

Д. Д. ЧУМИЧОВ

аспірант кафедри інформаційних технологій та комп'ютерної інженерії  
Національний технічний університет «Дніпровська політехніка»  
ORCID: 0009-0005-2729-0735

## ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ ПРОСТОРОВИХ ТРАНСФОРМЕРІВ У ЗАДАЧІ НЕЙРОМЕРЕЖЕВОЇ КЛАСИФІКАЦІЇ СІЛЬСЬКОГОСПОДАРСЬКИХ УГІДЬ НА ОСНОВІ ДАНИХ СУПУТНИКІВ SENTINEL

У даній роботі досліджено доцільність використання просторових трансформерних механізмів у задачі нейромережевої класифікації сільськогосподарських угідь на основі мультичасових супутникових даних Sentinel-1 та Sentinel-2. Запропоновано вдосконалення існуючої інформаційної технології шляхом поєднання часової уваги для адаптивної агрегації багаточасових спостережень із просторовим трансформерним модулем, призначеним для моделювання глобальних просторових залежностей у картах ознак.

Існуюча інформаційна технологія передбачає використання нейронної мережі з архітектурою U-Net з енкодером EfficientNetV2-L, попередньо навченої на датасеті ImageNet-21k, із застосуванням механізмів уваги, що агрегують часові ознаки у процесі обробки багатоспектральних і радарних знімків.

У роботі розглянуто декілька варіантів інтеграції часової та просторової уваги: послідовні схеми (із розміщенням трансформера до або після механізму часової агрегації) та паралельні схеми з різними способами об'єднання ознак (конкатенація з  $1 \times 1$  проєкцією, шлюзоване адитивне злиття та зважене сумування). Експериментальні дослідження проведено на вибірці, що охоплює понад 90 тис. земельних ділянок у регіоні Берлін-Бранденбург (Німеччина), із поділом на орні та неорні землі, що забезпечує репрезентативність отриманих результатів.

Отримані результати свідчать, що використання просторового трансформера не завжди призводить до підвищення точності класифікації. Найкращий показник досягнуто у варіанті з включенням трансформера перед механізмом часової уваги, тоді як інші схеми інтеграції демонструють зниження якості. Це вказує на необхідність доцільного та контекстно обґрунтованого використання трансформерних механізмів у задачах класифікації сільськогосподарських угідь, де домінують локальні текстурні та спектральні ознаки.

**Ключові слова:** багатоканальні космічні зображення, Sentinel, класифікація сільськогосподарських угідь, глибинне навчання, нейронна мережа, трансформер, часова увага, просторова увага.

D. D. CHUMYCHOV

Postgraduate Student at the Department of Information Technologies  
and Computer Engineering  
Dnipro University of Technology  
ORCID: 0009-0005-2729-0735

## USING SPATIAL TRANSFORMER TECHNOLOGY IN THE TASK OF NEURAL NETWORK CLASSIFICATION OF AGRICULTURAL FIELDS BASED ON SENTINEL SATELLITE IMAGES

This paper investigates the feasibility of using spatial transformation mechanisms for neural network classification of agricultural land using multi-temporal Sentinel-1 and Sentinel-2 satellite data. It is proposed to improve existing information technology by combining temporal attention for adaptive aggregation of multi-temporal observations with a spatial transform module that models global spatial dependencies in feature maps.

The existing information technology involves a neural network with the U-Net architecture and the EfficientNetV2-L encoder, pre-trained on the ImageNet-21k dataset, that uses attention mechanisms to aggregate temporal features when processing multispectral and radar images.

The paper considers several options for integrating temporal and spatial attention: sequential schemes (with the transformer placed before or after the temporal aggregation mechanism) and parallel schemes with different feature-combination methods (concatenation with  $1 \times 1$  projection, gated additive merging, and weighted summation). Experimental studies were conducted on a sample comprising more than 90,000 land plots in the Berlin-Brandenburg region (Germany), divided into arable and non-arable lands, ensuring the representativeness of the results. The results indicate that using a spatial transformer does not always increase classification accuracy. The best results were achieved with a transformer placed before the temporal attention mechanism, whereas other integration schemes showed a decrease



*in quality. This indicates the need for appropriate and context-based use of transformer mechanisms in agricultural land classification tasks, where local textural and spectral features dominate.*

**Key words:** High-resolution optical satellite images, Sentinel, classification of agricultural lands, deep learning, neural network, transformer, attention pooling, spatial attention.

### Постановка проблеми

Задача класифікації сільськогосподарських угідь за даними дистанційного зондування Землі є складною через високу просторову роздільність зображень, багатоспектральний характер сигналів та наявність багаточасових спостережень із нерівномірною інформативністю. Попри активне впровадження трансформерних архітектур у комп'ютерному зорі, питання приросту точності класифікації, стабільності навчання та обчислювальної доцільності їх застосування в аграрних задачах залишається відкритим. Зокрема, недостатньо досліджено, за яких умов просторові трансформери дійсно покращують результати класифікації, а коли їх застосування призводить до деградації якості через надмірну складність моделі або невідповідність індуктивних упереджень структурі даних. Тож, актуальним є дослідження впливу різних способів інтеграції трансформерних механізмів на точність класифікації сільськогосподарських угідь із використанням супутникових знімків.

### Аналіз останніх досліджень і публікацій

Застосування трансформерних підходів у задачах супутникової сегментації та класифікації останнім часом стало активною областю досліджень. У дослідженні [1] запропоновано трансформер-орієнтований фреймворк для семантичної сегментації віддалених знімків, що поєднує згорткові шари з self-attention для підвищення точності класифікації. Схожий підхід використано у роботі [2], де трансфер-навчання із ієрархічними трансформерами показує значне покращення точності сегментації землекористування, що підтверджує доцільність глобального контекстного моделювання. Для вирішення проблем багатомасштабної сегментації в аграрних регіонах. У статті [3] запропоновано нейронну мережу для виявлення сільськогосподарських угідь за знімками Landsat із використанням ефективного багатомасштабного поєднання ознак.

