

БРИКСІН В. О., к.т.н., доцент кафедри інформаційних технологій, Український державний університет залізничного транспорту,

ЗАРИЦЬКИЙ А. М., аспірант кафедри інформаційних технологій, Український державний університет залізничного транспорту



Визначення ефективності використання методів глибокого навчання для виявлення полів технічних культур за допомогою аналізу супутникових зображень

Анотація. У статті описано дослідження методів комп'ютерного зору, які використовують для розпізнавання супутникових зображень. У дослідженні зазначено, що згорткові нейронні мережі (CNN), UNet і Mask R-CNN є найефективнішими моделями для сегментації супутникових зображень. Ці моделі допомагають виділити поля сільськогосподарських культур із необроблених супутникових зображень. Інший метод, згаданий у статті, передбачає класифікацію з використанням RGB-зображень чи зображень із різних спектральних діапазонів, засновану на трансферному навчанні з інших доменів, таких як ImageNet.

Ключові слова: комп'ютерний зір, розпізнавання зображень, згорткові нейронні мережі (CNN), сегментація зображень, трансферне навчання.

Вступ.

Тема дослідження обумовлена актуальністю завдань визначення термінів визрівання посівів технічних культур та обсягів урожаю, зокрема в контексті застосування штучного інтелекту на основі аналізу супутникових зображень. Великі обсяги даних і неспроможність повного і своєчасного їх отримання потребують упровадження ефективних методів аналізу для вдосконалення планування логістики, передбачення та мінімізації ризиків, а також максимізації прибутків.

За традиційних методів обробки даних стикаються з викликами, пов'язаними з неможливістю своєчасного отримання та обробки даних про обсяги посівних кампаній, запланованих і реалізованих великою кількістю не пов'язаних між собою організацій, що призводить до обмежень у точності та швидкості обрахунків. Використання сучасних моделей аналізу зображень за допомогою штучного інтелекту дає змогу швидко надати впорядковану інформацію для ухвалення обґрунтованих рішень на основі використання доступних джерел супутникових зображень.

Визначення мети та завдання дослідження.

Метою дослідження є визначення ефективності використання методів глибокого навчання для виявлення полів технічних культур за допомогою аналізу супутникових зображень, а також визначення найбільш перспективного підходу.

Основна частина дослідження.

Нижче наведено опис і дослідження декількох запропонованих підходів, пов'язаних з оцінюванням площі полів технічних культур, на основі аналізу супутникових зображень із використанням сучасних методів штучного інтелекту, зокрема комп'ютерного зору (computer vision) [1-3].

Перший підхід використовує глибинне навчання та методи контрольованої сегментації для аналізу супутникових зображень, що дає змогу точно виділяти поля. Найефективнішою групою моделей аналізу зображень за допомогою штучного інтелекту для нашого дослідження є підходи глибинного навчання (deep learning [4]) для комп'ютерного зору. Вони найбільш корисні для першого етапу виявлення полів. Таку задачу можна розв'язати за допомогою повністю згорткової нейронної мережі (fully convolutional neural network [5]) або UNet [6] чи Mask R-CNN [7], що є методами контрольованої сегментації зображень (supervised image segmentation) (рис. 1).

