Applied Computing and Geosciences 18 (2023) 100119

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

Where are the outcrops? Automatic delineation of bedrock from sediments using Deep-Learning techniques

Alexandra Jarna Ganerød [a,b,*], Vegar Bakkestuen [c], Martina Calovi [a], Ola Fredin [d], Jan Ketil Rød [a]

[a] NTNU, Department of Geography, Norway
[b] NGU, Geological Survey of Norway, Norway
[c] NINA, Norwegian Institute for Nature Research, Norway
[d] NTNU, Department of Geoscience and Petroleum, Norway

ARTICLE INFO

Keywords:
Deep learning
Bedrock
Sediment
U-net
Google Earth Engine (GEE)
Cloud computing

ABSTRACT

The delineating of bedrock from sediment is one of the most important phases in the fundamental process of regional bedrock identification and mapping, and it is usually manually performed using high-resolution optical remote-sensing images or Light Detection and Ranging (LiDAR) data. This task, although straightforward, is time consuming and requires extensive and specialized labor. We contribute to this line of research by proposing an automated approach that uses cloud computing, deep learning, fully convolutional neural networks, and a U-Net model applied in Google Collaboratory (Colab). Specifically, we tested this method on a site in southwestern Norway using both a set of explanatory variables generated from a 10 m resolution digital elevation model (DEM) and, for comparison, cloud-based Landsat 8 data. Results show an automatic delineation performance measured by an F1 score between 77% and 84% for DEM terrain derivatives against a manually-mapped ground truth. Overall, our automated bedrock identification model reveals very promising results within its constraints.

1. Introduction

Bedrock exposure identification provides crucial geological and geotechnical information which is used in infrastructure project planning, hazard mapping, and other research fields within Earth Science (Nordgulen, 2020). The first geological map of England and Wales, dated 1815, is one of the first examples of geological maps created through manual mapping. Since then, the mapping process has not undergone significant change; in general, the mapping geologist becomes familiar with the field area of interest, performs field work to observe and describe the area, and then analyzes and processes collected data and observations by manually digitizing, or drawing, the different geological features (Lisle et al., 2011).

Even though Geographic Information Systems (GIS) have made the digitization, storage, and aggregation of different datasets relatively easy, geological mapping is still a time-consuming and subjective (operator or mapper-dependent) practice. Therefore, there is significant untapped potential in using automated machine- and deep learning techniques to make geological mapping more efficient (Caté et al., 2017; Karpatne et al., 2019; Sircar et al., 2021; Xiong et al., 2018). Computer-based methods are rapidly emerging as powerful, efficient, and viable tools to analyze, extract, and synthesize large data sets (Chen and Lin, 2014; Dargan et al., 2020; Zhu et al., 2017). These methods can thus reduce costly and time-consuming manual labor, both in the preparation for geological fieldwork and afterwards during the post data interpretation and digitalization phase. The group of machine-learning algorithms called “deep learning”, a subfield of artificial intelligence, is rapidly becoming essential in geosciences (Zhang et al., 2016). Deep learning techniques are indeed largely used to classify or predict patterns in large datasets (Abiodun et al., 2018; Alavi et al., 2016; Baraniuk et al., 2020; Sarker, 2021; Zhou et al., 2019) Deep learning algorithms can process and deliver predictions with much less human effort and often yield very good performance with the use of multi-dimensional data. It has the potential to become a widely used method that increases the efficiency of future mapping of large areas in combination with expert revision of predictions (Donahue et al., 2017; Sengupta et al., 2020; Shelhamer et al., 2017). Previous studies on automated classification in geology focus mostly on digital soil mapping (Drăguţ et al., 2006; Grinand et al., 2008; Kerry and Oliver, 2011; MacMillan et al., 2004). However, this literature does not consider any automatic

* Corresponding author. NTNU, Department of Geography, Norway.
E-mail addresses: alexandra.jarna@ntnu.no (A.J. Ganerød), Vegar.Bakkestuen@nina.no (V. Bakkestuen), martina.calovi@ntnu.no (M. Calovi), ola.fredin@ntnu.no (O. Fredin), jan.rod@ntnu.no (J.K. Rød).

https://doi.org/10.1016/j.acags.2023.100119
Received 15 October 2022; Received in revised form 11 May 2023; Accepted 11 May 2023
Available online 20 May 2023

delineation and identification of bedrock from sediment. These studies rather show the use of high-quality landcover data or multi-scale neighborhood geometry, combined with different terrain derivatives (e.g., slope) (Behrens et al., 2018; Drăgut et al., 2011). Bedrock is usually characterized by a distinct rough and fractured terrain surface. It is therefore easy to detect from visual representations of high-resolution Digital Terrain Models (DTM) (e.g., hillshade). Although automatic classification using deep learning seems promising, there is a lack of stringent tests using this method. Indeed, the majority of the existing publications either still employ manual mapping or use statistics-based machine learning. Among the available publications, for example, is a study that uses image classification based on a Random Forest classifier on legacy land use data, using 43 different layers as predictors (such as various terrain derivatives, NDVI, ASTER and Landsat 7) (Scarpone et al., 2017). Milodowski et al. (2015) tried to identify areas of rock exposure from high resolution LiDAR data by using SVM classification, based on short-wavelength topographic roughness, by studying the local variability of surface normal vectors in Sierra Nevada, California (Milodowski et al., 2015). Harris et al. developed a supervised classification using an algorithm known as the Robust Classification Method (), and applied it to a variety of remotely sensed data, including Landsat 7, Landsat 8, Spot 5, Aster imagery and airborne magnetic data, producing predictions of bedrock lithology and Quaternary cover in Victoria Island, Canada (Harris et al., 2012, 2014), and also shows the use of two satellite-derived data layers (Landsat, ASTER) in a Random Forest model, over British Columbia (Canada).

While multiple studies, including those above and many others, have used supervised classification and various classifiers with many predictors, most pertain only to soil and sediment mapping, and only very few have tried to explore automatic delineation of bedrock outcrops. DiBiase et al. (2012) for example, identified rock exposures from LiDAR-derived slope measurements with the use of high-resolution panoramic photographs. However, the main goal of their study is not to automatically delineate bedrock from soil, but to define bedrock in steep hillslopes in order to check hillslope response to tectonic forcing, using slope maps generated from 1m LiDAR DEM. We believe that given the absence of deep learning techniques used for automated bedrock identification in the pertinent literature, this study proposes a novel and interesting model that will be of use in many areas of geological research.

Our study tests whether it is possible to automatically differentiate bedrock from sediments, with high precision, using U-Net architecture, Fully Convolutional Neural Networks (FCNN), Google Earth Engine (GEE) cloud computation, and Google Colab. The selected study area is on the coast in southern Norway, where both sediments and bedrock are well represented. Although geologists use Landsat 8 images in many mapping contexts, it is seldom used for soil and bedrock mapping. Previous studies (Harris et al., 2014; Scarpone et al., 2017) have claimed that satellite images can be used to improve bedrock delineation, and our study therefore seeks to add to the literature on this hypothesis. We expect a low-efficiency prediction when based only on Landsat 8 images as many of the outcrops in our study area are covered by vegetation. We therefore test the same ground truth data against two separate predictors: freely available cloud-based Landsat 8 images (USGS, 2021a), and seven terrain derivatives from a 10 m resolution DEM. Although a higher-resolution DEM is freely available for Norway (1 m ground sampling distance), as well as LiDAR data, this is not the case for all countries. We therefore decided to use a 10 m resolution DEM, to test the applicability and reproducibility of our model over other study areas in countries with scarcer data availability. Furthermore, we evaluated whether the resulting predictions vary according to different properties in the training data, such as distribution, amount, size, sampling design, and weighing the sampling intensity according to the percentage of bedrock in randomly created training rectangles by calculating precision, recall, F1-score and Matthew's correlation coefficient (MCC) scores. The goals of this study were twofold: first, to show whether deep learning can be used for delineating bedrock automatically to improve future geological mapping and second, to identify the ideal training model, by testing the effect on the automatic delineation performance of using Landsat 8 data exclusively versus additional terrain derivatives, and by testing the use of different settings of randomized training rectangles.

