



Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageoAutomatic detection of ditches and natural streams from digital elevation models using deep learning[☆]Mariana Dos Santos Toledo Busarello^{a,*} , Anneli M. Ågren^a , Florian Westphal^b ,
William Lidberg^a ^a Swedish University of Agricultural Sciences, Skogsmarksgränd, 901 83, Umeå, Sweden^b Jönköping University, Gjuterigatan 5, 551 11 Jönköping, Sweden

ARTICLE INFO

Keywords:

Streams

Ditches

Deep learning

LiDAR

Semantic segmentation

ABSTRACT

Policies focused on waterbody protection and restoration have been suggested to European Union member countries for some time, but to adopt these policies on a large scale the quality of small water channel maps needs considerable improvement. We developed methods to detect and classify small stream and ditch channels using airborne laser scanning and deep learning. The research questions covered the influence of the resolution of the digital elevation model on channel extraction, the efficacy of different terrain indices to identify channels, the potential advantages of combining indices, and the performance of a U-net model in mapping both ditches and stream channels. Models trained in finer resolutions were more accurate than models trained with coarser resolutions. No single terrain index consistently outperformed all others, but some combinations of indices had higher MCC values. Natural stream channels were not classified to the same extent as ditches. The model trained on the 0.5 m resolution had the most balanced performance using a combination of indices trained using the dataset with both types of channel separately. The deep learning model outperformed traditional mapping methods for ditches, increasing the recall from less than 10% to over 92%, while the recall for natural channels was around 71%. However, despite the successful detection of ditches, the models frequently misclassified streams as ditches. This poses a challenge, as natural channels are protected under land use management practices, while ditches are not.

1. Introduction

The primary objective of the United Nations Agenda 2030 for Sustainable Development is the protection of the planet from further environmental degradation (United Nations General Assembly, 2015), highlighting the importance of protecting and restoring water-related ecosystems. A similar goal is present in the European Water Framework Directive (where policy changes implemented in 2000 brought an integrated approach to the management and protection of aquatic environments) adopted throughout the European Union. Furthermore, a proposal for new targets of nature restoration is currently being drawn up by the European Commission, aiming at successful restoration of 20% of the target area by 2030, and 90% by 2050 (Council of the European Union, 2023). However, the management strategies for applying these initiatives differ among countries.

Most countries use different sizes of riparian buffer zones to protect surface waters during land-use operations, but these policies vary when it comes to small streams. In Finland, for example, stream channels are protected through a forest buffer of minimum width (Ring et al., 2018). In Sweden, the Swedish Forest Act (Skogsstyrelsen, 2013) also prescribes forest water protection through riparian buffers of variable width (Hasselquist et al., 2020). This is a necessary measure because over 75% of the total river network is estimated to be small streams (Bishop et al., 2008), and therefore even small changes in the network can impact downstream channels dramatically. Even so, the data shows that after 2004 as few as 25% of the small streams in Sweden were protected in such a manner, and when a buffer is present it usually has a width of 4 ± 0.4 m (Kuglerová et al., 2020), despite the recommended 5–30 m width of no-harvesting zones.

Some laws only address watercourses in general and do not

[☆] Link to the code: <https://github.com/mbusarello/Automatic-Detection-of-Ditches-and-Natural-Streams-from-Digital-Elevation-Models-Using-Deep-Learning>.

* Corresponding author.

E-mail addresses: mariana.busarello@slu.se, mariana.busarello@gmail.com (M.D.S.T. Busarello), anneli.agren@slu.se (A.M. Ågren), florian.westphal@ju.se (F. Westphal), william.lidberg@slu.se (W. Lidberg).

<https://doi.org/10.1016/j.cageo.2025.105875>

Received 18 March 2024; Received in revised form 21 January 2025; Accepted 23 January 2025

Available online 27 January 2025

0098-3004/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

differentiate between natural channels and those altered or made by humans, while other laws go into more depth on different types of watercourses. For example, according to the Swedish Forestry Act (September 1, 2022) ditches are divided into two categories: “ditches” and “protective ditches”. Simple “ditches” are dug for permanent soil drainage to change the land use of an area. “Protective ditches,” on the other hand, are temporarily dug to mitigate groundwater level rise following clear-cutting. Protective ditches must not be cleaned, as they are temporary, and should not be more than 50 cm deep. No permit is needed to clean ditches, while digging new ones does require official permission (Swedish PEFC, 2023). The idea is that ditches should gradually fill in with sediments and vegetation, eventually disappearing with time. The management of ditches can also include damming/plugging them to restore wetlands (Nieminen et al., 2018). Because of this variability in the practices which are allowed by law, knowing if a channel is natural or altered by man determines the best management choice.

Within the context of environmental impact, forest ditches can be strong anthropogenic emitters of greenhouse gases (Peacock et al., 2021b), with methane offsetting the uptake from terrestrial CH₄; they also transport suspended solids, which impacts water quality (Nieminen et al., 2018). Even though the differences between ditches and small natural streams are not always clear, factors such as morphology and hydrology do stress the distinction between channel types. Some of these attributes can also influence the quantity of methane being emitted (Peacock et al., 2021a), resulting in an annual flux slightly higher for ditches than for streams.

There is wide recognition of the importance of hydrological variability to the ecology of small streams (Huryn and Wallace, 1987; Lanka et al., 1987; Wohl, 2017), after all, the characteristics of meandering, pools, and rapids can define habitats (Beschta and Platts, 1986; Wiens, 2002; Martínez et al., 2013), nutrient cycling (Alexander et al., 2007; Claessens et al., 2010), and water quality (Cox et al., 2023). Yet, the mapping of small water channels (<6 m wide) on Sweden’s traditional digital maps was poor: 55% of the natural streams and 91% of ditches were not detected in the Swedish property map (Flyckt et al., 2022). Plus, the simplified digitized line from this dataset (Lantmäteriet, 2014) has limited usefulness for research in ecology when working across the landscape scale with geographic information system methods. Still, the number of mapped ditches was increased from 9% to 86% by Lidberg et al. (2023) using deep learning (LeCun et al., 2015) and remote sensing, turning the once laborious manual task with a substantial investment of cost and time into an automated process. Many countries have already been scanned with airborne laser scanning (ALS), and, using the latest return data, digital elevation models (DEMs) can be constructed, revealing small-scale channels (Raber et al., 2002).

Deep learning approaches have been used to map stream channels based on satellite images and Digital Elevation Models (Mazhar et al., 2022; Fei et al., 2022; Isikdogan et al., 2017). However, the main focus of these studies has been on larger rivers, while deep learning applications in small streams is limited. Koski et al. (2023) mapped small channels but did not separate between ditches and natural streams, while others have focused only on ditches based on ALS data (Du et al., 2024; Lidberg et al., 2023), or aerial photos (Robb et al., 2023). Despite these efforts, a research gap remains for small natural streams – the headwaters. Headwater streams are like the capillary system in the body – just as the health of the whole organism depends on a functioning capillary system, the health of larger streams and rivers depend upon an intact headwater stream network (Kuglerová et al., 2017), hence there is a large societal need for improving the mapping of the headwaters. Traditionally, headwaters are mapped from DEMs by calculating flow accumulation and applying a threshold to determine where streams begin (Ågren et al., 2015). However, the high natural variability in stream initiation thresholds makes these networks unreliable (Paul et al., 2023). Additionally, channel networks derived from flow accumulation are subject to further uncertainties because flow accumulation

requires extensive preprocessing to of the DEM which introduces more uncertainties especially at stream/road crossings (Lidberg et al., 2017). Therefore, the goal of this study was to develop a method for mapping channels in the landscape without including upstream areas or considering the presence of water. Instead, the focus was on detecting the physical structure of the channel, specifically the elongated depression visible in the DEM.

Building on the successful use of deep learning to map ditches in Lidberg et al. (2023), this article extends the methodology by incorporating the digitization of small natural stream channels into a dataset that was previously limited to ditches and adding one more study area. Topographic indices derived from ALS data and the manually mapped channels were used to train a U-net model to detect small-scale channels (both ditches and natural streams). Here, we explore for the first time if deep learning can be used to detect small streams from the high-resolution DEM considering not only the channels’ location, but also their variable width instead of just buffering them. The following research questions were answered:

- 1) How important is the resolution of the DEM for detecting ditches and natural channels? Here we explore two resolutions: 0.5 m and 1 m.
- 2) When highlighting the channels using digital terrain indices, is there a best one? Is the same index best for natural channels and ditches, or do they differ?
- 3) When detecting channels, is it better to work with just one terrain index, or to combine the information from many indices?
- 4) Can a U-net model be used to detect natural channels as well as ditches? Is it better to include ditches and natural channels in the same model, or to make separate models?

2. Methodology

Digital terrain indices were extracted from the DEM obtained from the high-resolution ALS data. These terrain indices were combined to form a database of manually mapped water channels, this then being used to train a deep neural network to detect and classify small-scale channels.

2.1. Study areas

We used remote sensing data and field data from the 12 regions described by Lidberg et al. (2023). The original dataset was exclusively composed of ditches; smaller (<6 m width) natural streams were added later by Paul et al. (2023). This data were revised and updated by comparing the location of the channels directly to orthophotos with a resolution ranging from 0.17 to 0.5 m (Lantmäteriet, 2021a) and the High-Pass Median Filter (HPMF) terrain analysis, increasing the length of channels to 2235 km of ditches and 315 km of natural streams.

Following Paul et al. (2023), these sites illustrated the diversity of the country’s landscape properties, with land use mainly represented by forests covering 86–99% of the area, and agriculture ranging from 0 to 13.2% coverage among sites. Variability in characteristics such as soil type, tree species, runoff, and topography were considered in the site selection process. Overall, the Swedish landscape has been heavily ditched, tripling the originally unaltered channel length density, with the majority of the channels built being forest ditches. Most of the natural channel heads can be found in the northern areas, but transition points (i.e., the connection between a natural channel and an upstream ditch network) happened more often in the south. Small natural channels in Sweden are meandering and blend with the surrounding terrain, as boulders in their course minimize stark contrasts (Fig. 2B). Ditches are instead straight and smooth-looking, with generally well-defined borders resulting from the removal of boulders during the digging process. Most of the ditches in the dataset were forest ditches (56%), with road ditches in second (25%), and agricultural ditches last (6%, Paul et al. (2023)).

