи.
 - Перевірити, чи перекриті крани подачі води, газу, вакууму.
 - Знезаразити та утилізувати лабораторні відходи відповідно до встановлених правил Зб.
 - Зняти лабораторний халат та інший ЗІЗ, розмістити їх у спеціально відведених місцях.
 - Обов'язково вимити руки з милом перед виходом з лабораторії.
 - Переконатись, що всі вікна зачинені, вимкнене освітлення (якщо немає потреби в цілодобовому освітленні).
 - Закрити двері лабораторії.

4.7 Вимоги безпеки в аварійних ситуаціях

- Загальні дії: При виникненні будь-якої аварійної ситуації (пожежа, витік газу, розлив небезпечних речовин, ураження електричним струмом, нещасний випадок):
 - Зберігати спокій, не панікувати.
 - Негайно припинити виконання робіт, якщо це не становить додаткової небезпеки.
 - За можливості та дотримання особистої безпеки, вжити заходів для ліквідації джерела небезпеки (наприклад, вимкнути електроприлад, перекрити кран).
 - Повідомити про подію керівника лабораторії або іншого відповідального працівника.
 - За необхідності, викликати відповідні екстрені служби: пожежна охорона (101), поліція (102), швидка медична допомога (103), газова служба (104). Чітко повідомити адресу та характер події.

- Надати першу домедичну допомогу постраждалим, використовуючи аптечку [38].
- Діяти відповідно до плану евакуації та вказівок керівника.
- Пожежа або загоряння:
 - При виявленні ознак загоряння (дим, запах горіння) негайно повідомити про це [38].
 - При загорянні електрообладнання – негайно знеструмити його.
 - Використовувати первинні засоби пожежогасіння відповідно до типу пожежі: вогнегасники (порошкові, вуглекислотні), пісок [38]. Не використовувати воду для гасіння електрообладнання під напругою, легкозаймистих рідин, лужних металів.
 - При сильному задимленні рухатись, пригинаючись до підлоги, захищаючи органи дихання вологою тканиною.
 - При загорянні одягу на людині – збити полум'я, використовуючи покривало, пісок або обливаючи водою (якщо це не хімічний опік та не горить метал), не намагатись бігти [38].
- Розлив небезпечних хімічних речовин:
 - При невеликих розливах неагресивних речовин – засипати абсорбентом (пісок, вермикуліт) та зібрати у спеціальну ємність.
 - При розливі їдких або токсичних речовин – негайно залишити приміщення, забезпечити його вентиляцію, повідомити керівника. Вживати заходів з ліквідації тільки після оцінки ризиків та з використанням відповідних ЗІЗ.
 - При потраплянні хімікатів на шкіру або в очі – негайно промити уражене місце великою кількістю проточної води (для очей використовувати фонтанчик для промивання очей) протягом щонайменше 15 хвилин та звернутись за медичною допомогою [38].

- Травми:
 - При порізах, опіках, інших травмах – негайно надати першу допомогу (обробити рану антисептиком, накласти пов'язку, охолодити опік) та звернутись за медичною допомогою.
 - При ураженні електричним струмом – перш за все, усунути контакт постраждалого з джерелом струму, використовуючи сухий неелектропровідний предмет. При необхідності провести реанімаційні заходи та викликати швидку допомогу [38].

ВИСНОВКИ

1. Проведено пошук наукової літератури за тематикою дослідження та показано, що обмеженість існуючих методів, зокрема трудомісткість і низька пропускна здатність електронної мікроскопії, ускладнюють визначення форми біогенних магнітних наночастинок магнітотаксисних бактерій, що підкреслює критичну важливість розробки нових підходів.

2. Сформовано та підготовлено набір даних про наявність генів *tamA*, *tamB*, *tamE*, *tamI*, *tamK*, *tamM*, *tamP*, *tamQ*, *tamH*, *tamF*, *tamS*, *tamT*, *tamC*, *tamD*, *tamG*, *tamR*, *tamL*, *tamO*, *tamN*, *tamX*, *tamZ*, *tamY*, *tamJ*, *tamU*, *tamV*, *tamW*, *mms5*, *mms6*, *mms6-L*, *mms36*, *mms48*, а також численні гени груп *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) в їх геномах, а також фізико-хімічні властивості кодованих цими генами білків.

