Doctoral theses at NTNU, 2024:62

ALEXANDRA JARNA GANERØD

Between data-science and geosciences: exploring the use of deep learning in automated mapping

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

ALEXANDRA JARNA GANERØD

Between data-science and geosciences: exploring the use of deep learning in automated mapping

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2024

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Geography

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Social and Educational Sciences
Department of Geography

ISBN 978-82-326-7724-5 (printed ver.)
ISBN 978-82-326-7723-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:62

Printed by NTNU Grafisk senter

PROMOTERS

DR. JAN KETIL RØD
DEPARTMENT OF GEOGRAPHY
NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

DR. MARTINA CALOVI
DEPARTMENT OF GEOGRAPHY
NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

DR. OLA FREDIN
DEPARTMENT OF GEOSCIENCE AND PETROLEUM
NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

TABLE OF CONTENTS

Preface

This thesis is submitted in fulfilment of the requirements for the degree of Philosophiae Doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU). The work has been carried out at, and funded by, the Department of Geography (IGE) with Professor Jan Ketil Rød as the main supervisor and Associate Professor Martina Calovi and Professor Ola Fredin as co-supervisors. The Department for Geography funded the three-month research stay at the Bruno Kessler Foundation FBK Trento, and DEEP: Norwegian Research School for Dynamics and Evolution of Earth and Planets funded the one-month research stay at the UC Berkeley at the Department of Astronomy and Statistics.

Acknowledgments

I sincerely thank my supervisors, especially my main supervisor Jan Ketil Rød, who always stood next to me, and co-supervisors Martina Calovi for all kinds of support, and Ola Fredin for responding to my endless geological questions.

Thanks to Vegar Bakkestuen, who helped start the collaboration with NINA, introduced me to cloud computing and collaborated in finalizing two scientific articles. Martina Calovi, who created the bridge with the Bruno Kessler Foundation (Trento, Italy), and made it possible to visit it for three months, which started a great collaboration with the Data Science for Industry and Physics research unit. Huge thanks to Gabriele Franch and Marco Cristoforetti for helping me to further develop my work and skills as an experienced data scientist.

DEEP: Norwegian Research School for Dynamics and Evolution of Earth and Planets allowed me to join three great Ph.D. courses, a summer school in Iceland, and gave the opportunity for a three-week exchange visit at UC Berkeley. During all these activities, I met many incredible people – I wish I could have joined it much earlier. Thanks to Gabriela Spakman and her excellent teaching week in Oslo on DL & ArcPy possibilities which led me to publish two articles thus far.

Mathew Reeve helped me along with writing the very first article while taking his course *Writing with the Results* (DEEP). He always had time to meet with me and help structure my thoughts regarding my Ph.D. thesis's superstructure. Thanks to Danielle Robert for English proofreading help with two articles and the superstructure. Many thanks to my friend and colleague Erin Lindsay for her support during the Ph.D., for sharing great ideas, and for collaborating on the landslide topic (Article III, IV, V). Thanks also to Mikis van Boeckel, with whom we efficiently worked the entire 2023, but mainly for the incredible sprint in summer 2023 to finish Article VI about the mapping of gullies. Thanks to the Finse summer ML school and all the participants and connections I made with UC Berkeley and Ellie Abrahams. The entire NGU family during this process, but mostly my GIS team, helped me grow personally and professionally and supported me all the time: Nicole Beaten and Fredrik Høgaas provided mental support on the way to the finish line; Tone that made it possible to

start with this Ph.D.; Frank Haugan always supported me from my first day at NGU in 2013 and was a great reference to get the Ph.D. position; and Per Arne Stavnås, who selected me for the Ph.D. position when he was head of the Geography Department in 2019.

All the reviewers with critical comments contributed to making the manuscripts better. 😊 Of course, also a huge thanks to all my family members, especially my mom, who always believed in me and supported me in all my choices in my life. Moreover, thank you for all the support and understanding during this period, and especially for always being there for me when needed. To Morgan, the one and only – I wish everyone could have a Morgan – it all would never have been possible without you – your constant support, help, comments, and expertise in field of geology. Thank you for surviving it all with me and keeping me sane with my feet on the ground – even on the most challenging days.

Superhero Troy (4 years), who made me start later but also finish earlier and made me realize how fast time flies and that we never get it back – you taught me to live for the moment!

Trym and Thyra, who helped me think of things other than my Ph.D. and showed me all the other beautiful things that everyday life brings.

Corona and home office gave "a quick start" to my Ph.D. by starting 16.3.2020 (1st day of my Ph.D.)

Finally, to all those many other people who helped me in some way on this journey, thank you for your contribution and company. Collaboration is the key to success. And, lucky me, I met and made many good friends along this path.

Sincerely,
Alexandra

Abbreviation

AI	Artificial intelligence
ANN	Artificial Neural Networks
CART	Classification and Regression Tree
CUDA	Compute Unified Device Architecture
DEM	Digital Elevation Model
DL	Deep Learning
dNDVI	Differenced Normalized Difference Vegetation Index
FCNN	Fully Convolutional Neural Network
FBK	Fondazione Bruno Kessler
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Networks
GEE	Google Earth Engine
GIS	Geographic Information Systems
GPU	Graphic Processor Units
IGE	Department of Geography
LiDAR	Light Detection and Ranging
MCC	Matthews correlation coefficient
MDP	Markov decision process
ML	Machine Learning
NDVI	Normalized Difference Vegetation Index
NGU	Geological Survey of Norway
NINA	Norwegian Institute for Nature Research (Norsk institutt for naturforskning)
NTNU	Norwegian University of Science and Technology
NVE	The Norwegian Water Resources and Energy Directorate (Norges vassdrags- og energidirektorat)
OBIA	Object-Based Image Analysis
RQ	Research questions
SAR	Synthetic Aperture Radar
TPI	Topographic Positioning Index
TP	True Positive
TP	True Negative

Abstract (English)

Artificial intelligence (AI) is one of the fastest-growing disciplines in information technology and has been successfully applied in many different fields. Where conventional approaches rely heavily on manual labour, data analysis, and expert knowledge, deep learning has emerged as a transformative tool for data analysis and its interpretation. Currently, the technology is available for everyone and allows to perform complex operations using large dimensional dataset but remaining challenging due to a required background in data science. This challenge opens for new strategies and has practical implications for the enhancement of the field itself and the development of new user-friendly approaches that can help geoscientists that do not have an extensive background in data science. This Ph.D. research takes up the challenge and aims to bridge the disciplinary gap between data science and geoscience as it demonstrates how non-experts can harness and integrate deep learning in applications based on the use of freely available data (e.g., terrain derivates, Landsat 8, Sentinel-1, and Sentinel-2) to simplify daily tasks and to improve future work.

The overarching goal of this Ph.D. study is to explore accessible and user-friendly approaches that enable geoscientists to apply deep learning techniques a to the automation of mapping outcrops, wetlands, landslides, or ravines. The achieved outcomes demonstrate and contribute to the broader enhancement and understanding of the application of DL techniques in geoscience, highlighting the importance of a continuous exploration and integration of these techniques in this field, and emphasizing the need for interdisciplinary collaboration between data scientists and geologists in order to drive innovation and maximize the potential of DL for the benefit of the geoscience community, and hence society at large.

Abstrakt (Norwegian)

Kunstig intelligens (AI) er en av de raskest voksende disiplinene innen informasjonsteknologi og brukes med suksess på mange forskjellige felt. Der konvensjonelle tilnærminger er avhengige av manuelt arbeid, dataanalyse og ekspertise, har dyp læring dukket opp som et transformativt verktøy for dataanalyse og tolkning. I dag er denne teknologien tilgjengelig for alle og gjør det mulig å utføre komplekse operasjoner ved hjelp av et stort dimensjonalt datasett, men forblir utfordrende på grunn av den nødvendige bakgrunnen innen datavitenskap. Denne utfordringen åpner for nye strategier og har praktiske implikasjoner for å forbedre selve feltet og utvikle nye brukervennlige tilnærminger som kan hjelpe geovitere som ikke har en omfattende bakgrunn innen datavitenskap. Denne Ph.D. forskning har tatt denne utfordringen og har som mål å bygge bro over det disiplinære gapet mellom datavitenskap og geovitenskap ved å vise hvordan ikke-eksperter kan bruke og integrere dyp læring i applikasjoner basert på bruk av fritt tilgjengelige data (f.eks. terrengderivater, Landsat 8, Sentinel - 1 og Sentinel-2) for å forenkle daglige oppgaver og forbedre fremtidig arbeid.

Hovedmålet med denne Ph.D. Målet med studien er å utforske rimelige og brukervennlige tilnærminger som lar geoforskere bruke dyplæringsteknikker for å automatisere kartleggingen av utspring, våtmarker, jordskred eller raviner. Resultatene som er oppnådd demonstrerer og bidrar til en bredere forbedring og forståelse av anvendelsen av DL-teknikker i geovitenskapene, fremhever viktigheten av kontinuerlig utforskning og integrering av disse teknikkene i feltet, og understreker behovet for tverrfaglig samarbeid mellom dataforskere og geologer i for å fremme innovasjon og maksimere potensialet til DL til fordel for geovitenskapssamfunnet og dermed samfunnet som helhet.

Abstrakt (Slovak)

Umelá inteligencia (AI) je jednou z najrýchlejšie rastúcich disciplín v oblasti informačných technológií a úspešne sa uplatňuje v mnohých rôznych oblastiach. Tam, kde sa konvenčné prístupy do veľkej miery spoliehajú na manuálnu prácu, analýzu údajov a odborné znalosti, sa hlboké učenie objavilo ako transformačný nástroj na analýzu údajov a ich interpretáciu. V súčasnosti je táto technológia dostupná pre každého a umožňuje vykonávať zložité operácie s použitím veľkého rozmerového súboru údajov, ale zostáva náročná vzhľadom na požadovanéznalosti v oblasti údajovej vedy. Táto výzva otvára nové stratégie a má praktické dôsledky na zlepšenie samotnej oblasti a vývoj nových užívateľsky prívetivých prístupov, ktoré môžu pomôcť geovedcom, ktorí nemajú rozsiahle zázemie v oblasti dátovej vedy. Táto Ph.D. práca sa chopila tejto výzvy a jej cieľom je preskúmať disciplinárnu priepasť medzi vedou o údajoch a geovedou, pretože ukazuje, ako môžu neodborníci využiť a integrovať hlboké vzdelávanie v aplikáciách založených na použití voľne dostupných údajov (napr. terénne deriváty, Landsat 8, Sentinel -1 a Sentinel-2) na zjednodušenie každodenných úloh a zlepšenie budúcej práce.

Hlavným cieľom tejto Ph.D. štúdie je preskúmať dostupné a užívateľsky prívetivé prístupy, ktoré umožňujú geovedcom aplikovať techniky hlbokého učenia na automatizáciu mapovania skalné podloží, mokradí, zosuvov pôdy alebo roklín. Dosiahnuté výsledky demonštrujú a prispievajú k širšiemu zlepšeniu a pochopeniu aplikácie hlbokého vzdelávania v geovedách, zdôrazňujú dôležitosť neustáleho skúmania a integrácie týchto techník v tejto oblasti a podčiarkujú potrebu interdisciplinárnej spolupráce medzi dátovými vedcami a geológmi v s cieľom podporiť inovácie a maximalizovať potenciál hlbokého vzdelávania v prospech geovedeckej komunity, a tým aj spoločnosti ako celku.

Prologue

Geosciences have existed since long before the new era of Machine Learning (ML) and Artificial Intelligence (AI). The methods used by geoscientists throughout the history of their research have relied on manual analysis, empirical interpretations, and human intuition. There was a time when the Earth's mysteries were unlocked by collecting small pieces and the puzzling of these together. Maps were drawn by hand and were perceived as an art. Cartographers, equipped with their tools and expertise, explored the Earth's surface, relying on years of training and craftsmanship, collecting what amounts to an incredible amount of data. These maps even today are crucial for such things as building planning and risk assessment.

This tedious process of revealing connections and patterns was changed by the introduction of Geographic Information Systems (GIS) in the 1960s. GIS formed the new backbone of geographic science, including spatial analysis and visualization. These were times when the potential of technologies was unexplored, waiting to unlock their analytic power. The researcher wishes to recognize and to highlight the dedication, insight, and progress of earlier geoscientists in unravelling geographical enigmas before the era of ML.

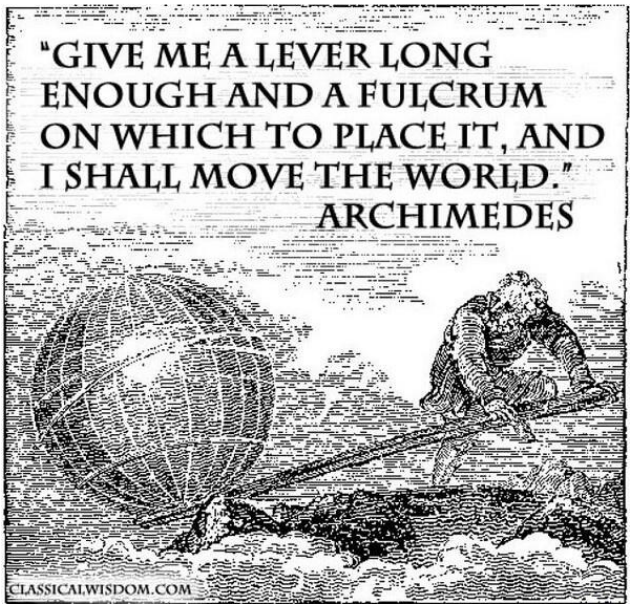
This thesis joins in this continuous progress, recognizing the significance of innovation and the today necessity to enhance human capabilities and our understanding of the world's complexity. However, the researcher wishes also to acknowledge that without geologists and scientists helping to understand the data before it enters the algorithm, to interpret the predictions, and then actually solve the problem and use the predictions further in the field or during map production, deep learning predictions make no sense.

Nowadays, one of the powerful tools of GIS, ML, is one of the fastest-growing areas of computer science. It has become a part of our daily activities, helping us with different tasks from unlocking our mobiles, to opening bank systems with fingerprints, to finding and classifying photos and driving cars.

Progresses in the introduction of Machine Learning frameworks in geoscience opened the field for advanced remote sensing classification techniques, as well as the development of prediction modelling algorithms.

Timely land cover maps help governments and the private sector alike to make well-informed decisions and respond quickly to a changing environment. Convenient and cost-efficient solutions, able to detect and map different geohazard features, are crucial. We witness the emerging power of computer-based methods, which can help produce results with possible replicability (using the same model and data) or reproducibility (using the same model and a new study area and new data).

Seeing the unseen, exploring, and making sense of the imperceptible, shedding light on powerful imagery that informs more than just plain pictures – is that a future in geoscience mapping? Can these data deepen our understanding of the world around us? Is this giving us possibilities we never had before? Can seeing more help us to understand more?



1. Introduction

More than 2200 years ago, Archimedes laid the foundation for what we now recognize as "simple machines" (winch, pulley, lever, wedge, and screw—known to antiquity). While machines have undergone significant changes since then, the main goal remains unchanged: we wish to develop devices that enhance efficiency in our daily lives. Fortunately, technology, such as Machine Learning algorithms, offer some possibilities to help us towards that goal.

Machine learning (ML) and its variants, in the form that we know today in the geoscientific community, have already been applied in many studies. These studies reveal the usefulness and potentialities of ML in geo-research fields, investigating some classical mapping problems, such as seismic hazards (M. Liu & Grana, 2019; Ma & Mei, 2021), reservoir characterization (Bihani et al., 2022; Okon & Appah, 2021), slope stability and landslide prediction (Ghorbanzadeh et al., 2019; Nava et al., 2022; Prakash et al., 2021; Tehrani et al., 2022) and mineral exploration (Ching et al., 2019; Cai Liu et al., 2023; Xiong et al., 2018). In the last few years, ML techniques, including Deep Learning (DL) methods, have become popular, demonstrating promising performance, especially if compared to conventional mapping and image classification methods (Alzubaidi et al., 2021; Choudhary et al., 2022). In this introduction, I provide a brief overview of ML techniques and their potential applications in geoscience, which I explored during my Ph.D.

In recent years, the size of data sets has significantly increased, surpassing the capabilities of traditional data processing and ML techniques in many cases. Fortunately, the application of DL techniques has been explored as a solution (X. Wang et al., 2020). It shows promising results in terms of high accuracy and performance when trained with huge amounts of data and more and more affordable computational power. Therefore, this thesis concentrates on an examination of classical mapping problems from a DL perspective, highlighting the benefits and presenting the achieved results through the application of different algorithms.

Ultimately, this research contributes to the broader understanding and utilization of ML and DL techniques by extension, as will later be explained in geoscience (Figure 1). By shedding light on the possibilities, it is meant to be a valuable resource for researchers and geoscientists seeking to leverage the power of DL to enhance their everyday work and propel the field forward with outstanding performance.

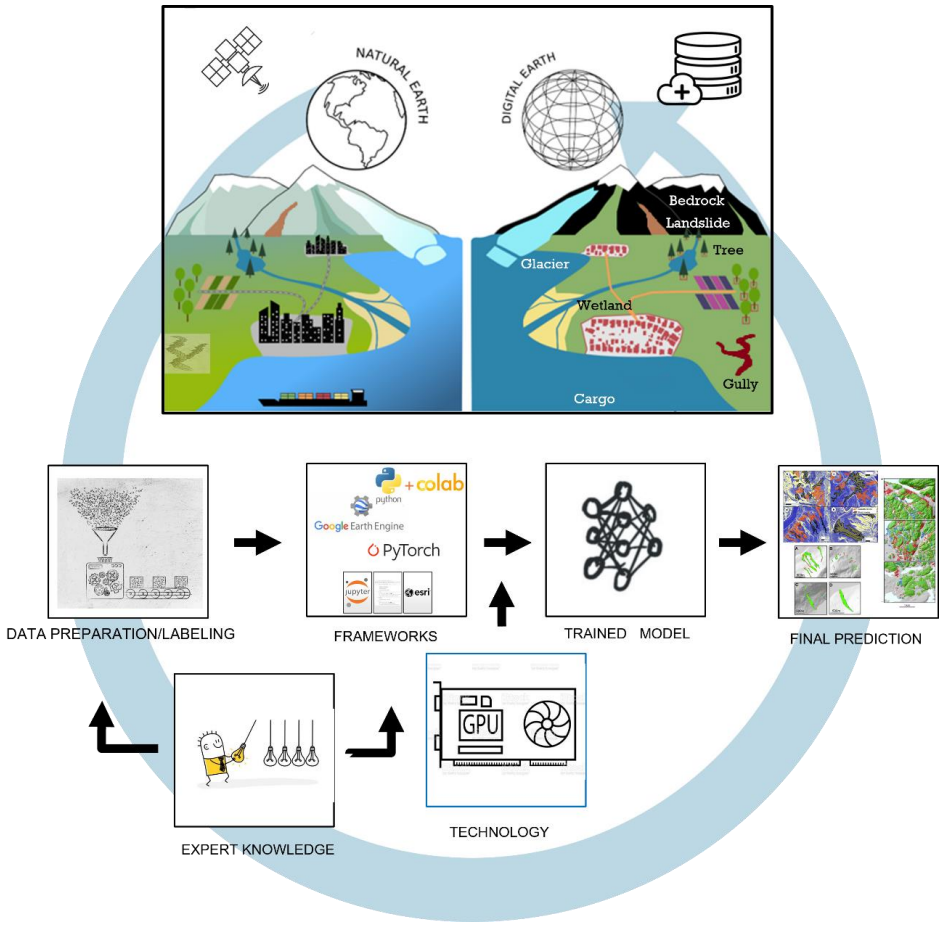


Figure 1: At the top are environmental phenomena represented digitally, which is the necessary first step in data preparation and labelling. Representing the real world in a GIS facilitates further DL steps as follows: (1) data preparation (based on expert knowledge), (2) deciding the most suitable framework for the defined problem (using available and most suitable technology) for (3) training the model and visualization of the final prediction.

Recognizing the interdisciplinary nature of DL in geoscience and geology, this research emphasizes the importance of collaboration between geoscientists and data

science experts (Emden et al., 2006). The main results of the multidisciplinary collaboration between NTNU and different international organizations (Geological Survey of Norway (NGU), NINA, FBK Trento, and Berkeley University) are presented in this thesis. The goal is to connect different fields and create methods and tools that enable geoscientists without expert data science backgrounds to apply the algorithms and the results of this Ph.D. research in their daily work. This will help them utilize these powerful tools more efficiently.

This thesis is composed by six articles. Articles I and II show how to use a cloud-based approach using Google Collaboratory. This platform is based on a Jupyter Notebook environment which allows running code entirely on the cloud (Dwivedi, 2023). Articles III and VI present a compilation of examples of different features (landslide, ravines) using the locally based platform with Jupyter Notebook, ArcPy libraries, and ArcGIS Pro with Image Analyst extension. Article IV explores the data preparation with the most suitable data band combination for DL training in global-scale landslide delineation. Lastly, article V shows the use of a globally compiled dataset for DL computation, that compares different segmentation models and uses an ensemble model, exploring and demonstrating its potential.

From this thesis, two main conclusions emerge. Firstly, the multidisciplinary collaborative approach of powerful data-driven technique led to successful applications of geoscientific problems. Secondly, the thesis reveals successful use of these powerful techniques for the improved future application in geoscience and geology. The future of DL in geoscience remains very bright, and this thesis contributes to further understanding and making better use of it for non-data scientists.

1.1. Research questions (RQ)

The main objective of this thesis is to investigate methods and possibilities for utilizing Machine- and Deep Learning techniques in geoscience, demonstrating their use also for non-data scientists. Figure 2 illustrates how the various articles published the following research questions analysed in this Ph.D.

1. **RQ1: Among the available platforms, identify optimal ML platforms for geospatial applications (Figure 2, Section 3.2):**

I) server-based cloud computing using Jupyter Notebook, Google Colab, and Google Earth Engine (GEE) (in collaboration with NINA);

II) locally based ArcPy libraries and Jupyter Notebook (in collaboration with NGU and NTNU);

III) server-based computing in the JupyterHub (in collaboration with UC Berkeley);

IV) locally based PyTorch segmentation models comparison on a global scale (in collaboration with FKB Trento).

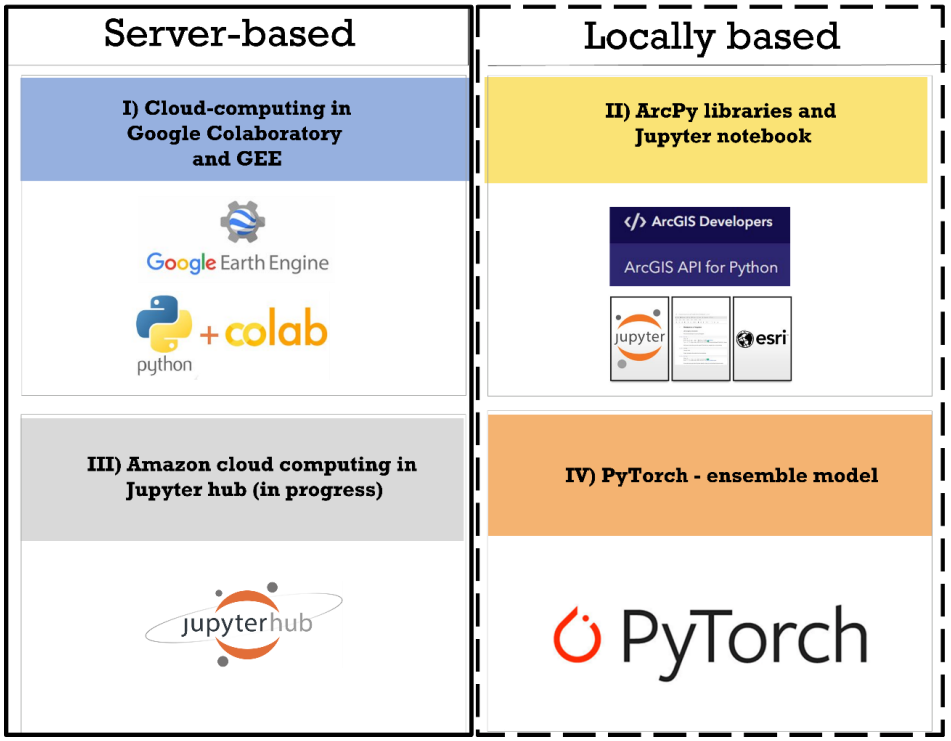


Figure 2: Used and presented platforms for evaluating the optimal platform and showing possibilities for geospatial applications in the scope of this thesis. The colours represent the following methods (used again in Figure 3): blue and grey (server-based computing), yellow and orange (locally based computing).

2. **RQ2: What methods are commonly used in geospatial applications for automatic classification currently, and where can they be improved with DL? (Figure 3, Section 3.3)**

I) How to compare Object-Based Image Analysis (OBIA), Machine Learning (ML), and Deep Learning (DL) for Geospatial Applications?

II) What are the strengths and weaknesses of OBIA, ML, and DL in geoscientific contexts?

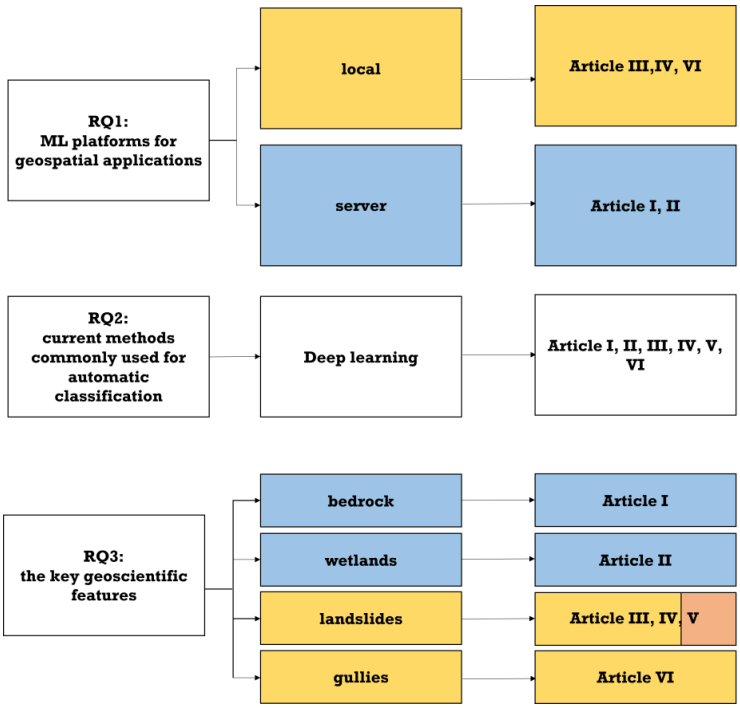


Figure 3: Publication and contribution overview, where yellow shows locally based computation uses in articles, blue for server-based computations. Article V is locally trained, but ensemble approach, therefore orange (same as in Figure 4).

3. **RQ3: Identify critical geoscientific features needed to evaluate the usefulness of automatic identification using DL (Figure 3, Section 1.2):**

I) What key features are used in existing automation systems?

II) What would be useful to automate in today's mapping? (Section 7, for each article)

The work explores server vs. locally based platforms (RQ1) exclusively in the field of deep learning (RQ2) for delineating various geological features that can help geoscientists with more efficient mapping (RQ3) (Figure 3). RQ3 suggests the features that might be beneficial for geoscientists to automate with the use of freely available data and images. The differences, 'state of the art,' and possible improvements on previous studies are described in the articles.

1.2 Publications

The thesis is structured as a collection of articles (Figure 4) and a summary outlining the performed work, theoretical background, methodology, main findings, conclusions, and future research developments. The final work is a result of a collaboration established during my Ph.D. studies at the Department of Geography, NTNU, with different regional actors (NTNU, NGU, NINA) and internationals (FBK Trento, US Berkeley). The colours used in Figure 3 correspond to those used in Figure 2, and represent the method applied in each of the articles: blue (cloud-computing with GEE), yellow (locally trained ArcPy libraries and Jupyter notebook), orange (locally trained PyTorch ensemble approach), green (only presents data preparation for DL, no approach used from Figure 2) and grey is the last method explored based on cloud computing, however not presented in the article section (in progress, Section 5.1).

Table 1: Publications included in the thesis, showing the topic, the methods used from Section 1.1, Figure 2, and the collaboration that contributed to finalizing the article. Article IV does not use I-IV methods but presents data preparation for Article V based on Articles III and IV findings.

	Topic	Method (Figure 2)	Collaboration
Article I	Where are outcrops?	I	NGU, NINA
Article II	Can we map wetlands? (co-author)	I	NINA
Article III	Landslide detection – DL model comparison (global vs. local)	II	NTNU
Article IV	Landslide signatures in SAR images (co-author)	-	NTNU
Article V	Landslide detection globally trained ensemble model	IV	FBK Trento, NTNU
Article VI	Can we map gullies?	II	NGU

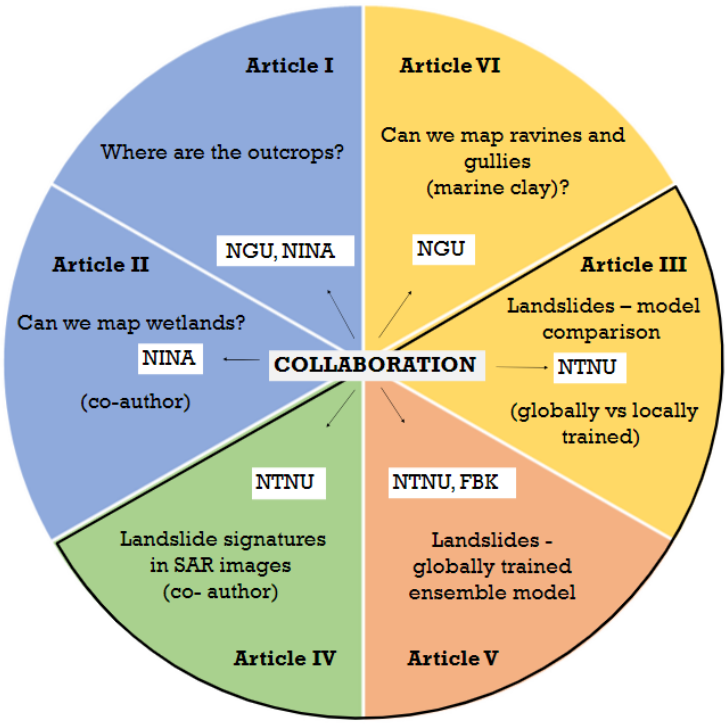


Figure 4: The six articles included in the thesis, where colours represent methods applied in the articles presented in Figure 2. Green colour for Article IV means that none of the aforementioned methods were used in this article.

Article References:

Article I

Garnerød, A. J., Bakkestuen, V., Calovi, M., Fredin, O., & Ketil, J. (2023). Where are the outcrops? Automatic delineation of bedrock from sediments using Deep-Learning techniques. 18(May). https://doi.org/10.1016/j.acags.2023.100119https://doi.org/10.1016/j.acags.2023.100119

Article II

Bakkestuen, V., Venter, Z., Ganerød, A. J., & Framstad, E. (2023). Delineation of Wetland Areas in South Norway from Sentinel-2 Imagery and Light Detection and Ranging (LiDAR) Using TensorFlow, U-Net, and Google Earth Engine. Remote Sensing, 15(5), 1–17. https://doi.org/10.3390/rs15051203

Article III

Ganerød, A. J., Lindsay, E., Fredin, O., Myrvoll, T. A., Nordal, S., & Rød, J. K. (2023). Globally vs. Locally Trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape. Remote Sensing, 15(4). https://doi.org/10.3390/rs15040895https://doi.org/10.3390/rs15040895

Article IV

Lindsay, E., Devoli, G., Reiche, J., Frauenfelder, R., Jarna Ganerød, A., Nordal, S., & Tokle, L.-C. (2023). Understanding Landslide Expression in SAR Backscatter Data: A Global Study. PrePrint. https://doi.org/10.20944/preprints202302.0390.v1https://doi.org/10.20944/preprints202302.0390.v1

Article V

Garnerød, A. J.,Franch, G., Lindsay, E., Calovi, M. (2023). Automating global landslide detection with heterogeneous ensemble deep-learning classification. PrePrint. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4571303

Article VI

Ganerød, A. J., van Boeckel, M., & Solberg, I.-L. (2023). Where are the ravines? A Case Study of Gully Landscapes in Norway Using Deep Learning. PrePrint. https://doi.org/10.22541/au.169230564.41279479/v2https://doi.org/10.22541/au.169230564.41279479/v2

1.3 Contribution at Glance

This thesis has contributed to the literature on the intersection of geoscience and data science in the following ways:

- It proposes the novel idea of using cloud-computing to delineate bedrock from sediments and wetlands;

- It demonstrates that U-Net successfully delineated trivial and complex geospatial problems (bedrocks, wetlands, landslides, gullies);
- It compares the globally pre-trained landslide model with two locally trained Machine and Deep Learning models;
- It proposes a method for the regular monitoring of landslides based on three predictors exclusively using the pre-trained model;
- It compares nine different segmentation models for creating a global ensemble pre-trained model;
- It created a functioning pre-trained model for delineating gullies in the gully landscape;
- It develops a method to map the spatial extent and condition of red-listed landforms in Norway and to develop monitoring systems for future landscape change.

1.4 Outline

The thesis is structured as follows: Section 2 contains the background on the topics covered in the publications. Section 3 presents the methods and data preparation parts and explains the different terms. Section 4 discusses the contributions of each article as a part of the thesis, and Section 5 sheds light on the conclusions, along with some reflections on possible future work. Section 6 presents the development and flow of my Ph.D. in time and space. Finally, in Section 7, the articles are presented as published journal articles (Article I, II, III) and three preprints (Article IV, V, VI).

2. Background

Machine Learning (ML algorithms build mathematical models based on sample data, called "training data," to make predictions. The primary advantage of ML, when compared to human decision-making, is in utilizing historical data. The Machine Learning algorithms can be trained to recognize trends and patterns. Using this knowledge as a foundation, the algorithm can apply the same process to present data to make predictions about the present or even the future. It can establish rules that produce desired outcomes with the assumption that the inputs come from the same or a very similar statistical distribution as the one the training data are drawn from (Daume III & Marcu, 2006; Goodfellow et al., 2020).

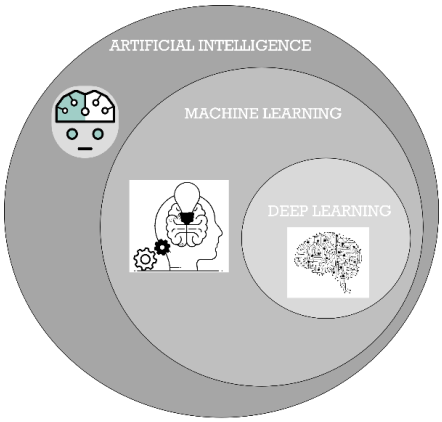
The ML process, in the same way that humans learn (Burkov, 2019), varies, considering the change in prediction output when inputs are slightly changed. ML models mostly lack the adaptability and generalization capabilities natural to humans (Sinz et al., 2019). For example, a model trained on data related to bedrock or landslides in vegetated areas may not perform well when applied to bare land areas (Article I, III, V). This demonstrates a need for replicability in new, unknown areas. However, given the speed in technology development and trends in ML and Deep Learning (DL), there is a possibility for revolutionizing efficiency and high potential to use DL in addition to the techniques used in today's mapping tasks for increasing productivity and making further high-quality data available. There is enormous progress in possibilities of powerful computation, many more people are applying DL into their studies, the field is expanding by exchanging of experiences by publishing new findings. The gap between data scientists and experts is getting narrower, because they starting to understand that the key is to collaborate and work together. Research and technological development continue to advance hand in hand. Therefore, there are opportunities to refine and improve DL algorithms and reduce the gap between human decision-making and machine performance. This research summarizes the ongoing efforts to explore the efficiency and decision-making capabilities of DL in the geoscience community.

2.1 Explanation versus prediction

(Conventional Machine Learning versus deep learning)

Statisticians, over a couple of decades, have debated the scientific value of predictive models versus explanatory and descriptive models (Geisser, **1975**; Wallis, **1980**; Breiman, **2001**; Shmueli *et al.*, **2010**; Parzen & Mukhopadhyay, **2012**). Lately, this debate has taken on a new dimension mostly due to the emergence of ML techniques in the computer science community as powerful predictive methods compared to classical statistics-based methods. By the start of the millennium, this resulted in two primary cultures in data-driven analysis: (1) data modelling and (2) algorithmic modelling, where the common aim is to gain information from data to be better able to make predictions. Shmueli (2010) discusses the difference between explanatory and predictive modelling. He questions whether an explanatory model must have some predictive power to be considered scientifically useful. Data modelling is designed to produce the facts, whereas algorithmic modelling is a predictive method, which comes up with possible outcomes based on the available data. Therefore, most of the ML techniques use algorithmic modelling producing predictions.

