

МЕТОДИ МОНІТОРИНГУ ТА АНАЛІЗУ ЗДОРОВ'Я ЛІСІВ ЗА СУПУТНИКОВИМИ ДАНИМИ

В. Е. Іорданова^{1,а}, Г. О. Яйлимова¹

¹ Навчально-науковий Фізико-технічний інститут

Анотація

У даній роботі розглядаються методи моніторингу здоров'я лісових масивів із використанням супутникових даних Sentinel-2. В основу дослідження покладено аналіз великої кількості даних, що включають багатоспектральні зображення та маски лісів, які містять інформацію про здорові та пошкоджені ділянки. З використанням алгоритмів машинного навчання та вегетаційних індексів був проаналізований стан лісів та порівняно результати роботи моделей.

Ключові слова: Sentinel-2, вегетаційні індекси, SVM, Random Forest, CNN, DNN, MLP, Logistic Regression, XGBoost, Gradient boosting

Вступ

Останнім часом спостерігається значна зміна кліматичних умов. Це призводить до навали шкідливих лісових організмів та пожеж через спекотнішу погоду і, як наслідок, знищення гектарів лісів – легенів нашої планети. Також ліси потерпають від різних хвороб та вирубки зі сторони людини.

Жуки-коріди є серйозною проблемою для лісів. Епізоди екстремальних погодних умов, пов'язані з антропогенними змінами клімату, збільшили частоту та масштаби хвилювань жуків-корідів, що спричинило загибель лісів у всьому світі в останні десятиліття [1]. Нещодавні спалахи жука-коріда в Європі привернули велику увагу, оскільки вони вразили великі ділянки хвойних лісів, спричинивши загибель дерев і зниження продуктивності екосистеми [2].

Лісові пожежі не лише знищують великі ділянки лісів, але впливають на біорізноманіття та безпосередньо на людину.

Тому, зважаючи на всі ці фактори, використання супутникових даних та моделей машинного навчання може значно поліпшити здатність до виявлення хвороб, навали шкідників та лісових пожеж.

Дослідження проводилось відповідно до наукових цілей європейського проекту «Satellites for Wilderness Inspection and Forest Threat Tracking» (SWIFTT).

1. Дані та їх підготовка

1.1. Постановка задачі та опис даних

Задані супутникові зображення, на основі яких треба визначити, чи є ліс хворим чи ні.

Для дослідження пошкоджених лісів навчальною та тестовою територією слугувала територія Украї-

ни. Пілотна територія складається з 505 тисяч гектарів і розташована в трьох різних частинах України – на півдні країни в Херсонській області (3 локації) країни в Херсонській області (3 локації) загальною площею 75 тис. га, на сході в Луганській, Донецькій та Харківській областях (12 ділянок) загальною площею 325 тис. гектарів, та на півночі України у Київській та Житомирській областях (3 локації) загальною площею 105 тис. гектарів [3]. Для роботи були відібрані тестові та навчальні дані. Їх можна побачити на рисунку 1.

Використовуються надані [3] зображення з супутника Sentinel-2 на рисунку 2, що містять інформацію в 12 спектральних каналах. Для аналізу стану лісів враховуються задані маски на рисунку 3, які точно визначають розташування лісових ділянок та маски уражених зон на рисунку 4, де спостерігаються ознаки пошкодження лісів через хвороби або шкідників.

Значення масок наведені в таблиці 1.

Значення маски	Значення
1	Природна пожежа, яка НЕ супроводжувалась санітарною вирубною
11	Лісові пожежі в результаті воєнних дій
2	Підсихання лісу, яке НЕ супроводжувалось санітарною вирубною

Таблиця 1. Значення масок [3]

1.2. Підготовка даних та створення датасету

Загальні маски лісів були додатково оброблені для виокремлення ділянок, засаджених хвойними

