International Cartographic Association
Association Cartographique Internationale

Classifying challenges in machine learning of small watercourse recognition

Justus Poutanen, Pyry Kettunen, Anssi Jussila, Christian Koski*
Finnish Geospatial Research Institute (FGI, NLS), justus.poutanen@nls.fi, pyry.kettunen@nls.fi, anssi.jussila@nls.fi, christian.koski@nls.fi

* Corresponding author

Keywords: Machine learning, watercourse, point cloud, hydrography, topographic data

Abstract:

As lidar data becomes more precise and accessible, the use of machine learning (ML) in Geographic Information System (GIS) applications becomes more viable. Automatic recognition of watercourses from high resolution point clouds, digital elevation models (DEMs) and other remote sensing data by convolutional neural networks (CNN) allows for highly efficient production of geospatial data (Paul, Ågren & Lidberg 2021). While our proposed CNN based framework demonstrates promising results, several challenges persist, predominantly linked to the flaws in the input data (Koski et al., 2023). National Land Survey of Finland (NLS) is aiming at renewing its watercourse mapping from high resolution point clouds. In this study we classify challenges that arise in automatic recognition of small watercourses to help better understand them.

This study was conducted across two 36 km² study areas in central Finland, one located near Ristijärvi and the other near Heinävesi. For the label data of the ML model, all watercourses less than five meters wide were manually digitized by a professional mapper. The digitization was done in QGIS with the help of seven layers to highlight the watercourses from the background: 1) a DEM at 0.25 m resolution created from a lidar point cloud with density of 20 per square meter (20p) dataset, 2) a DEM-derived hillshade, 3) relative topographic position (RTP), 4) flow accumulation (FA), 5) slope aspect, 6) NLS Orthophotos, and 7) the NLS Topographic Map. The ML model was trained with the digitized dataset and a DEM at 0.25 m resolution derived from the 20p lidar dataset.

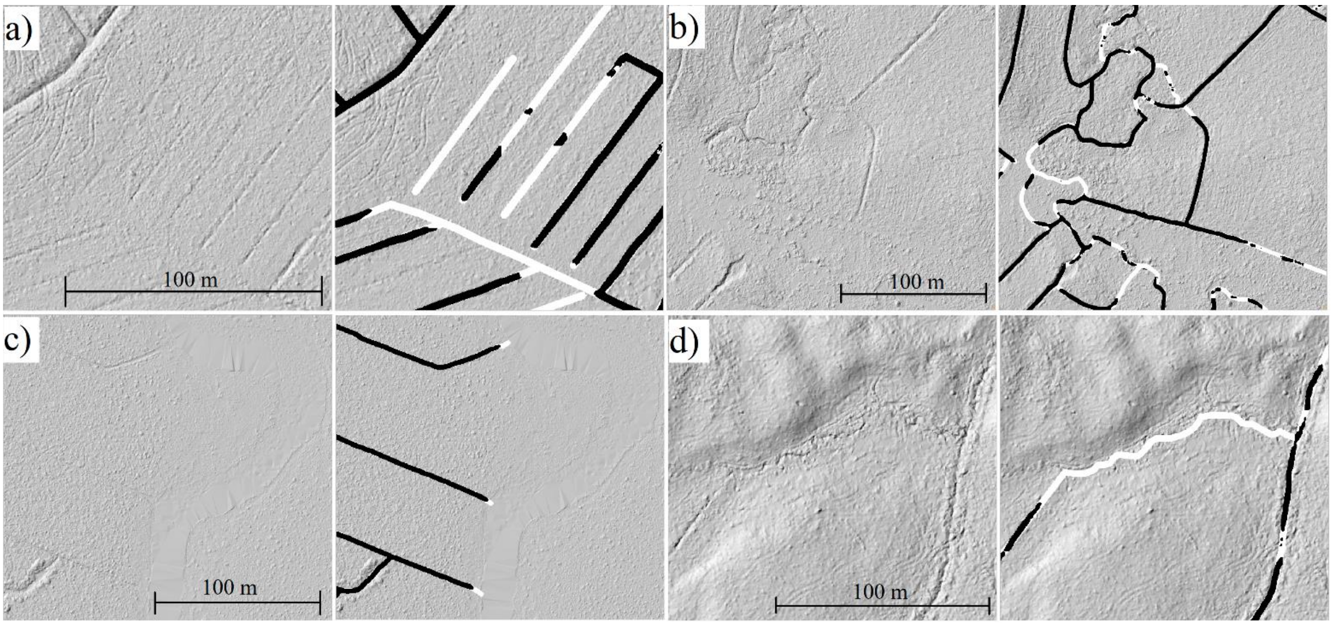

Figure 1. Main challenges of watercourse recognition with ML methods; **a)** watercourses that are very shallow; **b)** watercourses that widen; **c)** watercourses that touch a waterbody; **d)** natural streams with small meanders. Predicted watercourses (black) are shown on top of the labeled ones (white).

Challenging watercourse sections can be identified with an overlay analysis to locate where the ML prediction surface fails to cover features in the label data. There are four main challenges that create false negatives in automatic watercourse

Abstracts of the International Cartographic Association, 10, 234, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-234-2025

recognition (Figure 1): 1) watercourses that are very shallow, 2) watercourses that widen, 3) watercourses that touch a waterbody, and 4) natural streams with small meanders. Very shallow watercourses are difficult for the ML model to recognize, since even as shallow as 10 centimeters of elevation change can be detected from the DEMs. Also, in practice, these extremely shallow features are not relevant for many needs of watercourse network data. Watercourses that begin narrow but spread too wide are challenging for the model as well, because our training label data consisted of watercourses less than five meters wide. In these cases, this classification creates a challenge to maintain watercourse network integrity. In cases where the watercourse touches a lake or river, it is unclear even for humans to determine where the watercourse ends, and waterbody begins. Marshlands make this distinction even more difficult. Lastly, natural streams are more difficult for the model to recognize due to irregularities in their geometry, and tendency to meander. Moreover, unlike drainage ditches, natural streams often do not have clear edges, making them more challenging to recognize.

Most importantly, all these challenges are subject to temporality. Depending on the season and recent weather events before the airborne laser scanning (ALS) campaign, changes in water level can be significant. During the spring when the snows are melting, or after heavy rainfall, water is likely to run in even the shallowest watercourses. In late summer, water levels can be lower, and in late autumn, shallow watercourses can fill with leaves, concealing some watercourses from the ALS's laser pulses.

Addressing these challenges could further improve accuracy and thus the viability of ML methods in the automatic recognition of watercourses. Future research could focus on determining which watercourses have a minor significance for the larger watercourse network.

Acknowledgements

This research has been conducted as part of the development work of the Basic Spatial Data Services of the National Land Survey of Finland and using the CSC high-performance computing by Geoportti.

References

Paul, S. S., Ågren, A. M., & Lidberg, W. (2021, April). Detection of drainage ditches using high-resolution LIDAR data in the Swedish forest. In *EGU General Assembly Conference Abstracts* (pp. EGU21-2226).

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. *Remote Sensing*, *15*(11), 2776.

https://doi.org/10.5194/ica-abs-10-234-2025