У роботі [4] показано, що комбінація CNN і трансформерних блоків може суттєво підвищити точність сегментації за рахунок покращення глобальних залежностей і локальних деталей. Автори дослідження [5] продемонстрували застосування механізмів уваги до сегментації сільськогосподарських культур із Sentinel-2, що підтверджує ефективність уваги у задачах агромоніторингу. Комплексний огляд останніх трендів у глибокому навчанні для просторово-часової інтеграції різнорідних даних подано у статті [6], де виділено важливість об'єднання інформації за допомогою механізмів уваги.

Особливу увагу зосереджено на стратегіях інтеграції ознак: у праці [7] представлено гібридну архітектуру, яка комбінує CNN і Transformer для побудови представлень із довготривалими залежностями та мультисенсорним вирівнюванням даних. Результати порівняння CNN і трансформерних моделей у сегментації сільськогосподарських угідь наведено у матеріалі [8], де Swin Transformer показав високу точність на рентабельних даних Sentinel-2.

Систематичний огляд з трансформерів у задачах дистанційного зондування проведено авторами статті [9], який узагальнює тенденції застосування self-attention у сегментації та класифікації, вказуючи як позитивні сторони – глобальне контекстне моделювання, так і виклики – підвищені вимоги до обчислювальних ресурсів.

Ці роботи підтверджують, що сучасні трансформерні архітектури та механізми уваги являють собою потужний інструмент для задач сегментації та класифікації супутникових даних.

### Формулювання мети дослідження

Метою роботи є експериментальне дослідження впливу просторових трансформерних механізмів на точність у задачі нейромережевої класифікації сільськогосподарських угідь на основі мультимасштабних супутникових даних Sentinel.

Для досягнення поставленої мети поставлені та вирішені наступні задачі:

- інтегрувати просторовий трансформер у інформаційну систему нейромережевої класифікації космічних знімків для виявлення орних та неорних земельних ділянок;
- провести систематичне порівняння альтернативних варіантів об'єднання часових і просторових ознак на єдиному датасеті.

### Викладення основного матеріалу дослідження

Технологія нейромережевої класифікації космічних знімків для виявлення неорних земельних ділянок описана у [10] та [11]. Однією з ключових особливостей цієї архітектури є використання механізмів уваги (attention), які дозволяють підсилювати релевантні часові ознаки, агрегація яких підвищує якість класифікації об'єктів на основі часових змін. Узагальнена схема алгоритму зображена на рисунку 1.

При створенні нейромережевої моделі використовуються модулі інтеграції часової інформації та механізми формування виразних просторових представлень на основі трансформерної архітектури. У задачах сегментації на багаточасових даних (наприклад, супутникові знімки з різних дат) ключовою проблемою є наявність

