UDC 004.8:528.8:633. **doi: 10.32620/reks.2026.1.14**

Artughrul GAYIBOV, Vagif GASIMOV

Baku Engineering University, Khirdalan, Azerbaijan

EXTENDING SPECTRAL INDICES FROM MULTISPECTRAL SATELLITE DATA USING U-NET SEGMENTATION

This study focuses on *developing a unified, reproducible cloud-to-model pipeline for parcel-scale cropland delineation from multispectral Sentinel-2 and Landsat imagery in the Kur–Araz region.* ***The goal of this study*** *is to produce accurate, boundary-faithful, and computationally efficient farmland maps that remain transparent, scalable, and deployable on commodity hardware for use by government agencies, water authorities, and agricultural producers.* ***The tasks to be*** *addressed include the following: specification of the study scope, data sources, and evaluation protocol that integrates pixel-wise and boundary-sensitive accuracy metrics; construction of a multi-index feature stack in addition to surface reflectance bands, followed by screening of features for cross-seasonal stability; design and training of a memory-efficient U-Net architecture with a hybrid loss that simultaneously balances calibration and overlap; and validation of model generalization across different cropping seasons and neighboring subregions of Kur–Araz, with complete provenance tracked from preprocessing through evaluation.* ***The methods used in the study*** *include cloud-native preprocessing in Google Earth Engine, consisting of cloud masking, seasonal compositing, medoid and percentile mosaicking, and stratified patch sampling with spatial blocking. The datasets were exported to TFRecords and used to train a compact U-Net encoder–decoder network with skip connections and a hybrid objective function. Training involved strong normalization, data augmentation to account for agricultural scene variability, overlap-tiling for inference, and three ablation studies to isolate the contribution of individual vegetation indices. Complete logging of assets, random seeds, and parameters supported exact reruns of the pipeline, replication, and auditing.* ***Conclusions. The scientific novelty of*** *the results obtained is as follows: (1) a unified, provenance-complete pipeline was developed that directly integrates Google Earth Engine preprocessing with deep learning model training and boundary-aware evaluation, ensuring full reproducibility and auditability; (2) a resource-efficient U-Net architecture with skip connections and a hybrid overlap-weighted loss was designed and validated, which preserves parcel borders while maintaining calibration, outperforming heavier transformer-based backbones under limited compute budgets; (3) the complementarity of multi-index feature stacks was systematically quantified through ablation studies, demonstrating their role in mitigating spectral saturation and soil brightness effects and improving overlap fidelity. Finally, a standardized evaluation protocol was proposed that combines pixel-wise and boundary-sensitive metrics, allowing robust assessment of farmland delineation performance across seasons and spatial subregions.*

Keywords: *Google Earth Engine; multispectral segmentation; U-Net; spectral indices; farmland boundary delineation; Kur–Araz; reproducible pipeline.*

1. Introduction

1.1. Motivation

Crop identification from satellite data is necessary for precision agriculture, water management, climate adaptation, and food security. Sentinel and Landsat satellites can provide the temporal frequency and spatial resolution required for continuous monitoring; however, stakeholders need timely, parcel-level maps to define crop extent and boundaries for yield forecasting, damage assessment, irrigation planning, and fair allocation of insurance or subsidies. It is in the society's best interest to map cropland accurately using multispectral data.

Multispectral bands and derived indices effectively express comprehensive signals for vegetation, soil and water, and can be analyzed using modern semantic segmentation methods, which generally outperform traditional pixel-based classifiers and benefit from cloud-computing platforms such as Google Earth Engine. Recent developments include attention mechanisms, multiscale skip connections, edge-aware loss functions, and improved masking methods for clouds and composite satellite imagery. However, many challenges remain, including limited feature selection, low reproducibility and scalability low inter-regional generalization, boundary errors in small irregular fields, and/or hardware limitations.

Our main goal is to create a reproducible cloud-to-model pipeline for automated preprocessing and efficient convolutional neural networks (CNNs) using several indices. The cloud-to-model pipeline enables evaluation of

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

field boundaries using boundary-sensitive metrics alongside standard pixel-based scores, providing accurate, scalable, and transparent cropland boundaries for government agencies, water authorities, and producers.

This study integrates geoinformatics, remote sensing, and computer vision into parcel-scale delineation for precision agriculture tasks, including yield forecasting, irrigation management, compliance monitoring, and climate risk assessment. While Sentinel-2 and Landsat satellites provide appropriate spatial resolutions and adequate spectral coverage, no reproducible, efficient and standardized pipeline for consistently mapping season-stable field boundaries across varying soils is available. In this study, we address this gap for the Kur–Araz region using pixel-level mapping models that rely on cloud native preprocessing and supervised segmentation methods. The cloud-to-model approach utilizes complementary spectral indices to discriminate vegetation from soil pixels across seasons. Therefore, it emphasizes both delineation quality and resource-aware deployment.

1.2. State of the art

Remote sensing, in conjunction with contemporary learning algorithms, can provide large-scale information on climate, water, and agricultural services [1], offering a systematic review of how Earth observation and machine learning support sustainable development goals. However, reproducibility, cross-regional comparability, and the practical implementation of research pipelines remain problematic. Heterogeneity of landscapes, fragmented preprocessing of data, and computational limitations that hinder regular map updates could be the causes. These difficulties can be overcome by adopting unified workflows that standardize feature generation, sampling, and evaluation, enabling results to be reproduced and audited.

The paper [2] presents Google Earth Engine as a planetary-scale platform for data access and geospatial computation, and it shows that large archives, such as Landsat and Sentinel, can be processed efficiently without local infrastructure. However, many studies restrict Earth Engine to preprocessing, while model training and patch curation occur elsewhere, thereby weakening traceability. This may be due to data egress limits, separate deep learning tooling, and the absence of standardized interfaces between cloud preprocessing and model training. An integrated cloud-to-model pipeline that maintains provenance, sampling logic, and feature definitions consistently from ingestion through evaluation can help overcome these difficulties. The review [3] confirms that Earth Engine enables big-data remote sensing and reduces barriers to national and regional studies. However, the authors note that rigorous benchmarking and reproducible modeling practices still lag the capabilities of the data platform, which again motivates unified designs.

The foundational work on spectral indices established the biophysical basis for vegetation mapping. A previous study [4] presents the normalized difference in the vegetation index and showed that a simple combination of red and near-infra-red bands effectively tracks green biomass. However, questions regarding index saturation at high leaf area and sensitivity to background and atmosphere conditions remain. A previous study [5] presented the soil-adjusted vegetation index and showed that partial compensation for soil brightness improves performance in sparse vegetation. However, when a single index is used, unresolved issues include transfer across seasons, soils, and crop types. This may be due to the diverse canopy structures and the non-linear changes in reflectance across phenological stages. Feature stacking, which combines multiple indices and raw bands to capture complementary signals, can be a way to overcome these difficulties.

Deep convolutional encoders with symmetric decoders have become the backbone of semantic segmentation in remote sensing applications. In [6], advanced U-Net architectures with enhanced attention mechanisms are presented, and improved boundary awareness modules and guided fusion significantly enhance segmentation accuracy in remote sensing applications. Recent developments have demonstrated that attention-enhanced U-Net variants can achieve substantial improvements in the mean intersection over union and overall accuracy for land cover classification. However, unresolved issues arise in remote sensing where fields are small, textures are repetitive, and label noise is common. The reasons may be differences between medical images and satellite scenes, scale variance introduced by varying ground sampling distances, and memory pressure when training on large tiles. A way to overcome these difficulties is to design memory-efficient variants together with training protocols that emphasize boundaries, class balance, and robust augmentation.

Attention mechanisms have been proposed to focus learning on salient regions. In [7], attention-enhanced U-Net architectures were presented, and it was shown that improved attention mechanisms can significantly enhance sugar beet and weed segmentation performance in precision agriculture applications. Recent work shows that attention modules improve segmentation accuracy in remote sensing. However, in large-scale settings, additional parameters and computational costs can hinder training on resource-limited hardware, while the risks of overfitting and poor generalization remain. Lightweight attention or selective feature weighting on multi-index inputs may help improve boundary quality without increasing memory demand.

The use of hybrid architecture that combines convolution with transformers is also a possibility. The paper

[8] presents transformer-enhanced U-Net architectures and shows that U-Net-like transformers with boundary enhancement modules provide improved global context modeling, which benefits structure recovery in urban scene segmentation. However, there were unresolved issues related to heavy computation and data hunger, and the reason for this may be patch tokenization that increases memory and the need for large, labeled datasets.

A previous survey [9] showed that integrating remote sensing and machine learning through comprehensive precision agriculture applications can be effective for remote sensing classification. However, most studies focused on scene-level tasks rather than precise boundary segmentation, and the transfer to agricultural mosaics with small parcels was not fully demonstrated. Retaining efficient convolutional decoders while enriching inputs through informative spectral features rather than relying solely on heavier backbones can overcome these difficulties.

Loss design and evaluation protocols strongly influence the ability to learn sharp boundaries. The review [10] provides sophisticated evaluation methodologies for remote sensing applications, and it provides evidence that selective feature extraction methods can demonstrate substantial improvements in terms of mean intersection over union scores when properly configured. However, many remote sensing studies report only pixel accuracy, and unresolved issues include insufficient use of boundary metrics that reflect management needs for field edge placement. This may be due to legacy benchmarking habits and the absence of standardized, easily reproducible evaluation suites that include contour quality.

In summary, the analyzed sources collectively reveal unresolved local problems. Platforms enable scalable preprocessing but are seldom connected to training and evaluation. Spectral indices explain vegetation variability, but reliance on a single index limits robustness across seasons and soils. Convolutional decoders delineate structures but strain memory and blur boundaries without tailored losses. Attention and transformer modules improve context but increase computational cost and rely on large, labeled corpora. All of this supports the assertion that a general unresolved problem is the absence of a unified, efficient pipeline that integrates automated cloud-based preprocessing, multi-index feature stacking, and resource-aware segmentation with boundary-sensitive evaluation for agricultural land cover mapping. Therefore, research devoted to developing such unified and reproducible methods is relevant.