2.2. Training data

2.2.1. Topographic indices

The ALS data (Lantmäteriet, 2021b) were collected by an aircraft flying at a height of 2888–3000 m with a compact laser-based system onboard (Leica ALS80-HP-8236) generating point clouds with a density of 1–2 points per square meter. LiDAR Tin Gridding from Whitebox Tools was used to create DEMs with 0.5 m and 1 m resolutions over the study areas, totaling 430 km². We selected seven topographic indices that could visually highlight small-scale channels present in the DEMs (Fig. 2) as a proxy for the differences in elevation. Many indices could have been experimented on, but there is a limitation in the number of variables that could be used in the study considering the amount of time and effort involved in calculating new indices and preparing them as input for training the models. It was also observed that larger moving windows provided excessive smoothing, blending small channels in the landscape, while small scales introduced a high amount of noise. This is why the choice in scale relied on the visual evaluation for the cases where the size of the moving window was not arbitrarily defined by the tool in use.

The topographic indices were normalized between zero and one before being divided into chips of 500 × 500 pixels for input to the deep learning algorithm (Fig. 1B and C). Whitebox Tools was used to calculate

all topographic indices, except for the Sky-view Factor, which was obtained using the Relief Visualization Toolbox v. 2.2.0 (Kokalj et al., 2016).

2.2.2. High-Pass Median Filter

The HPMF (Lindsay, 2016) emphasizes short-range variability, subtracting the pixel value from the median value of the other pixels inside a window. The window size kernel is user-defined; this study used 11 in both X and Y directions. The data were normalized by applying the Min–Max Normalization. Negative values indicate depressions and can be used to highlight channels, i.e. elongated depressions in the soil. This index is similar to the topographic position index, which is obtained through the subtraction of the mean value of the area covered by a moving window, however, HPMF was chosen due to the previously successful application in Lidberg et al. (2023), and because the median is more resistant to extreme values in the data.

2.2.3. Hillshade

The shaded relief (Wilson and Gallant, 2000) makes it possible to visualize a three-dimensional surface considering its slope and aspect, with shadows distributed according to the illumination source position (altitude and azimuth). This study has used the fixed altitude of 30° and the azimuths 0°, 45°, 90°, and 135°. The values were normalized

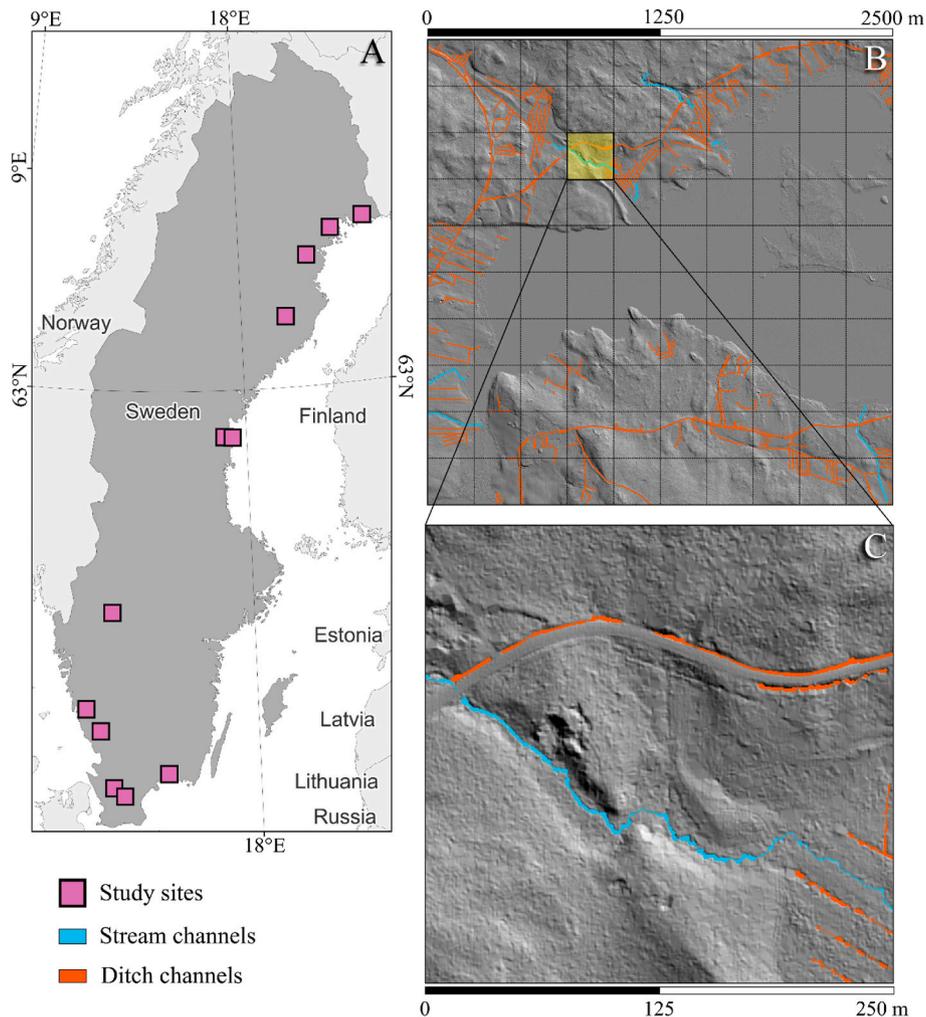


Fig. 1. Study areas. (A) 12 regions spread across Sweden where all ditches and streams were manually digitized; (B) Study regions split into 2.5 km × 2.5 km tiles. Locations of manually mapped water channels were separated by type, with ditches in orange and natural channels in turquoise, drawn over the hillshaded elevation model. Each grid cell represents chips with sides of 500 × 500 pixels. (C) An example of a 0.5 m resolution image chip obtained after splitting the tile. These chips are the images that the deep learning models will use as training data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

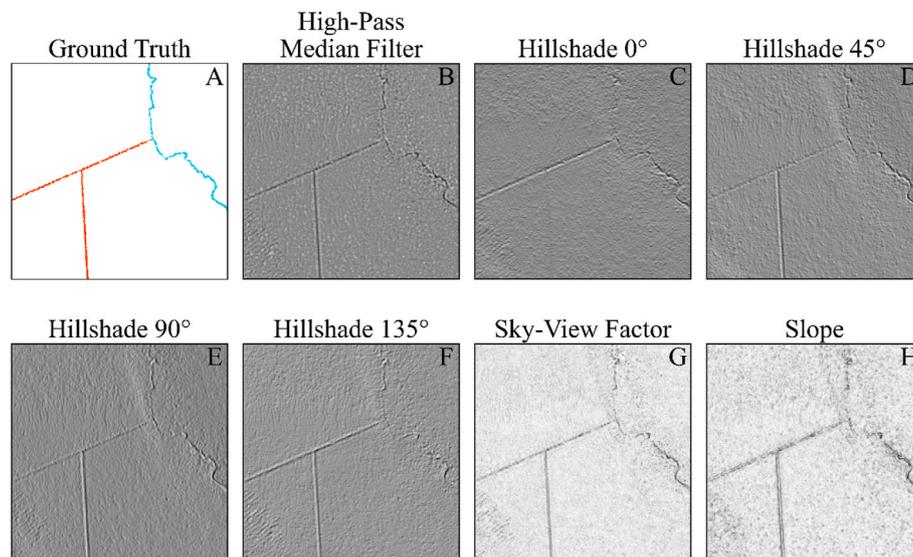


Fig. 2. Examples of the ground truth and topographic indices. Orange represents ditches and turquoise represents natural streams. Images displayed represent an area of 250 m × 250 m. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

afterward through their division by the maximum value. The bottom of a channel would be shaded unless it was hit by sunlight along the direction of the channel. To address this issue, we included hillshades from four different angles.

2.2.4. Sky-view factor

This index is defined by the ratio between the radiation received at a specific grid cell and the one emitted through the whole hemispheric environment around it (Zakšek et al., 2011). Considering a visual observation of the channels, the chosen radius was 5 m with 16 directions.

2.2.5. Slope

This topographic index represents the change in elevation between every pixel in the DEM with a moving window sized 5 × 5 for increased accuracy and stronger reduction of high-frequency noise (Florinsky, 2016), with the inclination displayed in degrees. To perform the normalization, all values were divided by the theoretical maximum value of 90°.

2.2.6. Labels

When the word “channel” is used in this article, it includes both ditches and natural streams. Using Whitebox Tools, we started by obtaining the flow accumulation (O’Callaghan and Mark, 1984). First, we filled the single cell depressions in the DEM (*FillDepressions*), then burning streams at roads using data from the Swedish Property map (Lantmäteriet, 2014) to ensure stream continuity across roads (*BurnStreamsAtRoads*). Remaining larger depressions were breached (*BreachDepressionsLeastCost*) to keep the flow continuity, using this as the input to calculate the D8 flow accumulation (*D8FlowAccumulation*). Streams were extracted (*ExtractStreams*) using the lowest stream initiation threshold from the distribution observed for natural channel heads in Paul et al. (2023): 2 ha.

Following this methodology, the channel heads and connections to the ditch network were identified, and downstream stream paths manually marked and edited. Ditches were visually identified from HPMF and orthophotos, being manually mapped as vector lines by a team of experts. We have utilized the HPMF values within the channels to give these lines a variable width, creating structures that more closely resemble the actual shape of the channels. Based on the method described in Lidberg et al. (2023), the HPMF analysis had its pixels reclassified based on the threshold of -0.075 (determined through

visual inspections), receiving the label 0 when they are above it, and 1 when below. A 3 m buffer surrounding the vector lines was generated, later overlapping the relabeled data and extracting the non-null pixels within it. Finally, we applied the majority filter to these selected pixels to remove strays, preserving the continuity of the channels (Fig. 2A).

Eight different datasets were created (Fig. 3), initially separated by how the channels were represented:

- Channels: all channels, merged to a combined dataset with no separation of ditches and streams. Two class labels; channel and background (Fig. 3A and E)
- Ditches: a separate dataset of only ditches. Two class labels; ditch and background (Fig. 3B and F).
- Streams: a separate dataset of only streams. Two class labels; streams and background (Fig. 3C and G)
- Ditches&Streams: a combined dataset with three class labels; ditches, streams, and background (Fig. 3D and H)

Each type of representation was calculated for both 0.5 m and 1 m resolution to analyze how this impacted the results; each one is noted as an added “0.5” or “1” the dataset names.

The datasets exhibited significant class imbalance. To compensate for that, only the chips containing more than 250 pixels with the positive label were selected for the analysis, resulting in 4615 chips in total. From these, 1.1% of the total pixels were ditches and 0.1% were streams. Not all chips contained both types of channels, so datasets with only streams or ditches had fewer chips (Busarello et al., 2024).