3. Розроблено алгоритм для передбачення форми біогенних магнітних наночастинок магнітотаксисних бактерій на основі алгоритмів машинного навчання з використанням відкритої бази даних GenBank NCBI.

4. Написано скрипт-програму за створеним алгоритмом на основі мови програмування Python та її пакетів Biopython, scikit-learn та pandas.

5. Виконано оцінювання точності та стабільності розроблених моделей за допомогою методу перехресної перевірки. Встановлено, що модель випадкового лісу має більшу робастність та генералізаційну здатність, оскільки її точність складає $90 \pm 9\%$, в той час як модель дерева рішень має точність $82 \pm 11\%$.

6. Проведено тестування моделей на незалежних експериментальних даних для перевірки їхньої узагальнювальної здатності. Передбачено форму вигнутої кулі біогенних магнітних наночастинок для *Candidatus Magnetominusculus xianensis* HCN-1 та кубічно-октаедричну – для *Magnetospirillum* sp. SS-4. Знайдено 4 штами *Alpha proteobacterium* LM-2, LEMS, KR-1 та CB-1, що не мають відкритих даних секвенування повних геномів, але є потенційними продуцентами призматичних біогенних магнітних наночастинок.

7. Описано основні заходи щодо охорони праці під час роботи із персональним комп'ютером та у мікробіологічній лабораторії.

ПЕРЕЛІК ПОСИЛАНЬ

1. Magnetogenetics: remote activation of cellular functions triggered by magnetic switches / S. Del Sol-Fernández et al. *Nanoscale*. 2022. Vol. 14, no. 6. P. 2091–2118. DOI: 10.1039/d1nr06303k.
2. Lefevre C. T., Bazylnski D. A. Ecology, Diversity, and Evolution of Magnetotactic Bacteria. *Microbiology and Molecular Biology Reviews*. 2013. Vol. 77, no. 3. P. 497–526. DOI: 10.1128/mubr.00021-13.
3. Magnetogenetics as a promising tool for controlling cellular signaling pathways / A. A. Latypova et al. *Journal of Nanobiotechnology*. 2024. Vol. 22, no. 1. DOI: 10.1186/s12951-024-02616-z.
4. Uncovering a possible role of reactive oxygen species in magnetogenetics / M. I. Brier et al. *Scientific Reports*. 2020. Vol. 10, no. 1. DOI: 10.1038/s41598-020-70067-1.
5. Remote control of ion channels and neurons through magnetic-field heating of nanoparticles / H. Huang та ін. *Nature Nanotechnology*. 2010. Т. 5, № 8. С. 602–606. DOI: 10.1038/nnano.2010.125.
6. Wu J., Goyal R., Grandl J. Localized force application reveals mechanically sensitive domains of Piezo1. *Nature Communications*. 2016. Т. 7, № 1. DOI: 10.1038/ncomms12939.
7. Vogt N. Manipulating neurons with magnetogenetics. *Nature Methods*. 2016. Vol. 13, no. 5. P. 394. DOI: 10.1038/nmeth.3857.
8. Magnetic hyperthermia / S. Ruta et al. *Magnetic Sensors and Actuators in Medicine*. 2023. P. 185–226. DOI: 10.1016/b978-0-12-823294-1.00001-4.
9. The interaction between nanoparticles and immune system: application in the treatment of inflammatory diseases / J. Liu et al. *Journal of Nanobiotechnology*. 2022. Vol. 20, no. 1. DOI: 10.1186/s12951-022-01343-7.
10. Biocompatibility Research of Magnetosomes Synthesized by *Acidithiobacillus ferrooxidans* / B.-Q. Wu et al. *International Journal of Molecular Sciences*. 2025. Vol. 26, no. 9. P. 4278. DOI: 10.3390/ijms26094278.

