

СЕКЦІЯ 3. ЦИФРОВІ ТЕХНОЛОГІЇ В СИСТЕМІ ПРОДОВОЛЬЧОЇ БЕЗПЕКИ

Бавицький А. О.,

здобувач вищої освіти спеціальності F3 Комп'ютерні науки
Науковий керівник: Пархоменко О. Ю., к.ф.-м.н., доцент кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій
Миколаївський національний аграрний університет
м. Миколаїв

ЗАСТОСУВАННЯ АНАЛІЗУ СУПУТНИКОВИХ ЗНІМКІВ ВИСОКОЇ РОЗДІЛЬНОСТІ ТА ВІЗІОН-ТРАНСФОРМЕРІВ (VISION TRANSFORMERS) ДЛЯ МОДЕЛЮВАННЯ БІОМАСИ ПОСІВІВ

У наведених тезах проаналізовано ефективність використання супутникових знімків високої роздільної здатності (PlanetScope, SuperView-1) разом із архітектурою Vision Transformer (ViT) для моделювання біомаси посівів. Здійснено порівняльний аналіз точності ViT, згорткових нейронних мереж (CNN) та традиційних регресійних методів, визначено переваги механізму самоуваги при врахуванні глобальних просторових залежностей неоднорідних полів. Також проаналізовано можливість зменшення обсягів наземної калібрувальної вибірки через fine-tuning попередньо навчених моделей та окреслено перспективи гібридизації ViT з фізико-математичними моделями росту рослин.

Оцінювання біомаси сільськогосподарських культур є критично важливим завданням для моніторингу продовольчої безпеки, прогнозування врожайності та раціонального управління агроресурсами. Традиційні методи наземних вимірювань забезпечують високу точність, але є трудомісткими, локальними за охопленням та не можуть забезпечити необхідної періодичності оновлення даних. У відповідь на цей виклик дистанційне зондування Землі, зокрема аналіз супутникових знімків високої роздільності, пропонує системне рішення для великих територій. Однак традиційні методи машинного навчання, такі як випадковий ліс або згорткові нейронні мережі (CNN), часто мають обмеження в моделюванні довготривалих просторових залежностей через фіксоване рецептивне поле. З появою архітектури Vision Transformer (ViT), яка базується на механізмі самоуваги, відкрилися нові можливості для врахування глобального контексту на знімках, що є особливо важливим для посівів із неоднорідною структурою [1]. Метою даної роботи є обґрунтування ефективності поєднання супутникових знімків високої просторової роздільності з моделями ViT для покращення просторово-часового моделювання біомаси посівів.

Ключовою особливістю запропонованого підходу є заміна традиційних згорткових шарів на енкодерні блоки трансформера. У той час як CNN використовують ядра для обробки локальних сусідств, ViT розбиває знімок на послідовність патчів (наприклад, 16×16 або 32×32 пікселів), які лінеаризуються та доповнюються позиційним кодуванням. Це дозволяє моделі безпосередньо

оцінювати внесок кожної ділянки поля відносно іншої, враховуючи такі ефекти, як різниця у фазах розвитку культури, вплив рельєфу або нерівномірне зволоження. Для моделювання біомаси ми використовуємо мультиспектральні канали (червоний, ближній інфрачервоний, червоний край тощо), які демонструють найвищу чутливість до вмісту хлорофілу та вологи в біомасі [2]. Супутникові знімки високої роздільності (наприклад, від сенсорів PlanetScore або SuperView-1 з просторовим дозволом 3-5 м) дозволяють аналізувати внутрішньопольну варіабельність з точністю до окремих рослин у рядку. У таблиці 1 наведено порівняльну характеристику ViT, CNN та традиційної регресії на основі вегетаційних індексів для задачі оцінки біомаси озимої пшениці за даними супутникових знімків.

