

ПЕРЕДБАЧЕННЯ ФОРМИ БІОГЕННИХ МАГНІТНИХ НАНОЧАСТИНОК У МАГНІТОТАКСИСНИХ БАКТЕРІЯХ ЗА ДОПОМОГОЮ ДЕРЕВ РІШЕНЬ ТА ВИПАДКОВИХ ЛІСІВ

Хахно К.Ю., Горобець С.В.

КПІ ім. Ігоря Сікорського, khakhnokryl@gmail.com

Abstract

Predicting biogenic magnetic nanoparticle shape is crucial for magnetogenetics. This study explores machine learning, specifically Decision Trees and Random Forests, for efficient shape classification of magnetosome crystals. Cross-validation demonstrates high accuracy, suggesting machine learning's potential for rapid shape prediction in bio-magnetic applications.

Keywords: *biogenic magnetic nanoparticles; magnetogenetics; decision tree classifier; random forest classifier.*

Вступ. Зі стрімким розвитком міждисциплінарних досліджень на межі біології та магнетизму виникла магнітогенетика – інноваційна галузь, що відкриває перспективи неінвазивного керування клітинними процесами за допомогою магнітних полів. Використовуючи магнітні поля для впливу на спеціальні рецептори у клітинах, магнітогенетика обіцяє стати потужним інструментом у нейронауках, біомедицині та багатьох інших галузях. На відміну від оптогенетики, що потребує інвазивного введення світла, та хемогенетики, залежної від хімічних речовин, магнітогенетика пропонує неінвазивний та глибоко проникний метод впливу на клітинну активність, використовуючи магнітні поля, які легко проникають крізь біологічні тканини та не пошкоджують їх [1].

В основі магнітогенетичного підходу лежить використання магніточутливих елементів, здатних перетворювати магнітний сигнал у біологічну відповідь. Особливу увагу привертають біогенні магнітні наночастинки (БМН), зокрема магнітосоми, що виробляються магнітотаксисними бактеріями (МТБ). Ці унікальні органели, що складаються з кристалів магнетиту або грейгіту, оточених мембраною, наділяють МТБ здатністю орієнтуватися та рухатися вздовж геомагнітного поля Землі [2]. Саме магнітосоми, завдяки своїм винятковим магнітним властивостям та біосумісності, розглядаються як перспективні компоненти для розробки магнітогенетичних інструментів.

Розуміння та контроль властивостей магнітних наночастинок є критично важливим для ефективного застосування магнітогенетики. Дослідження [3, 4] підкреслюють вплив магнітних полів на біологічні системи, зокрема на експресію генів. Хоча ці роботи зосереджені на ширшому спектрі впливу магнітних полів, вони підтверджують фундаментальну взаємодію між магнітними полями та живими системами, що є основою для магнітогенетичних підходів. Важливо відзначити, що форма магнітних наночастинок відіграє ключову роль у визначенні їх магнітних характеристик та, як наслідок, їхньої ефективності у магнітогенетичних застосуваннях.

Враховуючи важливість форми біогенних магнітних наночастинок для магнітогенетики, розробка методів точного передбачення цієї характеристики є актуальною науковою задачею. Традиційні методи визначення форми, такі як електронна мікроскопія, є трудомісткими та потребують значних ресурсів. У цьому контексті, застосування методів машинного навчання, зокрема дерев рішень та випадкових лісів, пропонує перспективний шлях для швидкого та ефективного передбачення форми біогенних магнітних наночастинок на основі їх інших характеристик. Метою дослідження є передбачення форми БМН, що дасть змогу оптимізувати процес виділення та модифікації наночастинок для конкретних магнітогенетичних застосувань, а також поглибити розуміння процесів біомінералізації у магнітотаксисних бактеріях.

Матеріали і методи. Для дослідження можливості передбачення форми біогенних магнітних наночастинок за допомогою методів машинного навчання було використано набір даних зі статті [5]. Набір даних включає інформацію про наявність генів *tamA*, *tamB*, *tamE*, *tamI*, *tamK*, *tamM*, *tamP*, *tamQ*, *tamH*, *tamF*, *tamS*, *tamT*, *mms5*, *mms6*, *mms6-L*, *tamC*, *tamD*, *mms36*, *mms48*, *tamG*, *tamR*, *tamL*, *tamO*, *tamN*, *tamX*, *tamZ*, *tamY*, *tamJ*, *mcaA*, *mcaB*, *tamU*, *tamV*, *tamW*, *mad* (від *mad1* до *mad31*) та *man* (від *man1* до *man6*) у різних геномах МТБ та відповідну форму магнітних наночастинок. Додатково було здійснено пошук генів *tamO* та *tamL* у базі даних GenBank NCBI та встановлено властивості білків, які вони кодуєть, за допомогою модуля ProtParam пакету Biopython (молекулярну вагу, ароматичність, ізоелектричну точку).