© БРИКСІН В. О., ЗАРИЦЬКИЙ А. М., 2025

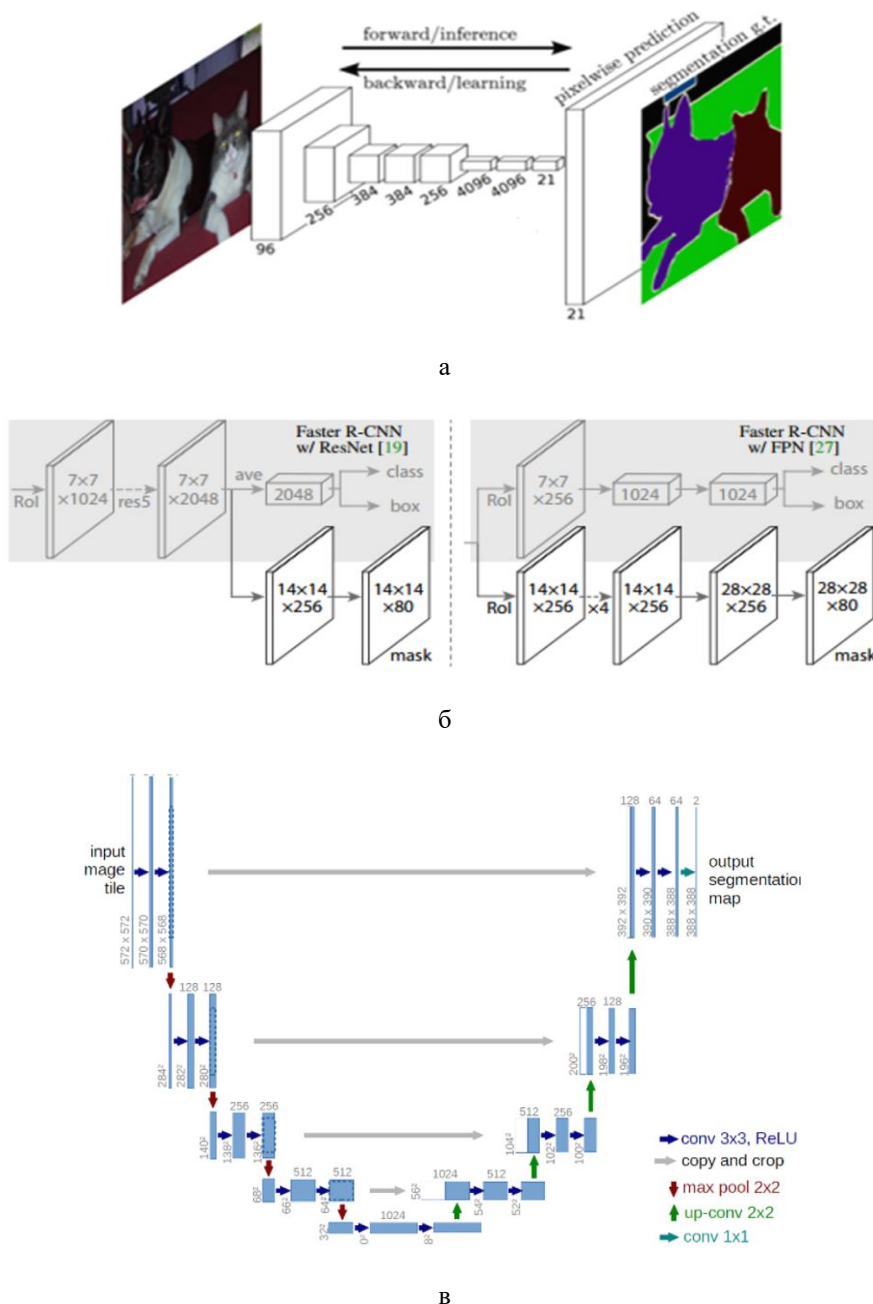


Рис. 1. Методи контрольованої сегментації зображень:
а – Convolutional neural network; б – Mask R-CNN; в – U-Net

Усі ці моделі сфокусовані на піксельному витягу об'єктів/областей поля із необроблених (у нашому випадку супутникових) зображень, які можна отримати з різних агрегаторів супутникових зображень. Для такого підходу необхідний позначений набір даних, що має бути репрезентативним (описувати більшість варіантів змін зображень із різними кутами обертання, різними умовами навколишнього середовища, змінами освітлення протягом циклу вирощування та врожайності та ін.) і містити достатню кількість зразків. Для деяких ефективних архітектур глибокого навчання потрібно тисячі позначених полів

(наприклад PASCAL VOC dataset [8]) із полігонами. Зазвичай цей набір даних можна розширити за допомогою певних методів нарощення (augmentation).

Це основна частина такого підходу, і вона є найбільш складною. Після неї застосовують типові задачі, такі як оцінювання площі на піксель (використовуючи масштаб, відомі перепади висоти та ін.). Після цього можна легко підсумувати всі площі. Але враховуючи те, що, по-перше, ми маємо реальні площі за багато років (тобто вибірка історичних даних є доволі великою), а також можемо перевіряти супутникові дані в динаміці для підвищення точності, то можна зробити висновок, що таке рішення не є

оптимальним. По-друге, усі отримані в результаті гектари не завжди рівнозначні з точки зору врожайності (рис. 2).