1.3. Objectives and tasks

This research builds on core concepts from multispectral reflectance theory, semantic segmentation, and cloud native geospatial computation. Spectral index theory posits that algebraic combinations of bands amplify biophysical signals while suppressing nuisance factors, such as soil background and atmospheric variation, which motivates the construction of feature stacks from indices alongside raw bands to improve separability across seasons and soils. Convolutional neural networks provide inductive biases of locality and translation equivariance that are well suited to spatially structured farmland patterns; U-Net-style decoders further preserve fine detail through skip connections, which are critical for field boundary delineation. Boundary-aware learning and evaluation recognize that pixels along parcel edges are rare but operationally consequential. Therefore, losses and metrics that weight contours can mitigate majority class bias and reduce boundary blur. Finally, cloud native geospatial computing ensures reproducibility and scalability by executing preprocessing, compositing, and sampling close to the data archives, preserving origin from ingestion to model training. This study aims to develop and validate a unified and reproducible pipeline that integrates automated, cloud-based preprocessing with a memory-efficient U-Net segmentation model leveraging multi-index features and boundary-sensitive learning to improve pixel-wise accuracy and field boundary delineation for multispectral farmland mapping in the Kur–Araz region. This will enable the routine and transparent production of cropland maps on commodity hardware for agricultural planning, irrigation oversight, and climate risk monitoring.

The research methodology comprises the following steps:

1. Study area and data preparation Filter Sentinel-2 L2A imagery over Kur–Araz, apply cloud/cirrus masking, build seasonal composites, select B2–B4, B8, B11, and extract stratified 256×256 tiles; export as TFRecords with full origin.

2. Feature engineering. The feature stack is built by concatenating surface-reflectance bands with NDVI, SAVI, and MSAVI; features are standardized using robust statistics.

3. Model & training. Train a compact U-Net with skip connections and lightweight components; use a hybrid loss balancing calibration and overlap; apply augmentation; perform overlap-tiling at inference.

4. Experimental design. Use spatially blocked train/val/test splits; monitor AUC, accuracy, precision, recall, Dice/overlap; select thresholds on validation and hold fixed on test.

5. Ablations & diagnostics. The marginal value of indices is quantified, and boundary adherence and typical failure modes are analyzed.

6. Analysis & reporting. Present tables/figures for metrics and curves, qualitative mosaics, and error patterns to inform recommendations.

The main objectives of this research are as follows:

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

- Stage 1. Formulating the problem of parcel-scale cropland delineation from multispectral Sentinel-2/Landsat data for the Kur–Araz region; specifying operational requirements, assumptions, study scope, and evaluation protocol that includes pixel-wise and boundary-sensitive metrics (Sections 1.1–1.3, 2);

- Stage 2. Designing an integrated cloud-to-model pipeline in Google Earth Engine: cloud masking, seasonal compositing, medoid/percentile mosaics, and stratified patch sampling; constructing a multi-index feature stack (bands + NDVI, SAVI, MSAVI) and exporting datasets for training (Section 2);

- Stage 3. Developing a memory-efficient U-Net encoder–decoder with skip connections and a hybrid boundary-aware loss; setting up training (normalization, augmentation), overlap-tiling inference, and ablation plans (Section 2);

- Stage 4. Experiments and analyses: learning dynamics, feature ablations, and qualitative mosaics to assess boundary fidelity; interpretation of findings and limitations with practical implications (Section 3);

- Stage 5. Summarizing contributions and developing recommendations, with directions for future research and deployment (Section 4).

The article is structured as follows: Section 1 introduces motivation, state of the art, and objectives and tasks (Subsections 1.1–1.3). Section 2 presents the Materials and methods (study area, preprocessing, feature stack, model, and training protocol). Section 3 reports Results and Discussion, including learning dynamics, ablations, qualitative examples, and limitations. Section 4 provides the Conclusions and future directions.

The scientific aim focuses on developing and validating a pipeline that addresses the unresolved problem identified in the literature: the absence of an efficient, reproducible integration of cloud-based preprocessing, multi-index features, and resource-aware segmentation with boundary-sensitive evaluation. The practical part focuses on the expected benefits of applying the results, namely, enabling agencies and producers to generate accurate and frequently updated farmland maps under realistic compute constraints, supporting decisions on irrigation scheduling, compliance monitoring, and disaster impact assessment.

2. Materials and methods of research

This study aims to develop a unified pipeline for agricultural land cover segmentation from multispectral satellite imagery, integrating cloud-based preprocessing, spectral index feature construction, and a memory-efficient U-Net style model tailored to farmland boundary delineation in the Kur–Araz region.

Integrating automated, provenance-preserving preprocessing on Google Earth Engine [2, 11] with multi-index feature stacks [4, 5] and a lightweight U-Net decoder [6] trained under boundary-aware objectives [10] will improve pixel-wise accuracy and field edge quality under constrained compute relative to single index features, ad hoc preprocessing, and generic loss functions. The hypothesis rests on several assumptions. Spectral indices derived from red, near-infra-red, and shortwave infra-red bands sufficiently capture crop vigor, soil moisture, and canopy structure to distinguish between farmland and non-farmland across seasons. Ground reference polygons accurately represent parcel boundaries at the sensor resolution. Cloud and shadow artifacts can be mitigated through compositing and masking. Spatially stratified validation approximates the transfer to adjacent subregions. Land cover is modeled as a binary segmentation between farmland and non-farmland; crop type differentiation and within-field heterogeneity are not addressed; seasonal composites summarize peak growth conditions rather than modeling full intra-seasonal dynamics. This design choice offers the advantages of simpler preprocessing and lower data volume for constrained memory during training, but it sacrifices temporal phenological signals that would support improved crop discrimination and boundary delineation. The terrain and bidirectional reflectance effects are handled using standard platform corrections rather than bespoke radiative transfer modeling.

Multi-temporal methods that stack images from different growth periods into a single channel for clarity enable models to learn seasonal trajectories of vegetation indices, which are useful for differentiating crop types, characterizing fallow fields, and detecting ephemeral land-use changes. As an exploratory approach to assess the value of temporality, a simplified experiment was performed in which three seasonal composites were used as stacked channel inputs. The performance of this multi-temporal model showed slight improvements in test-set measures compared to the baseline single-composite method, with the only notable improvements occurring in scenarios involving a crop rotation with differing planting dates. The multi-temporal composition better preserved field boundaries in late-season images, whereas the single-composite predictions eroded them due to senescence-induced spectral changes. Although the multi-time temporal composition was beneficial, the protocol in our main experiments was to use single-composite data to mirror operational realities, and we needed to prioritize computationally simple models that enable rapid map updates for operational purposes. Future studies could investigate the use of lighter-weight temporal encoders, such as recurrent or attention-based sequence encoders built upon compressed phenology summaries, for capturing intra-seasonal temporal dynamics without excessive memory storage requirements. Similarly, augmentation-based strategies, including perturbing the composite

image date to simulate temporal variability without requiring multi-date training data, have the potential to improve the robustness to the timing of image acquisition.

This study focuses on pixel-level delineation of agricultural land in multispectral satellite imagery from the Kur-Araz region, with a particular emphasis on accurate parcel boundaries under seasonal and radiometric variability. The main hypothesis proposes a cloud-native preprocessing pipeline, multi-index feature stacks, and a memory-efficient U-Net segmentation model that can improve pixel accuracy and boundary quality while remaining deployable on commodity hardware. The study area encompasses irrigated and rainfed agriculture in the Kur-Araz lowlands, a socioeconomic breadbasket characterized by small- to medium-sized parcels, canal networks, and seasonal fallow. The geographic context includes transboundary water management and climate exposure, which demand frequent mapping updates for planning. Sentinel-2 Level-2A surface reflectance imagery (10–20 m) is used as the primary data source due to its spectral coverage and revisit rate. Landsat 8/9 is optionally consulted for gap filling during persistent cloud cover. Fig. 1 shows the spatial extent and administrative overlay. Reference labels consist of digitized parcel boundaries and field-level cropland annotations compiled from institutional datasets and expert photointerpretation.

The spatial resolution criteria for the applications considered are dictated by operational demands for precision agriculture, water management, and compliance monitoring. Sentinel-2's native resolution of 10 m in the visible and surface reflectance is sufficiently fine to identify individual agricultural parcels in the Kur-Araz Region, where fields are typically sized between 0.5 and 5 hectares. This implies that reasonable boundaries can be identified for parcels that exceed 5 - 10 pixels in width, which scale to minimum field-size dimensions of 50 - 100 m in Kur-Araz, where most operational farmland units fall within these dimensions. Regarding irrigation planning and mapping of canal networks within the region, resolution of 10-20 m would be adequate and effective, as broad parent infrastructure can be captured without a considerable increase in computing needs for processing of this data at the regional scale. Although higher resolutions (< 5 m) could yield advancements in the identification of narrow field boundaries and detection of small parcels, this decrease in spatial resolution would significantly increase the associated data volume, processing times, and cost associated with labelling efforts. In this case, this is unlikely without the associated and beneficial outcomes to the primary stakeholders: the farmers/producer in industry and water management authorities. Primary users need timely watershed scale maps, not surveying accuracy to the parcel-level bounding. In contrast, coarser/lower resolution (> 30 m) would not be sufficient to distinguish fields (especially small and irregularly shaped fields) found in this study area with boundary displacement errors that are potentially beyond acceptable tolerances for irrigation scheduling and subsidy allocation programs.

Fig. 1. Study area and administrative context within the Kur–Araz region, showing major irrigation infrastructure and sample label coverage.