2.3. Semantic segmentation

The convolutional neural network (CNN) U-net (Ronneberger et al., 2015) (Fig. 4) was chosen for having successful real-world applications in different scientific fields such as medicine (Siddique et al., 2021), geology (Gao et al., 2022), and forestry (Korznikov et al., 2021), being both robust and versatile. It also has the advantage of concatenating the feature maps of the downsampling path to the upsampling path, preventing the loss of information during downsampling. A limitation of the study was the amount of chips available in the datasets: CNN models usually require thousands of training data examples, and for this reason, acquiring training data is the most challenging part of the process. The use of data augmentation (Tanner and Wong, 1987) increased the number and diversity of training images by adding slightly different

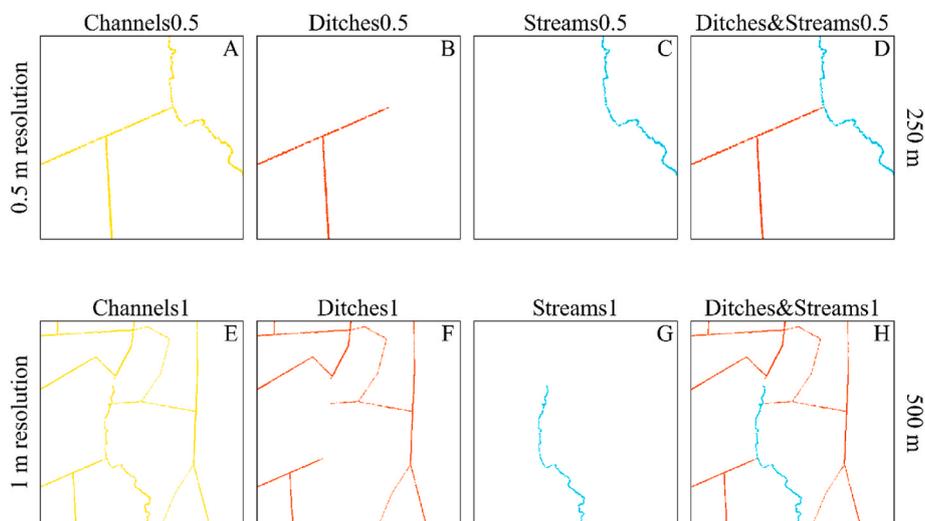


Fig. 3. Training data chip examples of both resolutions. Top row represents 0.5 m, and bottom row shows 1 m resolution. Chip size is 250 m × 250 m for 0.5 m resolution and 500 m × 500 m for 1 m resolution. Yellow lines in dataset Channels represent channels, without distinction between stream channels and ditch channels. Ditch channels are represented in orange in the datasets Ditches and Ditches&Streams. Turquoise represents stream channels in datasets Streams and Ditches&Streams. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

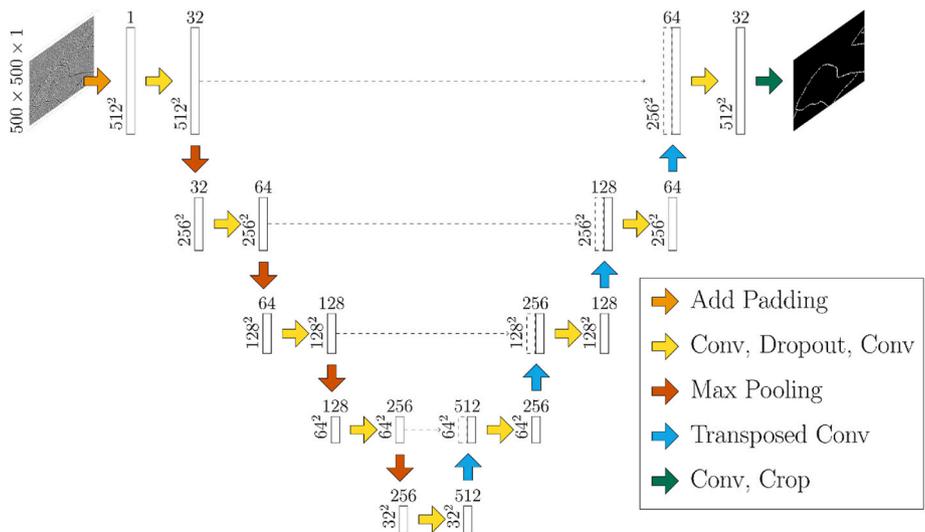


Fig. 4. U-net architecture. The left side shows the encoding/down-sampling process, where the main features are extracted while the input is compacted. On the right side is the decoding/up-sampling path, which upscales the features until it reaches the same size as the input.

copies of them to the dataset, obtained through transformations. The geometric transformations used in this work were the random rotation and random flips (horizontal and vertical). Random rotation rotated images in a random angle within the specified range of 0°–360°, helping improve the model’s generalization by increasing the pattern recognition regardless of the object orientation in the image. The horizontal random flipping rotated the image along its vertical axis, swapping left and right, while the vertical random flipping flipped across the horizontal axis, swapping the top and bottom of the image instead.

Considering that the proportion between the classes of pixels showed considerable imbalance, median frequency balancing (Eigen and Fergus, 2015) was used to establish the class weights used for training. Adam (Kingma and Ba, 2015) was used as the optimization algorithm, and the chosen batch size was 16. In the beginning, the topographical indices were used individually as input to train the first models, while the last model combines all indices, resulting in 64 different models. Later, all possible combinations were used as training data for the dataset Ditches&Streams to determine if combining indices is a better option

than using them individually.

The general architecture of U-net incorporates two paths: encoding and decoding. During the encoding phase, hierarchical features are extracted by a combination of convolutions and the pooling of feature maps, down-sampling the data resulting in a compact representation of the input, with an increased number of channels. Subsequently, in the decoding phase, transposed convolutions are applied to upscale the spatial dimension until the output matches the input original size. After each transposed convolution, a skip connection happens between corresponding layers in both paths. This allows the network to keep fine-grained details in the up-sampling process. The final convolution reduces the number of channels, producing the final segmentation map. In it, each pixel is assigned a probability of belonging to a class.

The processing time for calculating the topographic indices was tracked, as well as the inference time, being further extrapolated for the whole area of Sweden to estimate how long it would take to detect the location of channels throughout the entire country. Training and inference were done using an NVIDIA RTX A6000 GPU and AMD Ryzen

Threadripper 3990X Processor.

2.4. Evaluation

The data were split into two parts, 80% for training and 20% to evaluate the performance of all models, comparing the ground truth pixels with the detected pixels. Precision, Recall, F-score, and Matthews correlation coefficient (MCC; Matthews, 1975) were the key metrics used to evaluate the models, along with information retrieval tables. Precision is the metric that accounts for the accuracy of positive predictions from a model, being affected by the number of false positives. It assesses how much of the detection and classification made by the model was right. Recall, on the other hand, accounts for how much of the ground truth was correctly detected. F-score is the harmonic average of precision and recall, and MCC is a special case of the phi coefficient. The F-score was calculated to easily compare the performance of this study with other publications, but MCC reports the overall quality of the classification performed by the model, being more reliable for imbalanced datasets (Chicco and Jurman, 2020). The Precision-Recall curves were plotted to display the tradeoff between recall and precision in the highest-ranking models. In addition to these metrics, we also used models with the highest MCC values from each dataset to illustrate the location of detected channels. For the final evaluation, the inference of the best-performing models was compared to the ground truth in order to account for how much of each type of channel was detected by them.

2.5. Benchmark

We have used the traditional flow accumulation method of the 0.5m resolution as a benchmark to compare with our deep learning approach and our manually labeled dataset. The process to obtain the flow accumulation has been described in section 2.2.2, but now we have included the other two stream initiation thresholds of 6 ha and 10 ha, also observed in Paul et al. (2023). To make the comparison fair, the extracted streams went through the same described process to create the labels with natural contours: buffering, multiplying the buffer with the reclassified HPMF data, majority filtering, and combination with rasterlines. Additionally, the Swedish property map (1:12 500) was also used for comparison. It was rasterized (VectorLinesToRaster) and underwent the same process described in section 2.2 to create natural contours. All of this data was compared pixel by pixel to the labeled dataset, counting how many pixels labeled as ditch or natural stream were identified as channel by the flow accumulation.

Furthermore, the inference results from the deep learning model from Lidberg et al. (2023) was also compared to our ground truth data. Despite their model being trained exclusively on ditches, it indirectly detected some natural channels, allowing for a relevant comparison. To ensure we did not evaluate on data that the previous model might have been trained on, we used data from the newly added study site for this process, as it was not included in the previous model’s training data.

3. Results and discussion

3.1. Importance of DEM scale for the modeling of channels using deep learning

The precision and recall values were higher for datasets with a 0.5 m resolution than for the 1 m counterpart. This was the case for all datasets and topographic indices (Fig. 5). Despite this, some models displayed higher values at either metric individually, and some overlap between the resolutions has been observed. This is partially in line with previous research on mapping terrain features with deep learning and DEM data, where higher resolution had better results (Chowdhuri et al., 2021) but also showed that the difference in performance between resolutions was not very strong (Robson et al., 2020).

The recall had different values for all datasets at 1 m resolution, with

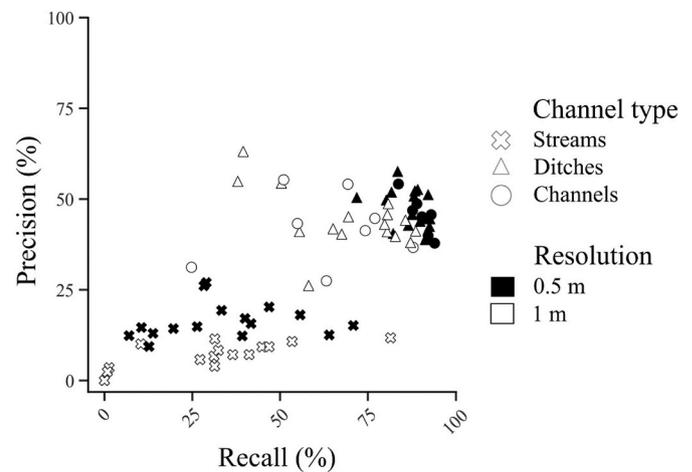


Fig. 5. Precision by Recall plot of the trained models, grouped by resolution and channel type. Black represents the 0.5 m resolution, while the 1 m resolution is represented by the white color. When referring to dataset Ditches&Streams0.5, the channel types were analyzed separately. The “Ditches” identified in the legend refers to this class in dataset Ditches and Ditches&Streams, while “Streams” addresses this class in dataset Streams and Ditches&Streams. “Channels” describes the models trained with the dataset Channels, combining ditch and stream channels in the same class.

small differences between ditches and channels. The precision was similar for either resolution, with a variation of around 10%. We can assume that the performance of the models with 1 m resolution was impacted by the topographic index used in the training process. This impact was also observed in the 0.5 m resolution but to a lesser extent, which could indicate that models trained on a higher resolution were stable. The stability was not present on channels labeled as streams: in both scales and with any dataset, as seen in the black crosses in Fig. 5, the recall values were different while the precision was similar, not going over 25%.