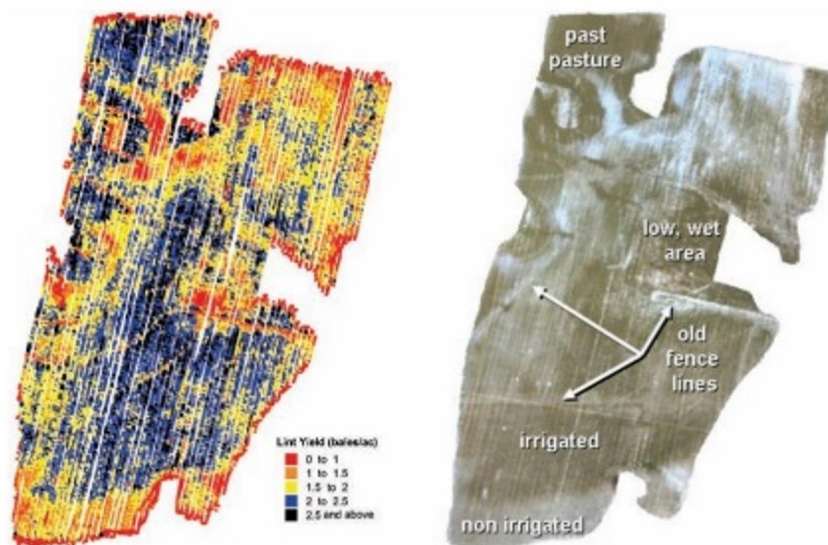


Рис. 2. Приклад невідповідності результатів використання моделі відносно реальних факторів впливу на врожайність

Отже, усі отримані в результаті гектари не дорівнюють реальній площі чи площі, оціненій у результаті поділу на врожайність. Тому на фінальному етапі краще додавати до суми гектарів або їх розподілу певний комбінований коефіцієнт, за допомогою якого можна відкорегувати отримані дані для співставлення з реальними історичними даними.

З іншого боку, навчання на такому невеликому наборі даних може призвести до значних відхилень у навчанні: модель навчатиметься певним ознакам, які є загальними лише для невеликих підмножин і не є

відмінностями від інших типів полів чи супутникових об'єктів, а також які можуть не характеризувати поля вибраної технічної культури. Існує певна можливість використовувати інше супутникове подання якості зразків із дистанційного зондування, але вони також мають бути перевірені на більшому наборі даних. Іншим шляхом може бути знаходження певних часових інтервалів (з 25-тижневого циклу врожаю), у межах яких можуть бути деякі відмінності як у RGB-зображеннях, так і записих з інших частотних діапазонів (рис. 3).



Рис. 3. Приклад RGB-зображень із різних часових інтервалів із відмінностями

Існує **другий підхід**, за якого обробляють усі області, що включають поля вибраної технічної культури, використовуючи деякий пороговий рівень імовірності на основі певного діапазону частот, спектрів RGB чи коефіцієнтів інтервалів. Для цих

областей можна провести агрегацію (із можливістю трансферного навчання (transfer learning) із поточних натренованих коефіцієнтів (estimators), адже існує гіпотеза, що коефіцієнти, які можуть виділяти типові форми з несупутникових або немаповних зображень (наприклад ImageNet [9, 10]), можуть бути

використані в цьому випадку). Згодом цю агрегацію можна співставити з історичними даними про площі чи об'єднати в певну згорткову (convolutional) архітектуру з неперервною (лінійною) активацією для вихідних даних.

Однак цей підхід не настільки ефективний, як попередні методи, оскільки навіть для задач класифікації (що є простішою задачею, ніж виявлення відмінностей між класами) глибокі нейронні мережі (DNN) [11] забезпечують близько 80-90 % точності за наявності великих наборів даних і достатнього часу на навчання. Окрім того, всередині виявлених областей будуть інші поля, які матимуть свої власні ознаки, відмінні від наших результатів. Кількість історичних даних (навіть із 1978 року, коли супутники ще навіть не були доступні) недостатня для запропонованого рішення.

Основні вимоги або важливі моменти для застосування цього підходу:

- достатня кількість позначених полів;
- агрегація великої кількості полів/поверхонь потребує багатьох зразків (що відповідає рокам спостережень);

Порівняння підходів за критеріями

Критерій	Підхід 1: глибинне навчання з контрольованою сегментацією	Підхід 2: порогова агрегація з використанням коефіцієнтів
Точність	Висока (за умови великого позначеного набору даних)	Середня (80-90 % у класифікаційних задачах)
Вимоги до даних	Велика кількість репрезентативних даних	Необхідність довготривалих історичних даних
Гнучкість	Чутливий до змін умов знімання (кут, освітлення)	Використовує спектральний аналіз, але може містити хибні об'єкти
Обсяг обробки	Потрібне глибоке навчання, потребує часу та ресурсів	Агрегація може бути швидшою, але менш точною
Можливість нарощування	Підтримує аугментацію для розширення набору даних	Додаткове навчання може допомогти, але ефект обмежений
Складність реалізації	Висока, потребує сегментаційних моделей	Нижча, але агрегація потребує аналізу точності
Якість сегментації	Висока за достатнього набору даних	Може бути низькою через відмінності полів

Валідація гіпотези. Гіпотеза: бавовняні поля можна відрізнити за значеннями пікселів спектральних зображень і нормалізованого диференційного індексу вегетації (NDVI)*.