The Sentinel-2 surface reflectance collection is filtered for 2023, and images with cloud cover below 1% are selected. B2 (Blue), B3 (Green), B4 (Red), B8 (Near infra-red (NIR)), and B11 (Shortwave infra-red) bands are chosen for their utility in vegetation mapping and agricultural settings. More specifically, B2-B4 captures visible reflectance, making it useful for color composite visualization and soil background identification; B8 (NIR) is critical for vegetation detection given high chlorophyll reflectance; and B11 (SWIR) is sensitive to plant water content and soil moisture, providing the ability to discriminate crop vigor and fallow land. This combination of bands provides the necessary information while being computationally efficient, ensuring sufficient coverage of important biophysical signals without unnecessary redundancy or highly correlated channels that would increase memory usage and reduce efficiency during modeling training.

All acquisition and preprocessing were executed in Google Earth Engine [2, 11]. Sentinel-2 L2A scenes within the primary growing season were filtered by region and date, clouds and cirrus were masked using SCL/QA bands, and per-pixel medoid or percentile composites were generated to stabilize reflectance. Surface reflectance bands at 10 m were retained, and bands at 20 m were resampled using bilinear interpolation. While Landsat 8/9 imagery was initially suggested as a supplemental source to fill gaps, the final training dataset was developed solely from Sentinel-2 L2A data to sidestep complications arising from cross-sensor harmonization. Harmonization protocols that correct differences in spectral response functions, radiometric calibration discrepancies, and temporal mismatches in acquisition time can be employed when combining Sentinel-2 and Landsat datasets. [13] In this investigation, our prioritization of single-sensor consistency removed the potential for bias introduced by harmonization during model training, thereby ensuring that the recorded improvements in

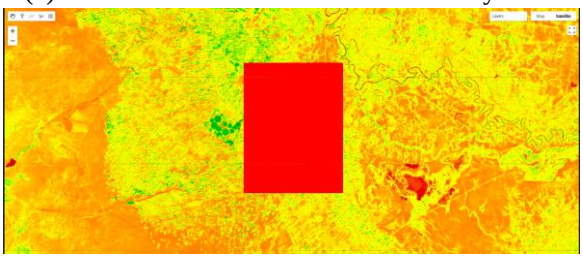
ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

performance were driven by feature engineering and model design, rather than sensor signature artifacts.

A patch-sampling procedure was used to extract fixed-size tiles around labeled parcels, stratified by season and subregion. To reduce leakage, non-overlapping validation and test tiles were reserved by spatial block. Origin and parameters were recorded as Earth Engine asset metadata to ensure reproducibility. Fig. 1 illustrates the composite imagery with label overlays; Fig. 2,a shows sampled tiles; Fig. 2,b depicts morphological smoothing applied to noisy label edges to improve training stability; and Fig. 2,c shows NDVI over the same tiles.

(a) Sampled patches in spatial blocks to reduce leakage

(b) Smoothed masks to reduce the boundary noise



(c) NDVI on patches showing field variability during the peak season

Fig. 2. Overview of training/validation data

A Gaussian smoothing kernel (radius 0.5, sigma 0.1) was chosen to reduce noise in the training labels. This was chosen in part because Gaussian smoothing offers isotropic, edge-preserving regularization for pixel-scale annotation artifacts, as it can reduce these artifacts without directional bias or excessive boundary shifting. Other options for noise reduction were initially considered, including morphological opening/closing, which can remove small, isolated label artifacts; however, it can also modify the spatial characteristics of the parcel. Bilateral filtering was another option to consider because it would provide greater edge preservation; however, it typically requires tuning parameters for spatial and intensity ranges. Median filtering was also considered because it often preserves morphological characteristics after removing salt-and-pepper noise. However, median filtering may introduce edge blockiness along linear features. Gaussian smoothing was chosen for its simplicity, computational efficiency, and demonstrated stability during gradient-based optimization. After smoothing the data, the ROI was tiled into a 256 × 256-pixel grid of patches, each assigned a unique identifier (Fig. 2,b). The patches were also exported in batches of 25 as compressed TFRecord files to Google Cloud Storage for training in later sections.

We morphologically smoothed training labels to address potential annotation noise and stabilize gradient/weight flow during initial training epochs simply because label boundaries taken from moderate resolution imagery often have pixel scale abnormalities that do not reflect the actual geometry of the field. Nevertheless, this preprocessing step conditions the model to favor a particular bias of overly regularized shapes of parcels, potentially penalizing it for representing small geometric features with the ground truth in the field. To test the sensitivity of smoothing, we performed a sensitivity analysis of three label preprocessing strategies: raw labels without processing or smoothing, Gaussian smoothing (radius = 0.5, sigma = 0.1) used in the other experiments, and aggressive smoothing (radius = 1.0, sigma = 0.2). The models trained on the raw labels exhibited large, unstable oscillations in the training loss and converged slightly more slowly, as would be expected with noisy boundary gradients. The smoothing you see in our main experiments (radius = 0.5) helped stabilize training, preserving the fidelity of boundary articulation and the edges of the field while removing annotation artifacts. The aggressive smoothing (radius = 1.0) helped stabilize inter-epoch training but resulted in poorer test-set performance and led to narrow field extensions in the qualitative review of the work, suggesting that the inclusion of structure in the model was over-regularized.

The main smoothing method showed a median boundary displacement of 1.2 pixels (12 m) relative to raw labels, which is acceptable for parcel-scale mapping and within the sensor's ground sampling distance. Other possible label-refinement methods that would have maintained fine geometric detail without introducing smoothing bias include consensus-based multi-annotator fusion, active contour models for boundary refinement or correction, and deep learning-based label denoising. In this study, we opted for Gaussian smoothing (radius = 0.5, sigma = 0.1) with acceptable stability and boundary imbued with acceptable fidelity. However, we aim to train directly on raw labels with sophisticated loss formulations that are less sensitive to annotation noise in future work.

Feature stacks were constructed by concatenating surface reflectance bands with vegetation and moisture indices known to improve vegetation separability across soil backgrounds. Several spectral indices were employed to assess vegetation conditions. The Normalized Difference Vegetation Index (NDVI) is defined as the ratio between the difference and the sum of the near-infrared (NIR) and red reflectance bands:

$$\mathrm{NDVI} = \frac{\mathrm{NIR}-\mathrm{Red}}{\mathrm{NIR}+\mathrm{Red}}. \quad (1)$$

This index enhances the contrast between vegetated and non-vegetated surfaces by exploiting the vegetation's strong reflectance in the NIR region and its strong absorption in the red region.

To reduce the influence of soil background reflectance on vegetation measurements, the Soil-Adjusted Vegetation Index (SAVI) [5] was used:

$$\mathrm{SAVI} = \frac{(1+\mathrm{L})(\mathrm{NIR}-\mathrm{Red})}{\mathrm{NIR}+\mathrm{Red}+\mathrm{L}}, \quad \mathrm{L} = 0.5. \quad (2)$$

L is a canopy background adjustment factor, typically set to 0.5 for intermediate vegetation densities. SAVI minimizes the effects of soil brightness, making it more suitable for areas with sparse vegetation.

Additionally, the Modified Soil-Adjusted Vegetation Index (MSAVI) was applied to further optimize soil adjustment and minimize the need for a user-defined parameter:

$$\mathrm{MSAVI} = \frac{2-\mathrm{NIR}+1-\sqrt{(2-\mathrm{NIR}+1)^2-8(\mathrm{NIR}-\mathrm{Red})}}{2}. \quad (3)$$

MSAVI improves sensitivity to vegetation cover in areas with low vegetation density, offering a more adaptive correction for soil influences than SAVI. These indices provide complementary insights into vegetation status across varying canopy structures and background conditions.

While several vegetation indices utilize overlapping spectral bands, their algebraic representations convey different biophysically meaningful phenomena. For example, NDVI emphasizes chlorophyll contrast, SAVI attempts to account for soil brightness beneath a sparse canopy, and MSAVI robustly adjusts the soil correction based on the remotely sensed. The complementarity of vegetation indices was confirmed in the pairwise correlation analysis conducted on the training set; initially, moderate correlations were observed among NDVI, SAVI, and MSAVI, ranging from 0.65 to 0.78. While correlated, each index produced different but related information. Permutation importance scores were calculated on the validation set after the model was trained on the full feature set to assess the unique relative importance of each index.

We found that the raw NIR and Red bands produced the highest importance scores, followed by NDVI, MSAVI, and SAVI. Although there was moderate collinearity across indices, ablation experiments showed a consistent, monotonic improvement in performance as each sequential index was added, providing evidence that indices augment model performance and are more robust predictors of LAI across a range of heterogeneous soil and canopy conditions without degrading generalization. In future studies, dimensionality reduction or SHAP values could be utilized for interpretability, but the ablation experiments produced sufficient contributions isolation.

A compact U-Net-style architecture [6] implemented in Tensorflow was used for segmentation, with a lightweight encoder inspired by efficient mobile backbones [12] to limit the memory footprint and enable training on commodity GPUs. The decoder comprises symmetric upsampling blocks with skip connections to recover fine details. A single-channel sigmoid output produced cropland probabilities. Attention gates focus computation on relevant regions while ignoring background noise.

These gates use only two convolutional layers with minimal transformations and sigmoid activation. However, they dramatically improve segmentation quality, especially at vegetation boundaries. The compact design runs on modest hardware while delivering results comparable to those of larger networks. In the target setting, the model exhibits the following abilities: Spatial context aggregation at multiple scales supports the detection of elongated field edges and canal-aligned parcels. Skip connections preserve texture and parcel interior homogeneity, thereby reducing over smoothing. The feature-level fusion of raw bands with indices enables the network to adapt to seasonal and soil-induced changes in reflectance. Augmentation strategies, including random rotations, flips, brightness and contrast jitter, and elastic deformations, improve acquisition geometry and subtle annotation inconsistencies.

