Erosion-SAM: semantic segmentation of soil erosion by water

Hadi Shokati, Andreas Engelhardt, Kay Seufferheld, Ruhollah Taghizadeh-Mehrjardi, Peter Fiener, Hendrik P. A. Lensch, Thomas Scholten

Angaben zur Veröffentlichung / Publication details:

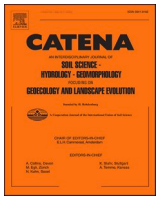
Shokati, Hadi, Andreas Engelhardt, Kay Seufferheld, Ruhollah Taghizadeh-Mehrjardi, Peter Fiener, Hendrik P. A. Lensch, and Thomas Scholten. 2025. "Erosion-SAM: semantic segmentation of soil erosion by water." CATENA 254: 108954. https://doi.org/10.1016/j.catena.2025.108954.

Catena 254 (2025) 108954

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena



Erosion-SAM: Semantic segmentation of soil erosion by water

Hadi Shokati [a,*], Andreas Engelhardt [b], Kay Seufferheld [c], Ruhollah Taghizadeh-Mehrjardi [a], Peter Fiener [c], Hendrik P.A. Lensch [b,d], Thomas Scholten [a,d]

[a] University of Tübingen, Department of Geosciences, Soil Science and Geomorphology, Germany
[b] University of Tübingen, Department of Computer Graphics, Germany
[c] Institute of Geography, Augsburg University, Augsburg, Germany
[d] University of Tübingen, Cluster of Excellence Machine Learning: New Perspectives for Science, Germany

ARTICLE INFO

Keywords:
Erosion monitoring
Fine-tuning
RADOLAN
SAM
Soil erosion
Transfer learning

ABSTRACT

Soil erosion (SE) by water threatens global agriculture by depleting fertile topsoil and causing economic costs. Conventional SE models struggle to capture the complex, non-linear interactions between SE drivers. Recently, machine learning has gained attention for SE modeling. However, machine learning requires large data sets for effective training and validation. In this study, we present Erosion-SAM, which fine-tunes the Segment Anything Model (SAM) for automatic segmentation of water erosion features in high-resolution remote sensing imagery. The data set comprised 405 manually segmented agricultural fields from erosion-prone areas obtained from the rain gauge-adjusted radar rainfall data (RADOLAN) for bare cropland, vegetated cropland, and grassland. Three approaches were evaluated: two pre-processing techniques— resizing and cropping — and an improved version of the resizing approach with user-defined prompts during the testing phase. All fine-tuned models outperformed the original SAM, with the prompt-based resizing method showing the highest accuracy, especially for grassland (recall: 0.90, precision: 0.82, dice coefficient: 0.86, IoU: 0.75). SAM performed better than the cropping approach only on bare cropland. This discrepancy is attributed to the tendency of SAM to overestimate SE by classifying a large proportion of fields as eroded, which increases recall by covering most of the eroded pixels. All three fine-tuned approaches showed strong correlations with the actual SE severity ratios, with the prompt-enhanced resizing approach achieving the highest R^2 of 0.93. In summary, Erosion-SAM shows promising potential for automatically detecting SE features from remote sensing images. The generated data sets can be applied to machine learning-based SE modeling, providing accurate and consistent training data across different land cover types, and offering a reliable alternative to traditional SE models. In addition, erosion-SAM can make a valuable contribution to the precise monitoring of SE with high temporal resolution over large areas, and its results could benefit reinsurance and insurance-related risk solutions.

1. Introduction

Soil erosion (SE) by water poses a significant threat to agricultural sustainability worldwide, as it removes fertile topsoil and leads to extensive soil degradation (Eltner et al., 2015; Quinton and Fiener, 2024; Scholten and Seitz, 2019). Beyond its impact on agriculture, SE imposes significant economic burdens and challenges the profitability of farming practices (Pimentel et al., 1995). To address these challenges, numerous traditional SE models have been developed (Jetten and Favis-Mortlock, 2006; Mitasova and Mitas, 2001), ranging from empirical to physically based models, varying in spatial scale from sub-plot to catchment level (Borrelli et al., 2021). The regression-based empirical Universal Soil Loss Equation (USLE, (Wischmeier and Smith, 1978)) and the revised USLE (RUSLE, (Renard, 1997)) are the most commonly used SE models (Karydas et al., 2014). However, collecting data to calibrate and validate such SE models is still a major challenge. Typically, runoff and sediment delivery measurements for individual events taken over extended periods at plot outlets with natural or artificial rainfall (Schindewolf and Schmidt, 2012) or at watershed outlets (Kohrell et al., 2023) are used to compare predicted and observed soil loss rates. However, when using outlet data for model validation, especially in large watersheds, the internal process dynamics of the watershed are

* Corresponding author.
E-mail address: hadi.shokati@uni-tuebingen.de (H. Shokati).

https://doi.org/10.1016/j.catena.2025.108954
Received 22 November 2024; Received in revised form 1 March 2025; Accepted 15 March 2025
Available online 23 March 2025

often overlooked. To address this, long-term erosion assessments using radionuclides as tracers (Walling et al., 2003; Wilken et al., 2020) or event-based high-resolution remote sensing data (Cândido et al., 2020) can be used to test and validate these internal dynamics. However, high-resolution remote sensing data has primarily been used in SE models to incorporate variables such as vegetation coverage, slope length and slope gradient (Ganasri and Ramesh, 2016) and to identify visible occurrence of erosion processes at the regional scale (Polovina et al., 2024) rather than for model validation and training. Fischer et al., (2018) advanced this field by implementing a targeted monitoring approach that tracks potential SE events over the Bavarian Tertiary Hills. Their study visually classified aerial orthophotos into four soil loss classes representing the spatial extent of SE damage within fields. These classes were used to validate event-based soil loss estimates using soil loss ratios from USLE-technology. The study showed a strong and statistically significant correlation between the visually classified and the predicted soil loss at the field scale. While the use of visually classified orthophotos proved to be effective for validating SE estimates, the manual classification process is still labor-intensive and time-consuming. Machine learning models offer a promising way to automate this process and extract more detailed information from remote sensing data. Machine learning models have been used in various studies to assess and map soil erosion (Agnihotri et al., 2021; Mosavi et al., 2020; Wang et al., 2021). However, traditional machine learning models occasionally provide suboptimal predictions (Bui et al., 2020; Gautam et al., 2021; Saha et al., 2021). Recent studies have therefore suggested a shift towards deep learning models, emphasizing that they are likely to be less constrained and more flexible than traditional machine learning models (Prasad et al., 2022). Previous studies have shown the significant potential of deep learning approaches in SE studies (e.g. Khosravi et al., 2023; Liu et al., 2023). However, deep learning models also require a considerable amount of ground-truth data not only for validation but also for effective training, while remotely sensed data labeled with SE features are scarce.

Recent advances in deep learning have introduced several innovative techniques to reduce reliance on big data. These include few-shot learning, self-supervised learning, transfer learning, and data augmentation, which data augmentation and transfer learning have become very popular in the field of deep learning applications in remote sensing (Safonova et al., 2023). In data augmentation, a dataset is artificially enlarged by creating new samples through transformations such as rotation, scaling and flipping. This technique improves the generalizability and robustness of the model and reduces overfitting (Shorten and Khoshgoftaar, 2019; Wagner et al., 2023). Transfer learning, on the other hand, uses previously learned features to improve the generalization ability of the data-based machine learning model. The importance of transfer learning in processing small remote sensing datasets has been demonstrated in various studies (Character et al., 2021; Hou et al., 2022; Lv et al., 2022; Saadeldin et al., 2022; Wang et al., 2018; Zhang et al., 2019). Building on this foundation, the Segment Anything Model (SAM) has emerged as a pioneering approach in image segmentation, offering substantial potential for transfer learning applications. Trained on an extensive dataset of over 1 billion masks derived from 11 million images, SAM demonstrates strong generalization capabilities across diverse datasets and generates precise segmentation masks (Kirillov et al., 2023). For instance, Song et al., (2024) refined SAM by fine-tuning it for crop segmentation, demonstrating its superior performance compared to five other advanced segmentation methods. Similarly, the ClassWise-SAM-Adapter (CWSAM) was developed by Pu et al., (2024) through fine-tuning SAM for land cover classification using spaceborne Synthetic Aperture Radar (SAR) imagery. Their results showed that CWSAM outperformed traditional state-of-the-art semantic segmentation algorithms.

Despite these advances in deep learning and transfer learning techniques, automated segmentation of SE at the pixel level has not been tested so far. For the first time, this study fine-tunes the pre-trained SAM as a transfer learning technique using aerial orthophotos to automatically identify the erosion and deposition forms at the pixel level. The proposed Erosion-SAM model tests different data preprocessing methods and incorporates pixel-based predictions, thereby enhancing both the processing and prediction accuracy. Overall, this study aims to 1. develop a method to automate pixel-based SE feature segmentation, thereby improving the scalability and accuracy of SE detection, and 2. evaluate the accuracy of the fine-tuned model across different types of land cover, namely grassland, vegetated cropland, and bare cropland. By customizing SAM with aerial images of agricultural fields affected by erosive rainfall, we aim to qualitatively identify and delineate visible traces of SE processes at a high spatial resolution of 20 cm with minimal training data.