*NDVI (Normalized Difference Vegetation Index) розраховують на основі видимого та ближнього

• наявність супутникових даних за вибрані історичні періоди;

• очікувана продуктивність гірша, ніж у попереднього методу;

• агрегація супутникових даних потребує багато часу;

• сегментація може бути виконана з низькою якістю (через подібність і/або суттєві відмінності між полями різних технічних культур).

Ці рішення є найбільш чіткими та мають практичне обґрунтування, проте існують інші гіпотези та підходи.

Детальне порівняння підходів за критеріями наведено в таблиці.

інфрачервоного світла, що відбиває рослинність. Здорова рослинність поглинає більшість видимого світла, яке на неї падає, і відбиває велику частину ближнього інфрачервоного світла. Нездорова чи розріджена рослинність відбиває більше видимого світла і менше ближнього інфрачервоного (рис. 4).

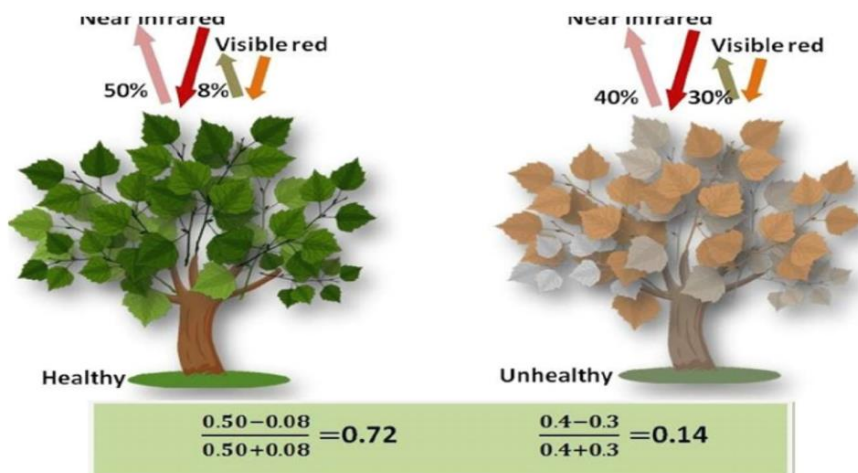


Рис. 4. Принцип, на якому базований NDVI індекс

Контрольні точки було нанесено на мапу (рис. 5).

Дані: набір контрольних точок (Ground Truth Points (GTP)) із їхніми координатами. 83 контрольні точки (51 – бавовна, 32 – інші культури, такі як помідори, кукурудза тощо).