^аiordanova.valeriia@lil.kpi.ua

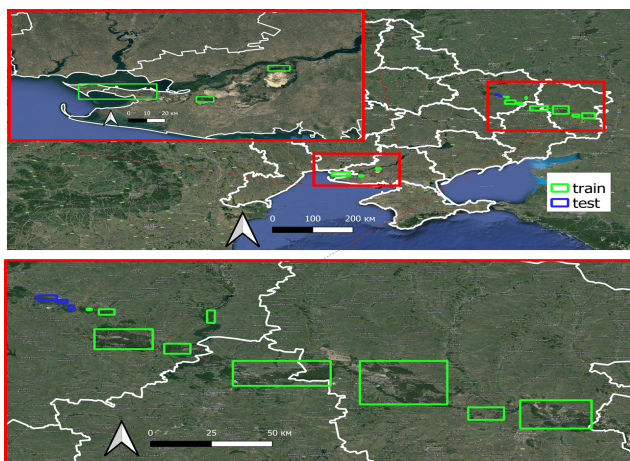


Рис. 1. Досліджувана зона: зелені області – train, сині – test



Рис. 2. Супутникове зображення

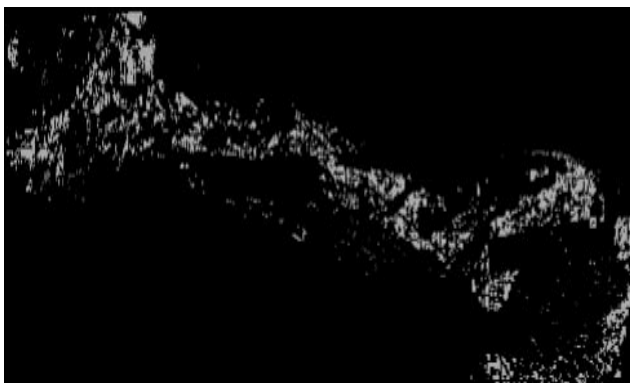


Рис. 3. Маска лісу

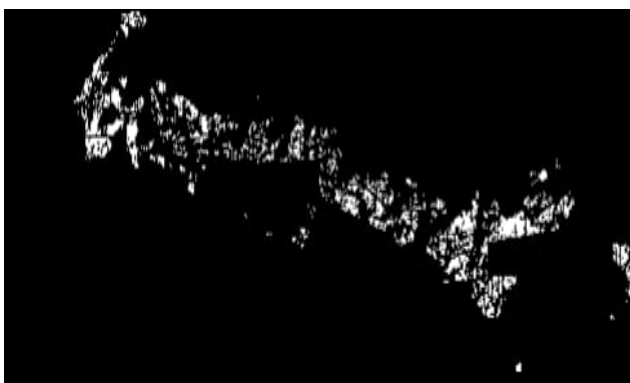


Рис. 4. Маска ураженого лісу

видами. Значення з таблиці 1 було перетворено та отримано, що будь-яке пошкодження приймає значення 1, а якщо пошкодження немає – 0. Створення датасету для аналізу здоров'я лісів було здійснено шляхом попіксельного вилучення даних з супутникових зображень та відповідних масок ураження хвойного лісу. Спочатку для кожного пікселя були визначені географічні координати за допомогою заданого геопросторового перетворення, що дозволило співставити піксельні дані з реальними локаціями на поверхні Землі. Потім інтенсивність усіх спектральних каналів для кожного пікселя була зчитана, і дані були агреговані разом з відомостями про ушкодження від масок. Ця інформація була організована у вигляді датасету, де кожен рядок представляє унікальний піксель, а стовпці відповідають за координати, інтенсивність каналів та стан ураження.

2. Методи моніторингу здоров'я лісів

2.1. Загальний огляд

У контексті цих значущих функцій лісів, моніторинг їх здоров'я стає ключовою задачею не тільки для збереження природних екосистем, але й для захисту людського здоров'я та благополуччя. Методи моніторингу здоров'я лісів, особливо за допомогою супутникових даних, набувають все більшого значення у цьому процесі. Зі статті [4] серед основних методів можна виділити:

- Використання дистанційного зондування. Техніки дистанційного зондування, включаючи аналіз спектральних даних та інші способи сенсорного спостереження, є ключовими для оцінки стану лісів.
- Вимірювання концентрацій SO_2 і NO_2 . Ці показники допомагають виявити зміни у якості повітря, що можуть вплинути на здоров'я лісових екосистем.
- Вегетативні індекси. Індекси дозволяють оцінювати здоров'я рослинності на основі аналізу зображень.

2.2. Вегетаційні індекси

Певні області електромагнітного спектру мають особливий зв'язок із здоровими зеленими рослинами. У видимому спектрі хлорофіл поглинає значну кількість енергії, головним чином для фотосинтезу [5]. Він інтенсивно поглинає світло в синьому та червоному діапазонах для фотосинтезу, в той час як відбиття підвищується в ближньому інфрачервоному спектрі. Це відбивання і поглинання визначаються за допомогою спектральних сенсорів на супутниках і обробляються для розрахунку вегетаційних індексів, які можуть відображати стан здоров'я рослин і допомагати в різноманітних сферах моніторингу.