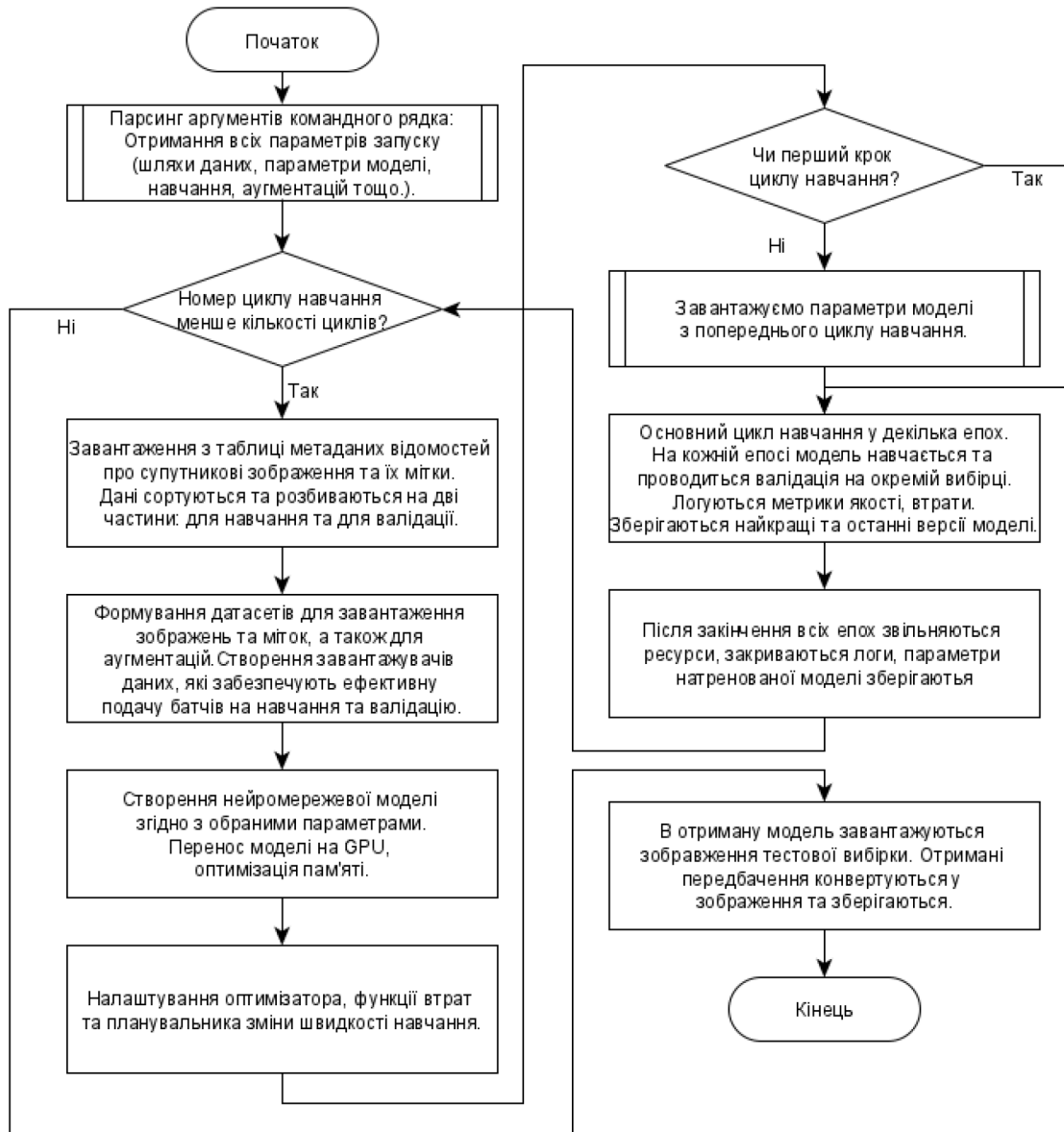


Рис. 1. Схема алгоритму нейромережевої класифікації сільгоспугідь з використанням знімків Sentinel-1 та Sentinel-2

нерівномірної інформативності часових кроків: окремі спостереження можуть містити хмари, сезонні артефакти або пропуски, тоді як інші – чітко відображають цільові об’єкти. Тому агрегування часових ознак простим усередненням або конкатенацією не завжди є оптимальним і може призводити до спотворення корисних сигналів.

Для адаптивної інтеграції багаточасових супутникових даних у роботі використано механізм часової уваги (Attention Pooling), який дозволяє моделі навчатися зважувати внесок окремих часових зрізів у підсумкове просторове представлення. Механізм уваги забезпечує селективне виділення інформативних спостережень і зменшення впливу зашумлених або частково відсутніх даних, що є типовим для мультичасових супутникових серій.

На вхід модуля подається тензор ознак  $X \in R^{B \times T \times C \times H \times W}$ , де  $B$  – розмір батчу,  $T$  – кількість часових зрізів,  $H = W = 256$  – просторові розміри,  $C = 15$  – кількість каналів. Метою модуля є агрегація часової осі  $T$  в єдине просторове представлення  $Y \in R^{B \times C \times H \times W}$  використовуючи увагу по часу окремо для кожної просторової позиції.

Для кожної просторової позиції  $p = (i, j)$ , де  $i \in \{1..H\}$ ,  $j \in \{1..W\}$  та об’єкта  $b$  обчислюється скалярний логіт уваги  $a_{b,i,p}$  через невеликий багаточасовий перцептрон (MLP):

$$h_{b,i,p} = GELU(LayerNorm(W_1 x_{b,i,p} + b_1)), \quad a_{b,i,p} = W_2 h_{b,i,p} + b_2 \quad [12], \quad (1)$$

де  $x_{b,i,p}$  – вхідний вектор ознак для конкретного часу і просторової позиції,  $W_1 \in R^{H_d \times C}$ ,  $b_1 \in R^{H_d}$  – ваги та зсув першого лінійного шару,  $GELU$  – активаційна функція лінійної одиниці гаусової похибки,  $LayerNorm$  – нормалізація

шару MLP по вектору ознак розмірності  $H_d$ ,  $W_2 \in R^{1 \times H_d}$ ,  $b_2 \in R$  – ваги та зсув другого лінійного шару,  $H_d$  – розмір прихованого шару MLP.

Далі обчислюються ваги уваги нормованою експоненційною функцією (softmax) по часовій осі для кожного  $b, p$ :

$$\alpha_{b,t,p} = \frac{\exp(a'_{b,t,p})}{\sum_{\tau=1}^T \exp(a'_{b,\tau,p})}, \quad \sum_{t=1}^T \alpha_{b,t,p} = 1 [13], \quad (2)$$

де,  $\alpha_{b,t,p}$  – нормалізована вага уваги, яка визначає відносну значущість часової ознаки  $t$  для конкретної просторової позиції  $p$  та об'єкта  $b$ . Підсумкове представлення формується як зважена сума ознак по часовій осі:

$$y_{b,p} = \sum_{t=1}^T \alpha_{b,t,p} \cdot x_{b,t,p} [14], \quad (3)$$

де  $x_{b,t,p} \in R^C$  – вектор ознак у момент часу  $t$ . Після цього підсумкові ознаки  $y_{b,p}$  повертається у тензорний формат  $Y \in R^{B \times D \times H \times W}$ .

Для коректної роботи з нерегулярними часовими рядами додатково використовується маска валідності, яка виключає відсутні або непридатні спостереження з нормалізації softmax, забезпечуючи їх нульовий внесок у результат.

Для моделювання глобальних просторових взаємозв'язків у картах ознак застосовуються трансформери. Механізм self-attention дозволяє кожній просторовій позиції адаптивно інтегрувати інформацію з усього поля зору, формуючи контекстуалізовані представлення, що особливо важливо для узгодженої сегментації протяжних об'єктів і структур зі складною геометрією.

У модулі трансформера просторовий вимір  $H \times W$  інтерпретується як послідовність токенів довжини  $L = H \cdot W$ , де кожен токен відповідає окремій просторовій позиції. Оскільки механізм self-attention є інваріантним до перестановок токенів, для збереження геометричної структури зображення до вхідних ознак явно вводиться позиційна інформація двома комплементарними способами.

По-перше, локальний просторовий контекст моделюється за допомогою згорткового позиційного кодування (Convolutional Positional Encoding, CPE), реалізованого як depthwise-згортка розміру  $3 \times 3$  з подальшим резидуальним додаванням:

$$x' = x + DWConv_{3 \times 3}(x) \in R^{B \times C \times H \times W}, [15] \quad (4)$$

Такий підхід вводить локальну індуктивну упередженість, дозволяючи кожній позиції враховувати найближче оточення, не порушуючи семантику каналів і з мінімальними обчислювальними витратами.

По-друге, глобальна координатна структура задається фіксованим двовимірним синусно-косинусним позиційним кодуванням, яке додається до токенів після розгортання простору.