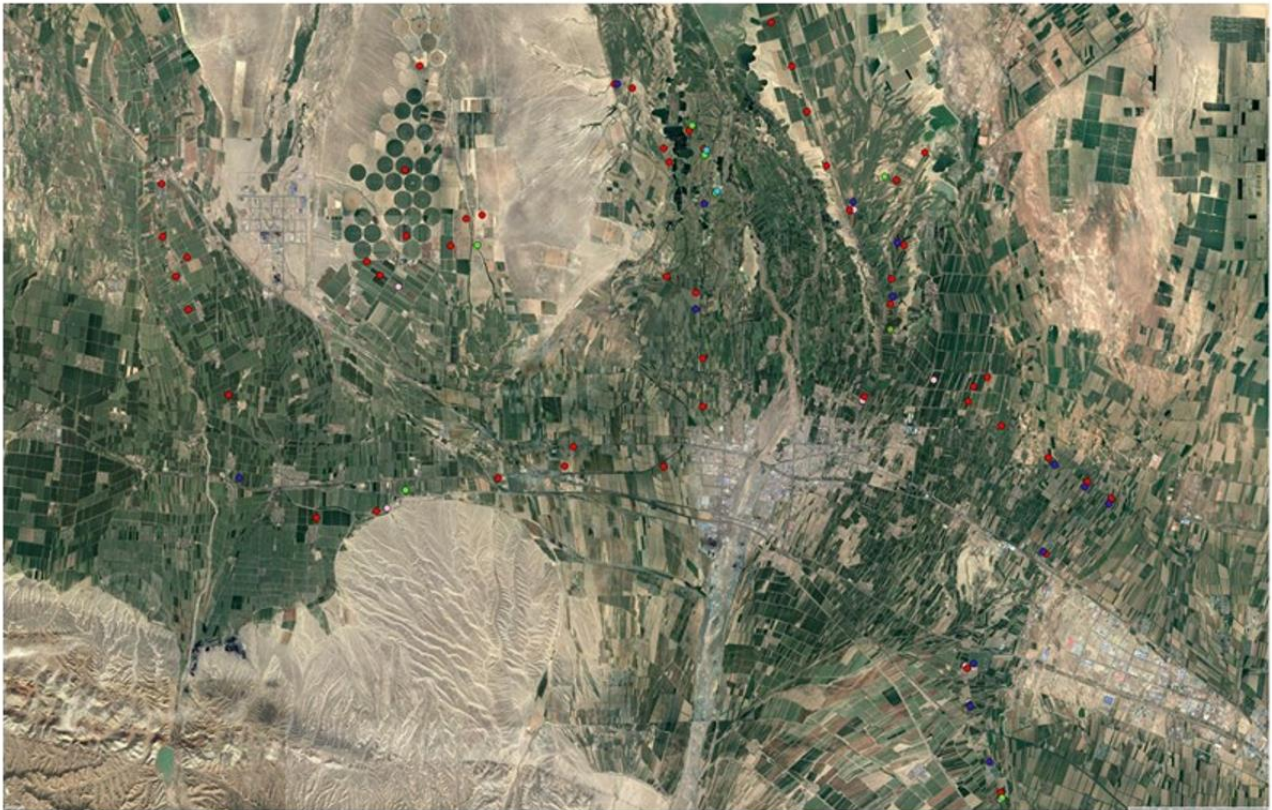


Рис. 5. Контрольні точки (GTP), нанесені на мапу Google Maps

Було вибрано супутник Landsat-8 як основне джерело зображень (рис. 6) та отримано 19 зображень Landsat-8 за 2021 рік у такий спосіб, що:

- кожне зображення містить 11 спектральних діапазонів;

- кожне зображення має приблизну роздільну здатність 8000x8000 пікселів;
- кожен піксель відповідає площі 30 м².

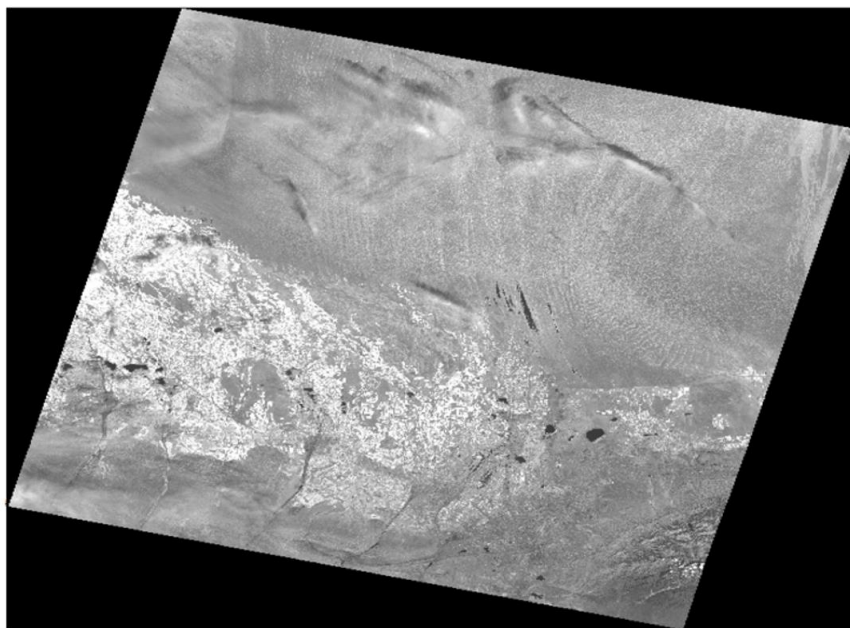


Рис. 6. Приклад зображення, отриманого з Landsat-8

Попередня обробка (рис. 7):

- кожна координата була скорегована так, щоб перебувати щонайменше за 150 м від межі поля;
- для кожної з 83 точок було вирізано прямокутники розміром 10x10 пікселів (300x300 м)

навколо кожної точки. Цей процес було виконано для кожного спектрального діапазону;

- зображення було згруповано в папки відповідно до дати отримання.