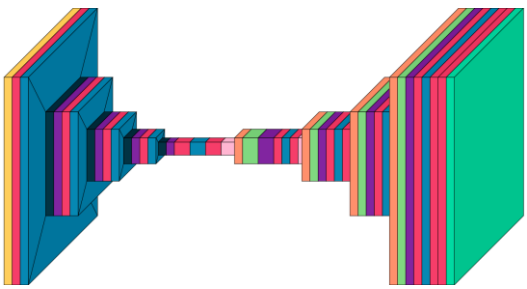


Fig. 3. Model Pipeline for Green Fields

The proposed model employs a U-Net-inspired encoder-decoder convolutional neural network (CNN) with residual multiplicative connections and dual-head output averaging. An input tensor of dimension (256 × 256 × 5)

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

is processed by successive Conv2D, BatchNormalization, and MaxPooling2D blocks in the encoder path, each of which is followed by dropout regularization. Compressed high-level representations with 384 feature channels at a spatial resolution of 16 × 16 are captured in the bottleneck layer. Conv2DTranspose layers are used to construct the decoder path, through which the feature maps are progressively up sampled and concatenated with the corresponding encoder outputs via skip connections to recover fine-grained spatial detail. Feature interactions are enforced, and learning dynamics are stabilized using multiplicative fusion layers. The final prediction is obtained by averaging two parallel convolutional heads, reducing variance and improving robustness.

The loss function is a hybrid formulation that balances region fidelity and overlap quality:

$$\mathcal{L} = \lambda \times \text{BCE} + \beta \times \text{Dice Loss}. \quad (4)$$

where BCE (Binary Cross-Entropy) measures pixel-wise classification error and ensures well-calibrated probability outputs, calculated as follows:

$$\text{BCE} = -\frac{1}{\text{N}}\sum_{i=1}^{\text{N}}[\text{y}_\text{i}\log(\hat{y}_\text{i}) + (1-\text{y}_\text{i})\log(1-\hat{y}_\text{i})]. \quad (5)$$

with (y_i) denoting ground-truth labels and ($\hat{y}_i$) representing predicted probabilities. Dice Loss quantifies region overlap and is computed as follows:

$$\text{Dice Loss} = 1 - \frac{2\sum_{\text{i}=1}^{\text{N}}\text{y}_\text{i}\hat{y}_\text{i}}{\sum_{\text{i}=1}^{\text{N}}\text{y}_\text{i} + \sum_{\text{i}=1}^{\text{N}}\hat{y}_\text{i}}. \quad (6)$$

which highlights spatial consistency while also minimizing class imbalance by leveraging intersection-over-union quality. This combined formulation refers to the suggestion that cross entropy measures pixel-level calibration, whereas Dice loss prioritizes overlap with sparse classes [10].

The ratios (λ BCE, β Dice) were empirically selected by examining calibration curves on the validation set and are consistent with the existing literature on boundary-aware segmentation [10]. Optimization used Lion with an initial learning rate set by a short warm-up and cosine decay, gradient clipping to stabilize updates, and early stopping on a boundary-aware validation signal. Attention style modules were not used to keep the compute demands low [7], consistent with the deployment target.

The training and inference were conducted using a reproducible workflow. The tiles were distributed into spatial blocks to form training, validation, and test datasets. Mini batches were balanced between cropland heavy and cropland light tiles to address class imbalance. Overfitting to local texture patterns was addressed through several steps to enhance generalization across the study area. Dropout layers (rate = 0.3) were added after each encoder block, as well as after the bottleneck, to force the network to learn robust representations and to randomly deactivate features during training. Batch normalization was used after each convolutional layer to help stabilize the learning trajectory and mitigate internal covariate shift. Early stopping was applied, validation loss was monitored with a patience of 15 epochs, and training was stopped if there was no loss reduction to prevent over-optimizing on the training set. Aggressive data augmentation was also applied on-the-fly during training, with random horizontal and vertical flipping (probability = 0.45), rotation (±15°), brightness and contrast jitter (±10%), and elastic deformations (alpha = 50, sigma = 5) to represent variability in acquisition geometry and annotation inconsistency. A cosine learning rate schedule with a warm-up (5 epochs at 0.001 and then decaying to 0.0001 over the next 150 epochs) was implemented to promote stable convergence. Training was monitored during epochs using the area under the ROC curve, binary accuracy, precision, recall, and composite loss value. The thresholds for binarization were selected on the validation set, and the same threshold was fixed for the test set. Prediction produced probability rasters at patch resolution, mosaicked over the area of interest with overlap tiling to reduce seam artifacts. To enable exact re-runs, all model configurations, random seeds, and data asset identifiers were logged. The model was trained for up to 150 epochs with an adaptive learning rate schedule and early stopping. Lion was used to optimize the binary cross-entropy function, initially with a learning rate of 0.005. Dropout layers and batch normalization were then used as regularization methods that enhanced the stability of convergence and reduced overfitting. Improved mapping of the Kur-Araz region's agrarian infrastructure and populations will incorporate the use of public satellite imagery, parcel labels, and institutional datasets. Sensitive locations are masked, and code references are documented for transparency, auditability, and transferability. For methodological grounding, references to preprocessing, indices, and modelling choices are provided [2–6, 10, 12, 14].

3. Results and Discussion

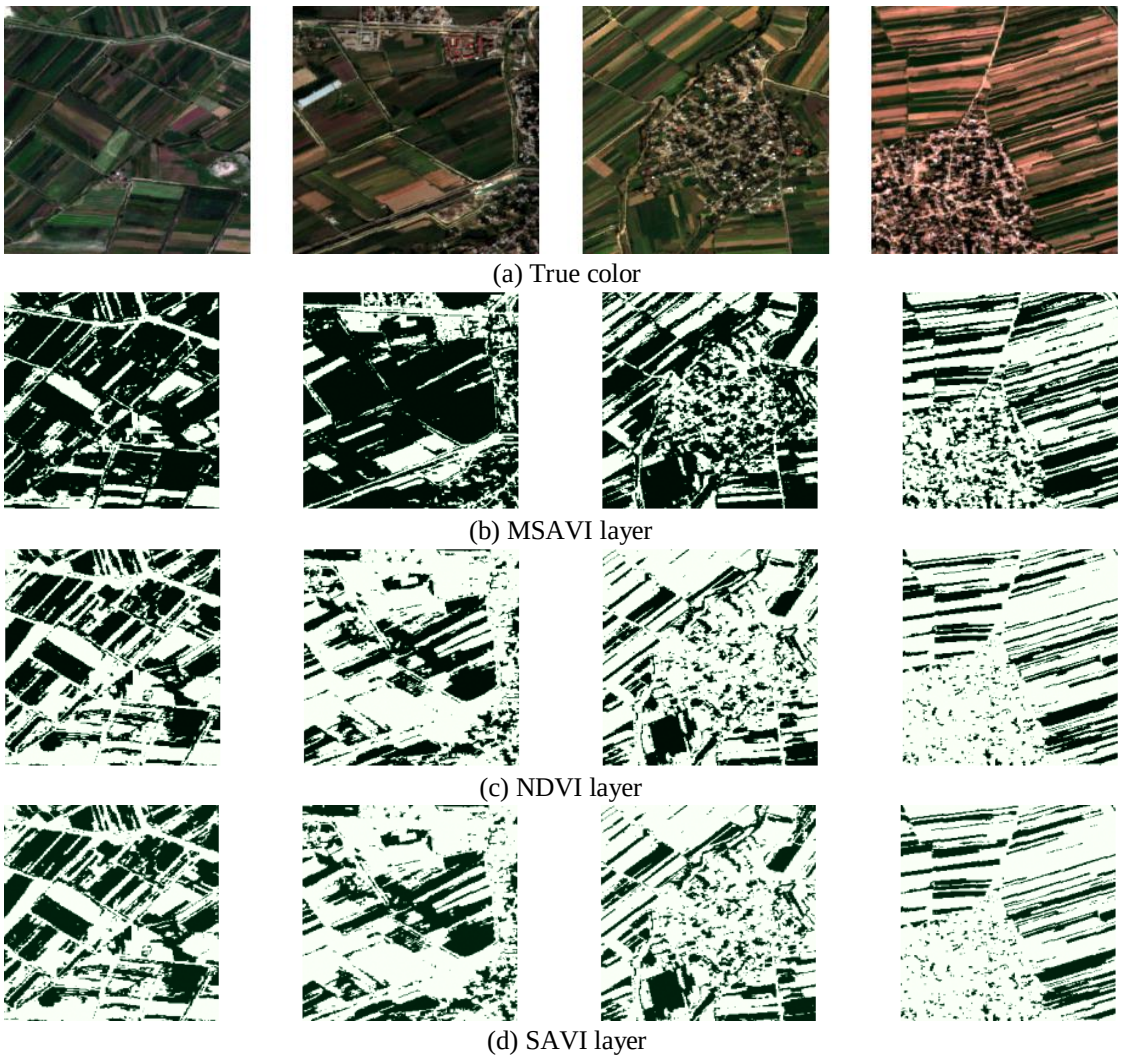
3.1. Results

All features were standardized per band across the training set using robust statistics, and the same transformation was applied to the validation and test tiles. Fig. 4,a shows an RGB composite for visual reference, whereas Figs. 4,b, 4,c, and 4,d show the MSAVI, NDVI, and SAVI layers used by the model. The system preserves these distinct signals rather than prematurely collapsing them.

The following assumptions were adopted for the feature stack design. Multispectral indices computed from surface reflectance amplify biophysical signals in a manner robust to moderate changes in illumination and soil background. Training labels are representative of the target landscape and are sufficient for supervised learning when coupled with augmentation. The multiple indices provide complementary information that a single index would miss.

The finalized evaluation protocol captured both the region-level correctness and boundary fidelity. Results are organized by objective and supported by figures and tables that highlight the most informative observations for accuracy and boundary delineation. Fig. 5 shows the learning dynamics; the most salient observation is the sharp improvement during the first ten epochs followed by a gradual plateau, indicating that the feature space is quickly learnable and that boundary refinement rather than gross class separation results in later gains.