2. Materials and methods

2.1. Study area

The study area is located in the south-eastern part of Bavaria in south-eastern Germany (Fig. 1). The region was selected based on two main criteria: (i) availability of high-resolution orthophotos (0.2 m x 0.2 m) taken after high-intensity rainfall events, and (ii) generally intensive agricultural practices in hilly terrain comprising cropland and grassland. The test sites within the study area shown in the orthophotos include locations with a mean annual precipitation of 800 to 1200 mm/year (Auerswald et al., 2009) and a mean temperature between 6 °C and 11 °C (German Weather Service, DWD[1]). Dominant soils are Cambisols and Luvisols (Gocke et al., 2021; Iuss and Fao, 1999), whereas slopes range between 0 % and 40 %. Typical crops are winter wheat, maize, winter barley and potatoes, with the erosion-prone maize fields dominating the data set.

2.2. Orthophotos

The Orthophotos used in this study were acquired in a research project of the Bavarian Agency for Agriculture, as documented in Fischer et al., (2018). The general idea was to acquire orthophotos after heavy rainfall events in erosion-prone areas for the period between May and September in 2011 and 2012. To decide for the timing of the data acquisition, the RADOLAN rainfall data of the DWD with a spatial resolution of 1×1 km^2 were used. In case of rainfall events (at least 10 mm total rain depth or maximum 30 min rain intensity higher than 10 mm/h) that indicates a high erosion potential, flights with a small airplane were performed not longer than 30 days after the determined erosive rainfall event. The aerial images were taken at a flight height of about 900 m with an onboard Nikon D200 camera. This resulted in an image pixel resolution of 0.2 m x 0.2. The images were georeferenced using ArcGIS Pro 3.1.3 to ensure accurate spatial alignment with InVeKoS data (Integrated Administration and Control System). We manually identified and masked 405 individual parcels (Fig. 1) from the aerial imagery that exhibited erosion and deposition features such as ephemeral gullies, rills and sediment fans. The fields were categorized into grassland (n = 128) and cropland (n = 277), whereas the latter was then subdivided into vegetated cropland (n = 131) and bare cropland (n = 146).

2.3. Erosion classification

For each field in the categories, an expert performed a visual interpretation to manually segment SE features within the respective field (Fig. 1) using ArcGIS Pro 3.1.3, delineating areas where erosion and deposition had occurred irrespective of the land cover type. Pixels showing erosion or deposition on cropland or grassland were classified and labeled as “erosion-deposition”, while all other pixels within the

[1] https://www.dwd.de.

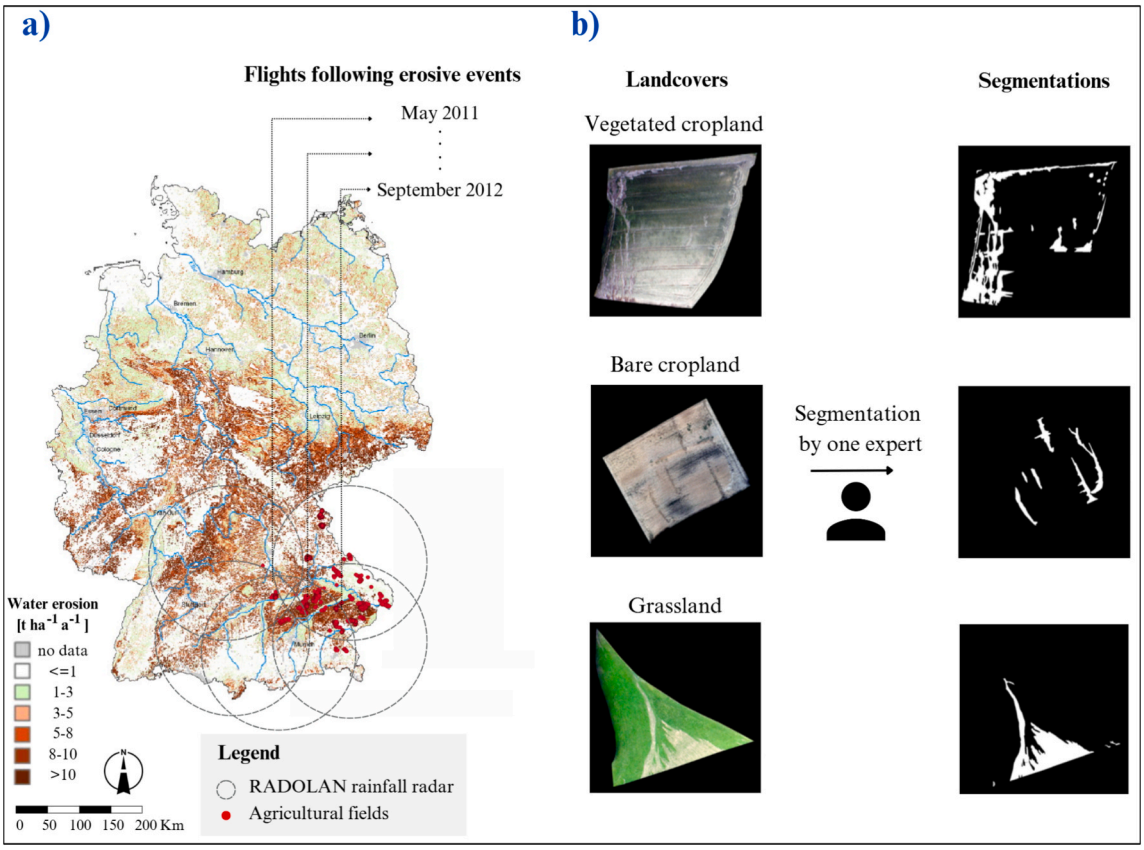


Fig. 1. A) soil erosion map of Germany in a 75 × 75 m raster (Auerswald et al., 2009) and b) examples of manual segmentation of agricultural fields with erosive rainfall, showing erosion and deposition features across different land covers.

field were labeled as "undisturbed areas". From each land cover type, 12 fields were randomly selected and segmented by three independent human classifiers to quantify the accuracy of the manual segmentation. The area with erosion and deposition features varied by 8 % for grassland, 13 % for vegetated cropland and 17 % for bare cropland between the four replications.

2.4. Model construction

2.4.1. The Segment Anything Model (SAM)

The Segment Anything Model (Meta AI Inc., USA) marks a significant advancement in image segmentation technology. SAM consists of three main components (Fig. 2). The image encoder, built on a robust Vision Transformer (ViT) backbone, extracts image features. The prompt encoder, a lightweight embedding module, generates both sparse and dense embeddings from various prompt inputs (points, boxes or masks). Finally, the mask decoder combines the outputs of the image encoder and the prompt encoder to generate the final segmentation masks.

SAM is available in three variants: base (ViT-B), large (ViT-L) and huge (ViT-H) with 91, 308 and 636 million parameters respectively (Kirillov et al., 2023). These models differ in terms of computational requirements and architectural complexity. In this study, we used the ViT-B variant of the SAM to reduce computational demands.

2.4.2. Fine-tuning SAM

Despite SAM's high accuracy and little to no training data requirements, its performance can be limited under specific domain conditions. To address this, we employed the fine-tuning technique, a form of transfer learning. Fine-tuning involves adapting a pre-trained model (including its architecture and weights) to a specific task instead of training a new model from scratch. Fine-tuning enhances performance, reduces the need for large training data sets, lowers computational costs, and saves time (Rawat and Wang, 2017). Pre-training on a broad data set is particularly effective when the task-specific data set is small or lacks extensive labeled data.

When fine-tuning SAM, we focused on fine-tuning the mask decoder (Fig. 2) due to the complexity and the high number of parameters of the image encoder. The mask decoder has a lightweight design (Sun et al., 2024), i.e., it contains fewer parameters and a simpler architecture, which makes fine-tuning easier, faster, and more memory efficient.

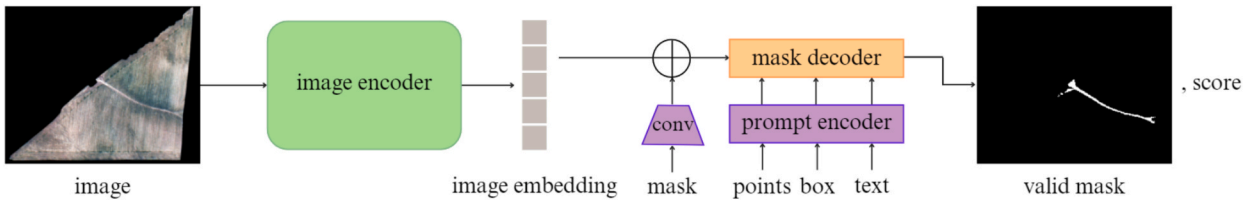


Fig. 2. Overview of the Segment Anything Model (SAM) (adapted from Kirillov et al., (2023)). SAM uses a heavyweight image encoder to create an image embedding, which is combined with input prompts and passed through a prompt encoder. The features are then decoded into object masks, with corresponding confidence scores.

2.4.3. Input prompts

The architecture of SAM enables the integration of human prompts, which increases the effectiveness of human-in-the-loop annotation. These prompts can be multimodal: points on the area to be segmented, a bounding box around the object or textual descriptions of what is to be segmented. For segmentation, we used point prompts in the training phase of our fine-tuned SAM. Specifically, 80 points were used, each separated by a minimum radius of 10 pixels. However, in the testing phase, we used two different approaches to comprehensively evaluate the performance of the model. First, we tested the model without prompts, as manual SE segmentation is challenging and in real-world scenarios often no pre-existing masks are available. Second, we considered a scenario where only minimal user input is possible in real-world applications. In this approach, point prompts were used in testing to mirror the process used in training.