Отримана послідовність токенів подається на трансформерний енкодер, що складається з  $N$  однакових шарів. Ключовим компонентом кожного шару є багатоголова самоувага (Multi-Head Self-Attention, MHSA).

Для кожної пари токенів обчислюється скалярна міра схожості, нормалізована за допомогою softmax, що формує матрицю уваги:

$$A_h = softmax(S_h) \in R^{B \times L \times L}, \quad (5)$$

де  $S_h \in R^{L \times L}$  – матриця попарних оцінок схожості між всіма просторовими позиціями (токенами) в голові уваги  $h$ , а оператор softmax застосовується по кожному рядку, формуючи нормалізовані ваги уваги.

Вихід кожної голови визначається як зважена сума значень, після чого результати всіх голів конкатенуються та проєктуються назад у простір розмірності  $C$ . Вихід MHSA інтегрується з вхідними ознаками через резидуальне з'єднання.

Другим компонентом шару є позиційно-інваріантна двошарова мережа прямого поширення (Feed-Forward Network, FFN) з нелінійністю GELU, яка незалежно обробляє ознаки кожного токена та також інтегрується за резидуальним принципом. Після проходження  $N$  шарів формується глобально контекстуалізована послідовність, яка згортається назад у просторову карту ознак  $y \in R^{B \times C \times H \times W}$ .

Таким чином, трансформерний модуль зберігає форму вхідного тензора, але кожна просторова позиція містить інформацію з усього поля зору.

У задачі сегментації за мультитимовими супутниковими спостереженнями вхідні дані подаються у вигляді послідовності зображень, отриманих у різні моменти часу. Після оброблення енкодером формується багаторівнева піраміда ознак, на рівні  $l$  якої маємо тензор  $F^{(l)} \in R^{B \times T \times C_l \times H_l \times W_l}$ , де  $B$  – розмір батча,  $T$  – кількість часових кроків, а  $(H_l, W_l)$  – просторові розміри карти ознак.

Метою інтеграції є отримання агрегованого просторового представлення  $Y \in R^{B \times C_l \times H_l \times W_l}$ , яке одночасно враховує часову динаміку спостережень і глобальний просторовий контекст. Залежно від порядку та способу їх поєднання розглядається кілька варіантів інтеграції:

## 1. Послідовний варіант А: AttentionPooling → SpatialTransformer

У цьому варіанті спочатку виконується часова агрегація, після чого просторовий трансформер застосовується до вже агрегованої карти:

$$P^{(l)} = \text{AttentionPooling}(F^{(l)}), \quad Y^{(l)} = \text{SpatialTransformer}(P^{(l)}). \quad (6)$$

Перевагою цього підходу є менша обчислювальна складність, оскільки трансформер працює лише з однією картою на батч. Недоліком є ризик втрати дрібних часових деталей до застосування просторової уваги.

## 2. Послідовний варіант В: SpatialTransformer → AttentionPooling

У даному варіанті просторовий трансформер застосовується до кожного часового кроку окремо, а часова агрегація виконується після цього:

$$\tilde{F}^{(l)} = \text{SpatialTransformer}(F^{(l)}), \quad Y^{(l)} = \text{AttentionPooling}(\tilde{F}^{(l)}). \quad (7)$$

Такий підхід краще зберігає просторову структуру перед агрегацією у часі, однак є дорожчим за обчислювальними ресурсами.

Далі розглядатимуться паралельні варіанти об'єднання. У паралельних схемах часова увага та просторовий трансформер обчислюються незалежно, після чого їх результати поєднуються на рівні ознак. Такий підхід дозволяє уникнути передчасної втрати інформації, оскільки ні часова агрегація, ні просторове моделювання не обмежують одна одну на етапі обчислення. Нехай  $P^{(l)} \in R^{B \times C_l \times H_l \times W_l}$  – карта ознак, отримана механізмом часової уваги, а  $Q^{(l)} \in R^{B \times C_l \times H_l \times W_l}$  – карта ознак після просторового трансформера. Розглядаються три способи їх об'єднання.

3. Конкатенація з подальшою  $1 \times 1$  проєкцією (Concat +  $1 \times 1$  conv)

Є найбільш прямолінійним способом інтеграції. У цьому випадку карти  $P^{(l)}$  та  $Q^{(l)}$  конкатенуються по каналному виміру, після чого застосовується лінійна проєкція у вигляді згортки  $1 \times 1$ :

$$Y^{(l)} = \text{Conv}_{1 \times 1}([P^{(l)}, Q^{(l)}]) [16]. \quad (8)$$

Такий підхід забезпечує високу гнучкість комбінування ознак з обох гілок, однак через відмінності у статистиках і семантиці каналів часто призводить до нестабільності навчання та перенавчання.

## 4. Шлюзоване адитивне об'єднання (Gated-add fusion)

Ґрунтується на принципі залишкових з'єднань і розглядає часову агрегацію як базове представлення, яке може бути покращене просторовим трансформером лише за необхідності. У цьому випадку спочатку обчислюється просторово-каналний гейт:

$$G^{(l)} = \sigma(\text{Conv}_{1 \times 1}([P^{(l)}, Q^{(l)}])) [17], \quad (9)$$

після чого формується вихідна карта ознак:

$$Y^{(l)} = P^{(l)} + G^{(l)} \odot Q^{(l)} [17]. \quad (10)$$

Така схема зберігає базове представлення незмінним за відсутності корисного внеску просторового трансформера та забезпечує просторово-каналну адаптивність, що зазвичай підвищує стабільність навчання й узагальнювальну здатність порівняно з простою конкатенацією.

