

# Piloting the use of machine learning methods for automatic mapping of streams and ditches in Finland

Christian Koski <sup>a, \*</sup>, Pyry Kettunen <sup>a</sup>, Justus Poutanen <sup>a</sup>, Lingli Zhu <sup>a</sup>, Juha Oksanen <sup>a</sup>

<sup>a</sup> Finnish Geospatial Research Institute (FGI, NLS), christian.koski@nls.fi, pyry.kettunen@nls.fi, justus.poutanen@nls.fi, juha.oksanen@nls.fi

\* Corresponding author

**Keywords:** Machine learning, Hydrography, Streams, Ditches

## Abstract:

National mapping authorities aim to increase automation in their mapping processes. Automation increases the rate of updating features in topographic databases (TDBs), improves the quality in terms of regional homogeneity and logical consistency between the mapped features, and lowers costs. Recently, the National Land Survey of Finland (NLS) has achieved countrywide coverage for its digital elevation model with a 2-meter resolution (NLS-DEM2m). During the production of the NLS-DEM2m, the NLS has initiated several pilot projects to find automatic methods for extraction of topographic features from the DEM, including, contour lines (Kettunen, Koski and Oksanen 2017), cliffs, and hydrographic features. While producing encouraging results, these projects have highlighted the inconsistency between automatically generated features from the NLS-DEM2m and manually mapped features in the NLS TDB. For example, this is apparent when overlaying automatically generated contour lines and the current manually mapped hydrographic features of the TDB (Figure 1).

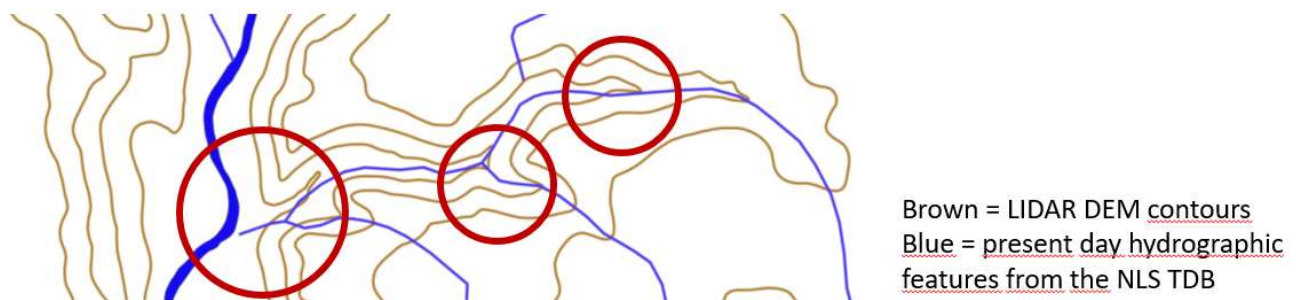


Figure 1. Streams poorly match contour lines that are automatically produced from NLS-DEM2m.

Increased computing power, more data, and better models have made machine learning (ML) popular for automation. In May 2021, NLS started piloting automated mapping of hydrographic features from the NLS-DEM2m with the use of ML. Using the NLS-DEM2m to map hydrographic features to the NLS TDB can improve their positional accuracy and completeness and make them consistent with automatically generated contour lines.

Previous piloting of automated mapping of hydrographic features with non-ML methods from high-resolution DEMs has not yet achieved a high enough quality for them to be viable options for fully automated production of hydrographic features (Eraña-Beitia 2013, Oksanen 2014). The main challenges have been to correctly identify and map small streams and ditches. Therefore, in this project, we will focus mainly on ML-based extraction of hydrographic features. The objectives of automated mapping of hydrographic features in the project are as follows:

1. To assess the quality of the ditch and stream vectors that are derived from the results of ML models.
2. To identify best practices for producing test and validation data of ML models that detect ditches and streams from the NLS-DEM2m.
3. To identify the most critical areas of the stream and ditch extraction process that need to be improved.
4. To improve the used methods by developing new and innovative solutions for automated mapping of hydrographic features from DEMs.

Recently, convolutional neural networks (CNNs) have been used to identify both streams (Xu et al. 2021) and ditches (Paul, Ågren, and Lidberg 2021) from DEMs. These CNN approaches have displayed promising results using DEMs with resolutions as high as 1 and 0.5 meters as input data. Results of the CNN models have been assessed based on recall and

precision on the pixel level. We will apply these CNN approaches and assess results achieved with them, using the NLS-DEM2m as input data as this is the highest resolution country-wide DEM that is available. The results will then be used to generate ditch and stream vectors, which will be assessed based on their completeness and positional accuracy.