2.4.4. Image pre-processing approaches

To address the variability in the sizes of agricultural fields, we considered two methods of pre-processing to standardize the image sizes for the model.

Uniform Resizing: This method involves resizing all field images to 256 × 256 pixels before inputting them into the model. After the model has made its predictions, the images are resized back to their original dimensions. While this approach contributes to the standardization of the input dimensions, it can lead to distortions, especially in hilly terrain, which may result in unrealistic erosion patterns or the loss of important details during downscaling. These effects can affect the reliability of model predictions. To minimize such problems, it is important to choose a resizing dimension that minimizes distortion and ensures that the adjusted sizes remain close to the actual dimensions of the fields.

Image Cropping: In this approach, the training images are divided into multiple 256 × 256 image patches with a step size of 256 pixels. To ensure compatibility with the specified patch size during image processing, we used zero padding, in which the image size is expanded by appending zeros to facilitate the division into uniform patches. To balance the input between erosion and non-erosion pixels, patches consisting entirely of non-erosion pixels are excluded from the modeling process. Each image patch is processed individually by the model. After prediction, the image patches are reassembled to reconstruct the original image. To avoid jagged predictions at the corners of the test image patches, the step size for cropping the test data was reduced to 128 pixels, creating an overlap between the merged patches. A blending technique using a simple second-order spline window function is then applied. This method weights the pixels in the overlapping areas when merging the patches and provides smoother transitions between neighboring patches.

Both resizing and cropping approaches aim to ensure consistent image input sizes for the model while accommodating the varying dimensions of the agricultural fields. These approaches were compared to determine which yields the best performance and accuracy.

2.4.5. Data augmentation

The choice of data augmentation technique depends on the quantity, quality and type of data (Safonova et al., 2023). In this study, due to the pixel values being zero in the corners of the input images, certain data augmentation methods, such as random rotation or random cropping, might result in augmented images consisting entirely of zero values, potentially causing errors. Therefore, we used horizontal and vertical flips methods of data augmentation, which do not encounter this issue.

2.4.6. Field-based soil erosion classification for model testing

Erosion-SAM labels each pixel of a high-resolution orthophoto as erosion–deposition or undisturbed areas based on the presence or absence of erosion and deposition features. Even though this approach provides a detailed spatial resolution, field-based assessment is a valuable complementary to the analysis. A field-based SE classification provides critical insights by identifying which land cover types are most susceptible to erosion and deposition, evaluating how different management practices affect soil loss and accumulation on fields, and assessing how different environmental conditions influence erosion processes. Field-based classifications have also been explored by researchers such as Fischer et al., (2018). In this study, we build on this concept by implementing a field-based approach to create a data set for testing our segmentation model using 15 % of our images (n = 60). The percentages of eroded areas within each agricultural field are calculated using our labeled ground-truth data pixels, and fields are subsequently categorized into three SE severity classes: Class 1 (0–10 % erosion), Class 2 (10–30 % erosion), and Class 3 (more than 30 % erosion). However, it is important to note that our pixel-based classification cannot directly compare to the original classification based on estimating the area per field affected. The same classification method is then applied to the SE maps produced by the model, allowing a direct comparison between the predicted and manually segmented results.

2.4.7. Implementation details

The data from all 405 agricultural fields were randomly divided into training, validation and test sets. Specifically, 70 % of the fields from each land cover category were included in the training set (n = 285), 15 % in the validation set (n = 60) and 15 % in the test set (n = 60). This stratified splitting approach ensured that the distribution of land cover types in the training, validation and test subsets was balanced, resulting in a representative data set for model evaluation. In the cropping approach, to avoid spatial correlations, we first divided the agricultural fields into training, validation and test sets and then performed cropping to ensure that patches from the same field did not appear in different sets. During the training phase, the model adjusts its internal parameters (e.g. weights and biases) to learn patterns in the data using back-propagation. After each iteration, the validation set, which is kept separate from the training data, is used to evaluate the model's performance. This step helps in optimizing hyperparameters (e.g., learning rate and batch size) and preventing overfitting (Vabalas et al., 2019; Ying, 2019). The model with the lowest validation loss is selected as the best-performing model, as it is likely to generalize well to unseen data. Finally, the selected model is evaluated using the test set which has remained isolated from all previous stages, providing an unbiased measure of its ability to generalize to new data.

To further increase the robustness of the model, horizontal and vertical flips were applied to the training set as data augmentation methods. To further increase the robustness of the model, horizontal and vertical flips were applied to the training set as data augmentation methods. By applying data augmentation, the number of training data was doubled. Data augmentation was only applied to the training set so that the validation and test data remained unchanged in both resizing and cropping approaches. In the resizing approach, the number of training data increased from 285 to 570 samples after applying the data augmentation methods, while the number of validation and test samples remained unchanged at 60 each. Similarly, in the cropping approach, the number of training samples increased from 1,774 to 3,548, while the number of validation and test samples remained unchanged at 314 and 1,060 respectively. The model was trained for 50 epochs, with each epoch representing a complete pass of the dataset through the segmentation algorithm. To optimize the parameters in the decoder module, the combination of dice and cross-entropy losses (DiceCELoss) was used as the loss function, which uses their average for the total loss and is a standard choice for training (Hadlich et al., 2023). The cross-entropy loss is given by:

$$L_{CE} = -\frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c \log S_i^c \quad (1)$$

where g_i^c is the ground truth binary indicator of the class label c of pixel i, and S_i^c is the corresponding predicted segmentation probability. The dice loss is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c S_i^c}{\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c + \sum_{c=1}^{C}\sum_{i=1}^{N} S_i^c} \quad (2)$$

We used the Adam optimizer, which has been very popular in the field of machine learning in recent years due to its easy fine-tuning and strong adaptability (Liu et al., 2023). This optimization algorithm adaptively adjusts the learning rate, enabling efficient and effective model training. The model was trained with a batch size of 4 and an initial learning rate of 0.001. To further refine the training process, we used the learning rate scheduler, which dynamically adjusted the learning rate based on the validation loss. Whenever the validation loss did not improve in five consecutive epochs, the learning rate was reduced by a factor of 0.1. This adaptive approach allowed us to maintain an optimal balance between convergence speed and model performance throughout the training process.

All experiments were conducted using Python and the PyTorch framework on a computer configured with an NVIDIA GeForce RTX 4070 Ti GPU and running the Windows 11 operating system. The parameter settings used during the fine-tuning process are summarized in Table 1. These parameters were carefully chosen to optimize the performance of the model while overcoming the particular challenges posed by the erosion dataset and computational resources.

2.5. Evaluation metrics

To evaluate the accuracy of the model in segmenting SE features, four evaluation metrics including precision, dice coefficient, recall and intersection-over-union (IoU) were utilized as defined by the following equations (Vinayahalingam et al., 2023):

$$\text{precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{dice coefficient} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (6)$$

Where TP, FP and FN represent the number of true positives, false positives and false negatives pixel labels, respectively.

Precision, dice coefficient, recall and IoU are key performance metrics in image segmentation. Precision describes how well the model identifies erosion pixels as positive predicted values, indicating the proportion of correctly predicted erosion pixels. In addition, the dice coefficient measures the similarity between the ground truth and the predicted data set and is commonly used in image segmentation. It ranges from 0 to 1, where 1 represents a perfect overlap and 0 represents no overlap. Similarly, IoU, also known as the Jaccard index, includes the data of TP, FP and FN. Our segmentation process for erosion features involves two steps, localization of the area in an orthophoto and classification of an area as erosion or non-erosion. In addition, IoU is scale-invariant and evaluates how well the predicted erosion area matches the ground truth. Finally, recall, which reflects the sensitivity of the model, describes the proportion of correct erosion pixels in the total number of erosion pixels in the ground truth data. Fig. 3 shows the flowchart of the work.

Table 1
The parameter settings used during the fine-tuning process.

Parameter	Settings
Data split	Training: 70 % (n = 285), Validation: 15 % (n = 60), Testing: 15 % (n = 60)
Loss function	DiceCELoss (Dice + Cross-Entropy loss)
Optimizer	Adam
Learning rate	Initial: 0.001, Decay: 0.1
Batch size	4
Epochs	50
Data augmentation	Horizontal and vertical flips
Weight initialization	Pretrained weights (SAM)
Framework	PyTorch (Python)
Hardware	NVIDIA GeForce RTX 4070 Ti, Windows 11

To assess the field-based predictions, two metrics including the coefficient of determination (R^2) and mean absolute error (MAE) were also calculated as follows:

$$R^2 = 1 - \frac{ResidualSumofSquares(RSS)}{TotalSumofSquares(TSS)} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (7)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (8)$$

Where n is the number of data, y_i is the observed value, $\hat{y}_i$ is the corresponding predicted value and $\bar{y}$ is the mean of the observed values.