## 5. Зважене сумування з навчуваними коефіцієнтами (Weighted-sum fusion)

Є найбільш обмеженим і водночас найбільш стабільним способом інтеграції. У цьому випадку дві гілки змішуються через навчуваний коефіцієнт  $\alpha^{(l)}$ , спільний для всіх просторових позицій, але індивідуальний для кожного каналу:

$$\alpha^{(l)} = \sigma(\beta^{(l)}), \quad Y = (1 - \alpha^{(l)}) \odot P^{(l)} + \alpha \odot Q^{(l)} [18], \quad (11)$$

де  $\beta^{(l)} \in R^{1 \times C_l \times 1 \times 1}$  – навчуваний параметр. Цей підхід озволяє канално оцінити відносну користь просторового трансформера за мінімальної кількості додаткових параметрів і високої стабільності навчання, однак не забезпечує просторової адаптивності, оскільки ваги злиття є сталими для всіх позицій.

Вплив запропонованих варіантів об'єднання механізму часової уваги та просторового трансформера на точність класифікації було перевірено експериментально на знімках супутників Sentinel. Вхідні супутникові зображення були отримані з відкритого сервісу Copernicus Browser [19] та поділені на окремі фрагменти фіксованого розміру  $256 \times 256$  пікселів.

Для побудови навчальних та тестових масок використовувався відкритий векторний набір даних Field Boundaries for Agriculture (fboa) [20]. Цей набір містить полігональні контури сільськогосподарських угідь разом з атрибутивною інформацією щодо типу землекористування. На основі аналізу атрибутів було відібрано ділянки, які є орними та не орними, після чого їхні векторні межі було перетворено у растрові маски.

У межах дослідження застосовано регіональний піднабір Field Boundaries for Berlin/Brandenburg (Germany), що охоплює близько 90 тис. офіційно зареєстрованих меж сільськогосподарських ділянок на території федеральних

земель Берлін і Бранденбург. З метою спрощення задачі класифікації всі категорії землекористування було агреговано у два узагальнені класи: орні та неорні угіддя (рис. 2).