У цій роботі використовуються такі індекси, як MCARI, RENDVI, NDI45, GNDVI та NDII, бо вони зазвичай показують найвищі точності.

MCARI реагує на концентрацію хлорофілу в листі та коефіцієнт відбиття землі. Зазвичай високі значення MCARI вказують на низький вміст хлорофілу в листі [6].

$$MCARI = ((B5 - B4) - 0.2 \times (B5 - B3)) \times \left(\frac{B5}{B4} \right) \quad (1)$$

GNDVI – це показник фотосинтетичної активності рослинного покриву; найчастіше використовується для оцінки вмісту вологи та концентрації азоту в листі рослин за мультиспектральними даними, які не мають крайнього червоного каналу [7].

$$GNDVI = \frac{B8 - B3}{B8 + B3} \quad (2)$$

NDI45 відображає рівень зеленого листя, дозволяючи оцінити загальний фотосинтетичний потенціал рослинності.

$$NDI45 = \frac{B5 - B4}{B5 + B4} \quad (3)$$

NDII фокусується на вмісті води в листі, що допомагає виявляти водний стрес у рослин та є важливим для моніторингу засух.

$$NDII = \frac{B8 - B11}{B8 + B11} \quad (4)$$

2.3. Методи машинного навчання

Логістична регресія – це статистичний метод, який застосовується для моделювання ймовірності певної класифікації або події, зазвичай для бінарної відповіді (тобто здоровий ліс чи заражений). Вона використовується для визначення ймовірності того, що дана змінна належить до певної категорії.

Random Forest – це алгоритм, який використовує численні дерева рішень для вибудовування загального модельного прогнозу. Кожне дерево навчається на випадково вибраному піднаборі даних, а їхні прогнози комбінуються для підвищення точності та уникнення перенавчання.

Градiєнтний бустінг послідовно вибудовує дерева рішень, кожне з яких намагається виправити помилки попередніх. Цей метод зосереджується на мінімізації втрати функції, яка є мірою того, наскільки добре модель передбачає реальні дані.

SVM – це метод класифікації, який знаходить гіперплощину в багатовимірному просторі, що найкраще розділяє дані на класи.

CNN є передовим підходом глибинного навчання, який широко використовується для вирішення складних завдань. Основна перевага CNN у порівнянні з її попередниками полягає в тому, що вона автоматично визначає відповідні функції без будь-якого контролю з боку людини [8].

DNN – це розширені нейронні мережі з великою кількістю прихованих шарів, що дозволяє моделювати складні не лінійні залежності. Шари у DNN перетворюють вхідні дані шляхом ієрархічного вилучення ознак, починаючи з найпростіших і завершуючи найскладнішими.

MLP – це класична форма нейронної мережі, яка складається з одного вхідного шару, одного або декількох прихованих шарів, і одного вихідного шару. Шари пов'язані зваженими з'єднаннями, які коригуються під час процесу навчання з використанням зворотного поширення помилки. Він може моделювати складні не лінійні залежності між вхідними та вихідними даними.

3. Порівняння та аналіз результатів

Метою цього дослідження було порівняння різних методів машинного навчання. Серед метрик, що беруть участь у порівнянні, – точність (Accuracy), F1-score, час тренування моделей та кількість записів, використаних у навчанні. Результати порівняння можна побачити у таблиці 2.

Модель	Кіл-ть записів	F1	Accuracy	Час, с
Logistic regression	5898692	0.68	0.89	185
XGBoost	5898692	0.76	0.94	20
Gradient boosting	5898692	0.75	0.93	3691
DNN	5898692	0.82	0.95	14439
MLP	5898692	0.79	0.95	3516
CNN	5898692	0.71	0.94	14360
SVM	2000000	0.80	0.95	535
Random Forest	2000000	0.53	0.79	91

Таблиця 2. Порівняння моделей машинного навчання

Логістична регресія показала відносно високу точність з невеликим часом тренування, тоді як XGBoost демонструє хороший F1-score та високу точність з усіх моделей при значно меншому часі тренування. Градiєнтний бустінг має схожі показники до XGBoost, але з більшим часом на тренування. DNN та MLP показали високі результати як за F1-score, так і за точністю, але вимагають значно більше часу на тренування в порівнянні з іншими моделями. CNN, хоч і мають трохи нижчу точність та F1-score порівняно з DNN та MLP, також потребують великої кількості часу для тренування. SVM виявилася однією з найбільш точних моделей з порівняно коротким часом тренування. Але слід врахувати, що для SVM та для Random Forest було зменшено кількість записів для навчання, що було пов'язано з доволі довгим процесом тренування. Random Forest має найнижчий F1-score і точність серед усіх моделей.