3. Results and discussion

3.1. Training and validation losses

The model was trained and validated over 50 epochs (Fig. 4) reflecting the number of times the data set passes through the segmentation algorithm. The results showed that for the resizing approach, the training loss dropped significantly during the first epochs. After this sharp decrease, the rate of decrease slowed down and finally reached a value below 0.23 in the last epoch. The validation loss for the resizing approach also showed a general downward trend until epoch 20, indicating that the model generalizes well to unseen data, while it stabilized after this point, indicating an overfitting to the training data after this inflection point (Shorten and Khoshgoftaar, 2019).

In the cropping approach, training losses followed a similar pattern to the resizing approach, with a general decrease across epochs, reaching approximately 0.26 at epoch 50. Similarly, validation losses displayed a consistent downward trend with a minimum of 0.25 at epoch 30. Following epoch 30, the validation loss stabilized, indicating that additional training would likely not improve performance and suggesting overfitting to the training data beyond this point. The validation loss curve showed signs of overfitting later for the cropping method (after 30 epochs) than for the resizing method (after 20 epochs). This discrepancy stems from the different pre-processing approaches. In the resizing method, the images were resized to 256 × 256 pixels, whereas in the cropping method, each image was divided into multiple 256 × 256 patches depending on the original dimensions. As a result, the cropping approach generated a larger training dataset (3548 patches for the cropping approach compared to 570 patches for the resizing approach, both after applying data augmentation techniques), delaying the occurrence of overfitting (Elhage et al., 2022; Safonova et al., 2023; Shorten and Khoshgoftaar, 2019).

The lowest validation loss for the resizing approach was achieved in epoch 20 with a value of 0.24, while it was achieved for the cropping approach in epoch 30 with a value of 0.28. Consequently, the models from these respective epochs were selected as optimal models for the resizing and cropping approaches. Overall, the training and validation losses for the resizing approach were slightly lower than for the cropping approach, indicating a better generalization performance for the

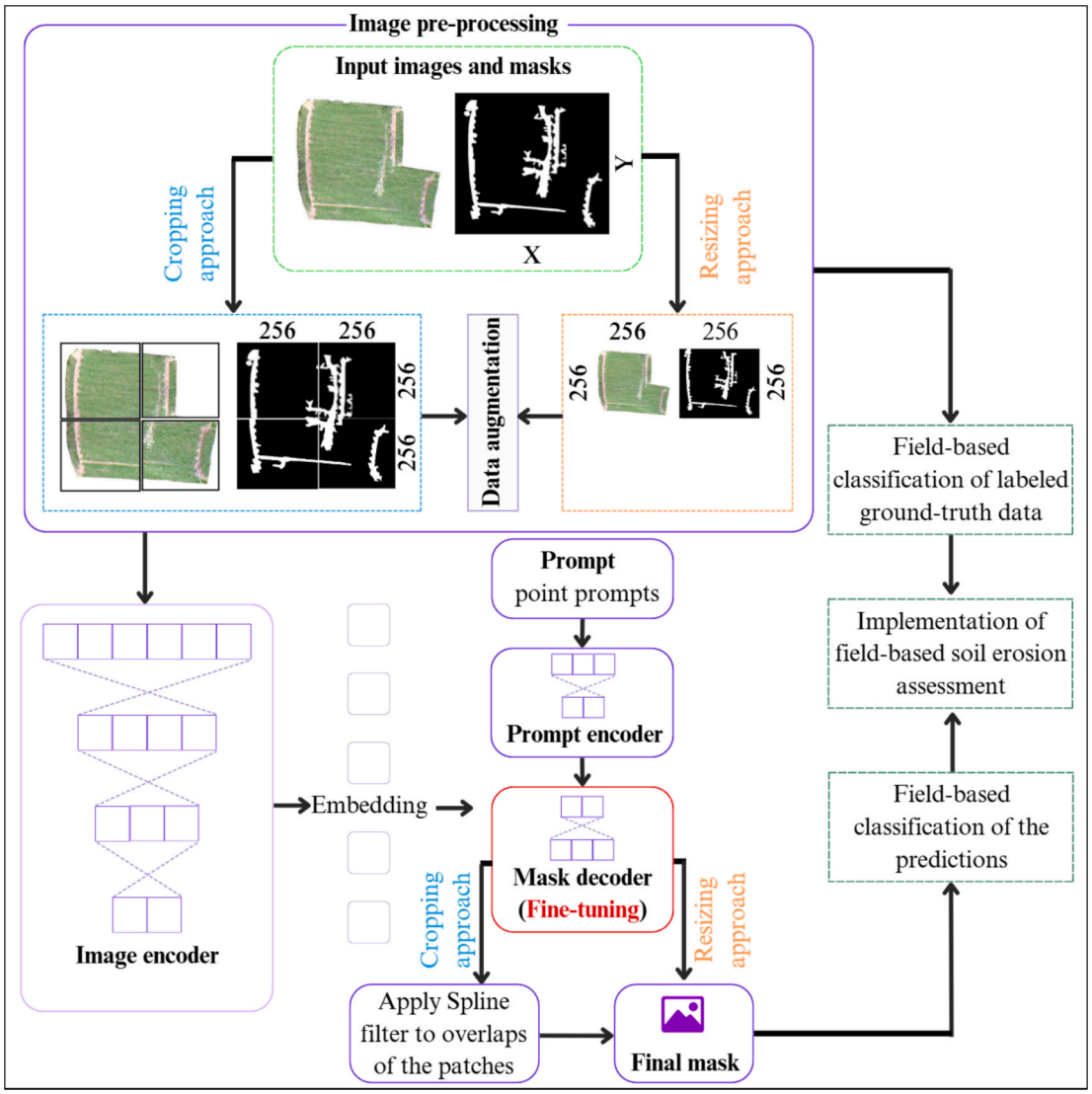


Fig. 3. Overview of the network for fine-tuning the mask decoder of the Segment Anything Model. The input images and masks are preprocessed using two approaches: Cropping to 256 × 256 patches (blue box) or resizing to 256 × 256 pixels (orange box). Both approaches are followed by data augmentation. The images are then passed through the image encoder, where they are transformed into embeddings. These embeddings, in combination with point prompts, are fed into the fine-tuned mask decoder (red box) to generate the final segmentation mask. In the cropping approach, after generating a mask for each patch, they are merged, and a spline filter is applied to the overlaps of the patches to provide smoother transitions between neighboring patches.

resizing method.

3.2. Performance of the model on the validation set

The values of the performance metrics across the training epochs for both the resizing and cropping approaches are shown in Fig. 5. Both approaches showed a consistent ascending course in their metrics, indicating a progressive improvement in model accuracy throughout the training process. Remarkably, the improvement reached a plateau in middle training epochs and behaved similarly to the loss function. This stabilization indicates that the models have achieved optimal performance and further training beyond this point would likely not yield significant improvements.

In the first few epochs, considerable fluctuations were observed in all metrics, which can be attributed to the weight adjustments of the model and the extensive learning phases. As training progressed, the fluctuations decreased, indicating that the learning process became more refined and the model gradually recognized patterns in the data. For the resizing approach, precision was consistently higher than the other metrics throughout training, reaching a value of 0.78 at the best epoch (epoch 20). At the same point, the dice coefficient and recall were 0.77 and 0.76, respectively. IoU remained the lowest with a value of 0.63, indicating a relatively high overlap between predicted and actual SE pixels. For the cropping approach, precision also remained the highest of all evaluation metrics, reaching 0.81 in the best epoch (epoch 30), followed by dice coefficient, recall and IoU at 0.79, 0.76 and 0.66, respectively.

3.3. Performance analysis of the SAM and fine-tuned models across different land covers

The performance of the models on test data for grassland, vegetated cropland and bare cropland and the overall performance are shown in Table 2. In vegetated cropland, SAM showed its weakest performance compared to bare cropland and grassland, with a recall of 0.52, a precision of 0.19, a dice coefficient of 0.28 and an IoU of 0.16. The significant discrepancy between recall and precision indicates that although SAM detected a large proportion of eroded areas on vegetated

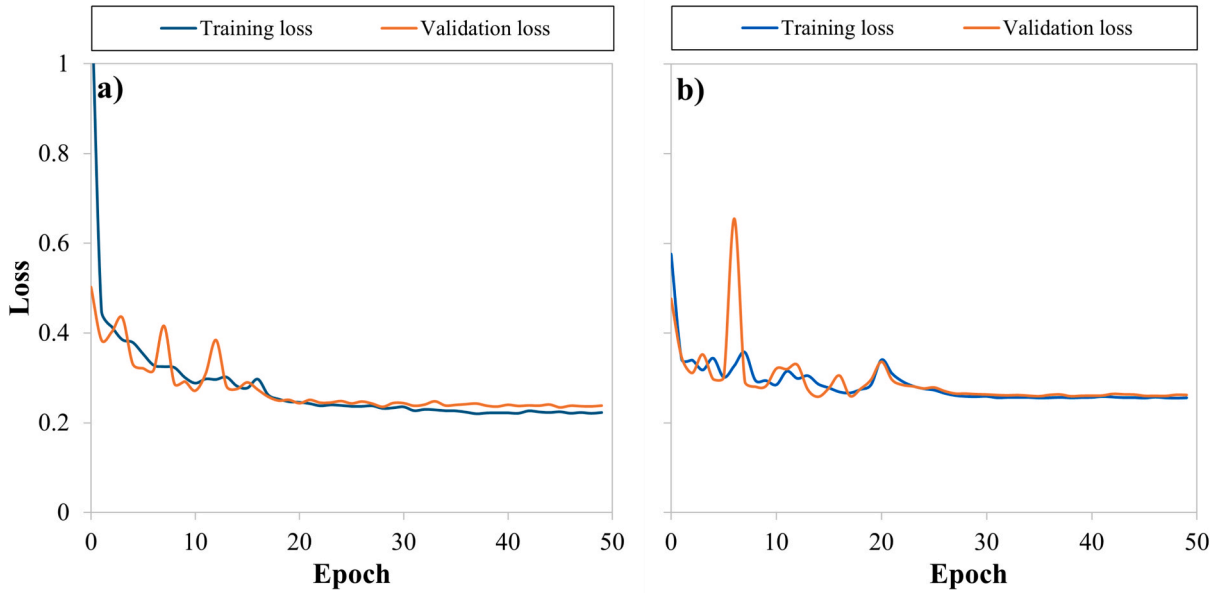


Fig. 4. Training and validation losses over the training epochs for a) resizing and b) cropping approaches.