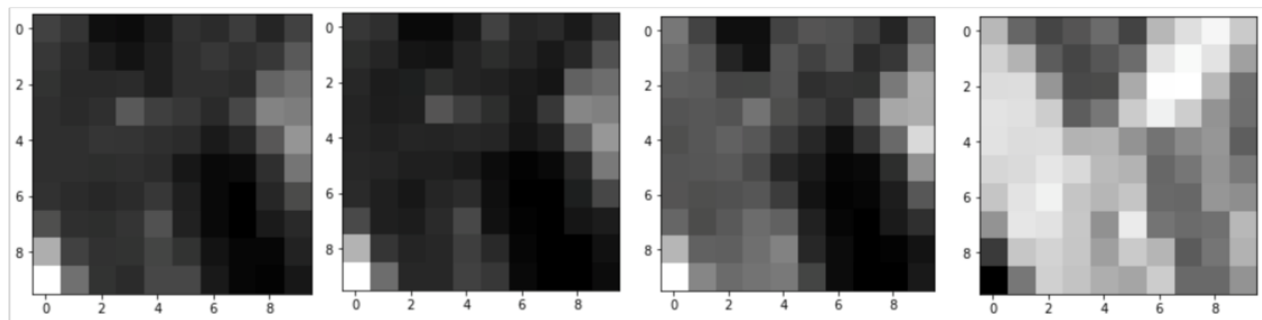


Рис. 7. Приклади патчів для конкретного поля

спектральних діапазонів (11 спектральних діапазонів для 19 дат) (рис. 8);

Використані ознаки:

- кожен вектор ознак (feature vector) було побудовано з використанням значень пікселів для кожної часової позначки і всіх доступних

- до векторів ознак також додано NDVI для кожної часової позначки.

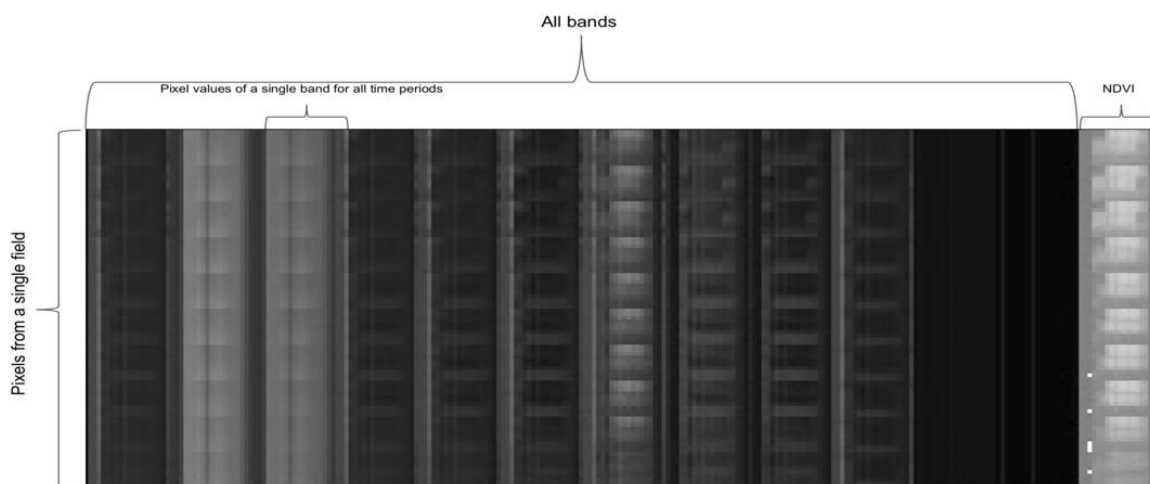


Рис. 8. Візуалізація векторів ознак для конкретного поля

Результати валідації:

- було використано кілька різних алгоритмів машинного навчання для тренування моделі (див. журнал нижче). Найбільш точні результати, за оцінкою F1 (F1 score**) (це середньозважене значення точності та повноти. Тому цей показник враховує як хибно-позитивні, так і хибно-негативні результати. Інтуїтивно зрозуміти цей показник складніше, ніж точність, але оцінка F1 зазвичай є більш корисною, ніж проста точність, особливо якщо у вас є нерівномірний розподіл класів)), було досягнуто за

допомогою техніки градієнтного посилення (gradient boosting);

- найвища поточна оцінка F1 - 0.75 із такими параметрами: точність (precision) = 0.61, повнота (recall) = 0.98 відповідно;
- усі валідації виконували на пікселях із полів, які не було використано в тренувальному наборі.