The estimated processing time required to both extract the topographical indices and apply the model differed substantially between the DEM resolutions (Table 1). The Sky-view Factor in particular was computationally demanding compared to the other topographical indices, regardless of resolution. This happens because the source-code for the RVT library was written in python, which is an interpreted language. The tools from WBT, on the other hand, were coded in Rust – a compiled language. Compiled programs are faster than those that have to be interpreted (Kwame et al., 2017), and one way to have similar processing times would be to have all the processing steps written in a compiled language. Furthermore, parallelizing the codes for execution on the GPU could potentially mean a considerable speed improvement. The inference time of the deep learning model was about the same for

Table 1

Time spent to calculate each topographic index individually and in combinations, and the time spent to apply a deep learning model on new data (inference): both in two resolutions and measured in seconds by square kilometers. It was also estimated how long it would take (in days) to calculate the topographic index(es) and apply the model to the processed data for the whole surface area of Sweden (447 425 km²). Hillshade had the same processing time regardless of the angle.

Topographic Index	Processing time (s/km ²)		Inference time (s/km ²)		Estimated time for Sweden (days)	
	0.5 m	1 m	0.5 m	1 m	0.5 m	1 m
HPMF	0.30	0.09	6.71	1.68	36	9
Hillshade	0.25	0.07	6.69	1.67	36	9
Slope	0.27	0.08	6.69	1.67	36	9
Sky-view Factor	3.01	0.68	6.64	1.67	50	12
Combination (all)	4.58	1.13	6.81	1.70	59	15

models trained with one index or several combined.

As models trained on the 0.5 m resolution datasets had the highest recall, the rest of this work focused on the models trained on topographical indices with a 0.5 m resolution. The analyses for 1 m resolution are in Appendices A.1, A.2, A.3, and A.4.

3.2. Impact of different terrain indices for detecting ditch and stream channels

We did not find a particular topographical index that consistently outperformed the others in this study. Models trained on Hillshades had the highest recall, while models trained on HPMF and Hillshade 0° had the highest precision using the datasets Channels0.5 (Fig. 6A) and Ditches0.5 (Fig. 6B). The model trained on the dataset Streams0.5 had the highest recall when trained on a combination of all topographical indices (Fig. 6C). That model had a recall of 70%, but the precision was still low at 20%. The highest recall for ditches with dataset Ditches&Streams0.5 was from the combination of all indices with 92% and 7% for streams using Hillshade 90° (Fig. 6D). The precision for the model trained on this dataset was highest with the HPMF for ditches, and Slope for streams. We believe that MCC gives the most balanced measure of the overall model performance, but there was no clear winner among models trained on different digital terrain indices (Table 2).

Indices not used in our work were listed as the most effective ones in studies focused on channels and fluvial features using the DEM (Du et al., 2019; Koski et al., 2023), or topographic positive openness for ditches (Du et al., 2024). Koski et al. (2023) detected channels using deep learning and several terrain indices besides the DEM, finding recall and precision values ranging 16–77% and 43–86%, respectively, while the F-score varied 0.23–0.81. The best terrain indices in our study for this type of dataset scored higher recall (83–93%), but lower precision (range 42–52%, Fig. 6A) and lower F-score (0.54–0.63, Table 3). The reasons for the differences are analyzed in section 3.4. Similarly, Du et al. (2024) detected ditches with deep learning, combining topographic and other features. Recall and precision were in the range of 73–76% and 63–69%, respectively, and F-score 0.69–0.71. Meanwhile, our similar dataset had higher recall (72–92%), lower precision 42–52% (Fig. 6B), and lower F-score 0.57–0.66 (Table 3). This difference could

be because of the U.S. study having a higher resolution (0.3 m against our 0.5 m). Lidberg et al. (2023), however, obtained a higher MCC value than this study using the HPMF (0.78), which could be due to the different deep learning architecture.

The variation in the performance of the hillshade indices could be explained by the variation in channel orientation. In Fig. 2C, for example, part of the stream and the vertical ditch do not show because they were parallel to 0°, while the channels oriented perpendicularly were highlighted. Therefore, no matter the amount of data acquired and data augmentation performed, when using an index there is a chance that the channels might not be visible at all. This further motivated our choice to combining them.

3.3. Combining topographic indices

Combining all of the topographic indices did not result in a higher MCC compared to using them individually as input training data for most datasets, except Streams0.5 (Table 2). This dataset (Fig. 6C) and Ditches&Streams0.5 (Fig. 6D) had higher recall values.

However, when all of the possible combinations between the indices with dataset Ditches&Streams0.5 were analyzed (Appendix B) we observed that, for ditches, the HPMF was surpassed by the combination of Sky-view Factor + Slope in the ditches class (MCC = 0.69 (Table 3) against 0.74 (Fig. 7)) and the streams class (MCC = 0.09 against 0.31). Furthermore, for streams the Slope was surpassed by the combination of Hillshade 45° + Hillshade 90° + Hillshade 135°, not in the ditch class (MCC = 0.63 against 0.63) but in the stream class (MCC = 0.28 against 0.36). These results are in line with Kazimi et al. (2020) and Du et al. (2019), where a combination outperformed the single index, even though both studies used a coarser resolution (50 m) to detect fluvial structures (among others). We believe that the resolution did not influence this difference between combining indices or not, since these results also matched the coarser one analyzed by us (Appendix A.4).

Additionally to the observed trend that no single index was better than a combination of indices, we noted that the best performing combinations are those that combined two or three indices (Fig. 7). This appears reasonable since each index extracted different information from the DEM and as such may not contain all necessary information.

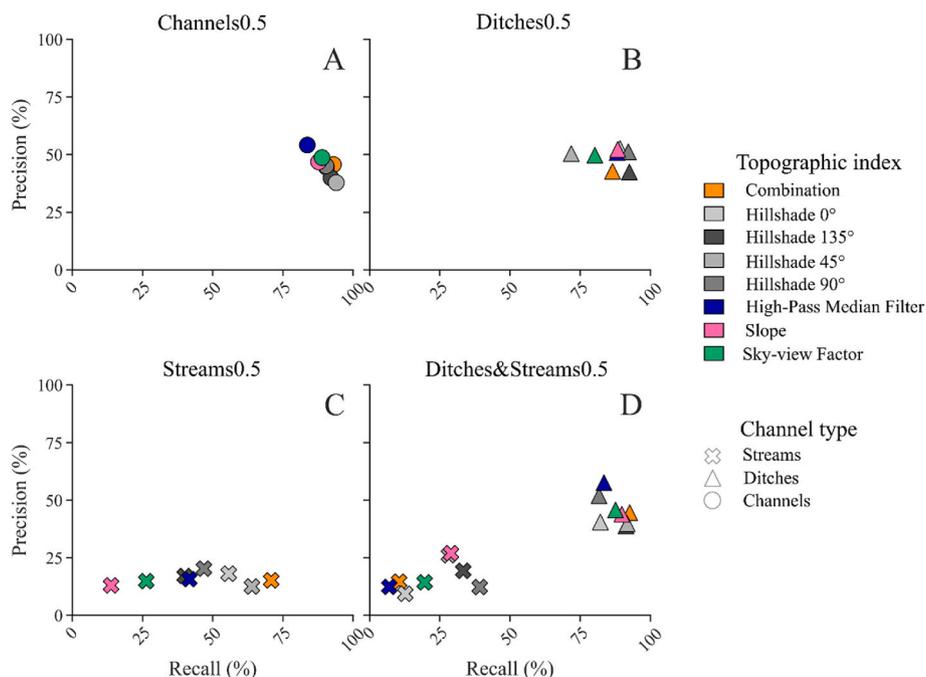


Fig. 6. Precision by Recall plots separated by dataset with 0.5 m resolution. Each color represents a topographic index, and each symbol represents a channel type. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2

MCC values for all datasets with the 0.5 m resolution. The terrain indices with the highest MCC are highlighted in bold.

Topographic Indices	Channels0.5	Ditches0.5	Streams0.5	Ditches&Streams0.5 (ditches)	Ditches&Streams0.5 (streams)
Combination	0.65	0.61	0.32	0.64	0.12
Hillshade 0°	0.63	0.68	0.31	0.57	0.11
Hillshade 45°	0.59	0.60	0.28	0.60	0.27
Hillshade 90°	0.64	0.69	0.30	0.65	0.22
Hillshade 135°	0.60	0.63	0.26	0.59	0.25
HPMF	0.67	0.67	0.25	0.69	0.09
Slope	0.64	0.68	0.13	0.63	0.28
Sky-view Factor	0.66	0.63	0.19	0.63	0.17

Table 3

Evaluation metrics for each model, dataset, and its highest-performing topographic index. The recall, precision, F-score, and MCC values are also presented.

Model	TP	FP	TN	FN	Recall	Precision	F-score	MCC
Channels0.5 High-Pass Median Filter	2395624	2028771	224358343	467262	83.7%	54.1%	0.66	0.67
Ditches0.5 Hillshade 90°	2396057	2282997	203616321	204625	92.1%	51.2%	0.66	0.69
Streams0.5 Combination	185753	1040722	54697074	76451	70.8%	15.1%	0.25	0.32
Ditches&Streams0.5 High-Pass Median Filter (ditches)	2170656	1597868	224812682	430026	83.4%	57.6%	0.68	0.69
Ditches&Streams0.5 High-Pass Median Filter (streams)	18318	130158	226965020	243886	6.9%	12.3%	0.12	0.09

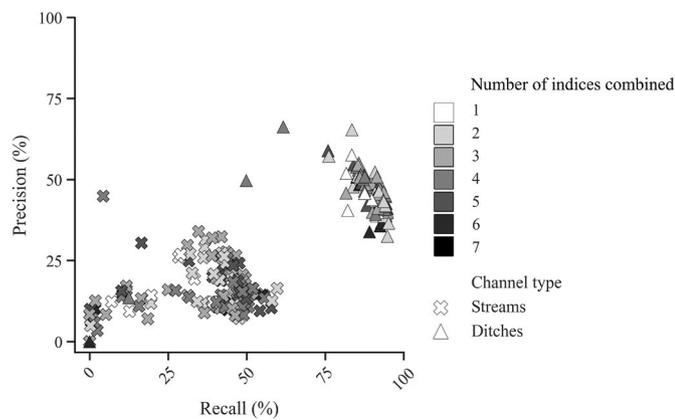


Fig. 7. Precision and Recall plots for all of the possible combinations of topographic indices. The color indicates the number of combined indices, and the shape represents the type of channel.