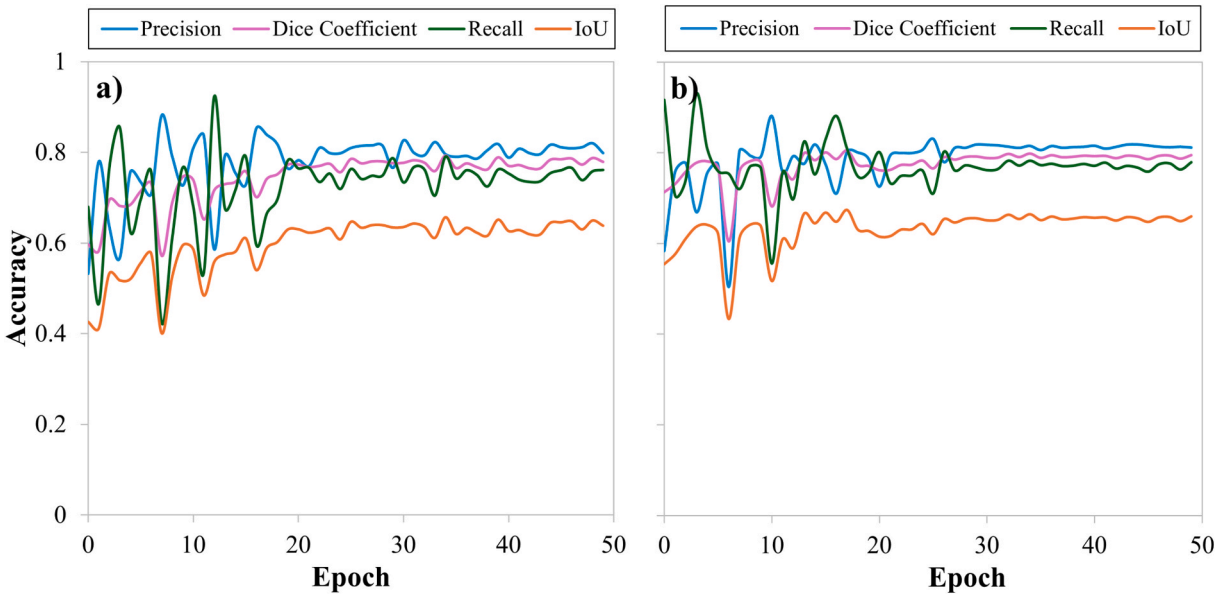


Fig. 5. Performance of the fine-tuned model on the validation set over the training epochs for a) resizing and b) cropping approaches.

cropland, many of the identified regions were false positives, as the model often misclassified non-eroded areas as erosion areas, such as crops, bare patches in between or tractor tracks. This leads to an overestimation of eroded areas (Ishikawa et al., 2021; Sun et al., 2024) (Fig. 6 a).

The fine-tuned model, which utilized the cropping approach for preprocessing the data, achieved its best performance in grassland with recall, precision, dice and IoU values of 0.77, 0.68, 0.72 and 0.56, respectively. This can be attributed to the more uniform ground surface in grassland compared to bare cropland and vegetated cropland, which facilitates the detection of SE. In addition, the deposition areas in the grassland form a clearer contrast to the vegetated, non-eroded or sediment-covered areas. However, for vegetated cropland, segmentation accuracy decreased in all evaluation metrics, with recall, precision, dice, and IoU values of 0.59 each and an IoU of 0.42. This decrease in performance can be explained by the presence of crops, tractor tracks or bare patches, which the model occasionally misclassified as SE. The lowest performance of the cropping approach was observed for bare cropland where the recall, precision, dice, and IoU values were 0.41, 0.31, 0.35 and 0.21, respectively. This lower accuracy is due to the similarity between eroded and natural soil, which makes it difficult for the model to accurately segment SE in bare cropland.

In the resizing approach without using input prompts in the testing phase, the performance for grassland stood out with the highest score, including a recall of 0.84, a precision of 0.77, a dice coefficient of 0.80 and an IoU of 0.67. These metrics indicate that the model performs well in detecting SE in grassland areas, achieving high sensitivity and decent precision. On the other hand, the resizing approach yielded moderate results in vegetated cropland with a recall of 0.57, precision of 0.65, a dice coefficient of 0.61, and IoU of 0.43. While the higher precision suggests that the model is quite accurate in predicting SE, the lower recall suggests that it may miss certain eroded areas due to complex

Table 2
The performance of the models on test data across different land cover types.

Model	Land cover	recall	precision	dice	IoU
Original SAM	grassland	0.63	0.48	0.54	0.37
	vegetated cropland	0.52	0.19	0.28	0.16
	bare cropland	0.53	0.37	0.44	0.28
	total	0.56	0.35	0.43	0.27
Cropping approach	grassland	0.77 (+ 0.14)	0.68 (+ 0.20)	0.72 (+ 0.18)	0.56 (+ 0.19)
	vegetated cropland	0.59 (+ 0.07)	0.59 (+ 0.40)	0.59 (+ 0.31)	0.42 (+ 0.26)
	bare cropland	0.41 (− 0.12)	0.31 (− 0.06)	0.35 (− 0.09)	0.21 (− 0.07)
	total	0.58 (+ 0.02)	0.52 (+ 0.17)	0.55 (+ 0.12)	0.38 (+ 0.11)
Resizing approach (without prompt)	grassland	0.84 (+ 0.21)	0.77 (+ 0.29)	0.80 (+ 0.26)	0.67 (+ 0.30)
	vegetated cropland	0.57(+ 0.05)	0.65 (+ 0.46)	0.61 (+ 0.33)	0.43 (+ 0.27)
	bare cropland	0.50 (− 0.03)	0.46 (+ 0.09)	0.48 (+ 0.04)	0.31 (+ 0.03)
	total	0.63 (+ 0.07)	0.62 (+ 0.27)	0.63 (+ 0.20)	0.46 (+ 0.19)
Resizing approach (with prompt)	grassland	0.90 (+ 0.27)	0.82 (+ 0.34)	0.86 (+ 0.32)	0.75 (+ 0.38)
	vegetated cropland	0.80 (+ 0.28)	0.74 (+ 0.55)	0.77 (+ 0.49)	0.62 (+ 0.46)
	bare cropland	0.73 (+ 0.20)	0.72 (+ 0.35)	0.72 (+ 0.28)	0.57 (+ 0.29)
	total	0.81 (+ 0.25)	0.76 (+ 0.41)	0.78 (+ 0.35)	0.64 (+ 0.37)

landscape features such as crops, bare patches or tractor tracks, leading to under-detection rather than misclassification. The lowest performance of the resizing approach was observed in bare cropland with a recall of 0.50, a precision of 0.46, a dice of 0.48 and an IoU of 0.31. This is consistent with previous observations in the cropping approach regarding the difficulty of detecting SE in bare cropland.

The resizing approach, having demonstrated superior performance across all land cover types, was selected as the optimal model for SE segmentation. To further improve the practical applicability of the model, user-specific prompts were integrated during the testing phase, allowing the user's knowledge to be incorporated for better identification of SE sites. This improvement takes advantage of the practical fact that field operators often have valuable contextual information about potential SE areas. The implementation of these prompts showed varying effectiveness across different land cover types. In grassland, the approach reached its optimal performance with impressive values: a recall of 0.90, a precision of 0.82, a dice coefficient of 0.86 and an IoU of 0.75. For vegetated cropland, the performance metrics dropped to 0.80, 0.74, 0.77 and 0.62 for recall, precision, Dice coefficient and IoU, respectively. Bare cropland proved to be the most challenging environment where the approach achieved its lowest performance: a recall of 0.73, a precision of 0.72, a dice coefficient of 0.72 and an IoU of 0.57. Nevertheless, the integration of point prompts significantly improved the performance of the model in detecting SE areas, especially in bare cropland environments where identification was previously a major challenge. This marked difference in performance between land cover types suggests that the environmental context significantly influences the segmentation capabilities of the model, with the resizing approach using point prompts effectively mitigating many land cover-specific challenges.