The quality of results produced by CNN models is dependent on high-quality training and validation data, as well as input data. The positional accuracy and completeness of the data are critical for producing high quality results. However, mapping ditches and streams, especially those in forested areas can be difficult, even when done manually from the available source datasets (Figure 2). Apart from the positional accuracy and completeness of the training data, we need to assess how much training data is needed, what types of test areas are needed considering different types of terrain in Finland, and how to produce the most accurate data possible considering available resources and source data.

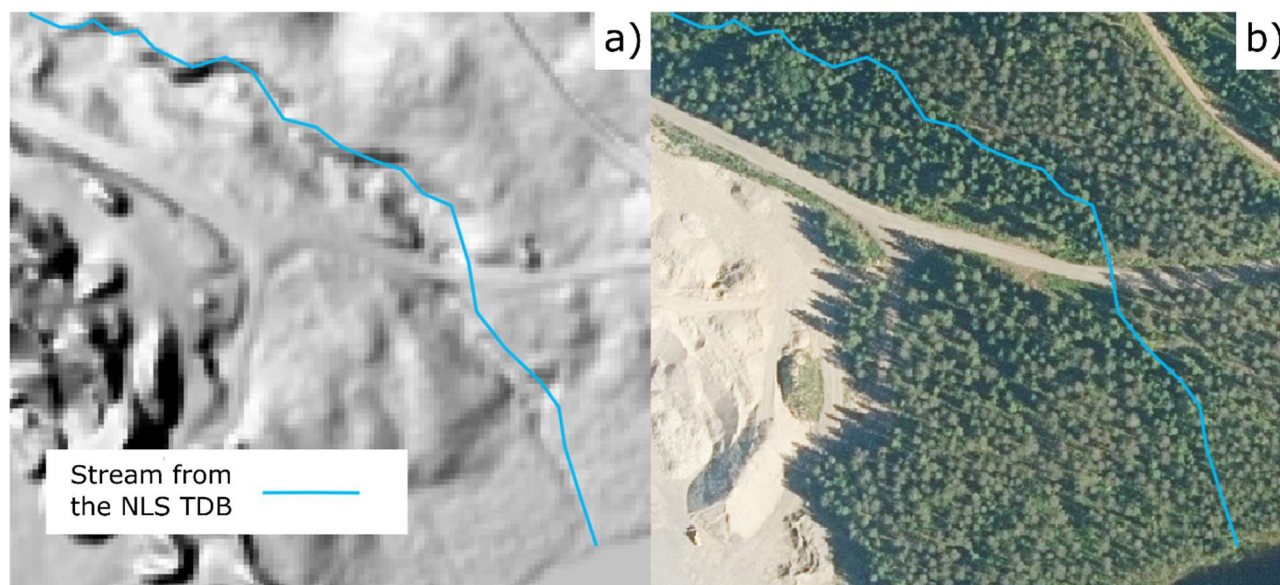


Figure 2. It is difficult to determine the exact pixels that belong to streams and ditches when manually digitizing them from high-resolution datasets (e.g. hillshading raster derived from the NLS-DEM2m) (a). Aerial images with a resolution of 0.5 meters can provide better accuracy but provide no information in forest areas (b).

In the final steps of the project, we aim to identify areas that need improvements and develop methods to improve the outcome. Possible areas of improvement include quality of training data, resolution of input DEM, CNN model architecture, and automated vectorization of the CNN results. We will identify the most promising areas of improvement by testing higher resolution DEMs produced from denser point clouds, by assessing the impact of the quality of training and validation data on the results, by testing different CNN models, and by comparing different methods for vectorization of the CNN model results.

## References

- Eraña-Beitia, X. (2013). Updating of hydrographic features in topographic base information based on airborne laser scanning data. Master of Science thesis, Aalto University.
- Kettunen, P., Koski, C., & Oksanen, J. (2017). A design of contour generation for topographic maps with adaptive DEM smoothing. *International Journal of Cartography*, 3(1), 19-30.
- Oksanen, J. (2014). VERSATILE - Valtakunnallisen laserkeilausaineiston monipuolinen hyödyntäminen, loppuraportti (Versatile use of the country-wide airborne LIDAR point clouds, final report) (In Finnish). Unpublished report.
- 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).
- Xu, Zewei, Shaowen Wang, Lawrence V. Stanislawski, Zhe Jiang, Nattapon Jaroenchai, Arpan Man Sainju, Ethan Shavers et al. "An attention U-Net model for detection of fine-scale hydrologic streamlines." *Environmental Modelling & Software* 140 (2021): 104992.