Для розробки моделей класифікації використовували алгоритми дерева рішень (Decision Tree) та випадкового лісу (Random Forest) з бібліотеки scikit-learn мови програмування Python. Набір даних було розділено на тренувальний та тестовий набори у співвідношенні 80:20 для оцінки узагальнюючої здатності моделей. Для більш надійної оцінки продуктивності застосовували процедуру крос-валідації ShuffleSplit із 3 розділеннями та розміром тестової вибірки 20%. Для дослідження стабільності моделей оцінку точності (accuracy) крос-валідації проводили для 100 різних значень початкового значення (seed).

Для візуалізації розподілу оцінок точності використовували стовпчасті діаграми, побудовані за допомогою бібліотеки seaborn мови програмування Python. Додатково розраховували середнє значення та стандартне відхилення оцінок точності для кількісної характеристики продуктивності та стабільності кожної моделі.

З метою перевірки натренованих моделей на даних, що не входять до початкового набору, було проведено пошук додаткових даних. У результаті було ідентифіковано та обрано два штами магнітотаксисних бактерій, які не були представлені у вихідному наборі даних, використаному для навчання та внутрішньої валідації моделей. Цими штамами були *Magnetococcus sp.* PR-3 та *Candidatus Magnetominusculus xianensis* HCH-1.

Результати та обговорення. Результат побудови дерева рішень для прикладу візуалізовано на рис. 1.

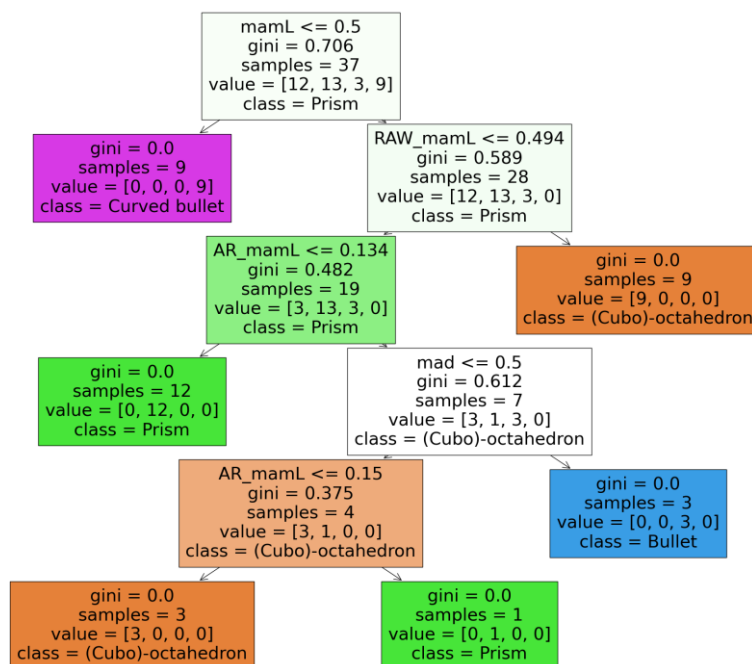


Рис. 1. Графічне представлення дерева рішень.

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. Інші 2 класи (октаедричних та призматичних БМН) неможливо розділити за однією ознакою.

Результати крос-валідації продемонстрували високу точність передбачення форми біогенних магнітних наночастинок як для дерева рішень, так і для випадкового лісу. Середня точність для моделі випадкового лісу склала 91% зі стандартним відхиленням 9%. Модель дерева рішень показала дещо нижчу середню точність 82% та вище стандартне відхилення 12%, що вказує на меншу стабільність порівняно з моделлю випадкового лісу.

Візуалізація результатів за допомогою стовпчастих діаграм (рис. 2) підтвердила варіативність оцінок точності для обох моделей за різних випадкових розділень даних, особливо для дерева рішень.