3.4. Comparison of the effectiveness of SAM and fine-tuned models in detecting soil erosion

A comprehensive evaluation showed that the resizing approach, when enhanced with point prompts during the testing phase, consistently outperformed all other tested models across different land covers (Fig. 6 b and c). Without prompts, the resizing approach remained in second place across all land covers, followed by the cropping approach, with a notable exception in bare cropland environments, where the original SAM performed better than the cropping approach. Despite the higher accuracy in bare cropland, the main problem of SAM was that it tended to overestimate SE by misclassifying a large proportion of fields as eroded, leading to numerous false positives and over-segmentation of SE (Ishikawa et al., 2021; Sun et al., 2024) (Fig. 6 d). This limitation was emphasized by the significant discrepancy between recall and precision observed in SAM across all land covers. In contrast, the resizing and cropping approaches showed a more balanced relationship between recall and precision, indicating higher accuracy in SE classification.

On average across all land covers, the prompt-enhanced resizing approach delivered the most balanced and effective results, achieving the highest scores for recall (0.81), precision (0.76), dice coefficient (0.78) and IoU (0.64). These results revealed a clear hierarchy of performance: the prompt-enhanced resizing approach ranked first, followed by without prompt resizing, cropping and finally the original SAM. These results are consistent with those of Sun et al., (2024), who had previously demonstrated the superiority of fine-tuned SAM implementations over the original SAM. The higher performance of the resizing approach compared to the cropping approach can be attributed to the ability of the resizing approach to better preserve spatial relationships and important features, as it processes entire fields at once, as opposed to the cropping approach, which divides the field into several smaller patches. Erosion and deposition often cause large, visible features that extend across entire agricultural fields extending from one side of the image to the other. In the cropping approach, these features are usually spread across multiple image patches and not limited to a single patch. Since the predictions in the cropping approach are made independently for each patch and later combined, this method does not preserve spatial relationships and important features despite the expected advantages of increasing the sample size and preserving details. Consequently, the resizing approach proves to be more effective in identifying signs of erosion and sedimentation.

3.5. Implementation of field-based soil erosion assessment

The field-based correlation analysis (Fig. 7) revealed that the SE ratios predicted by the resizing approach with point prompts showed the highest agreement with the ground truth data ($R^2 = 0.93$, MAE = 1.5), followed by the resizing approach without prompts ($R^2 = 0.70$, MAE = 3.19) and the cropping approach ($R^2 = 0.58$, MAE = 3.74).

Confusion matrices show the classification performance for SE severity in three different classes (Fig. 8). The prompt-based resizing approach showed superior accuracy, especially for classifying the lower severity classes. For Class 1 (minimal erosion), an accuracy of 97.5 %

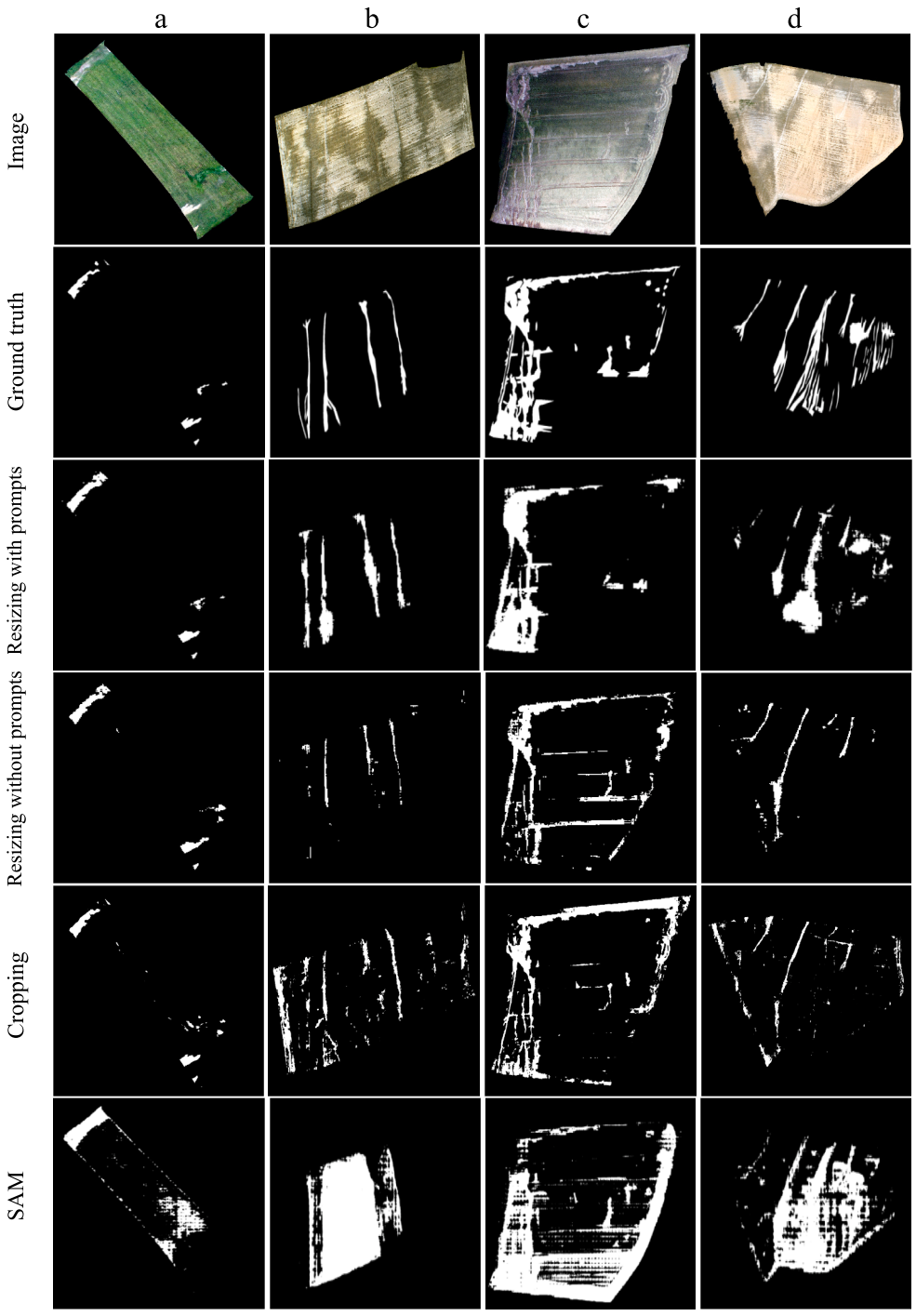


Fig. 6. Overview of soil erosion segmentation with the original SAM and the fine-tuned models.

was achieved, with only 2.5 % misclassified as Class 2. Class 2 (moderate erosion) also performed solidly, with only 5.9 % of fields being misclassified as Class 1. For Class 3 (severe erosion), the model achieved a moderate accuracy of 66.7 %, with the remaining 33.3 % being misclassified as Class 2.

The standard resizing approach (without point prompts) maintained high accuracy for Class 1 (95.0 %), with 5.0 % misclassification as Class 2. However, performance for Class 2 decreased, with only 64.7 % of fields correctly classified and 35.3 % incorrectly classified as Class 1, suggesting that it is difficult to distinguish between Class 2 and Class 1. The performance for Class 3 was similar to the prompt-enhanced approach with 66.7 % accuracy and 33.3 % misclassification as Class 2.

The cropping approach showed a comparatively lower performance for Class 1 with an accuracy of 90.0 %. For Class 2, the accuracy of the model decreased, with 76.5 % of fields correctly classified as Class 2 and 23.5 % incorrectly classified as Class 1, again indicating some confusion between Class 2 and Class 1. The accuracy for Class 3 remained the same as the other approaches at 66.7 %, with 33.3 % incorrectly classified as Class 2.

Overall, it can be concluded that both standard resizing and the cropping approaches tend to underestimate the severity of SE compared to the ground truth values. In contrast, the prompt-enhanced resizing approach achieved consistently higher classification accuracy at all severity classes, excelling particularly in distinguishing between Class 1

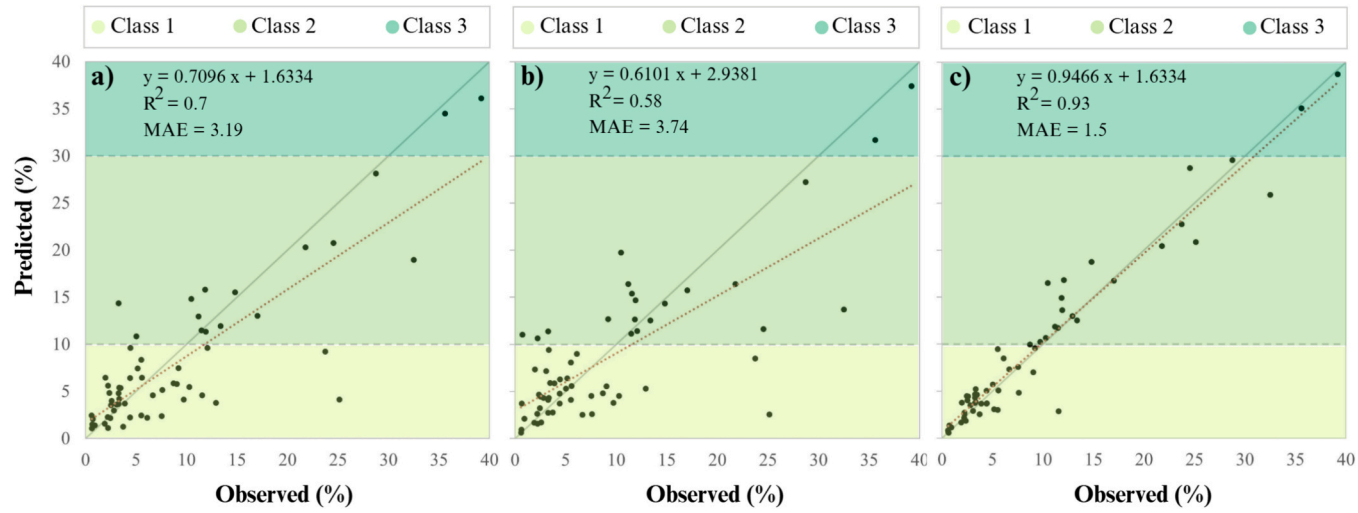


Fig. 7. Prediction of field-based soil erosion ratios using a) without prompt resizing, b) cropping and c) with prompt resizing approaches.