For example, ditches running from north to south were difficult to see in Hillshade 0° (Fig. 2C). However, adding indices to the considered combination, which introduce only slight variations of the information already provided by the considered indices, harmed performance, since it made the learning problem more difficult. This issue has been observed by others, for example by Yang et al. (2023) and Koski et al. (2023), who trained models directly on the DEM. These models performed similar or better than models trained on the DEM combined with indices derived from it, since all required information was already included in the DEM. Still, we argue that it is reasonable to assume that a model trained on topographic indices can generalize better due to the more uniform representation of the relevant topographic features.

Processing time could affect the decision to use multiple topographical indices, considering that it can increase greatly with a higher resolution. It seems that combining multiple topographical indices derived from the same LiDAR data could be beneficial, and so, including aerial photographs in the topographical data is something that might be worth exploring. Robb et al. (2023) obtained a higher F-score than our study (0.79 against 0.66) using orthophotos with a 0.25 m resolution to detect ditches, but this was not observed by Koski et al. (2023), where combining the orthophotos had the worst performance detecting channels. The aerial imagery data used by the Finnish study had a coarser resolution (0.5 m; NLS (2023)) which could be creating this difference. Koski et al. (2023) also points out that the extent of tree coverage

hindered the performance of this input data to some extent, something that seems not to have happened in the UK publication, judging by the fact that the study area was less forested.

3.4. Evaluating model performance with different datasets

The models with the highest MCC values were selected for further evaluation under section 3.4. By “datasets” we mean if the model was trained to identify channels, streams, and/or ditches. The models trained with dataset Ditches0.5 with the highest MCC had a recall of 92.1%, while models trained on the dataset Channels0.5 had a recall of 83.7% (Table 3). The same observation was made in the Precision-Recall curves, with AP = 0.76 for Channels0.5 (Fig. 8A) versus AP = 0.82 for Ditches0.5 (Fig. 8B). A Finnish dataset similar to Channels0.5 was used by Koski et al. (2023), with lower recall values (77.3%) but notably greater precision (85.6% against our 54%, Table 3). Starting in the 1950s, the ditching process in peatlands that was conducted in Finland altered the shape of most small natural channels (Muotka et al., 2002), with a low number of unaltered small streams left. This could mean that the uncertainty brought in by natural channels was smaller, as unaltered streams might be rarer in Finland, resulting in a higher precision. This could be an indication that when streams and ditches had the same label, uncertainty was introduced in the training process, blurring the detection and classification of channels. With the streams labeled as background, the separation became clearer and more channels were detected (despite the number of false positives also increasing).

The precision-recall curves strengthen the observations from Table 3. The average precision values reported were higher than the ones seen in the table because this metric is an approximation of the area under the precision-recall curve (Aslam et al., 2005), i.e., a summary of the precision-recall performance across all thresholds. However, we could still see similarities in the overall poor performance of the stream class in the Streams0.5 dataset (AP = 0.22, Fig. 8C), Ditches&Streams0.5 trained with HPMF (AP = 0.06, Fig. 8D), and the improvement brought to it by combining Sky-View Factor and Slope (AP = 0.28, Fig. 8E). Overall, the ditch label performed better across all datasets, showing that whichever high-ranking model was chosen, their detection would be similar. The differences, though, could be seen in the inferences (Fig. 9), where the interruption in channels happened more often within Channels0.5 (Fig. 9B) than Ditches0.5 (Fig. 9C).

For the model where the channels were trained with three labels (ditches, streams, and background (Fig. 3D)) we evaluated the ditches and streams separately. Ditch channels were correctly classified

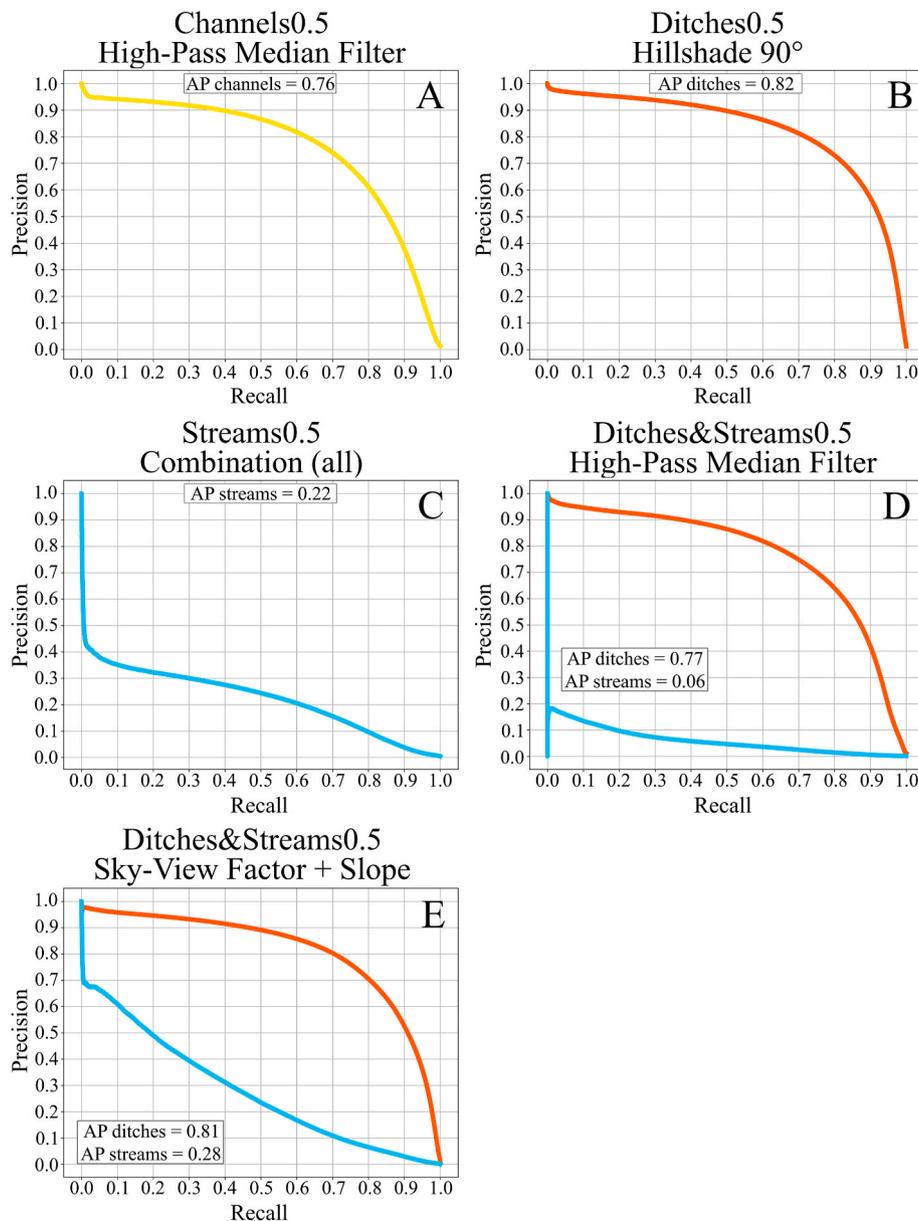


Fig. 8. Precision-Recall curves of the highest-ranking trained models and their average precision (AP).

frequently, which could mean that these channels had morphological attributes that made them more easily recognized by the neural network, while streams did not. Comparing this dataset (3-class) to dataset Ditches0.5 (binary), the recall was lower (83.4% against 92.1%), a result similar to [Phinzi et al. \(2020\)](#) when comparing the performance of a binary and a multiclass dataset to detect gullies with machine learning.

Models trained with the binary datasets had false positives more often, meaning that labeling streams and ditches separately in the training process could have helped distinguish both from the background data. A visual analysis of the detection ([Fig. 9E](#)) demonstrates that the models were not able to separate ditches and streams, but the number of false positives for the stream channels and ditches was low (0.06% and 0.7%, respectively; [Table 3](#)). For the dataset Ditches0.5 ([Fig. 9C](#)), stream channels were mainly misclassified as ditches despite being detected, while in the dataset Streams0.5 ([Fig. 9D](#)) the opposite happened, with frequent channel interruptions. This discontinuity was also observed in dataset Channels0.5 ([Fig. 9B](#)).

The channel interruptions were observed in small sections where the

width was narrower than the average 3 m. In the ground truth data, these gaps were absent because the original polyline shapefile was converted to raster format and merged with HPMF-extracted values. This provided channel continuity, but limited their width to a single pixel. Gaps in the channel network are not unusual due to not only natural processes like sedimentation, falling trees and logs, but also to anthropogenic modifications such as culverts, bridges, road embankments ([Lindsay and Dhun, 2015](#)), which would explain why parts of the channel would be absent in ground truth. With that, they would not be detected in the inference either.

The highest-ranking models ([Table 3](#)) detected channels but were not as effective when classifying them, so we have calculated how much of each channel type was detected by each model regardless of the model's classification ([Table 4](#)). For the binary datasets, "detection" was the same as recall ($TP/TP + FN$), while "classification" was the same as precision ($TP/TP + FP$). However, we also used the multilabel ground truth (with pixels labeled 0, 1, or 2) to evaluate the performance of the models on channels, calculating how much of each channel type was detected. For the multilabel dataset (Ditches&Streams0.5), "detection"

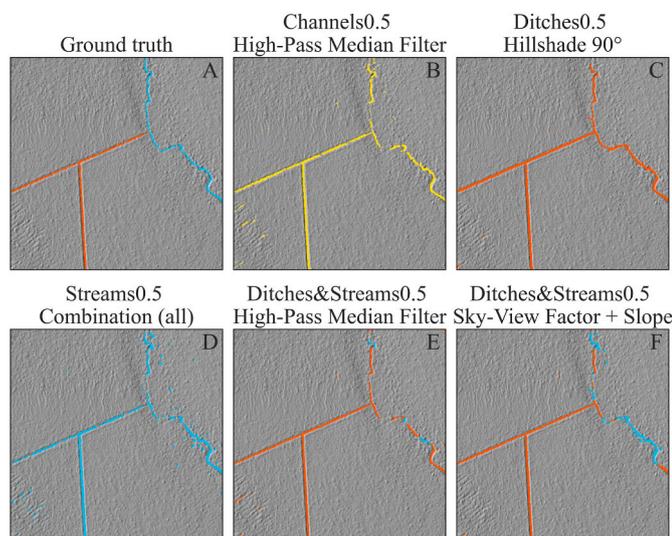


Fig. 9. Detected channels by the highest performing model from every dataset using the 0.5 m resolution, plotted over the hillshade. The colors represent channel type: ditch channels are orange, stream channels are turquoise, and combined channels are yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4

Amount of channel pixels detected by each model, separated by channel type. The last two columns are only relevant to the multilabel dataset and describe the quantity of detected channels that were correctly classified by the model as their ground truth channel type.