DL is a subset of ML (Figure 5). The main distinction between DL and conventional ML algorithms is based on the modality of the data or how many different types of data algorithms the model is learning from. By conventional ML algorithm, I mean fundamental algorithm structures that solve a given problem. DL is primarily based on Artificial Neural Networks (ANN), while conventional ML methods include algorithms and rule-based approaches based on statistics (Burkov, 2019; Tseng et al., 2008).



Artificial Intelligence is a whole knowledge field (like biology/chemistry).

Machine Learning is a part of artificial intelligence.

Neural Network is one type of machine learning.

Deep Learning is a modern method of building, training, and using neural networks.

Figure 5: Deep learning as a subset of Machine Learning and Machine Learning as a subset of Artificial Intelligence

In conventional ML algorithms, labelled or unlabelled data has certain features, parameters, variables, or attributes. Labelled data are the data with meaningful labels or assigned classes to the raw collected data. Moreover, features or variables are individual properties from a data set used as input to ML models. Features are often represented as numerical or string columns in data sets. The number of features might be crucial, depending on data quantity and the utilized ML and DL algorithm (Figure 6).

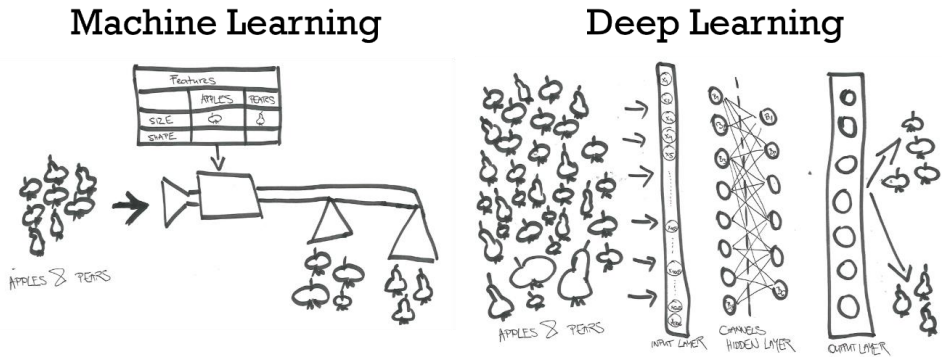


Figure 6: Workflow difference of using conventional ML (left) versus DL (right).

Through the training process, a conventional ML algorithm learns to find patterns in the data based on features. On the other hand, in DL, the input data (e.g., image, text, video, or time series) is directly sent to ANN, where each network hierarchically

learns specific features about the dataset from the input data. Therefore, DL requires vast amounts of training data because of a higher amount of hyperparameters (Avci et al., 2021; Tseng et al., 2008), and extensive computing power Graphic Processor Units (GPUs). In fact, we only know the input and output. A hidden layer (Figure 6) is between the input and output layers, where artificial neurons take in a set of weighted inputs and produce an output such as separating apples from pers. They are "hidden" because the actual valid values are unknown in the training dataset. The more hidden are the layers between input and output, the deeper the algorithm is. Each neural network has at least one hidden layer, and a neural network with two or more hidden layers is defined as a deep neural network. These are further used to find a pattern that connects the input data to a specific label for a decision.

2.2 Learning methods

ML techniques can solve various problems based on the type of available data and research question (Sarker, 2021a). The main ML methods are (1) supervised, (2) unsupervised (semi-supervised), and (3) reinforcement learning (Table 2).

Table 2: Summarize Supervised Learning, Unsupervised Learning and Reinforcement Learning.

Information	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance	Works on interacting with the environment
Data type	Labelled data	Unlabelled data	Not predefined data
Problem	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, SVM, KNN	K – Means, Density-Based Spatial Clustering	Q – Learning
Aim	Calculate outcomes	Explore patterns	Learn a series of action (trial and error)
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self-Driving Cars, Gaming, Healthcare

Supervised learning algorithms are used on data that consists of a set of inputs (independent features) and their corresponding outputs (labels). The training process continues until the model achieves the desired level of accuracy. This generally gives the potential to solve a problem but has some limitations: it is based on the biases in which it is supervised. That means it learns how to perform the task independently based on the information it has received. Additionally, supervised learning depends on a substantial manual effort in terms of label preparation. The more manual labels that are provided, the more effectively the algorithm can use the training data. The model's validity is assessed by evaluating the model on unused data (a test set). Usual methods of supervised learning algorithms include Decision Trees (Random Forest, Gradient Boosting algorithms), Support Vector Machines, and ANN, including supervised DL algorithms. Supervised learning was applied in all the articles (IBM Data and AI Team, 2023; Chao Liu et al., 2022; Sarker, 2021b).

Unsupervised learning does not use given labels to make predictions; rather, it tries to "make sense" of given data. The expected outcome is not directly defined. Instead, the algorithm is trained to find those outcomes. The algorithm summarizes the data

meaningfully and then tries to solve the problem based on inferences. The conventional use of unsupervised learning is clustering. The unsupervised learning algorithms are represented using K-means, Density-Based Spatial Clustering, and DL algorithms (autoencoders and Generative Adversarial Networks (GANs)) (Goodfellow et al., 2020; Morales & Escalante, 2022).

In reinforcement learning, the algorithm is trained to make specific decisions. The computer learns from experience and tries to identify the most useful information to accurately make specified decisions. There is no need for labelled input. Instead, the focus is on finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge) (Botvinick et al., 2019; Kaelbling et al., 1996; Mousavi et al., 2018). The environment often uses a Markov Decision Process (MDP) because many reinforced learning algorithms use dynamic programming techniques (Hutter et al., 2019; Otterlo & Wiering, 2012).

2.3 Challenges for Earth Observation

In remote sensing, 'data fusion' is the typical challenge, which consists in integrating data from multiple sources to produce specific, comprehensive, unified data about an entity with the aim to solve a problem (Jia et al., 2021; Kawulok et al., 2020).

A typical example would be:
1) combining images having different resolutions (such as in Articles IV and V, which required data normalization);
2) joint processing of imagery and a resolution DEM to address the surface topography (as in Articles II, III, IV, V, and VI);
3) the fusion of optical and radar data (clouds, radar shadows) (as in Articles III and V).

Challenges with data fusion that were encountered in this work are presented more closely in section 3.1 Data - image processing and data preparation.

2.4 Case studies and main objectives

All six articles that make up this thesis involve an accessible DL process which is closely described in the Method section. However, each presents different specific case studies (Figure 7), which are closely described in each article. Here I highlight the study areas and the specific main objectives of each article:

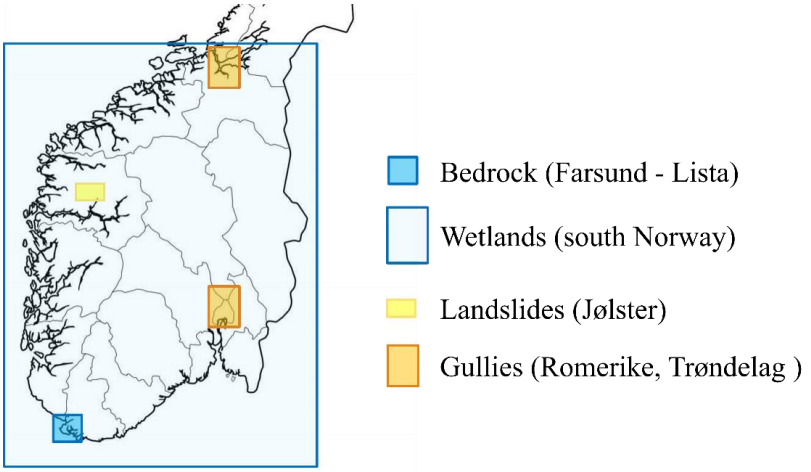


Figure 7: Study areas and objects of study presented in this thesis. Article I: bedrock (dark blue), Article II: wetlands (light blue), Article III: landslides (yellow), Article VI: gullies (orange); Article IV and V study 30 case studies spread globally.

Article I

The main objective is delineating bedrock for mapping purposes in an area (Farsund – Lista, Norway) characterized by bedrock and sediments. Using a Digital Elevation Model (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., because an area is covered by vegetation), and mapping could potentially be done over larger areas within the same resource restrictions because of more efficient mapping. We applied an automated approach that uses cloud computing, Deep Learning, fully convolutional neural networks, and a U-Net model applied in Google Collaboratory (Colab).

Article II

The main objective is to delineate wetlands over large areas, specifically in southern Norway. The current wetland mapping techniques in Norway are tedious and costly, and remote sensing provides an opportunity for large-scale mapping. We highlight the potential of satellite-based wetland maps for the ecosystem, able to account for changes in wetlands over time, which is not feasible with traditional mapping methods. We implemented U-Net through Google Earth Engine (GEE) with Sentinel-2 and LiDAR data images over southern Norway.

Article III, IV, V

Articles III, IV, and V have the same objective: moving towards an automated landslide monitoring system. Landslide detection is crucial for effective disaster response and hazard mapping. Traditionally, landslide inventories have been created manually, relying on expert visual interpretation of remote sensing imagery and field observation. There is much to gain from automated detection solutions, especially in remote areas. Article III tests a DL algorithm on the well-mapped local landslide area of Jølster. Article IV builds on Article III and expands to a globally spread case study areas. By exploring trends in the spatial and temporal signatures of over 1000 landslides in 30 diverse case studies, we build on our knowledge and understanding of the landslide expression in Synthetic-aperture radar (SAR) backscatter data. The knowledge was then further applied in Article V, where we build the robust ensemble Deep Learning model as a next step towards the development of an automatic landslide monitoring system at a global scale that combines different segmentation models (explain in Section 3.5).

Article VI

The main objective of this article is the detection of gullies in gully landscapes such as those found in the Romerike district and Trøndelag county, Norway. Gullies in marine clay are important landforms indicative of soil erosion, natural hazards, and high conservation value given their ability to support high plant and animal diversity. Gully areas are not properly mapped yet in high resolution, so DL is here tested for

this purpose. This research aims to develop monitoring systems for future landscape change.

3. Methods – Detection and mapping

3.1 Data-image processing and data preparation

Data compilation processes have slightly varied throughout the different articles. Data has been used as follows: DEM 10m (Article I), 1m, and their derivates, LiDAR 1m (Article II, VI), Sentinel-1 (SAR) (Article III, IV, V), Sentinel-2 (Articles II, III, IV, V), and Landsat 8 (Article I).

In **Article I,** a raster dataset (tiff format) delineating bedrock and sediments over the study area was set as ground truth data. The dataset was based on a recent quaternary map manually produced by NGU at a scale of 1:50.000, and published as a vector format (Fredin et al., 2015). 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. As a predictor, we compared two sets of data: (1) a cloud-free Landsat 8 composite scene (both optical and thermal bands), and (2) derivatives from a DEM 10m (slope, elevation, slope sum, relative relief, valley depth, and topographical position index). We calculated the terrain derivatives using QuantumGIS (a free and open-source desktop GIS) and uploaded the different layers into GEE for computing, while Landsat 8 was cloud-based, and directly available on GEE.

In **Article II,** we selected 3 Sentinel-2 bands (B3-green, B8-near infrared, B12-red edge) and ten indices (such as Normalized Difference Vegetation Index (NDVI), the normalized burn ratio, the normalized difference red/green, the plant senescence reflectance index, etc.), as well as the mean canopy height model from LiDAR (accessed from hoydedata.no) as a substitute for tree height. All the bans and indices are presented and explained in Article II. Using the cloud-masked Sentinel-2 imagery, we derived 13 bands. In Norway, LiDAR data is freely available from hoydedata.no (accessed 23.04.2022), and both terrain and surface models from LiDAR data were uploaded and processed into GEE.

In **Article III**, both bi-temporal and multi-temporal composite change images (Figure 8) were used. These were based either on bitemporal, i.e., one pre-event image and one post-event image, or multi-temporal, i.e., stacks of images used from pre- and post-events. Change images (multi-temporal) were prepared using Sentinel-1 (SAR) and Sentinel-2 data in GEE.

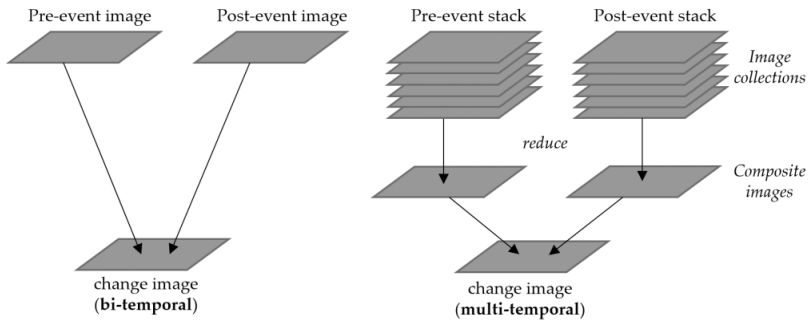


Figure 8: Bi-temporal and multi-temporal approaches for creating change images.

This required creating separate pre- and post-event image collections (stacks) filtered by date and location. In addition, the Sentinel-1 image was filtered by polarization (VV and VH separately) and satellite acquisition geometry (ascending and descending separately) (Figure 9). VV is a mode that transmits vertical waves and receives vertical waves to create the SAR image. VH is a mode that transmits vertical waves and receives horizontal waves to create the SAR image (CapellaSpace, n.d.). For Sentinel-1, all stack images were terrain-corrected (Vollrath et al., 2020). Finally, the pre-event composite images were subtracted from the post-event composite images to create dVV and dVH images (Figure 9). For Sentinel-2, NDVI (ScienceDirect, 2018; Weier & Herring, 2000) band was added to each image in the filtered stack. Then a 'greenest pixel composite' (maximum NDVI) image was produced using the quality mosaic tool. Finally, the pre-event composite was subtracted from the post-event composite to produce a differenced NDVI (dNDVI) image (Lindsay et al., 2022).This required creating separate pre- and post-event image collections (stacks) filtered by date and location. In addition, the Sentinel-1 image was filtered by polarization (VV and VH separately) and satellite acquisition geometry (ascending and descending separately) (Figure 9). For Sentinel-1, all stack images were terrain-corrected (Vollrath et al., 2020). Finally, the pre-event

composite images were subtracted from the post-event composite images to create dVV and dVH images (Figure 9). For Sentinel-2, an NDVI (ScienceDirect, 2018; Weier & Herring, 2000) band was added to each image in the filtered stack. Then a 'greenest pixel composite' (maximum NDVI) image was produced using the quality mosaic tool. Finally, the pre-event composite was subtracted from the post-event composite to produce a dNDVI image (Lindsay et al., 2022).

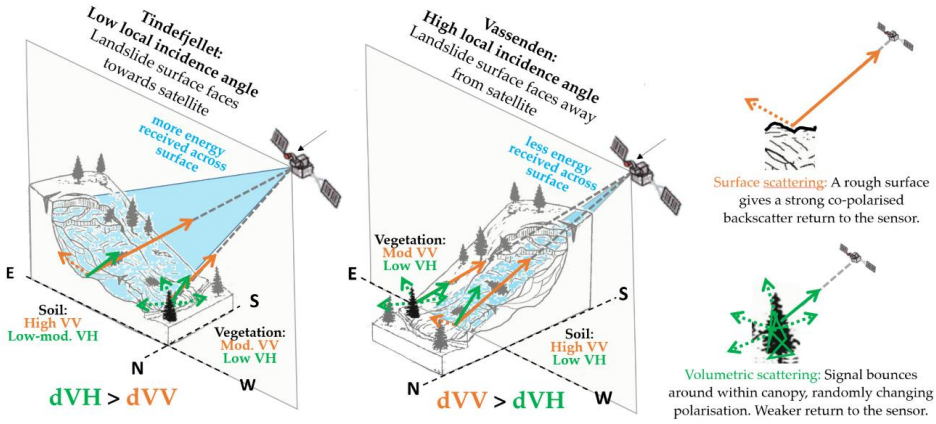


Figure 9: Conceptual model showing how Local Incidence Angle (LIA) affects the visibility of landslides in VV and VH polarisations (from Lindsay *et al.*, 2022).

In **Article IV**, data was calculated and exported from GEE with the final goal of creating the dataset for use in **Article V. Article V** create different band settings based on the dataset and apply them further to train a deep learning algorithm. We used multi-temporal composite change images exported using GEE (Gorelick et al., 2017). The Sentinel-1 images are already available on the GEE platform as pre-processed (calibrated and ortho-corrected) and in 10m resolution. A Machine-Learning-based land cover classification was created for each case study using the ee.smile.CART algorithm (Classification and Regression Tree) in GEE (Breiman et al., 2017) using Sentinel-2 image with minimal cloud cover from before the landslide event, along with the DEM and pre-event Sentinel-1 composite images. Finally, the following data was exported for each case study: (1) landslide polygons and (2) geotiff raster images, including 14 bands. These are described in detail in Article IV.

In **Article VI,** we used terrain derivatives from a high-resolution DEM (1 m resolution), which comes as a derived product from LiDAR, accessible at Kartverket

(Kartverket, 2016, 2021). The terrain derivatives comprise Slope, Topographic Positioning Index (TPI), and Terrain roughness stacked into one composite band.

3.2 Computation platform and method

Altogether, in all six presented articles in the scope of this thesis, we compared two different computational platforms: (1) local and (2) server-based (Figure 2). Each of them has its pros and cons. In this section, are presented some of the benefits of each application using Deep Learning in geoscientific mapping and is up to each specific task to make decision which fits best.

When choosing the computation method (server vs. a local device), capability, computation cost, complexity, and model size are important factors to take into consideration. The remote server-based computation needs to send, store, and process data on the remote server. The model uses the server for training and then sends the output back. Network connection and delays are important factors during this process. Implementing the operations, visualizing, downloading, and also the costs are important considerations. However, a remote server is generally faster than a local device since it can process more data in a shorter time (Ghasemi et al., 2022). GPUs have high memory and the ability to conduct numerous parallel computations. Therefore, GPU became a widely accepted method for training DL models (Shaikh, 2019), and for this, GPU quality can significantly affect the performance of any DL models. In contrast, using a powerful local device to implement DL computations is more convenient and flexible. Potentially, it requires lighter algorithms to be executed effectively or just a longer computational time (Pandey et al., 2022).
Altogether, this thesis applied DL techniques in four different computational platforms, shown in Figure 2. Two are server-based (Articles I and II), and three are locally based platforms (Articles III, V, and VI). Article IV calculated and processed datasets using GEE (server) and exported for the use locally in Article V.

3.2.1 Cloud computing using Jupyter Notebook and GEE (Article I & II)

Google Earth Engine (GEE) is a commonly used server-based platform that allows to analyse remote sensing data and to provide high-performance computing resources for processing large geospatial datasets online, without the needs of downloading and processing the imagery locally (H. Li et al., 2019). The GEE platform provides a user-friendly and cloud-base processing environment of free and available data, able to use the power of thousands of computers located in the Google data centres around the world (Prasai et al., 2021; Tamiminia et al., 2020). We used GEE and Google Colaboratory (Google Colab) to store, customize, and export the ground truth data for cloud-based DL modelling. Google Colab allowed us to write and execute Python code through the browser connected to GEE.

3.2.2 ArcPy libraries using Jupyter Notebook and ArcGIS Pro (Article III & VI)

We used Jupyter Notebook with the ArcPy libraries (Tobergte & Curtis, 2013) in combination with ArcGIS Pro functionalities using the Geoprocessing and the Image Analyst extensions ('Export Training Data for Deep Learning') for preparing the training dataset (ArcGIS Pro, 2022). This approach was applied using a local platform with CUDA Graphics Processing Units (GPU) NVIDIA GeForce RTX 3080 Ti 12GB. The entire model was trained in Jupyter Notebook, and the final trained model was saved as a 'Deep Learning Package' (.dlpk format), the standard format used to deploy DL models on the ArcGIS Pro platform. The developed pre-trained model (ESRI, 2023; Han et al., 2021) can be applied for further implementation in different untouched areas using the 'Classify Pixels using Deep Learning' package (Article VI).

3.2.3 PyTorch coding and ensembled model (Article V)

In Article V, PyTorch, nine different segmentation models (Section 3.5) and five different loss functions have been used. Specifically, BCE, Dice, Focal, Jaccard, and Lovasz Loss, as these are known to be popular loss functions for semantic segmentation (Jadon, 2020). The loss function is the function that computes the distance between the current output of the algorithm and the expected output and

evaluate it. The choice relied on the models provided by the open-source library segmentation models PyTorch (SMP - https://smp.readthedocs.io/), which implements nine different segmentation models (many of which have shown state-of-the-art results when presented). We evaluated four different settings of ensemble learning to calculate the performance of an ensemble model based on 10, 20 and 40 best single performing models on validation data. This approach was applied using a local platform with a powerful GPU: Nvidia RTX 4090 GPU (24GB VRAM). The models are trained on a global case study, and results show that the developed framework can be applied to pre-trained models on untouched hazardous areas.

3.2.4 Cloud computing using JulyterHub

(work in progress - Article VII, Section 5.1)

JupyterHub is a server-based approach that enables multiple users to collaborate by providing an individual Jupyter Notebook server for each user. Once a cluster is created, JupyterHub is installed and configured and can be accessed from a browser. JupyterLab has support for real-time collaboration, where multiple users are working with the same Jupyter server and can see each other's edits, which makes it a great tool for collaboration (JupyterHub - Sphinx, 2023; Oracle, 2023).

3.3 Pixel- versus object-based image analysis (OBIA)

Pixel-based detection is performed with pixels both as input and output. This method requires parametric tuning and precise geometrical correction to be applied to large areas (Sameen & Pradhan, 2019). The pixel values are often binary, labelled as absence (0) or presence (1) by the experts (as used in Articles I, II, III, V). The labelled pixels, and their corresponding signatures from bands of input images, are used to train the algorithms. Afterward, the classification uses metrics derived from the confusion matrix to evaluate the models' performance (as in Articles I, III, V, VI).

On the other hand, object-based methods within the framework of Object-Based Image Analysis (OBIA) include two major steps: (1) image segmentation and (2) classification of the emerged segments. OBIA offers extra features to distinguish

decided features from other objects (Belgiu & Drăguţ, 2014; Jarna et al., 2019; L. Wang et al., 2018). Optimization of segmentation parameters is needed (Myint et al., 2011), and therefore the degree of automation can be lower compared to the pixel-based methods (Sameen & Pradhan, 2019). The OBIA was assessed in the early stage of this Ph.D. thesis and evaluated as not suitable for mapping features in this work.

3.4 What is Deep learning?

Deep Learning (DL) is a subbranch of Machine Learning (ML) and Artificial Intelligence (AI) and is considered to be, together with robotics, 3D printing, and quantum computing, a core technology of today's Fourth Industrial Revolution (4IR or Industry 4.0) (Sarker, 2021a; Schwab, 2016; Zuo et al., 2021). We tested the applicability of DL (RQ2) to geoscience mapping in all the articles presented in this Ph.D. thesis.

Neural Networks (NNs), as a backbone of DL algorithms, were introduced in the 1980s. The number of node layers, or depth, is what sets apart a single NN from a DL algorithm. A NN that consists of more than three layers—which would be inclusive of the inputs and the output—can be considered to be a DL algorithm (IBM Data and AI Team, 2023). However, even though they were successfully used already in 1980s for many applications, the research on NN decreased until 2006 because of a lack of technology development (Hardesty, 2017; Standford, 2023). In 2006, the concept "Deep Learning" was introduced by Hinton, Osindero, and Teh (2006) and related to the concept of Artificial Neural Networks (ANN). When properly trained, the DL networks proved to produce successful results in various classification and regression challenges (Alzubaidi et al., 2021; Sarker, 2021b).

Therefore, DL technology is considered a hot topic in Machine Learning, Artificial Intelligence, data science, and analytics, for its learning capabilities from the given input data (ground truth). The performance comparison between DL and conventional ML algorithms is illustrated in Figure 10. Deep learning increases its performance with the increase of the amount of data. Therefore, DL is useful when dealing with a large amount of data (. Even though the availability of spatial data in

Norway is very high, DL techniques at the beginning of my journey in 2020 were not properly explored within the topics and the scope of this thesis, and most of the work in this thesis focus on DL applications.

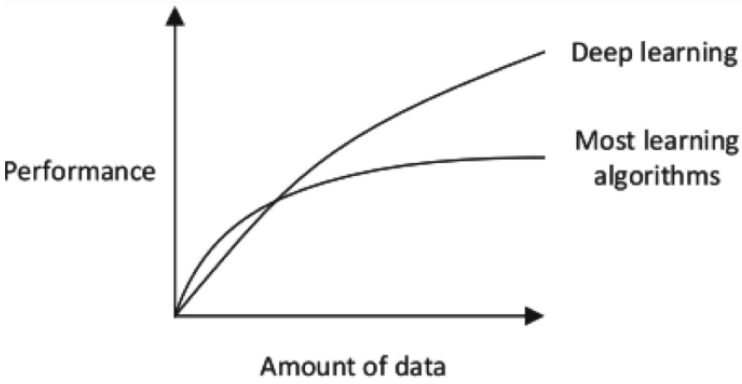


Figure 10: An illustration of the performance comparison between Deep Learning (DL) and other Machine Learning (ML) algorithms. DL modelling from large amounts of data can increase the performance, from (Sarker, 2021b).

DL methods can be categorized into three groups, which by order of complexity are: (1) image classification, (2) object detection, and (3) semantic segmentation (Figure 11).

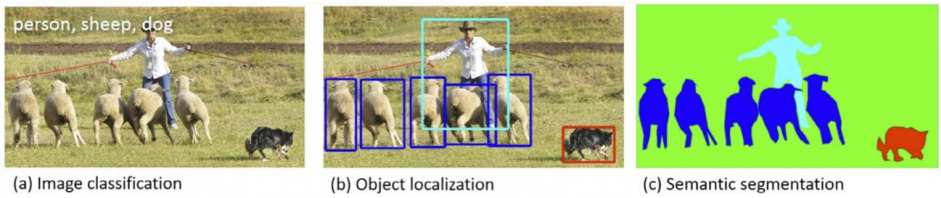


Figure 11: Shows DL methods as follows: (a) image classification, (b) object detection, and (c) semantic segmentation.

The main goal in image classification is, for each pixel, to determine the value of the label (presence or absence). This is done accordingly to specific rules and results in groups of pixels that have similar label values. In object detection, we aim to identify objects in an image with the help of bounding boxes (Ghasemi et al., 2022; Tan et al., 2021; Zaidi et al., 2022). Image semantic segmentation is trying to accurately determine the exact boundary of the objects in the image (Rzhanov et al., 2012). In semantic segmentation, each pixel in an image is assigned to a certain class, which can be considered a classification problem per pixel (Mary et al., 2020).

3.5 Segmentation models

Image segmentation aims at dividing an image into multiple parts, or regions, that belong to the same class. This task is based on specific criteria, such as colour, texture, and mean spectral value, weighted by their size and shape (Laio & Torino, 2011). Image segmentation can also be called pixel-level classification (X. Liu et al., 2019). Semantic Segmentation Models are a class of methods that address semantically segmenting (explained in 3.4, Figure 11c) an image into different object classes. The semantic segmentation model list is long and continuously updated (PapersWithCode, 2023). In all Articles, we used the U-Net model for semantic segmentation. However, in the globally trained model for landslide detection (Article V), we compared all the models listed in Table 3. The image segmentation algorithms based on DL models have achieved impressive performance in various image segmentation tasks.

Minaee et al. (2022) divide DL network architectures into the following categories:

1) Fully convolutional networks
2) Convolutional models with graphical models
3) Encoder-decoder-based models
4) Multiscale and pyramid network-based models
5) R-CNN-based models (for instance, segmentation)
6) Dilated convolutional models and DeepLab family
7) Recurrent neural network-based models
8) Attention-based models
9) Generative models and adversarial training
10) Convolutional models with active contour models

Table 3: Segmentation models used for the fully convolutional neural network (FCNN).

Segmentation Models	Meaning	Categories (Minaee et al., 2022)	Specialty	Source
U-Net	U shape	encoder-decoder (3)	works with fewer training images	(Ronneberger et al., 2015)
U-Net++	U shape ++	nested encoder-decoder (3)	more complex decoder	(Zhou et al., 2018)
MA-Net	Multi-scale Attention Net	Multiscale and Pyramid Network-Based Model (4)	Two blocks: position-wise and multi-scale fusion attention block	(Fan et al., 2020)
Linknet	LinkNet	encoder-decoder (3)	uses sum for fusing decoder blocks	(Chaurasia & Culurciello, 2018)
FPN	Feature Pyramid Network	Multiscale and Pyramid Network-Based Model (4)	without the need to compute image pyramids - suitable for small objects	(Li *et al.*, **2019**)
PSP-Net	Pyramid Scene Parsing Network	Multiscale and Pyramid Network-Based Model (4)	not suitable for small objects-based context aggregation.	(Zhao et al., 2017)
PAN	Pyramid Attention Network	The Regional CNN (R-CNN) (5)	data-to-text generation	(Jiang et al., 2020)
DeepLabV3	DeepLab version 3	Dilated Convolutional Model (6)	uses dilated convolutions	(Chen, Papandreou, et al., 2018)
DeepLabV3+	DeepLab version 3+	Dilated Convolutional Model (6)	adds simple yet effective decoder module	(Chen, Zhu, et al., 2018)

3.6 Ensemble learning

Ensembling is the process of combining multiple learning algorithms to obtain their combined performance, improve the overall output and obtain a robust and reliable model to make predictions. When combining several models, we often observe one reliable, more robust model (Yang et al., 2022) (same as Articles V, VI). As shown in Figure 12, models are stacked together to improve their performance and get one final prediction.

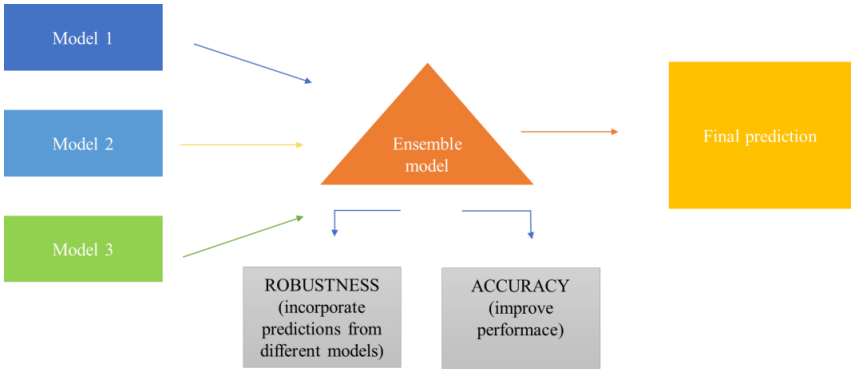


Figure 12: Shows that raw data is given as input or ground truth to the ensemble model, and then the training is done to get the final prediction.

The ensemble concept has existed forever, in the sense that most cultures have expressions for the 'wisdom of the crowd' (Wu & Levinson, 2021). Ensemble models have been applied in many disciplines and applications (Caté et al., 2017; Merghadi et al., 2020; Shrestha et al., 2022). However, most studies still rely on a single well-performing model. The performances of different segmentation models (Section 3.5) are usually presented to show variety in predictions connected to different features when using various single models. Still, DL models individually proved competent in most applications (Articles I, II, III). Additionally, there is always the possibility to use a group of DL models to perform the same task as an ensembling approach to increase their performance (Article V, VI).

3.7 Evaluation assessment

The resulting predictions in all Articles were evaluated quantitatively by comparing pixels of the sampled ground throughout the data. We calculated metrics of precision, recall and F1-score, accuracy, and Matthews Correlation Coefficient (MCC) to evaluate the performance of the models. Precision is a measure of how many of the positive predictions are made correctly (true positives) (Table 4). At the same time, recall measures how many positive cases were correctly predicted over all the positive cases in the data. F1-Score is a measure that combines both precision and recall. A satisfactory F1 score means that there are few false positives and few false negatives. An F1 score is solid when having a value close to 1. Accuracy presents the number of correct predictions divided by the total number of predictions. A high accuracy is the result of overfitting. Overfitting means that the training data size is too small and lacks enough samples to represent all possible input data values accurately. Therefore, MCC scores were calculated (Table 5) in some cases (Article I, III). The MCC score is considered the most appropriate metric for comparing results, which produces a high score only if the prediction shows good results in all four confusion matrix categories true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values (Chicco and Jurman, 2020).

Table 4. Explained values of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values.

Prediction	Actual value	Type	Explanation
1	1	True Positive (TP)	Predicted Positive and was Positive
0	0	True Negative (TN)	Predicted Negative and was Negative
1	0	False Positive (FP)	Predicted Positive but was Negative
0	1	False Negative (FN)	Predicted Negative but was Positive

Table 5. The evaluated performance metrics precision, recall, F1-score, accuracy, and Matthew's correlation coefficient (MCC) were used in different combinations to assess evaluation.

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. Discussion - Main findings

This thesis is organized as a compilation of six articles that are either already published (Articles I, II, and III) or have been submitted (Articles, IV, V and VI) (Figure 4). In this chapter, I present a more thorough discussion regarding how the articles have contributed developing the literature of the use of Machine Learning in geoscience (Section 1.3).

4.1 Where are the outcrops?

Article I investigate the possibilities of using the Deep Learning (DL) cloud based and Fully Convolutional Neural Network (FCNN) model to efficiently and automatically identify bedrock from sediments to produce outcrop maps. The proposed framework uses a U-Net architecture and cloud-computing DL predictive algorithms. Moreover, the open code and cloud computation makes our framework accessible to anyone who wants to test the algorithm's predictions using their data from their study areas using cloud functions. Our DL approach proved to differentiate bedrock from sediments much more quickly than manual mapping and with high precision measured by an F1 score between 77% and 84% for DEM terrain derivatives. We also confirmed that using vegetated Landsat 8 data alone was insufficient for predictions on variable terrain (despite its proven utility in exposed mountainous areas from other studies). However, a combination of terrain derivates, Landsat 8, or optical datasets was not tested. We concluded that we could generate high-value predictions based on a DEM and its derivatives with a 10m resolution. Using DEM and its derivatives has essential implications for cost-effective geological mapping. Fieldwork would be more efficient because it could be targeted toward the areas where the predictions are uncertain, and mapping could potentially be done over larger areas within the same restricted resources. While this study did explore some of the limitations in applying such methods to all kinds of Norwegian contexts (such as bedrock variety and elevation differences). We successfully applied the pre-trained model to areas similar to the Farsund study area (coastal bedrock). Further

research is necessary to address these issues in expanding the study to different regions with different topography.

4.2 Can we map wetlands?

Article II contains a valuable mapping prediction over wetland areas in southern Norway. The classification of ecosystems and land cover classes based on satellite and repeated airborne remote sensing imagery offers some significant advantages over in situ and manual reference mapping: (1) it covers large areas and multiple years consistently and comparably; (2) it is scalable, to some extent (can extrapolate over time and space); (3) it is continually updateable and therefore allows for ongoing monitoring and surveillance (however, LiDAR data need to be continuously updated and therefore updates must not rely on this source); (4) it performs better than existing approaches for reference data that are validated by regional unused for training ground truth data. Whether the DL approach we proposed provides better and more effective classification results must be explored in future studies. A focus on change detection using the proposed methods, able to account for changes in wetland extents over time, would be particularly interesting and critical for effective wetland conservation in the future.

4.3 Landslides – model comparison (global vs. local)

Article III documents the performance of five different Machine Learning models, including three globally and two locally trained models, using various input data combinations, on the Jølster case study. The prediction results of the best performing globally (Prakash CNN) and locally (U-net CNN) trained models are shown in Figure 13.

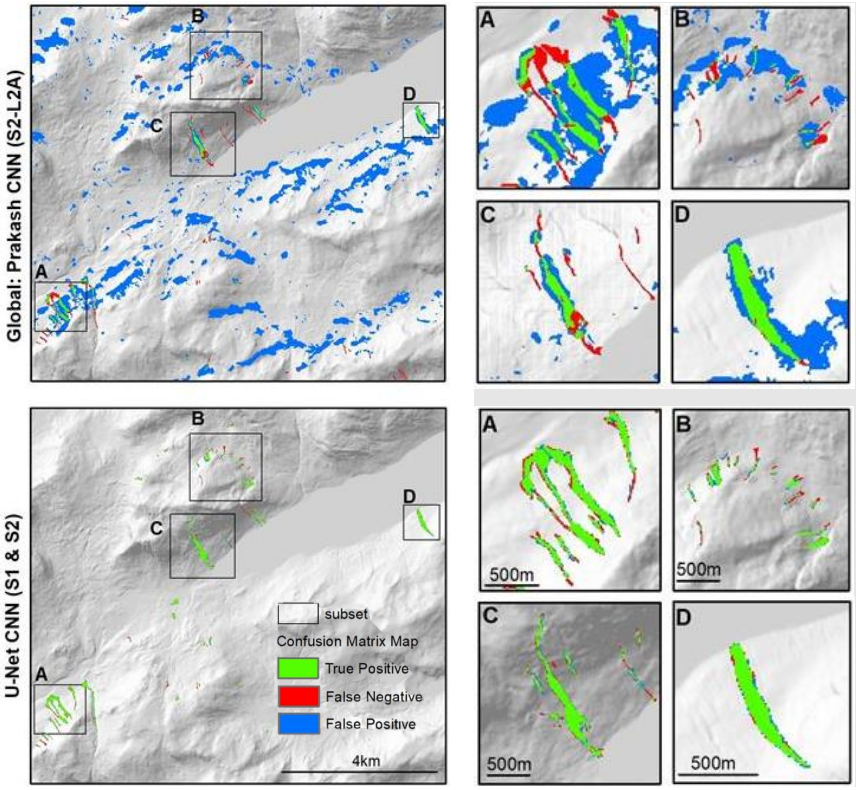


Figure 13: Best predictions of the globally trained models (top) and locally trained models (bottom).