11. Small-Molecule Modulators of Mitochondrial Channels as Chemotherapeutic Agents. *Cellular Physiology and Biochemistry*. 2019. Vol. 53, S1. P. 11–43. DOI: 10.33594/000000192.
12. Magnetic Iron Oxide Nanoparticles: Synthesis, Stabilization, Vectorization, Physicochemical Characterizations, and Biological Applications / S. Laurent et al. *Chemical Reviews*. 2008. Vol. 108, no. 6. P. 2064–2110. DOI: 10.1021/cr068445e.
13. Tuning the Magnetic Properties of Nanoparticles / A. Kolhatkar et al. *International Journal of Molecular Sciences*. 2013. Vol. 14, no. 8. P. 15977–16009. DOI: 10.3390/ijms140815977.
14. Large-Scale Synthesis and Medical Applications of Uniform-Sized Metal Oxide Nanoparticles / H. J. Kwon et al. *Advanced Materials*. 2018. Vol. 30, no. 42. P. 1704290. DOI: 10.1002/adma.201704290.
15. Alphandéry E. Applications of Magnetosomes Synthesized by Magnetotactic Bacteria in Medicine. *Frontiers in Bioengineering and Biotechnology*. 2014. Vol. 2. DOI: 10.3389/fbioe.2014.00005.
16. Faivre D., Schüler D. Magnetotactic Bacteria and Magnetosomes. *Chemical Reviews*. 2008. Vol. 108, no. 11. P. 4875–4898. DOI: 10.1021/cr078258w.
17. Expanding magnetic organelle biogenesis in the domain Bacteria / W. Lin et al. *Microbiome*. 2020. Vol. 8, no. 1. DOI: 10.1186/s40168-020-00931-9.
18. Key gene networks that control magnetosome biomineralization in magnetotactic bacteria / P. Liu et al. *National Science Review*. 2022. Vol. 10, no. 1. DOI: 10.1093/nsr/nwac238.
19. Komeili A. Molecular Mechanisms of Magnetosome Formation. *Annual Review of Biochemistry*. 2007. Vol. 76, no. 1. P. 351–366. DOI: 10.1146/annurev.biochem.74.082803.133444.
20. Applications of Magnetotactic Bacteria, Magnetosomes and Magnetosome Crystals in Biotechnology and Nanotechnology: Mini-Review / G. Vargas et al. *Molecules*. 2018. Vol. 23, no. 10. P. 2438. DOI: 10.3390/molecules23102438.
21. Set-up of a pharmaceutical cell bank of *Magnetospirillum gryphiswaldense* MSR1 magnetotactic bacteria producing highly pure magnetosomes / T. Chades et al.

- Microbial Cell Factories. 2024. Vol. 23, no. 1. DOI: 10.1186/s12934-024-02313-4.
22. The effect of particle design on cellular internalization pathways / S. E. A. Gratton et al. Proceedings of the National Academy of Sciences. 2008. Vol. 105, no. 33. P. 11613–11618. DOI: 10.1073/pnas.0801763105.
23. Albanese A., Tang P. S., Chan W. C. W. The Effect of Nanoparticle Size, Shape, and Surface Chemistry on Biological Systems. Annual Review of Biomedical Engineering. 2012. Vol. 14, no. 1. P. 1–16. DOI: 10.1146/annurev-bioeng-071811-150124.
24. Binnig G., Quate C. F., Gerber C. Atomic Force Microscope. Physical Review Letters. 1986. Vol. 56, no. 9. P. 930–933. DOI: 10.1103/physrevlett.56.930.
25. Vokoun D., Samal S., Stachiv I. Magnetic Force Microscopy in Physics and Biomedical Applications. Magnetochemistry. 2022. Vol. 8, no. 4. P. 42. DOI: 10.3390/magnetochemistry8040042.
26. Biomedical applications of magnetosomes: State of the art and perspectives / G. Ren et al. Bioactive Materials. 2023. Vol. 28. P. 27–49. DOI: 10.1016/j.bioactmat.2023.04.025.
27. Barber-Zucker S., Keren-Khadmy N., Zarivach R. From invagination to navigation: The story of magnetosome-associated proteins in magnetotactic bacteria. Protein Science. 2015. Vol. 25, no. 2. P. 338–351. DOI: 10.1002/pro.2827.
28. Scikit-learn: Machine Learning in Python / F. Pedregosa et al. Journal of Machine Learning Research. 2011. Vol. 12. DOI: 10.48550/arXiv.1201.0490.
29. scikit-learn. Supervised learning. scikit-learn. URL: https://scikit-learn.org/stable/supervised_learning.html.
30. scikit-learn. ShuffleSplit. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.ShuffleSplit.html.
31. Diversity of Magneto-Aerotactic Behaviors and Oxygen Sensing Mechanisms in Cultured Magnetotactic Bacteria / C. T. Lefèvre et al. Biophysical Journal. 2014. Vol. 107, no. 2. P. 527–538. DOI: 10.1016/j.bpj.2014.05.043.