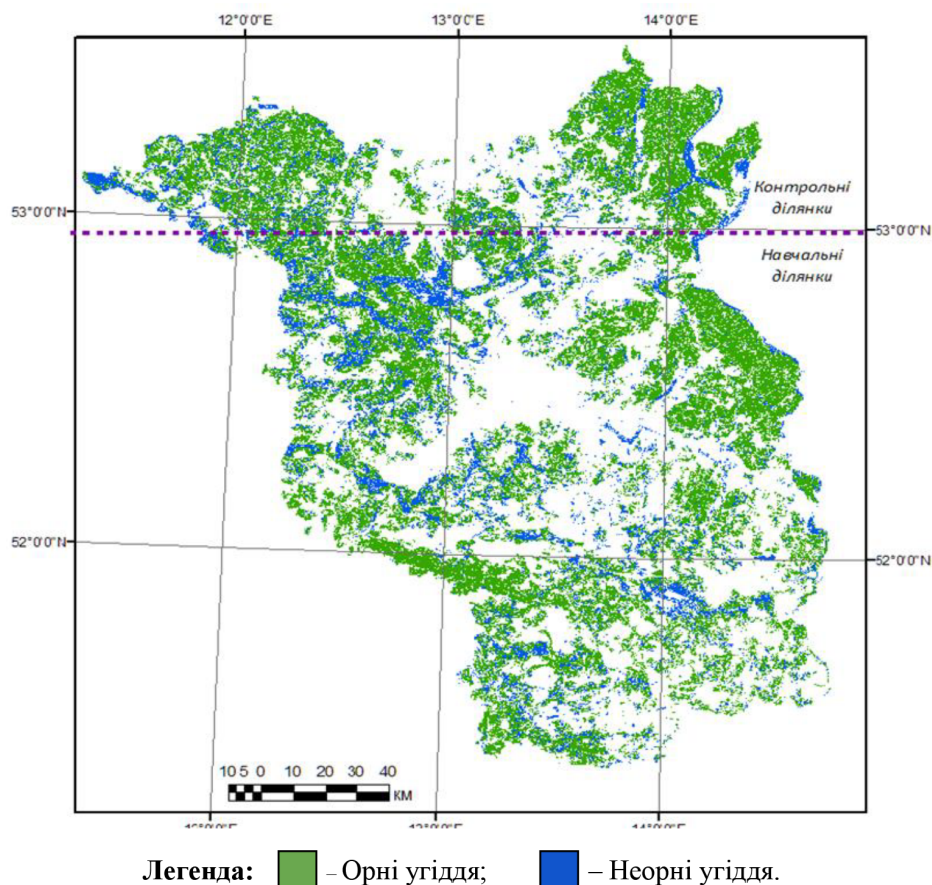


Рис. 2. Досліджувана територія федеральних земель Берлін та Бранденбург

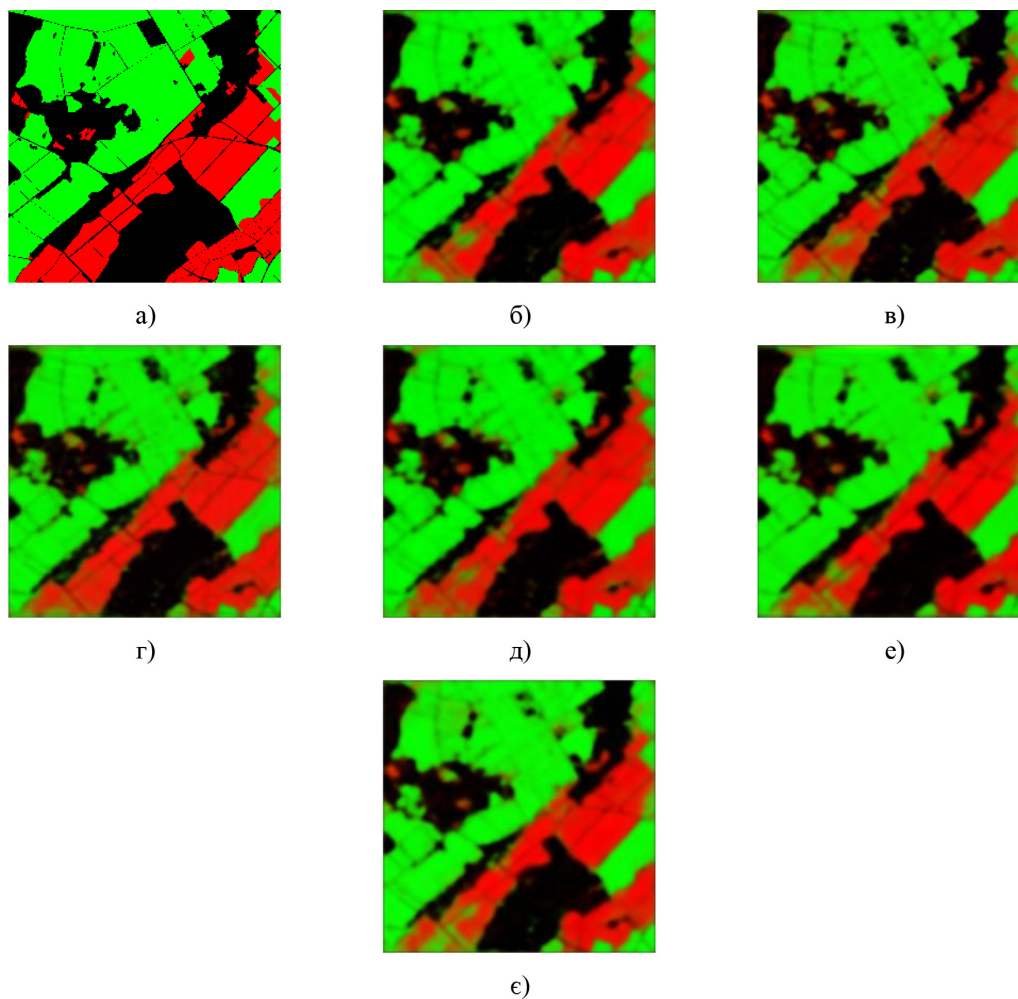
Для аналізу впливу трансформерних компонентів у запропонованій архітектурі було проведено серію експериментів із використанням мультиспектральних і радарних даних Sentinel-2 та Sentinel-1 відповідно. Навчальний набір складався з 3690 тайлів, тоді як для валідації моделі використовувалися 922 зображення. Загалом у межах досліджуваної території було розглянуто 90 770 окремих земельних ділянок, з яких 43 736 відповідали класу неорних угідь. Для забезпечення просторово незалежної оцінки якості моделі приблизно 20 % ділянок, розташованих у північній частині регіону, було виділено в окремий тестовий набір, тоді як решта використовувалася для навчання.

На рисунку 3 наведено порівняльні приклади результатів роботи моделі з різними варіантами інтеграції трансформерних блоків: еталонні класифікаційні маски тестових зображень (а), прогнозні карти, отримані без використання трансформерів (б), з включенням трансформера після механізму часової уваги (в), з включенням трансформера перед механізмом часової уваги (г), а також при паралельному використанні трансформера та механізму часової уваги з різними способами об'єднання ознак – `concat_conv` (д), `gated_add` (е) та `weighted_sum` (є).

Для кількісної оцінки точності класифікації було використано 18154 тестові ділянки. Для кожної ділянки були отримані прогнозні маски, які порівнювалися з еталонними картами тестової вибірки. У разі якщо більш ніж 50 % пікселів в межах ділянки класифікувались правильно, то вся ця ділянка вважалась класифікованою правильно. Основні метрики точності класифікації ділянок з тестового набору у відсотках наведена у таблиці 1.

Отримані результати свідчать, що інтеграція трансформерного блоку в архітектуру моделі не завжди забезпечує підвищення точності класифікації порівняно з базовим варіантом без трансформера.

Найкращий результат (85,32 %) досягнуто у випадку включення трансформера перед механізмом часової уваги, що перевищує показник базової моделі (85,17 %). Використання трансформера після механізму уваги, а також паралельно з ним із різними схемами злиття ознак призводить до зниження точності класифікації. Це вказує на те, що в межах розглянутої архітектури трансформерні механізми у деяких випадках можуть ускладнювати процес навчання. Отримані результати підкреслюють доцільність зваженого застосування трансформерних архітектур у задачах оброблення супутникових знімків та необхідність їх адаптації до специфіки просторово-часових ознак даних дистанційного зондування.



Легенда: ■ – Орні угіддя; ■ – Неорні угіддя.

Рис. 3. Результати роботи моделі:

*a – еталонна маска; б – без використання трансформера; в – трансформер до часової уваги; г – трансформер перед часовою увагою; д, е, є – паралельне об’єднання трансформера й часової уваги (concat\_conv, gated\_add, weighted\_sum)*

Таблиця 1

**Метрики точності класифікації ділянок**

Результат роботи моделі	Точність (Accuracy)	Прецизійність (Precision)	Чутливість (Recall)	F1-score	Специфічність (Specificity)
без трансформера	85.17	84.65	79.92	81.67	91.67
трансформер до часової уваги	85.32	84.65	80.30	81.95	91.81
трансформер після часової уваги	85.02	84.11	79.79	81.42	91.65
паралельне об’єднання (Concat + 1 × 1 conv)	84.87	84.51	79.45	81.28	91.48
паралельне об’єднання (Gated-add fusion)	84.81	83.96	79.60	81.24	91.53
паралельне об’єднання (Weighted-sum fusion)	84.07	82.40	79.22	80.50	91.23

**Висновки**

У роботі досліджено вплив просторових трансформерів на якість класифікації сільськогосподарських угідь за супутниковими знімками. Встановлено, що найкращий результат класифікації (85,32 %) досягається у випадку включення трансформерного блоку перед механізмом часової агрегації, що незначно, але стабільно перевищує показник базової моделі (85,17 %). Натомість застосування трансформера після часової уваги, а також його паралельне використання з різними схемами злиття ознак призводить до зниження точності, що свідчить про можливе ускладнення процесу навчання в межах розглянутої архітектури.