2. Study area

The study area encompasses the Farsund-Lista area in the most southern part of Norway (Fig. 1). This area is characterized by jointed and fractured Meso-to-Neoproterozoic granites and gneisses (Falkum, 1982). Quaternary sediments are generally restricted to valleys and troughs where glacial till, glaciofluvial outwash, and peat are found (Fredin et al., 2015). Bare bedrock outcrops dominate higher-lying areas with minimal sediment and organic (peat) cover. It is assumed that Quaternary glaciations have scoured the landscape to a large extent, possibly because of the proximity to the highly erosive Norwegian channel ice stream during the late stages of the last glaciation (Sejrup et al., 2003). The landscape is characterized by Quaternary fjords and valleys which cut into bedrock weakness zones. The mostly bare bedrock is characterized by small-to-medium topographic relief (0–500 msl) and is dissected by brittle fault structures and lineaments (Fig. 1). An exception to this pattern is the southwestern part of the Lista peninsula, which is covered by thick (i.e., several tens of meters of) drumlinized glacial till and marine sediments. The area was manually mapped during the years of 2012–2015 using high-quality LiDAR data (1 m ground sampling distance) and orthophotos, in combination with extensive field observations. The resulting 1:50 000 map clearly separates bedrock from sediments (Fredin et al., 2015). Consequently, we consider the area suitable for a study on whether this labor-intensive mapping can be done using automated methods, or more specifically, whether deep learning methods can be used to differ between bedrock and sediments with high levels of accuracy. Additionally, the area shows sufficient variation to provide both less-than-ideal and very good data, which is desirable both to challenge the model and to test the reliability of Landsat 8 data in such a zone. The size of this study area is about 307 km^2 counting only land area.

3. Methods – data and computing

3.1. Data input

3.1.1. Data and sampling design for generating training and evaluating dataset

A raster dataset (tiff format) that delineates bedrock and sediments over the study area was set as ground truth data. It was based on a recent quaternary map manually produced by the Geological Survey of Norway (NGU) at a scale of 1:50 000 (Fredin et al., 2015). This is the most precise manmade and published map of the area depicting bare rock and gives us an opportunity to predict bedrock at the same scale (i.e., 1:50 000). On the map, sediments are classified according to genesis. Two of the classes on the map cover 53% of the study area (shown in Fig. 2) and were chosen as reference data for training and validation of our model: "bare rock" and "thin organic soil on bedrock". The "bare rock" sediment class consists of exposed bedrock with minimal cover, although a small amount of vegetation and organic remains may be present in some cases. The bare rock signature, with visible bedrock slabs, joints, and fractures, is readily visible in LiDAR data and in the field. The "thin organic soil on bedrock" sediment class consists of bare bedrock with 10–30 cm of organic cover (plant remains), but no other sediments. Bare bedrock and thin organic soil on bedrock have very similar signatures and cannot be distinguished from each other using elevation model derivatives. However, when using aerial photographs or other optical remote sensing data, the "thin organic soil on bedrock" class has a signature resembling that of other vegetation-covered sediments, but that is beyond the scope

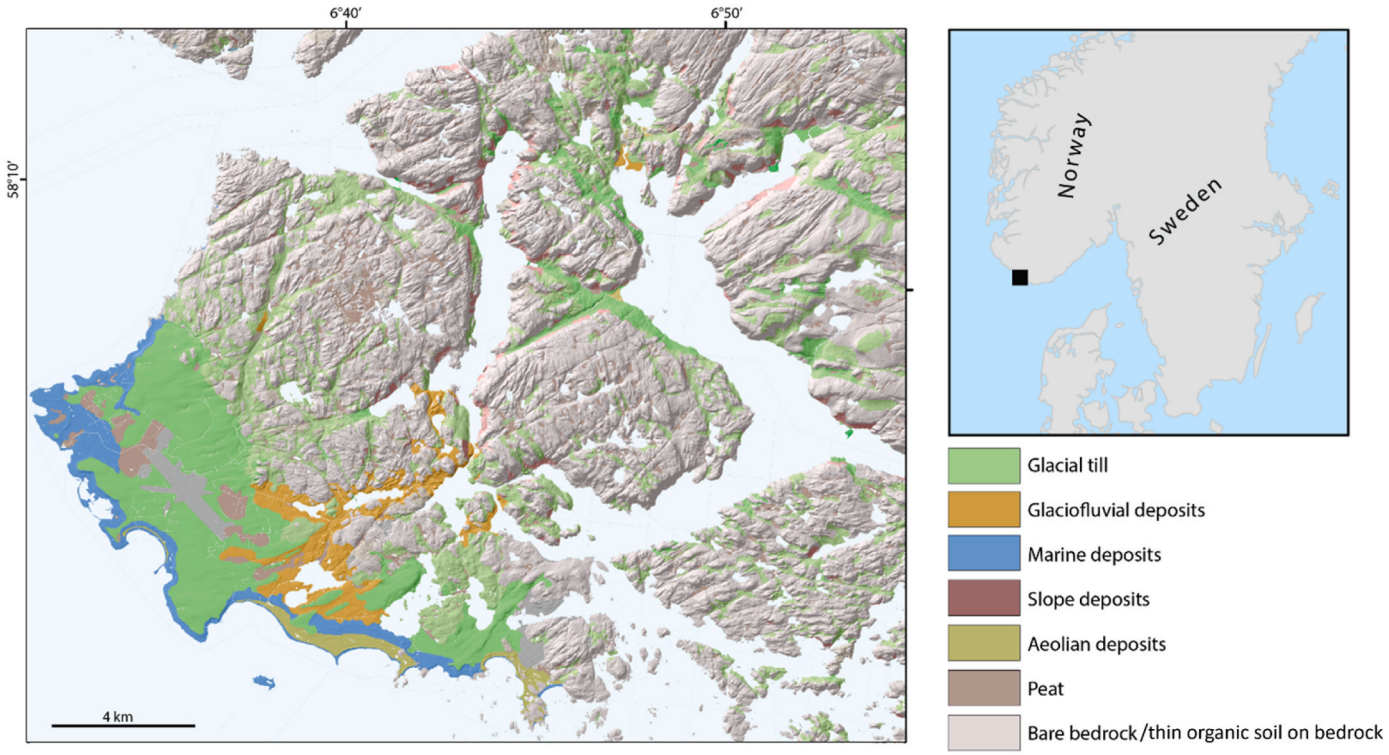

Fig. 1. Bare bedrock and different sediment types superposed on a hillshade in the investigated area (image closely corresponds with study area). Large amounts of sediments (till, glaciofluvial) were deposited during the last glaciation. Marine, slope, aeolian deposits and peat were formed in post-glacial time (around the last 12–14 ka).

of this study. For all practical purposes, bare bedrock and thin organic soil on bedrock can be considered as similar categories, and because the main attempt here was performed using the elevation model for automated classification, the two categories "bare bedrock" and "thin organic soil on bedrock" were merged. The ground truth dataset was converted to a raster dataset and superimposed with sea and water layers (Kartverket, 2018) to exclude water bodies from the calculations. For further calculations, the quaternary map was reclassified and given the value 1 (presence) for the bedrock/bare rock and 0 (absence) for sediments.

Fig. 2 shows two sets of rectangles: 10 evaluating rectangles in red and 10 training rectangles in blue. Training rectangles defined the areas used by the algorithm to learn how to differentiate between bedrock and sediments. Evaluating rectangles are used by the algorithm to check that features meet the conditions defined in the learning stage, to improve the quality. We generated the rectangles using a python script ensuring random location, random size (within a predetermined range), and some additional limiting conditions outlined below, to see how different sampling approaches influenced the prediction results. In addition, the rectangles had a common restricting condition, wherein they should all be entirely within the boundaries of the study area. The settings were as follows:

1) 10 training and 10 evaluating rectangles (within a range of 1.5–4 km^2) randomly spread across the study area;
2) 10 training and 10 evaluating rectangles (within a range of 1.5–4 km^2) randomly spread across the study area where the rectangles contain a minimum of 70% bedrock and do not intersect fjords;
3) 15 smaller training and 15 smaller evaluating rectangles (within a range of 0.3–0.8 km^2) randomly spread over the study area;
4) 15 smaller training and 15 smaller evaluating rectangles (within a range of 0.3–0.8 km^2) randomly spread across the study area where the rectangles contain a minimum of 70% bedrock and do not intersect fjords.