Висновки

Інфекції короїда, хвороби лісів, пожежі не тільки призводять до втрати великих площ лісу, але й впливають на біорізноманіття, водні ресурси, а також мають значні соціо-економічні наслідки та впливають безпосередньо на людину. З огляду на

ці загрози, точне і своєчасне виявлення хвороб лісів, навали шкідників та інших загроз стає критично важливим для ефективного управління лісовими ресурсами та збереження біорізноманіття.

Отримані результати показують, що моделі машинного навчання мають великий потенціал у виявленні здоров'я лісів. З порівняльного аналізу випливає, що найкращий результат за точністю та часом показав XGBoost, що робить його найбільш ефективним методом для визначення, чи є ліс здоровим чи ні.

Перелік використаних джерел

1. Jaime L., Batllori E., Lloret F. Bark beetle outbreaks in coniferous forests: a review of climate change effects // European Journal of Forest Research. — 2023. — Листоп. — Т. 143, № 1. — С. 1—17. — ISSN 1612-4677. — DOI: [10.1007/s10342-023-01623-3](https://doi.org/10.1007/s10342-023-01623-3). — URL: <http://dx.doi.org/10.1007/s10342-023-01623-3>.
2. Andresini G., Appice A., Malerba D. : An eXplainable Framework to Map Bark Beetle Infestation in Sentinel-2 Images // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — 2023. — Т. 16. — С. 10050—10066. — ISSN 2151-1535. — DOI: [10.1109/jstars.2023.3312521](https://doi.org/10.1109/jstars.2023.3312521). — URL: <http://dx.doi.org/10.1109/JSTARS.2023.3312521>.
3. A multimodal dataset for forest damages detection using machine learning / H. Yailymova, B. Yailymov, Y. Salii, V. Kuzin, A. Odruzhenko, S. Sydorenko, A. Shelestov, N. Kussul // IEEE International Geoscience and Remote Sensing Symposium. — Athens, Greece, 2024. — Лип. — Accepted.
4. Abas A. A systematic literature review on the forest health biomonitoring technique: A decade of practice, progress, and challenge // Frontiers in Environmental Science. — 2023. — Січ. — Т. 11. — ISSN 2296-665X. — DOI: [10.3389/fenvs.2023.970730](https://doi.org/10.3389/fenvs.2023.970730). — URL: <http://dx.doi.org/10.3389/fenvs.2023.970730>.
5. Improving Forest Detection Using Machine Learning and Remote Sensing: A Case Study in Southeastern Serbia / I. Potić, Z. Srdić, B. Vakanjac, S. Bakrač, D. Đorđević, R. Banković, J. M. Jovanović // Applied Sciences. — 2023. — Т. 13, № 14. — ISSN 2076-3417. — DOI: [10.3390/app13148289](https://doi.org/10.3390/app13148289). — URL: <https://www.mdpi.com/2076-3417/13/14/8289>.
6. Nagler P., Daughtry C., Goward S. Plant Litter and Soil Reflectance // Remote Sensing of Environment. — 2000. — Лют. — Т. 71, № 2. — С. 207—215. — ISSN 0034-4257. — DOI: [10.1016/S0034-4257\(99\)00082-6](https://doi.org/10.1016/S0034-4257(99)00082-6). — URL: [http://dx.doi.org/10.1016/S0034-4257\(99\)00082-6](http://dx.doi.org/10.1016/S0034-4257(99)00082-6).
7. Soft.Farm. Vegetation Indices: NDVI, EVI, GNDVI, CVI, True Color. — 2022. — URL: <https://www.soft.farm/en/blog/vegetation-indices-ndvi-evi-gndvi-cvi-true-color-140> ; Accessed on: 25.04.2024. Available online.
8. Recent Advances in Convolutional Neural Networks / J. Gu [та ін.]. — 2015. — DOI: [10.48550/ARXIV.1512.07108](https://doi.org/10.48550/ARXIV.1512.07108). — URL: <https://arxiv.org/abs/1512.07108>.