Обсяг експериментального матеріалу є достатнім для формування обґрунтованих висновків, оскільки оцінювання виконано на великій кількості незалежних земельних ділянок із просторовим розділенням навчальної та тестової вибірок, що мінімізує ризик просторового витоку інформації та підвищує надійність отриманих результатів.

Погіршення результатів при використанні механізмів просторової уваги у низці конфігурацій пояснюється тим, що для сільськогосподарських угідь визначальними є локальні спектрально-текстурні ознаки та сезонна динаміка, тоді як глобальні просторові залежності часто є слабо вираженими або надлишковими. За таких умов трансформерні механізми можуть моделювати неінформативні далекі взаємозв'язки, що призводить до внесення шуму в представлення ознак. Отримані результати підтверджують, що просторові трансформери не є універсальним засобом підвищення якості класифікації супутникових знімків і потребують зваженого застосування.

### Список використаної літератури

1. Boulila W., et al. A transformer-based approach empowered by self-attention for semantic segmentation in remote sensing. *Heliyon*. 2024. № 10(8). e29396. DOI: <https://doi.org/10.1016/j.heliyon.2024.e29396>
2. Chen M., Li L. Hierarchical transfer learning with transformers to improve semantic segmentation in remote sensing land use. *Remote Sensing*. 2025. № 17(2). 290. DOI: <https://doi.org/10.3390/rs17020290>
3. Chen H., He G., Peng X., Wang G., Yin R. A multi-scale feature fusion deep learning network for the extraction of cropland based on Landsat data. *Remote Sensing*. 2024. № 16(21). 4071. DOI: <https://doi.org/10.3390/rs16214071>
4. Chen X., Li D., Liu M., Jia J. CNN and transformer fusion for remote sensing image semantic segmentation. *Remote Sensing*. 2023. № 15(18). 4455. DOI: <https://doi.org/10.3390/rs15184455>
5. Gao M., Lu T., Wang L. Crop mapping based on Sentinel-2 images using semantic segmentation model of attention mechanism. *Sensors*. 2023. № 23(15). 7008. DOI: <https://doi.org/10.3390/s23157008>
6. Lian Z., Zhan Y., Zhang W., Wang Z., Liu W., Huang X. Recent advances in deep learning-based spatiotemporal fusion methods for remote sensing images. *Sensors*. 2025. № 25(4). 1093. DOI: <https://doi.org/10.3390/s25041093>
7. Gong Y., Chen C., Zheng Y. Hybrid deep learning model for multi-source remote sensing data fusion: integrating DenseNet and Swin Transformer for spatial alignment and feature extraction. *Informatica*. 2025. № 49(24). DOI: <https://doi.org/10.31449/inf.v49i24.8395>
8. Xu H., Song J., Zhu Y. Evaluation and comparison of semantic segmentation networks for rice identification based on Sentinel-2 imagery. *Remote Sensing*. 2023. № 15(6). 1499. DOI: <https://doi.org/10.3390/rs15061499>
9. Wang R., Ma L., He G., Johnson B. A., Yan Z., Chang M., Liang Y. Transformers for remote sensing: a systematic review and analysis. *Sensors*. 2024. № 24(11). 3495. DOI: <https://doi.org/10.3390/s24113495>
10. Чумичов Д., Нікулін С. Технологія нейромережевої класифікації космічних знімків для виявлення неорних земельних ділянок. *Електротехнічні та інформаційні системи*. 2025. № 108. С. 114–127. DOI: <https://doi.org/10.32782/EIS/2025-108-15>
11. Чумичов Д., Нікулін С. Використання контурів об'єктів супутникових знімків Sentinel для класифікації сільськогосподарських земель за допомогою нейронних мереж. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*. 2025. № 61. С. 213–226. DOI: <https://doi.org/10.36910/6775-2524-0560-2025-61-30>
12. Safari P., India M., Hernando J. Self-attention networks in speaker recognition. *Applied Sciences*. 2023. № 13(11). 6410. DOI: <https://doi.org/10.3390/app13116410>
13. Slimani N., Jdey I., Kherallah M. Improvement of satellite image classification using attention-based vision transformer. *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*. 2024. С. 80–87. DOI: <https://doi.org/10.5220/0012298400003636>
14. Gackstetter D., Yu K., Körner M. Self-attention and frequency-augmentation for unsupervised domain adaptation in satellite image-based time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2025. № 224. С. 113–132. DOI: <https://doi.org/10.1016/j.isprsjprs.2025.03.024>
15. Gao Y., Jiang X., Li Z., Song X., Li W. M3LNet: multi-frequency multi-scale multi-modal learning for multisource image classification. *Proceedings of the 2nd Asia Symposium on Image and Graphics*. 2025. С. 85–91. DOI: <https://doi.org/10.1145/3718441.3718454>
16. Liu Y., Guo Y., Georgiou T., et al. Fusion that matters: convolutional fusion networks for visual recognition. *Multimedia Tools and Applications*. 2018. № 77. С. 29407–29434. DOI: <https://doi.org/10.1007/s11042-018-5691-4>
17. Fan Y., Niu L., Liu T. Multi-branch gated fusion network for image quality improvement in maritime perception systems. *Journal of Marine Science and Engineering*. 2022. № 10(12). 1839. DOI: <https://doi.org/10.3390/jmse10121839>
18. Yuan J., Shi Z., Chen S. Feature fusion in deep-learning semantic image segmentation: a survey. *Science and Technologies for Smart Cities*. Cham : Springer, 2022. С. 261–276. DOI: [https://doi.org/10.1007/978-3-031-06371-8\\_18](https://doi.org/10.1007/978-3-031-06371-8_18)
19. Copernicus Browser [Електронний ресурс]. URL: <https://browser.dataspace.copernicus.eu> (дата звернення: 28.11.2025)
20. Field Boundaries for Agriculture (fiboa) [Електронний ресурс]. URL: <https://fiboa.org> (дата звернення: 30.11.2025)