Both the locally trained models, using multi-temporal image composites as input data, significantly outperformed the globally trained models, with the best Matthews Correlation Coefficient (MCC) (Section 3.7) score of 89%, achieved using the U-net CNN DL model, with a combination of Sentinel-1 and -2 images as input. Therefore, results indicated great potential for using a DL based model to improve landslide inventories in Norway, using multi-temporal image composites from Sentinel-1 and -2 as input data. However, this study did not address some major limitations as for example, applying such methods in different Norwegian conditions, such as snow cover, seasonal variations, and possible false positives from non-landslide-induced vegetation loss. To address these issues, further research is necessary in order to expand the study to a regional level. We recommend developing landslide detection models for mapping and inventory purposes, using Sentinel-1 and -2 as input data,

while for rapid detection in disaster response, Sentinel-1 can be used along with DL based detection models.

4.4 Landslide signatures in SAR images

In order to understand the reason for the discrepancy, or lack of similarity, in landslides signatures (Mondini et al., 2021), a systematic global investigation of a diverse selection of 30 case studies was performed. We used Sentinel-1 multi-temporal composite change images presented in **Article IV.** Out of the 30 cases investigated, 26 (86%) were identified as visible or partially visible. It was possible to identify repeating patterns between different cases with similar characteristics of morphology and land cover, by compiling the images from many cases. The main findings contribute to an improved understanding of how landslides appear in SAR backscatter intensity images. In addition, the scripts to produce the change images can rapidly produce change images for new areas after new Sentinel-1 images are available online within the Google Earth Engine catalogue. The globally spread dataset was a pre-study to understand landslides footprints better and is further applied in **Article V,** where we compared globally trained DL models and created an ensemble model of the best-performing single models.

4.5 Landslides globally trained ensemble model

Article V has the overarching objective to mitigate landslide hazards and by that minimize risks to society, by implementing different segmentation models with different parameters. Here we applied nine segmentation models with five different loss functions (Section 3.2.3), and two different learning rates on four different settings based on various combinations of Sentinel-1 and Setting-2 data. Our study explored the potential and performance of ensemble learning on globally trained DL models.

The combined use of Sentinel-1 and Sentinel-2 bands emerged as the most effective configuration. Results, performed by using only two bands based on Sentinel-2 data

only, show promising application of this model for developing a continuous landslide monitoring system. The outcomes also emphasize the significance of ensemble size, with an optimal range of 10 to 20 models providing the best compromise between calculated scores and computational efficiency.

The integration of various data sources, and the identification of context-specific applicability, presets the potential for innovative geohazard management strategies. By embracing the complexities of ensemble size, our research contributes to the ongoing efforts of enhancing disaster preparedness and minimize the impact of landslides on society.

4.6 Where are Ravines in Gully Landscape?

Article VI uses the same methodology proposed in Article III, combining the U-net model, Jupyter Notebook, and ArcPy library. This study further explored the possibilities of the ArcGIS Pro platform and its DL extension. It bases on a well-working pre-trained model able to delineate gullies in gully areas. Two different settings, in combination with different terrain derivates based on 1m resolution DEM, were performed after testing different input combinations. For the areas mapped previously, as for the future – yet unexplored areas, the combination of composite bands based on only three terrains derivates (Slope, Roughness, and Topological Position Index) is the most suitable. The ensemble process is based on the compilation of four predictions for the same area, synthesized into a single final map of gullies. The results are promising; the final produced map has a high visual predictive accuracy compared to the individual predictions. This could be a simple and efficient method to automatically delineate gullies when producing large-scale geological maps with similar geological settings. This study shows that the post-processing of combined prediction models significantly increased the applicability of DL models in geomorphological mapping routines. If necessary, adjustments in mapping routines in postprocessing and production of geological mapping could be established in future work.

5. Conclusion

This thesis explores the application of Deep Learning techniques in the field of geoscience. However, rapid changes bring new promising trends in this area, such as federated or collaborated learning, Generative Adversarial Networks (GAN), Explainable Artificial Intelligence, reinforcement learning, and transfer learning. However, the scope of this thesis and the main objective are to enhance efficiency in future mapping, by exploring simple and accessible uses of Deep Learning methods for the geoscientific community to apply in their regular tasks, without the necessity of extensive data science familiarity. Through comprehensive research and analyses, several key findings and contributions have led to the following conclusions:

First, the potential of DL in geoscience is demonstrated, showing its ability to support traditional manual approaches and offering new insights into complex geological mapping. The integration of DL algorithms has shown promising results in addressing classical mapping problems and understanding of geological hazards, such as landslides.

Second, it has been recognized that for geologists without expertise in data science, applying the results of DL research to their daily work can be challenging. The need for accessible tools, workflows, and guidelines tailored specifically for non-experts in data science has been emphasized, and some of these tools have been developed during my Ph.D. and presented in this work. By addressing this gap and providing user-friendly solutions, geoscientists can apply the power of DL to turn their daily tasks more efficient and impactful.

Third, Articles V and VI suggest that the improved results are obtained using ensemble models, which can be seen as an evolution in my research over the last three years.

Furthermore, the limitations of DL have been acknowledged, particularly in terms of replicability and reproducibility to new and unknown areas (specifically with

different types of terrain and vegetation). While DL models may not fully replicate human learning and adaptability, ongoing advancements in technology and trends in ML and DL, as mentioned above, offer opportunities for further improvements and refinement.

Looking ahead, the future of DL in geoscience appears promising. Continued research and efforts lead to the development of robust DL algorithms that integrate knowledge coming from multiple fields and try to overcome limitations. One of the well-known limitations is the need for enough good-quality data. The time used for data preparation is beneficial. Therefore, collecting, cleaning, and augmenting data are necessary. Another limitation can be linked to computational resources; this challenge can be overcome by optimizing the model and by using cloud-based platforms that provide affordable solutions for Deep Learning, as we illustrated in this thesis. Having reduced as much as possible many of the challenges that DL poses to the non-expert, this thesis shows how it is possible to enable geoscientists in applying the potential of DL in their daily work, ultimately enhancing everyone's understanding of geological processes. DL can improve hazard assessment and support new mapping strategies by applying the new techniques in everyday mapping tasks, as well as in postprocessing and production of geological mapping.
In conclusion, this thesis contributes to the broader understanding and application of DL in geoscience. I aim to highlight its potential to revolutionize efficiency and different angles on decision-making processes together with exploring different methods (cloud-computing, powerful GPU stationary computers). The main hope of this work is to join the forces of powerful technology, computational power, and expert-based knowledge, not to replace experts' abilities; on the contrary, researchers should bring their knowledge to the table, to develop models that can contribute to making geoscientists' days easier through automated pattern recognition. By allowing geoscientists to apply its results to their daily work, thereby bridging the gap between expert data science knowledge and geoscience expertise, this research aims to make significant strides in contributing to a more sustainable future for expert-based decisions.

Overall, the findings of this dissertation underscore the importance of continued exploration and integration of DL techniques in geoscience, emphasizing the need

for interdisciplinary collaboration between data scientists and geologists to drive innovation and maximize the potential of DL for the benefit of the geoscience community.

Further directions

This thesis shows the technical aspects and presents different platforms of Deep Learning models. The main aim is to assess the applicability of Deep Learning models in combination with geoscientific datasets, in order to make procedures as simple as possible for non-data scientists. DL presents many possibilities, and geoscience is finally starting to utilize them. However, there is still a lot of unknown and unexplored, and there are many suggestions on how Deep Learning could help in many aspects of geoscientific questions. However, the direction that I did not have time to explore is how this work can influence the work of geoscientists in the future. By working on this topic, I realized how helpful these tools can be in future mapping and solving geoscientific problems. However, the applicability and actual use of the models remains a big question. There is a possible need for change in the mindset. Mapping processes can benefit by using the potential for effective use of these tools. Further studies should address the applicability and actual use of the employed models. Questions like, what needs to be done to change the mapping habits of geoscientists? and how can we be more open to the computational power of Deep Learning models? remain open.

The thesis revealed some possible answers and solutions and opens possibilities for future collaboration. One of the still ongoing collaborations is with UC Berkeley. We are currently working on the development of a model for delineating lineaments (Article VII). This project has been awarded a grant (Peder Sæther grant) for the next two years, strengthen even more the collaboration between the Department of Geography at NTNU, and the Department of Astronomy and Statistics at UC Berkeley.

Article I supported the idea of creating a solution of bedrock delineation for the entire Norway at scale of 1:50.000 and testing chosen areas for higher quality at 1:10.000.

There is a future potential use of high-quality predictions in the fieldwork where bedrock is hidden under the vegetation. Also, the results of Article I led to a collaboration with the Centre for Exploration and Targeting School of Earth Sciences, University of Western Australia. This joint study aims to develop lithological classification guided by geophysical datasets with a multiple-point statistics algorithm. The expected predictions resulting from this project can be used as additional application areas in order to assess the method's performance.

Thanks to the ideas, conceptualized in Articles III and IV, we were reached out to by The Norwegian Water Resources and Energy Directorate (NVE) to apply DL methods to map out the landslides caused by the Hans extreme weather event (7.-9.8.2023 in Norway).

The finalization of Article V revealed a clear need for further work with a global dataset for developing a monitoring system using only Sentinel-1 predictor layers. We need to create a larger dataset with more study areas in order to prove that the ensemble model works efficiently when separating the dataset into herbaceous and forested areas. The next step is to develop a continuous monitoring system using Sentinel-2 predictors.

Epilogue

As this thesis journey concludes, it is evident that the exploration of Machine Learning techniques within geoscience has opened up new ideas about understanding and potential. The collaborative efforts and interdisciplinary approach that connect this research have shed light on the pathways of data-driven methodologies in exploring geological challenges.

The epilogue of this thesis serves as a reflection on the journey by tracing the evolution from the ideas to the summary of findings. Integrating Deep Learning methods and their application in geoscience has revealed the promise of improved performance and innovative views to the field.

However, it is equally important to acknowledge the limitations that remind us of the need for continuous research and development of ML and DL models within geoscience. The crossroads between data science and geoscience draw the fictional bridge between experts in data science and geoscientists. It is crucial to recognize that the compilation of this thesis is not the end of a journey but a milestone that marks the transition from academic exploration to broader landscape applications. The developed tools serve as viewpoints guiding geologists and researchers toward a future where the integration of ML techniques facilitates more efficient and impactful work.

In closing, the epilogue invites all those who read this thesis to join in the ongoing progress. As technology evolves and new horizons emerge, the journey to fully benefit from the use of Machine Learning in geoscience is closer than we can imagine. This thesis is a tiny puzzle in the scientific progress, threading dedication, collaboration, and belief that interdisciplinarity in Machine Learning and geoscience holds the power to see the unseen and shape the bright future of informed decisions and enhanced understanding.

References

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *Journal of Big Data* (Vol. 8, Issue 1). Springer International Publishing. https://doi.org/10.1186/s40537-021-00444-8

ArcGIS Pro. (2022). *Export Training Data For Deep Learning (Image Analyst)*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/export-training-data-for-deep-learning.htm

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical Systems and Signal Processing*, *147*, 107077. https://doi.org/10.1016/j.ymssp.2020.107077

Belgiu, M., & Drăguţ, L. (2014). Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *96*, 67–75. https://doi.org/10.1016/j.isprsjprs.2014.07.002

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*(October 2021). https://doi.org/10.1016/j.cageo.2021.104952

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*, *23*(5), 408–422. https://doi.org/10.1016/j.tics.2019.02.006

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215. https://doi.org/10.1214/ss/1009213726

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. https://doi.org/10.1201/9781315139470

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. https://doi.org/10.1080/15228053.2020.1766224

CapellaSpace. (n.d.). *Sentinel-1 Polarization*. https://support.capellaspace.com/hc/en-us/articles/360044738831-Sentinel-1-Polarization#:~:text=VV is a mode that,vertical and horizontal wave patterns.

Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). Machine learning as a tool for geologists. *Leading Edge*, *36*(3), 215–219. https://doi.org/10.1190/tle36030215.1

Chaurasia, A., & Culurciello, E. (2018). LinkNet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE Visual Communications and Image Processing, VCIP 2017*, *2018-Janua*, 1–4. https://doi.org/10.1109/VCIP.2017.8305148

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., … Carranza, E. J. M. (2019). Deep learning and its application in geochemical mapping. *Journal of the Royal Society Interface*, *192*(March), 1–14. https://doi.org/10.1098/rsif.2017.0387

Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P., & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. *Npj Computational Materials*, *8*(1). https://doi.org/10.1038/s41524-022-00734-6

Daume III, H., & Marcu, D. (2006). Domain Adaptation for Statistical Classifiers. *Journal of Artificial Intelligence Research*, *26*, 101–126. https://doi.org/10.1613/jair.1872

Dwivedi, H. (2023). *How to Use Google Colab for Deep Learning – Complete Tutorial*. https://neptune.ai/blog/how-to-use-google-colab-for-deep-learning-complete-tutorial

Emden, Z., Calantone, R. J., & Droge, C. (2006). Collaborating for New Product Development: Selecting the Partner with Maximum Potential to Create Value. *Journal of Product Innovation Management*, *23*(4), 330–341. https://doi.org/10.1111/j.1540-5885.2006.00205.x

ESRI. (2023). *Pretrained Deep Learning Models*. https://www.esri.com/en-us/arcgis/deep-learning-models

Fan, T., Wang, G., Li, Y., & Wang, H. (2020). Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, *8*, 179656–179665. https://doi.org/10.1109/ACCESS.2020.3025372

Geisser, S. (1975). The Predictive Sample Reuse Method with Applications. *Journal of the American Statistical Association*, *70*(350), 320–328. https://doi.org/10.1080/01621459.1975.10479865

Ghasemi, Y., Jeong, H., Choi, S. H., Park, K. B., & Lee, J. Y. (2022). Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry*, *139*, 103661. https://doi.org/10.1016/j.compind.2022.103661

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning

convolutional neural networks for landslide detection. *Remote Sensing*, *11*(2). https://doi.org/10.3390/rs11020196

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., … Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, *2*, 225–250. https://doi.org/10.1016/j.aiopen.2021.08.002

Hardesty, L. (2017). *Explained: Neural networks (Ballyhooed artificial-intelligence technique known as "deep learning" revives 70-year-old idea.)*. MIT News. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated Machine Learning Methods, Systems, Challenges. In *Automated Deep Learning Using Neural Network Intelligence*. https://doi.org/10.1007/978-1-4842-8149-9_2

IBM Data and AI Team. (2023). *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?* https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/

Jadon, S. (2020). A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*. https://doi.org/10.1109/CIBCB48159.2020.9277638

Jarna, A., Baeten, N. J., Elvenes, S., Bellec, V. K., Thorsnes, T., & Diesing, M. (2019). Semi-automatic versus manual mapping of cold-water coral carbonate mounds located offshore Norway. *ISPRS International Journal of Geo-Information*, *8*(1). https://doi.org/10.3390/ijgi8010040

Jia, D., Cheng, C., Song, C., Shen, S., Ning, L., & Zhang, T. (2021). A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions. *Remote Sensing*, *13*(4), 1–33. https://doi.org/10.3390/rs13040645

Jiang, N., Chen, J., Zhou, R. G., Wu, C., Chen, H., Zheng, J., & Wan, T. (2020). PAN: Pipeline assisted neural networks model for data-to-text generation in social internet of things. *Information Sciences*, *530*, 167–179. https://doi.org/10.1016/j.ins.2020.03.080

JupyterHub - Sphinx. (2023). *Real-time collaboration without impersonation*. https://jupyterhub.readthedocs.io/en/stable/tutorial/collaboration-users.html

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

Kartverket. (2016). *Nasjonal detaljert høydemodell*. https://www.kartverket.no/geodataarbeid/nasjonal-detaljert-hoydemodell

Kartverket. (2021). *Høgdedata og djupnedata*. https://www.kartverket.no/en/api-and-data/terrengdata

Kawulok, M., Benecki, P., Piechaczek, S., Hrynczenko, K., Kostrzewa, D., & Nalepa, J. (2020). Deep Learning for Multiple-Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters*, *17*(6), 1062–1066. https://doi.org/10.1109/LGRS.2019.2940483

Laio, F., & Torino, P. (2011). *2.18 Statistical Hydrology* `.

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. *Environmental Modelling and Software*, *112*(March 2018), 16–22. https://doi.org/10.1016/j.envsoft.2018.11.004

Li, X., Lai, T., Wang, S., Chen, Q., Yang, C., & Chen, R. (2019). Weighted feature pyramid networks for object detection. *Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, 1500–1504. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217

Lindsay, E., Frauenfelder, R., Rüther, D., Nava, L., Rubensdotter, L., Strout, J., & Nordal, S. (2022). Multi-Temporal Satellite Image Composites in Google Earth Engine for Improved Landslide Visibility: A Case Study of a Glacial Landscape. *Remote Sensing*, *14*(10). https://doi.org/10.3390/rs14102301

Liu, Cai, Wang, W., Tang, J., Wang, Q., Zheng, K., Sun, Y., Zhang, J., Gan, F., & Cao, B. (2023). A deep-learning-based mineral prospectivity modeling framework and workflow in prediction of porphyry–epithermal mineralization in the Duolong ore District, Tibet. *Ore Geology Reviews*, *157*(April), 105419. https://doi.org/10.1016/j.oregeorev.2023.105419

Liu, Chao, Gao, C., Xia, X., Lo, D., Grundy, J., & Yang, X. (2022). On the Reproducibility and Replicability of Deep Learning in Software Engineering. *ACM Transactions on Software Engineering and Methodology*, *31*(1), 1–46. https://doi.org/10.1145/3477535

Liu, M., & Grana, D. (2019). Accelerating geostatistical seismic inversion using TensorFlow: A heterogeneous distributed deep learning framework. *Computers and Geosciences*, *124*(October 2018), 37–45. https://doi.org/10.1016/j.cageo.2018.12.007

Liu, X., Deng, Z., & Yang, Y. (2019). Recent progress in semantic image segmentation. *Artificial Intelligence Review*, *52*(2), 1089–1106. https://doi.org/10.1007/s10462-018-9641-3

Ma, Z., & Mei, G. (2021). Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth-Science Reviews*, *223*, 103858. https://doi.org/10.1016/j.earscirev.2021.103858

Mary, S. P., Ankayarkanni, Nandini, U., Sathyabama, & Aravindhan, S. (2020). A Survey on Image Segmentation Using Deep Learning. *Journal of Physics: Conference Series*, *1712*(1). https://doi.org/10.1088/1742-6596/1712/1/012016

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, *207*(May), 103225. https://doi.org/10.1016/j.earscirev.2020.103225

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Mondini, A. C., Guzzetti, F., Chang, K.-T., Monserrat, O., Martha, T. R., & Manconi, A. (2021). Landslide failures detection and mapping using Synthetic Aperture Radar: Past, present and future. *Earth-Science Reviews*, *216*, 103574. https://doi.org/10.1016/j.earscirev.2021.103574

Morales, E. F., & Escalante, H. J. (2022). A brief introduction to supervised, unsupervised, and reinforcement learning. In *Biosignal Processing and Classification Using Computational Learning and Intelligence* (pp. 111–129). Elsevier. https://doi.org/10.1016/B978-0-12-820125-1.00017-8

Mousavi, S. S., Schukat, M., & Howley, E. (2018). Deep Reinforcement Learning: An Overview. *Lecture Notes in Networks and Systems*, *16*, 426–440. https://doi.org/10.1007/978-3-319-56991-8_32

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, *115*(5), 1145–1161. https://doi.org/10.1016/j.rse.2010.12.017

Nava, L., Monserrat, O., & Catani, F. (2022). Improving Landslide Detection on SAR Data Through Deep Learning. *IEEE Geoscience and Remote Sensing Letters*, *19*(Ml). https://doi.org/10.1109/LGRS.2021.3127073

Okon, E. I., & Appah, D. (2021, August 2). *Application of Machine Learning Techniques in Reservoir Characterization*. https://doi.org/10.2118/208248-MS

Oracle. (2023). *Using Jupyterhub*. https://docs.oracle.com/en-us/iaas/Content/bigdata/notebooks-jupyterhub.htm#:~:text=Using Jupyterhub 1 Prerequisites Before Jupyterhub can be,Permissions ... 8 Integrating with Object Storage

Otterlo, M., & Wiering, M. (2012). Reinforcement Learning and Markov Decision Processe. In *Adaptation, Learning, and Optimization* (Volume 12).

Pandey, M., Fernandez, M., Gentile, F., Isayev, O., Tropsha, A., Stern, A. C., & Cherkasov, A. (2022). The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, *4*(3), 211–221. https://doi.org/10.1038/s42256-022-00463-x

PapersWithCode. (2023). *Semantic Segmentation*. https://paperswithcode.com/task/semantic-segmentation

Parzen, E., & Mukhopadhyay, S. (2012). *Modeling, dependence, classification, united statistical science, many cultures*. 31. http://arxiv.org/abs/1204.4699

Prakash, N., Manconi, A., & Loew, S. (2021). A new strategy to map landslides with a

generalized convolutional neural network. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-89015-8

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. *Ecological Informatics*, *66*(April), 101474. https://doi.org/10.1016/j.ecoinf.2021.101474

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rzhanov, Y., Fonseca, L., & Mayer, L. (2012). Construction of seafloor thematic maps from multibeam acoustic backscatter angular response data. *Computers and Geosciences*, *41*, 181–187. https://doi.org/10.1016/j.cageo.2011.09.001

Sameen, M. I., & Pradhan, B. (2019). *Landslides: Theory, Practice and Modelling* (Vol. 50). http://link.springer.com/10.1007/978-3-319-77377-3

Sarker, I. H. (2021a). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Sarker, I. H. (2021b). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. In *SN Computer Science* (Vol. 2, Issue 6). Springer. https://doi.org/10.1007/s42979-021-00815-1

Schwab, K. (2016). *The Fourth Industrial Revolution*. Currency; Illustrated edition (January 3, 2017).

ScienceDirect. (2018). *Normalized Difference Vegetation Index*. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/normalized-difference-vegetation-index

Shaikh, J. (2019). *GPU became a widely accepted method for training deep learning models*. https://www.analyticsvidhya.com/blog/2017/05/gpus-necessary-for-deep-learning/

Shmueli, G., van Bruggen, Gerrit; Lilien, G. L. . K. M., & Parzen, E. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Shrestha, S., Binod Bhatta, Talchabhadel, R., & Virdis, S. G. P. (2022). Integrated assessment of the landuse change and climate change impacts on the sediment yield in the Songkhram River Basin, Thailand. *Catena*, *209*(P2), 105859. https://doi.org/10.1016/j.catena.2021.105859

Sinz, F. H., Pitkow, X., Reimer, J., Bethge, M., & Tolias, A. S. (2019). Engineering a Less Artificial Intelligence. *Neuron*, *103*(6), 967–979. https://doi.org/10.1016/j.neuron.2019.08.034

Standford. (2023). *History: The 1980's to the present*. Standford - Neural Networks. https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-

networks/History/history2.html

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*(March), 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001

Tan, J., Zuo, J., Xie, X., Ding, M., Xu, Z., & Zhou, F. (2021). MLAs land cover mapping performance across varying geomorphology with Landsat OLI-8 and minimum human intervention. *Ecological Informatics*, *61*(November 2020), 101227. https://doi.org/10.1016/j.ecoinf.2021.101227

Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine learning and landslide studies: recent advances and applications. *Natural Hazards*, *114*(2), 1197–1245. https://doi.org/10.1007/s11069-022-05423-7

Tobergte, D. R., & Curtis, S. (2013). The ArcGIS Imagery Book. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Tseng, M. H., Chen, S. J., Hwang, G. H., & Shen, M. Y. (2008). A genetic algorithm rule-based approach for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(2), 202–212. https://doi.org/10.1016/j.isprsjprs.2007.09.001

Vollrath, A., Mullissa, A., & Reiche, J. (2020). Angular-based radiometric slope correction for Sentinel-1 on google earth engine. *Remote Sensing*, *12*(11), 1–14. https://doi.org/10.3390/rs12111867

Wallis, W. A. (1980). The Statistical Research Group, 1942–1945. *J. Amer. Statist. Assoc.*, *75*(370), 320–335. https://doi.org/81j:62004

Wang, L., Marzahn, P., Bernier, M., & Ludwig, R. (2018). Mapping permafrost landscape features using object-based image classification of multi-temporal SAR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*, 10–29. https://doi.org/10.1016/j.isprsjprs.2018.03.026

Wang, X., Zhao, Y., & Pourpanah, F. (2020). Recent advances in deep learning. *International Journal of Machine Learning and Cybernetics*, *11*(4), 747–750. https://doi.org/10.1007/s13042-020-01096-5

Weier, J., & Herring, D. (2000). Measuring Vegetation (NDVI & EVI). *NASA Earth Observatory, Washington DC.*

Wu, H., & Levinson, D. (2021). The ensemble approach to forecasting: A review and synthesis. *Transportation Research Part C: Emerging Technologies*, *132*(September), 103357. https://doi.org/10.1016/j.trc.2021.103357

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

Yang, Z., Li, L., Xu, X., Kailkhura, B., Xie, T., & Li, B. (2022). on the Certified Robustness for Ensemble Models and Beyond. *ICLR 2022 - 10th International Conference on Learning Representations*.

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal*

Processing: A Review Journal, *126*, 103514. https://doi.org/10.1016/j.dsp.2022.103514

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 6230–6239. https://doi.org/10.1109/CVPR.2017.660

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation BT - Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. *Miccai*, *11045*(2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5

Zuo, Z., Zhao, K., Gebauer, J., Skinner, R., Haupt, R., Kremer, L., van der Pal, H., Michel, G., Armstrong, G. T., Hudson, M. M., Hjorth, L., Lehnert, H., Langer, T., & Sarker, I. H. (2021). The more multidisciplinary the better? - The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of Informetrics*, *2*(9), 736–756. https://doi.org/10.1530/EC-22-0083

6. Publications

This Section contains reprints of the publications which constitute the thesis. The publications are formatted to fit the format of the thesis.

Article I.

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

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

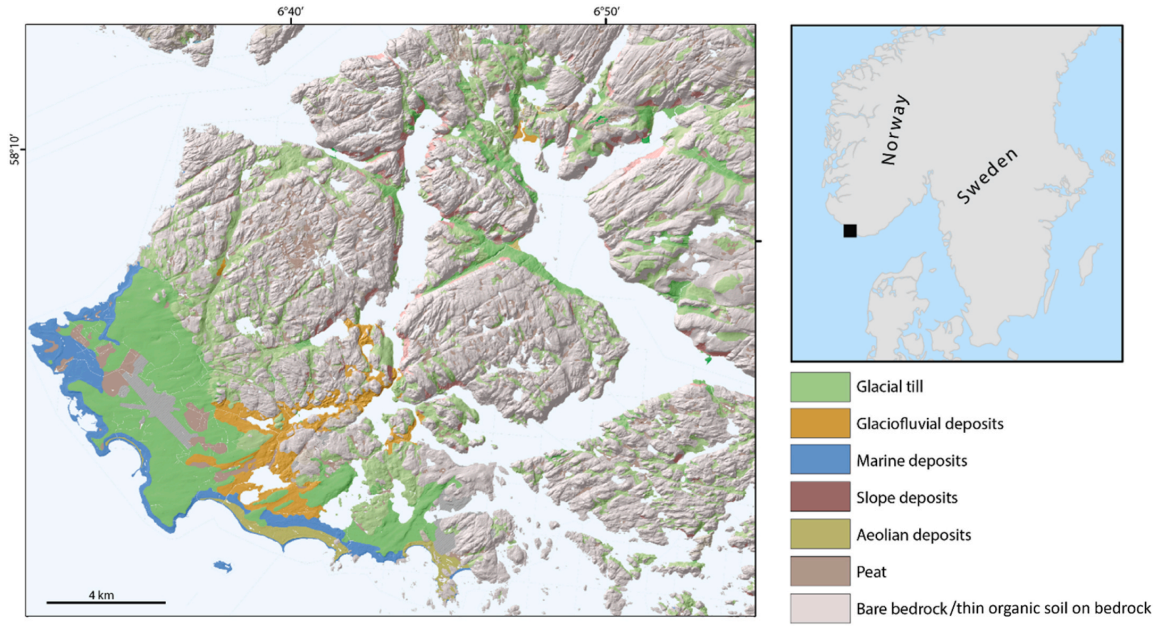


Fig. 1. Bare bedrock and different sediment types superposed on a hillshade in the investigated area (image closely corresponds with study area). Large amounts of sediments (till, glaciofluvial) were deposited during the last glaciation. Marine, slope, aeolian deposits and peat were formed in post-glacial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

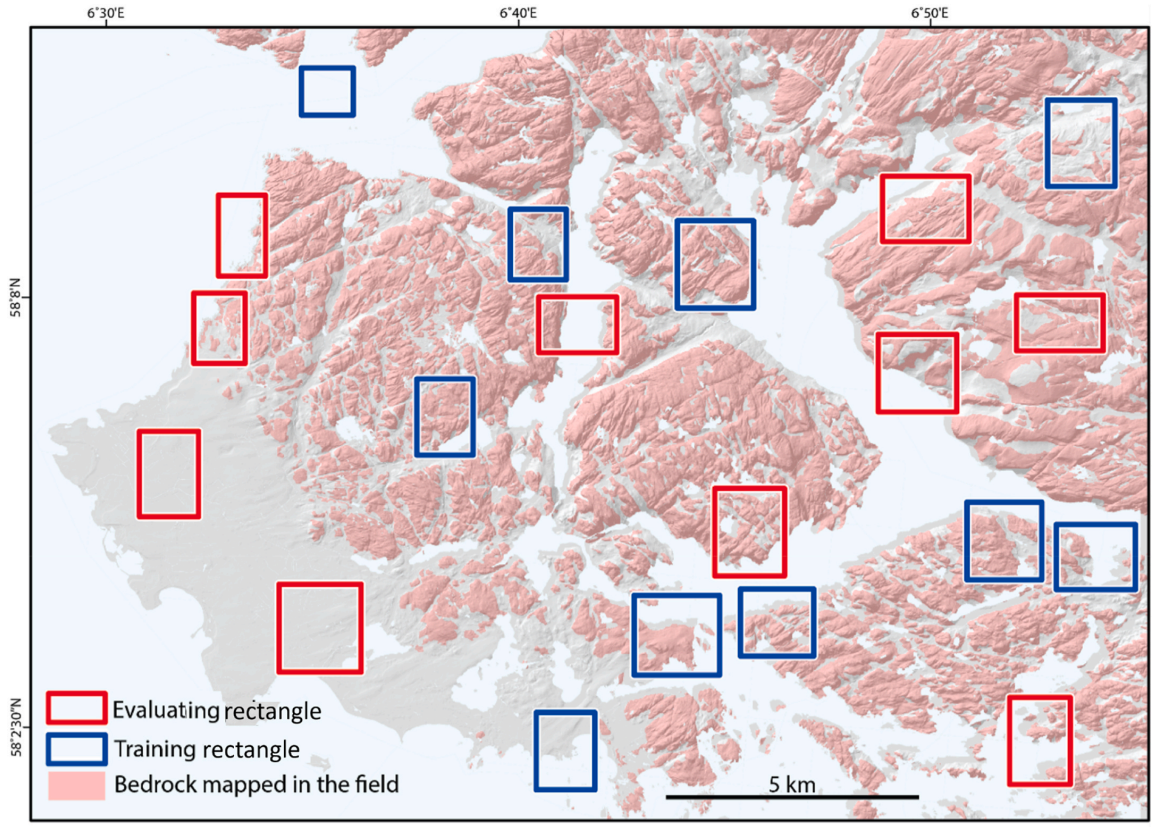


Fig. 2. Bare rock (and thin organic soil on bedrock) ground truth dataset based on a 1:50 000 quaternary map (Fredin et al., 2015) with training (blue) and evaluating (red) rectangles superposed. Example of 1 case (Setting 1). (For interpretation of the references to colour in this figure legend, the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

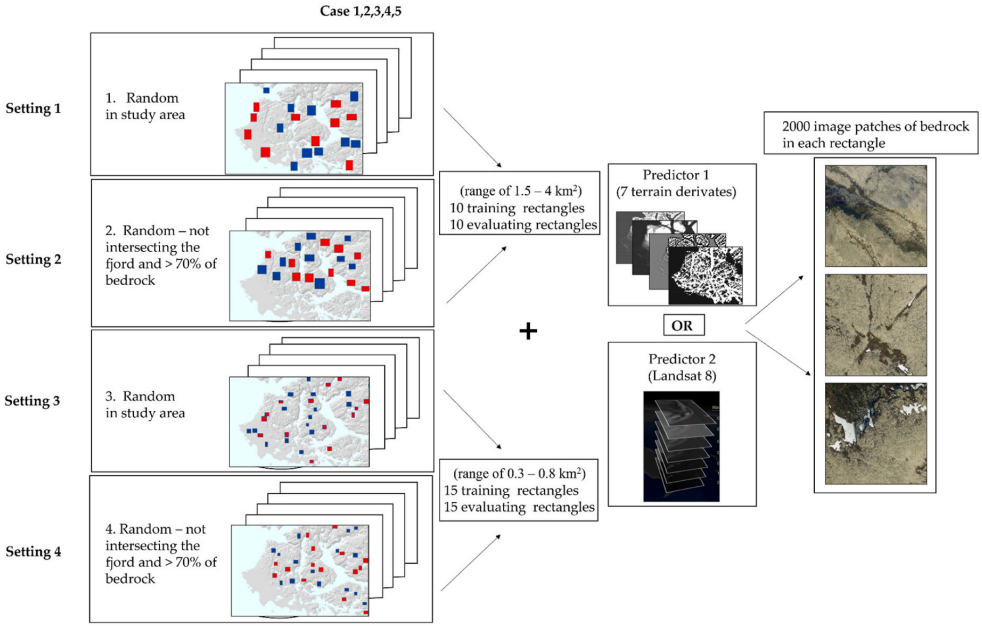


Fig. 3. Sampling strategy: four different settings were applied to produce five sets (cases) each of training and evaluating rectangles, combined with one of two predictors, creating 2000 image patches of bedrock for further calculation from each rectangle where possible (see 3.2.2 for more details on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

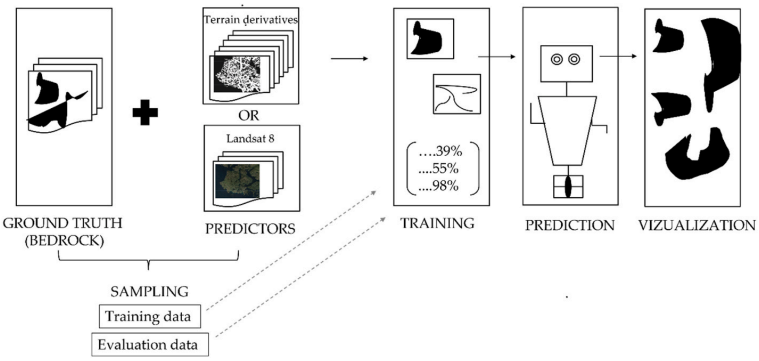


Fig. 4. Workflow applied in GEE: (1) sampling, (2) training and evaluating, (3) training the model, resulting in metrics on accuracy, (4) prediction, and (5) visualization in GEE.