32. Ectosymbiotic bacteria at the origin of magnetoreception in a marine protist / C. L. Monteil et al. *Nature Microbiology*. 2019. Vol. 4, no. 7. P. 1088–1095. DOI: 10.1038/s41564-019-0432-7.
33. Comparative genomic analysis of magnetotactic bacteria from the Deltaproteobacteria provides new insights into magnetite and greigite magnetosome genes required for magnetotaxis / C. T. Lefèvre et al. *Environmental Microbiology*. 2013. P. n/a. DOI: 10.1111/1462-2920.12128.
34. Insight into the Evolution of Magnetotaxis in *Magnetospirillum* spp., Based on *mamGene* Phylogeny / C. T. Lefèvre et al. *Applied and Environmental Microbiology*. 2012. Vol. 78, no. 20. P. 7238–7248. DOI: 10.1128/aem.01951-12.
35. Magnetotactic bacteria from diverse Pseudomonadota families biomineralize intracellular Ca-carbonate / C. C. Mangin et al. *The ISME Journal*. 2025. DOI: 10.1093/ismejo/wrae260.
36. Detection of interphylum transfers of the magnetosome gene cluster in magnetotactic bacteria / M. Uzun et al. *Frontiers in Microbiology*. 2022. Vol. 13. DOI: 10.3389/fmicb.2022.945734.
37. Repeated horizontal gene transfers triggered parallel evolution of magnetotaxis in two evolutionary divergent lineages of magnetotactic bacteria / C. L. Monteil et al. *The ISME Journal*. 2020. Vol. 14, no. 7. P. 1783–1794. DOI: 10.1038/s41396-020-0647-x.
38. Інструкція № 2 з охорони праці для працюючих в лабораторії. ІКБГІ. URL: http://icbge.org.ua/re/images/4/4c/Instruction_OP_02_Lab_2022-06-21.pdf.

ДОДАТОК А

Код програми для навчання моделі дерева рішень:

```

import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score, ShuffleSplit
import matplotlib.pyplot as plt
from sklearn import tree

seed = 0
# Прочитання даних із Excel-файла
df = pd.read_excel(
    'Descriptors.xlsx', # Назва Excel-файла
    'Sheet2'           # Назва листа у файлі
)
# Створення списку назв колонок
features = list(df.columns.drop(['strains', 'shape']))

# Кортеж із можливих форм БМН
shapes = (('Cubo)-octahedron', 'Prism', 'Bullet', 'Curved bullet')
# Створення словника {форма БМН: індекс}
# Це необхідно для перетворення текстових даних на числа
d = {shape: index for index, shape in enumerate(shapes)}

# Перетворення даних у колонці 'shape' із назв форм на числові відповідники
df['shape'] = df['shape'].map(d)
# Розділення датафрейму на X та y
y = df['shape']           # Цільовий вектор
X = df.drop(['strains', 'shape'], axis=1) # Матриця початкових даних