For each of these four settings, we repeated the process five times to create five different cases (such as depicted in Fig. 2), to be able to observe any patterns and to compare variation in the determined setting. As illustrated in Fig. 3, this sample design resulted in twenty cases drawn from the 4 settings.

Rectangles of different sizes were used in an attempt to increase the automated randomization of the data. The advantage to this method lies in its ability to show more reliably how well the predictions perform, both in higher terrain and on the coast/in areas of little-to-no bedrock. We also wanted to determine if using more rectangles increased the quality of the results, and if using smaller rectangles was in fact sufficient. A further goal was to determine whether rectangles with little-to-no bedrock data would be less useful to the machine learning than those in >70% bedrock areas.

3.1.2. *Predictors - terrain derivatives from DEM and Landsat 8*

We compared two sets of predictors: (1) a cloud-free Landsat 8 composite scene and (2) derivatives from a DEM. Our first predictor was a three-month cloud-free composite cloud-based Landsat 8 dataset providing seasonal coverage of the global landmass (Catalog, 2021; NASA, 2022). We used both optical and thermal bands in our approach: (a) optical bands with 30 m resolution, B1 (Coastal aerosol), B2 (Blue), B3 (Green), B4 (Red), B5 (Near Infrared), B6 (SWIR 1), and B7 (SWIR 2) and (b) thermal bands with 100 m resolution, B10 (Thermal Infrared 1) and B11 (Thermal Infrared 2) (USGS, 2021b). The 100 m resolution bands were resampled to 30 m. We tested the relevance and usefulness of this predictor in our area on its own, to test the applicability of our model using easily accessible cloud-based satellite data. Our second predictor, the DEM and its derivatives, had a resolution (grid size) of 10 × 10 m and an accuracy of ± 2–3 m standard deviation in height (Kartverket, 2013). We calculated the terrain derivatives shown in Table 1 using QGIS (a free and open-source desktop GIS) and uploaded the set of layers to the Google Earth Engine (GEE) for computing. Seven different terrain derivatives were created based on the 10 m DEM

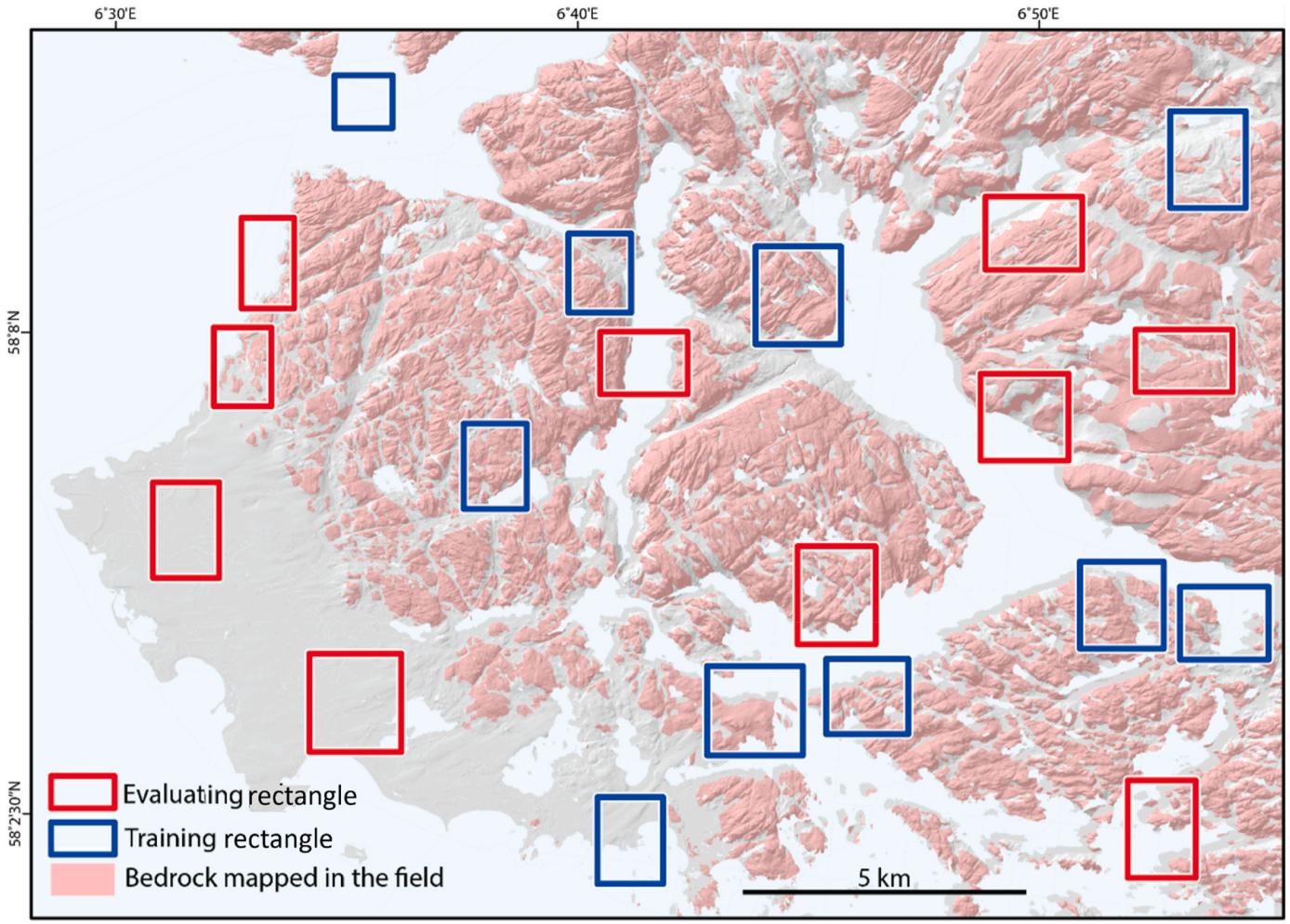

Fig. 2. Bare rock (and thin organic soil on bedrock) ground truth dataset based on a 1:50 000 quaternary map (Fredin et al., 2015) with training (blue) and evaluating (red) rectangles superposed. Example of 1 case (Setting 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(Table 1). We tested these separately, but the best results were achieved when using a combination of all of them. All training and evaluating rectangles, and both sets of predictors, were uploaded, and stored in the cloud for further analysis and modeling.

3.2. Method – cloud computing (sampling, training, prediction, and visualization)

As mentioned, we chose to use a method relatively new in geology, a type of machine learning and artificial intelligence called *deep learning*. In contrast to traditional supervised classification in machine learning which can use training points as ground truth, in our case, deep learning uses image patches (for example, 128 × 128 pixels) as its training basis. Each image patch must contain predictor layers (Fig. 3) and a ground truth layer (bedrock) which shows how the deep learning algorithm should sort each pixel into two classes: presence (bedrock) or absence (sediment). Here, we provide a complete field design for collecting data, where samples for training were produced (section 3.2.2), and finally how predictions were generated (section 3.2.3). The deep learning algorithm itself (U-Net) is written in Python in Google Colab and the model is trained in the computing cloud (section 3.2.3).