## References

1. Boulila, W., et al. (2024). A transformer-based approach empowered by self-attention for semantic segmentation in remote sensing. *Heliyon*, 10(8), e29396. <https://doi.org/10.1016/j.heliyon.2024.e29396>
2. Chen, M., & Li, L. (2025). Hierarchical Transfer Learning with Transformers to Improve Semantic Segmentation in Remote Sensing Land Use. *Remote Sensing*, 17(2), 290. <https://doi.org/10.3390/rs17020290>
3. Chen, H., He, G., Peng, X., Wang, G., & Yin, R. (2024). A Multi-Scale Feature Fusion Deep Learning Network for the Extraction of Cropland Based on Landsat Data. *Remote Sensing*, 16(21), 4071. <https://doi.org/10.3390/rs16214071>
4. Chen, X., Li, D., Liu, M., & Jia, J. (2023). CNN and Transformer Fusion for Remote Sensing Image Semantic Segmentation. *Remote Sensing*, 15(18), 4455. <https://doi.org/10.3390/rs15184455>
5. Gao, M., Lu, T., & Wang, L. (2023). Crop Mapping Based on Sentinel-2 Images Using Semantic Segmentation Model of Attention Mechanism. *Sensors*, 23(15), 7008. <https://doi.org/10.3390/s23157008>
6. Lian, Z., Zhan, Y., Zhang, W., Wang, Z., Liu, W., & Huang, X. (2025). Recent Advances in Deep Learning-Based Spatiotemporal Fusion Methods for Remote Sensing Images. *Sensors*, 25(4), 1093. <https://doi.org/10.3390/s25041093>
7. Gong, Y., Chen, C., & Zheng, Y. (2025). Hybrid Deep Learning Model for Multi-Source Remote Sensing Data Fusion: Integrating DenseNet and Swin Transformer for Spatial Alignment and Feature Extraction. *Informatika*, 49(24). <https://doi.org/10.31449/inf.v49i24.8395>
8. Xu, H., Song, J., & Zhu, Y. (2023). Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery. *Remote Sensing*, 15(6), 1499. <https://doi.org/10.3390/rs15061499>
9. Wang, R., Ma, L., He, G., Johnson, B. A., Yan, Z., Chang, M., & Liang, Y. (2024). Transformers for Remote Sensing: A Systematic Review and Analysis. *Sensors*, 24(11), 3495. <https://doi.org/10.3390/s24113495>
10. Chumyochov, D., & Nikulin, S. (2025). Neural network classification technology of space images to identify non-arable land plots. *Electrical and Information Systems*, (108), pp. 114–127. <https://doi.org/10.32782/EIS/2025-108-15>
11. Chumyochov, D., & Nikulin, S. (2025). Using contours of objects from Sentinel satellite images for classifying agricultural lands using neural networks. *Computer-integrated technologies: education, science, production*, (61), pp. 213-226. <https://doi.org/10.36910/6775-2524-0560-2025-61-30>
12. Safari, P., India, M., & Hernando, J. (2023). Self Attention Networks in Speaker Recognition. *Applied Sciences*, 13(11), 6410. <https://doi.org/10.3390/app13116410>
13. Slimani, N., Jdey, I. & Kherallah, M. (2024). Improvement of Satellite Image Classification Using Attention-Based Vision Transformer. In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence – Volume 3: ICAART*; ISBN 978-989-758-680-4; ISSN 2184-433X, SciTePress, pp. 80-87. <https://doi.org/10.5220/0012298400003636>
14. Gackstetter, D., Yu, K., Körner, M. (2025) Self-attention and frequency-augmentation for unsupervised domain adaptation in satellite image-based time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 224, pp. 113-132. <https://doi.org/10.1016/j.isprsjprs.2025.03.024>
15. Gao, Y., Jiang, X., Li, Z., Song, X., & Li, W. (2025) M3LNet: Multi-frequency Multi-scale Multi-modal Learning for Multisource Image Classification. In *Proceedings of the 2024 2nd Asia Symposium on Image and Graphics (ASIG '24)*. Association for Computing Machinery, New York, NY, USA, pp. 85–91. <https://doi.org/10.1145/3718441.3718454>
16. Liu, Y., Guo, Y., Georgiou, T., et al. (2018) Fusion that matters: convolutional fusion networks for visual recognition. *Multimed Tools Appl*, 77, pp. 29407–29434. <https://doi.org/10.1007/s11042-018-5691-4>
17. Fan, Y., Niu, L., & Liu, T. (2022). Multi-Branch Gated Fusion Network: A Method That Provides Higher-Quality Images for the USV Perception System in Maritime Hazy Condition. *Journal of Marine Science and Engineering*, 10(12), 1839. <https://doi.org/10.3390/jmse10121839>
18. Yuan, J., Shi, Z., & Chen, S. (2022). Feature Fusion in Deep-Learning Semantic Image Segmentation: A Survey. In: Paiva, S., et al. *Science and Technologies for Smart Cities. SmartCity 360 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 442. Springer, Cham. [https://doi.org/10.1007/978-3-031-06371-8\\_18](https://doi.org/10.1007/978-3-031-06371-8_18)
19. Copernicus Browser. (n.d.). Retrieved November 28, 2025, from: <https://browser.dataspace.copernicus.eu>
20. Field Boundaries for Agriculture (fiboa). (n.d.). Retrieved November 30, 2025, from: <https://fiboa.org>

Дата першого надходження статті до видання: 21.01.2026

Дата прийняття статті до друку після рецензування: 23.02.2026

Дата публікації (оприлюднення) статті: 30.04.2026