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

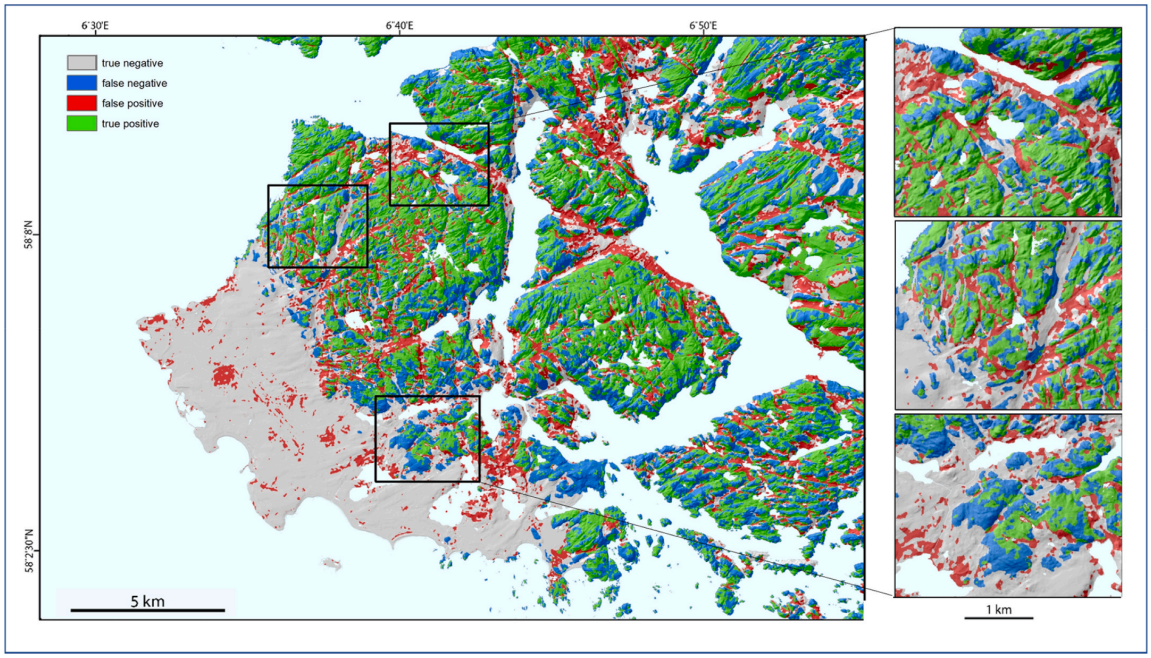


Fig. 5. Landsat 8-based prediction: The map shows the spatial distribution of the modelled prediction of bedrock areas in comparison with the ground truth bedrock areas from the geological map (Setting 1, case 4 with 99% training match and 86% evaluating) visualized as true positive (green), false positive (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

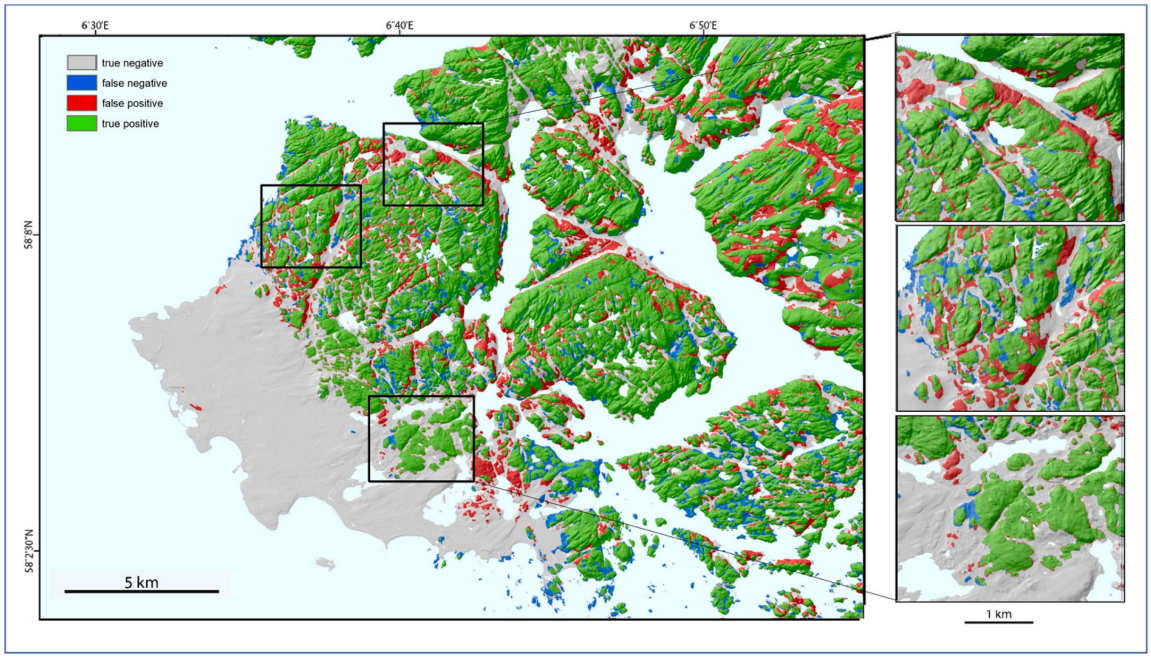


Fig. 6. Spatial distribution of the DEM terrain derivatives-based prediction. The map shows the modelled prediction of bedrock areas in comparison with the ground truth bedrock areas from the geological map (Setting 1, case 5 with 97% training match and 83% evaluating) visualized as true positive (green), 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.

Drăguț, 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.

Drăguț, 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.

Article II.

Delineation of Wetland Areas in South Norway from Sentinel-2 Imagery and LiDAR Using TensorFlow, U-Net, and Google Earth Engine

Vegar Bakkestuen [1], Zander Venter [1], Alexandra Jarna Ganerød [2,3], Erik Framstad [1]

1 Norwegian Institute for Nature Research, Sognsveien, 0855 Oslo, Norway
2 Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway
3 Geological Survey of Norway (NGU), 7040 Trondheim, Norway

remote sensing

MDPI

Article

Delineation of Wetland Areas in South Norway from Sentinel-2 Imagery and LiDAR Using TensorFlow, U-Net, and Google Earth Engine

Vegar Bakkestuen [1,*], Zander Venter [1], Alexandra Jarna Ganerød [2,3] and Erik Framstad [1]

[1] Norwegian Institute for Nature Research, Sognsveien, 0855 Oslo, Norway
[2] Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway
[3] Geological Survey of Norway (NGU), 7040 Trondheim, Norway
* Correspondence: vegar.bakkestuen@nina.no

Abstract: Wetlands are important habitats for biodiversity and provide ecosystem services such as climate mitigation and carbon storage. The current wetland mapping techniques in Norway are tedious and costly, and remote sensing provides an opportunity for large-scale mapping and ecosystem accounting. We aimed to implement a deep learning approach to mapping wetlands with Sentinel-2 and LiDAR data over southern Norway. Our U-Net model, implemented through Google Earth Engine and TensorFlow, produced a wetland map with a balanced accuracy rate of 90.9% when validated against an independent ground-truth sample. This represents an improvement upon manually digitized land cover maps in Norway, which achieved accuracy rates of 46.8% (1:50,000 map) and 42.4% (1:5000 map). Using our map, we estimated a total wetland coverage area of 12.7% in southern Norway, which is double the previous benchmark estimates (5.6%). We followed an iterative model training and evaluation approach, which revealed that increasing the quantity and coverage of labeled wetlands greatly increases the model performance. We highlight the potential of satellite-based wetland maps for the ecosystem accounting of changes in wetland extents over time—something that is not feasible with traditional mapping methods.

Keywords: remote sensing; neural network; deep learning; land cover; wetland; machine learning

Citation: Bakkestuen, V.; Venter, Z.; Ganerød, A.J.; Framstad, E. Delineation of Wetland Areas in South Norway from Sentinel-2 Imagery and LiDAR Using TensorFlow, U-Net, and Google Earth Engine. *Remote Sens.* **2023**, *15*, 1203. https://doi.org/10.3390/rs15051203

Academic Editor: Isam Shahrour

Received: 22 January 2023
Revised: 17 February 2023
Accepted: 18 February 2023
Published: 22 February 2023

1. Introduction

Wetlands are ecosystems that are permanently or periodically saturated or inundated with water and include habitats in the transition between terrestrial and freshwater or marine ecosystems. They offer a wide range of ecosystem services, including water purification, flood control and carbon sequestration [1,2]. Wetlands hold the highest density of soil carbon among terrestrial ecosystem types [3]. This makes them efficient and cost-effective nature-based solutions to climate change, as they sequester atmospheric carbon and are important in the long-term storage of carbon [4]. Despite this, wetlands are constantly under pressure from human activities [5–7], and wetland habitat losses due to land-use change often lead to decreases in local biodiversity and altered hydrological cycles, which result in carbon emissions.

It has been estimated that about 9% of the mainland of Norway is covered by open wetlands [8]. Open peatlands, which include open bogs and fens, are the wetland types with the highest coverage rate in Norway (approximately 8% of land cover; [8]. However, the coverage of wetlands over Norway remains uncertain, as different data sources yield different estimates (e.g., 3.8% in 1:5000 maps [9] vs. 6.0% in 1:50,000 [10] and vs. 5.3% in official statistics [11]. This is partly because previous mapping efforts have been based on manual in situ mapping procedures, which require substantial financial investment and adopt differing definitions of wetland habitats. There are, however, several concerns regarding these map layers, including: (1) the landcover classification system suitable

for topographic mapping at a scale of 1:50,000 is coarse; (2) small polygons are kept or removed selectively for cartographic reasons; (3) the main method for the 1:50,000 layer is aerial photo interpretation without field validation; and (4) the 1:5000 layer is not mapped in mountainous areas, where many wetlands are excluded from national estimates [8]. Furthermore, employing fieldworkers to digitize habitat types introduces a sampler bias, which makes the resulting map vulnerable to spatial and temporal inconsistencies [12–15] Mapping instructions and methods can also change over time, making it difficult to discern whether changes in wetland cover are real or merely an artifact of changes in mapping methodology. Apart from the mostly single-timepoint aerial coverage estimates, there is little active monitoring and surveillance of wetland conditions or changes over time. However, the active monitoring and annual or biennial mapping of wetlands will become important given the revised management plan for the restoration of wetlands in 2021 [16].

Due to the importance of wetlands for ecosystem services, and the threats of anthropogenic disturbance, monitoring and mapping wetlands is important to aid in their conservation. The remote sensing of wetlands has seen increased attention in recent years, with many large-scale studies illustrating the ability to map wetlands from satellite images using machine learning techniques [17,18]. The current wetland mapping techniques in Norway are tedious and costly, but satellite and airborne remote sensing methods including optical sensing, radar sensors, and LiDAR provide an opportunity for large-scale mapping and ecosystem accounting. So far, Norway has not supplemented in situ wetland inventories and mapping with remote sensing data and machine learning classification workflows, which stands in contrast to countries such as Canada [19]. Perhaps the most important thing to note is that Norway until recently has lacked access to high-quality ground truths, which are necessary for training, calibrating, and validating satellite-based wetland maps [20].

Few studies have applied deep learning models to wetland classification [17], and to the best of our knowledge only [21] has done so using a fusion of Sentinel-1 and Sentinel-2 data. In a meta-analysis of more than 200 publications, [22] found that the median accuracy for classifying land use and land cover using deep learning models was 91%, and that there was no other tree- or kernel-based classifier that achieved a median accuracy over 90%. Therefore, although less sophisticated models are more efficient in terms of the training data requirement and inference speeds, deep learning models ultimately achieve higher accuracies for land cover classification.

In the literature review by [18], which also included the data base from [17], we found very few examples on the deep learning classification of wetlands by use of Sentinel-1 or -2. One exemption was [21], who used deep learning for land use and land cover (LULC) classification based on U-Net based on Sentinel-1 and -2. Another exemption was also [23], who used deep U-Net convolutional networks and Sentinel-2 imagery for coastal wetland classification. By comparing deep learning and shallow learning for large-scale wetland classification in Alberta, Canada, [24] found that the accuracy of the convoluted neural networks (CNN) product was about 80%. [24] also concluded that CNNs may be better able to capture natural complexities within wetland classes. In this study, we explored the potential of using deep learning to map wetlands at a regional scale using U-Net convolutional networks and cloud computation on Sentinel-2 and LiDAR data.

To distinguish wetlands from other land use and land cover types (LULC), another type of deep learning method called semantic segmentation is needed. In the process of semantic segmentation, the purpose is to delineate areas with the desired content, in our approach wetlands, from all other unwanted information (i.e., the other LULC classes). This type of deep learning with semantic segmentation often requires the manual digitization of wetlands as polygons, preferably as a wall-to-wall approach in limited areas where training and validation image patches can be cut out and made available for a deep learning training process.

Deep learning has recently been proposed as one of the most promising methods for classifying LULC classes from remote sensed imagery [25]. It has been argued that deep learning can also solve more difficult classification issues such as wetland delin-

eation [26]. Despite the potential for deep learning models, the most common machine learning framework used to generate wetland maps has been decision trees (e.g., Random Forest), followed by support vector machines (SVM) [17]. However, the type of machine learning model adopted has so far had no discernible effect on map accuracy [18].

To meet the demand for wall-to-wall, frequently updatable wetland maps in southern Norway, and to further the international state of wetland remote sensing, our aim was to generate a deep learning-based map of wetlands in Norway using Sentinel-2 imagery and LiDAR data. The integration between Google Earth Engine (GEE) [27] and TensorFlow in the cloud made it possible to apply and train a U-Net model on Sentinel-2 imagery and LiDAR. In this paper, we describe a pipeline on how to classify a land cover type (wetland) using Google Colab [28], TensorFlow, and GEE. Further, we discuss the accuracy and potential of this approach related to other methods.

2. Methods

2.1. Study Area

The study area included the mainland of southern Norway, from 58 to 63°N and from 5 to 13°E, from sea level to 2469 m (Figure 1). It spanned a wide range of environmental gradients relevant to wetland biogeography, including temperature and precipitation regime, solar radiation, geology, and topography gradients [29,30].

Norway is located on the western border of the Baltic Shield, dominated by Precambrian rocks and the Caledonian Mountains with sedimentary, metamorphic bedrock [31]. In July, the average temperatures vary from 17 °C around the Oslo Fjord to 0 °C in the highest peaks of the Jotunheimen Mountain area (Norwegian Meteorological Institute 2023). The monthly average temperatures for January range from −15 °C in Jotunheimen to +3 °C in the outer coastal areas in the south and west. The outer and central areas on the west coast receive the most precipitation, with more than 2000 mm in several areas; however, the interior parts east of the central mountain range receive as little as 200 to 400 mm of annual average rainfall.

In Norway, there are estimated to be approximately 20,500 square kilometers of wetland, i.e., about 5.3% of the total land area [11]. Other major land cover types are open land areas, which includes mountains (38%), forests (37%), fresh water (5%), and agricultural land (3%) [32]. The variation in the mire peatland types in Norway is high and unique, even in a global context [29]. The habitat classification system Nature in Norway 2.2.0 [33] (NiN, in English EcoSyst) includes 13 main habitat types for the wetland ecosystem. These are open fen (V1), bog (V3), mire and swamp forest (V2), tidal and alluvial swamp forest (V8), wet snowbed and snowbed spring (V4 and V5), spring (V6), arctic permafrost wetland (V7), semi-natural fen (V9) and semi-natural wet meadow (V10), peat quarry (V11), drained mire (V12), and artificial wetlands (V13) (see details in https://www.artsdatabanken.no/Pages/172028/Vaatmarkssystemer, in Norwegian) (accessed on 5 January 2023). In our study, we did not include the mire and swamp forest classes (classes V2 and V8). Open peatlands, which constitute bogs and fens, are the most common wetland types in Norway (accounting for approximately 95% of open wetland cover; [8].

In the rectangles shown in Figure 1, there is wall-to-wall digitization of wetland polygons as ground-truth data. Some red and blue rectangles are smaller than others, mainly because they have a higher density and amount of wetland polygons, meaning a lot more effort was needed to map wall-to-wall in these areas. The green rectangles are generally smaller than the red and blue and were created to improve the first deep learning model.

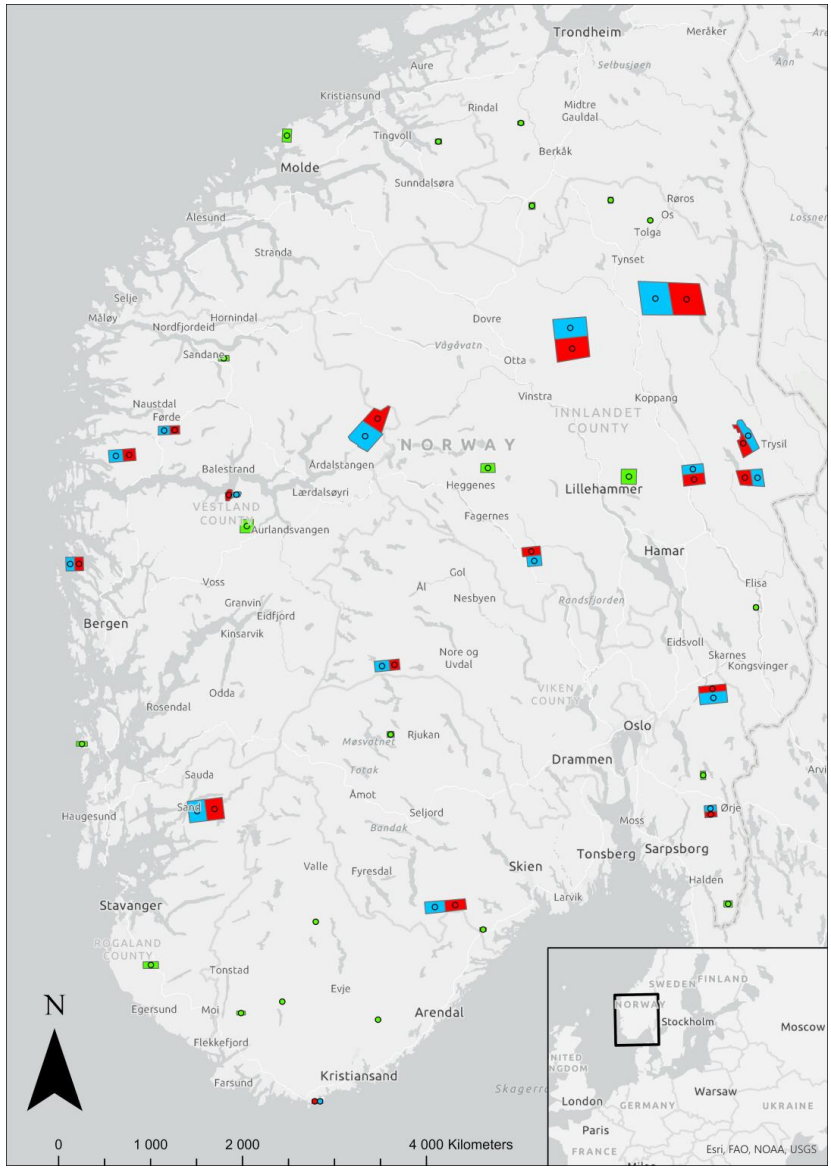


Figure 1. Study area and rectangles for collection of image patches for evaluation (red), original training set (blue), and additional training set (green). The study area is shown in the North-European map in the lower right corner.

2.2. Preparation of the Predictor Stack

The regional classification of wetlands from Sentinel satellite imagery using deep learning requires access to a large-capacity platform and infrastructure for storing and analyzing large amounts of data. TensorFlow is a software library based on open-source machine learning developed by Google. Pytorch and Keras are examples of other frameworks and libraries. The GEE was used for storing, customizing, and exporting our input data for cloud-based deep learning modeling. Customized files in TensorFlow format

(TFRecord-format) were then exported to Google Cloud Storage and used as an input for a virtual machine to process the deep learning scripts in Google Colab.

We processed all Sentinel-2 optical scenes over Norway during 1 August to 31 October 2020. The Sentinel-2 data were used to derive spectrotemporal features as predictor variables [34]. The spectrotemporal features were used to capture both the spectral and temporal (e.g., phenology including autumn color senescence [35,36]) characteristics of land cover classes to offer enhanced model prediction accuracy compared to single-timepoint image classification [37,38].

After trial and error, we selected 3 Sentinel-2 bands and 10 indices, as well as the mean canopy height model from LiDAR (accessed from hoydedata.no) as a substitute for tree height, as our explanatory variables. Using the cloud-masked Sentinel-2 imagery [34], we derived 13 bands or indices (Table 1). For most of these, we used median mosaic values, except for the normalized difference vegetation index (NDVI), where we used the 25th percentile mosaics across the three-month time stack of images.

Table 1. List of input variables from Sentinel-2 in the deep learning model showing the names, abbreviations, equations, and references for equations. See footnotes for the various Sentinel-2 bands that were used.

Number	Name	Abbreviation	Equation	Statistics	Reference
1	The green band 3	B3	b3	Median	
2	The near infrared band 8	B8	b8	Median	
3	The red edge band swir4 band 12	B12	b12	Median	
4	The normalized difference vegetation index	NDVI	((b8 − b4)/(b8 + b4))	The 25th percentile	[39]
5	The normalized burn ratio	NBR	((b8 − b12)/(b8 + b12))	Median	[40]
6	The normalized difference red/green	REDGREEN	((b4 + b3)/(b3 − b4))	Median	
7	The plant senescence reflectance index	PSRI	((b4 − b2)/b6))	Median	[36]
8	The green-red vegetation index	GRVI	((b3 − b4)/(b3 + b4))	Median	[41]
9	The red-edge ratio vegetation index	RERVI	b5/b8	Median	[42]
10	The enhanced vegetation index	EVI	(2.5 * ((b8 − b4)/(b8 + 6 * b4 − 7.5 * b2 + 1))	Median	[43]
11	The carotenoid reflectance index 1	CRI1	((1/b2) − (1/b3))	Median	[44]
12	The green normalized difference vegetation index	GNDVI	((b8 -b3)/(b8 + 3))	Median	[45]
13	The pigment specific simple ratio	PSSR	b8/b4	Median	[46]

* Sentinel-2 bands used: b2 blue (490nm), b3 green (560nm), b4 red (665nm), b5 near infrared 1 (705nm), b6 near infrared 2 (740nm), b8 near infrared 4 (783nm), b12 short wave near infrared 4 (2190nm).

In Norway, LiDAR data are possible to download free of cost from hoydedata.no (accessed 23.04.2022). Both terrain and surface models from LiDAR data were uploaded to GEE. The terrain data can be found on asset "users/vegar/dtm1/dtmcoll", while the surface model can be found on asset "users/vegar/dom1/domcoll". The items can be visualized using this link: 6b56518b215b373147969757a695b4b6—Earth Engine Code Editor (google.com). The index we used was the surface model subtracted by terrain. This index indicates forests, buildings, and other objects that protrude above the terrain. The hypothesis is that open wetlands do not have high index values here.

2.3. Preparation of the Training Data

As a learning basis for the model, wetland polygons had to be annotated or digitized throughout the study area to create image chips or patches that could be used in the deep learning process. The advantage and big difference from using training points as a learning basis, for example, is that deep learning uses large pixel neighborhoods to decide whether a given pixel is a wetland area or something else. In theory, a pixel with some spectral deviation on a huge wetland surface will end up in the wetland class and not create the salt and pepper classification output that is often seen in other classification algorithms.

To avoid creating many thousands of image patches manually, all wetland polygons were digitized wall-to-wall within delimited rectangles. We initially made 17 rectangles over the total study area in southern Norway (see Figure 1). These rectangles were selected based

on previous knowledge and field work experience in these areas and to cover a cross-section of different wetland types and regional geographical and climate gradients. As a starting basis for digitizing the wetlands, two sets of orthophotos and existing wetland polygons from the 1:50,000 and 1:5000 map series were used. All polygons and parts of polygons that contained forest were removed either by mapping these via intensive fieldwork or by checking them against new orthophotos or LiDAR data in GIS. Previously unmapped wetland polygons were added to the map, either via intensive field work, via digitization in GIS from orthophotos, or preferably both. The levels of generalization both within and between the existing map bases varied from area to area. The 1:5000 maps also do not have coverage in the mountains. In addition, these maps have different refresh rates. Because of this, a great deal of work was required to create a harmonized data basis for models.

The 17 rectangles varied in size from approximately 25 km^2 to 600 km^2. They were divided into two, so that there were 17 squares that could be used for training purposes and 17 for evaluation. Based on these rectangles, image patches were systematically created through an algorithm made using Google Earth Engine Team (TensorFlow example workflows, Google Earth Engine, Google Developers, accessed on 24 April 2022) measuring 1280 × 1280 m (128 × 128 pixels for our 10-m resolution predictor stack, where the 20-m resolution bands in Sentinel-2 were resampled to 10 m), consisting of the 14 layers from satellite and LIDAR data as well as a rasterized version of the digitized wetlands as a training and evaluation base.

We generated 2000 samples (image patches/chips) for every training and evaluation polygon. All image patches contained one layer with the feature to be predicted or segmented (i.e., wetland) stacked upon all the other layers in the predictor stack (i.e., the 14-layer stack). The image patches from each polygon were merged into a single export and stored in Google Cloud Storage as a TFRecord. The TFRecord file contains patches of pixel values in each record.

2.4. Computing

GEE was used for storing, customizing, analyzing, and exporting and importing data for cloud-based deep learning modeling. Customized files in TensorFlow format (TFRecord-format) were exported to Google Cloud Storage and used as inputs for the virtual machine to process the deep learning scripts. The model used was a fully convolutional neural network (FCNN) for semantic classification. The goal of semantic segmentation is to label each pixel with a probability estimate of the input image with the class that represents a specific object [47,48].

We used the Keras implementation of the U-Net model [49]. The U-Net model takes 128 × 128-pixel patches as inputs and outputs the per-pixel class probability. The U-Net model is based on the architecture of a TensorFlow workflow made by the Google Earth Engine Team and can be inspected online (TensorFlow example workflows, Google Earth Engine, Google Developers). We used 0.1 as the learning rate, with a batch size of 16, 50 epochs, and 500 steps per epoch. U-Net consists of five encoder and five decoder convolutional layers, each consisting of 32, 64, 128, 256, and 512 channels, plus one center layer with 1024 filters.

For the model settings, for the training gain and evaluation accuracy metrics, we used the 'Adam' as the optimizer, 'Intersection-Over-Union as the loss, and the 'mean Intersection-Over-Union' metric for the accuracy evaluation. The latter meant that 17 samples from the training dataset were used to estimate the error gradient before the model weights were updated. The model was trained with 50 epochs to reach stable output performance. The trained model was then applied to the whole study area by exporting the predictor layers from this area to the cloud storage, where the calculations were performed.

The first U-Net deep learning model performed better near the training and evaluation rectangles than in areas far from these. Therefore, we ran an overlay analysis between our model and existing wetland maps from 1:5000 and 1:50,000 to find areas with major discrepancies. The main discrepancies could be "bad performance" from the deep learning

model or that the 1:5000 or 1:50,000 map bases were poorly updated. We found 23 areas where we wanted to improve the model and made an extra set of training rectangles and a similar wall-to-wall map of these (see Figure 1). The size of these additional rectangles was significantly smaller than the original 17 rectangles, and only the conformities were checked and corrected. This process was much faster than the first. The whole workflow is shown in Figure 2.

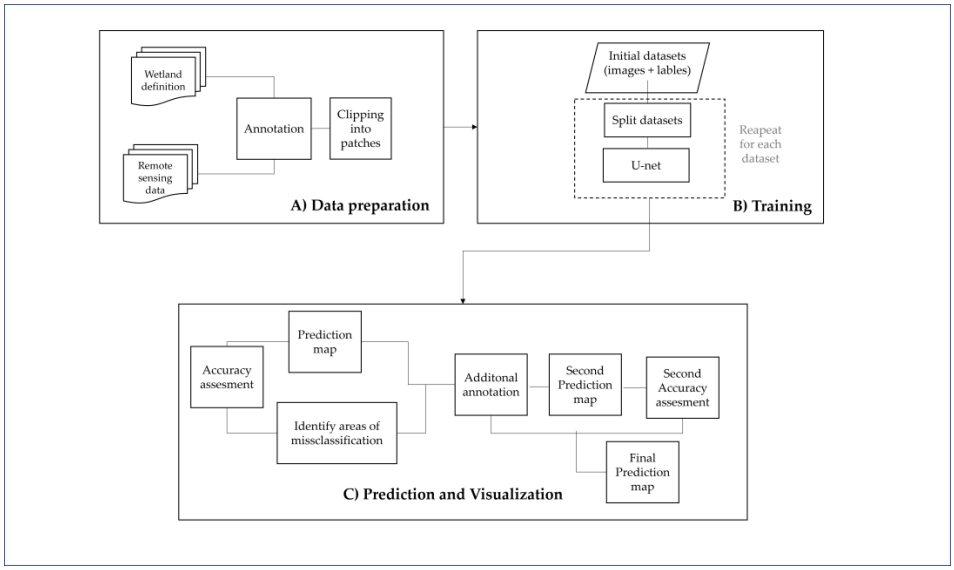


Figure 2. Workflow of the process of predicting open wetlands from Sentinel-2 imagery using TensorFlow and U-Net.

2.5. Unseen Validation Dataset

The wetland model was evaluated with independent ground truths from the national monitoring project ANO (area representative for monitoring in Norway). The ANO consists of 1000 randomly selected 500 × 500 m squares across mainland Norway (Figure 3). The 500 × 500 m squares were randomly drawn from a regular grid across Norway, and the only criterion was that they should not share a common line or corner point. Each ANO square contains 18 circles measuring 250 m^2 (radius 8.21 m), which are systematically located within the 500 × 500 m square (Figure 3). In the center of each circle, a 1 × 1 m vegetation plot was analyzed, where among other things the coverage of all vascular plants was recorded with the percentage coverage rates.

In both the 250 m^2 circles and in the 1 m^2 squares, dominant nature types are registered according to the Nature in Norway (NiN) system developed by [33] Halvorsen et al. (2020). Wetlands are divided into 13 different classes according to NiN, where the criterium is mostly related to the species composition. Our wetland model includes all of these NiN wetland classes, except the mire and swamp forest classes (classes V2 and V8).

For the NiN classes that include wetland edges, we accepted an uncertainty margin of one Sentinel-2 pixel (10 m) for a correct hit on a wetland edge. This was done by considering that Sentinel-2 is georeferenced with an inaccuracy rate corresponding to 90% of the pixels having an inaccuracy rate of less than 10 m [50].

A total of 4966 ANO ground truths were available in our study area at both the 250 m^2 scale level and the associated 1 m^2 vegetation plot level at the midpoint of the circle. In the circles at the 250 m^2 level, only the dominant nature type is specified, along with what percentage of the circle this nature type covers. At the 1 m^2 level, the dominant habitat type is also indicated but without a percentage indication, even if several nature types

are present. Note that nature types at the 250 m^2 level and 1 m^2 level may be different from each other, although they represent parts of the same area. As conditions for a valid wetland ground truth, we used as criteria that a wetland must cover more than 50% at the 250 m^2 level and that it must either be classified as a wetland at the 1m^2 level or that *Sphagnum* mosses have been registered in the vegetation plot.

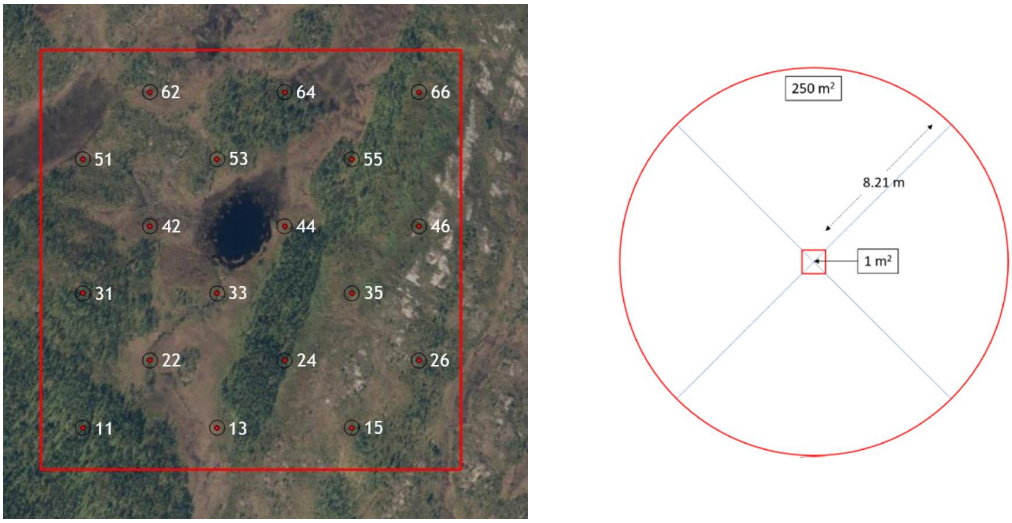


Figure 3. Field design of the unseen validation dataset named as the area representative for monitoring in Norway (ANO), first established in 2019. To the left is an example of a 500 × 500 m square that contains 18 monitoring circles in a systematically positioned geographical system. To the right is an example of a 250 m^2 monitoring circle with a vegetation sample plot of 1 m^2 in the middle. The main dominating LULC class (percent cover) is allocated to the circle in the field, while both the main LULC class and species estimates are given to the 1 m^2 plot.

3. Results

The first TensorFlow model, based on the 17 original selected training rectangles, was run over 50 epochs until achieving stable performance as measured by inspecting the training gain and evaluation accuracy metrics. The best model had a training mIOUaccuracy rate of 95.6% and validation accuracy rate of 94.9%.

The model output can be inspected here: https://vegar.users.earthengine.app/view/deeplearningmodel1 (accessed on 19 January 2023).

After the identification of areas of mismatches between the first TensorFlow model and existing reference data, a completely new deep learning model was trained on additional annotated data. This model was based on 36 training areas and 17 evaluation areas (rectangles) (Figure 1) and was trained over 50 epochs for stable performance. The new model had 98.51% accuracy for the training data and 98.41% accuracy for the evaluation data. This model corrected most of the weaknesses of the first deep learning model and was considered a better model.

The results of this model can be inspected here: https://vegar.users.earthengine.app/view/deeplearningmodel2 (accessed on 19 January 2023).

Of the 4966 independent ground truths, 547 of these satisfied our criteria as wetlands (Table 2).

Table 2. Estimated error matrix for the final classification with estimates for the user's accuracy (UA) and producer's accuracy (PA).

		Reference			
		Wetland	**Non-Wetland**	**Total**	**UA (%)**
Prediction	Wetland	491	351	842	58.3
	Non-wetland	56	4068	4124	98.6
	Total	547	4419	4966	
	PA (%)	89.8	92.1		

Here, we accepted an uncertainty rate of up to 10 m for NiN wetland edge types. The sensitivity rate of the model was 491/(491 + 56), equaling 89.8%m, and the specificity rate was 4068/(4068 + 351), equaling 92.1%. The balanced accuracy equation was (sensitivity + specificity)/2, i.e., (89.8 + 92.1)/2, equaling 90.9% accuracy.

The corresponding accuracy rates with independent ANO ground truths for the existing reference data for wetlands in Norway were 36.6% for the 1:5000 map basis and 42.2% for the 1:50,000 map. Note that the 1:5000 map does not cover the whole area of Norway and lacks coverage, especially in mountains. Therefore, after correcting for this unmapped area, the true positive rate was 46.8% for these areas according to the ANO ground truths.

We found 351 false positives out of a potential 4419 for our wetland model, which equated to 7.1% misclassifications of non-wetland ground truths. The most common misclassified nature types were nutrient-poor forest types and open heathlands, which are common neighboring nature types to wetlands and often appear together in different mosaics at different scale levels.

Our model estimated that approximately 12.7% of the land area in southern Norway is covered by wetlands (except mires and swamp forests). This was higher than previous estimates (please see the Introduction). The official Norwegian map on a scale of 1:50,000 contains only 5.6% mire areas for all of Norway.

4. Discussion

Our results indicate that as much as 12.7% of the land surface area in southern Norway is covered by wetlands. This is more than double the area of what is mapped in official maps for LULC in Norway. In another study using a point-based survey, [8] estimated that open wetlands in Norway probably cover 8.9% of the total area. Their study was based on an area frame survey of LULC types in a regular 18 × 18 km network (grid) across Norway, where an area of 0.9 km^2 was mapped at each intersection of the grid. The mapping was, however, mainly based on the interpretation of aerial photographs [8].

In the area representative for monitoring (ANO) program that we used for independent ground truths, 11.7% of the area was considered to be wetlands (mire and swamp forest not included) based on the registrations made in the 250 m^2 circles. When we only considered the registrations made in the center vegetation plot on the circles that are at the 1 m^2 scale level, 13.4% of these were assessed as wetlands.

Moreover, there are clear regional patterns where our model predicts more, but also less, wetland areas than the existing maps (see Figures 4 and 5). Our model predicts more wetlands in the mountains and on the west coast of Norway. This agrees with the findings of [8]. There are also regions in Norway where our model has smaller areas with wetland than the existing maps. This is particularly the case in southeastern Norway (see both Figures 4 and 5). This pattern was also observed by [8]. The cause is probably related to the ditching of bogs in the decades after World War II and a general reduction in wetland areas due to construction and other development close to heavily populated areas. The ditching of bogs has led to the bogs being desiccated, planted with trees, and now overgrown by forests. The official maps, thus, appear to be outdated in these areas.