3.2.1. Cloud computing and Google Earth Engine (GEE)

Recently, GEE is commonly used in various research communities (Arruda et al., 2021; Liang et al., 2020; Shaharum et al., 2020). Its ability to analyze remote sensing data and to provide high-performance computing resources for processing large geospatial datasets online, without downloading and processing the imagery locally, makes it easy to access and use for everybody (Li et al., 2019). The GEE platform provides a user-friendly environment and cloud-based processing of free and available data using the power of thousands of computers located in Google data centers. We used GEE as well as Colab to store, customize, and export the ground truth data for cloud-based deep learning modeling. Colab allowed us to write and execute Python code through the browser connected to GEE. Deep learning is an important element of data science, which includes statistics and predictive modeling and consists of three steps: training, prediction, and visualization, as illustrated in Fig. 4.

3.2.2. Sampling - training and evaluating

A general recommendation is to use, on average, at least 1000 samples per class when training a deep learning algorithm (Theophano, 2019). In our case, we trained the model to predict one class only, that is, bedrock. What was not predicted as bedrock was marked as undefined but contained mostly sediments. Therefore, to obtain a reliable and representative number of samples with which to train and evaluate our model, the algorithm was directed to create a total of 2000 sample image patches of bedrock (where bedrock was present) of 128 × 128-pixels for every training and evaluating rectangle in each test. All image patches contained one layer with the feature to be predicted (i.e., bedrock)

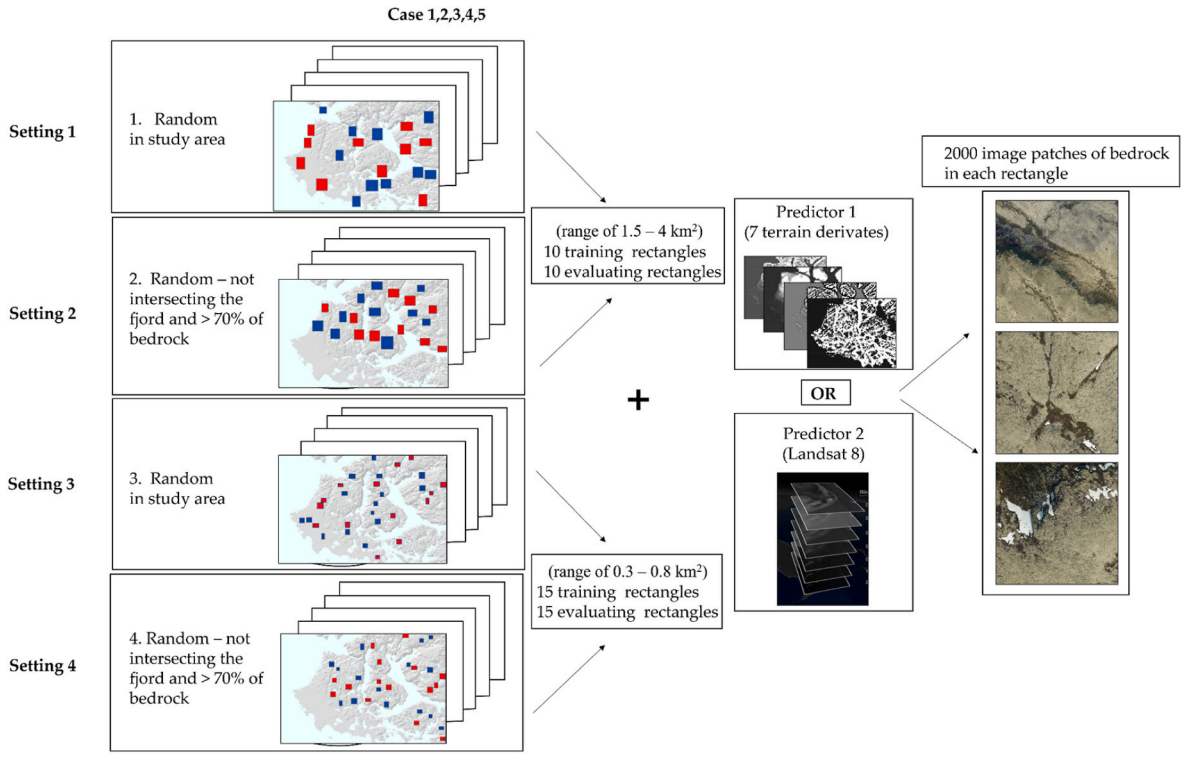

Fig. 3. Sampling strategy: four different settings were applied to produce five sets (cases) each of training and evaluating rectangles, combined with one of two predictors, creating 2000 image patches of bedrock for further calculation from each rectangle where possible (see 3.2.2 for more details on image patches).

Table 1
Terrain derivatives used for digital bedrock extraction; Relative relief in 12-pixel neighborhood; TPI3 and TPI9 (Topographic position index in 3- and 9-pixel neighborhoods).

Variable/terrain derivative	Definition
Slope	angle of inclination to the horizontal plane
Elevation	elevation above the sea (geodetic datum)
Slope_sum	sum of slope values in the 12-pixel neighborhood
Relative relief	relative difference in elevation between a morphological feature and the features surrounding it
Valley depth	difference between the elevation and an interpolated ridge level
TPI3 (Topographic position index)	difference between a central pixel and the mean of its surrounding cells by 3 pixels
TPI9	difference between a central pixel and the mean of its surrounding cells by 9 pixels

stacked upon all the other layers in the predictor stack (i.e., the stack with DEM derivatives or the Landsat 8 stack). It resulted in a total of 20 000 samples for settings (1) and (2) and resulting in 30 000 samples for settings (3) and (4) (Fig. 3). The image patches from each rectangle were merged into a single export file and stored in Google Cloud Storage as TFRecord files, which contained patches of pixel values in each record.

3.2.3. Training the model, prediction, and visualization

The chosen model is a scaled-down version of a deep learning architecture called U-Net with Keras implementation. Keras is open-source software that provides a python interface for artificial neural networks and for the TensorFlow 2 library[1] (Huang and Le, 2021). U-Net is a convolutional network architecture with a unique U-shaped architecture. The network is based on a fully convolutional neural network (Shelhamer et al., 2017). The input image was propagated through the entire path of the U-Net architecture, resulting in a classified map (Educative Answers Team, n.d.). The goal of a semantic segmentation is to label each pixel of the ground truth image with the class that represents a specific object (Bihani et al., 2022; Du et al., 2021; Zhuang et al., 2019). U-Net has proven to be very powerful in scenarios with limited data, having no restrictions regarding ground truth image size (Isensee et al., 2021; Yadav, 2017; Zou et al., 2021; Zunair and Hamza, n.d.). Our model was trained with 15 epochs to reach a stable output performance and to avoid overfitting (Brownlee, 2022). An epoch in machine learning is one complete pass of the training dataset through the algorithm and can be compared to a "for-loop" common in programming. After 15 iterations, our model was trained and was applied to the whole study area (Brownlee, 2022).

3.2.4. Performance evaluation

The resulting predictions were evaluated quantitatively, as follows. The bedrock vector layer (ground truth) was converted to a binary raster to compare it with the automated bedrock predictions. In order to evaluate the performance of the proposed model and to compare the outcome predictions, based on different parameters and different predictors (Landsat 8 vs. DEM derived terrain derivatives), a map of confusion matrix values was created, showing true positive (TP), false positive (FP), false negative (FN) and true negative (TN) values. From these, the performance metrics precision, recall, F1-score, accuracy, and Matthew's correlation coefficient (MCC) scores were calculated (Table 2). The MCC score is considered to be the most appropriate metric for comparing the results which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (TP,FP,FN,TN) (Chicco and Jurman, 2020). For a binary model, the MCC gave a score between 0 and 1; with 0 indicating a model with no correlation (random predictions) and 1 indicating a perfect correlation (all correct predictions).

[1] TensorFlow 2 is a free and open-source software library developed particularly for the training of ML algorithms.