Детальний журнал (log) валідації:

```

('Accuracy for', 'Decision Tree', 'is', 0.5916666666666667)
('Precision for', 'Decision Tree', 'is', 0.56321839080459768)
('Recall for', 'Decision Tree', 'is', 0.8166666666666666)
('F1 score for', 'Decision Tree', 'is', 0.6666666666666665)
('confusion_matrix for', 'Decision Tree', 'is', array([[220, 380],
[110, 490]]), '\n')
('Accuracy for', 'Random Forest', 'is', 0.6324999999999999)
('Precision for', 'Random Forest', 'is', 0.58054711246200608)
('Recall for', 'Random Forest', 'is', 0.9549999999999999)
('F1 score for', 'Random Forest', 'is', 0.7221172022684309)
('confusion_matrix for', 'Random Forest', 'is', array([[186, 414],
[ 27, 573]]), '\n')
('Accuracy for', 'AdaBoost', 'is', 0.6025000000000000)
('Precision for', 'AdaBoost', 'is', 0.56825749167591566)
('Recall for', 'AdaBoost', 'is', 0.8533333333333333)
('F1 score for', 'AdaBoost', 'is', 0.68221185876082613)
('confusion_matrix for', 'AdaBoost', 'is', array([[211, 389],
[ 88, 512]]), '\n')
('Accuracy for', 'Gradboosting', 'is', 0.6683333333333333)
('Precision for', 'Gradboosting', 'is', 0.60390946502057619)
('Recall for', 'Gradboosting', 'is', 0.9783333333333333)
('F1 score for', 'Gradboosting', 'is', 0.7468193384223919)
('confusion_matrix for', 'Gradboosting', 'is', array([[215, 385],
[ 13, 587]]), '\n')
('Accuracy for', 'Decision Tree', 'is', 0.5333333333333333)
('Precision for', 'Decision Tree', 'is', 0.52493765586034913)
('Recall for', 'Decision Tree', 'is', 0.7016666666666666)
('F1 score for', 'Decision Tree', 'is', 0.6005706134094152)
('confusion_matrix for', 'Decision Tree', 'is', array([[219, 381],
[179, 421]]), '\n')
('Accuracy for', 'Random Forest', 'is', 0.6241666666666666)
('Precision for', 'Random Forest', 'is', 0.57784743991640541)
('Recall for', 'Random Forest', 'is', 0.9216666666666666)
('F1 score for', 'Random Forest', 'is', 0.71034039820166983)
('confusion_matrix for', 'Random Forest', 'is', array([[196, 404],
[ 47, 553]]), '\n')
('Accuracy for', 'Random Forest1', 'is', 0.6391666666666666)
('Precision for', 'Random Forest1', 'is', 0.58546571136131009)
('Recall for', 'Random Forest1', 'is', 0.9533333333333333)
('F1 score for', 'Random Forest1', 'is', 0.72542802790107785)
('confusion_matrix for', 'Random Forest1', 'is', array([[195, 405],
[ 28, 572]]), '\n')
('Accuracy for', 'Random Forest2', 'is', 0.6516666666666666)
('Precision for', 'Random Forest2', 'is', 0.59266802443991851)
('Recall for', 'Random Forest2', 'is', 0.9699999999999999)
('F1 score for', 'Random Forest2', 'is', 0.73577749683944371)
('confusion_matrix for', 'Random Forest2', 'is', array([[200, 400],
[ 18, 582]]), '\n')
('Accuracy for', 'AdaBoost', 'is', 0.5258333333333333)
('Precision for', 'AdaBoost', 'is', 0.51964512040557664)
('Recall for', 'AdaBoost', 'is', 0.6833333333333333)
('F1 score for', 'AdaBoost', 'is', 0.59035277177825773)
('confusion_matrix for', 'AdaBoost', 'is', array([[221, 379],
[190, 410]]), '\n')
('Accuracy for', 'Gradboosting', 'is', 0.7491666666666666)
('Precision for', 'Gradboosting', 'is', 0.592173017507724)
('Recall for', 'Gradboosting', 'is', 0.8583333333333333)
('F1 score for', 'Gradboosting', 'is', 0.73201782304264806)
('confusion_matrix for', 'Gradboosting', 'is', array([[304, 296],
[125, 475]]), '\n')
('Accuracy for', 'Gradboosting1', 'is', 0.7633333333333333)
('Precision for', 'Gradboosting1', 'is', 0.68040983606557374)
('Recall for', 'Gradboosting1', 'is', 0.7966666666666666)
('F1 score for', 'Gradboosting1', 'is', 0.72365482233502533)
('confusion_matrix for', 'Gradboosting1', 'is', array([[310, 290],
[114, 486]]), '\n')
('Accuracy for', 'Gradboosting2', 'is', 0.6783333333333333)
('Precision for', 'Gradboosting2', 'is', 0.61053719008264462)
('Recall for', 'Gradboosting2', 'is', 0.9849999999999999)
('F1 score for', 'Gradboosting2', 'is', 0.75382653061224492)
('confusion_matrix for', 'Gradboosting2', 'is', array([[223, 377],
[ 9, 591]]), '\n')

```

Висновки.