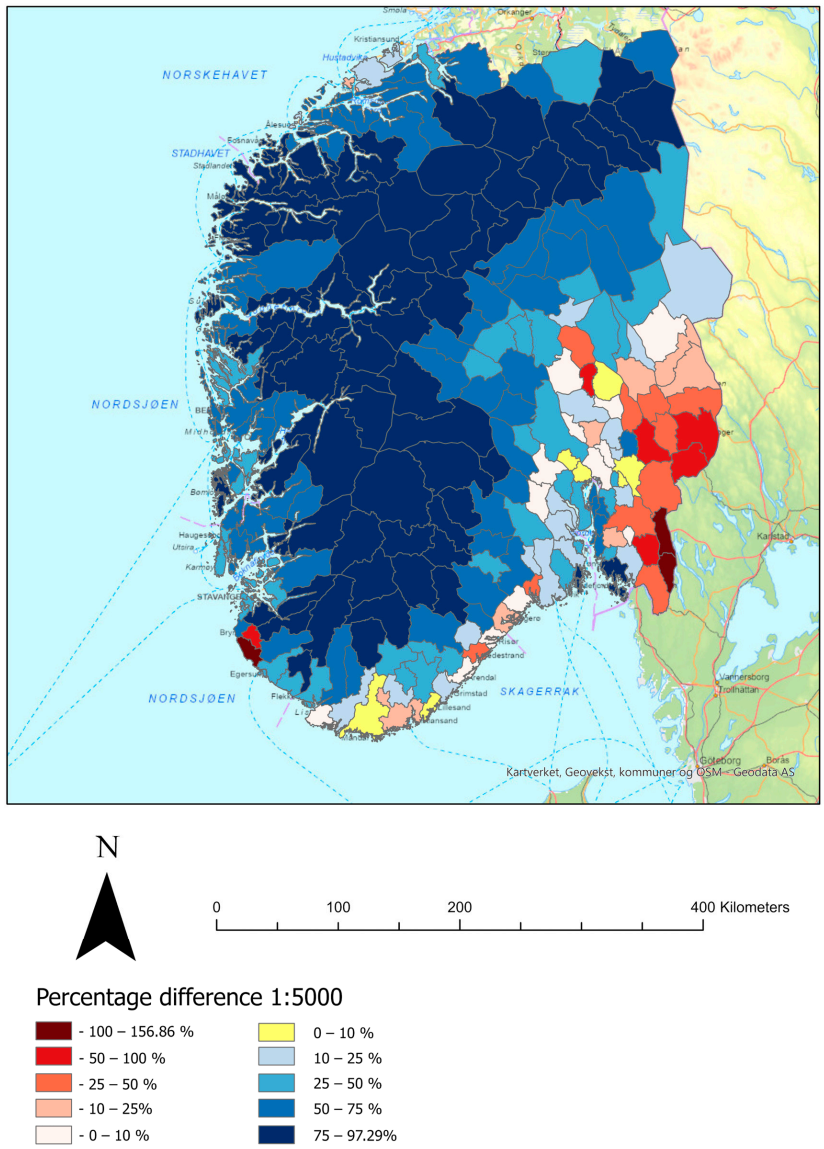


Figure 4. Visualization of the differences in predicted wetland coverage and existing reference data at the scale of 1:5000. Note that the 1:5000 maps do not cover mountain areas, and the large discrepancies in the central part of southern Norway are partly due to this lack of coverage (but see also Figure 5).

It was an extensive process and task to create a satisfactory map of wetlands in southern Norway. This was largely due to the lack of satisfactory annotated data, and most of the job involved digitizing wall-to-wall wetland polygons in the rectangles for the collection of image patches. However, the result and the internal deep learning evaluation accuracy of the final product (98.41%) must be said to be very good compared to earlier studies that aimed to classify wetlands by means of remote sensing. Accuracy rates of 70–95% have been reported in literature studies performed by [17,18,26]. However, only when we introduced an unseen validation dataset for evaluation could a credible confusion matrix be established. We found using this unseen and independent dataset that our model

is approximately 90.9% correct for our study area, which is far better than existing reference map data, which is below 50%.

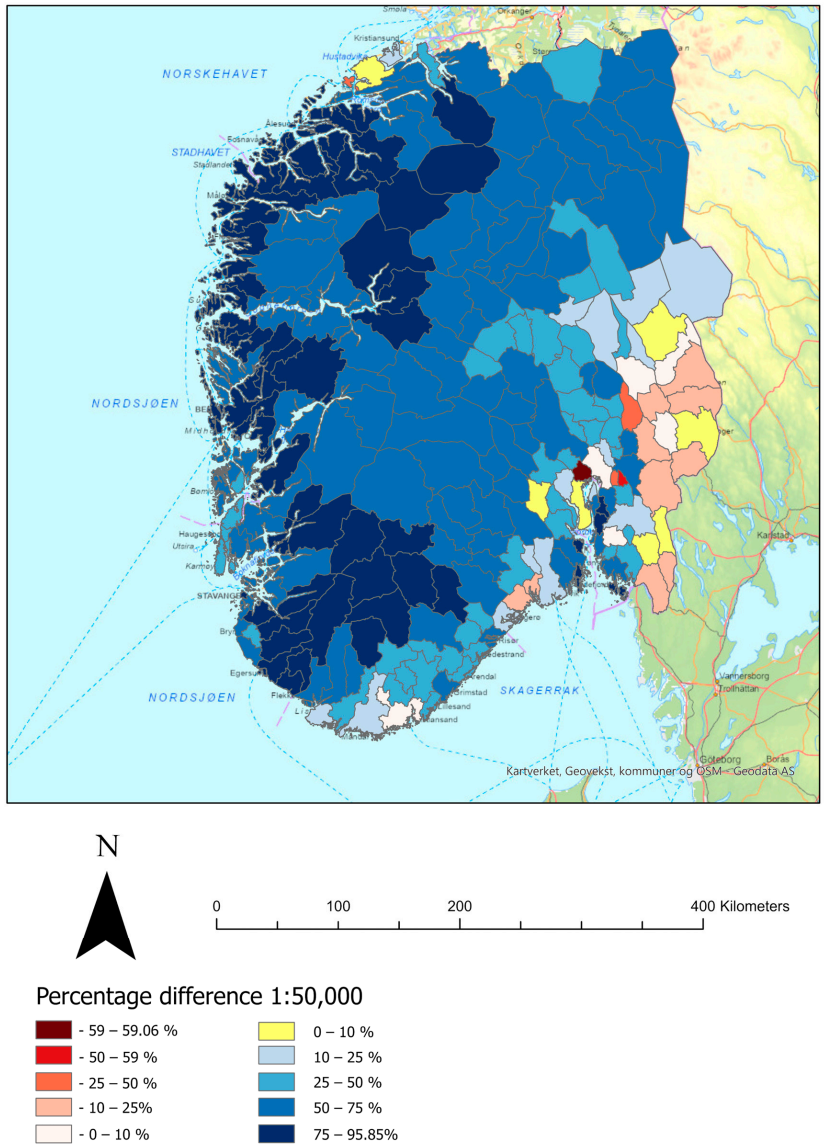


Figure 5. Visualization of the differences in predicted wetland coverage and existing reference data at the scale of 1:50,000.

Our model has a balanced accuracy rate of 90.9%, and depending on the further use of the model, users might explore additional filters and post-processing steps to remove some obvious misclassifications and unwanted areas. It will be possible to use the slope to mask out steep areas with a low possibility of being wetlands [51,52]. Additionally, it will be possible to set a threshold on heights above the nearest drainage network. Agricultural land is also annually mapped with a high degree of accuracy in Norway. This agricultural land layer may also be used to mask out misclassified wetlands in these areas.

Initially, a lot of time was spent looking for either spectral bands or indices that formed visual patterns that were in accordance with the wetland delineation we wanted to create. It quickly became clear that single bands or indices alone would not be able to contribute fully to the wetland delineation along the east–west gradient in Norway, nor to different types of wetlands. For instance, indices that seemed to work well in continental areas showed a lower spatial pattern correlation with wetland types in western Norway. It was only when we began to experiment with autumn colors and limited the search for Sentinel-2 images to the period of August–October that the patterns became clearer throughout the study area. The colors yellow, orange, and red, in addition to green, are the most common autumn colors in Norway. The indices that distinguished these colors from each other visually also had the clearest patterns that could distinguish wetlands from other habitat types.

The transfer value of our results is limited by the fact that we did not calculate variable importance scores for the satellite features that informed the deep learning model. Due to funding and time constraints, we were not able to calculate importance scores for the set of predictor variables. We identify this as an avenue for further research given that reducing the feature set can drastically improve the processing and inference times in deep learning architectures [53].

The process of creating the map was iterative, where the first deep learning model based on wetland polygons from 17 rectangles gave good results in the surrounding areas with roughly the same wetland types. However, the models deteriorated when transferred to areas further away from the training rectangles and became quite unreliable 100–200 km from the nearest annotated training area. We, therefore, supplemented the training dataset with new rectangles for wetland types in different regions along altitude, north–south, and east–west (continental) gradients at regular intervals to achieve comparable results with existing maps. This resulted in an updated wetland prediction map of southern Norway that is up to date regarding the input data and now captures new cottage developments and the large-scale ditching of bogs, for which the existing official maps are not yet updated.

It is possible that the use of less training data can give similarly good results by using augmentation methods [54]. We did not try this here, but it may be relevant if the model is to be extended to include northern Norway. However, we felt strongly that the use of a denser network of training rectangles to capture different and other wetland types was important, as the model's performance was significantly poorer with increasing distances from training (and evaluation) rectangles. It is also possible that different models could have been made for specific classes of different types of wetlands and for different regions, such as one model for wetlands in the mountains and one for the lowlands. These are issues that must be tested in future projects to further improve the models. Wetlands must also be said to be a difficult nature type (land cover) to classify [55]. Deep learning methods have also been used to classify other LULC classes with some success [56–58]

In recent times, and especially in the last two or three years, an increasing number of papers have been published on the use of U-Net and partly Tensor-Flow in semantic segmentations of imagery. However, these methods are mostly used on very high-resolution images and in remote sensing in products such as WorldView3 and others with a resolution of 1 m or even higher [18]. U-Net has hardly been used on remote sensing products other than RGB and hardly in combinations of Sentinel-1 and Sentinel-2 imagery. One of few exceptions was reported by [21], who used these methods for LULC classification based on U-Net on Sentinel-1 and -2 data. They stated the accuracy to be about 76%, which is quite low. They reported three reasons for this: (1) there were too many details in the classification system; (2) the spatial resolution was too coarse; (3) there were too few observations in the training set (annotations). As we attempted to predict only one LULC class, their first reason did not apply to us. To solve point (reason) 2, we tried to use relevant bands and indices that had as crisp and sharp boundaries between wetlands and other LULC classes as possible. To solve the last point, we invested a lot of time and effort into increasing the amount of annotated data. [21] also stated that the use of a low training basis with annotated data was an important reason for low accuracy in their results. They also stated

that they used augmentation to increase the training basis. [59] used BigEarthNet [60] and achieved accuracy rates ranging from 98% (impervious and agricultural areas) to 75% for wetlands. BigEarthNet is a large-scale Sentinel-2 dataset collected from a total of 125 Sentinel-2 tiles covering areas of 10 countries in Europe, with a total of 590,326 tiles measuring 120 × 120 pixels, which are annotated with land cover classification labels.

SAR (Sentinel-1) data were also tested as the input data in our initial phase. However, we ruled out Sentinel-1 bands as predictors as the same wetland (training) polygons often had very different band values in different internal parts. This gave erroneous results, where parts of the wetland polygon were not correctly predicted in the finished classification. U-Net was originally designed to find objects (segments) in very high-resolution images, especially RGB and other low-spectral products [61]. To us, it seems that the method works best on Sentinel imagery when the bands or indices form sharp boundaries and that the polygons' boundaries are in accordance with the boundaries we initially wanted to draw.

5. Conclusions

In our experience, and based on the evidence presented in this paper, the classification of ecosystems and land cover classes based on satellite and repeated airborne remote sensing imagery offers some significant advantages over in situ and manual reference mapping: (1) it covers large areas and multiple years in a consistent and comparable manner, in that it is objectively compared to manual mapping, which is performed by different individuals over different parts of the country; (2) it is scalable to some extent because one can extrapolate over time and space using the deep learning models; (3) it is continually updateable and often available in near real-time, thereby allowing for ongoing monitoring and surveillance (note that LiDAR data are not continuously updated and updates must not rely on this source); (4) it performed better than the existing approaches for reference data validated by regional unseen ground truth data. Whether the kind of deep learning approach presented here or other machine learning methods provide better and more effective classification results must be explored in future studies. A focus on change detection using the methods presented here to account for changes in wetland extents over time would be particularly interesting and critical for effective wetland conservation.

Author Contributions: Conceptualization, V.B.; Methodology, V.B. and A.J.G.; Validation, V.B., Z.V., A.J.G. and E.F.; Formal analysis, V.B.; Writing—original draft, V.B., Z.V., A.J.G. and E.F. Writing— review & editing, V.B., Z.V., A.J.G. and E.F.; Visualization, V.B., and A.J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The results can be viewed at https://vegar.users.earthengine.app/view/deeplearningmodel2 (accessed on 20 January 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, X.; Chen, M.; Yang, G.; Jiang, B.; Zhang, J. Wetland ecosystem services research: A critical review. *Glob. Ecol. Conserv.* **2020**, *22*, e01027. [CrossRef]
2. Magnussen, K.; Bjerke, J.W.; Brattland, C.; Nybø, S.; Vermaat, J. *Verdien av Økosystemtjenester fra Våtmark*; Menon-Publikasjon: Oslo, Norway, 2018. (In Norwegian)
3. Villa, J.A.; Bernal, B. Carbon sequestration in wetlands, from science to practice: An overview of the biogeochemical process, measurement methods, and policy framework. *Ecol. Eng.* **2018**, *114*, 115–128. [CrossRef]
4. Taillardat, P.; Thompson, B.S.; Garneau, M.; Trottier, K.; Friess, D.A. Climate change mitigation potential of wetlands and the cost-effectiveness of their restoration. *Interface Focus* **2020**, *10*, 20190129. [CrossRef]
5. IPBES. *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services*; Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES): Paris, France, 2019.
6. Sievers, M.; Hale, R.; Parris, K.M.; Swearer, S.E. Impacts of human-induced environmental change in wetlands on aquatic animals. *Biol. Rev.* **2018**, *93*, 529–554. [CrossRef]

7. Nybø, S.; Evju, M.; Fagsystem for Fastsetting av God Økologisk Tilstand. Forslag fra et Ekspertråd. *Ekspertrådet for Økologisk Tilstand, 247 s.* 2017. Available online: https://www.regjeringen.no/no/dokument/rapportar-og-planar/id438817/ (accessed on 20 January 2023).
8. Bryn, A.; Strand, G.-H.; Angeloff, M.; Rekdal, Y. Land cover in Norway based on an area frame survey of vegetation types. *Nor. Geogr. Tidsskr.* **2018**, *72*, 131–145. [CrossRef]
9. Ahlstrøm, A.P.; Bjørkelo, K.; Fadnes, K. AR5 Klassifikasjonssystem. Klassifisering av Arealressurser. NIBIO Bok: Ås, Norway, 2019; Volume 5. (In Norwegian)
10. Norwegian Map Authorities. *Produktspesifikasjon N50 Kartdata: Versjon April 2017*; Statens Kartverk: Hønefoss, Norway, 2017. (In Norwegian)
11. Statistics Norway. *09594: Arealbruk og Arealressurser, Etter Arealklasser (km^2) (K) (B) 2011–2022. Statistikkbanken (ssb.no)*; Statistics Norway: Oslo, Norway, 2023.
12. Eriksen, E.L.; Ullerud, H.A.; Halvorsen, R.; Aune, S.; Bratli, H.; Horvath, P.; Volden, I.K.; Wollan, A.K.; Bryn, A. Point of view: Error estimation in field assignment of land-cover types. *Phytocoenologia* **2018**, *49*, 135–148. [CrossRef]
13. Ullerud, H.A.; Bryn, A.; Halvorsen, R.; Hemsing, L.Ø. Consistency in land-cover mapping: Influence of field workers, spatial scale and classification system. *Appl. Veg. Sci.* **2018**, *21*, 278–288. [CrossRef]
14. Erikstad, L.; Strand, G.-H.; Bentzen, F.; Salberg, A.-B. *Arealrepresentativ Overvåking Basert på Fjernanalyse. Flyfototolkning i Fjell og Myrnatur—NINA Rapport 743*; Norsk Institutt for Naturforskning: Trondheim, Norway, 2011. (In Norwegian with English Abstract)
15. Burrough, P.A.; McDonnell, R.A. *Principles of Geographical Information Systems*; Oxford University Press: Oxford, Uk, 1998.
16. Norwegian Environment Agency. In *Wetland restoration plan, Norway (2021–2025)*; Report M-1903; NIBIO Bok: Oslo, Norway, 2021.
17. Mahdianpari, M.; Granger, J.E.; Mohammadimanesh, F.; Salehi, B.; Brisco, B.; Homayouni, S.; Gill, E.; Huberty, B.; Lang, M. Meta-analysis of wetland classification using remote sensing: A systematic review of a 40-year trend in North America. *Remote Sens.* **2020**, *12*, 1882. [CrossRef]
18. Venter, Z.S.; Nowell, M.S.; Bakkestuen, V.; Ruud, A.; Kruse, M.; Skrindo, A.B.; Kyrkjeeide, M.O.; Singsaas, F.T. *Literature Review of Wetland Remote Sensing and Mapping*; NINA Rapport 2014; Norsk Institutt for Naturforskning: Trondheim, Norway, 2021.
19. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Homayouni, S.; Gill, E. The first wetland inventory map of Newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote Sens.* **2019**, *11*, 43. [CrossRef]
20. d'Andrimont, R.; Yordanov, M.; Martinez-Sanchez, L.; Eiselt, B.; Palmieri, A.; Dominici, P.; Gallego, J.; Reuter, H.I.; Joebges, C.; Lemoine, G.; et al. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. *Sci. Data* **2020**, *7*, 352. [CrossRef]
21. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 3600. [CrossRef]
22. Ma, L.; Liu, Y.; Zhang, A.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
23. Dang, K.B.; Nguyen, M.H.; Nguyen, D.A.; Phan, T.T.H.; Giang, T.L.; Pham, H.H.; Nguyen, T.N.; Tran, T.T.V.; Bui, D.T. Coastal Wetland Classification with Deep U-Net Convolutional Networks and Sentinel-2 Imagery: A Case Study at the Tien Yen Estuary of Vietnam. *Remote Sens.* **2020**, *12*, 3270. [CrossRef]
24. DeLancey, E.R.; Simms, J.F.; Mahdianpari, M.; Brisco, B.; Mahoney, C.; Kariyeva, J. Comparing Deep Learning and Shallow Learning for Large-Scale Wetland Classification in Alberta, Canada. *Remote Sens.* **2020**, *12*, 2. [CrossRef]
25. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
26. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]
27. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
28. Bisong, E. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Apress: Berkeley, CA, USA, 2019. [CrossRef]
29. Moen, A. *National atlas of Norway. Vegetation*; Norwegian Mapping Authority: Hønefoss, Norway, 1999.
30. Bakkestuen, V.; Erikstad, L.; Halvorsen, R. Step-less models for regional environmental variation in Norway. *J. Biogeogr.* **2008**, *35*, 1906–1922. [CrossRef]
31. Sigmond, E.M.O. *Brukerveiledning til Berggrunnskart over Norge. Nasjonalatlas for Norge*; Statens Kartverk: Hønefoss, Norway, 1985. (In Norwegian)
32. Statistics Norway. *Skog, Fjell og Vidde Dominerer-SSB*; Statistics Norway: Oslo, Norway, 2017.
33. Halvorsen, R.; Skarpaas, O.; Bryn, A.; Bratli, H.; Erikstad, L.; Simensen, T.; Lieungh, E. Towards a systematics of ecodiversity: The EcoSyst framework. *Glob. Ecol. Biogeogr.* **2020**, *29*, 1887–1906.
34. Venter, Z.S.; Sydenham, M.A.K. Continental-Scale Land Cover Mapping at 10 m Resolution Over Europe (ELC10). *Remote Sens.* **2021**, *13*, 2301. [CrossRef]

35. Gómez-Giráldez, P.J.; Pérez-Palazón, M.J.; Polo, M.J.; González-Dugo, M.P. Monitoring Grass Phenology and Hydrological Dynamics of an Oak–Grass Savanna Ecosystem Using Sentinel-2 and Terrestrial Photography. *Remote Sens.* **2020**, *12*, 600. [CrossRef]
36. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-Destructive Optical Detection of Pigment Changes during Leaf Senescence and Fruit Ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]
37. Pflugmacher, D.; Rabe, A.; Peters, M.; Hostert, P. Mapping Pan-European Land Cover Using Landsat Spectral-Temporal Metrics and the European LUCAS Survey. *Remote Sens. Environ.* **2019**, *221*, 583–595. [CrossRef]
38. Griffiths, P.; Nendel, C.; Pickert, J.; Hostert, P. Towards National-Scale Characterization of Grassland Use Intensity from Integrated Sentinel-2 and Landsat Time Series. *Remote Sens. Environ.* **2019**, *238*, 111124. [CrossRef]
39. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
40. Snyder, A.; Fulé, P.; Crouse, J. Comparison of burn severity assessment using Differenced Normalized Burn Ratio and ground data. *Int. J. Wildland Fire* **2005**, *14*, 189–198. [CrossRef]
41. Motohka, T.; Nasahara, K.N.; Oguma, H.; Tsuchida, S. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sens.* **2010**, *2*, 2369–2387. [CrossRef]
42. Cao, Q.; Miao, Y.; Wang, H.; Huang, S.; Cheng, S.; Khosla, R.; Jiang, R. Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor. *Field Crops Res.* **2013**, *154*, 133–144. [CrossRef]
43. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
44. Gitelson, A.A.; Zur, Y.; Chivkunova, O.B.; Merzlyak, M.N. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochem. Photobiol.* **2002**, *75*, 272–281. [CrossRef]
45. Gitelson, A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
46. Maleki, M.; Arriga, N.; Barrios, J.M.; Wieneke, S.; Liu, Q.; Peñuelas, J.; Janssens, I.A.; Balzarolo, M. Estimation of Gross Primary Productivity (GPP) Phenology of a Short-Rotation Plantation Using Remotely Sensed Indices Derived from Sentinel-2 Images. *Remote Sens.* **2020**, *12*, 2104. [CrossRef]
47. Zhuang, J.; Yang, J.; Gu, L.; Dvornek, N. Shelfnet for fast semantic segmentation. In Proceedings of the 2019 International Conference on Computer Vision Work ICCVW, Seoul, Korea, 27–28 October 2019; pp. 847–856.
48. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
49. Chollet, F. Keras. 2015. Available online: https://keras.io. (accessed on 5 January 2023).
50. Stumpf, A.; Michéa, D.; Malet, J.-P. Improved Co-Registration of Sentinel-2 and Landsat-8 Imagery for Earth Surface Motion Measurements. *Remote Sens.* **2018**, *10*, 160. [CrossRef]
51. Halabisky, M.; Babcock, C.; Moskal, L.M. Harnessing the Temporal Dimension to Improve Object-Based Image Analysis Classification of Wetlands. *Remote Sens.* **2018**, *10*, 1467. [CrossRef]
52. Muro, J.; Varea, A.; Strauch, A.; Guelmami, A.; Fitoka, E.; Thonfeld, F.; Diekkrüger, B.; Waske, B. Multitemporal optical and radar metrics for wetland mapping at national level in Albania. *Heliyon* **2020**, *6*, e04496. [CrossRef]
53. Wang, J.; Bretz, M.; Dewan, M.A.A.; Delavar, M.A. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Sci. Total Environ.* **2022**, *822*, 153559. [CrossRef]
54. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
55. Mahdavi, S.; Salehi, B.; Granger, J.; Amani, M.; Brisco, B.; Huang, W. Remote sensing for wetland classification: A comprehensive review. *GIScience Remote Sens.* **2018**, *55*, 623–658. [CrossRef]
56. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
57. Hoeser, T.; Bachofer, F.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications. *Remote Sens.* **2020**, *12*, 3053. [CrossRef]
58. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
59. Ulmas, P.; Liiv, I. Segmentation of satellite imagery using U-Net models for land cover classification. *arXiv* **2020**, arXiv:2003.02899.
60. Sumbul, G.; Charfuelan MDemir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IEEE International Conference on Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5901–5904.
61. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351. [CrossRef]

Article III.

Globally vs. Locally trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape

Alexandra Jarna Ganerød [1,2, †], Erin Lindsay [3, *, †], Ola Fredin [4], Tor-Andre Myrvoll [5], Steinar Nordal [3], Jan Ketil Rød [1]

1 Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway
2 Geological Survey of Norway (NGU), 7040 Trondheim, Norway
3 Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway
4 Department of Geoscience and Petroleum, Norwegian University of Science and Technology, 7034 Trondheim, Norway
5 Department of Electronic Systems, Norwegian University of Science and Technology, 7034 Trondheim, Norway
* Author to whom correspondence should be addressed.
† These authors contributed equally to this work

remote sensing

MDPI

Article

Globally vs. Locally Trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape

Alexandra Jarna Ganerød [1,2,†], Erin Lindsay [3,*,†], Ola Fredin [4], Tor-Andre Myrvoll [5], Steinar Nordal [3] and Jan Ketil Rød [1]

[1] Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway
[2] Geological Survey of Norway (NGU), 7040 Trondheim, Norway
[3] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway
[4] Department of Geoscience and Petroleum, Norwegian University of Science and Technology, 7034 Trondheim, Norway
[5] Department of Electronic Systems, Norwegian University of Science and Technology, 7034 Trondheim, Norway
* Correspondence: erin.lindsay@ntnu.no; Tel.: +47-9156-4607
† These authors contributed equally to this work.

Abstract: Landslide risk mitigation is limited by data scarcity; however, this could be improved using continuous landslide detection systems. To investigate which image types and machine learning models are most useful for landslide detection in a Norwegian setting, we compared the performance of five different machine learning models, for the Jølster case study (30 July 2019), in Western Norway. These included three globally pre-trained models; (i) the continuous change detection and classification (CCDC) algorithm, (ii) a combined k-means clustering and random forest classification model, and (iii) a convolutional neural network (CNN), and two locally trained models, including; (iv) classification and regression Trees and (v) a U-net CNN model. Images used included Sentinel-1, Sentinel-2, as well as digital elevation model (DEM) and slope. The globally trained models performed poorly in shadowed areas and were all outperformed by the locally trained models. A maximum Matthew's correlation coefficient (MCC) score of 89% was achieved with a CNN U-net deep learning model, using combined Sentinel-1 and -2 images as input. This is one of the first attempts to apply deep learning to detect landslides with both Sentinel-1 and -2 images. Using Sentinel-1 images only, the locally-trained deep-learning model significantly outperformed the conventional machine learning model. These findings contribute to developing a national continuous monitoring system for landslides.

Keywords: NDVI; SAR; change detection; Norway; Sentinel-1; Sentinel-2; deep learning; U-Net; CCDC; Google Earth Engine

Citation: Ganerød, A.J.; Lindsay, E.; Fredin, O.; Myrvoll, T.-A.; Nordal, S.; Rød, J.K. Globally vs. Locally Trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape. *Remote Sens.* **2023**, *15*, 895. https://doi.org/10.3390/rs15040895

Academic Editors: Federico Raspini, Simon Plank, Matteo Del Soldato, Pierluigi Confuorto, Chiara Cappadonia and Mariano Di Napoli

Received: 23 December 2022
Revised: 27 January 2023
Accepted: 1 February 2023
Published: 6 February 2023

1. Introduction

Landslides are the most widespread geologic hazard, yet are amongst the least reported type of disasters. In the period 1998–2017, landslides affected an estimated 4.8 million people globally, resulting in over 18,000 fatalities [1]. Landslides can occur in soil or rock materials and include a variety of slope failure mechanisms such as falls, slides, spreads, and flows [2]. They may occur as single events, or multiple events sharing a common triggering event such as heavy rainfall or an earthquake and occur most frequently in regions with high hydrogeological or seismic hazard [3].

Accurate knowledge of past landslide events is needed to mitigate risk from future events. This knowledge is used to develop an understanding of the local hazard conditions, needed for accurate hazard and susceptibility mapping, spatial planning and landslide

early warning systems [4,5]. A lack of systematic information on the type, abundance, and distribution of historic landslides is a major limitation for landslide risk mitigation.

Landslides are generally detected from field observations or remotely sensed imagery [5]. There have been an increasing number of studies investigating automated methods for landslide detection and mapping using machine learning models and satellite images, particularly since 2017 [6]. Operational monitoring and alert systems using similar approaches exist for deforestation [7–9] and are being developed for other types of natural hazards, including flood [10] and wildfire detection [11,12]. Similar systems for landslide detection would be extremely valuable for obtaining timely and objective data on landslide events. This would lead to an improved understanding of the controlling factors and spatial distribution of past and future landslides and improved reliability of susceptibility and hazard maps [13].

Many of the same change detection methods and data types used for continuous monitoring of forest loss are also relevant for landslide detection, given that landslides often result in the removal of vegetation. Change detection with machine-learning techniques can be performed using temporal or spatial data from satellite images. Temporal methods can detect abrupt changes in time-series data due to a change in ground cover properties. For example, the continuous change detection and classification (CCDC) algorithm [14] can detect gradual and abrupt changes in land cover types. This involves detecting deviances from expected values based on patterns of historic seasonal spectral behavior for a given pixel. The original CCDC model has been run for all existing Landsat data globally, with results made available on Google Earth Engine [15]. We did not find any examples of automated landslide detection using similar time-series-based change detection methods.

Spatial methods on the other hand, are popular for both deforestation and landslide detection. Pixels showing vegetation loss can be identified from post-event, or from sets of pre- and post-event images, using various machine learning methods. Deep learning, and particularly, U-net architecture, has proven to be a powerful segmentation tool in scenarios with limited data, simple structure, and high recognition accuracy. These methods typically follow a workflow that involves training a model using an existing local landslide inventory. The pre-trained model is then used to predict landslides in the surrounding regions that are similar to the training area [16]. Recently U-Net has been widely used in landslide mapping, e.g., [17–21].

In terms of image types used, optical multispectral and LiDAR (light detection and ranging) data are common. However, event detection may be delayed by months due to persistent cloud cover. Hence, there has recently been increasing use of synthetic aperture radar (SAR) data for landslide detection [22–24] and continuous monitoring systems for deforestation [7,25,26]. SAR data are also useful for change detection in areas where there are strong seasonal variations, including snow, seasonal darkness, and lack of vegetative biomass (e.g., in temperate and cold climates). Using U-net architecture, a combination of both Sentinel-1 and Sentinel-2 input data, was found to achieve improved accuracy compared to optical data only, for detecting illegal logging events in both summer and winter in Ukraine. The input was stacks of optical and radar images in summer and spring, and radar images only in winter and autumn [27]. However, there are barriers to using SAR data for landslide detection, due to more complicated pre-processing, and a lack of understanding of how to interpret landslides in SAR backscatter data [24]. Therefore, most machine learning models for landslide detection use optical or multispectral images as input data [6].

However, even if cloud-free optical images are available shortly after a triggering event, applying U-Net models for rapid landslide mapping in emergencies is often not feasible. This is the case when there is a lack of historic and local landslide data, represented as polygon features, available to pre-train the model [3]. In response to this problem, there have been attempts to produce globally-trained generalized machine learning models capable of detecting and mapping landslides in previously unseen locations. The first attempt was by Prakash et al. (2021) with a convolutional neural network (CNN) model

that was trained on seven locations around the seismically active Pacific Ring, with high vegetation coverage [28]. Another example was by Tehrani et al. (2021), who developed an object-based method using k-means clustering to perform semantic segmentation, followed by random forest classifiers that determine whether the segments represent landslides or not [29]. This model was trained on data from 29 locations around the world.

In this study, our main goal is to determine which elements of existing automatic landslide detection and deforestation monitoring approaches could be feasible to include in a national landslide detection system in Norway. This represents one of the first attempts to use machine learning to detect and map landslides in Norway. With a glacially sculpted landscape with steep slopes, and strong seasonal variability, the environment in Norway is relatively unique, and it is unknown how well the generalized models will perform in such a setting. We investigate the performance of five different machine learning models using satellite images from Sentinel-1 and -2, along with elevation and slope rasters. The well-verified landslide inventory from the Jølster case study (30 July 2019) [30] is used to test which approaches could be adapted for larger-scale use in the future.

We test the performance of three pre-existing globally-trained models, including (i) the time-series-based CCDC algorithm, (ii) the object-based model from Tehrani et al., (2021), and (iii) the pixel-based CNN model from Prakash et al. (2021). Two locally-trained models were also tested, including: (iv) a classification and regression tree (CART) machine learning model [31] and (v) a CNN U-net deep learning model.

The following research questions are investigated:

1. How do globally pre-trained machine learning models for landslide detection perform in a glacial landscape?
2. Which locally-trained model and input data combination gives the best results?
3. Which elements of the investigated models could be implemented in an operational national landslide detection system?

In the following section, we describe the current situation in Norway in terms of landslide hazards and introduce the case study. In the results, we show that the globally trained models generally did not perform well in a glacial landscape, particularly for landslides on north-facing slopes. The locally trained deep learning model outperformed the machine learning model with all input data combinations, except for one. The best performance (MCC score: 89%) was achieved using combined Sentinel-1 and Sentinel-2 data as input. We did not attempt to retrain or modify the existing globally pre-trained models in this study, although we provide suggestions as to how their performance could be improved in the discussions.

2. Norwegian Setting and Case Study

Landslides occur almost daily in mountainous regions in Norway and are the natural hazard responsible for most fatalities [32]. In addition, they cause large economic losses due to damage to infrastructure and disruption to transportation [33]. In comparison to other countries in Europe, Norway has a relatively high proportion of land area that is susceptible to landslides, with over 70% of municipalities affected [34]. This is due to the geological landscape with high mountains, valleys with steep slopes, and post-glacial isostatic rebound that has resulted in sensitive clays in valley bottoms in coastal regions [32].

The most frequent types of landslides in Norway include rock fall, rock slides, debris avalanches, and debris flows [13]. In addition, there are unstable mountains and deep-seated landslides that can evolve into large rock avalanches and quick clay slides [32].

To mitigate the increasing risk to society due to landslide hazards, there are several national initiatives coordinated by the Norwegian Water Resources and Energy Directorate (NVE). These include, among others, the preparation of susceptibility and hazard maps and close communication with spatial planners at municipalities to protect inhabitants and key infrastructure already located in hazardous areas. NVE is also working with the prediction of hydro-meteorologically induced landslide occurrences through a national early forecasting and warning service [35]. The early warning system allows municipalities

and individuals to take timely action to reduce risk, including evacuations and closure of transport routes in areas with high hazards.

These initiatives rely on knowledge of historic landslide occurrence and are limited by the quality and completeness of historic landslide records [13]. The Norwegian Mass Movements Database (available from: https://nedlasting.nve.no/gis/, accessed on 20 December 2022) contained 84,768 reports at the time of writing, from the year 900 to 2022. Yet there are some significant limitations in the existing landslide dataset that make it unsuitable for spatial analyses; for instance, determining statistical relationships between landslide occurrence and the topographical, geological, hydrological, vegetation, or meteorological factors. These include low locational and qualitative (i.e., information on landslide type, size, and trigger) accuracy of older events that have been extracted from historic church and municipality records. These reports are generally limited to events that caused death or destroyed property.

While modern reporting is performed systematically by the road and rail authorities [36], reporting focuses on events that directly impact transport infrastructure. The given locations are typically represented by the point where a landslide impacted the road, and the initiation point is not usually specified. Although these data generally have high spatial and temporal accuracy, there remain some qualitative inaccuracies. Furthermore, compared to 11 other national landslide databases, there is a spatially biased distribution, with many reports located along roads but relatively few events reported in remote areas [37], as illustrated in Figure 1. NVE use aerial and satellite images to manually map polygons representing the landslides and periodically perform quality control of the existing landslide point data. However, detecting and mapping traces of small landslide events across large areas remain a tedious and labor-intensive process [35].

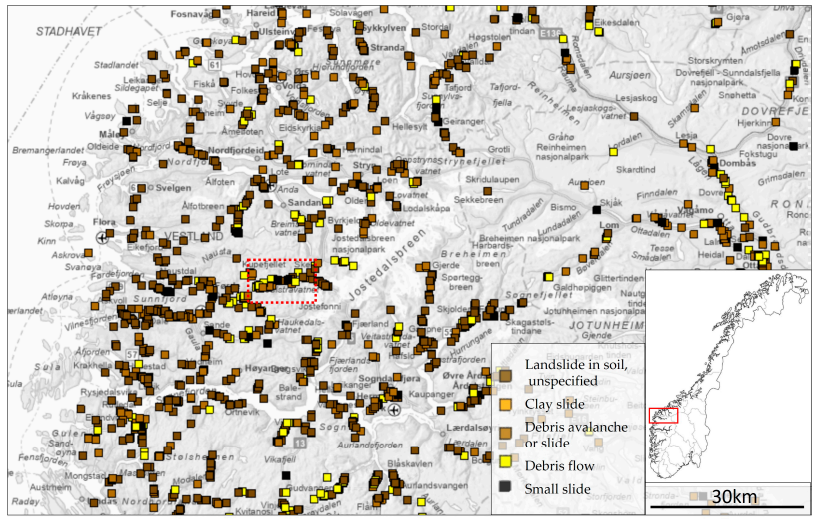


Figure 1. Registered landslide events in Western Norway have an inherent spatial bias towards roads. The location of the case study area, shown in the following figures, is indicated by the dashed red lines. Data come from www.skredregistering.no (accessed on 16 December 2022), showing registered landslide events from 1992 to 2022.

There is a strong need for improved landslide mapping techniques in Norway, which can provide objective and accurate spatial information, and allow the detection of events that occur away from populated areas and transport routes. Recent studies have demonstrated there is great potential to improve the detection of landslides in remote areas using satellite images [30,38].