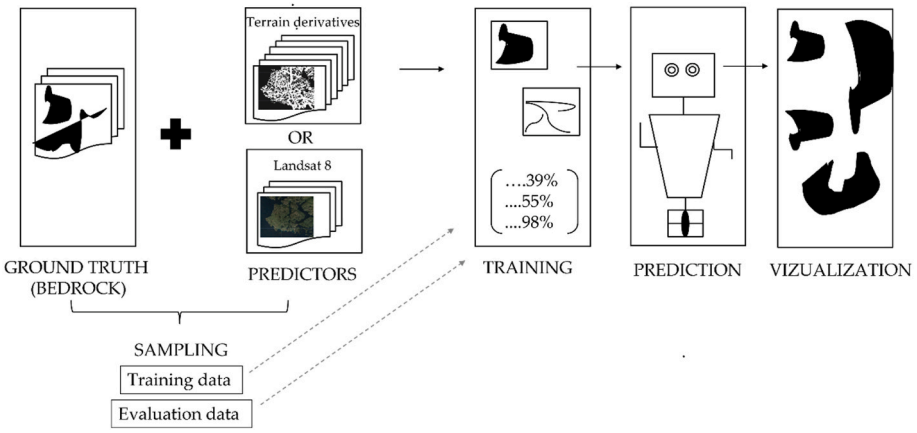

Fig. 4. Workflow applied in GEE: (1) sampling, (2) training and evaluating, (3) training the model, resulting in metrics on accuracy, (4) prediction, and (5) visualization in GEE.

Table 2
Equations for performance evaluation metrics from confusion matrix values.

Metric	Formula
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$\frac{2TP}{2TP+FP+FN}$
Accuracy	$\frac{(\mathrm{TP}+\mathrm{TN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})}$
MCC	$\frac{TP\times TN-FP\times FN}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$

4. Results

The performance of the U-Net model applied in a cloud-based environment is presented in this section. First, we aimed to evaluate whether the inclusion of additional parameters, such as the definition of a minimal amount of bedrock in both training and evaluating areas (Setting 1 and 3), influenced the resulting prediction. We can clearly see that there are only minimal differences when applying this parameter. The statistics presented in Table 3 show the performances when applying different settings. When comparing the size and amount of training and evaluating rectangles (Settings 1 and 2 vs. Settings 3 and 4), we can observe lower performances for both Landsat 8 and DEM predictors when the size of the rectangles is decreased, and the number increased (Table 3). Setting 1 for Landsat 8 shows an MCC value of 40%, while Setting 2 shows an MCC value of 27%. When using smaller rectangles, the MCC scores only 19% for Setting 3 and 21% for Setting 4. Different performances are also clear when looking at the statistics for DEM derivatives that show MCC score differences between Setting 1 and 2 of 74 vs 71%, while for Setting 3 and 4 MCC scores are equal to 63 and 68% respectfully.

Table 3
Performance metrics for bedrock detection using four sampling strategies (settings) for Landsat 8 (L8) and DEM terrain derivatives as predictors, showing F1-score, MCC, accuracy, prediction, and recall. The Matthews correlation coefficient (MCC) shows the most reliable statistical rate. It produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives) (Chicco and Jurman, 2020).

	Setting 1		Setting 2		Setting 3		Setting 4	
	L8	DEM	L8	DEM	L8	DEM	L8	DEM
F1-score	0,6	0,84	0,39	0,83	0,2	0,81	0,22	0,77
MCC	**0,4**	**0,74**	**0,27**	**0,71**	**0,19**	**0,68**	**0,21**	**0,63**
accuracy	0,73	0,87	0,67	0,85	0,67	0,85	0,67	0,83
precision	0,64	0,81	0,71	0,76	0,72	0,81	0,76	0,82
recall	0,56	0,88	0,26	0,9	0,12	0,81	0,13	0,72

All the 20 predictions (see Fig. 3) using Landsat 8 images showed low accuracy performance. Fig. 5 depicts one of the most representative 20 cases and shows the geographical distribution of the false positive and false negative errors. True positive results are shown in green, false positive in red and false negatives in blue. The black boxes highlight the areas of the poorest predictive performance, namely large parts of the coastal areas, and the area around the airport. These areas have been in fact erroneously predicted as bedrock. The low predictive performances of Landsat 8 are confirmed also by the F1 score values presented in Table 3. Setting 4 and 5 show values around 20%.

Contrary to the poor prediction performances obtained by using Landsat 8 data, when using the same ground truth data in combination with the terrain derivatives, a spatial match with higher statistical values is obtained (see Table 3). In these cases, when different parameters have been introduced, the variance in the predictions is not statistically significant. Both visually and statistically the predictions performed similarly for all four settings. For this reason, we decided to show only the prediction result with the best result from each setting (Table 3), and the confusion map of the best result (based on evaluating and training accuracy values) among the 20 predictions (see Fig. 6). The errors in all the predictions consist mostly of false positives, represented by the red patches in Fig. 6. The accuracy assessment of the predictions is performed by comparing our prediction results with ground truth, as shown in Tables 3 and in Fig. 6.

5. Discussion

This study has presented an automated approach for delineating bedrock from sediments as an alternative to traditional methods, based on the application of freely available cloud computing, and deep learning techniques (Fully Convolutional Neural Networks) using a U-Net model applied in Google Collaboratory (Colab). Our approach is novel in three ways. First, we tested the potential of cloud computing and using the Google Earth Engine (GEE) interface on bedrock. Second, we compared the performance of freely available cloud-based Landsat 8 images versus a set of explanatory variables generated from terrain parameters uploaded to the cloud. Third, we applied a deep learning approach and U-Net architecture for delineating bedrock outcrops, using sediment maps as ground truth data (Fredin et al., 2015).

When using an automatic method to train a new model, there will always be slight variations in the results when compared to a traditional manual approach on the exact same ground truth, because of the inner