```

```

# Розділення набору даних на тренувальний та тестовий
X_train, X_test, y_train, y_test = train_test_split(
    X,          # Матриця початкових даних
    y,          # Цільвий вектор
    test_size=0.2, # 20% даних будуть виділені на тестування
    random_state=seed # Випадковий посів для відтворюваності
)
# Ініціалізація об'єкту класу DecisionTreeClassifier
dtree = DecisionTreeClassifier(random_state=seed)

# Навчання класифікатора на тренувальному наборі даних
dtree = dtree.fit(X_train, y_train)
# Створення вікна фігури
fig = plt.figure(figsize=(25,20))

# Створення дерева рішень
tree.plot_tree(
    dtree,          # Натренована модель класифікатора типу дерево рішень
    feature_names=features, # Назви властивостей (білків чи груп білків)
    class_names=shapes,    # Можливі форми БМН
    filled=True          # Заповнення прямокутників кольорами
)
# Використання натренованої моделі для передбачення за тестовими даними форми БМН
y_pred = dtree.predict(X_test)

# Створення оберненого словника (до словника d), щоб із чисел отримати назви форм
rev = {index: shape for index, shape in enumerate(shapes)}

```

```
# Заміна чисел у y_pred на відповідні текстові назви форм
test = list(map(lambda x: rev[x], y_pred))
# Створення нової колонки 'result' датафрейму X_test із результатами тестування
X_test['result'] = test

# Видалення всіх колонок, крім 'shape'
X_test = X_test.drop(features, axis=1)

# Створення нової колонки 'real' датафрейму X_test із дійсними формами БМН у
числовій формі
X_test['real'] = X_test.index.map(dict(zip(df.index, df['shape'])))

# Заміна чисел у колонці 'real' на відповідні назви форм БМН
X_test['real'] = X_test['real'].map(rev)

# Створення нової колонки 'strain' датафрейму X_test із назвами штамів
X_test['strain'] = X_test.index.map(dict(zip(df.index, df['strains'])))

# Упорядкування колонок у датафремі X_test
X_test = X_test[['strain', 'real', 'result']]

# Створення нової колонки 'is_predict_correct' датафрейму X_test
# із значеннями True/False відповідно до коректності передбаченої форми
X_test['is_predict_correct'] = X_test['real'] == X_test['result']
# Вивід кінцевої таблиці
X_test
# Прочитання даних із Excel-файла
df_test = pd.read_excel(
    'Descriptors.xlsx', # Назва Excel-файла
    'Sheet7'          # Назва листа у файлі
```

```

)
strains = df_test['strains'].tolist()
df_test = df_test.drop(['strains'], axis=1)
# Використання натренованої моделі для передбачення за тестовими даними форми
БМН
y_pred = dtree.predict(df_test)

# Заміна чисел у y_pred на відповідні текстові назви форм
test = list(map(lambda x: rev[x], y_pred))
for st, sh in zip(strains, test):
    print(f'Strain: {st}, Predicted shape: {sh}')
# Ініціалізація класифікатора DecisionTreeClassifier
dtree = DecisionTreeClassifier(random_state=seed)

# Параметри крос-валідації
n_splits = 3
test_size = 0.2
num_seeds = 100
seed_values = range(num_seeds)
all_scores = []

# Запуск крос-валідації для різних seed
for seed in seed_values:
    cv = ShuffleSplit(n_splits=n_splits, test_size=test_size, random_state=seed)
    scores = cross_val_score(dtree, X, y, cv=cv, scoring='accuracy')
    all_scores.extend(scores)

mean_accuracy = np.mean(all_scores)
std_accuracy = np.std(all_scores)

```

```
print(f"\nСередня точність: {mean_accuracy:.2f}")  
print(f"Стандартне відхилення точності: {std_accuracy:.2f}")
```