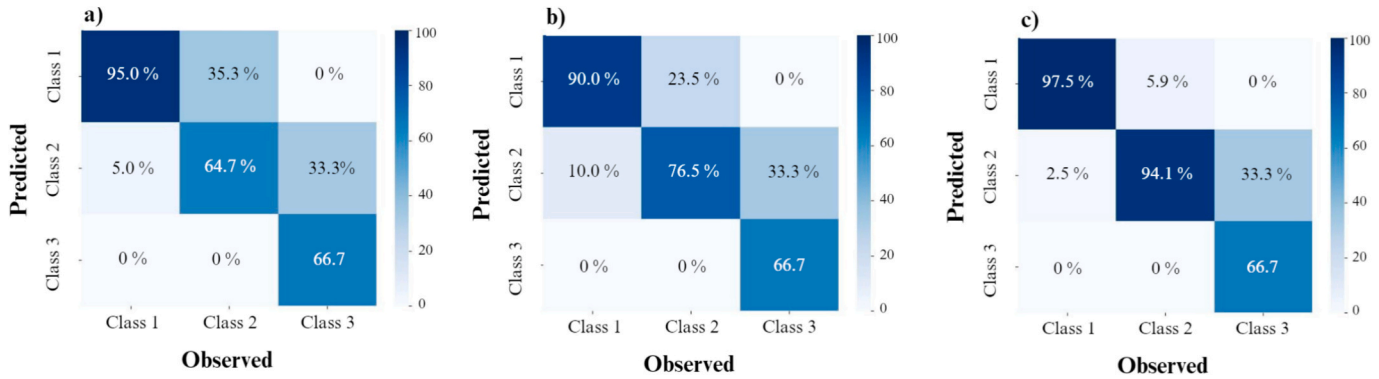


Fig. 8. Confusion matrix showing the percentage of correctly predicted and incorrectly predicted field-based soil erosion classes using a) without prompt resizing, b) cropping and c) with prompt resizing approaches.

and Class 2 categories.

4. Conclusion

This study introduces the innovative Erosion-SAM model, which automatically detects erosion and deposition features for each pixel of high-resolution orthophotos. By fine-tuning the Segment Anything Model for bare cropland, vegetated cropland, and grassland, Erosion-SAM overcomes significant challenges in SE detection. Three methodological approaches were evaluated: resizing, cropping, and an improved resizing method with user-defined prompts during testing. Our comparative analysis showed that fine-tuning significantly improves the accuracy of the SAM model in detecting areas of erosion and deposition and enables accurate segmentation of SE features in remote sensing data. These results highlight the potential of the Erosion-SAM model in overcoming the challenges of SE detection, especially when limited labeled data is available. It provides the unique opportunity to obtain validation data for traditional physically-based and empirical SE models as well as large training data sets for data-based erosion models using machine learning approaches. This innovation enables not only detailed environmental monitoring but also information for policy decisions, agricultural practices and risk assessments for erosion-prone areas, including applications in insurance and land management. Although the results of SE segmentation are promising, this study is the first of its kind and further research is needed. Future studies should investigate the training of the Erosion-SAM model with larger and more accurate data sets. Since our current labeled data were segmented by only one expert, incorporating multiple expert annotations would standardize the data and potentially improve model accuracy. In addition, the relatively small data set used for training underscores the need for expanded data sets to improve the robustness and generalizability of the model. Future research could investigate the integration of high-resolution satellite remote sensing data such as WorldView-3, GeoEye-1 and KOMPSAT-3 which could extend the applicability of Erosion-SAM for large-scale studies. Such efforts would enable monitoring with higher temporal and spatial resolution and provide important insights for long-term environmental management. Further comparative studies are also recommended to evaluate the performance of Erosion-SAM against traditional and machine learning-based SE models in both localization and classification of erosion and deposition areas. These studies would deepen the understanding of the strengths and limitations of the model and support its integration into broader environmental monitoring systems. In summary, Erosion-SAM represents a significant advance in the automatic detection and segmentation of SE. Its potential to revolutionize SE monitoring at fine spatial and temporal scales makes it a valuable tool for addressing global challenges related to land degradation, food security and climate change adaptation.

Funding

This work was supported by the Deutsche Forschungsgesellschaft (DFG) [No. SCHO 739/25-1].

CRediT authorship contribution statement

Hadi Shokati: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andreas Engelhardt:** Writing – review & editing, Validation, Software, Methodology. **Kay Seufferheld:** Writing – review & editing, Data curation. **Ruhollah Taghizadeh-Mehrjardi:** Conceptualization. **Peter Fiener:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hendrik P.A. Lensch:** Methodology, Conceptualization. **Thomas Scholten:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The aerial images are subject to license restrictions and cannot be shared. However, other data will be made available upon request.

References

Agnihotri, D., Kumar, T., Jhariya, D., 2021. Intelligent vulnerability prediction of soil erosion hazard in semi-arid and humid region. Environ. Dev. Sustain. 23, 2524–2551. https://doi.org/10.1007/s10668-020-00685-2.

Auerswald, K., Fiener, P., Dikau, R., 2009. Rates of sheet and rill erosion in Germany — a meta-analysis. Geomorphology 111, 182–193. https://doi.org/10.1016/j.geomorph.2009.04.018.

Borrelli, P., Alewell, C., Alvarez, P., Anache, J.A.A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., De Girolamo, A.M., Gessesse, G.D., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., Fiener, P., Freppaz, M., Gentile, F., Gericke, A., Haregeweyn, N., Hu, B., Jeanneau, A., Kaffas, K., Kiani-Harchegani, M., Villuendas, I.L., Li, C., Lombardo, L., López-Vicente, M., Lucas-Borja, M.E., Märker, M., Matthews, F., Miao, C., Mikoš, M., Modugno, S., Möller, M., Naipal, V., Nearing, M., Owusu, S., Panday, D., Patault, E., Patriche, C.V., Poggio, L., Portes, R., Quijano, L., Rahdari, M.R., Renima, M., Ricci, G.F., Rodrigo-Comino, J., Saia, S., Samani, A.N., Schillaci, C., Syrris, V., Kim, H.S., Spinola, D.N., Oliveira, P.T., Teng, H., Thapa, R., Vantas, K., Vieira, D., Yang, J.E., Yin, S., Zema, D.A., Zhao, G., Panagos, P., 2021. Soil erosion modelling: a global review and statistical analysis. Sci. Total Environ. 780, 146494. https://doi.org/10.1016/j.scitotenv.2021.146494.

Bui, D.T., Khosravi, K., Tiefenbacher, J., Nguyen, H., Kazakis, N., 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci. Total Environ. 721, 137612. https://doi.org/10.1016/j.scitotenv.2020.137612.

Cândido, B.M., Quinton, J.N., James, M.R., Silva, M.L.N., de Carvalho, T.S., de Lima, W., Beniaich, A., Eltner, A., 2020. High-resolution monitoring of diffuse (sheet or interrill) erosion using structure-from-motion. Geoderma 375, 114477. https://doi.org/10.1016/j.geoderma.2020.114477.

Character, L., Ortiz JR, A., Beach, T., Luzzadder-Beach, S., 2021. Archaeologic machine learning for shipwreck detection using lidar and sonar. Remote Sens. (Basel) 13 (9), 1759. https://doi.org/10.3390/rs13091759.

Elhage, N., Hume, T., Olsson, C., Schiefer, N., Henighan, T., Kravec, S., Hatfield-Dodds, Z., Lasenby, R., Drain, D., Chen, C., Grosse, R., McCandlish, S., Kaplan, J., Amodei, D., Wattenberg, M., Olah, C., 2022. Toy Models of Superposition. https://doi.org/10.48550/arXiv.2209.10652.

Eltner, A., Baumgart, P., Maas, H.-G., Faust, D., 2015. Multi-temporal UAV data for automatic measurement of rill and interrill erosion on loess soil. Earth Surf. Proc. Land. 40, 741–755. https://doi.org/10.1002/esp.3673.

Fischer, F.K., Kistler, M., Brandhuber, R., Maier, H., Treisch, M., Auerswald, K., 2018. Validation of official erosion modelling based on high-resolution radar rain data by aerial photo erosion classification. Earth Surf. Proc. Land. 43, 187–194. https://doi.org/10.1002/esp.4216.

Ganasri, B.P., Ramesh, H., 2016. Assessment of soil erosion by RUSLE model using remote sensing and GIS-A case study of Nethravathi Basin. Geosci. Front. 7, 953–961.