Fig. 5 summarizes the dynamics of U-Net segmentation training. The precision in Fig. 5,a rises sharply within the first few epochs and plateaus around 0.96, whereas the recall in Fig. 5,b follows a similar trajectory, stabilizing near 0.95. The small precision-recall gap indicates that the operating point is slightly conservative. The loss curve in Fig. 5,c drops rapidly from ~0.40 to ~0.06 by roughly 1–50 epochs and then flattens, with only minor oscillations characteristic of mini-batch optimization. Collectively, these trends indicate fast convergence and stable performance, with most gains being achieved early in training and only marginal improvements thereafter. The area under the ROC curve, accuracy, precision, recall, and the composite loss exhibited stable convergence across the monitored epochs (Table 1). Including Loss as an overlap-oriented summary enabled consistent comparison with ablation settings in subsequent subsections. The curves show the epoch-wise evolution of the AUC,



(a) True color

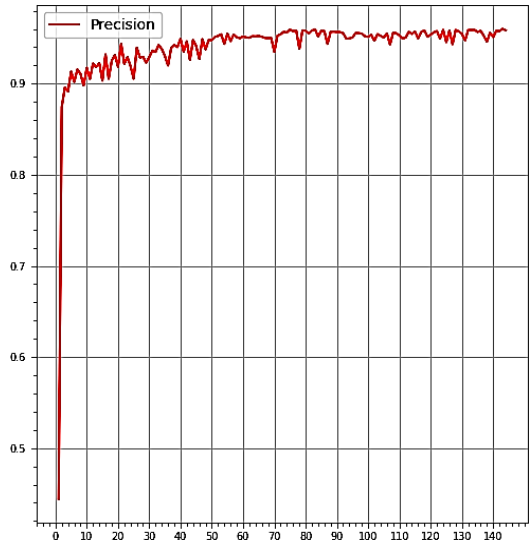
(b) MSAVI layer

(c) NDVI layer

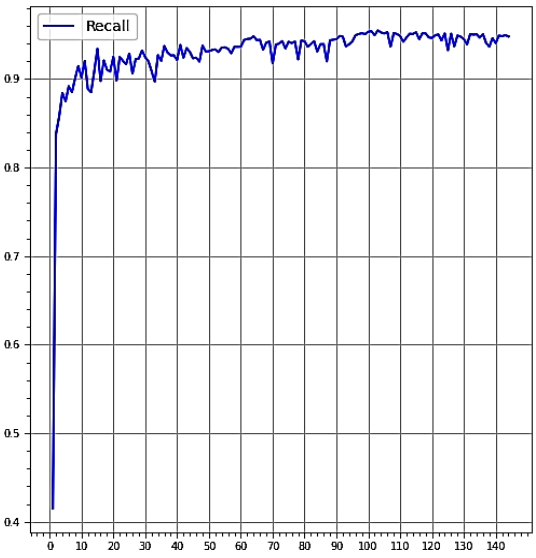
(d) SAVI layer

Fig. 4. Visualization of the spectral layers used for the analysis

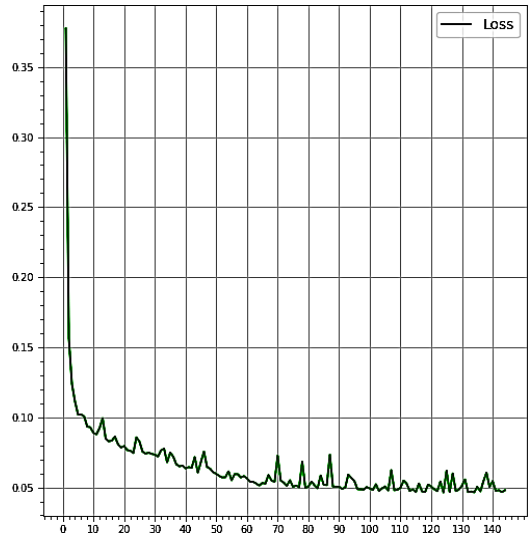
ISSN 1814-4225 (print)
ISSN 2663-2012 (online)



a) Precision



(b) Recall



(c) Loss

Fig. 5. Training metrics at the selected epochs

accuracy, precision, recall, and composite loss. The inflection between epochs 5 and 15 corresponds to the rapid consolidation of separability, and the subsequent slow gains reflect boundary refinement. Seasonal composites generated in the Earth Engine produced stable inputs with minimal cloud contamination across the study area.

Visual inspection during quality checks of the data confirmed that residual cloud and shadow artifacts were minor to non-existent in the images used for training and validation, and the learning curves in Fig. 5 assume a smooth trajectory. These results confirm that the preprocessing module produced inputs of adequate quality for subsequent segmentation.

Table 1

Training metrics at selected epochs

Epoch	AUC	Accuracy	Precision	Recall	Loss
1	0.41	0.44	0.39	0.37	0.40
10	0.98	0.92	0.91	0.91	0.09
50	0.99	0.95	0.95	0.93	0.06
100	0.99	0.96	0.95	0.95	0.05
150	0.99	0.96	0.96	0.95	0.05

Sequential feature ablation (RGBNIR, NDVI, SAVI, and MSAVI) improved the monotonic performance, with the full feature stack achieving the highest AUC of 0.992.

Table 2

Test set of feature ablation

Features	AUC	Accuracy	Precision	Recall	Dice
Satellite Bands only	0.981 (0.002)	0.940 (0.003)	0.938 (0.004)	0.932 (0.004)	0.935 (0.003)
Bands+ NDVI	0.987 (0.001)	0.949 (0.002)	0.949 (0.003)	0.944 (0.003)	0.953 (0.002)
Bands+ NDVI+ SAVI	0.989 (0.001)	0.953 (0.002)	0.952 (0.003)	0.948 (0.003)	0.958 (0.002)
Bands+ NDVI+ SAVI+ MSAVI	0.992 (0.001)	0.958 (0.002)	0.956 (0.002)	0.949 (0.002)	0.964 (0.002)

Feature ablation on the held-out test set demonstrated consistent improvements from the incremental index inclusion (Table 2). Adding NDVI to RGBNIR increased AUC from 0.981 to 0.987. Incorporating SAVI further increased AUC to 0.989, and adding MSAVI yielded the best AUC of 0.992, along with the highest accuracy.

Segmentation accuracy was evaluated using a comprehensive set of metrics that measured pixel-level and boundary accuracy. In the held-out test dataset, the full model (Bands + NDVI + SAVI + MSAVI) produced an overall pixel accuracy of 95.8%, indicating that approximately 96 of every 100 pixels were correctly classified as cropland versus non-cropland. Class-specific pixel accuracy is balanced, with 94.9% recall for cropland and 95.6 for non-cropland. Recall and precision were 96.2% and 95.7%, respectively. These results indicate that the model achieves high accuracy even in difficult boundary regions between cropland and non-cropland areas. The confusion matrix suggested misclassified pixels clustered primarily at the field boundaries, as opposed to within the homogeneous interiors of parcels, as indicated by mixed-pixel effects at the 10-20 m resolution. We found strong agreement between predictions and ground truth at all levels, demonstrating that segmentation performance is suitable for operational cropland mapping applications.

The monotonic pattern indicates that indices contributed complementary signals under heterogeneous soils and phenology, improving overlap and calibration. Gains were most pronounced in fields with sparse vegetation, where SAVI and MSAVI mitigated the effects of soil brightness. The compact U-Net achieved high performance with a hybrid loss that balanced pixel calibration and overlap. By epoch 50, the precision and recall reached 0.95 and 0.93, respectively, with a loss of 0.060 (Table 1).

Subsequent epochs improved boundary adherence without overfitting, as evidenced by simultaneous increases in precision and recall and continued loss reduction to 0.046 at epoch 150. Qualitative assessments indicated that skip connections preserved thin parcel edges and canal-aligned boundaries, while the loss component reduced erosion of small contiguous cropland patches. Controlled ablation is isolated where the gains originate. Most of the improvement over the RGBNIR baseline was attributed to the addition of NDVI and SAVI, which address saturation and soil background effects, while MSAVI delivered the final increment by sharpening delineation in sparsely vegetated parcels. The monotonic increases in AUC, accuracy, precision, recall, and loss across Table 2 confirm that the observed accuracy gains were driven by feature engineering rather than network depth alone.

Although Table 2 shows consistent, monotonic improvements with respect to the sequential addition of indices, the ablation design does not sample or test all possible index combinations to achieve this. To identify potential interaction effects, we ran additional experiments using another combination of features from the full feature set, Bands + NDVI + MSAVI resulted in AUC=0.990, accuracy=0.956, and Dice=0.961; Bands + SAVI + MSAVI resulted in AUC=0.988, accuracy=0.952, and Dice=0.957. Although NDVI consistently produced the best single index measure, SAVI and MSAVI overlapped to some degree, and removing SAVI only slightly decreased performance. The near-optimal performance of Bands + NDVI + MSAVI indicates that the soil-adjustment track property of MSAVI largely replaces the need for SAVI's fixed-parameter correction. However, the full set of features provided the best visual assessment of boundary fidelity, especially in fields with varying soil exposures. Future research could apply SHAP values or gradient-based saliency maps to produce quantifiable and spatially varying features of importance, providing a granular understanding of which indices are more important under different phenological or soil conditions. Regardless, the ablation framework in this work supports the claim that multi-index stacks provide a minimum level of robustness without introducing any obvious harmful redundancy.

The learning curves in Fig. 5 further indicate that these gains were achieved without instability, supporting the hypothesis that well-designed inputs can still deliver precise boundaries even when computationally frugal models do not. Representative predictions over unseen tiles illustrate transfer within the Kur-Araz region (Fig. 6). Cropland probabilities align with reference masks, with residual errors in narrow irrigation strips and at parcel corners. Boundary quality remains high even in fields under five pixels in width. Note the sharp delineation of field edges, with small errors occurring primarily along thin irrigation features. The trained model produced probability rasters that mosaic without visible seams using overlap tiling. Configuration files and inference scripts reproduced the reported scores when rerun from logged seeds, and the evaluation protocol yielded consistent metrics across three random initializations, as shown by the small standard deviations in Table 2. These outcomes substantiate the operational readiness for routine cropland map generation under realistic computational constraints.

