

# ПІДВИЩЕННЯ ТОЧНОСТІ ВИЗНАЧЕННЯ ПЛОЩІ ВОДОЙМ МЕТОДАМИ ЩІЛЬНІСНОЇ КЛАСТЕРИЗАЦІЇ СПЕКТРАЛЬНИХ МАСОК ВОДНИХ ПІКСЕЛІВ

Вінницький національний технічний університет

## Анотація

У роботі проаналізовано недоліки сучасних методів виділення водних об'єктів із супутникових зображень, заснованих на аналізі спектральних характеристик поверхонь. Продемонстровано застосування індексу NDWI для отримання маски водних об'єктів в районі міста Вінниця. Завдяки використанню алгоритмів кластеризації на основі щільності (DBSCAN та HDBSCAN) підвищено точність розрахунку площі водойм за рахунок усунення зашумленості маски водних пікселів. Результати свідчать про ефективність та доцільність застосування алгоритмів кластеризації в задачах розрахунку площі водних поверхонь.

**Ключові слова:** кластеризація, NDWI, площа води, DBSCAN, HDBSCAN.

## Abstract

The paper analyzes the limitations of modern methods for extracting water bodies from satellite imagery based on the analysis of surface spectral characteristics. The application of the NDWI index to obtain a water body mask in the area of Vinnytsia city is demonstrated. By employing density-based clustering algorithms (DBSCAN and HDBSCAN), the accuracy of water body area calculation is enhanced through the elimination of noise in the water pixel mask. The results demonstrate the efficiency and feasibility of using clustering algorithms in tasks related to water surface area calculation.

**Keywords:** clustering, NDWI, water surface area, DBSCAN, HDBSCAN.

## Актуальність дослідження

Моніторинг води відіграє важливу роль в управлінні водними ресурсами. Сучасні методи картографування та обчислення площі водних об'єктів засновані на аналізі оптичних та радіолокаційних супутникових зображень. У цих методах враховуються спектральні властивості води, низьке відношення відбиття та інше. Зокрема, вода має високе поглинання в інфрачервоному діапазоні та високе відбиття в видимому зеленому спектрі. Аналіз цих спектральних відмінностей дозволяє автоматизувати процес визначення меж водойм та розрахунку їх площі на великих територіях. Наприклад, вирішення такої задачі в контексті завдань водогосподарського балансу автори даної роботи разом зі співавторами продемонстрували у своїй статті [1].

Існує низка алгоритмів [2], які для виділення водних об'єктів оцінюють спектральні характеристики води (наприклад, індекси NDWI, MNDWI, AWEI). Вони базуються на підсиленні контрасту води у порівнянні з іншими типами поверхні, після чого застосовується певне значення порогу, що відповідає водному пікселю. Ці індекси прості у реалізації, проте їх недоліком є залежність від якості зображення: тіні, кут освітлення, сезонність, хмарність. Також, як зазначають автори статті [3], ці методи помилково позначають водою поверхні з низьким альбедо (низькою відбивною здатністю), такі як тіні від висотних будівель, мокре асфальтове покриття та темні дахи, що особливо відчутно при аналізі зображень міських забудов.

Метою даної роботи є підвищення точності розрахунку площі водойм (на прикладі р. Південний Буг та озер у межах м. Вінниця) шляхом використання індексу NDWI та застосування алгоритмів щільнісної кластеризації DBSCAN та HDBSCAN. Такий підхід дозволяє виділити цілісні водні об'єкти та відфільтрувати міський шум.

## Результати дослідження

Програмна реалізація використання індексу NDWI в парі з алгоритмами кластеризації виконана мовою Python із використанням бібліотек Scikit-learn і розміщена у хмарному середовищі Kaggle [4].

На рисунку 1 зображена ділянка, з якої досліджувалася поверхня водних об'єктів.



Рис. 1. Супутникове зображення міста Вінниці та околиць

Для обчислення індексу NDWI використовується значення видимого зеленого спектру та ближній інфрачервоний діапазон:

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

Головною складністю у визначенні водних пікселів є вибір коректного порогового значення. Якщо значення буде занадто низьке, то більшість пікселів, що дійсно відповідають воді, будуть визначені правильно, але загальний результат буде сильно зашумлений через міську забудову. Якщо ж поріг встановити занадто великим, то більшість міського шуму видалиться, але й втраяться тонкі ділянки річки, невеликі притоки та прибережні пікселі. На рисунку 2 зображена маска водних пікселів з пороговим значенням 0.05, що емпірично було визначене як компромісне.