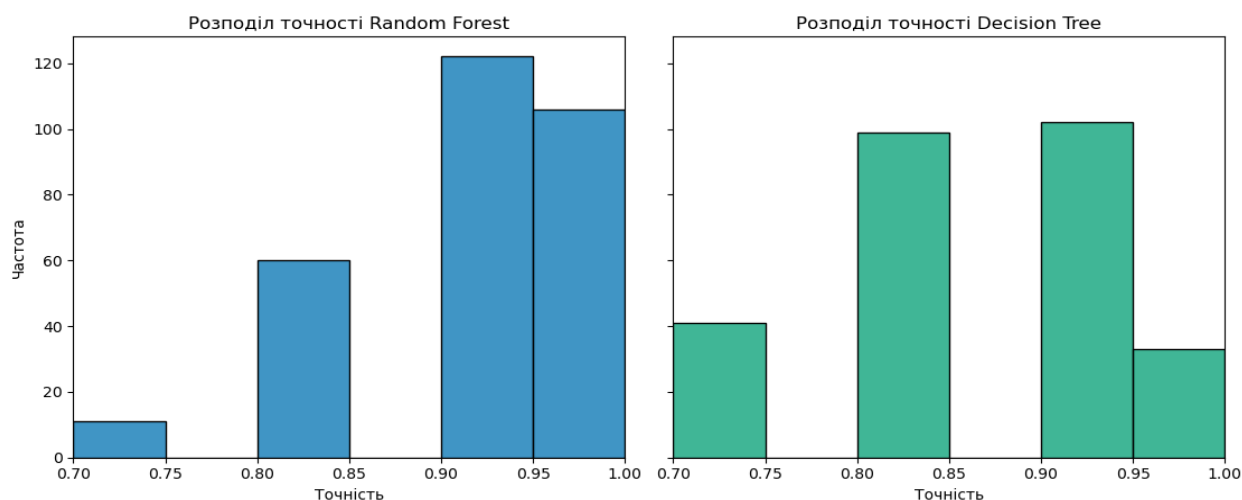


Рис. 2. Порівняння розподілу точності моделей Random Forest та Decision Tree за результатами крос-валідації.

Проте, загальний розподіл оцінок для випадкового лісу був більш концентрованим у області високих значень точності, що свідчить про його кращу узагальнювальну здатність.

Перевірка на нових даних також показала перевагу методу Random Forest. Дерево рішень визначило форму БМН для штаму PR-3 як октаедричну, а випадковий ліс – як призматичну. За літературними даними [6] *Magnetococcus sp.* PR-3 має саме призматичну форму біогенних магнітних наночастинок. А от форму вигнутої кулі БМН штаму HCH-1 змогли передбачити обидві моделі, оскільки дескриптор наявності Mad-білків явно виділяє цей клас магнітотаксисних бактерій.

Отримані результати підкреслюють потенціал використання методів машинного навчання для ефективного та швидкого передбачення форми біогенних магнітних наночастинок. Висока точність класифікації, досягнута за допомогою випадкового лісу, робить цей метод перспективним для застосування у дослідженнях магнітогенетики, даючи змогу прискорити процес відбору та оптимізації наночастинок з потрібними магнітними характеристиками.

Висновки. У даній роботі продемонстровано ефективність застосування алгоритмів машинного навчання, зокрема випадкового лісу та дерева рішень, для задачі передбачення форми біогенних магнітних наночастинок. Модель випадкового лісу показала високу середню точність та стабільність для практичного застосування у галузі магнітогенетики. Використання машинного навчання дає змогу значно спростити та прискорити процес визначення форми наночастинок порівняно з традиційними методами, відкриваючи нові можливості для розробки та оптимізації магнітогенетичних інструментів. Подальші дослідження можуть бути спрямовані на розширення набору даних, дослідження інших алгоритмів машинного навчання та пошук дескрипторів для досягнення ще вищої точності та надійності передбачення.

Список використаної літератури:

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. <https://doi.org/10.1039/d1nr06303k>
2. Blakemore R. Magnetotactic bacteria. *Science*. 1975. Vol. 190, no. 4212. P. 377–379. <https://doi.org/10.1126/science.170679>
3. Effects of Static and Low-Frequency Magnetic Fields on Gene Expression / V. Zablotskii et al. *Journal of Magnetic Resonance Imaging*. 2025. <https://doi.org/10.1002/jmri.29726>
4. Bencardino D. A., Zaitsev M. Commentary on the “Effects of Static and Low-Frequency Magnetic Fields on Gene Expression”. *Journal of Magnetic Resonance Imaging*. 2025. <https://doi.org/10.1002/jmri.29765>
5. Liu P., Zheng Y., Zhang R., Bai J., Zhu K., Benzerara K., Menguy N., Zhao X., Roberts A. P., Pan Y., Li J. Key gene networks that control magnetosome biomineralization in magnetotactic bacteria. *National Science Review*. 2023. Vol. 10. No. 1. <https://doi.org/10.1093/nsr/nwac238>
6. 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. <https://doi.org/10.1016/j.bpj.2014.05.043>