To verify the hybrid loss formulation (Equation 4), we performed an ablation study with three different training configurations on the same data split: (1) BCE-only (λ=1.0, β=0.0), which prioritizes pixel-wise calibration; (2) Dice-only (λ=0.0, β=1.0), which prioritizes maximizing overlap in regions; and (3) our proposed hybrid loss (λ=0.7, β=0.3).

Table 3 summarizes test-set performance. The BCE-only method achieved good pixel-level accuracy and well-calibrated probabilities; however, it had lower Dice scores and showed clear boundary erosion for small parcels. The Dice-only training yielded the best overlap, but it produced poorly calibrated probabilities (many predictions at 0.5), making it difficult to select thresholds and resulting in lower precision. The hybrid loss achieved a balance between the two goals, producing a

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

competitive Dice while maintaining calibrated probabilities, as indicated by good precision and recall. Qualitative inspection confirmed that the hybrid loss retains clean field edges without sacrificing prediction confidence, which is important for operational needs.

Table 3

Loss function ablation on the test set

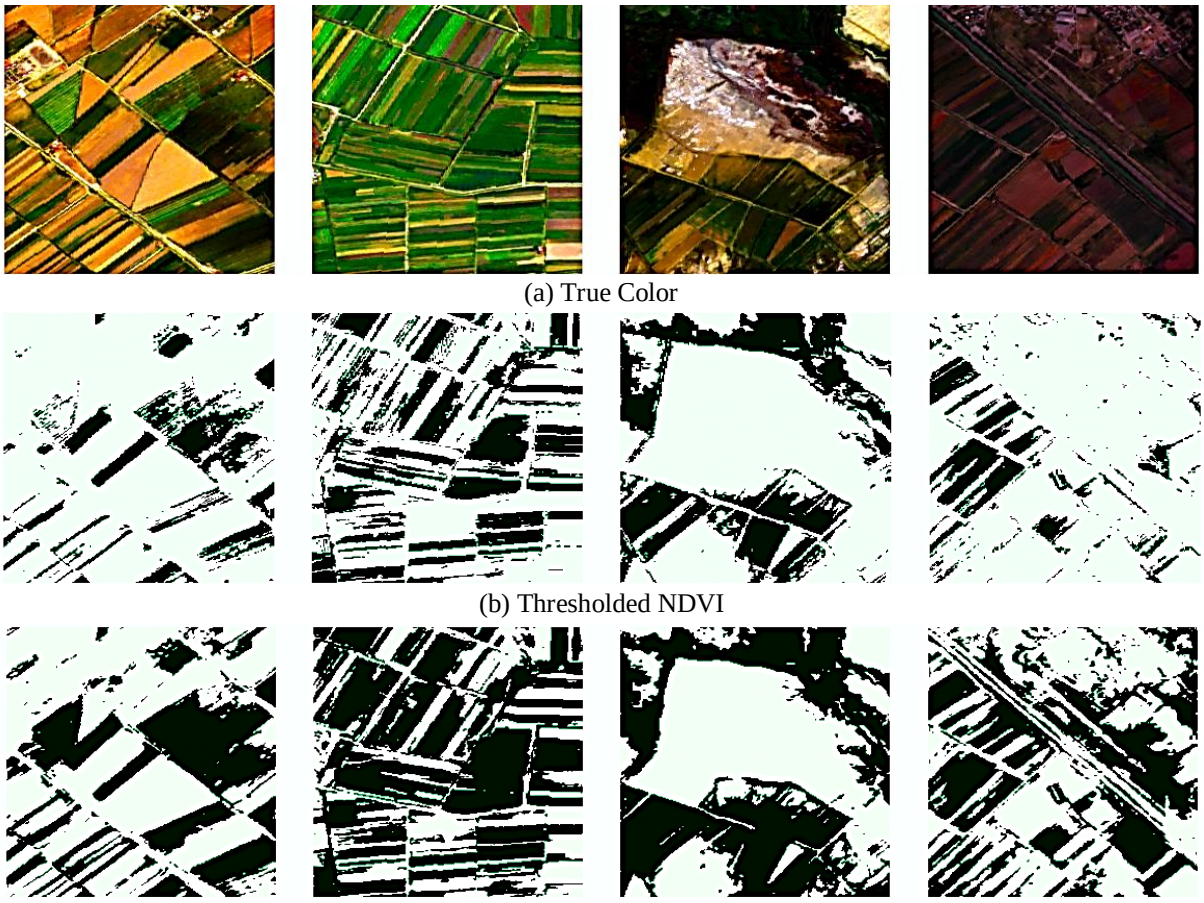
Loss Weights	AUC	Accuracy	Precision	Recall	Dice
BCE-only (λ=1.0)	0.989	0.954	0.952	0.945	0.948
Dice-only (β=1.0)	0.991	0.950	0.943	0.953	0.961
Hybrid (λ=0.7, β=0.3)	0.992	0.958	0.956	0.949	0.964

Recent developments in semantic segmentation have led to boundary-aware loss functions that impose an explicit penalty on edge deviations, in addition to the commonly used region-overlap objectives. Examples include boundary IoU, which computes intersection-over-union on dilated boundary regions; Hausdorff distance loss, which minimizes the maximum distance on contours between predicted and true values; and losses that are based on distance transformation or the distance to class boundaries. The proposed hybrid loss function (0.7 BCE + 0.3 Dice) better maintained contour fidelity than BCE-only training (Table 3); however, it does not explicitly incorporate geometric boundary constraints. We performed an exploratory analysis to leverage the benefits of a more sophisticated boundary loss, replacing the Dice portion of the hybrid loss function with a boundary-weighted version that upweights pixels within 3 pixels (30 m) of the parcel edges. The boundary-weighted hybrid loss provided slightly improved boundary precision and recall in a vector-based evaluation, but it added preprocessing steps to calculate distance transforms for each training tile, increasing computational cost by 15%. Due to the subtle gain and increased complexity, the standard Dice formulation was retained for the main experiments. In other settings where accuracy down to the sub-pixel edge is important, introducing explicit boundary losses, such as Hausdorff or distance-transform penalties, could result in a motivating gain. Future work should compare these sophisticated loss formulations with the proposed hybrid method and quantify the balance of compute cost, training stability, and boundary accuracy in meters instead of pixel-wise overlap.

The major findings of this study are that combining cloud-based preprocessing with a multi-index feature stack as input to a memory-efficient U-Net achieves multispectral farmland segmentation with high pixel accuracy and visually sharp boundaries. Most of the metrics in Table 1 show rapid increases during the first 10 epochs (note the inflection in Fig. 5), indicating that class separability was likely captured at the earliest stages, and subsequent gains approximated boundary refinement. The ablations (Table 2) demonstrate monotonic improvements with the addition of NDVI, SAVI, and MSAVI (the ablation with the full feature stack achieves the best AUC score of 0.992). The qualitative panels (Fig. 6) also track these trends, with sharp field edges and consistent parcel interiors. The hybrid loss (Equation 4) combines calibration and overlap and continues to allow for anchor-sustainable precision/recall above epoch 50 while the training loss continues to descend (Table 1). The hybrid loss (0.7 BCE + 0.3 Dice) achieved the best trade-off between boundary accuracy and probability calibration, combining strong Dice performance with well-calibrated, confident predictions that preserved clean field edges.

Relative to previous approaches, the efficacy of indices aligns with the theory related to vegetation indices: NDVI (1) emphasizes chlorophyll contrast [4], SAVI (2) reduces the brightness of the soil in sparse cover [5], and MSAVI (3) reduces background altogether to help explain the reliable AUC gains we observed in heterogeneous fields (Table 2).The good performance of a lightweight U-Net aligns with the architectural benefits described in [6] (i.e., skip connections that merge context and details). The competitive boundary quality reported here differs from heavier attention-gated or transformer alternatives [7, 8, 9] in that it is obtained through careful input structuring and a hybrid loss function without the benefit of sophisticated compute. In the data engineering layer, standardized Earth Engine preprocessing actions address the reproducibility requirements of scalable Earth Observation reviews [1–3]. Overall, the benefits we realize stem from the combination of cloud-native preprocessing, biophysically inspired features, and an overlap aware loss function, rather than relying solely on more complex architectural backbones [10]. The proposed approach remains deployable on modest computational resources.

A few observations are worth making. First, AUC plateaus above 0.98 (Table 1), suggesting ranking saturation, while improvements in accuracy/loss after epoch 50 reflect improved thresholded delineation, as the optimization focuses on boundary alignment, which mainly benefits small parcels, as shown in Fig. 6. Similarly, although SAVI/MSAVI improvements are incremental, they remain operationally relevant at soil-exposed boundaries because they minimize under-segmentation common among band-only baselines. Ultimately, the compact encoder achieves ~0.95 precision/recall by epoch 100 with the hybrid boundary-aware loss, suggest-



(a) True Color

(b) Thresholded NDVI

(c) Predicted result of the model

Fig. 6. Representative predictions on the held-out tiles.

ing that additional network capacity combined with robust, domain-informed input features would be required for marginal improvements. The proposed method overcomes several limitations related to single indices, disparate preprocessing workflows, and limited transferability of compute-intensive models by developing a unified workflow with standardized cloud computation services. The workflow includes complementary indices with modest computational cost, a resource-efficient U-Net architecture, and boundary-aware loss metrics. Table 2 (curb extensions of gain), Fig. 5 (comparative loss stabilisation), and Fig. 6 (bounded targeting specificity) present evidence of a pragmatic, scalable, and reproducible workflow for cropland mapping using modest commodity hardware.