Dataset used to train the model	Detected ditches	Detected streams	Classified as ditches	Classified as streams
Channels0.5	85.9%	61.4%	–	–
Ditches0.5	92.1%	55.9%	–	–
Streams0.5	81.5%	70.8%	–	–
Ditches&Streams0.5	83.8%	54.9%	99.5%	12.7%

meant not being predicted as background (label 0). At the same time, “classification” verified how many of the channel type predictions were correct, i.e., ditch pixels predicted to be ditches and stream pixels predicted to be streams. This was done because, despite a pixel being classified incorrectly as either ditch or stream, as long as it was not classified as “background” (label 0), it was still counted as a channel per the definition we use in this work: the combination of ditches and streams.

When both channel types had the same label (Channels0.5), the detection was higher than when they were labeled separately in the same dataset (Ditches&Streams0.5). Models with only one channel type (Ditches0.5 and Streams0.5) detected the other class, and in the case of Streams0.5 more ditches were detected than stream channels. Streams can be characterized by relative depth, continuity, and high sinuosity. Ditches are also characterized by relative depth and continuity, and low sinuosity (more straight). However, not all streams are meandering and not all ditches are straight. These similarities make it challenging to distinguish between both channel types, while the straight aspect tends to simplify the recognition of ditches. Furthermore, we employed median frequency balancing (Eigen and Fergus, 2015), which assigns larger class weights to less frequent classes, leading to larger errors when pixels of these classes are mislabeled. With this in mind, we observed how different labeling strategies affect the tradeoff different models made between precision and recall (Fig. 8).

In the three binary classification datasets (Channels0.5, Ditches0.5, and Streams0.5), misclassifying background as the respective positive class is comparatively inexpensive, due to the small class weight for the

background class. Thus, the models favored higher recall despite an increase in false positives. In the 3-class dataset (Ditches&Streams0.5), labeling uncertain pixels as a minority class was costly due to the large class weights assigned to the ditch and the stream class. Mislabeled stream pixels as ditches incurred a significant penalty, while correctly identifying a small number of additional ditch pixels had limited benefits given their rarity. Conversely, background pixels offered the lowest relative cost, as they outnumbered the other two classes significantly. This led to higher precision but lower recall for ditches and streams. Additionally, the different recall values for ditches and streams in the three binary classification datasets were presumably due to the difficulty of identifying streams compared to ditches. When only streams were labeled (Stream0.5), the model needed to account for the meandering, sometimes nearly interrupted pattern of streams (Fig. 9A), which appeared to push the model toward recognizing other features in the landscape which have a similar pattern (Fig. 9D). This did not happen when only ditches were labeled (Fig. 9C), presumably because the model exploits the linear aspect of ditches, which allowed it to ignore other landscape features. When ditches and streams were labeled as channels (Channels0.5), the model needed to find a tradeoff between only focusing on the linear aspect, to allow it to find more streams than the ditch model, and recognizing too many landscape features, to achieve a better precision than the stream model. It appears to find this tradeoff by detecting more meandering interrupted features of the landscape as channels, while labeling more uncertain pixels as background, leading to more interrupted ditches (Fig. 9B).

Furthermore, because Ditches&Streams0.5 was a multilabel dataset we could verify how much of one label is classified as the other. In this case, from the number of ditches detected (83.8%), 99.5% were ground truth ditches. Meanwhile, only 12.7% of the streams detected by the model (54.9%) were correctly classified as streams. The difference in performance between stream and ditch channels in this dataset could be partially explained by the imbalance in the datasets. While the number of pixels with ditch labels was around 1.11% of the data, the stream pixels were underrepresented, with 0.01%. Contrasting class prior probabilities is a common occurrence in real-world data, and some techniques could be used to overcome it (Kotsiantis et al., 2006). In this work, the use of median frequency balancing (Eigen and Fergus, 2015) was motivated by its successful application in other studies such as Xu et al. (2022) and Kampffmeyer et al. (2016). However, despite the positive impact it had on the ditch class, an increase in performance of the stream class was not observed to the same extent. This represents a model limitation because the incorrect classification of streams as ditches is a regular occurrence. Adding more training data containing small natural streams would be an option to try to reduce the data imbalance, while an alternative would have been to perform a chip-based sampling, choosing chips that have more stream than ditch pixels in it. This would require further manual labor, though, where choices to reduce the costs of data acquisition could be explored, such as the use of semi-automated methods for labeling (Desmond et al., 2021) and crowdsourcing, despite the limitations that may arise regarding those who are not domain-specific experts (Clough et al., 2013).

3.5. Comparison to the benchmark

Our model (Streams0.5) had a recall of 70.8% of stream pixels, while the flow accumulation had the highest recall rate of stream pixels for 2 ha and 6 ha of initiation threshold (Table 5). 76.0% of the natural stream network was detected by the flow accumulation of 2 ha of the catchment area, 71.3% by 6 ha, and 70.2% by 10 ha. The Swedish Property map had a recall of 27.5% of pixels from the same channel type, which could be explained by the fact that the stream headwaters have been digitized from grainy black-and-white orthophotos in this data, being often obscured by canopy cover, which impacted its performance. Meanwhile, Lidberg et al. (2023) had an indirect recall (i.e., how much of the label “stream” was detected despite the model being trained with only

Table 5

Comparison between the recall performance of different methods of channel detection separated by type of channel pixels. “Recall of ditch pixels” refers to how many ditch pixels could be detected when compared to the ground truth. “Recall of stream pixels” refers to how many stream pixels were detected. All methods were evaluated on the same study areas, except Lidberg et al. (2023), which was evaluated on the study area that was not included in its training data. The MCC values listed were calculated with only the streams as the positive class to make a fair comparison between the methods.

Method	Recall of ditch pixels	Recall of stream pixels	MCC of ditches	MCC of natural streams
Swedish property map	8.1%	27.5%	0.16	0.28
Flow accumulation (2 ha)	33.8%	76.0%	0.32	0.21
Flow accumulation (6 ha)	21.5%	71.3%	0.26	0.26
Flow accumulation (10 ha)	17.3%	70.2%	0.24	0.29
Deep learning (Lidberg et al., 2023)	82.1%	25.7%	0.63	0.09
Deep learning (Ditches0.5)	92.1%	55.9%	0.68	0.29
Deep learning (Streams0.5)	81.5%	70.8%	0.59	0.32

ditches) of 16.9%.

For ditches, our model Ditches0.5 had the highest recall: 92.1% against the reported 86.0% of ditch pixels from Lidberg et al. (2023); 33.8% (2 ha), 21.5% (6 ha), and 17.3% (10 ha) from the flow accumulation; and 27.5% from the Swedish property map. We believe that the differences between our deep learning model and the one from Lidberg et al. (2023), for either channel type, comes from the resolution: their model used 1 m, whereas our data was at a finer 0.5 m one. The lower recall rates of ditch pixels from the flow accumulation and Swedish property map could be explained by the absence of the ditch network, reported to be 91% missing from Swedish maps (Flyckt et al., 2022) before the use of deep learning.

Despite having a high recall rate for stream pixels, the MCC values had a low performance in both the baseline data and deep learning models, showing that there could be a bias towards finding positives at the expense of accuracy. In conclusion, our deep learning-based method for detecting channels outperformed traditional methods regarding ditches, where the recall reached 92.1%, but did not outperform the detection of natural streams. However, while one might argue that missing 29.2% of headwaters still requires further improvement, these results demonstrate that deep learning holds significant promise for improving automatic headwater mapping.

3.6. Limitations and future research

We believe that more studies are needed to improve the performance of class separation. Extracting additional features to the channels and training a separate model with them might improve the classification, especially with attributes related to drainage. The use of hydrological features in the future might answer whether the channel contains water or not and improve the network connectivity, avoiding the interruption of channels in the inference (Fig. 9). However, defining the banks of low relief channels can be particularly challenging if there are wetlands along the river course (Wohl, 2017), something that was observed in the study areas, causing the interruption of visible channels in the HPMF. Adding future information about culverts (Lidberg, 2025) and bridges might impact the inference connectivity as well. To deal with these occurrences, traditional topographic modeling could be applied, and with techniques such as burning and breaching, it might be possible to create the missing connectivity in the ground truth.

The dense canopy cover could have impacted the classification of small streams, potentially affecting the comparison of resolution performance too. The number of laser points is directly related to the resolution of the calculated DEM, however, as the canopy coverage becomes more dense in forested areas, the number of laser points that are able to penetrate it decreases (Chasmer et al., 2004). This could result in wrong terrain elevation estimates for densely covered areas, lowering the performance of the classification of small natural streams. With a higher amount of training data, it would be possible to separate the forested areas from the open ones to train the models, evaluating how much the tree tops were impacting the resolution performance. However, doing so with these datasets would result in a lower performance overall.

At the same time, while adding more data for this type of channel might seem like a solution, Yang et al. (2022) showed that this might not necessarily improve the models. Not only that, but the most time-consuming and expensive part of training a model with machine learning is acquiring the ground truth data, which in this study is due to the manual labeling and classification of channels relying on the terrain data and orthophotos. However, in dense vegetation covered sites, the orthophotos were not helpful, requiring an expert to visit the location and evaluate the channel type, which in turn increased the costs and time of the process. Despite these difficulties, the inclusion of aerial photographs and other data sources combined with ALS might be beneficial to the models, adding new characteristics to the channels.

Forwarding ruts were not observed in our dataset, but we acknowledge that this could be a cause for false positives. Some publications have focused on their identification using image data from drones (Bhatnagar et al., 2022) or conventional cameras (Pierzchała et al., 2016) unlike our study, which was based on the DEM. Another issue is that the vegetation can hinder the visual identification of these structures, making it hard to remove them from the data.

3.7. Water channel management and policies

Knowing the ambitious scope of the suggested actions by Agenda (2030) regarding water ecosystems, the management of both types of channels needs to be addressed. The measures allowed depend on the type of channel: riparian buffers are prescribed around streams, while ditch channels can be cleaned without permits (Swedish PEFC, 2023). Most ditches were detected in this study; however, streams were often misclassified as ditches. This is a cause for concern as streams have stronger protection policies than ditches during forest management. For example, crossing streams with heavy forest machinery should be avoided according to best management practices (Skogsstyrelsen, 2016) to avoid disturbing soils near and in the stream; such disturbance causes downstream sedimentation (Bishop et al., 2009). Meanwhile, ditches are not protected, and the full length of the ditch can be dug out and cleaned, also causing downstream sedimentation (Bishop et al., 2009); management procedures applied on natural channels would negatively change their characteristics, such as flow patterns and retention potential of detritus input (Muotka et al., 2002). Therefore, streams misclassified as ditches on maps could lead to the deterioration of both local and downstream environments if these maps were unquestioningly trusted by practitioners.