In July 2019, an extremely heavy rainfall event triggered multiple landslides in the (formerly named) Jølster municipality in Western Norway [38]. The maximum recorded rainfall was 113 mm in 24 h, exceeding the 200-year event magnitude at the two nearest precipitation weather stations, Botnen and Haukedalen, in the neighboring municipality of Førde [39]. The road authority reported 14 landslides on this date, while mapping from Sentinel-2 images detected 120 events, with only 30% being located within 500 m of a road, compared to 100% of those registered by the road authority [30].

The study area is shown in Figure 2. The landscape consists of steep glacial valleys, lakes, and mountains up to 1666 m. The town of Vassenden is located in a tempered climate zone with relatively mild winters and wet summers due to its proximity to the coast. The mean annual precipitation over the past five years is 2800 mm/yr at the Botnen weather station, and temperatures vary from −25 °C to 31 °C, with an annual mean of 5 °C (https://seklima.met.no/, accessed on 27 January 2023). The hydro-meteorologically induced landslides that pose a risk to these areas are expected to become more frequent due to an increase in extreme precipitation events [40].

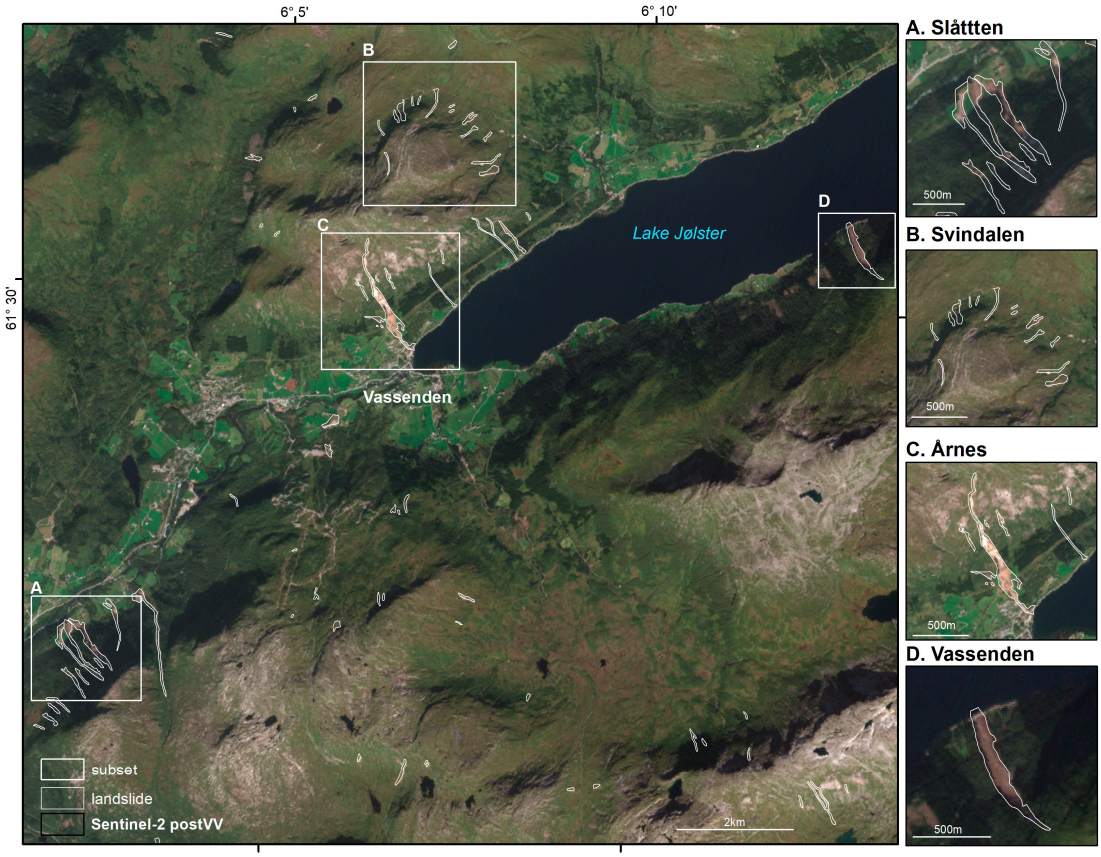


Figure 2. Case study overview showing four subsets with ground truth landslide outlines in white.

The bedrock geology is predominantly granitic (banded and augen) orthogneiss and quartz-monzonite. The geomorphology is shaped by old faults and glacial erosion, with a quaternary surface cover typically consisting of highly consolidated moraine material overlying the bedrock, with a looser veneer of colluvium on valley slopes. The surface cover is thin to non-existent at high altitudes and increases to several meters thick in lower areas close to the lake. The vegetation ranges from sparse moss and shrubs or light birch

forest at high elevations to spruce forest and agricultural fields lower in the valleys. Roads and built areas are mainly located in the flatter main valleys. The area is very susceptible to landslides due to the steep slopes and wet climate, with over 40 historic landslides recorded in the national database [30].

3. Methods

We compare the performance of five different models: (i–iii) generalized globally trained predictive models (CCDC, Tehrani, and Prakash), (iv) a locally trained supervised machine learning model in Google Earth Engine (GEE) (smile.Cart classifier), and (v) a locally trained pixel-based deep learning model (U-net). For verification of the results, we used a set of 120 manually mapped landslides [30].

3.1. Generalized Globally-Trained Predictive Models

To run the three generalized globally trained predictive models, the steps in the methods and accompanying documentation on GitHub were followed, with some modifications made where necessary. A summary of these methods and any deviations are described here.

(i) CCDC time-series model [15]: The CCDC model results are available for visualization purposes as an app on GEE. The results have been pre-calculated for the Landsat bands (not including NDVI). The SWIR1 band was chosen for the change detection analysis, as this is known to be sensitive to changes in vegetation. The changes within the period 1 July 2019 to 31 August 2019 were displayed for the study area using the app.

(ii) Tehrani machine learning model [29]: Pre-processing of the input data is performed automatically using a script run in GEE [41]. The script takes a table of landslide coordinates and dates, and generates sets of Geotiff images for each point, which are then used as input for the model. Sentinel-2 Level 1C images with low cloud coverage are selected within three months before and after the landslide event date. If no cloud-free images are found in that period, a composite image is made using images from one year. The pre-processing involves the normalization of the images, and the addition of brightness, NDVI, and GNDVI (green-NDVI) bands, and the output is three images for each landslide point; pre-event, post-event, and difference (see Figure 3I). Modifications made to this process included uploading a shapefile to GEE, instead of a Google Fusion table, which has been discontinued. Further, the 10 m resolution Norwegian DEM was used instead of the global 30 m resolution ALOS DEM, because the ALOS DEM does not cover Norway.

The outputs are raster images with labeled segments in KEA file format [42] and a list of the segments that were classified as landslides.

(iii) Prakash CNN deep learning model [16]: The required inputs are three bands (R, G, B) pre- and post-event images, single-band slope, hillshade, DEM, bounding box, and no-data mask rasters. In the accompanying article [16], it was not specified if Sentinel-2 Level 1C or Level-2A products were used. Pre-processing the input images involved selecting a Sentinel-2 image at the landslide location with the lowest cloud cover within one month of the landslide date, clipping to the area of interest, and then manually creating a mask of snow and clouds. Again, we used the Norwegian DEM instead of a global DEM, from which slope and hillshade rasters were created.

Greenest-pixel composite image: One modification to the methodology described in Prakash et al. [16] was to use a greenest-pixel composite as input, as this method can reduce noise from clouds and agriculture. These were produced using one month of images from before and after the landslides, using the S2 cloudless algorithm for cloud filtering and the SCL (scene classification) band for snow filtering. Using the quality mosaic function, a composite image was then created based on NDVI, in which for each pixel—the pixel with the maximum NDVI is taken, along with the corresponding values from the other bands from the same date. This gives a 'greenest pixel' composite that is cloud-free and gives the least snow cover and shadow within the specified date range.

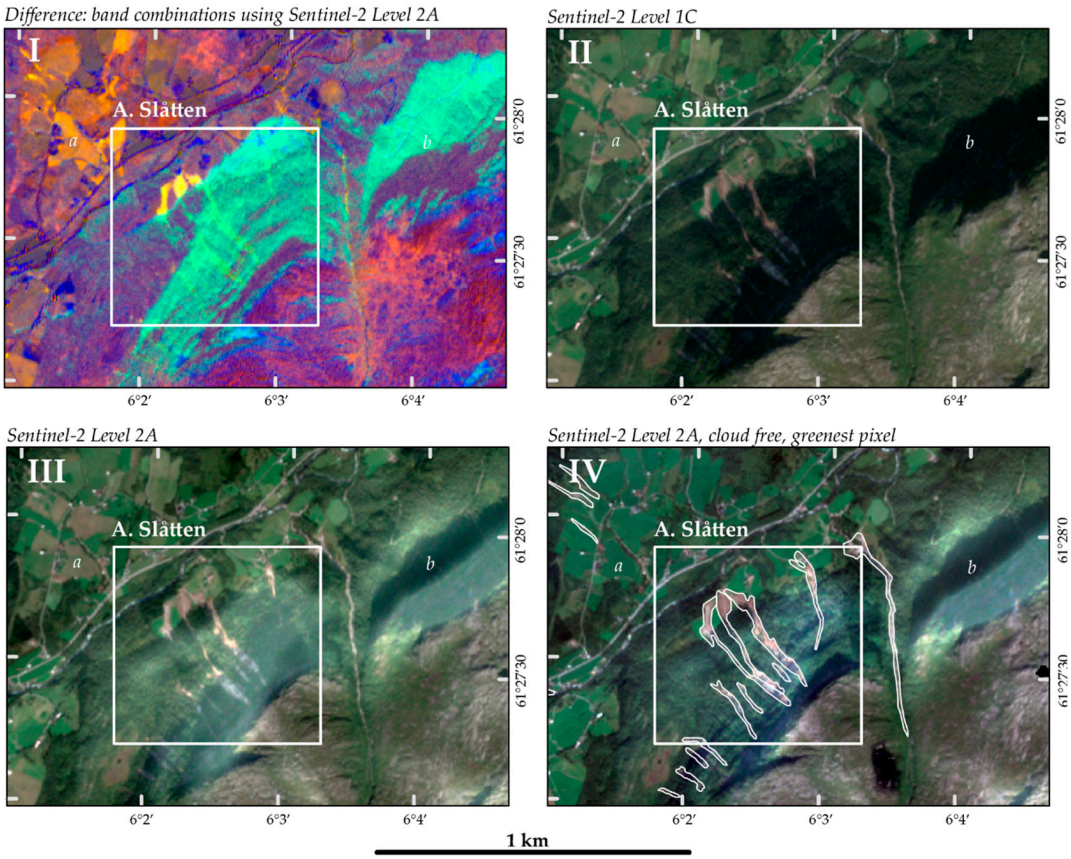


Figure 3. Optical image inputs derived from Sentinel-2 images shown for subset A. Slåtten. The letters a and b indicate areas of agriculture and shadows respectively, where a difference is observable between the four types of input images. (**I**) Difference image with three bands (brightness, red-over-green, and NDVI derived from Level-2A images) used for the Tehrani model. (**II**) Sentinel-2 Level 1A Top of Atmosphere (TOA). (**III**) Sentinel- Level 2A, with atmospheric correction applied to the Level-1C TOA image. Note that the shadowed areas at point b have been brightened. (**IV**) Cloud-filtered, greenest-pixel composite produced from Sentinel-2 Level2A images. Images (**II**–**IV**) were used as inputs in the Prakash model, while the locally trained model based on U-net architecture used only image (**IV**), along with Sentinel-1 images.

3.2. Locally-Trained Machine and Deep Learning Models

(iv) smile.CART machine learning model: Landslide predictions were performed in Google Earth Engine using the ee.Classifier.smileCart algorithm [31], which uses a CART (Classification and Regression Trees) classifier. This involved the following steps.

First, the images were pre-processed. For Sentinel-2, one month of Level-2A images from before and after the event were used to create cloud-filtered, greenest-pixel composites. Cloud filtering was performed using the s2cloudless algorithm to remove cloudy pixels [43]. Greenest-pixel composites were then created from the pre- and post-event image collections, using the quality mosaic function. All bands from the same image the selected pixel with the highest NDVI value was taken from are included in the output. For Sentinel-1, again, one month of images from pre- and post-event were used to create terrain-corrected median composites. The terrain correction was performed using the volumetric scattering

model [44]. Then, median composites for VV and VH bands separately were created from each of the pre- and post-event image collections, using both ascending and descending orbit geometries. Finally, the Sentinel-1 and -2 bands were combined, along with elevation and slope, into a single 13-band image.

Secondly, the supervised classification was performed following the tutorial by S. Levick [45]. This involved selecting training points from which to train the classifier—18 points were manually selected in the landslide class, and 112 points were from seven different non-landslide classes (water, snow/ice, bare rock, agriculture, forest, alpine scrub, and urban). Care was taken to sample from diverse slope aspects, elevations, and within shadow areas. The 13-band image was then sampled at each point, and these values were used to train the classifier and perform a classification across the whole image. In addition, classifications were performed using the same points for a 3-band, and 2-band subsets of the full 13-band image, as shown in Table 1.

Table 1. Input data used in locally trained models for four different settings.

Model Run	No. of Bands	Bands
S1, S2, and DEM	13	Sentinel-1: pre-VV, post-VV, diff-VV, pre-VH, post-VH, diff-VH
		Sentinel-2: post-R, post-G, post-B, post-NIR, dNDVI
		Terrain: elevation, slope
S1 (VV) and S2	3	Sentinel-1: pre-VV, post-VV
		Sentinel-2: dNDVI
S1 (VV) only	2	pre-VV, post-VV
S2 only	5	post-R, post-G, post-B, post-NIR, dNDVI

The results were inspected to see if any misclassification was apparent. Then, finally, a binary image of landslide–non-landslide was produced by combing the non-landslide classes, and salt-and-pepper noise was reduced using the focal mode function.

(v) U-net CNN deep learning model: The entire algorithm was implemented in a Jupyter Notebook using ArcPy, Keras, and TensorFlow 2. The chosen model is a scaled-down version of a deep-learning architecture called U-net, for automatic semantic segmentation [46] with Keras implementation. The U-net is a convolutional network architecture for fast, effective, and precise segmentation of images with its symmetric U-shape. U-net has proven to be a powerful segmentation tool in scenarios with limited data, simple structure, and high recognition accuracy. The network is based on the fully convolutional neural network (FCNN) for semantic segmentation [47,48].

The same input dataset was used for the GEE smile.CART model is described in Section 3.2 (Table 1). We exported random samples as classified tiles for all four settings by generating a minimum of 10,000 samples using an Image Analyst license and ArcGIS Pro [49]. The most suitable tile size in our case was 128 × 128 pixels, and stride (the distance to move in the x- and y-directions when creating the next image chips) of 64 × 64, to have 50% overlap in each sample tile. The output was a dataset of classified image tiles, the format primarily used for pixel classification. During the training process, an input image flows through the CNN network that recognizes it with a set of trainable kernels, resulting in a group of feature maps [50]. The dataset was divided into training, validation, and test subsets. The trained model was saved as a 'Deep Learning Package' ('.dlpk' format), which is the standard format used to deploy deep learning models on the ArcGIS platform and can be used further as a pre-trained model.

3.3. Performance Evaluation

The landslide inventory produced from the Sentinel-2 dNDVI [51] image was used for verifying the results of the other approaches.

Qualitative and quantitative analyses: The results of the CCDC and Tehrani models are briefly described in a qualitative manner, as both these methods produced limited landslide predictions. Additionally, it was not possible to download the CCDC model results from the GEE app; therefore, it was not possible to do quantitative pixel scale analyses on these results.

The Prakash and locally trained modelpredictions were evaluated quantitatively, as follows. The landslide polygons mapped with Sentinel-2 dNDVI were converted to a binary raster of landslide or non-landslide pixels. This was used to validate the automated landslide detection model outputs. Following the approach in ref [16], 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, and MCC scores were calculated (see Table 2). Since landslides represent only a tiny fraction of all the pixels in the study area, the learning problem is highly imbalanced towards non-landslide pixels. Therefore, the accuracy score can become unreliable due to the large proportion of true negatives. The MCC score is considered to be the most appropriate metric for comparing the results [52]. For a binary model, the MCC gives a score between 0 to 1, with 0 indicating a model with no correlation (random predictions) and 1 indicating a perfect correlation (all correct predictions).

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	$\frac{2TP}{2TP+FP+FN}$
MCC	$\frac{TP \times TN - FP \times FN}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$

4. Results

The performance of the three globally-trained and two locally-trained machine learning models in the Jølster case study is presented in this section.

4.1. Globally Trained Models

(i) CCDC time-series model:

The CCDC time-series model detected the large Vassenden landslide quite precisely (Figure 4), along with one smaller debris flow to the east. The large landslides at Årnes and Slåtten were not detected, nor were any of the other smaller landslides in the study area.

(ii) Tehrani machine learning model:

Overall, no landslides were detected using this model. The large landslide at Vassenden was partially segmented (see Figure 4), although the initiation zone was missed, and some nearby fields were included. The deposits of the landslides at Slåtten were also segmented. However, no segments were classified as a landslide.

(iii) Prakash CNN deep learning model:

The Prakash model was run with three different variations of Sentinel-2 images shown in Figure 3 (1) Level 1C, (2) Level 2A, and (3) a cloud-free greenest-pixel composite. The results are shown as confusion matrix maps in Figure 5. The most striking differences between the runs were firstly, that many false positives (wrongly predicted as a landslide) appear in the Level 2A products, and secondly, in the Level 1C product, there are many false negatives (missed landslides).

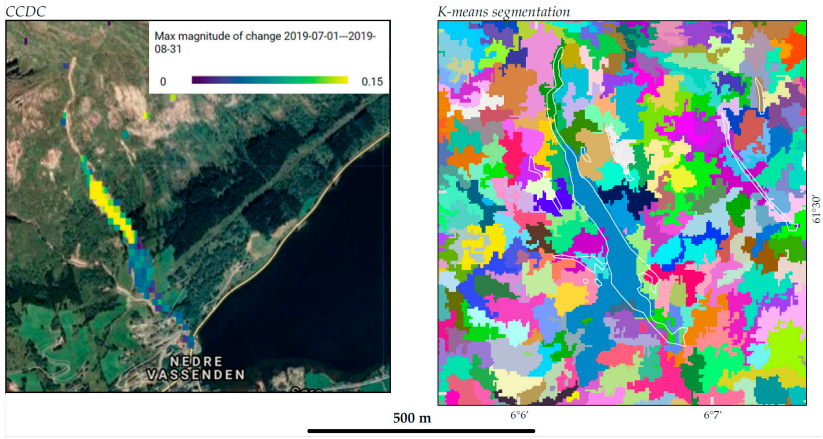


Figure 4. Results of CCDC and Tehrani models. (**Left**): CCDC results: NDVI band change detection between 1 July and 31 August 2019. CCDC results included with permission (Arévalo, P.; pers. comm. 2022). Background: Google. (**Right**): Output results of k-means segmentation. White polygon outlines show the manually mapped landslides used for verification.

After inspecting the results, it was noticed that the false positives in the Level 2A results appeared to be related to unnaturally bright areas on the northern slopes. This turned out to be due to an anomaly resulting from the terrain correction used in processing the Level 2C products, which results in a blueish appearance in shadowed areas in true color composite images and inaccurate surface reflectance values [53]. The problem has been reported to the Sentinel-2 Quality Working Group (December 2021) [54]. An outcome of their analysis is expected in the near future; changes will be reported in the Sentinel 2A Data Quality Report (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library, accessed on 15 December 2021). Due to the noise introduced by these artifacts, comparing the performance metrics of the model runs over the entire study area was not very insightful. Therefore, for a more detailed comparison, the metrics were calculated for the sub-plots shown in Figure 5. These are shown in Table 3.

For the entire study area, the MCC scores are below 1%. In the subplots, there are false positives likely caused by the over-correction artifact in the results for A, B, and D, which are on north-facing slopes. Despite this, the best score was 51% for model Input 2—S2_L2A in subplot D (Årnes landslide). The second best was 43% for model Input 2—S2_L2A in subplot C (Vassenden landslides). Subplots B (Svidalen) and A (Slåtten) had poor results across all runs.

Using Level-1C images, overall, the model failed to detect landslides. Only a small part of the Vassenden landslide was detected. The large landslides on north-facing slopes were not detected at all. There were some false positives, mainly related to changes in agricultural areas. Using Level 2A images (2. single date, and 3. greenest-pixel composite), the model predicted the Vassenden and Årnes landslides fairly well. The landslides in subplot A (Slåtten) were partially detected with the Level-2A images. However, most of the predictions on the steep north-facing slopes are false positives due to noise, while the deposit of the western-most of the three debris flows seems to have been detected meaningfully. It is interesting that that particular deposit was detected, and not the other two, given that from field observations, the deposits of the western-most debris flow were noticeably different from the others. The western-most deposit was a very thin layer of soil, with a high concentration of washed-out light-colored boulders and stones, whereas the other two were much thicker (up to 2 m high) deposits consisting of darker soil and forest debris (seen https://www.nibio.no/nyheter/skogsdrift-ikke-medvirkende-arsak-til-jordras-i-jolster, accessed on 22 December 2022)). In subplot B (Svidalen), there is a

significant difference in the number of false positives, with much fewer in the greenest-pixel composite from run 3, compared to run 2. Again, it is not clear if this difference is due to the artifacts or pre-processing. There appears to be just one pixel that has been correctly identified in all three runs. However, overall, the model was not able to detect the smaller landslides. In subplot C (Vassenden), there are more false positives using the greenest-pixel composite in agricultural areas than the single date image. These results are more likely to be meaningful because the slope is south-facing and not affected by the over-correction artifacts. Finally, in subplot D (Årnes), there are slight differences in the number of false positives between the two input image types; however, it is difficult to say whether the difference is related to the artifacts or to the difference between the manually masked image (single date), and the greenest-pixel composite.

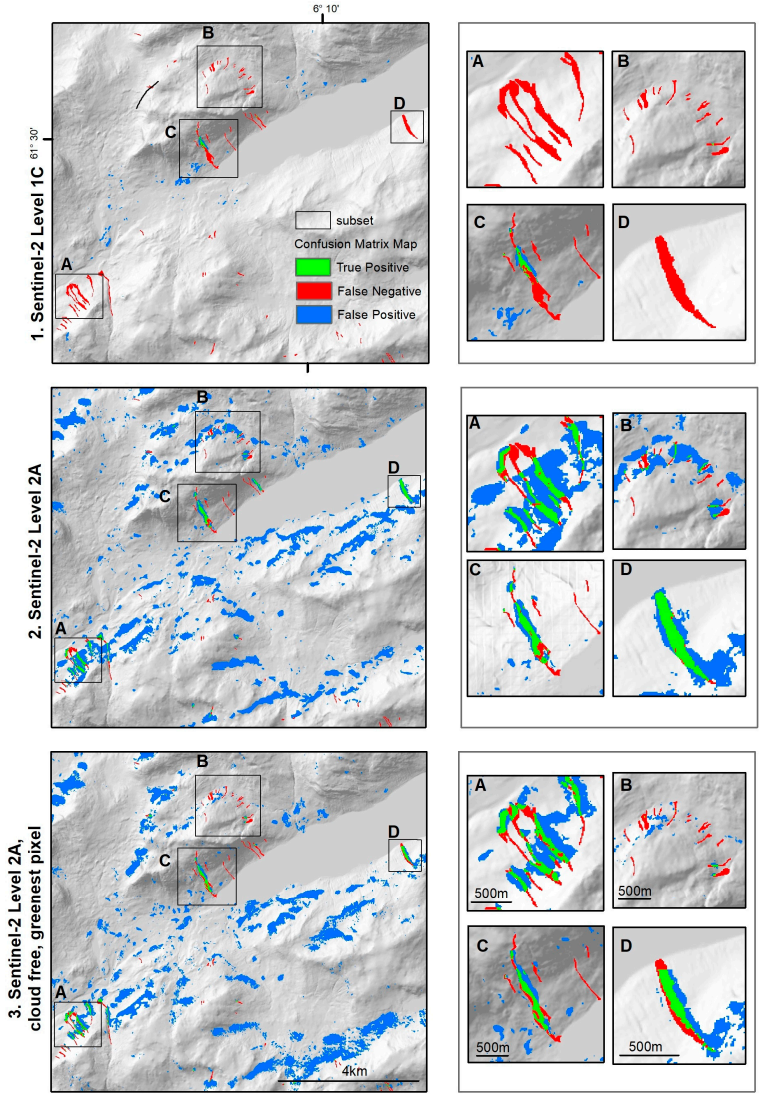


Figure 5. Performance results from the Prakash CNN deep learning as a confusion image, from the three different layer settings: (**1**) Level 1C, (**2**) Level 2A, and (**3**) a cloud-free greenest-pixel composite.

Table 3. Prakash CNN deep learning model (iii) performance metrics. The metrics were calculated for the entire study area, as well as for the four subplots shown in Figure 5. The model was run with three different input image types; 1. Level-1C images, 2. Level-2A images, and 3. Level-2A images as a cloud-free, greenest pixel composite. The MCC score (Matthew's Correlation Coefficient) is considered the most representative metric for the imbalanced problem of landslide classification [52].

Location	Input Image	Precision %	Recall %	F1-Score %	MCC %
Entire area	1—S2_L1C	5	4	4	4
	2—S2_L2A	2	45	5	9
	3—S2_L2A_gr	2	37	4	7
A. Slåtten	1—S2_L1C	40	0	0	2
	2—S2_L2A	19	60	29	20
	3—S2_L2A_gr	30	58	40	33
B. Svidalen	1—S2_L1C	86	1	1	8
	2—S2_L2A	6	28	9	8
	3—S2_L2A_gr	8	6	7	5
C. Vassenden	1—S2_L1C	25	17	21	18
	2—S2_L2A	40	51	45	43
	3—S2_L2A_gr	35	46	40	37
D. Årnes	1—S2_L1C	-	0	0	-
	2—S2_L2A	33	96	49	51
	3—S2_L2A_gr	35	60	44	41

The mediocre performance in these model runs is mainly due to introduced image artifacts in shadowed areas, thererfore we find the Prakash CNN deep-learning model is worth further investigation for use in an operational landslide detection system. With different adjustments, such as using input images without the over-corrected shadow areas and including NDVI or Sentinel-1 bands, to make the classification more robust in shadowed areas, the model performance could likely be improved.

4.2. Locally Trained Models

(iv) smile.CART machine learning model:

The supervised machine learning model in GEE using the ee.smile.Cart classifier was tested with different layer settings. We observed that some of the input data combinations yielded promising predictions; particularlywith setting 2 (dNDVI, pre-event S1-VV images, post-event S1-VV images), which had an MCC score of 73% (Figure 6, Table 4). The poorest result was obtained by the third combination using only Sentinel-1 VV-polarised SAR images as input. Although, the landslides were detected equally well as with only Sentinel-2 data as input (recall 72%), the overall MCC score was only 20%, due to the abundant false positives from speckle noise.

(v) U-Net CNN Deep learning model

The final approach, the locally trained deep learning model showed the best overall predictions. The values for the MCC score varied from 51–89%, and the precision results were between 80–85%. Setting 2: (using dNDVI, pre-event S1-VV images, and post-event S1-VV images showed the highest values for MCC score (89%), recall (79%), F1 score (81%) and and the best visual prediction (Table 2, Figure 7). Setting 3 (Sentinel-1 VV-polarised data only) was second best, with an MCC score of 79%. Although, with the use of only Sentinel-1 images in Setting 3, in subplot A. Slåtten, the upper part of the landslides is visibly not predicted. Most of the landslides are predicted correctly however, with some missing pixels around that were not predicted as a landslide (FN). Including the DEM (Setting 1, all 13 bands) introduced significant amounts of mostly false negatives (red) with poor prediction results. None of the small-sized landslides were predicted in this setting.

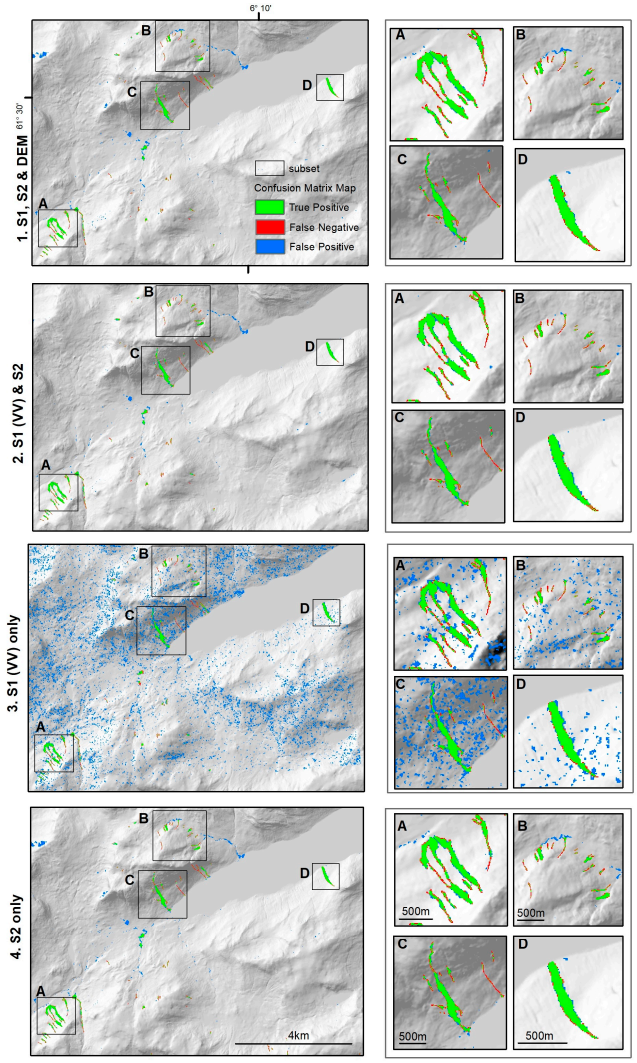


Figure 6. Perfomance results from the locally trained smile.CART machine learning model as a confusion image, from four different input data combinations. Setting: (**1**) full version (all 13 bands) (**2**) dNDVI, preVV, postVV (**3**) preVV, postVV (**4**) post-R, post-G, post-B, post-NIR, and dNDVI.

Table 4. Performance metrics for landslide detection for the locally trained models using four different input data combinations.

MODEL		Setting 1	Setting 2	Setting 3	Setting 4
		S1, S2 & DEM	S1 (VV) & S2	S1 (VV) only	S2 only
(iv) CART	precision %	62	72	6	59
	recall %	73	74	72	72
	F1 %	67	73	11	65
	MCC	63	73	20	65
(v) U-Net CNN	precision %	80	83	85	84
	recall %	33	79	74	73
	F1 %	47	81	79	78
	MCC	51	89	79	78

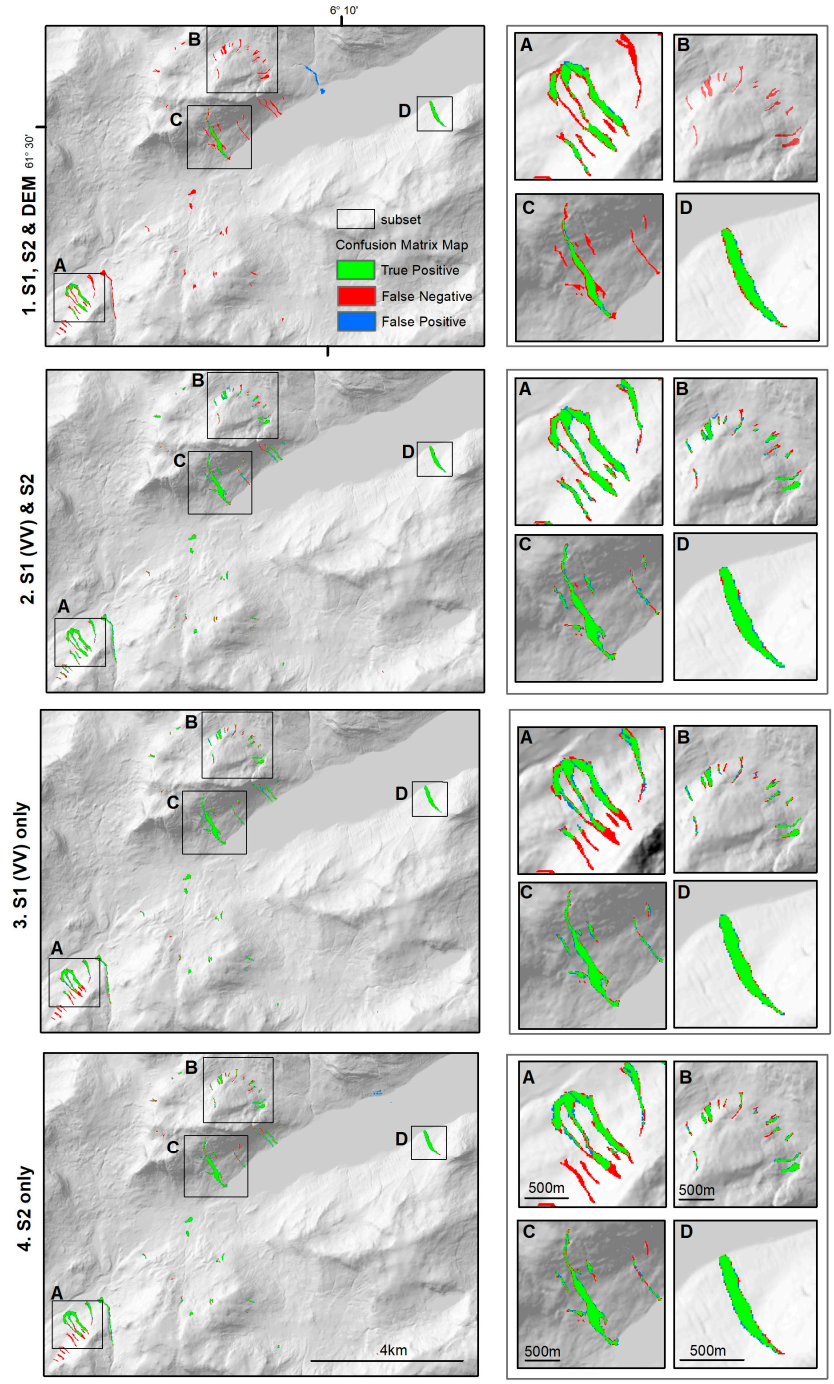


Figure 7. Performance results from the locally trained U-Net CNN deep learning model, as a confusion image from four different input data combinations. Setting: (**1**) full version (all 13 bands) (**2**) dNDVI, preVV, postVV (**3**) preVV, postVV (**4**) post-R, post-G, post-B, post-NIR, and dNDVI.

5. Discussion

5.1. Performance of Globally Pre-Trained Machine learning Models in a Glacial Landscape

Overall, the generalized models tested did not perform very well. Only the largest landslides were detected by these models. In most of the tests, the results appeared to be affected by the slope aspect and over-correction of shadow artifacts on north-facing slopes in the Sentinel-2 Level-2A products.

The CCDC model (i), despite not being designed specifically for landslide detection, showed good potential for applying time-series-based change detection methods for continuous landslide monitoring. The large landslide at Vassenden was outlined quite precisely, within the 30 m resolution of the Landsat data. However, it failed to detect the large landslides on north-facing slopes (i.e., in subsets A. Slåtten, and D. Årnes) and only detected one other landslide clearly. The results were very simple to view using the Google Earth Engine app [15]. Furthermore, by extending the time period visualized, the app allowed the user to quickly identify other landslides outside of the study area which occurred within the past 20 years. Although CCDC is designed for monitoring land cover changes generally [14], some modifications (e.g., running with NDVI, Sentinel-2, and perhaps Sentinel-1 data) could enable it to be used as part of a continuous landslide monitoring service.

Using the Tehrani model (ii), only the large landslide at Vassenden was visible in the segmentation results, although it was not classified as a landslide. This method used the Sentinel-2 Level-1C images as input. From the different runs with the Prakash model, it was observed that the landslides are detected more frequently when using the atmospherically corrected Level-2A products compared to Level-1C, especially for landslides on north-facing slopes. Thus, it can be speculated that the landslide detection on north-facing slopes may have been improved by using the Level-2C product. However, as seen from the results of the Prakash model runs with the Level-2C product, the anomalies caused by terrain over-correction on shadowed areas using the Level-2C product may also have introduced false positive predictions. The Tehrani model was also trained using landslides that were over 1000 km^2, and the minimum size of pixel clusters was 80. Including more small landslides in the training data set and adjusting the minimum size of pixel clusters may improve the detection of smaller landslides. Adjusting the number of k-means, or perhaps training with different indices, may improve the performance of the random forest classification.