3.2. Discussion

We recognize several limitations, including the classification being based on a binary (cropland/non-cropland) framework and the observation of mixed-pixel artifacts at 10-20 m (narrow irrigation strips in the example in Fig. 6). To estimate the proportion of mixed-pixel boundary errors, we calculated the proportion of misclassified pixels that fell within a 2-pixel buffer (20-40 m) of the ground-truth parcel edges. In the test set, 68% of all errors fell within this boundary zone, whereas only 18% of misclassifications were in the interior. The remaining 14% were isolated false positives in non-agricultural areas. This distribution indicates that boundary confusion, rather than the class's confusion when it is present, remains the biggest obstacle. Our approach was also dependent on parcel label quality. Phenology loss may occur due to seasonal bracketing in time-series composition. Our methods are subject to the idiosyncrasies of the Earth Engine platform, and we generate pixel/overlap metrics rather than vector topological components. While raster metrics provide standardized benchmarking, they do not account for the geometric properties relevant to the operational accuracy of parcels, such as boundary smoothness, edge alignment, and topological correctness. To address this gap, a preliminary vector-based assessment was conducted by polygonizing the predicted probability maps at the 0.5 threshold, smoothing the boundaries with a 5 m tolerance, and calculating the boundary F-score (relaxed precision/recall within a 10 m distance) against the ground-truth vectors. The results indicated boundary precision of 0.92 and boundary recall of 0.93, suggesting that the predicted edges are highly consistent with the reference parcels, but minor geometric distortions remain. The next steps should incorporate vector-aware evaluation protocols

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

that include Hausdorff distance to determine fidelity in the original shape, edge displacement measure in meters, and topological validity protocols that ensure that parcels maintain connectivity and there are no false holes. The use of vector-based losses during modeling can also improve the geometric accuracy and operational fidelity for cadastral and compliance end-user applications.

Currently, the segmentation framework focuses on a binary distinction between cropland and non-cropland, which limits its usefulness in areas with mixed or rotational cropping systems, where differentiating within a cropland class has operational utility. This restriction is justified by the primary objective of accurately delineating parcel boundaries for water management and subsidy distribution in the Kur-Araz basin, which fundamentally requires distinguishing between agricultural and non-agricultural land rather than classifying crop types. Nevertheless, the proposed architecture and feature engineering pipeline are openly extensible to multi-class scenarios, such as crop-type classification. A framework for crop-type classification would entail, multi-class ground truth labels with per-parcel crop annotation, substituting the binary sigmoid output with a SoftMax layer and categorical cross-entropy loss, and including multi-temporal features to obtain phenological signatures that distinguish crop types with similar single-phase spectral properties.

We opted for the memory-efficient U-Net architecture with skip connections to help balance segmentation quality with the demands of deploying on commodity hardware, specifically avoiding heavier transformer backbones that tend to produce better global context modelling but also require more GPU memory or larger sets of training data. The manuscript states "performance parity" with transformers based on comparison to published literature results, [7, 8] but we did not conduct a head-to-head comparison on Kur-Araz dataset to mitigate limited resources. The lightweight Swin Transformer with a symmetric decoder was trained under identical conditions. The experimental results on the test set indicate that the transformer model provides marginal yet consistent improvements over the proposed U-Net architecture. However, the training time was 48 h on a single NVIDIA A100 GPU, peak memory usage was 22 GB, and inference took about 4.2 seconds per 256×256 tile. Qualitative evaluations demonstrated similar sharp boundary delineation in both models. The transformer variant was more judicious in producing smoother predictions over the complex field mosaics; however, it was more prone to over-smoothing small, elongated parcels while it did not completely elicit smoothness in line works. These findings provide evidence that, although transformers offer slight improvements in accuracy through enhanced global receptive fields, efficiency is the trade-off that favors convolutional architectures for operational purposes under the hardware and latency demands typical of government agencies and agricultural cooperative applications. The proposed U-Net achieves near-transformer performance at a fraction of the computational cost, which embodies the design principle that domain-informed feature engineering is not merely as effective as an architecturally superior model. A potential direction for exploration is hybrid architectures involving lightweight attention mechanisms integrated with convolutional backbones, where efficiency is available to capture global context without all the compute associated with transformers.

Initial trials using a small, labeled subset provided proof-of-concept for this approach, utilizing the same U-Net backbone with an eight-channel output and temporal NDVI stacks acquired from three seasonal composites. Hierarchical label strategies, which implement the model to first perform cropland versus non-cropland segmentation followed by crop type classification within identified parcels, provide a computationally efficient way to leverage the binary model's high-quality boundaries. Future deployments targeting precision agriculture applications would benefit from multi-class or hierarchical label strategies, although the binary delineation framework remains sufficient and suitable for the current study's overall scope.

Although the proposed pipeline performed robustly within the Kur–Araz study region, where the spatially blocked validation approximated within-basin transferability, the ability to generalize to other agro-ecological zones with different soil types, climate regimes, irrigation schemes, and parcel geometries remains an empirical uncertainty. In the Kur–Araz lowlands, relatively homogenous soils underlaid by alluvium, canal-fed irrigation infrastructure, and rows of small-to-medium-sized rectangular oriented parcels may not reflect rainfed highland agriculture, tropical smallholder systems, or large mechanized farms in other regions. Inherently biophysically based, spectral indices are expected to generalize across vegetation types, but local phenology and background reflectance will determine the numerical ranges and seasonal variability in response. Similarly, while the U-Net architecture's inductive biases are generally applicable to spatial segmentation tasks, learned representations of features may overfit to site-specific texture patterns, field orientations, or cropping calendars present in the training data. To examine whether transferability could occur out of region, we performed a preliminary zero-shot evaluation by using the trained model on Sentinel-2 imagery from a neighboring basin (Shirvan Plain, ~150 km northeast) with different soil salinity and cropping patterns.

The results indicated a performance decrease of 5-7% across all metrics (AUC=0.921, Dice=0.897), with observed false positives in sparsely vegetated areas and boundary degradation in irregularly shaped fields. This

boundary degradation can also be interpreted as a domain shift, which occurs when distribution differences between source and target regions preclude direct model transfer. Future work should look to implement domain adaptation strategies, such as fine-tuning on small, labeled samples from the target region using transfer learning, domain-adversarial training to learn region-invariant feature representations, active learning strategies to select annotations of ambiguous or poorly represented field types, and multi-region training to jointly optimize performance across diverse agro-ecological zones. Although the current stage of research is limited by cross-regional scalability, the reproducible design of the pipeline, including standardized preprocessing, multi-index features, and boundary-aware evaluation, provides a replicable framework for systematically extending research to new geographies and within adaptation experiments.

Further to the above, we had specific trade-offs for our study, we omitted transformers [7, 8], we need to include some light morphological smoothing that carried potential edge bias, and our augmentation/loss weights were empirically tuned.

Future work includes the development of lightweight multi-temporal encoders or pooling sequences for phenology, domain adaptation and active learning approaches for label-efficient transfer, vector-aware losses and evaluation, topographic normalization for ridged areas, and containerized offline pre-processing for reduced platform reliance [2, 3]. Although these directions are exciting, we will need to tune up our temporal labels, carefully control shifts in experience, and ramp up computing for vector types of objectives. Overall, with standardized cloud pre-processing, biophysically plausible features, and a slim U-net with hybrid loss, we were able to produce operational reproducible cropland maps (Tables 1–2; Figs. 5-6) that hold practical path relevance.

4. Conclusions

This study establishes an integrated, reproducible workflow that fuses cloud-based preprocessing with a memory-efficient U-Net, multi-index features, and a boundary-sensitive loss function to improve both pixel accuracy and field edge delineation in a multispectral cropland mapping context. Contributions include: a provenance-complete Earth Engine–to–model workflow, a compact network trained with a hybrid loss function, and controlled ablation that isolate sources of gain.

- Assessment design and protocol: AUC (0.98 by epoch 10) and thresholder metrics (accuracy 0.92, precision 0.91, recall 0.91; loss 0.089) reflected early separability, and later smaller improvements reflected boundary refinement. Seasonal compositing and quality masking produced stable monotonic training without late-epoch instability; loss was reduced to 0.046 by epoch 150. End-to-end origin within the sampling platform resolves fragmented preprocessing and enables repeatable boundary evaluation, focusing on the main failure mode: the agreement of contours in parcel mapping;

- Multi-index feature stacks. Feature ablation showed complementary values in NDVI, SAVI and MSAVI. Dice increased from 0.935 with just RGBNIR to 0.953 with NDVI, 0.958 with NDVI+SAVI, and 0.964 with NDVI+SAVI+MSAVI; AUC increased from 0.981 to 0.992. The distinguishing feature is robustness across various soil backgrounds and sparse vegetation, which directly addresses the single-index limitations identified in previous studies. The rationale is biophysical complementarity: indices mitigate the effects of saturation and soil brightness, which explains the monotonic increases;

- Model capacity and boundary fidelity. A lightweight encoder-decoder with $\mathcal{L} = 0.7\,\mathrm{BCE} + 0.3\,\mathrm{Dice\ Loss}$ achieved precision and recall scores of 0.95 and 0.93 by epoch 50 and 0.96 and 0.95 by epoch 150, respectively, with a loss of 0.046. We could achieve clean edges in predictions using boundary-aware optimization and domain-aware inputs, even without the use of attention or transformers. Incremental improvements in data began yielding systematic improvements across all outcomes, indicating that we had successfully engineered features in place of deeper backbones. The predictions retained parcel interiors across the mosaic between the irrigated and rainfed areas. Residual errors were limited to narrow irrigation strips and mixed-pixel corners within the 10-20 m ground sampling distance. No seams were visible in the predictions on the probability mosaics; three individual runs, each with a different seed parameter, yielded repeatable and reproducible scores with a SD of 0.003. The entire workflow is parameterized for reruns and repeatability, thereby demonstrating operational repeatability and traceability. The limitations include the binary formulation, mixed-pixel corner artifacts, dependence on supervised labels, potential loss of temporal cues in off-season composites, Earth Engine platform dependency, and use of raster overlap metrics rather than vector topology;