Результати досліджень ефективності використання методів глибокого навчання для виявлення полів технічних культур за допомогою аналізу супутникових зображень показали, що з двох основних підходів для виявлення полів за допомогою супутникових зображень і ШІ більш перспективним є перший підхід (на основі глибинного навчання та контрольованої сегментації). Він є більш точним і структурованим для розв'язання задачі, а також має високу продуктивність за умов наявності великого позначеного набору даних.

Другий підхід, заснований на пороговій агрегації, є менш точним, оскільки використано

узагальнені порогові значення й історичні дані, що не завжди відповідають реальним умовам. Крім того, він може містити випадкові об'єкти, що впливають на якість сегментації.

Зважаючи на переваги та недоліки обох методів, підхід на основі глибинного навчання більш ефективний, оскільки забезпечує точніше виділення полів і має більший потенціал для розширення і удосконалення.

Аналіз початкової гіпотези показав, що найуспішніші результати в класифікації полів були отримані за допомогою алгоритмів градієнтного посилення, досягнувши F1 - оцінки 0,75, з високою чутливістю, але нижчою точністю. Також дослідження показали важливість добре позначеного

набору даних, історичних даних і подальшого вдосконалення методів обробки зображень для досягнення оптимальних результатів.

Список використаних джерел

1. What is computer vision? British Machine Vision Association (BMVA). 2016. (дата звернення: 10.11.2022).
2. Demush R., Hacker Noon. A Brief History of Computer Vision (and Convolutional Neural Networks). February 27, 2019. URL: <https://hackernoon.com/a-brief-history-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3> (дата звернення: 18.09.2024).
3. Linda G. Shapiro, George C. Stockman. Computer Vision, 1st Edition. February 2. 2001.
4. Holdsworth J., Scapicchio M. What is deep learning? International Business Machines Corporation (IBM). June, 17. 2024. URL: <https://www.ibm.com/topics/deep-learning>.
5. Long J., Shelhamer E., Darrel T. Fully Convolutional Networks for Semantic Segmentation. UC Berkeley. 2015.
6. Yao Wei, Zeng Zhigang, Lian Cheng, Tang Huiming. Pixel-wise regression using U-Net and its application on pansharpening. October, 24. 2018.
7. Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick. Mask R-CNN. Facebook AI Research (FAIR). January, 24. 2018.
8. Mark Everingham, Luc van Gool, Chris Williams, John Winn, Andrew Zisserman. The PASCAL VOC project. 2014. URL: <http://host.robots.ox.ac.uk/pascal/VOC/> (дата звернення: 06.05.2023).
9. Deng J., Berg A., Satheesh S., Su H., Khosla A. and Fei-Fei L. ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012). URL: <http://www.image-net.org/challenges/LSVRC/2012/>.
10. Krizhevsky A., Sutskever I. and Hinton G. ImageNet classification with deep convolutional neural networks. NIPS, 2012.
11. Szegedy C., Toshev A. and Erhan D. Deep neural networks for object detection. NIPS, 2013.

Determining the efficiency of using deep learning methods to detect fields of industrial crops using satellite image analysis

Abstract. The article describes the analysis and research of computer vision techniques used for satellite imagery recognition. The research highlights that convolutional neural networks (CNNs), UNet, and Mask R-CNNs are the most effective models for pixel-wise segmentation of satellite images. These models help extract the areas corresponding to crop fields from raw

satellite pictures. Another method described involves classification using RGB images or different spectral bands, relying on transfer learning from other image domains like ImageNet. This approach, however, is less effective due to limited historical data and the complexity of accurately classifying small field areas. The initial hypothesis was that crop fields can be differentiated by the pixel values of spectral images and the normalized difference vegetation index (NDVI), which reflects vegetation health. The most successful results in classifying the fields came from gradient boosting algorithms, achieving an F1 score of 0.75, with high recall but lower precision.

Keywords: computer vision, image recognition, convolutional neural networks (CNN), image segmentation, transfer learning.

Бриксін Володимир, кандидат технічних наук, доцент кафедри інформаційних технологій, Український державний університет залізничного транспорту, Харків, Україна.

E-mail: vladimir.bryksin@gmail.com. ORCID ID: 0000-0002-8036-8811. Scopus Author ID: 57195837705.

Зарицький Андрій, аспірант кафедри інформаційних технологій, Український державний університет залізничного транспорту, Харків, Україна.

E-mail: azarytskyy@gmail.com.

Bryksin Volodymyr, Associate Professor of the Department of Information Technologies, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

E-mail: vladimir.bryksin@gmail.com. ORCID ID: 0000-0002-8036-8811. Scopus Author ID: 57195837705.

Zarytskyy Andriy, PhD Student of the Department of Information Technologies, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

E-mail: azarytskyy@gmail.com.