The performance of the Prakash model (iii) was strongly affected by the Sentinel-2 product type, with very limited correct landslide detection with Level-1C (high levels of false negatives), and improved landslide detection with Level-2A, however with the introduction of significant areas of false positives due to the terrain over-correction anomaly. Due to these false positives, the difference between using the single image inputs (run 2), compared to the greenest-pixel cloud composite inputs (run 3), was not clear, even when examining the image at the resolution of the subsets. The landslide predictions were not as precise as in the CCDC and Tehrani models. To better understand the performance of this model using different inputs, it is recommended to wait for the reprocessing of the Sentinel-2 Level 2 images. This model could potentially be improved by retraining the classifier with more Norwegian landslide data and by including a greater range of bands and vegetation indices.

5.2. Comparison of Locally Trained Machine Learning and Deep Learning Models and Input Data Combinations

The U-net deep learning model-(v) outperformed the CART machine learning model-(iv) for three out of four input data combinations. These findings are in agreement with similar machine learning vs. deep learning model comparison studies for landslide detection [52,53]. The best MCC score achieved for our study area was 89%, using the three-band combination of pre-VV and post-VV from Sentinel-1 and dNDVI from Sentinel-2. In both the Sentinel-1-only and Sentinel-2-only input data settings, we found that the model could not recognize the landslide signature in the initiation zones of the landslides at Slåtten (subset A). We did not find any other landslide detection studies in the literature

where both Sentinel-1 and -2 data have been used to train a deep learning model. However, our results are in agreement with a similar study on illegal logging detection [27].

For S2-only deep learning, the false negatives appear in the shadowed area. The signature of the landslides is very clear from the dNDVI image only, even where shadows are present (Figure 8). It is possible that the inclusion of RGB bands reduces the performance of the classifier in this area.

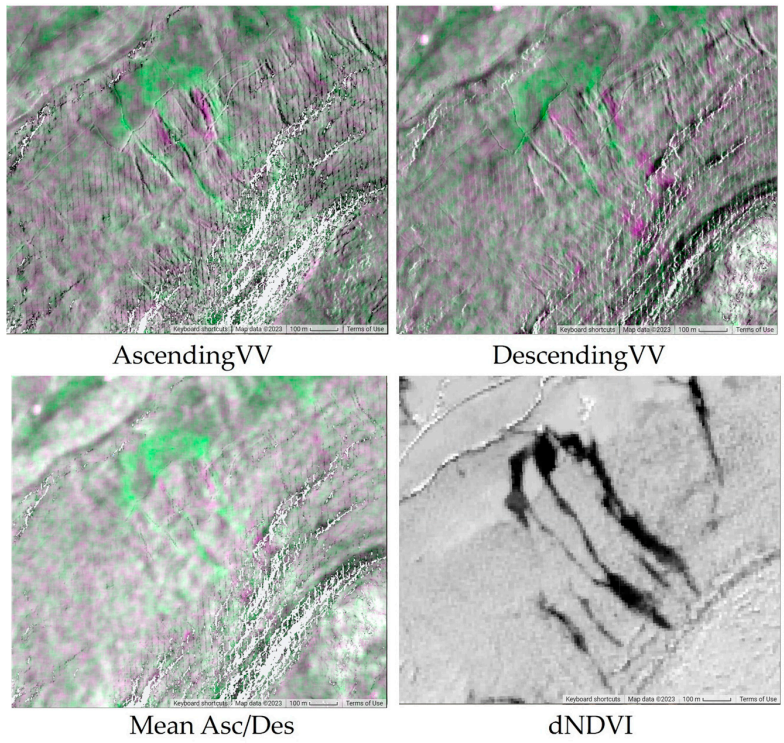


Figure 8. Landslides at Slåtten (subset A) in multi-temporal VV-polarized SAR backscatter intensity change images (ascending, descending, and mean) change in NDVI (**bottom right**). Green indicates a backscatter intensity increase; purple indicates a decrease. White outlines were mapped from the Sentinel-2 dNDVI image.

We believe, in the case of S1-only, that the false negatives are due to the landslide expression (i.e., the pattern of increase and decrease in backscatter intensity) in this location being different from other areas. When viewed separately in ascending and descending images (Figure 8), landslides in forested areas show both decreased backscatter intensity on the side of the landslide nearest to the sensor as well as a wide parallel band of increased backscatter intensity on the far side [51]. Averaging the ascending and descending images tends to produce a final post-event image that shows mainly increased backscatter intensity in the area of the landslides. Yet here, the landslides are expressed in the input SAR images by strongly decreased backscatter intensity relative to the pre-event image, and the decrease was not 'averaged out' in this case. It is also possible that geometric distortions in the descending image and DEM distortions also affect the results, as they produce gaps in the image.

We suspect that the performance of the classifier is strongly affected by the combination of ascending and descending images due to the simplification of the landslide signature and averaging out of most areas with decreased backscatter intensity. This is important to note for others considering following this approach, as local vegetation conditions,

landslide type, and geometry, as well as slope orientation relative to the sensor, can affect how landslides are expressed in SAR backscatter intensity data [51]. We did not test the U-net model using separate ascending and descending images as input; however, this would be interesting to compare.

The deep learning model had a significant advantage over the machine learning model for the Sentinel-1 only input data setting, with MCC scores of 73% and 20%, respectively. The machine learning model could detect the changes due to landslides with only Sentinel-1 data; however, there were many false positives due to speckle noise. Setting 2 (preVV, postVV, and dNDVI) performed better than Setting 1 (DEM, Sentinel-1, Sentinel-2), possibly because of overfitting when using all 13 bands or because the resolution of the slope map does not show the steep slopes on small-scale objects. Landslides have deposits in flat areas which makes it possible to detect landslides predicted in both flat and steep areas. In contrast, the deep learning model uses a sliding window approach and is capable of differentiating speckle noise from landslide signatures. This is because the deep learning model makes the decision whether each pixel is a landslide by taking into consideration the pattern of pixels in the patch surrounding the pixel being classified. In this way, whether the pixel is part of a cluster of pixels or is isolated and therefore, more likely to be random noise due to speckle. Mondini et al. [23] noted that while, in principle, SAR data are well suited for identifying landslides, SAR imagery remains underutilized for landslide detection. This is due in part to the reduced clarity of landslide signatures caused by speckle noise, as well as the side-looking sensor angle which also makes the images harder to interpret. The comparison of the conventional machine learning model with the deep learning model in this study shows how one of the main barriers to using SAR imagery for landslide detection can be reduced using a deep learning model.

5.3. Recommendations for an Operational Landslide Detection System and Future Research

Landslide detection and mapping are undertaken for different purposes, including (i) rapid emergency response; and (ii) inventory creation for use in spatial analyses (e.g., for hazard and susceptibility mapping or deriving local thresholds for early warning) or verification and improvement of landslide early warnings. Each of these situations has different priorities for the timeliness and accuracy of landslide data needed. The recommendations based on the findings of this study and relevant literature are organized accordingly.

(i) *Rapid emergency response:* the priority is to detect landslides as quickly as possible, while accurate delineation and mapping are of lower importance. In this situation, we recommend the use of SAR-only models, as there is no need to wait for cloud-free conditions at the time of writing; no globally trained SAR-based landslide detection models are available. Therefore, a locally trained model is needed. If a local landslide inventory of polygon data is available, then CNN models such as U-net give much higher performance than a conventional machine learning model due to their ability to differentiate speckle noise from landslide signatures. Using the methodology presented in this study, landslide predictions could be produced within three hours of the SAR image becoming available, but it requires computational power and a GPU. Where no local landslide inventory is available, the simple locally trained machine learning approach using Google Earth Engine performed in this study could be repeated for a new area in around 30 min after the image is available in GEE. This method requires only internet access and a free GEE account.

(ii) *Inventory creation:* the priority is for accurate and complete landslide data (including date, size information), while rapid detection is of lower importance. For automatically delineating landslides, optical or multispectral images combined with terrain-corrected multi-temporal SAR data with the best possible resolution are recommended. The locally trained U-net deep learning approach gave the best performance in a glacial setting. The globally trained models did not perform well in our study area due to shadows. The best performance would be achieved using images from a similar season and could be performed over large areas as an annual systematic survey. For obtaining date information,

a time-series approach based on SAR data would be useful, as it is possible to back-date landslide occurrences when the location is known.

A continuous monitoring system for landslide detection requires further research, particularly in terms of the spatial and temporal signatures of landslides in SAR data and how these vary in different environmental settings. Compared to deforestation, the problem of landslide detection is more complex because landslides can occur in a range of different land cover types, and their expression can also vary depending on seasonal conditions. Ongoing developments in data availability and pre-processing of images will provide many more options to explore. These include the NISAR satellites due to be launched in 2023 with L and X band SAR capabilities [54]. Additionally, improvements to the pre-processing of the Sentinel-2 images may result in better predictions in the generalized machine learning models we tested.

In working towards developing a system for continuous detection of landslides over large areas, the GEE platform is very suitable, as multiple datasets (e.g., optical, SAR, soil moisture, precipitation, slope, and land cover type) can be combined and analyzed performed quickly over large areas. Furthermore, there is a possibility to incorporate an external cloud-based TensorFlow model, as used by Prakash et al., within the workflow. The CCDC model is designed for continuous monitoring. Modifying the CCDC model (e.g., using Sentinel data and masking to show only vegetation loss) would be a good start. Training data should also include examples with areas likely to cause false negatives, e.g., with forestry or agricultural activity resulting in vegetation loss.

6. Conclusions

The locally trained models outperformed the globally trained models at detecting landslides in a glacial setting. The best result was achieved using the deep-learning approach with a U-net architecture and input data, including a difference in NDVI (normalized difference vegetation index) from Sentinel-2 and pre- and post-event SAR data (terrain-corrected, mean of multi-temporal ascending descending images in VV polarization) from Sentinel-1.

The generalized globally trained machine-learning-based models did not perform very well for landslide detection in a glacial landscape. The model from Prakash showed good potential to be applied in Norway; however, it would require retraining and further development to perform well in the local conditions. The model performance could be improved by retaining the NIR band, which is more robust in shadow areas.

High rates of false negatives (missed landslides) were the main source of error for the CCDC, Tehrani, and the Prakash model run using Sentinel-2 Level-1C images. In contrast, the Prakash model runs using Sentinel-2 Level 2A images resulted in high rates of false positives, mainly due to over-brightened artifacts on north-facing slopes introduced by a terrain over-correction. The results likely could be improved by (a) rerunning the tests when the reprocessed data are released by Copernicus, (b) including Norwegian training data, and (c) further development of the methods.

For the development of an operational landslide detection system, a SAR-only-based approach using a deep-learning model is recommended for rapid detection as part of an emergency response due to the capability to observe landslides despite the cloud cover. In contrast, for detailed mapping and back-dating of landslides, a combination of SAR and optical data can give improved performance over optical data alone, and the time-series approaches can be used for continuous monitoring or to back-date landslides.

Author Contributions: Conceptualization, E.L. and A.J.G.; methodology, E.L. and A.J.G.; software, E.L. and A.J.G.; validation, E.L. and A.J.G.; formal analysis, E.L. and A.J.G.; investigation, E.L. and A.J.G.; resources, E.L.; data curation, E.L. and A.J.G.; writing—original draft preparation, E.L. and A.J.G.; writing—review and editing, E.L., A.J.G. and J.K.R.; visualization, E.L. and A.J.G.; supervision, O.F., T.-A.M., S.N. and J.K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Council of Norway and several partners through the Centre for Research-based Innovation 'Klima 2050' (Grant No 237859) (see www.klima2050.no), as well as through the Norwegian Geotechnical Institute through its basic funding from the Norwegian Government (GBV 2020 & 2021).

Data Availability Statement: The locally trained GEE model is available: https://code.earthengine.google.com/91d0606e7797198754ec2a16d7333fb9 (accessed on 22 December 2022).

Acknowledgments: Thank you to Paulo Arevalo for providing the CCDC results for our study area. This paper contains modified Copernicus Sentinel data [2020] processed by Sentinel Hub, and Planet Scope Data provided by the European Space Agency and Planet under project: 61234—Landslide detection using satellite data. Gabriela Spakman-Tanasescu for introducing the possibilities of ArcGIS Pro and deep learning. Thank you to Regula Frauenfelder for reviewing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [CrossRef]
2. Tehrani, F.S.; Calvello, M.; Liu, Z.; Zhang, L.; Lacasse, S. Machine learning and landslide studies: Recent advances and applications. *Nat. Hazards* **2022**, *114*, 1197–1245. [CrossRef]
3. Reiche, J.; Mullissa, A.; Slagter, B.; Gou, Y.; Tsendbazar, N.-E.; Odongo-Braun, C.; Vollrath, A.; Weisse, M.J.; Stolle, F.; Pickens, A.; et al. Forest disturbance alerts for the Congo Basin using Sentinel-1. *Environ. Res. Lett.* **2021**, *16*, 24005. [CrossRef]
4. Hansen, M.C.; Krylov, A.; Tyukavina, A.; Potapov, P.V.; Turubanova, S.; Zutta, B.; Ifo, S.; Margono, B.; Stolle, F.; Moore, R. Humid tropical forest disturbance alerts using Landsat data. *Environ. Res. Lett.* **2016**, *11*, 34008. [CrossRef]
5. Vargas, C.; Montalban, J.; Leon, A.A. Early warning tropical forest loss alerts in Peru using Landsat. *Environ. Res. Commun.* **2019**, *1*, 121002. [CrossRef]
6. Katiyar, V.; Tamkuan, N.; Nagai, M. Near-real-time flood mapping using off-the-shelf models with SAR imagery and deep learning. *Remote Sens.* **2021**, *13*, 2334. [CrossRef]
7. Ban, Y.; Zhang, P.; Nascetti, A.; Bevington, A.R.; Wulder, M.A. Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning. *Sci. Rep.* **2020**, *10*, 1322. [CrossRef]
8. Zhang, P.; Ban, Y.; Nascetti, A. Learning U-Net without forgetting for near real-time wildfire monitoring by the fusion of SAR and optical time series. *Remote Sens. Environ.* **2021**, *261*, 112467. [CrossRef]
9. Devoli, G.; Bell, R.; Cepeda, J. *Susceptibility Map at Catchment Level, to Be Used in Landslide Forecasting*; Norwegian Water Resources and Energy Directorate: Oslo, Norway, 2019.
10. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]
11. Arévalo, P.; Bullock, E.L.; Woodcock, C.E.; Olofsson, P. A Suite of Tools for Continuous Land Change Monitoring in Google Earth Engine. *Front. Clim.* **2020**, *2*, 576740. [CrossRef]
12. Prakash, N.; Manconi, A.; Loew, S. A new strategy to map landslides with a generalized convolutional neural network. *Sci. Rep.* **2021**, *11*, 9722. [CrossRef] [PubMed]
13. Ghorbanzadeh, O.; Gholamnia, K.; Ghamisi, P. The application of ResU-net and OBIA for landslide detection from multi-temporal sentinel-2 images. *Big Earth Data* **2022**, 1–26. [CrossRef]
14. Nava, L.; Bhuyan, K.; Meena, S.R.; Monserrat, O.; Catani, F. Assessment of deep learning based landslide detection and mapping performances with backscatter SAR data. In Proceedings of the EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022. [CrossRef]
15. Bai, L.; Li, W.; Xu, Q.; Peng, W.; Chen, K.; Duan, Z.; Lu, H. Multispectral U-Net: A Semantic Segmentation Model Using Multispectral Bands Fusion Mechanism for Landslide Detection. In Proceedings of the 2nd Workshop on Complex Data Challenges in Earth Observation, Vienna, Austria, 25 July 2022.
16. Dong, Z.; An, S.; Zhang, J.; Yu, J.; Li, J.; Xu, D. L-Unet: A Landslide Extraction Model Using Multi-Scale Feature Fusion and Attention Mechanism. *Remote Sens.* **2022**, *14*, 2552. [CrossRef]
17. Fang, C.; Fan, X.; Zhong, H.; Lombardo, L.; Tanyas, H.; Wang, X. A Novel historical landslide detection approach based on LiDAR and lightweight attention U-Net. *Remote Sens.* **2022**, *14*, 4357. [CrossRef]
18. Nava, L.; Monserrat, O.; Catani, F. Improving Landslide Detection on SAR Data through Deep Learning. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
19. Kamiyama, J.; Noro, T.; Sakagami, M.; Suzuki, Y.; Yoshikawa, K.; Hikosaka, S.; Hirata, I. Detection of Landslide Candidate Interference Fringes in DInSAR Imagery Using Deep Learning. *Recall* **2018**, *90*, 94–95.
20. Mondini, A.C.; Guzzetti, F.; Chang, K.-T.; Monserrat, O.; Martha, T.R.; Manconi, A. Landslide failures detection and mapping using Synthetic Aperture Radar: Past, present and future. *Earth Sci. Rev.* **2021**, *216*, 103574. [CrossRef]

21. Bullock, E.L.; Healey, S.P.; Yang, Z.; Houborg, R.; Gorelick, N.; Tang, X.; Andrianirina, C. Timeliness in forest change monitoring: A new assessment framework demonstrated using Sentinel-1 and a continuous change detection algorithm. *Remote Sens. Environ.* **2022**, *276*, 113043. [CrossRef]
22. Doblas, J.; Reis, M.S.; Belluzzo, A.P.; Quadros, C.B.; Moraes, D.R.V.; Almeida, C.A.; Maurano, L.E.P.; Carvalho, A.F.A.; Sant'Anna, S.J.S.; Shimabukuro, Y.E. DETER-R: An operational near-real time tropical forest disturbance warning system based on Sentinel-1 time series analysis. *Remote Sens.* **2022**, *14*, 3658. [CrossRef]
23. Shumilo, L.; Kussul, N.; Lavreniuk, M. U-Net Model for Logging Detection Based on the Sentinel-1 and Sentinel-2 Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4680–4683.
24. Kirschbaum, D.B.; Adler, R.; Hong, Y.; Hill, S.; Lerner-Lam, A. A global landslide catalog for hazard applications: Method, results, and limitations. *Nat. Hazards* **2010**, *52*, 561–575. [CrossRef]
25. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: New York, NY, USA, 2017.
26. NGU Landslides. Available online: https://www.ngu.no/en/topic/landslides (accessed on 21 December 2022).
27. Luigi, S.; Guzzetti, F. Earth-Science Reviews Landslides in a changing climate. *Earth Sci. Rev.* **2016**, *162*, 227–252. [CrossRef]
28. Herrera, G.; Mateos, R.M.; Garcia-Davalillo, J.C.; Grandjean, G.; Poyiadji, E.; Maftei, R.; Filipciuc, T.-C.; Auflič, M.J.; Jež, J.; Podolszki, L.; et al. Landslide databases in the Geological Surveys of Europe. *Landslides* **2018**, *15*, 359–379. [CrossRef]
29. Jaedicke, C.; Lied, K.; Kronholm, K. Integrated database for rapid mass movements in Norway. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 469–479. [CrossRef]
30. Malamud, B.D.; Heijenk, R.A.; Taylor, F.E.; Wood, J.L. Road Influences on Landslide Inventories. In Proceedings of the EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022. [CrossRef]
31. Ruther, D.C.; Hefre, H.; Rubensdotter, L. Extreme precipitation-induced landslide event on 30th 3 July 2019 in Jølster, western Norway. *Nor. J. Geol.* **2022**, *102*, 202212.
32. Lindsay, E.; Frauenfelder, R.; Rüther, D.; Nava, L.; Rubensdotter, L.; Strout, J.; Nordal, S. Multi-Temporal Satellite Image Composites in Google Earth Engine for Improved Landslide Visibility: A Case Study of a Glacial Landscape. *Remote Sens.* **2022**, *14*, 2301. [CrossRef]
33. Meteorologisk Institutt. *Rapport om Intense Byger med store Konsekvenser i Sogn og Fjordane 30. juli*; Meteorologic Institute: Bergen, Norway, 2019; Available online: https://www.met.no/nyhetsarkiv/rapport-om-intense-byger-med-store-konsekvenser-i-sogn-og-fjordane-30.juli (accessed on 22 December 2022).
34. Devoli, G.; Colleuille, H.; Sund, M.; Wasrud, J. Seven Years of Landslide Forecasting in Norway-Strengths and Limitations. In *Understanding and Reducing Landslide Disaster Risk: Volume 3 Monitoring and Early Warning*; Casagli, N., Tofani, V., Sassa, K., Bobrowsky, P.T., Takara, K., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 257–264. ISBN 978-3-030-60311-3.
35. Hanssen-Bauer, I.; Drange, H.; Førland, E.J.; Roald, L.A.; Børsheim, K.Y.; Hisdal, H.; Lawrence, D.; Nesje, A.; Sandven, S.; Sorteberg, A.; et al. Climate in Norway 2100. 2009. Available online: https://www.researchgate.net/profile/Ingjerd-Haddeland/publication/316922280_Climate_in_Norway_2100/links/59194fab4585152e19a24c98/Climate-in-Norway-2100.pdf (accessed on 22 December 2022).
36. Tehrani, F.S.; Santinelli, G.; Herrera, M.H. Multi-Regional landslide detection using combined unsupervised and supervised machine learning. *Geomat. Nat. Hazards Risk* **2021**, *12*, 1015–1038. [CrossRef]
37. Herrera Herrera, M. *Landslide Detection Using Random Forest Classifier*; Delft University of Technology: Delft, The Netherlands, 2019.
38. Bunting, P.; Gillingham, S. The KEA image file format. *Comput. Geosci.* **2013**, *57*, 54–58. [CrossRef]
39. Braaten, J. Sentinel-2 Cloud Masking with s2cloudless. Available online: https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless (accessed on 22 December 2022).
40. Vollrath, A.; Mullissa, A.; Reiche, J. Angular-Based Radiometric Slope Correction for Sentinel-1 on Google Earth Engine. *Remote Sens.* **2020**, *12*, 1867. [CrossRef]
41. Levick, S.R. Lab 4-Image Classification-part 1. *Remote Sens.* **2017**, *9*, 329.
42. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
43. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
44. Ye, J.; Ni, J.; Yi, Y. Deep Learning Hierarchical Representations for Image Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2545–2557. [CrossRef]
45. Huang, S.-C.; Le, T.-H. Introduction to TensorFlow 2. In *Principles and Labs for Deep Learning*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 1–26.
46. Liu, Y.H. Feature Extraction and Image Recognition with Convolutional Neural Networks. *J. Phys. Conf. Ser.* **2018**, *1087*, 062032. [CrossRef]
47. ArcGIS Pro Export Training Data For Deep Learning (Image Analyst). Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/export-training-data-for-deep-learning.htm (accessed on 22 December 2022).

48. Lindsay, E.; Devoli, G.; Reiches, J.; Nordal, S. In Progress: Spatial and Temporal Signatures of Landslides in C-Band SAR Data. 2023.
49. Chicco, D.; Jurman, G. *The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–13.
50. Clerc, S. MPC-Team Terrain over-correction on shaded areas. In *S2 MPC Level 2A Data Quality Report*; ESA: Paris, France, 2022; Volume 45, p. 28.
51. Jackson, J. Clarification on Difference between L1C and L2A Data. Available online: https://forum.step.esa.int/t/clarification-on-difference-between-l1c-and-l2a-data/24940/12 (accessed on 22 December 2022).
52. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens.* **2019**, *11*, 196. [CrossRef]
53. Prakash, N.; Manconi, A.; Loew, S. Mapping landslides on EO data: Performance of deep learning models vs. Traditional machine learning models. *Remote Sens.* **2020**, *12*, 346. [CrossRef]
54. NASA Quick Facts. Available online: https://nisar.jpl.nasa.gov/mission/quick-facts/ (accessed on 22 December 2022).

Article IV.

Understanding Landslide Expression in SAR Backscatter Data: A Global Study

Erin Lindsay [1*], Alexandra Jarna Garnerød [2,3], Graziella Devoli [4], Johannes Reiche [5], Steinar Nordal [1], Regula Frauenfelder [6], Lars-Christian Tokle [7]

1* Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway; erin.lindsay@ntnu.no (E.L.); steinar.nordal@ntnu.no (S.N.)
2 Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway; alexandra.jarna@ntnu.no (A.J.G.);
3 Geological Survey of Norway (NGU), 7040 Trondheim, Norway; alexandra.jarna@ngu.no
4 Norwegian Directorate of Energy and Water, Oslo, Norway. g.de@nve.no (G.D.)
5 Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands; johannes.reiche@wur.nl (J.R.);
6 Norwegian Geotechnical Institute (NGI), 0806 Oslo, Norway; regula.frauenfelder@ngi.no (R.F.);
7 Department of Mathematics and Cybernetics, Norwegian University of Science and Technology, 7034 Trondheim, Norway; lars-christian.n.tokle@ntnu.no (LT)
* Corresponding author: erin.lindsay@ntnu.no

Article

Understanding Landslide Expression in SAR Backscatter Data: A Global Study

Erin Lindsay [1*], Alexandra Jarna Garnerød[2,3], Graziella Devoli[4], Johannes Reiche[5], Steinar Nordal[1], Regula Frauenfelder[6], Lars-Christian Tokle[7]

[1*] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7034 Trondheim, Norway; erin.lindsay@ntnu.no (E.L.); steinar.nordal@ntnu.no (S.N.)
[2] Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway; alexandra.jarna@ntnu.no (A.J.G.);
[3] Geological Survey of Norway (NGU), 7040 Trondheim, Norway; alexandra.jarna@ngu.no
[4] Norwegian Directorate of Energy and Water, Oslo, Norway. g.de@nve.no (G.D.)
[5] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands; johannes.reiche@wur.nl (J.R.);
[6] Norwegian Geotechnical Institute (NGI), 0806 Oslo, Norway; regula.frauenfelder@ngi.no (R.F.);
[7] Department of Mathematics and Cybernetics, Norwegian University of Science and Technology, 7034 Trondheim, Norway; lars-christian.n.tokle@ntnu.no (LT)
[*] Corresponding author: erin.lindsay@ntnu.no

Abstract: During disaster response, clouds or darkness can prevent the use of optical images for detecting consequences of natural disasters, including landslides. In these situations, radar images can be used to detect changes more rapidly. However, Synthetic Aperture Radar (SAR) backscatter intensity images are underutilized for landslide detection. Unfortunately, there remains a lack of understanding about how to interpret landslide signatures in SAR imagery. In this study, we investigate how the morphometric features and material properties of landslides, and preexisting land cover, control their expression in SAR backscatter intensity change images. Trends in the spatial and temporal signatures of over 1000 landslides in 30 diverse case studies are investigated, using multi-temporal composites and dense time-series of Sentinel-1 C-band SAR backscatter intensity data. The results show that the orientation of landslide surfaces relative to the sensor, pre-existing land cover, and the roughness of the landslide surface, determine whether landslides will produce an increase or decrease in backscatter intensity values. In certain cases, we can identify morphometric features of landslides (e.g. scarps, transit zone, deposits, ponding) and material properties. Generally, we see that landslides appear most clearly with a strong increase in intensity when they occur in herbaceous vegetation or non-vegetated ground surfaces, due to an increase in surface roughness. While in forested or densely vegetated areas, landslides produce a more complex signature with both decreases due to radar shadow and vegetation removal, and an adjacent edge of increased intensity due to double bounce and direct return from vertical tree trunks and convex edges. In most cases, rough deposits produce an increase in intensity, while smooth deposits (e.g. from mudslides) exhibit specular reflection, and thus show decreased values. Landslides are less visible in cases with pre-event very rough ground, or mixed vegetation conditions. The conceptual model developed can aid interpretation of landslides in SAR imagery, and provide domain knowledge needed to train models for automatic landslide detection.

Keywords: change detection; time-series; landslide detection; land cover; Sentinel-1; backscatter; Google Earth Engine;

 doi:10.20944/preprints202302.0390.v1

1. Introduction

Landslides pose an increasing risk to the human population due to urban expansion into unstable areas, and an increasing frequency of extreme precipitation events (Froude and Petley, 2018; Gariano and Guzzetti, 2016; Hanssen-Bauer et al., 2009). Landslides frequently occur in conjunction with other types of natural hazards including earthquakes, floods, volcanic eruptions and tsunamis, particularly in areas with high hydrological or seismic risk (Casagli et al., 2017). Consequences may include loss of life, damages to infrastructure and property, blocked transport routes and dammed rivers which create further risk of flash flooding to downstream areas (Kjekstad and Highland, 2009).

Satellite-based emergency mapping can provide highly beneficial information to the humanitarian response community on the location, size and density of landslides. If performed within hours to days after an event, this information can improve situational awareness for responders, particularly if conducted as part of a coordinated mapping effort (Kedia et al., 2022; Williams et al., 2018). The increasing availability of medium to high resolution satellite images, combined with cloud-based high performance computing capabilities and automatic image detection methods have facilitated recent development of automated landslide detection models (Ghorbanzadeh et al., 2019; Prakash et al., 2021). Preliminary studies have yielded promising results at a local scale particularly with deep-learning methods. However, further development is needed to achieve automatic landslide detection models that perform well in diverse environments (Ganerød et al., 2023). Eventually, operational real-time monitoring and alert systems, such as presently exist for detection of illegal deforestation (Reiche et al., 2021), could be developed for landslides.

The majority of the existing landslide detection models rely on optical images and vegetation indices (Guzzetti et al., 2012). However, such analyses require cloud-free images and sufficient light conditions. In tropical and high latitude regions of the earth, persistent cloud cover, or seasonal darkness, can lead to delays of months before changes to the ground surface are observable in optical satellite images (Lacroix et al., 2018; Sudmanns et al., 2020). This is especially problematic for rapid detection of rainfall-triggered landslides, which typically occur under cloudy conditions (Williams et al., 2018).

Synthetic Aperture Radar (SAR) is cloud penetrating, therefore ground surface reflectance properties can be measured irrespective of cloud cover or illumination from the sun (ASF, 2022). However, compared to optical imagery, SAR data are more difficult to process and interpretation is less intuitive. This is due to a complex backscatter signal, speckle noise, phase wrapping and geometric distortions (Meyer, 2019). This presents an obstacle for geoscientists and other operators in using SAR backscatter images for rapid landslide detection (Handwerger et al., 2022; Mondini et al., 2019). Although there is a well-established community using In-SAR for monitoring slow-moving deep-seated landslides (Cigna et al., 2014; Mondini et al., 2021; van Natijne et al., 2022), the same methods are not frequently applied for detecting rapid landslides due to a lack of coherence in the phase component of the returning radar signal in vegetated areas (Mondini et al., 2019).

(Guzzetti et al., 2012) asserted that the real part (amplitude, or backscatter intensity) of the SAR signal should be exploited in order to detect ground changes caused by landslides. A review of 54 journal articles about landslide failure detection using SAR imagery, published between 1995 and 2020 (Mondini et al., 2021), found that literature on the exploitation of SAR for landslide event detection remains limited. They conclude that in principle, the available SAR amplitude images are well suited

for landslide detection and mapping as the wavelength of the microwave SAR sensors is comparable to the length scales of morphometric elements typical of landslides. However, there remains a lack of understanding of the physical basis of backscatter response.

Recently, an increasing number of studies have begun to investigate the potential for using SAR-backscatter intensity change images for landslide detection and mapping using Sentinel-1 C-band SAR backscatter images. (Mondini et al., 2019) investigated the potential for detecting landslides in Sentinel-1 backscatter change images for 32 case studies. (Santangelo et al., 2022) explored how accurately two large single landslide events could be mapped using such images. Others, have used multi-temporal composites to improve landslide visibility in change images (Lindsay et al., 2022) and create heat maps of landslide density (Handwerger et al., 2022). (Burrows et al., 2022) used Sentinel-1 time-series to backdate monsoon-triggered landslides and explored the performance of different bands. There have also been a few attempts to automate this process using locally-trained deep-learning models (Ganerød et al., 2023; Nava et al., 2022).

While most of these studies test multiple band combinations (VV and VH) and orbit geometries (ascending or descending), interpretation of the spatial and temporal expression of landslides is still somewhat limited in detail. The majority of these studies focus on whether it is possible to detect, map and back-date landslides using SAR imagery. They do not systematically quantify the change in backscatter intensity, or relate it to the landslide type, morphometric elements, or environmental conditions. Most investigate landslides in densely forested areas, while areas with herbaceous or little to no vegetation, especially polar regions, are poorly represented (Mondini et al., 2021). Landslides are observed to produce both increases and decreases in backscatter intensity; however, a physically based explanation for these observations in relation to the ground properties is lacking.

In this study, we investigate the spatial and temporal signatures of different types of rapid landslides in diverse physiographical environments in 30 locations, using C-band SAR backscatter change images and time-series data. In doing so, we aim to improve understanding of how the physical changes to the ground surface relate to the change in backscatter intensity and scattering mechanisms, that determine the expression (pattern of increase or decrease) of landslides in SAR imagery. We consider the reflective properties of the pre-event ground surface, and the post-event landslide surface. In addition to identifying individual landslides, in some cases we are also able to distinguish separate morphometric features of landslides (i.e. scarps, sliding surfaces, deposits, and ponds) based on their specific expressions in the SAR change images. Based on these trends and new empirical data, we present a conceptual model that can be used to aid interpretation. The findings contribute towards developing standards for interpreting landslides in SAR backscatter change images. In addition, a diverse global dataset that can be used for training automatic landslide detection models was created.

The **research questions** are as follows:

1. What patterns and trends in the spatial and temporal expression of landslides in SAR backscatter intensity data can be identified from the 30 case studies?
2. Which factors control the visibility and expression of landslides in SAR data?

To answer these questions, we first consider how SAR theory can be related to landslide detection to identify the possible controlling factors. Then we examine the spatial and temporal signatures of

landslides, from 30 case studies with diverse landslide types and environmental conditions, and perform statistical analyses of the controlling factors. Based on the identified trends, and results from the statistical analyses, a physically-based conceptual model is presented to aid interpretation of landslides in SAR change images.

2. SAR backscatter theory applied to landslides

Changes to the ground surface caused by landslide erosion and deposition may change the surface roughness and dielectric properties of the ground surface and thus the intensity of backscatter. This information can be used to detect and map landslides. Several factors that can affect the visibility and expression of landslides in SAR backscatter intensity data are illustrated and described in this section. These include radar properties (wavelength **(**λ**)**, polarisation and incidence angle), and surface properties (terrain elevation and geometric distortions, local incidence angle (LIA), roughness (relative to λ**)**, and land cover type,) (Fung et al., 1992). Seasonal variations in vegetation or snow cover, and changes in moisture content, can also affect the backscatter intensity.

In Figure 1 (A), it is shown how the position of a landslide in the terrain relative to the sensor, determines whether the landslide signal will be affected by geometric distortions (including layover, shadow and foreshortening). It is seen that Landslide-I would be visible to the sensor, however distorted due to layover. While Landslide-II, would not be visible to the sensor as it is in the shadow zone. The distortion of Landslide-I can be corrected with a terrain correction, while the Landslide-II can be detected if images from both ascending and descending orbits are available.

In Figure 1 (B) the effect of the orientation of the landslide surface, relative to the sensor line of sight (LOS) is considered. Here, it is shown that the local incidence angle (LIA) affects the strength of the received backscatter signal, with surfaces with lower LIA generally returning stronger intensity signals than those with high LIA. The strength of the returned signal also depends on the ground surface properties, as will be shown in Figure 2, with surface scatters being more sensitive to the LIA than volumetric scatters. Figure 1 (C) illustrates how the wavelength of the sensor determines the height of irregularities (h [cm]) that the signal will be sensitive to, as well as the degree of penetration of vegetation and the elements of vegetation that the signal will interact with. Shorter wavelength signals (i.e. X- and C-band) are more sensitive to smaller changes in surface roughness (< 5cm), and are mainly reflected from the canopy. Longer wavelengths (i.e. L-band) are sensitive to larger scale changes in roughness (>10 cm), and penetrate leaves, thus are reflected from woody structures and the ground.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

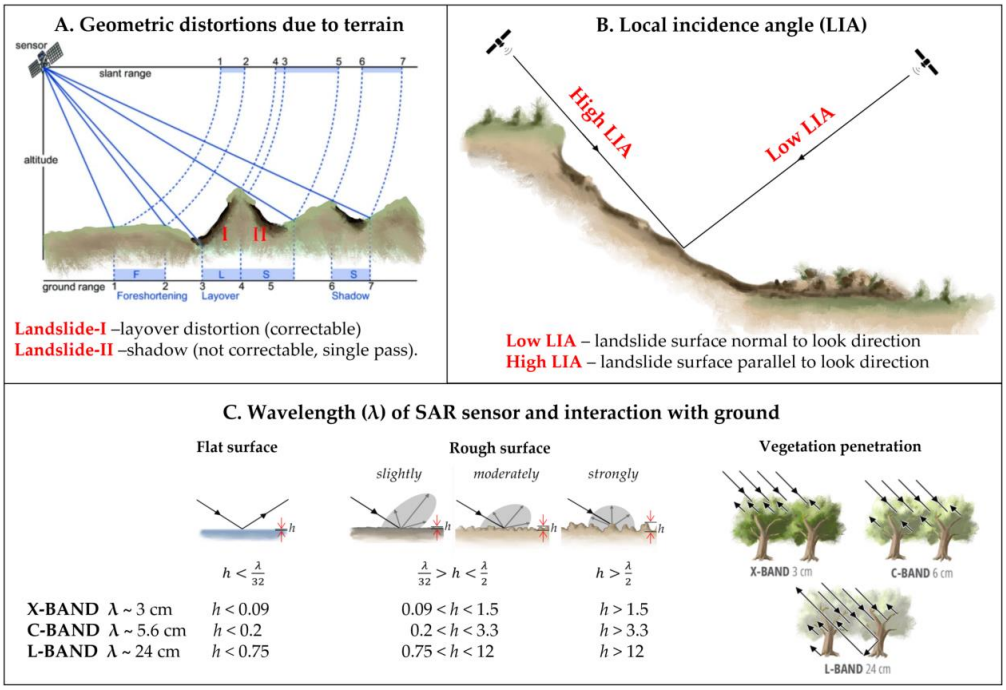


Figure 1 Factors affecting the expression of landslides in SAR data that remain static over time. These include the position of the landslide within the terrain relative to **(A)** terrain features, and **(B)** the sensor line of sight (LOS), and **(C)** the wavelength of the SAR signal (λ). Modified from (A) (Tempfli et al., 2009), (B) (Lindsay et al., 2022), (C) (Kellndorfer et al., 2019). The roughness is relative to the wavelength, *h* is the height of surface irregularities.