ДОДАТОК Б

Код програми для навчання моделі випадкового лісу:

```

import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score, ShuffleSplit
import matplotlib.pyplot as plt
from sklearn import tree

seed = 0
# Прочитання даних із Excel-файла
df = pd.read_excel(
    'Descriptors.xlsx', # Назва Excel-файла
    'Sheet2'           # Назва листа у файлі
)
# Створення списку назв колонок
features = list(df.columns.drop(['strains', 'shape']))

# Кортеж із можливих форм БМН
shapes = (('Cubo)-octahedron', 'Prism', 'Bullet', 'Curved bullet')
# Створення словника {форма БМН: індекс}
# Це необхідно для перетворення текстових даних на числа
d = {shape: index for index, shape in enumerate(shapes)}

# Перетворення даних у колонці 'shape' із назв форм на числові відповідники
df['shape'] = df['shape'].map(d)
# Розділення датафрейму на X та y
y = df['shape']           # Цільовий вектор
X = df.drop(['strains', 'shape'], axis=1) # Матриця початкових даних

```

```

# Розділення набору даних на тренувальний та тестовий
X_train, X_test, y_train, y_test = train_test_split(
    X,          # Матриця початкових даних
    y,          # Цільвий вектор
    test_size=0.2, # 20% даних будуть виділені на тестування
    random_state=seed # Випадковий посів для відтворюваності
)
# Ініціалізація об'єкту класу RandomForestClassifier
forest = RandomForestClassifier(random_state=seed)

# Навчання класифікатора на тренувальному наборі даних
forest = forest.fit(X_train, y_train)
fig, axes = plt.subplots(nrows = 2,ncols = 1, figsize = (25,40))
for index in range(0, 2):
    tree.plot_tree(forest.estimators_[index],
                   feature_names = features,
                   class_names = shapes,
                   filled = True,
                   ax = axes[index])

    axes[index].set_title('Estimator: ' + str(index), fontsize = 30)
# Використання натренованої моделі для передбачення за тестовими даними форми
БМН
y_pred = forest.predict(X_test)

# Створення оберненого словника (до словника d), щоб із чисел отримати назви
форм
rev = {index: shape for index, shape in enumerate(shapes)}

# Заміна чисел у y_pred на відповідні текстові назви форм

```

```
test = list(map(lambda x: rev[x], y_pred))
# Створення нової колонки 'result' датафрейму X_test із результатами тестування
X_test['result'] = test

# Видалення всіх колонок, крім 'shape'
X_test = X_test.drop(features, axis=1)

# Створення нової колонки 'real' датафрейму X_test із дійсними формами БМН у
числовій формі
X_test['real'] = X_test.index.map(dict(zip(df.index, df['shape'])))

# Заміна чисел у колонці 'real' на відповідні назви форм БМН
X_test['real'] = X_test['real'].map(rev)

# Створення нової колонки 'strain' датафрейму X_test із назвами штамів
X_test['strain'] = X_test.index.map(dict(zip(df.index, df['strains'])))

# Упорядкування колонок у датафремі X_test
X_test = X_test[['strain', 'real', 'result']]

# Створення нової колонки 'is_predict_correct' датафрейму X_test
# із значеннями True/False відповідно до коректності передбаченої форми
X_test['is_predict_correct'] = X_test['real'] == X_test['result']
# Вивід кінцевої таблиці
X_test
# Прочитання даних із Excel-файла
df_test = pd.read_excel(
    'Descriptors.xlsx', # Назва Excel-файла
    'Sheet7'           # Назва листа у файлі
)
```

```
strains = df_test['strains'].tolist()
df_test = df_test.drop(['strains'], axis=1)
# Використання натренованої моделі для передбачення за тестовими даними форми
БМН
y_pred = forest.predict(df_test)

# Заміна чисел у y_pred на відповідні текстові назви форм
test = list(map(lambda x: rev[x], y_pred))
for st, sh in zip(strains, test):
    print(f'Strain: {st}, Predicted shape: {sh}')
# Ініціалізація класифікатора RandomForestClassifier
forest = RandomForestClassifier(random_state=seed)

# Параметри крос-валідації
n_splits = 3
test_size = 0.2
num_seeds = 100
seed_values = range(num_seeds)
all_scores = []

# Запуск крос-валідації для різних значень seed
for seed in seed_values:
    cv = ShuffleSplit(n_splits=n_splits, test_size=test_size, random_state=seed)
    scores = cross_val_score(forest, X, y, cv=cv, scoring='accuracy')
    all_scores.extend(scores)

mean_accuracy = np.mean(all_scores)
std_accuracy = np.std(all_scores)
print(f"\nСередня точність: {mean_accuracy:.2f}")
print(f"Стандартне відхилення точності: {std_accuracy:.2f}")
```