Рис. 2. NDWI з порогом значенням 0.05

На рисунку 2 чітко видно озера та русло Південного Бугу, проте і присутні “зайві” пікселі, наприклад дахи деяких будівель.

Для очищення водної маски від шуму були використані алгоритми кластеризації DBSCAN та HDBSCAN. Вони належать до щільнісних методів кластеризації та здатні виділяти геопросторові об'єкти

довільної форми, маркуючи ізольовані пікселі як шум. Головною відмінністю DBSCAN від HDBSCAN є те, що перший працює з фіксованим радіусом сусідніх пікселів, а другий динамічно адаптується до змінної щільності, що на практиці дозволяє краще зберігати тонкі та протяжні озера та русла річок.

Алгоритм DBSCAN був застосований з такими параметрами: радіус сусідства – 3.5, мінімальна кількість точок для формування ядра кластера – 30. Додатково було відфільтровано кластери, менші за 120 пікселів, щоб прибрати великі об'єкти, які були визначені помилково як вода. На рисунку 3 представлений результат застосування DBSCAN.

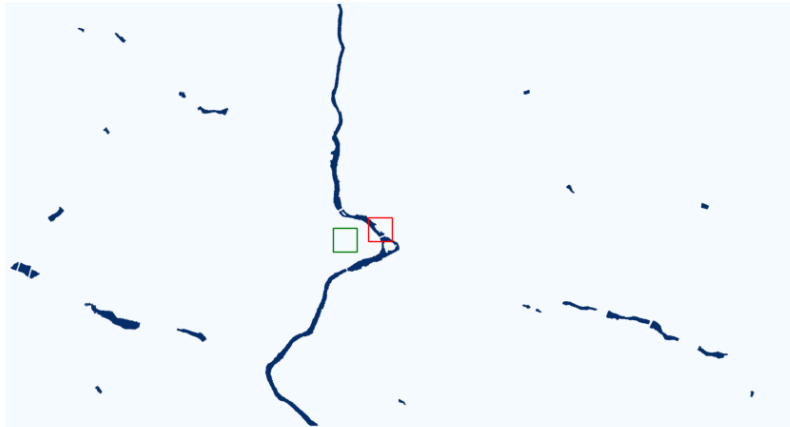


Рис. 3. Застосування DBSCAN поверх маски NDWI

Завдяки фільтрації по розмірності кластерів був відкинтий помилковий об'єкт – дах ТЦ Sky Park. На рисунку 3 він знаходиться в області, позначеній зеленим квадратом. На рисунку 2 після застосування лише NDWI цей об'єкт добре видно. Проте, через особливість алгоритму, тонка ділянка річки Південний Буг, на якій відсутня локальна щільність пікселів, була класифікована як шум (позначено червоним квадратом).

Для алгоритму HDBSCAN було встановлено: мінімальний розмір кластера – 5, поріг, який гарантує, що кластери не розділяться далі, – 1.5. Він використовується для створення великої кількості малих кластерів, для можливості наступної фільтрації за розміром. Після використання HDBSCAN кластери, що мають розмірність меншу за 120 пікселів, були теж прибрані. Результат представлений на рисунку 4.

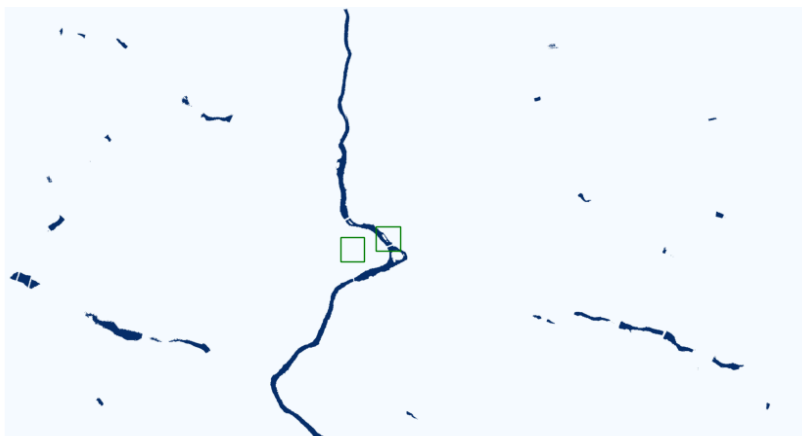


Рис. 4. Застосування HDBSCAN поверх маски NDWI

Можна бачити, що застосування HDBSCAN разом з фільтрацією вирішило одночасно дві проблеми: і очищення маски від великих зайвих об'єктів, і збереження витягнутої форми водойм.