We suggest caution then when implementing models trained on just ditches: our model trained on this dataset misclassified 50% of the stream channels as ditch channels. This advice also concerns the ditch map developed by Lidberg et al. (2023). We are confident that further studies on how to separate ditches and streams on maps are needed.

A restoration process is currently underway to turn some of the Finnish channelized streams back to their natural status, thus improving sport fisheries (Erkinaro et al., 2011), while demonstration restorations have also been done in a number of Swedish rivers (Gardeström et al., 2013). However, studies focused on the restoration of small stream channels (<6 m) of the sort that we investigated are still missing. A

better classification of natural streams can benefit these studies and practices, further helping us to reach the water goals set by the Agenda 2030.

4. Conclusion

With this work, we have identified several key findings:

- 1) Resolution impact: The 0.5 m resolution significantly improved the detection of both ditches and natural stream channels, leading to higher overall performance. However, the finer resolution also required more computing power for processing the training data, training and testing the model, and running inference. highlighting the need for parallelizing the code and executing it on the GPU.
- 2) Topographic Index Performance: The highest-scoring topographic index varied depending on the dataset. The High-Pass Median Filter performed best for Channels0.5 and Ditches&Streams0.5 (ditch label), while the Hillshade 90° was the top-ranking for Ditches0.5. For Streams0.5, Hillshade 0° ranked higher.
- 3) Combining indices: Using a combination of indices resulted in higher values of MCC than single indices, with the combination of Sky-view Factor and Slope having the highest value for the stream label.
- 4) U-net performance: Our deep learning model Ditches0.5 was able to detect ditches better than any previous method (Table 5). In comparison with traditional mapping methods, the detection for ditches increased from less than 40% to over 92%, while Streams0.5 could map 70.8% of stream pixels.

Hence, our study shows great potential for using deep learning for mapping small headwaters, whether natural or man-made. However, the detection of natural streams still needs improving as close to 30% of them are still missing on the resulting maps. Future research should focus on identifying shared morphological features between ditch and stream channels, exploring methods to reduce class imbalance, and incorporating additional data such as information on soils, catchment area, and channel morphology. Improving automatic channel detection and classification of natural and man-made channels can provide valuable support for future improved management decisions for surface waters and optimize resource allocation for landscape planning.

CRedit authorship contribution statement

Mariana Dos Santos Toledo Busarello: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation. **Anneli M. Ågren:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Florian Westphal:** Writing – review & editing, Visualization, Software, Methodology, Investigation. **William Lidberg:** Writing – review & editing, Supervision, Software, Resources, Methodology, Funding acquisition, Conceptualization.

Code availability section

Contact: Mariana Dos Santos Toledo Busarello, Skogsmarksgränd 901 83 Umeå, Sweden, mariana.busarello@slu.se.

The hardware requirements are an x86 CPU with at least 16 GB of RAM. All code was written in Python, and the software requirements are Python and Docker.

The open source codes (MIT) are available for download at: <http://github.com/mbusarello/Automatic-Detection-of-Ditches-and-Natural-Streams-from-Digital-Elevation-Models-Using-Deep-Learning>.

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.

Acknowledgements

We thank Liselott Nilsson, Eliza Maher Hasselquist, Elijah Ourth, Siddhartho Shekhar Paul, Amanda Jarefjäll, Gudrun Norstedt, Lars Strand, Anders Hejnebo, Björn Lehto, Magnus Martinsson, Marcus Björzell, Tobias Johansson, Andrew Landström, and Catarina Welin for digitizing the ditches and streams used to train the model in this study. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation, the Marcus and Amalia Wallenberg Foundation, Kempestiftelserna, the Swedish Forest Agency, Formas (Proj no. 2021-00115), and Future Silviculture (2018.0259) financed by the Knut and Alice Wallenberg Foundation.

Data availability

The data used in this project can be found at the link: <https://doi.org/10.5878/jrex-z325>.

References

- Ågren, A.M., Lidberg, W., Ring, E., 2015. Mapping temporal dynamics in a forest stream network – implications for riparian forest management. *Forests* 6, 2982–3001. <https://doi.org/10.3390/f6092982>.
- Alexander, R.B., Boyer, E.W., Smith, R.A., Schwarz, G.E., Moore, R.B., 2007. The role of headwater streams in downstream water quality 1. *JAWRA Journal of the American Water Resources Association* 43, 41–59. <https://doi.org/10.1111/j.1752-1688.2007.00005.x>.
- Aslam, J.A., Yilmaz, E., Pavlu, V., 2005. A geometric interpretation of r-precision and its correlation with average precision. In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Presented at the SIGIR05: the 28th ACM/SIGIR International Symposium on Information Retrieval 2005. Salvador, Brazil, pp. 573–574. <https://doi.org/10.1145/1076034.1076134>. ACM.
- Beschta, R.L., Platts, W.S., 1986. Morphological features of small streams: significance and function 1. *JAWRA Journal of the American Water Resources Association* 22 (3), 369–379.
- Bhatnagar, S., Puliti, S., Talbot, B., Heppelmann, J.B., Breidenbach, J., Astrup, R., 2022. Mapping wheel-ruts from timber harvesting operations using deep learning techniques in drone imagery. *Forestry* 95 (5), 698–710.
- Bishop, K., Buffam, I., Erlandsson, M., Fölster, J., Laudon, H., Seibert, J., Temnerud, J., 2008. *Aqua Incognita*: the unknown headwaters. *Hydrol. Process.* 22, 1239–1242. <https://doi.org/10.1002/hyp.7049>.
- Bishop, K., Allan, C., Bringmark, L., Garcia, E., Hellsten, S., Högbom, L., Johansson, K., Lomander, A., Meili, M., Munthe, J., Nilsson, M., Porvari, P., Skyllberg, U., Sørensen, R., Zetterberg, T., Åkerblom, S., 2009. The effects of forestry on Hg bioaccumulation in nemoral/boreal waters and recommendations for good silvicultural practice. *AMBIO A J. Hum. Environ.* 38, 373–380. <https://doi.org/10.1579/0044-7447-38.7.373>.
- Busarello, M.d.S.T., Lidberg, W., Ågren, A., Westphal, F., 2024. Automatic Detection of Ditches and Natural Streams from Digital Elevation Models Using Deep Learning (Version 1). Swedish University of Agricultural Sciences. <https://doi.org/10.5878/jrex-z325> [Data set].
- Chasmer, L., Hopkinson, C., Treitz, P., 2004. Assessing the three-dimensional frequency distribution of airborne and ground-based LiDAR data for red pine and mixed deciduous forest plots. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, 66–69.
- Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* 21, 6. <https://doi.org/10.1186/s12864-019-6413-7>.
- Chowdhuri, I., Pal, S.C., Saha, A., Chakraborty, R., Roy, P., 2021. Evaluation of different DEMs for gully erosion susceptibility mapping using in-situ field measurement and validation. *Ecol. Inf.* 65, 101425. <https://doi.org/10.1016/j.ecoinf.2021.101425>.
- Claessens, L., Tague, C.L., Groffman, P.M., Melack, J.M., 2010. Longitudinal and seasonal variation of stream N uptake in an urbanizing watershed: effect of organic matter, stream size, transient storage and debris dams. *Biogeochemistry* 98, 45–62.
- Clough, P., Sanderson, M., Tang, J., Gollins, T., Warner, A., 2013. Examining the limits of crowdsourcing for relevance assessment. *IEEE Internet Computing* 17, 32–38. <https://doi.org/10.1109/MIC.2012.95>.
- Council of the European Union, 2023. Interinstitutional File: 2022/0195 (COD). No. 15907/23). Brussels.
- Cox, E.J., Gurnell, A.M., Bowes, M.J., Bruen, M., Hogan, S.C., O'Sullivan, J.J., Kelly-Quinn, M., 2023. A multi-scale analysis and classification of the hydrogeomorphological characteristics of Irish headwater streams. *Hydrobiologia* 850 (15), 3391–3418.
- Desmond, M., Duesterwald, E., Brimijoin, K., Brachman, M., Pan, Q., 2021. Semi-automated data labeling. *J. Mach. Learn. Res.* 133, 159–169.

- Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. *ISPRS J. Photogrammetry Remote Sens.* 158, 63–75. <https://doi.org/10.1016/j.isprsjprs.2019.09.018>.
- Du, L., McCarty, G.W., Li, X., Zhang, X., Rabenhorst, M.C., Lang, M.W., Zou, Z., Zhang, Xuesong, Hinson, A.L., 2024. Drainage ditch network extraction from lidar data using deep convolutional neural networks in a low relief landscape. *J. Hydrol.* 628, 130591. <https://doi.org/10.1016/j.jhydrol.2023.130591>.
- Eigen, D., Fergus, R., 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015. IEEE International Conference on Computer Vision (ICCV), IEEE, Santiago, Chile, pp. 2650–2658. <https://doi.org/10.1109/ICCV.2015.304>.
- Erkinaro, J., Laine, A., Mäki-Petäys, A., Karjalainen, T.P., Laajala, E., Hirvonen, A., Orell, P., Yrjänä, T., 2011. Restoring migratory salmonid populations in regulated rivers in the northernmost Baltic Sea area, Northern Finland - biological, technical and social challenges: restoring migratory salmonid populations in regulated rivers. *J. Appl. Ichthyol.* 27, 45–52. <https://doi.org/10.1111/j.1439-0426.2011.01851.x>.
- Fei, J., Liu, J., Ke, L., Wang, W., Wu, P., Zhou, Y., 2022. A deep learning-based method for mapping alpine intermittent rivers and ephemeral streams of the Tibetan Plateau from Sentinel-1 time series and DEMs. *Rem. Sens. Environ.* 282, 113271.
- Florinsky, I.V., 2016. Digital terrain modeling. In: *Digital Terrain Analysis in Soil Science and Geology*. Elsevier. <https://doi.org/10.1016/B978-0-12-804632-6.00001-8>.
- Flyckt, J., Andersson, F., Lavesson, N., Nilsson, L., Ågren, A.M., 2022. Detecting ditches using supervised learning on high-resolution digital elevation models. *Expert Syst. Appl.* 201, 116961. <https://doi.org/10.1016/j.eswa.2022.116961>.
- Gao, K., Huang, L., Zheng, Y., 2022. Fault detection on seismic structural images using a nested residual U-net. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–15. <https://doi.org/10.1109/TGRS.2021.3073840>.
- Gardeström, J., Holmqvist, D., Polvi, L.E., Nilsson, C., 2013. Demonstration restoration measures in tributaries of the vindel river catchment. *Ecol. Soc.* 18, art8. <https://doi.org/10.5751/ES-05609-180308>.
- Hasselquist, E.M., Mancheva, I., Eckerberg, K., Laudon, H., 2020. Policy change implications for forest water protection in Sweden over the last 50 years. *Ambio* 49, 1341–1351. <https://doi.org/10.1007/s13280-019-01274-y>.
- Huryn, A.D., Wallace, J.B., 1987. Local geomorphology as a determinant of macrofaunal production in a mountain stream. *Ecology* 68 (6), 1932–1942.
- Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface water mapping by deep learning. *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 10 (11), 4909–4918.
- Kampffmeyer, M., Salberg, A.-B., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Las Vegas, NV, USA, pp. 680–688. <https://doi.org/10.1109/CVPRW.2016.90>.
- Kazimi, B., Thiemann, F., Sester, M., 2020. Detection of terrain structures in Airborne Laser Scanning data using deep learning. *ISPRS Annals of the Photogrammetry. Remote Sensing and Spatial Information Sciences* 493–500. <https://doi.org/10.5194/isprs-annals-v-2-2020-493-2020>.
- Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization. *Conference Track Proceedings*, San Diego, CA, USA.
- Kokalj, Ž., Zakšek, K., Ostir, K., Pehani, P., Cotar, K., Somrak, M., 2016. Relief Visualization Toolbox, Ver. 2.2.1 Manual.
- Korznikov, K.A., Kislav, D.E., Altman, J., Doležal, J., Vozmishcheva, A.S., Krestov, P.V., 2021. Using U-Net-Like deep convolutional neural networks for precise tree recognition in very high resolution RGB (red, green, blue) satellite images. *Forests* 12, 66. <https://doi.org/10.3390/f12010066>.
- Koski, C., Kettunen, P., Poutanen, J., Zhu, L., Oksanen, J., 2023. Mapping small watercourses from DEMs with deep learning—exploring the causes of false predictions. *Rem. Sens.* 15, 2776. <https://doi.org/10.3390/rs15112776>.
- Kotsiantis, S., Kanellopoulos, D., Pintelas, P., 2006. Handling imbalanced datasets: a review. *GESTS international transactions on computer science and engineering* 30 (1), 25–36.
- Kuglerová, L., Hasselquist, E.M., Richardson, J.S., Sponseller, R.A., Kreutzweiser, D.P., Laudon, H., 2017. Management perspectives on *Aqua incognita*: connectivity and cumulative effects of small natural and artificial streams in boreal forests. *Hydrol. Process.* 31, 4238–4244. <https://doi.org/10.1002/hyp.11281>.
- Kuglerová, L., Jyväsjärvi, J., Ruffing, C., Muotka, T., Jonsson, A., Andersson, E., Richardson, J.S., 2020. Cutting edge: a comparison of contemporary practices of riparian buffer retention around small streams in Canada, Finland, and Sweden. *Water Resour. Res.* 56, e2019WR026381. <https://doi.org/10.1029/2019WR026381>.
- Kwame, A.E., Martey, E.M., Chris, A.G., 2017. Qualitative assessment of compiled, interpreted and hybrid programming languages. *Communications on Applied Electronics* 7 (7), 8–13.
- Lanka, R.P., Hubert, W.A., Wesche, T.A., 1987. Relations of geomorphology to stream habitat and trout standing stock in small Rocky Mountain streams. *Trans. Am. Fish. Soc.* 116 (1), 21–28.
- Lantmäteriet, 2014. GSD-fastighetskartan, Vektor.
- Lantmäteriet, 2021a. Orthophoto.
- Lantmäteriet, 2021b. Laser Data.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>.
- Lidberg, W., 2025. Deep learning-enhanced detection of road culverts in high-resolution digital elevation models: improving stream network accuracy in Sweden. *J. Hydrol. Reg. Stud.* 57, 102148. <https://doi.org/10.1016/j.ejrh.2024.102148>.
- Lidberg, W., Nilsson, M., Lundmark, T., Ågren, A.M., 2017. Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrol. Process.* 31, 4660–4668. <https://doi.org/10.1002/hyp.11385>.
- Lidberg, W., Paul, S.S., Westphal, F., Richter, K.F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A., Ågren, A.M., 2023. Mapping drainage ditches in forested landscapes using deep learning and aerial laser scanning. *J. Irrigat. Drain. Eng.* 149, 04022051. <https://doi.org/10.1061/JIDEDH.IRENG-9796>.
- Lindsay, J.B., 2016. Whitebox GAT: a case study in geomorphometric analysis. *Comput. Geosci.* 95, 75–84. <https://doi.org/10.1016/j.cageo.2016.07.003>.
- Lindsay, J.B., Dhun, K., 2015. Modelling surface drainage patterns in altered landscapes using LiDAR. *Int. J. Geogr. Inf. Sci.* 29, 397–411. <https://doi.org/10.1080/13658816.2014.975715>.
- Martínez, A., Larranaga, A., Basaguren, A., Pérez, J., Mendoza-Lera, C., Pozo, J., 2013. Stream regulation by small dams affects benthic macroinvertebrate communities: from structural changes to functional implications. *Hydrobiologia* 711, 31–42.
- Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta Protein Struct.* 405, 442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9).
- Mazhar, S., Sun, G., Bilal, A., Hassan, B., Li, Y., Zhang, J., Lin, Y., Khan, A., Ahmed, R., Hassan, T., 2022. AUnet: a deep learning framework for surface water channel mapping using large-coverage remote sensing images and sparse scribble annotations from OSM data. *Rem. Sens.* 14 (14), 3283.
- Muotka, T., Paavola, R., Haapala, A., Novikmec, M., Laasonen, P., 2002. Long-term recovery of stream habitat structure and benthic invertebrate communities from in-stream restoration. *Biol. Conserv.* 105, 243–253. [https://doi.org/10.1016/S0006-3207\(01\)00202-6](https://doi.org/10.1016/S0006-3207(01)00202-6).
- Nieminen, M., Piirainen, S., Sikström, U., Löfgren, S., Marttila, H., Sarkkola, S., Laurén, A., Finér, L., 2018. Ditch network maintenance in peat-dominated boreal forests: review and analysis of water quality management options. *Ambio* 47, 535–545. <https://doi.org/10.1007/s13280-018-1047-6>.
- NLS, 2023. NLS orthophotos [WWW Document]. URL: <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/ortho-photos>.
- O’Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. *Comput. Vis. Graph Image Process* 28, 323–344.
- Paul, S.S., Hasselquist, E.M., Jarefjäll, A., Ågren, A.M., 2023. Virtual landscape-scale restoration of altered channels helps us understand the extent of impacts to guide future ecosystem management. *Ambio* 52, 182–194. <https://doi.org/10.1007/s13280-022-01770-8>.
- Peacock, M., Audet, J., Bastviken, D., Futter, M.N., Gauci, V., Grinham, A., Harrison, J. A., Kent, M.S., Kosten, S., Lovelock, C.E., Veraart, A.J., Evans, C.D., 2021a. Global importance of methane emissions from drainage ditches and canals. *Environ. Res. Lett.* 16, 044010. <https://doi.org/10.1088/1748-9326/abeb36>.
- Peacock, M., Granath, G., Wallin, M.B., Högbom, L., Futter, M.N., 2021b. Significant emissions from forest drainage ditches—an unaccounted term in anthropogenic greenhouse gas inventories? *J. Geophys. Res.: Biogeosciences* 126. <https://doi.org/10.1029/2021JG006478>.
- Phinzi, K., Abriha, D., Bertalan, L., Holb, I., Szabó, S., 2020. Machine learning for gully feature extraction based on a pan-sharpened multispectral image: multiclass vs. Binary approach. *International Journal of Geo-Information* 9, 252. <https://doi.org/10.3390/ijgi9040252>.
- Pierzchała, M., Talbot, B., Astrup, R., 2016. Measuring wheel ruts with close-range photogrammetry. *Forestry: Int. J. Financ. Res.* 89 (4), 383–391.
- Raber, G.T., Jensen, J.R., Schill, S.R., Schuckman, K., 2002. Creation of digital terrain models using an adaptive lidar vegetation point removal process. *Photogramm. Eng. Rem. Sens.* 68, 1307–1315.
- Ring, E., Lode, P.E., Libietė, Z., Oksanen, E., Gil, W., Šimkevičius, K., 2018. Good Practices for Forest Buffers to Improve Surface Water Quality in the Baltic Sea Region. *Arbetsrapport No. 995–2018*.
- Robb, C., Pickard, A., Williamson, J.L., Fitch, A., Evans, C., 2023. Peat drainage ditch mapping from aerial imagery using a convolutional neural network. *Rem. Sens.* 15, 499. <https://doi.org/10.3390/rs15020499>.
- Robson, B.A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., Schaffer, N., 2020. Automated detection of rock glaciers using deep learning and object-based image analysis. *Rem. Sens. Environ.* 250, 112033. <https://doi.org/10.1016/j.rse.2020.112033>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*. Springer International Publishing, pp. 234–241, 18.
- Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* 9, 82031–82057.
- Skogsstyrelsen, 2013. SKSFS 2013:2.
- Skogsstyrelsen, 2016. Nya Och Reviderade Målbilder För God Miljöhänsyn, No. 2016, p. 12.
- Swedish PEFC, 2023. Forest use standard. Technical Report PEFC SWE 002:5. Swedish PEFC.
- Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* 82 (398), 528–540.
- United Nations General Assembly, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development.
- Wiens, J.A., 2002. Riverine landscapes: taking landscape ecology into the water. *Freshw. Biol.* 47 (4), 501–515.

- Wilson, J.P., Gallant, J.C. (Eds.), 2000. *Terrain Analysis: Principles and Applications*. Wiley, New York.
- Wohl, E., 2017. The significance of small streams. *Front. Earth Sci.* 11, 447–456.
- Xu, S., Song, Y., Hao, X., 2022. A comparative study of shallow machine learning models and deep learning models for landslide susceptibility assessment based on imbalanced data. *Forests* 13, 1908. <https://doi.org/10.3390/f13111908>.
- Yang, J., Zhang, Z., Gong, Y., Ma, S., Guo, X., Yang, Y., Xiao, S., Wen, J., Li, Y., Gao, X., Lu, W., Meng, Q., 2022. Do deep neural networks always perform better when eating more data? *arXiv* 2022. [arXiv preprint arXiv:2205.15187](https://arxiv.org/abs/2205.15187).
- Yang, J., Xu, J., Lv, Y., Zhou, C., Zhu, Y., Cheng, W., 2023. Deep learning-based automated terrain classification using high-resolution DEM data. *Int. J. Appl. Earth Obs. Geoinf.* 118, 103249. <https://doi.org/10.1016/j.jag.2023.103249>.
- Zakšek, K., Oštir, K., Kokalj, Z., 2011. Sky-view factor as a relief visualization technique. *Rem. Sens.* 3, 398–415. <https://doi.org/10.3390/rs3020398>.