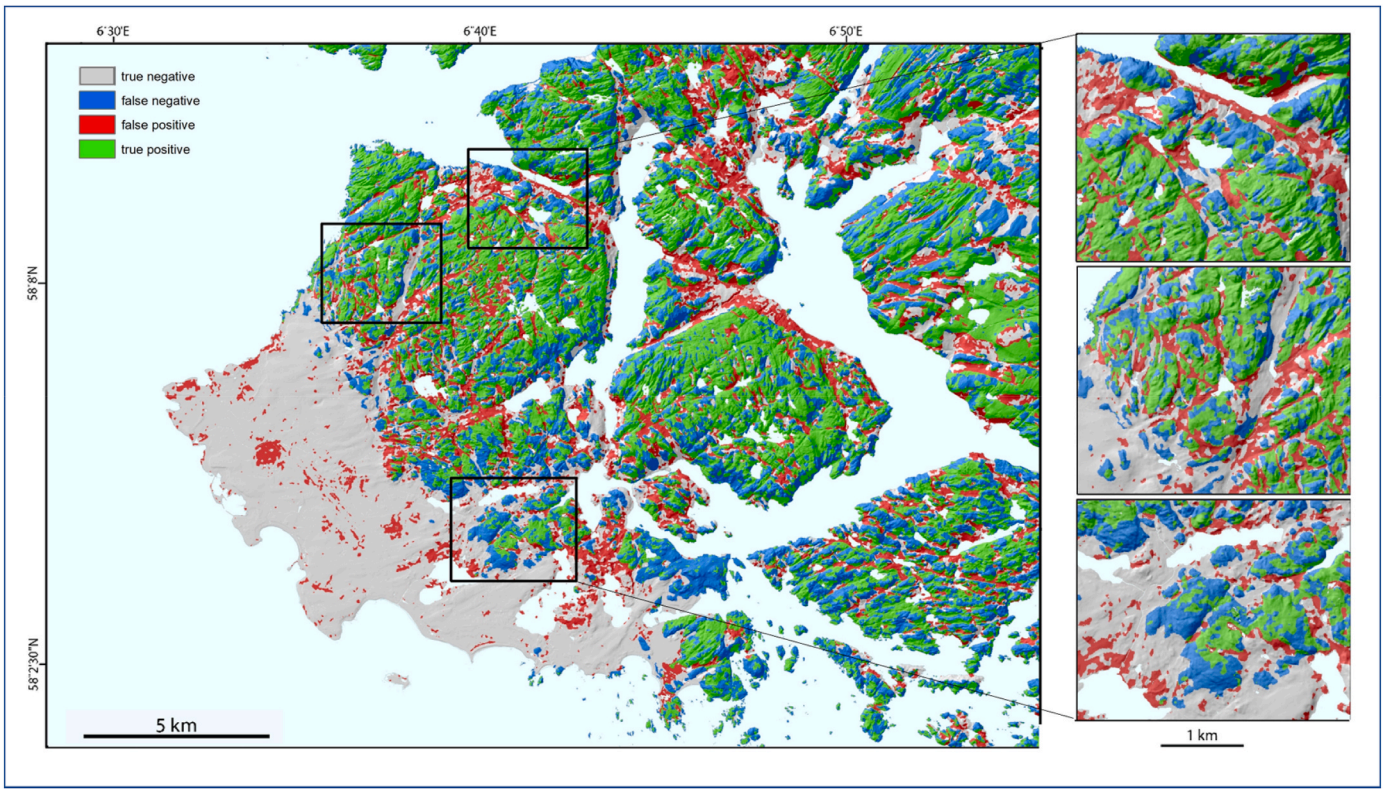

Fig. 5. Landsat 8-based prediction: The map shows the spatial distribution of the modelled prediction of bedrock areas in comparison with the ground truth bedrock areas from the geological map (Setting 1, case 4 with 99% training match and 86% evaluating) visualized as true positive (green), false positive (red), false negative (blue) and true negative (grey) over the entire study area (left) and with the inset areas (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

nature of deep learning procedures (Swedberg, 2016). It is a well-known characteristic that deep learning techniques introduce errors differentiating each new prediction from the previous one, when new training samples are introduced to the algorithm. All the prediction outputs are evaluated for the accuracy of the model's performance and guide the choice of ideal predictors. We tested, both visually and statistically, how changing and combining various settings for training and evaluating rectangles increases the quality of the predictions.

Our results show that there are only small variations in model predictions when using training rectangles of varying sizes, and with more or less bedrock composition, and that by using the 10 m DEM together with terrain derivatives we obtain excellent and robust mapping results when tested against the manually mapped validation dataset. We were expecting that ideal cases would possibly produce more reliable predictions. However, we observed that no additional parameters were required to increase the quality of the final prediction, but rather that the model creates enough well-defined samples from the available data when mapping bedrock areas in all settings. Final mapping results were not, however, satisfactory when using the Landsat 8 dataset alone in such variable terrain, despite Landsat 8 showing good potential for application in exposed, mountainous areas in other studies. (DiBiase et al., 2012; Milodowski et al., 2015). Scarpone et al. (2017) claimed that using high-resolution satellite images, such as Landsat 8, could improve predictions for isolating exposed bedrock from other sediment. We believe the difference in success may relate to two factors; first, the method used in Scarpone et al. (2017) is based on a random forest model while we are using a deep learning approach, and second, our study area is a coastal area with vegetated peaks and small elevation differences, while Scarpone et al. (2017) tested a machine-learning approach with apparently greater success in the southern part of British Colombia where the elevation difference is much greater (up to 2300 m), with jagged peaks and ridges dominating the landscape. We conclude then that Landsat 8 is not suitable on its own as a predictor in our type of landscape with a lot of forest, but we assume that satellite images in combination with terrain derivatives can improve prediction results in various areas, including those with more mountains.

Although cloud computing with the use of GEE is time consuming because of data preparation and training, the resources needed are small compared to those needed for traditional geological mapping of bedrock. The predictions of our study, supported by ground truth, found clear evidence for the possibility to automatically delineate bedrock from sediments for more efficient future production of geological maps. Our model creates predictions with a high success rate (precision up to 82% and recall up to 88% as shown in Table 3). The positive predictions therefore exceed the negatives and the limitations of computer-based techniques. These predictions give geologists a good foundation map of an area, which allows efficient planning of where field work is necessary to map areas of uncertainty, instead of mapping entire areas. This makes it possible to efficiently map larger areas within the same time frame.

The ground truth dataset we used is a complete map of sediments at a scale of 1:50 000. To automate the classification of bedrock, the resolution of the ground truth data determines to what extent it is possible to create maps of higher quality, as it is not possible to create predictions with a higher resolution than that provided by the ground truth data. While higher resolution sediment maps would be ideal, because current geological maps are produced at this scale, it is very helpful for geologists to have predictions adapted to exactly this scale for improving the quality and resolution of the mapping of the entire country. For the future publication, we have tested our model, as a pre-trained model, over other areas of Norway characterized by very similar geophysical conditions and have obtained solid and promising outputs. However, these preliminary tests have found that our model has some difficulties, in its current state, over inland areas with very different terrain, and

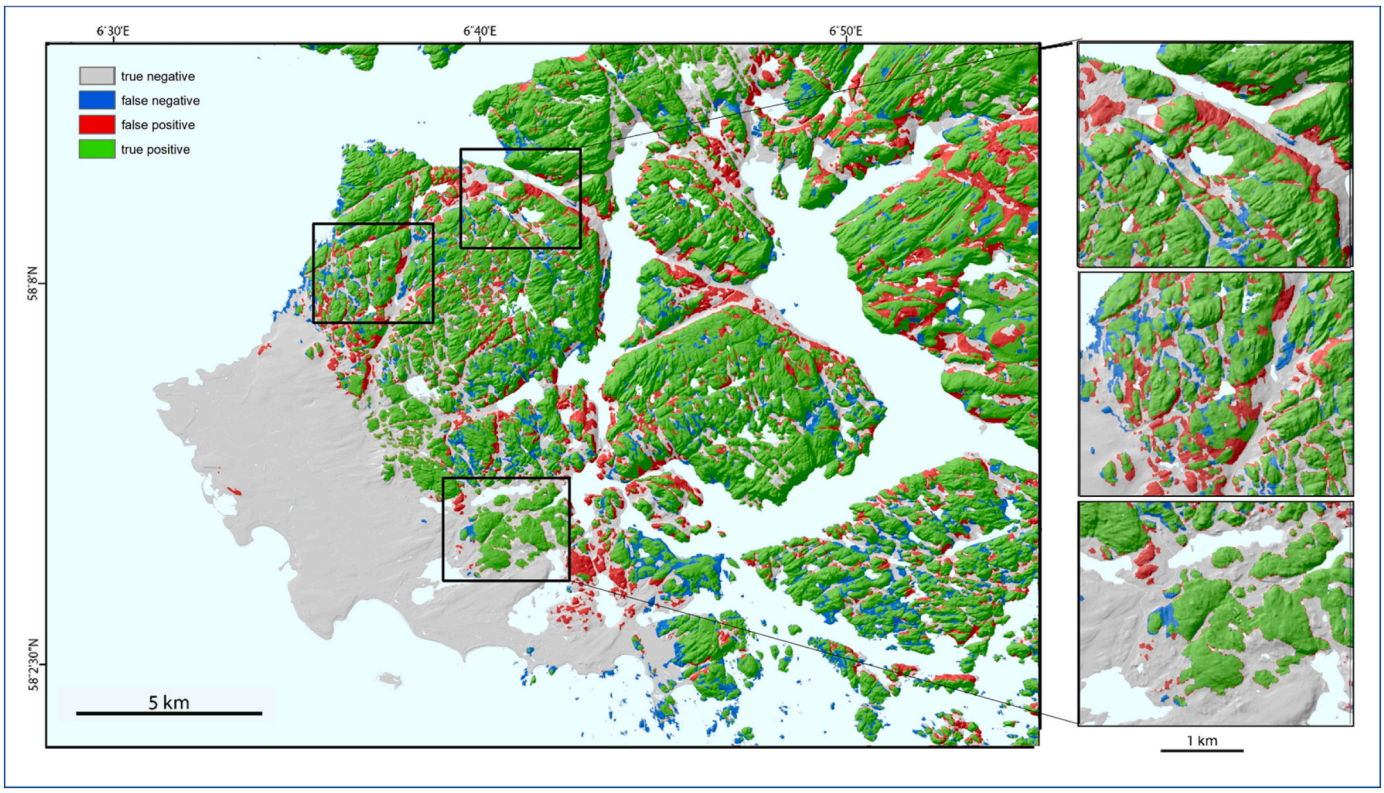

Fig. 6. Spatial distribution of the DEM terrain derivatives-based prediction. The map shows the modelled prediction of bedrock areas in comparison with the ground truth bedrock areas from the geological map (Setting 1, case 5 with 97% training match and 83% evaluating) visualized as true positive (green), false positive (red), false negative (blue) and true negative (grey) over the entire study area (left) and with the inset areas (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

further training on different ground truth data would be advantageous to increase the applicability of the model.