В таблиці нижче міститься кількісний аналіз та порівняння продуктивності обох алгоритмів разом з сирим NDWI. Варто зазначити, що точна площа води на досліджуваній ділянці невідома. Результати

оцінюються за ступенем фільтрації міського шуму в контрольних точках, збереженням зв'язності русла річки та присутності озер довільної форми після застосування алгоритмів кластеризації.

Таб. 1. Порівняння площі водних об'єктів, отриманих різними методами

Метод	Додатковий час на виконання, с	Фінальна кількість кластерів	Кількість пікселів	Площа, км <sup>2</sup>
NDWI	–	–	47887	2.349
DBSCAN	0.3352	28	42338	2.077
HDBSCAN	12.2044	33	43767	2.147

Як видно з таблиці, у порівнянні з DBSCAN, застосування HDBSCAN дозволило зберегти додаткові п'ять кластерів. Водночас обчислювальна складність побудови ієрархічного дерева в HDBSCAN потребує майже на порядок більше часу, що може бути критичним у системах прийняття рішень, які аналізують великі площі місцевості в реальному часі.

### Висновки

В роботі розглянуто застосування одного з найпоширеніших спектральних водних індексів, NDWI, для отримання маски водних пікселів на ділянці в межах міста Вінниця. Особливість цього алгоритму в тому, що для визначення, чи є конкретний піксель водою, використовуються спектральні властивості поверхонь. Оскільки деякі поверхні мають схожий альbedo з водою, алгоритм може давати хибно позитивні результати, що особливо відчутно при аналізі супутникових зображень міських забудов.

Для очищення маски водних пікселів від шуму були застосовані алгоритми кластеризації на основі щільності: DBSCAN та HDBSCAN. Розглянуті переваги та недоліки цих методів: проведено порівняння швидкості їх застосування, продемонстровано можливість адаптації до різних форм водних об'єктів. Емпірично отримано оптимальні вхідні параметри для цих алгоритмів, що дозволило відфільтрувати маску води від зайвих об'єктів та відкоригувати фінальну площу водної поверхні на ділянці.

Результати свідчать, що застосування методів щільнісної кластеризації для задач підрахунку площі водних об'єктів є ефективним та доцільним.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. А. О. Наюк, С. М. Крижановський, І. М. Штельмах, і О. О. Войцеховська, “Метод визначення площі дзеркал водних об'єктів водогосподарської ділянки з використанням супутникових зображень”, Вісник ВПІ, вип. 2, с. 71–78, Квіт. 2026. URL: <https://doi.org/10.31649/1997-9266-2026-185-2-71-78>
2. Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M., & Feng, X. “Satellite Detection of Surface Water Extent: A Review of Methodology”. Water 14, no. 7, p. 1148, 2022. URL: <https://doi.org/10.3390/w14071148>
3. G. Kaplan, U. Avdan. “Object-Based Water Body Extraction Model Using Sentinel-2 Satellite Imagery.” European Journal of Remote Sensing 50 (1): pp. 137–143, 2017. URL: <https://doi.org/10.1080/22797254.2017.1297540>
4. A. Naiuk. NDWI usage together with (H)DBSCAN. Kaggle. URL: <https://www.kaggle.com/code/andriinaiuk/ndwi-usage-together-with-h-dbscan> (date of access: 09.06.2026)

**Наюк Андрій Олексійович** – аспірант кафедри системного аналізу та інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: [anaiuk.stu@gmail.com](mailto:anaiuk.stu@gmail.com)

**Євгеній Миколайович Крижановський** – канд. техн. наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: [kruzhan@gmail.com](mailto:kruzhan@gmail.com)

**Naiuk, Andrii O.** – PhD in System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: [anaiuk.stu@gmail.com](mailto:anaiuk.stu@gmail.com)

**Kryzhanovsky, Evgeniy M.** – candidate of technical sciences, associate professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: [kruzhan@gmail.com](mailto:kruzhan@gmail.com)