ДОДАТОК В

```

import pandas as pd
from Bio.Align import PairwiseAligner
from Bio.SeqUtils import ProtParam
from Bio import Entrez, SeqIO
from Bio.SeqRecord import SeqRecord
Entrez.email = 'khakhno.kyryl@lil.kpi.ua'

MACG = ['mamA', 'mamB', 'mamE', 'mamI', 'mamK', 'mamM', 'mamP', 'mamQ']
MACGP = ['mamH', 'mamF', 'mamS', 'mamT']
ETC = ['mms5', 'mms6', 'mms6-L', 'mamC', 'mamD', 'mms36', 'mms48', 'mamG', 'mamR',
       'mamN', 'mamX', 'mamZ', 'mamY', 'mamJ', 'mcaA', 'mcaB', 'mamU', 'mamV',
       'mamW']
MAIN = ['mamL', 'mamO']
MAD = ['mad1', 'mad2', 'mad3', 'mad4', 'mad5', 'mad6',
       'mad7', 'mad8', 'mad9', 'mad10', 'mad11', 'mad12', 'mad13', 'mad14',
       'mad15', 'mad16', 'mad17', 'mad18', 'mad19', 'mad20', 'mad21', 'mad22',
       'mad23', 'mad24', 'mad25', 'mad26', 'mad27', 'mad28', 'mad29', 'mad30',
       'mad31']
MAN = ['man1', 'man2', 'man3', 'man4', 'man5', 'man6']

def is_exist_in_ncbi(prot_name: str, org_name: str) -> int:
    """Функція для перевірки наявності білка в NCBI
    Args:
        prot_name (str): назва білка
        org_name (str): назва організму
    Returns:
        int: 1, якщо білок знайдено, інакше 0

```

```

"""
handle = Entrez.esearch(
    db='protein', term=f'({prot_name}[Protein Name] OR "Magnetosome protein
{prot_name}"[Protein Name]) AND ({org_name}[Isolate] OR {org_name}[Strain] OR
{org_name}[Organism])')
rec_list = Entrez.read(handle)
id_list = rec_list['IdList']
hdl = Entrez.efetch(db='protein', id=id_list, rettype='gb')
recs = list(SeqIO.parse(hdl, 'gb'))

if len(recs) == 0:
    return 0
else:
    return 1

def search_to_dict(org_name: str) -> dict:
    """Функція для створення словника з наявностями білків у штамі
    Args:
        org_name (str): назва організму
    Returns:
        dict: словник з назвами білків та їх наявністю
    """
    prot_names = ETC + MAIN
    dict_features = {}

    has_macg = 0
    for prot_name in MACG:
        if is_exist_in_ncbi(prot_name, org_name) == 1:
            has_macg = 1