The computational requirements for training the samples and running the model with the use of deep learning techniques are not in the capacities of every computer. Our aim is to make the delineation of bedrock available for anybody, we therefore applied and tested the feasibility of using exclusively cloud-based solutions (GEE, Google Storage and Colab). Colab provides relatively cheap computing resources and a user-friendly interface (Almeida et al., 2021; Liang et al., 2020; Prasai et al., 2021). It is known that training time for deep learning algorithms takes many hours or even months (Gupta et al., 2015; Miranda and Von Zuben, 2015). However, with the use of cloud solutions and GPUs provided by GEE, all the procedures can be done in a reasonably efficient way that is available to everyone. GEE also provides support such as the cloud availability of the large datasets and the storage needed for the calculations, samples and predictions. This process was functional, but we found that deep learning with the usage of cloud computing has some limitations. Calculations being able to run continuously and the ability to create and export predictions for larger areas would be an interesting improvement. Also, the overall usage limits, timeout periods, and GPU types vary over time and Google does not guarantee unlimited usage of GEE, Google Storage and Colab (Google, 2021). The user is therefore limited by the abilities and restrictions of these resources. This may have influenced the final predictions in this study and will be an important consideration in future such studies.

6. Conclusions

Development in computing, deep learning algorithms and increased availability of high-quality and free data have the potential to automate many mapping problems in Earth sciences. Although Fully Convolutional Neural Networks (FCNN) have recently been used in various fields, such asautonomous driving (Aladem and Rawashdeh, 2021) and medical-image processing (Tajbakhsh et al., 2016), its application in geology for delineating bedrock from sediments has not yet been investigated thoroughly. The aim of our study was to investigate if the deep learning FCNN model could potentially be a simple and efficient method to automatically separate bedrock from sediments for the production of geological maps. The proposed framework is based on a U-Net architecture and cloud-computing deep learning predictive algorithms. Moreover, the open code and cloud-computation makes our framework accessible for anyone who wants to test the predictions of the algorithm with the use of their own data over their areas of study.

Our deep learning approach proved to differentiate bedrock from sediments much more quickly than manual mapping and with good precision. We observed that variable parameters to produce training and evaluating data have little influence on the quality of the final prediction, although Landsat 8 data alone was insufficient for predictions on variable terrain (despite its proven utility in exposed mountainous areas). We can however create high-value predictions based on a DEM and its derivatives with 10 m resolution. Using DEM and its derivatives has important implications for cost-effective geological mapping: fieldwork would be more efficient because it could be targeted towards the areas with uncertain predictions (e. g. those with abundant sediment cover), and mapping can potentially be done over larger areas within the same resource restrictions. It is also possible that operator or mapper bias might be reduced through the use of deep learning mapping.

Code availability section

The source code is available for download here: https://github.com/alexandra-jarna/EarthEngine/blob/9ea0d173ad32f1ab007b7a5c1ac61b33a51f9ea0/u-net_bedrock_sediment_calassification.ipynb

Program language

Python

Software required

Google Earth Engine, Google Colab, data preparation (ArcGIS Pro/QGIS).

Authorship contribution statement

Author 1, Author 2, Author 3, Author 4, and Author 5 initiated and designed the study together. Author 1 and Author 2 adapted deep learning concepts for this application. Author 1 performed coding, ran the model, and plotted figures. Author 3 contributed to the results and discussion after the peer review process. Author 4 provided ground truth mapping. All authors contributed to figure work and writing.

CRediT authorship contribution statement

Alexandra Jarna Ganerød: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Vegar Bakkestuen:** Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. **Martina Calovi:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Ola Fredin:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Jan Ketil Rød:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgments

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. This manuscript was benefited by proofreading of Danielle Robert. We are thankful to all reviewers for usefull advises and suggestions to improve message, strengths, and values of the manuscript.

References

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A.E., Arshad, H., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938.

Aladem, M., Rawashdeh, S.A., 2021. A single-stream segmentation and depth prediction CNN for autonomous driving. IEEE Intell. Syst. 36, 79–85. https://doi.org/10.1109/MIS.2020.2993266.

Alavi, A.H., Gandomi, A.H., Lary, D.J., 2016. Progress of machine learning in geosciences: preface. Geosci. Front. 7, 1–2. https://doi.org/10.1016/j.gsf.2015.10.006.

Almeida, L.P., Efraim de Oliveira, I., Lyra, R., Scaranto Dazzi, R.L., Martins, V.G., Henrique da Fontoura Klein, A., 2021. Coastal analyst system from space imagery engine (CASSIE): shoreline management module. Environ. Model. Software 140, 105033. https://doi.org/10.1016/j.envsoft.2021.105033.

Arruda, V.L.S., Piontekowski, V.J., Alencar, A., Pereira, R.S., Matricardi, E.A.T., 2021. An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna. Remote Sens. Appl.: Soci. Environ. 22, 100472 https://doi.org/10.1016/j.rsase.2021.100472.

Baraniuk, R., Donoho, D., Gavish, M., 2020. The science of deep learning. Proc. Natl. Acad. Sci. U. S. A 117, 30029–30032. https://doi.org/10.1073/pnas.2020596117.

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A., 2018. Multi-scale digital soil mapping with deep learning. Sci. Rep. 8, 2–10. https://doi.org/10.1038/s41598-018-33516-6.

Bihani, A., Daigle, H., Santos, J.E., Landry, C., Prodanović, M., Milliken, K., 2022. MudrockNet: Semantic segmentation of mudrock SEM images through deep learning. Computers and Geosciences 158. https://doi.org/10.1016/j.cageo.2021.104952.

Brownlee, J., 2022. Machine learning mastery with Python. Machine Learning Mastery.

Catalog, E.E.D., 2021. Landsat 8 [WWW Document]. URL. https://developers.google.com/earth-engine/datasets/catalog/landsat-8.

Caté, A., Perozzi, L., Gloaguen, E., Blouin, M., 2017. Machine learning as a tool for geologists. Lead. Edge 36, 215–219. https://doi.org/10.1190/tle36030215.1.

Chen, X.W., Lin, X., 2014. Big data deep learning: challenges and perspectives. IEEE Access 2, 514–525. https://doi.org/10.1109/ACCESS.2014.2325029.

Chicco, D., Jurman, G., 2020. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation 1–13.

Dargan, S., Kumar, M., Ayyagari, M.R., Kumar, G., 2020. A survey of deep learning and its applications: a new paradigm to machine learning. Arch. Comput. Methods Eng. 27, 1071–1092. https://doi.org/10.1007/s11831-019-09344-w.

DiBiase, R.A., Heimsath, A.M., Whipple, K.X., 2012. Hillslope response to tectonic forcing in threshold landscapes. Earth Surf. Process. Landforms 37, 855–865. https://doi.org/10.1002/esp.3205.

Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., Darrell, T., 2017. Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans. Pattern Anal. Mach. Intell. 39, 677–691. https://doi.org/10.1109/TPAMI.2016.2599174.

Drăguţ, L., Eisank, C., Fraser, O.L., Bailey, S.W., Ducey, M.J., McGuire, K.J., MacMillan, R.A., Pettapiece, W.W., Nolan, S.C., Goddard, T.W., Webster, T.L., Murphy, J.B., Gosse, J.C., Spooner, I., Weidner, L., Walton, G., Kromer, R., 2006. Predictive modeling of bedrock outcrops and associated shallow soil in upland glaciated landscapes. Geomorphology 376, 114495. https://doi.org/10.1016/j.enggeo.2019.105326.