Gautam, P., Kubota, T., Sapkota, L.M., Shinohara, Y., 2021. Landslide susceptibility mapping with GIS in high mountain area of Nepal: a comparison of four methods. Environ. Earth Sci. 80, 359. https://doi.org/10.1007/s12665-021-09650-2.

Gocke, M.I., Don, A., Heidkamp, A., Schneider, F., Amelung, W., 2021. The phosphorus status of German cropland—an inventory of top- and subsoils. J. Plant Nutr. Soil Sci. 184, 51–64. https://doi.org/10.1002/jpln.202000127.

Hadlich, M., Marinov, Z., Stiefelhagen, R., 2023. AutoPET Challenge 2023: Sliding Window-based Optimization of U-Net. DOI: 10.48550/arXiv.2309.12114.

Hou, Y., Ren, H., Lv, Q., Wu, L., Yang, X., Quan, Y., 2022. Radar-jamming classification in the event of insufficient samples using transfer learning. Symmetry 14, 2318. https://doi.org/10.3390/sym14112318.

Ishikawa, Y., Kasai, S., Aoki, Y., Kataoka, H., 2021. Alleviating Over-Segmentation Errors by Detecting Action Boundaries. Presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2322–2331.

Iuss, I., Fao, I., 1999. World reference base for soil resources.

Jetten, V.G., Favis-Mortlock, D., 2006. Modelling soil erosion in Europe. Soil Erosion in Europe. Wiley 695–716.

Karydas, C.G., Panagos, P., Gitas, I.Z., 2014. A classification of water erosion models according to their geospatial characteristics. Int. J. Digital Earth 7, 229–250. https://doi.org/10.1080/17538947.2012.671380.

Khosravi, K., Rezaie, F., Cooper, J.R., Kalantari, Z., Abolfathi, S., Hatamiafkoueieh, J., 2023. Soil water erosion susceptibility assessment using deep learning algorithms. J. Hydrol. 618, 129229. https://doi.org/10.1016/j.jhydrol.2023.129229.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollar, P., Girshick, R., 2023. Segment Anything. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026.

Kohrell, G.J., Mulla, D.J., Gelder, B., 2023. Calibration and validation of hillslope runoff and soil loss outputs from the Water Erosion Prediction Project model in Minnesota agricultural watersheds. JAWRA J. Am. Water Resour. Assoc. 59, 1529–1548. https://doi.org/10.1111/1752-1688.13148.

Liu, C., Li, H., Xu, J., Gao, W., Shen, X., Miao, S., 2023. Applying convolutional neural network to predict soil erosion: a case study of coastal areas. Int. J. Environ. Res. Public Health 20, 2513. https://doi.org/10.3390/ijerph20032513.

Lv, Q., Quan, Y., Feng, W., Sha, M., Dong, S., Xing, M., 2022. Radar deception jamming recognition based on weighted ensemble CNN with transfer learning. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2021.3129645.

Mitasova, H., Mitas, L., 2001. Multiscale Soil Erosion Simulations For Land Use Management. In: Harmon, R.S., Doe, W.W. (Eds.), Landscape Erosion and Evolution Modeling. Springer, US, Boston, MA, pp. 321–347. https://doi.org/10.1007/978-1-4615-0575-4_11.

Mosavi, A., Sajedi-Hosseini, F., Choubin, B., Taromideh, F., Rahi, G., Dineva, A.A., 2020. Susceptibility mapping of soil water erosion using machine learning models. Water 12, 1995. https://doi.org/10.3390/w12071995.

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., Crist, S., Shpritz, L., Fitton, L., Saffouri, R., Blair, R., 1995. Environmental and economic costs of soil erosion and conservation benefits. Science 267, 1117–1123. https://doi.org/10.1126/science:267.5201.1117.

Polovina, S., Radić, B., Ristić, R., Milčanović, V., 2024. Application of remote sensing for identifying soil erosion processes on a regional scale: an innovative approach to enhance the erosion potential model. Remote Sens. (Basel) 16, 2390.

Prasad, D.V.V., Venkataramana, L.Y., Kumar, P.S., Prasannamedha, G., Harshana, S., Srividya, S.J., Harrinei, K., Indraganti, S., 2022. Analysis and prediction of water quality using deep learning and auto deep learning techniques. Sci. Total Environ. 821, 153311. https://doi.org/10.1016/j.scitotenv.2022.153311.

Pu, X., Jia, H., Zheng, L., Wang, F., Xu, F., 2024. ClassWise-SAM-Adapter: Parameter Efficient Fine-tuning Adapts Segment Anything to SAR Domain for Semantic Segmentation.

Quinton, J.N., Fiener, P., 2024. Soil erosion on arable land: an unresolved global environmental threat. Prog. Phys. Geography: Earth Environ. 48, 136–161. https://doi.org/10.1177/03091333231216595.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29, 2352–2449. https://doi.org/10.1162/neco_a_00990.

Renard, K.G., 1997. Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). U.S. Department of Agriculture, Agricultural Research Service.

Saadeldin, M., O'Hara, R., Zimmermann, J., Mac Namee, B., Green, S., 2022. Using deep learning to classify grassland management intensity in ground-level photographs for more automated production of satellite land use maps. Remote Sens. Appl.: Soc. Environ. 26, 100741. https://doi.org/10.1016/j.rsase.2022.100741.

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., Ryo, M., 2023. Ten deep learning techniques to address small data problems with remote sensing. Int. J. Appl. Earth Obs. Geoinf. 125, 103569. https://doi.org/10.1016/j.jag.2023.103569.

Saha, S., Roy, J., Pradhan, B., Hembram, T.K., 2021. Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. Adv. Space Res. 68, 2819–2840. https://doi.org/10.1016/j.asr.2021.05.018.

Schindewolf, M., Schmidt, J., 2012. Parameterization of the EROSION 2D/3D soil erosion model using a small-scale rainfall simulator and upstream runoff simulation. CATENA, Exp. Earth Surf. Process Res. 91, 47–55. https://doi.org/10.1016/j.catena.2011.01.007.

Scholten, T., Seitz, S., 2019. Soil erosion and land degradation. Soil Systems 3, 68. https://doi.org/10.3390/soilsystems3040068.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.

Song, B., Yang, H., Wu, Y., Zhang, P., Wang, B., Han, G., 2024. A Multispectral remote sensing crop segmentation method based on segment anything model using multistage adaptation fine-tuning. IEEE Trans. Geosci. Remote Sens. 62, 1–18. https://doi.org/10.1109/TGRS.2024.3411398.

Sun, J., Yan, S., Yao, X., Gao, B., Yang, J., 2024. A Segment Anything Model based weakly supervised learning method for crop mapping using Sentinel-2 time series images. Int. J. Appl. Earth Obs. Geoinf. 133, 104085. https://doi.org/10.1016/j.jag.2024.104085.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PLoS One 14, e0224365. https://doi.org/10.1371/journal.pone.0224365.

Vinayahalingam, S., Berends, B., Baan, F., Moin, D.A., van Luijn, R., Bergé, S., Xi, T., 2023. Deep learning for automated segmentation of the temporomandibular joint. J. Dent. 132, 104475. https://doi.org/10.1016/j.jdent.2023.104475.

Wagner, F., Eltner, A., Maas, H.-G., 2023. River water segmentation in surveillance camera images: a comparative study of offline and online augmentation using 32 CNNs. Int. J. Appl. Earth Obs. Geoinf. 119, 103305. https://doi.org/10.1016/j.jag.2023.103305.

Walling, D.E., He, Q., Whelan, P.A., 2003. Using 137Cs measurements to validate the application of the AGNPS and ANSWERS erosion and sediment yield models in two small Devon catchments. Soil Tillage Res. 69 (1–2), 27–43. https://doi.org/10.1016/S0167-1987(02)00126-5.

Wang, F., Sahana, M., Pahlevanzadeh, B., Chandra Pal, S., Kumar Shit, P., Piran, M.J., Janizadeh, S., Band, S.S., Mosavi, A., 2021. Applying different resampling strategies in machine learning models to predict head-cut gully erosion susceptibility. Alex. Eng. J. 60, 5813–5829. https://doi.org/10.1016/j.aej.2021.04.026.

Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., Jiao, L., 2018. A deep learning framework for remote sensing image registration. ISPRS J. Photogrammetry Remote Sens., Deep Learning RS Data 145, 148–164. https://doi.org/10.1016/j.isprsjprs.2017.12.012.

Wilken, F., Ketterer, M., Koszinski, S., Sommer, M., Fiener, P., 2020. Understanding the role of water and tillage erosion from $^{239+240}$Pu tracer measurements using inverse modelling. Soil 6, 549–564. https://doi.org/10.5194/soil-6-549-2020.

Wischmeier, W.H., Smith, D.D., 1978. Predicting Rainfall Erosion Losses: A Guide to Conservation Planning. Department of Agriculture, Science and Education Administration.

Ying, X., 2019. An overview of overfitting and its solutions. J. Phys. Conf. Ser. 1168, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022.

Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q., Shen, C., 2019. Hyperspectral classification based on lightweight 3-D-CNN with transfer learning. IEEE Trans. Geosci. Remote Sens. 57, 5813–5828. https://doi.org/10.1109/TGRS.2019.2902568.