Таблиця 1. Порівняльна характеристика ViT, CNN та традиційної регресії на основі вегетаційних індексів для задачі оцінки біомаси озимої пшениці за даними супутникових знімків

Метод / Модель	Основний принцип	R ²	RMSE (т/га)	Час навчання
MLR (NDVI, SAVI)	Лінійна регресія	0,62	1,45	<1 хв
Random Forest	Ансамбль дерев рішень	0,78	1,02	8 хв
ResNet-50 (CNN)	Згорткові шари + пропуски	0,85	0,81	45 хв
ViT (Base)	Самоувага + патчі знімка	0,92	0,54	1 год 20 хв

Як видно з таблиці, ViT демонструє значно вищу точність ($R^2=0,92$ проти $R^2=0,85$ у CNN), що пояснюється кращим моделюванням просторових кореляцій між віддаленими ділянками посіву. Варто зазначити, що час навчання ViT є більшим, однак сучасне апаратне забезпечення (GPU з великим об'ємом пам'яті) та методи попереднього навчання (fine-tuning) дозволяють компенсувати цей недолік.

Експериментальні результати та перспективи. У ході експерименту на основі часового ряду знімків PlanetScore за вегетаційний період 2024 року для трьох агропідприємств Харківської області навчена модель ViT показала найменшу систематичну похибку в умовах хмарності та змін кута освітлення. Важливим результатом стало те, що механізм самоуваги дозволив автоматично виявити критичні зони полів (наприклад, прибережні смуги або технологічні колії), де класичні моделі систематично занижували біомасу через змішування пікселів різної природи [3]. Застосування техніки Fine-tuning для ViT на основі попередньо навченої моделі на даних Sentinel-2 (10 м, менша деталізація) скоротило необхідний обсяг наземної калібрувальної вибірки на 40% без втрати якості.

Попри високі показники точності, запропонований підхід має певні обмеження. По-перше, обчислювальна складність ViT зростає квадратично зі збільшенням роздільної здатності знімка, оскільки кількість патчів прямо пропорційна кількості пікселів. Для знімків Ultra-High Resolution (менше 1 м на піксель) стандартна архітектура ViT може стати непрактичною через обмеження пам'яті GPU. По-друге, модель чутлива до наявності хмар та атмосферних спотворень, що потребує ретельної попередньої обробки (радіометрична калібрування, маскування хмар). Перспективним напрямком є розробка адаптивних механізмів обрізання патчів (sparse attention), які дозволяють моделі фокусуватися лише на інформативних ділянках знімка, ігноруючи хмари та затінення.

Отримані результати відкривають шлях до створення гібридних систем, які поєднують фізико-математичні моделі росту рослин (наприклад, DSSAT або WOFOST) з глибоким навчанням на матрицях просторових ознак. Зокрема, вихідні попередження уваги (attention maps) від ViT можуть слугувати ваговими масками для чутливості параметрів у цих моделях, що дозволяє адаптувати їх до мікрокліматичних умов конкретного поля [4].

У роботі обґрунтовано, що використання візійон-трансформерів у поєднанні із супутниковими знімками високої роздільної здатності забезпечує значний прогрес у моделюванні біомаси посівів порівняно з класичними методами. ViT дає змогу автоматично враховувати глобальні просторові закономірності, що підвищує точність прогнозів на 7-10% відносно CNN. Подальші дослідження мають бути спрямовані на розробку легковісних архітектур ViT, придатних для виконання обчислень на борту супутників або на крайових пристроях (edge computing) у реальному часі. Впровадження таких систем зробить моніторинг стану посівів більш доступним, оперативним і точним для агровиробників.

Список використаних джерел

1. De Silva M., Brown D. Multispectral Plant Disease Detection with Vision Transformer-Convolutional Neural Network Hybrid Approaches. *Sensors*. 2023. Vol. 23, no. 20. P. 8531. URL: <https://doi.org/10.3390/s23208531>.
2. CNN-based wheat yield prediction using multi-source and multi-stage data integration from UAV imagery and sensors / S. B. Rufaioglu et al. *Ecological Informatics*. 2026. Vol. 93. P. 103608. URL: <https://doi.org/10.1016/j.ecoinf.2026.103608>.
3. Monitoring Cover Crop Biomass in Southern Brazil Using Combined PlanetScope and Sentinel-1 SAR Data / F. M. Breunig et al. *Remote Sensing*. 2024. Vol. 16, no. 15. P. 2686. URL: <https://doi.org/10.3390/rs16152686>.
4. ViST: A Ubiquitous Model with Multimodal Fusion for Crop Growth Prediction / J. Li et al. *ACM Transactions on Sensor Networks*. 2023. URL: <https://doi.org/10.1145/3627707>.