Drăgut, L., Eisank, C., Strasser, T., 2011. Local variance for multi-scale analysis in geomorphometry. Geomorphology 130, 162–172. https://doi.org/10.1016/j.geomorph.2011.03.011.

Du, B., Zhao, Z., Hu, X., Wu, G., Han, L., Sun, L., Gao, Q., 2021. Landslide susceptibility prediction based on image semantic segmentation. Computers and Geosciences 155, 104860. https://doi.org/10.1016/j.cageo.2021.104860.

Falkum, T., 1982. Beskrivelse Til Geologisk Kart I Norge - 1: 250 000 -Mandal.

Fredin, O., Romundset, A., Riiber, K., 2015. Kvartærgeologisk Kart (Quaternary Geological Map) Farsund Og Hidra 1311-2 Og 1311-3 M 1:50 000.

Google, 2021. Colaboratory [WWW Document]. URL. https://research.google.com/colaboratory/faq.html#usage-limits.

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190. https://doi.org/10.1016/j.geoderma.2007.11.004.

Gupta, S., Agrawal, A., Gopalakrishnan, K., Narayanan, P., 2015. Deep learning with limited numerical precision. 32nd international conference on machine learning. ICML 3, 1737–1746, 2015.

Harris, J., Rainbird, R.H., Canada, N.R., Behnia, P., Canada, N.R., 2014. Remote Predictive Mapping 6: A Comparison of Different Remotely Sensed Data for Classifying Bedrock Types in Canada's Arctic: Application of the Robust Classification Method and Ra. https://doi.org/10.12789/geocanj.2014.41.062.

Harris, J.R., Schetselaar, E., Behnia, P., 2012. Remote Predictive Mapping : an Approach for the Geological Remote Predictive Mapping : an Approach for the Geological Mapping of Canada ' S Arctic. https://doi.org/10.5772/25475.

Huang, S.-C., Le, T.-H., 2021. Introduction to TensorFlow 2. In: Principles and Labs for Deep Learning. Elsevier, pp. 1–26. https://doi.org/10.1016/B978-0-323-90198-7.00014-8.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2019. Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. 31, 1544–1554. https://doi.org/10.1109/TKDE.2018.2861006.

Kartverket, 2018. FKB Vann [WWW Document]. URL. https://kartkatalog.geonorge.no/metadata/fkb-vann/595e47d9-d201-479c-a77d-cbc1f573a76b.

Kartverket, 2013. DTM 10 Terrengmodell (UTM32) [WWW Document]. URL. https://kartkatalog.geonorge.no/metadata/dtm-10-terrengmodell-utm32/fd851873-f363-46f9-9fc6-bb1b403575df.

Kerry, R., Oliver, M.A., 2011. Soil geomorphology: identifying relations between the scale of spatial variation and soil processes using the variogram. Geomorphology 130, 40–54. https://doi.org/10.1016/j.geomorph.2010.10.002.

Li, H., Wan, W., Fang, Y., Zhu, S., Chen, X., Liu, B., Hong, Y., 2019. A Google Earth Engine-enabled software for efficiently generating high-quality user-ready Landsat mosaic images. Environ. Model. Software 112, 16–22. https://doi.org/10.1016/j.envsoft.2018.11.004.

Liang, J., Xie, Y., Sha, Z., Zhou, A., 2020. Modeling urban growth sustainability in the cloud by augmenting Google Earth Engine (GEE). Comput. Environ. Urban Syst. 84, 101542 https://doi.org/10.1016/j.compenvurbsys.2020.101542.

Lisle, R.J., Brabham, P., Barnes, J.W., 2011. Basic Geological Mapping, fifth ed., fifth ed. John Wiley.

MacMillan, R.A., Jones, R.K., McNabb, D.H., 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). Comput. Environ. Urban Syst. 28, 175–200. https://doi.org/10.1016/S0198-9715(03)00019-X.

Milodowski, D.T., Mudd, S.M., Mitchard, E.T.A., 2015. Topographic roughness as a signature of the emergence of bedrock in eroding landscapes. Earth Surf. Dyn. 3, 483–499. https://doi.org/10.5194/esurf-3-483-2015.

Miranda, C.S., Von Zuben, F.J., 2015. Reducing the Training Time of Neural Networks by Partitioning 1–10.

NASA, 2022. Landsat 8 mission details [WWW Document]. Landsat Science. URL https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-mission-details/#:~:text=The Landsat 8 satellite payload. . and 15 meters (panchromatic).

Nordgulen, Ø., 2020. BEDROCK GEOLOGY [WWW Document]. URL. https://www.ngu.no/en/topic/bedrock-geology-0.

Prasai, R., Schwertner, T.W., Mainali, K., Mathewson, H., Kafley, H., Thapa, S., Adhikari, D., Medley, P., Drake, J., 2021. Application of Google earth engine python API and NAIP imagery for land use and land cover classification: a case study in Florida, USA. Ecol. Inf. 66, 101474 https://doi.org/10.1016/j.ecoinf.2021.101474.

Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. SN Computer Sci. 2, 1–21. https://doi.org/10.1007/s42979-021-00592-x.

Scarpone, C., Schmidt, M.G., Bulmer, C.E., Knudby, A., 2017. Semi-automated classification of exposed bedrock cover in British Columbia's Southern Mountains using a Random Forest approach. Geomorphology 285, 214–224. https://doi.org/10.1016/j.geomorph.2017.02.013.

Sejrup, H.P., Larsen, E., Haflidason, H., Berstad, I.M., Hjelstuen, B.O., Jonsdottir, H.E., King, E.L., Landvik, J., Longva, O., Nygard, A., Ottesen, D., Raunholm, S., Rise, L., Stalsberg, K., 2003. Configuration, history and impact of the Norwegian channel ice stream. Boreas 32, 18–36. https://doi.org/10.1080/03009480310001029.

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V., Peters, A., 2020. A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl. Base Syst. 194 https://doi.org/10.1016/j.knosys.2020.105596.

Shaharum, N.S.N., Shafri, H.Z.M., Ghani, W.A.W.A.K., Samsatli, S., Al-Habshi, M.M.A., Yusuf, B., 2020. Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. Remote Sens. Appl.: Soci. Environ. 17, 100287 https://doi.org/10.1016/j.rsase.2020.100287.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 640–651. https://doi.org/10.1109/TPAMI.2016.2572683.

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., Oza, H., 2021. Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Res. 6, 379–391. https://doi.org/10.1016/j.ptlrs.2021.05.009.

Swedberg, R., 2016. Before theory comes theorizing or how to make social science more interesting. Br. J. Sociol. 67, 5–22. https://doi.org/10.1111/1468-4446.12184.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imag. 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302.

Theophano, M., 2019. How Do You Know You Have Enough Training Data? Towardsdatascience.

USGS, 2021a. Landsat 8 [WWW Document]. URL. https://www.usgs.gov/landsat-missions/landsat-8.

USGS, 2021b. What are the band designations for the Landsat satellites? [WWW Document]. URL. https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products.

Xiong, Y., Zuo, R., Carranza, E.J.M., 2018. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. Ore Geol. Rev. 102, 811–817. https://doi.org/10.1016/j.oregeorev.2018.10.006.

Zhang, Liangpei, Zhang, Lefei, Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Magazine 4, 22–40. https://doi.org/10.1109/MGRS.2016.2540798.

Zhou, J., Huang, B., Yan, Z., Bünzli, J.C.G., 2019. Emerging role of machine learning in light-matter interaction. Light Sci. Appl. 8 https://doi.org/10.1038/s41377-019-0192-4.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Magazine 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307.

Zhuang, J., Yang, J., Gu, L., Dvornek, N.. Shelfnet for fast semantic segmentation. https://doi.org/10.1109/ICCVW.2019.00113.