```

```
dict_features['MACG'] = has_macg

has_macgp = 0
for prot_name in MACGP:
    if is_exist_in_ncbi(prot_name, org_name) == 1:
        has_macgp = 1

dict_features['MACGP'] = has_macgp

for prot_name in prot_names:
    dict_features[prot_name] = is_exist_in_ncbi(
        prot_name, org_name)

has_mad = 0
for prot_name in MAD:
    if is_exist_in_ncbi(prot_name, org_name) == 1:
        has_mad = 1

dict_features['mad'] = has_mad

has_man = 0
for prot_name in MAN:
    if is_exist_in_ncbi(prot_name, org_name) == 1:
        has_man = 1

dict_features['man'] = has_man

return dict_features
```

```
def get_sequence(prot_name: str, org_name: str) -> SeqRecord:
```

```
    """Функція для отримання послідовності білка з NCBI
```

```
    Args:
```

```
        prot_name (str): назва білка
```

```
        org_name (str): назва організму
```

```
    Returns:
```

```
        SeqRecord: об'єкт SeqRecord з послідовністю білка
```

```
    """
```

```
    handle = Entrez.esearch(
```

```
        db='protein', term=f'({prot_name}[Protein Name] OR "Magnetosome protein  
{prot_name}"[Protein Name]) AND ({org_name}[Isolate] OR {org_name}[Strain] OR  
{org_name}[Organism])')
```

```
    rec_list = Entrez.read(handle)
```

```
    id_list = rec_list['IdList']
```

```
    hdl = Entrez.efetch(db='protein', id=id_list, rettype='fasta')
```

```
    recs = list(SeqIO.parse(hdl, 'fasta'))
```

```
    return recs[0]
```

```
def calculate_alignment_weight(reference_seq: str, target_seq: str) -> float:
```

```
    """Функція для розрахунку ваги вирівнювання відносно довжини референсної  
    послідовності
```

```
    Args:
```

```
        reference_seq (str): референсна послідовність
```

```
        target_seq (str): цільова послідовність
```

```
    Returns:
```

```
        float: відносна вага вирівнювання
```

```
    """
```

```
    aligner = PairwiseAligner()
```

```

best_alignment = aligner.align(reference_seq, target_seq)[0]
match_count = best_alignment.score
alignment_length = len(reference_seq)

return match_count / alignment_length

```

```

def get_props(dict_features: dict, prot_name: str, org_name: str) -> dict:

```

```

    """Функція для отримання властивостей білків

```

Args:

dict_features (dict): словник з назвами генів та їх наявністю

prot_name (str): назва білка, для якого потрібно отримати властивості

org_name (str): назва організму

Returns:

dict: словник з назвами білків та їх властивостями

```

    """

```

```

rec = get_sequence(prot_name, org_name).seq

```

```

X = ProtParam.ProteinAnalysis(rec)

```

```

dict_features[f'MW_{prot_name}'] = X.molecular_weight()

```

```

dict_features[f'AR_{prot_name}'] = X.aromaticity()

```

```

dict_features[f'IP_{prot_name}'] = X.isoelectric_point()

```

```

if prot_name == 'mamL':

```

```

    ref_seq = str(SeqIO.read("mamL_Cubo/mamL AMB-1.fasta", 'fasta').seq)

```

```

if prot_name == 'mamO':

```

```

    ref_seq = str(SeqIO.read("mamO_Cubo/mamO AMB-1.fasta", 'fasta').seq)

```

```

dict_features[f'RAW_{prot_name}'] = calculate_alignment_weight(

```

```

    ref_seq, str(rec))

```

```
return dict_features
```

```
def data_to_excel(org_name: str, file_name: str) -> None:
```

```
    """Функція для запису даних у Excel файл
```

```
    Args:
```

```
        org_name (str): назва організму
```

```
        file_name (str): назва файлу.
```

```
    """
```

```
    df = pd.read_excel(file_name, sheet_name='Sheet7', index_col=None)
```

```
    data = search_to_dict(org_name)
```

```
    if data['mamL'] == 1:
```

```
        data = get_props(data, 'mamL', org_name)
```

```
    if data['mamO'] == 1:
```

```
        data = get_props(data, 'mamO', org_name)
```

```
    df_test = pd.DataFrame(data, index=[0])
```

```
    df_test['strains'] = org_name
```

```
    df = pd.concat([df, df_test])
```

```
    with pd.ExcelWriter(file_name, engine='openpyxl', mode='a',
```

```
if_sheet_exists='replace') as writer:
```

```
        df.to_excel(writer, sheet_name='Sheet7', index=False)
```

```
if __name__ == '__main__':
```

```
    org_names = ['CR-1', 'FH-1', 'SS-4', 'UT-4', 'LM-5', 'CCP3-SC1AL1',
```

```
                'UT-2', 'LM-5', 'LM-2', 'LEMS', 'KR-1', 'CB-1']
```

```
    file_name = 'Descriptors.xlsx'
```

```
    for org_name in org_names:
```

```
print(f'Processing {org_name}...')  
data_to_excel(org_name, file_name)  
print(f'Finished processing {org_name}.')
```