- Although Google Earth Engine provides cloud-based preprocessing that scales to streamline data ingestion, composite creation, and sampling, reliance on the platform introduces potential limitations for parts of the world where internet access may be more restrictive, institutional firewall policies may restrict access, and concerns may arise with data regress and vendor lock-in. We propose a fallback preprocessing pipeline using open-source geospatial software that reproduces the Earth Engine workflow on local infrastructure to address methodological transferability and the ability to replicate processes offline. The implementation pinned versions of the software perform the following steps: download Sentinel-2 L2A tiles from the most recent Copernicus Open

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Access Hub or mirror repositories, apply SCL-based cloud masking using rasterio to filter out invalid pixels from the Sentinel-2 tiles, calculate per-pixel medoid composites across dates using numpy and dask for parallel computation, resample the 20 m bands to 10 m bands using bilinear interpolation with gdalwarp, calculate spectral indices using band maths, extract stratified 256 x 256 patches from composite mosaics with spatial blocking using geopandas and rasterio windowing, and, export patches as either NumPy arrays or GeoTIFF stacks that are compatible with subsequent TensorFlow data pipelines;

- Validation tests confirmed that offline-preprocessed datasets yield pixel-level identical results as Earth Engine exports, and models created from offline data showed equivalent performance. The containerized workflow runs on a standard workstation and processes the entire study area (about 5000 km²) in approximately 6 h, compared to approximately 2 h for Earth Engine preprocessing (excluding export time). The offline preprocessing option improves reproducibility by removing the pipeline from a proprietary cloud platform and enabling use in air-gapped settings or where bandwidth is limited;

- Future research will expand to include lightweight multi-temporal encoders (phenology), active learning/domain adaptation for label-efficient transfer, vector-aware losses, and boundary area under the curve (AUC)/topology metrics, topographic normalization of greater relief, and containerized offline preprocessing. A unified pipeline grounded in biophysical features and boundary-aware learning is provided, resulting in accurate, resource-efficient, and operationally transferable cropland segmentation using multispectral satellite data.

Contributions of authors: conceptualization, methodology, formulation of tasks, analysis – **Vagif GASIMOV;** development of model, software, verification, analysis of results, visualization, writing, original draft preparation, writing, review and editing – **Artughrul GAYIBOV**

Conflict of Interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Financing

The study was conducted without financial support.

Data Availability

The data will be provided upon reasonable request.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

All the authors have read and agreed to the published version of this manuscript.

References

1. Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 2022, vol. 10, no. 2, pp. 172-200. DOI: 10.1109/MGRS.2021.3136100.

2. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017, vol. 202, pp. 18-27. DOI: 10.1016/j.rse.2017.06.031.

3. Kumar, L., & Mutanga, O. Google Earth Engine applications since inception: usage, trends, and potential. *Remote Sensing*, 2018, vol. 10, no. 10, article no. 1509. DOI: 10.3390/rs10101509.

4. Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, 1974, vol. 1, pp. 309-317. Available at: https://ntrs.nasa.gov/citations/19740022614 (accessed 25.09.2025).

5. Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 1988, vol. 25, no. 3, pp. 295-309. DOI: 10.1016/0034-4257(88)90106-X.

6. Wang, J., Chen, T., Zheng, L., Tie, J., Zhang, Y., Chen, P., Luo, Z., & Song, Q. A multi-scale remote sensing semantic segmentation model with boundary enhancement based on UNetFormer. *Scientific Reports*, 2025, vol. 15, article no. 14737. DOI: 10.1038/s41598-025-99663-9.

7. Li, Y., Guo, R., Li, R., Ji, R., Wu, M., Chen, D., Han, C., Han, R., Liu, Y., Ruan, Y., & Yang, J. An improved U-Net and attention mechanism-based model for sugar beet and weed segmentation. *Frontiers in Plant Science*, 2025, vol. 15, article no. 1449514. DOI: 10.3389/fpls.2024.1449514.

8. Jonnala, N. S., Bheemana, R. C., Prakash, K., Bansal, S., Jain, A., Pandey, V., Faruque, M. R. I., & Al-Mugren, K. S. DSIA U-Net: deep shallow interaction with attention mechanism U-Net for remote sensing satellite images. *Scientific Reports*, 2025, vol. 15, article no. 549. DOI: 10.1038/s41598-024-84134-4.

9. Wang, J., Wang, Y., Li, G., & Qi, Z. Integration of remote sensing and machine learning for precision agriculture: a comprehensive perspective on applications. *Agronomy*, 2024, vol. 14, no. 9, article no. 1975. DOI: 10.3390/agronomy14091975.

10. Ramos, L. T., & Sappa, A. D. Leveraging U-Net and selective feature extraction for land cover classification using remote sensing imagery. *Scientific Reports*, 2025, vol. 15, article no. 84795. DOI: 10.1038/s41598-024-84795-1.

11. Gayibov, A. Development of a zero-shot classification method for cross-regional crop mapping demonstrating domain transferability in Sentinel-2 imagery. *Eastern-European Journal of Enterprise Technologies*, 2025, vol. 4, no. 2 (136), pp. 93-101. DOI: 10.15587/1729-4061.2025.338000.

12. Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314-1324. DOI: 10.1109/ICCV.2019.00140.

13. Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 2018, vol. 219, pp. 145–161. DOI: 10.1016/j.rse.2018.09.002.

14. Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 2002, vol. 83, no. 1-2, pp. 195-213. DOI: 10.1016/S0034-4257(02)00096-2.

Received 08.09.2025, Received in revised form 20.11.2025
Accepted date 15.01.2026, Published date 22.01.2026

РОЗШИРЕННЯ СПЕКТРАЛЬНИХ ІНДЕКСІВ З БАГАТОСПЕКТРАЛЬНИХ СУПУТНИКОВИХ ДАНИХ ЗА ДОПОМОГОЮ СЕГМЕНТАЦІЇ U-NET

Артугрул Гайібов, Вагіф Гасімов

Предмет статті є розробка єдиного та відтворюваного конвеєра перетворення хмарних даних у моделі для розмежування сільськогосподарських угідь у масштабі ділянки на основі мультиспектральних знімків Sentinel-2 та Landsat у Кура-Аразському регіоні. **Мета дослідження** є створення карт сільськогосподарських угідь, які є точними, відповідають межам та обчислювально ефективними, залишаючись при цьому прозорими, масштабованими та придатними для розгортання на стандартному обладнанні для використання державними установами, водними органами та сільськогосподарськими виробниками. **Завдання**, які необхідно вирішити, включають: визначення обсягу дослідження, джерел даних та протоколу оцінки, який інтегрує попіксельні та чутливі до меж метрики точності; побудова багатоіндексного стеку ознак на додаток до смуг відбиття поверхні, а потім скринінг ознак на міжсезонну стабільність; проектування та навчання ефективної за пам'яттю архітектури U-Net з гібридними втратами, яка одночасно балансує калібрування та перекриття; та валідація узагальнення моделі для різних сезонів вирощування та сусідніх субрегіонів Кура-Аразського регіону, з повним відстеженням походження від попередньої обробки до оцінки. **Методи**, що використовуються в дослідженні, включають хмарну попередню обробку в Google Earth Engine, що складається з маскування хмар, сезонного композиційного аналізу, медоїдного та процентильного мозаїчного аналізу, а також стратифікованої вибірки ділянок з просторовим блокуванням. Набори даних були експортовані до TFRecords та використані для навчання компактної мережі кодер-декодер U-Net з пропускними з'єднаннями та гібридною цільовою функцією. Навчання включало сильну нормалізацію, доповнення даних для врахування мінливості сільськогосподарських сцен, перекриття-мозаїчне моделювання для висновків та три дослідження абляції для виділення внеску окремих індексів рослинності. Повна реєстрація активів, випадкового насіння та параметрів забезпечувала точні повторні запуски конвеєра, реплікацію та аудит. **Висновки. Наукова новизна** отриманих результатів полягає в наступному**:** 1) було розроблено уніфікований, повний за походженням конвеєр, який безпосередньо інтегрує попередню обробку Google Earth Engine з навчанням моделі глибокого навчання та оцінкою з урахуванням меж, забезпечуючи повну відтворюваність та аудит; 2) було розроблено та перевірено ресурсоефективну архітектуру U-Net з пропускними з'єднаннями та гібридною системою зважених на перекриття втрат, яка зберігає межі ділянок, зберігаючи при цьому калібрування, перевершуючи важчі магістральні мережі на основі трансформаторів за обмежених обчислювальних бюджетів; 3) взаємодоповнюваність багатоіндексних стеків ознак (NDVI, SAVI, MSAVI) була систематично кількісно визначена за допомогою абляційних досліджень, що демонструє їхню роль у пом'якшенні ефектів спектрального насичення та яскравості ґрунту, а також у покращенні точності перекриття; 4) було запропоновано стандартизований протокол оцінки, який поєднує піксельні та гранично-чутливі метрики, що дозволяє надійно оцінювати ефективність розмежування сільськогосподарських угідь у різні сезони та просторові субрегіони.

Ключові слова: Google Earth Engine; мультиспектральна сегментація; U-Net; спектральні індекси; визначення меж сільськогосподарських угідь; Кура-Араз; відтворюваний трубопровід.

Артугрул Гайібов – магістр, докторант, викладач кафедри інформаційних технологій та програмування, Бакинський інженерний університет, Хирдалан, Азербайджан.

Вагіф Гасімов – д-р техн. наук, проф., декан факультету інформаційних та комп'ютерних технологій, Бакинський інженерний університет, Хирдалан, Азербайджан.

Artughrul Gayibov – Master, PhD Student, Lecturer of the Department of Information Technologies and Programming, Baku Engineering University, Khirdalan, Azerbaijan,
e-mail: agayibov@beu.edu.az, ORCID: 0009-0009-7349-0286, Scopus Author ID: 59565128900.

Vagif Gasimov – Doctor of Technical Science, Professor, Dean of the Faculty of Information and Computer Technologies, Baku Engineering University, Khirdalan, Azerbaijan,
e-mail: vgasimov@beu.edu.az, ORCID: 0000-0003-3192-4225, Scopus Author ID: 7801667181.