In Figure 2 (D), different ground-cover types and their associated scattering mechanisms are shown. Here we see that the intensity of the received backscatter increases with increasing surface roughness, with flat surfaces (e.g. still water, or snow) with oblique LIAs, reflecting the signal specularly in accordance with Snell's law (Ulaby et al., 2019). As surface roughness increases, diffusivity increases and more of the signal is reflected in all directions, including back towards the sensor. Co-polarised bands (e.g. VV) are more sensitive to variation in surface roughness. For volumetric scattering as occurs in vegetation, in C-band, the canopy provides a strongly reflective surface, with most of the energy received reflected volumetrically within the upper few cm of the canopy. Cross-polarised bands (e.g. VH) are more sensitive to variation in vegetation (volume scattering). For areas with mixed types of scatters, i.e. herbaceous vegetation, agriculture, or areas with both herbaceous and woody vegetation, the strength of the received signal depends on (i) the degree to which the signal penetrates the vegetation, and (ii) the roughness of the underlying soil surface. For thin herbaceous vegetation (e.g. grass, peat, or low bio-mass crops) the underlying surface roughness determines the intensity of the backscatter received, while for ground with predominantly volumetric scatters (e.g. dense leafy crops, small bushes) the biomass of the vegetation will have a stronger effect on the intensity. Finally, double bounce scatterers (near-vertically inclined surfaces, e.g. cliffs, or exposed tree trunks) produce the highest received backscatter intensity, and as surfaces, these features are most strongly observed in co-polarised bands.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

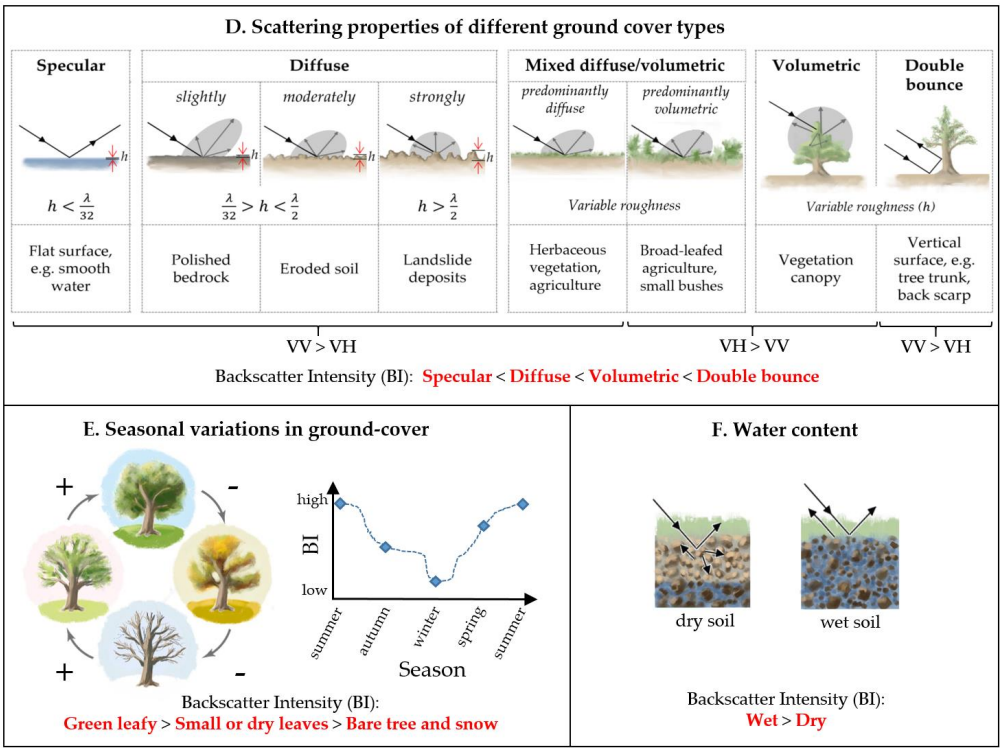


Figure 2 Factors affecting the expression of landslides in SAR data that may vary over time. These include **(D)** the type of ground cover and associated scattering mechanisms, **(E)** seasonal variations in ground cover, including vegetation changes and snow cover, and **(F)** water content

It is the change between the different types of ground cover types that determines the change in the observed intensity. For instance a change from a volumetric scatter (e.g. forest canopy) to a surface scatterer (e.g. soil surface) will result in a decrease in intensity and the edges of the remaining forest may also produce radar shadow (Bouvet et al., 2018). While for surfaces, an increase in roughness caused by deposition or erosion of the weathered soil surface, will produce an increase in intensity.

In Figure 2 (E) we see for a given ground cover type, the intensity can vary significantly if there are strong seasonal variations. For instance, intensity is highest when the leafy canopy reflects the SAR signal volumetrically, and is lower in autumn and spring when leaves are small or dry. Smooth snow reflects the signal away specularly, resulting in lower intensity. Finally in Figure 2 (F), we see that increased water content (of soil, snow, or vegetation) produces higher intensity than dry ground covers, due to increased conductivity, decreased penetration of the wave into the ground surface, and thereby increased reflectivity. The magnitude of intensity change that a landslide produces, will increase or decrease depending on seasonal variations of the preexisting ground cover.

By understanding how terrain, ground-cover, wavelength and polarisation produce backscatter signals of varying intensities, we can apply these ideas to predict how landslides may be expressed in change images and time-series data. Here we assume that landslides produce a change in ground surface cover, across an area of a size that is detectable within the resolution of the sensor. The types of ground-cover

observed within a landslide area can be related to the morphometric features of landslides, landslide material, as well as failure mechanisms, and LIA.

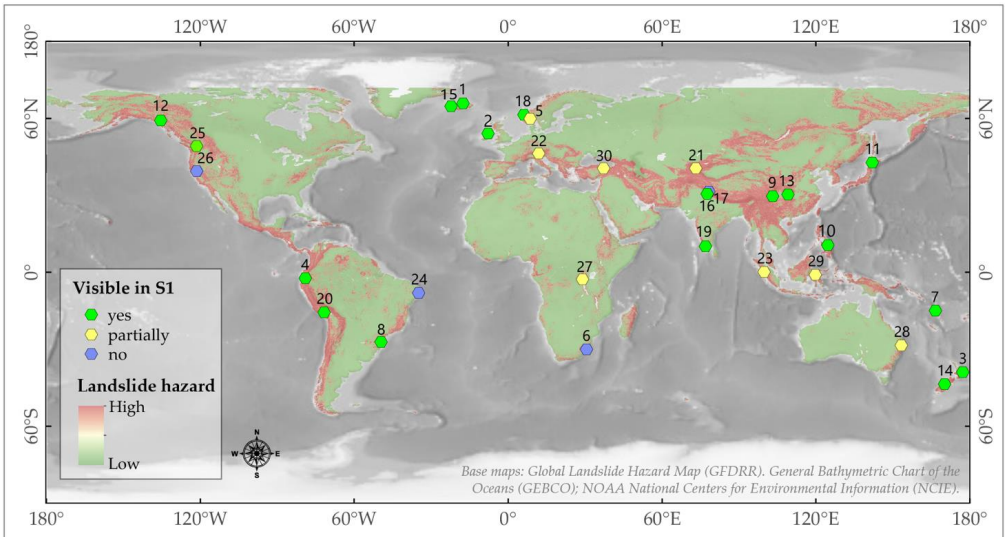
As illustrated above, several factors can affect the expression of landslides in change images. The roughness of the landslide surface can vary depending on the type of material (soil or rock), soil particle coarseness (silt to gravel, and presence of rocks, boulders or other debris), and whether material was completely or partially eroded from the surface, and the deposition mechanism. The strength of the received backscatter for a smooth scarp or eroded surface (height of irregularities, $h < 0.2$ cm, for C-band) is strongly affected by the LIA. Scarps facing away from the LOS, that dip more steeply than the incidence angle, will be in a radar shadow zone and show very low intensity. The height of the scarp can be estimated based on the width of the shadow zone and the sensor incidence angle (Arnold et al., 2018). Smooth scarps with a high LIA (approximately parallel to LOS will either reflect the signal away from the sensor, resulting in a low intensity, while rough surfaces ($h > 3$ cm, for C-band) will scatter the signal diffusely and are less sensitive to the LIA. Scarps with a low LIA (approximately normal to LOS) will reflect more signal back towards the sensor. Landslide deposits are typically extremely rough surfaces consisting of transported material including rock and soil, water, and eroded vegetation. An exception are earth flows (mudflows), which consist of very fine particles and can form very smooth deposits. Where the topographic change is less significant, the specific change in surface properties will determine whether landslides are expressed by an increase or decrease in intensity. Landslides may dam rivers or streams, resulting in upstream swamps, ponds, or lakes. These areas will show strongly decreased intensity.

3 Case studies

A selection of 30 case studies from different locations, with varying terrain types, orientation and size, ground cover, climate zones, geological materials and failure mechanisms were systematically analysed in this study. The case studies were mostly identified from reports in 'The Landslide Blog' (https://blogs.agu.org/landslideblog/) (Dave Petley, 2022), or from news reports, or journal articles found by internet search. The location and dates of the investigated landslide events are shown in Figure 3, while the properties of the landslides and local environment are shown in Table 1.

The data used to describe the landslide properties was retrieved from the following sources:

- **Location and date:** literature (various sources, see appendix 1)
- **Landslide type and trigger:** from the reports describing the article, or classified from the descriptions and images according to (Hungr et al., 2014).
- **Landslide size and aspect:** the outline of the largest landslide in the study area was measured.



Investigated events (Location, event date, coordinates)

1. Iceland	2021-10-06	(-17.5558, 65.8674)	16. India	2021-07-25	(78.30762, 31.41447)
2. Ireland	2021-07-04	(-7.85187, 54.15745)	17. India	2021-07-30	(77.658109, 30.589764)
3. New Zealand	2022-03-27	(177.4694, -38.9681)	18. Norway	2019-04-28	(8.6385996, 59.84685)
4. Ecuador	2021-02-12	(-78.93655, -2.2895)	19. India	2020-08-06	(77.01188, 10.166)
5. Norway	2019-07-30	(6.11699, 61.51734)	20. Peru	2020-06-16	(-71.685998, -15.66025)
6. South Africa	2022-04-12	(30.69701, -29.89816)	21. Kyrgzstan	2017-04-29	(73.290135, 40.587722)
7. Vanuatu	2020-04-06	(166.63854, -14.9281)	22. Italy	2019-11-23	(8.434000, 44.295000)
8. Brazil	2020-04-06	(-49.60777, -27.0774)	23. Indonesia	2022-02-25	(99.99442, 0.11412)
9. China	2021-04-05	(103.15055, 29.66474)	24. Brazil	2022-05-26	(-34.958299, -8.120723)
10. Phillipines	2022-04-10	(124.85286, 10.67658)	25. Canada	2021-11-14	(-121.67202, 49.31537)
11. Japan	2018-09-06	(141.966, 42.7876)	26. USA	2021-11-24	(-121.38802, 39.86029)
12. USA	2019-12-02	(-135.41194, 59.22035)	27. Burundi	2019-12-05	(29.080122, -2.619565)
13. China	2020-07-21	(109.29972, 30.36592)	28. Australia	2022-02-28	(153.39578, -28.51734)
14. New Zealand	2022-02-03	(170.17485, -43.57436)	29. Indonesia	2018-09-28	(119.91418, -0.93951)
15. Iceland	2022-02-02	(-22.109811, 64.78794)	30. Turkey	2019-05-16	(37.378026, 40.670806)

Figure 3 Map of case study locations and event dates. Base map: Landslide Hazard Map (GFDRR, 2023)

For each set of coordinates in Figure 3, values from the following maps were extracted using Google Earth Engine or ArcMap:

- **Geology:** Generalised Geology of the World, WMS V 1.3.0 (Chorlton, 2007),
- **Climate zone:** World Map of the Köppen-Geiger climate classification (Kottek et al., 2006)
- **Mean annual rainfall:** WorldClim BIO Variables V1 (Hijmans et al., 2005)
- **Land cover classification:** CORINE (EEA/Copernicus, 2012) and Copernicus Global Land Cover (Copernicus, 2019)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

Table 1: Properties of the investigated landslides and their local environment. The visibility of the landslides are also presented according to their assigned set. Set 1: not visible, set 2: partially visibly, recognisable only knowing a priori the location, and set 3: clearly visible.

	Landslide				Environment				Set
Location	Type	T	Size L x W [km]	Aspect	Geology	K.G. Climate	Rainfall [mm/yr]	Land cover Gl / EU	1 No 2 Part. 3 Yes
1. Iceland	DS	R	0.8 x 0.1	E	V	Cfc	672	Herb./ Moor	3
2. Ireland	PF	R	0.58 x 0.7	NW	S	Cfb	1358	Herb. / Peat	3
3. N. Zealand	DS	R	0.13 x 0.05	W	S	Cfb	1508	Herbaceous	2
4. Ecuador	EF	R	1.5 x 1.5	W-NW	S-V	Cfb	918	F|Unknown	3
5. Norway	DF, DA	R	0.11 x 0.03	mixed	M	Dfc	2285	Herbaceous	3
6. Sth. Africa	DF	R	0.5 x 0.2	W	M	Cfa	940	Herbaceous	1
7. Vanuatu	DS-DF	R	0.8 x 0.2	S	V	Af	3440	F|Broadleaf	2
8. Brazil	DF	R	1.6 x 0.02	NE	S	Cfa	1547	F|Broadleaf	2
9. China	DS-DF	R	1.34 x 0.92	S	S	Cwb	1297	F|Broadleaf	3
10. Philippines	MS	R	2.1 x 0.7	SW	S	Af	2915	F|Broadleaf	3
11. Japan	DS, DF	ER	0.22 x 0.13	mixed	S-V	Dfb	1131	F|Broad. dec.	3
12. USA	DA	R	1.7 x 0.18	N	P	Dsb	1282	F|Needle	3
13. China	DS	R	1.2 x 0.3	S	S	Cfa	1409	F|Unknown	3
14. N. Zealand	RA	R	1.8 x 0.28	SE	S	ET	4222	Snow	3
15. Iceland	RA	R	2.4 x 1.7	SE	V	Cfc	829	Herb. / Grass	3
16. India	RF	R	0.68 x 0.15	SW	M	Cwb	824	Herbaceous	1
17. India	DS	R	0.34 x 0.2	SE	S	Cwa	2183	F|Unknown	2
18. Norway	SF	S	1.35 x 0.95	E	V	Dfc	974	Herb. / Rock	1
19. India	DF	R	1.2 x 0.12	S	P	Am	2848	F|Needle	3
20. Peru	EF	U	0.6 x 1	NE	S-V	Dsb	506	Herbaceous	3
21. Kygysztan	CCS-EF	RS	5 x 0.6	NE	S	ET	394	Herbaceous	2
22. Italy	DF	R	0.35 x 0.07	SE	S	Dfc	886	Agriculture.	2
23. Indonesia	DF	E	6 x 0.3	NE	V	Af	2775	F|Broadleaf	2
24. Brazil	DS	R	0.06 x 0.03	SE	S	Am	1678	Urban	1
25. Canada	DF	R	0.85 x 0.32	SE	S-V	Cfb	1712	F|Needle	3
26. USA	RF	U	0.09 x 0.06	W	P	Csb	1560	Shrub	1
27. Burundi	DS, DF	R	0.4 xx 0.3	mixed	M	Aw	1519	F|Unknown	2
28. Australia	DS, DF	R	0.8 x 0.04	S	S	Cfa	2031	Agriculture	2-3
29. Indonesia	SLS	E	2.1 x 1.1	W	P	Af	1534	Urban	2
30. Turkey	RS	U	0.5 x 0.3	SE	S-V	Cfb	626	Agriculture	3

Acronyms: Type (according to Hungr et al., 2014): RF. Rock fall, RS. Rock rotational slide, DS. Gravel/sand/debris slide, CCS. Clay/silt compound slide, SLS. Sand/silt liquefaction spread, RA. Rock avalanche, DF. Debris flow, MF. Mud flow, DA. Debris avalanche, EF. Earthflow, PF. Peat flow, SF. Slush flow. **Trigger (T)**: R – rainfall, E-earthquake, S – snowmelt, U - unknown. **Geology:** S - mainly sedimentary terrane, P - plutonic terrane, M – metamorphic, S-V - mixed sedimentary-volcanic terrane, V - mainly volcanic terrane. **K.G. Climate zone:** A (Tropical) + f (Rainforest), m (Monsoon) w (Savanna, Dry winter) s (Savanna, Dry summer); C (Temperate) + w (Dry winter), f (No dry season), s (Dry summer) ||a (Hot summer), b (Warm summer), c (Cold summer); D (Continental) + w (Dry winter), f (No dry season) s (Dry summer) ||a (Hot summer) b (Warm summer) c (Cold summer) d (Very cold winter); E (Polar) + T (Tundra), F (Eternal frost (ice cap)). **Land Cover: Gl (Copernicus Global Land Cover):** F - Forest | dec. – deciduous, broadleaf. **Land cover: EU (CORINE Land Cover):** Grass - Natural grassland, Moor - Moors and heathland, Rock - Bare rocks, Peat - Peat bogs.

As shown in Table 1, of the 30 events included in this report, the most common types were debris slides and debris flows, with eight in each category. Half of the events had rapid flow type failure mechanisms. The next most frequent failure mechanism was sliding, including 10 events. Nearly all the landslides were rainfall triggered (25/30), with the remaining being earthquake triggered or unknown.

In terms of size, 10 were less than 0.1 km^3, and 12 were between 0.1 and 1 km^3. The remaining eight events were between 1-5 km^3. 12 different climate classes were included, with seven case studies in tropical climates (A), 15 in temperate climates (C), six in continental climates, and two in polar climates (E). For geology, 13 cases were located in sedimentary bedrock, five in volcanic, and four cases each had metamorphic, plutonic or mixed sedimentary-volcanic terrane bedrock. The mean annual rainfalls varied from 394 mm/yr in Kyrgyzstan (#21), to 4222 mm/yr in New Zealand (#14). The land cover types using the Copernicus Global Land Cover map included 15 cases in forest, 12 in herbaceous, cropland or shrubs, two in urban areas, and one in snow/ice.

4 Methods

The input data, pre-processing, visualisation and interpretation methods are described in the following section. The Sentinel-1 SAR change images and time-series were made for each case study using Google Earth Engine, (GEE), (Gorelick et al., 2017). Date ranges for the pre- and post-event image collections, and coordinates of the approximate event location were used as filter conditions, to produce pre- and post-event image stacks of both Sentinel-1 GRD product (Ground Range Detected) and Sentinel-2 (Level 2A) images. Time-periods of one, two, or 12 months either side of the event date were used, depending on the local image acquisition frequency, and seasonality. The date ranges and coordinate locations used for each case study (along with the change images produced) are provided in the electronic supplement. Pre- and post-event image composites were produced from the image stacks, for a 4 km^2 area about the defined point.

The Sentinel-1 images, available on GEE are available pre-processed (calibrated and ortho-corrected) in 10 m resolution. Image stacks were produced by filtering for ascending or descending obit pass, VV or VH receiver polarisation, and Interferometric Wide (IW) instrument swath mode. A terrain correction (Vollrath et al., 2020) was applied to each image in the stack using either a volumetric or a surface model, depending on the land cover type. For most cases, the 30 m resolution Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) available within GEE was used. However, local DEMs were needed for the Icelandic case (link: https://gee-community-catalog.org/projects/iceland_dem/, accessed: 11 December 2022) and Norwegian cases (https://hoydedata.no/LaserInnsyn2/, accessed: 11 December 2022). The Sentinel-1 composites were created by taking the median of the terrain-corrected image collections, and change images (post-minus pre-event image composite) produced from these.

For Sentinel-2, a Normalised Difference Vegetation Index (NDVI) band was added to each image in the pre- and post-event stacks, then a greenest-pixel composite was created (maximum NDVI), using the quality-mosaic tool. An NDVI change image (dNDVI) was then produced by subtracting the pre- from the post-event composite. These are shown in Figure 4. The dNDVI images were used for verification purposes and digitising outlines of the landslides.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

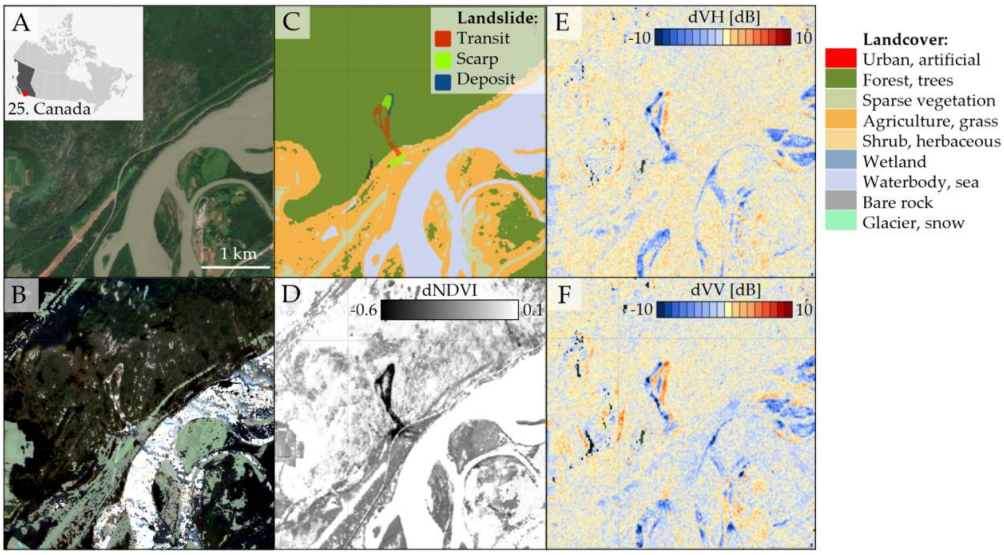


Figure 4 Data preparation for statistical analyses, showing case study 25. Ruby Creek landslide, British Columbia, Canada (14 November 2021). Input images including: A) Sentinel-2 Pre-event least cloudy image. B) Sentinel-2 Post-event greenest-pixel composite. C) Pre-event land cover produced using machine learning classifier, with pre-event Sentinel-1, -2 and DEM as input. D) Change in Normalised Difference Vegetation Index (dNDVI). E) Sentinel-1 Change in backscatter intensity (ascending VH). F) Sentinel-1 Change in backscatter intensity (ascending VV).

From the Sentinel-1 change images, it was determined firstly, whether the investigated landslides were visible. The case studies were assigned a set-number, following the approach by (Mondini et al., 2019): Set 1: not visible, set 2: partially visibly, recognisable only knowing a priori the location, and set 3: clearly visible. If visible (set 2 or 3), time-series plots were produced. This involved selecting a point within the landslide, and point adjacent to the landslide with similar but undisturbed vegetation. From these points, 30 m² square polygons were generated, and five years of terrain corrected S1 time-series data, with median backscatter intensity [dB] for ascending-VV and –VH, and descending-VV and -VH, were exported.

Next, in order to investigate the factors that influence the expression of landslides in the change images, plots of LIA, land cover class and backscatter intensity values were made. This involved manually mapping the landslide polygons for each case study, exporting the Shapefiles along with Geotiff raster images for each case study, and finally extracting and plotting the pixel values from within the polygons. The LIA was already calculated as part of the terrain correction algorithm, so the original set of outputs (VV, VH and angle) was modified to also include the LIA. Where possible, landslides were mapped according to morphometric feature class (scarp, transit or deposit zone). Additional ground-based or drone images of the landslides (internet search, journal articles) showing the shape of the landslide were used to identify these separate features, although there remained uncertainty in the exact boundaries. If no additional information was available to identify the morphometric features, then the entire landslide body was mapped in the transit zone class as the default, as this includes both erosion and deposition.

For the land cover class, existing land cover maps (Copernicus Global Land Cover, CORINE, Dynamic World) were first examined. These were deemed to be not of high enough accuracy at the pixel level

for statistical analyses. Therefore, we produced our own locally-trained land cover maps. For each case study a machine-learning based land cover classification was performed using the ee.smile.CART algorithm (Classification and Regression Tree) in GEE (Breiman et al., 2017). These were trained using a Sentinel-2 image with minimal cloud cover from before the event, along with the DEM, and pre-event Sentinel-1 composite images (VV and VH polarisation). The classifier was trained by manually selecting points within the following classes; Urban, artificial; Forest, trees, Scrub, herbaceous; Pasture, grass; Sparsely vegetated; Water body, sea; Wetland; Bare rocks; and Glacier, snow. A minimum of 10 points were selected in each of the classes present. Then the images were sampled at the selected points, the classifier was trained. The classification was thereafter performed over the entire area of interest and results viewed. If the classification was not satisfactory (judged by visual comparison of the results with the higher resolution base satellite images) then additional training points were added, or input data modified. The process was repeated until a satisfactory land cover map was achieved.

Finally, for each case study the following data was exported: 1) landslide polygons, (2) geotiff raster images including the pre-processed Sentinel-1 bands from ascending and descending mode where available (preVV, postVV, dVV, preVH, postVH, dVH, LIA) along a band of the land cover class. The pixel values within each of the landslide polygons were extracted to produce a dataset consisting of approximately 300,000 pixels. Using Seaborne (version 0.12.2) violin plots and 2-dimensional histogram plots were produced to display the distribution of the data. The default settings were used, except for the histograms where the argument stat="percent" was used instead of "count" (Waksom, 2022a, 2022b).

5 Results:

Over 1000 landslides were digitized from the 30 case studies. During this process we identified predictable patterns in the expression of landslides in SAR backscatter change images related to the different morphometric features of the landslides (scarp, transit zone and deposit zone) and land cover type, and performed statistical analyses of the pixel values from the landslides. Here, we present an overview of the trends that were identified, and present evidence from individual cases and the statistical analyses that show how different factors, including land cover and landslide orientation relative to the sensor, control this expression.

5.1 Trends identified in the expression of landslides in SAR data

Some examples of the trends identified are shown with contextual photos in Figure 5, SAR difference images in Figure 6 with five-year time-series plots in Figure 7. These trends are summarised as follows:

(A) Scarp: The expression of scarps varied depending on the look direction, with scarps angled away from the sensor look direction producing an abrupt decrease in backscatter intensity in both dVV and dVH images, while scarps facing towards the sensor produced slightly to moderately increased backscatter intensity. The time-series plots in Figure 7 show that the decrease was most clear in VV polarised data, with a magnitude of 7 to 12 [dB]. In some cases (seen quite clearly in case 20 in Peru), an edge of increased back scatter intensity was also observed slightly behind the scarp on the far side of the landslide from the sensor. Rock fall scarps were not clearly distinguishable in the cases we examined.

(B) Transit zone in herbaceous vegetation: the most easily distinguishable landslides were those that occurred in herbaceous vegetation (e.g. tundra, peat, grass, cultivated land). These

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

produced strong increases in backscatter intensity shown in time-series (Figure 7), most clearly seen in VV polarised data, in the order of 7 to 10 dB in the time-series data.

(C) Transit zone in forested area: a more complex, but quite distinct, pattern of backscatter intensity change was observed in most of the landslides that occurred in forested areas, seen most clearly in VH polarisation. As with the scarps, the pattern depends on the look direction of the sensor. For the cases shown in Figure 6 (including cases 5, 9, 10, 12, 23, and 25) it can be seen that moving away from the sensor - there is a sequence with first decreased backscatter intensity along the edge of the landslide closest to the sensor, and increased backscatter intensity on the far edge of the landslide. For wider landslides there may be a zone with moderately increased or decreased backscatter intensity in the centre of the transit zone. The decreases shown in the time-series plots are around 4 to 8 db.

(D) Deposits: in most of the cases we observed, deposits were observable by a moderately to strongly increased backscatter intensity (in VV polarisation) as seen in cases 2 and 14. Although in some specific cases the deposits were observable by areas of decreased backscatter intensity, as seen in cases 10 and 25. From the contextual photos in Figure 5, it appears that the deposits with decreased backscatter intensity relate to cases where fine sediments settled from still water caused by drainage blockage. Whereas, those showing increased backscatter intensity appear to relate to deposits consisting of coarser materials inferred to be deposited more rapidly from the turbulent landslide flow.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

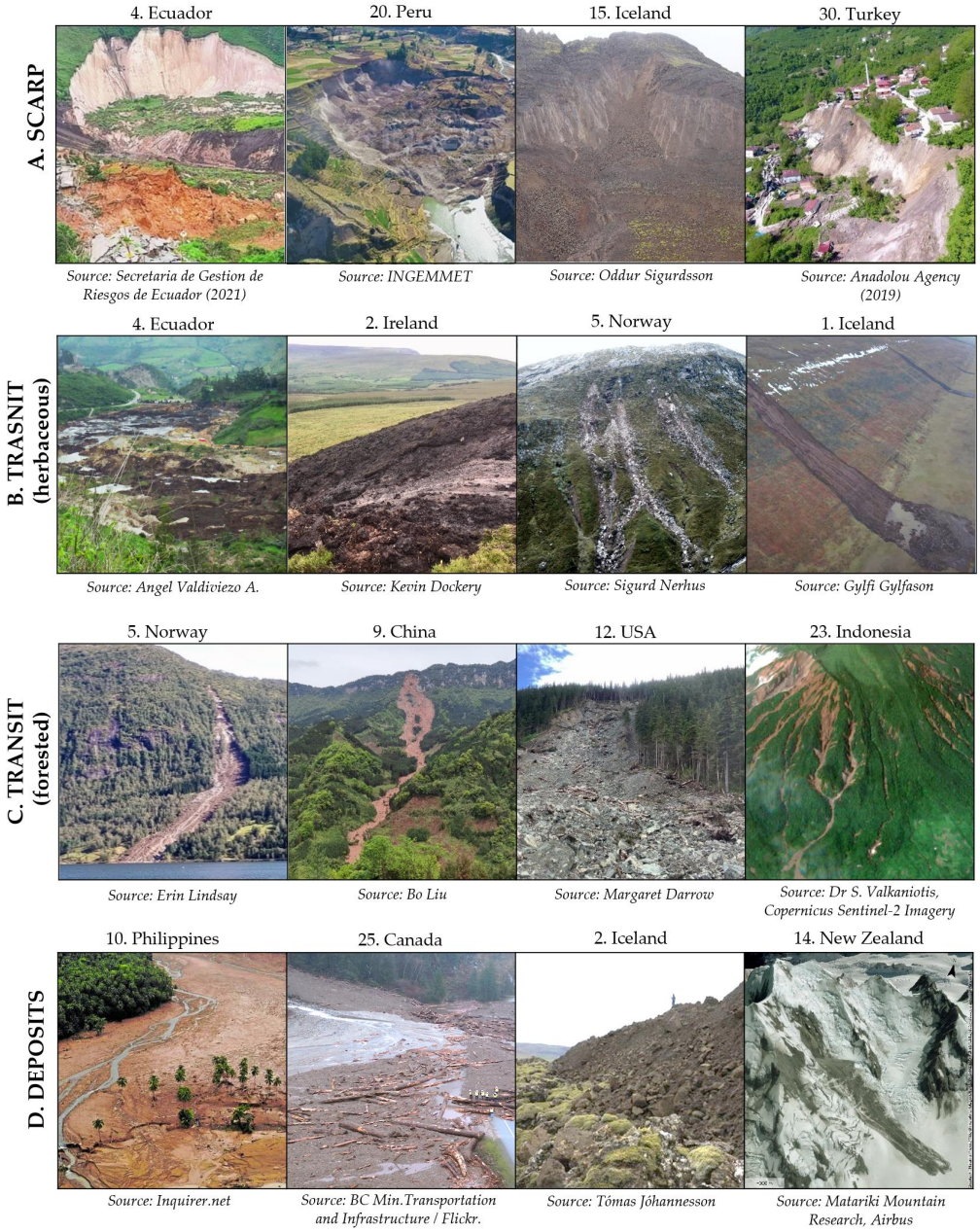


Figure 5 Context images corresponding to the landslides shown in Figure 7 and Figure 8, showing a variety of landslide types and environmental settings.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

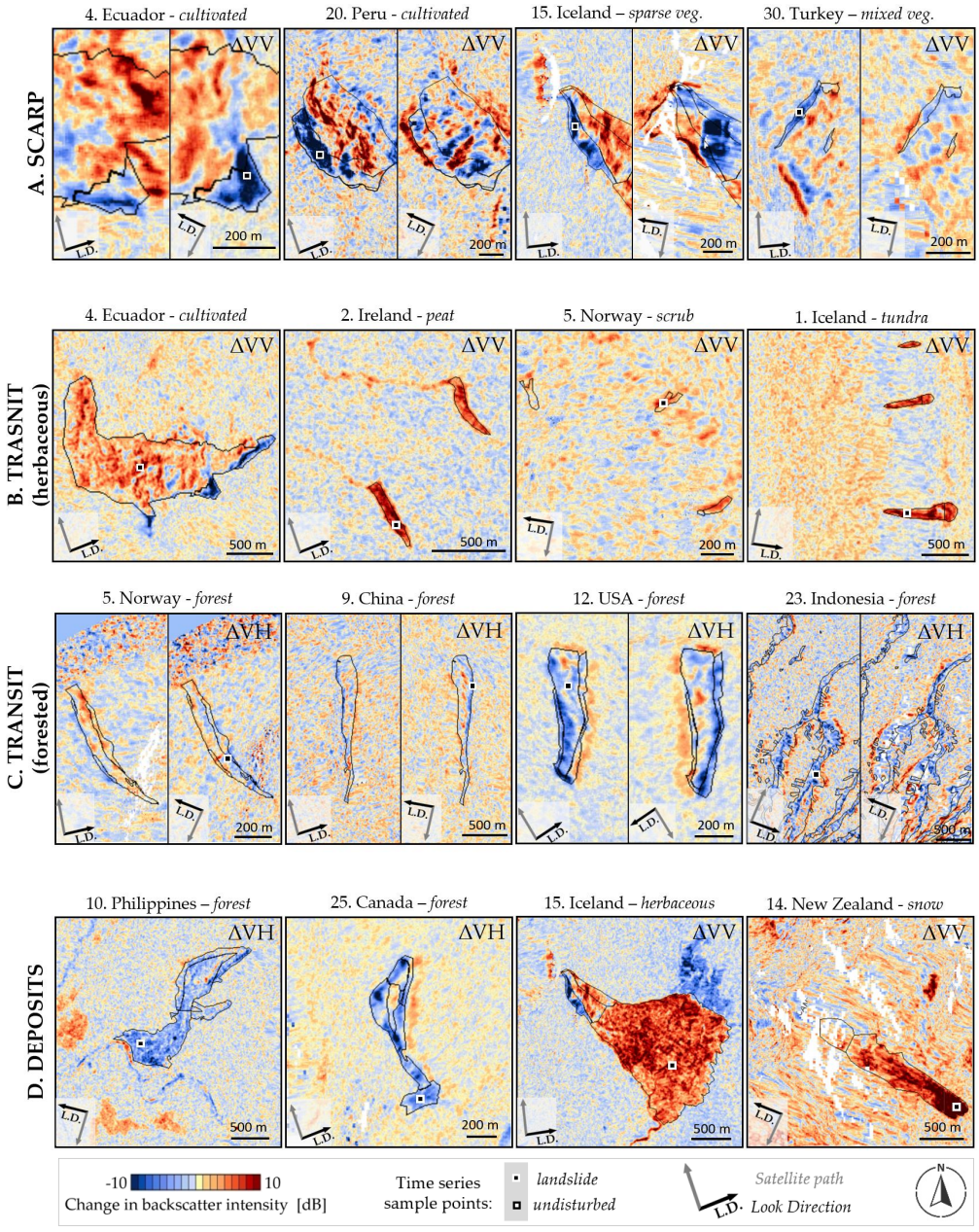


Figure 6 Collection of backscatter intensity change images showing a selection of landslides, demonstrating the trends identified during systematic mapping of landslides in all 30 case studies shown in Figure 3. The change images are produced from pre- and post-event median composite (from stacks consisting of 1 to 12 months of images, see appendix for details) of terrain-corrected Sentinel-1 SAR images. Center coordinates and dates of the landslides are provided in Figure 3.

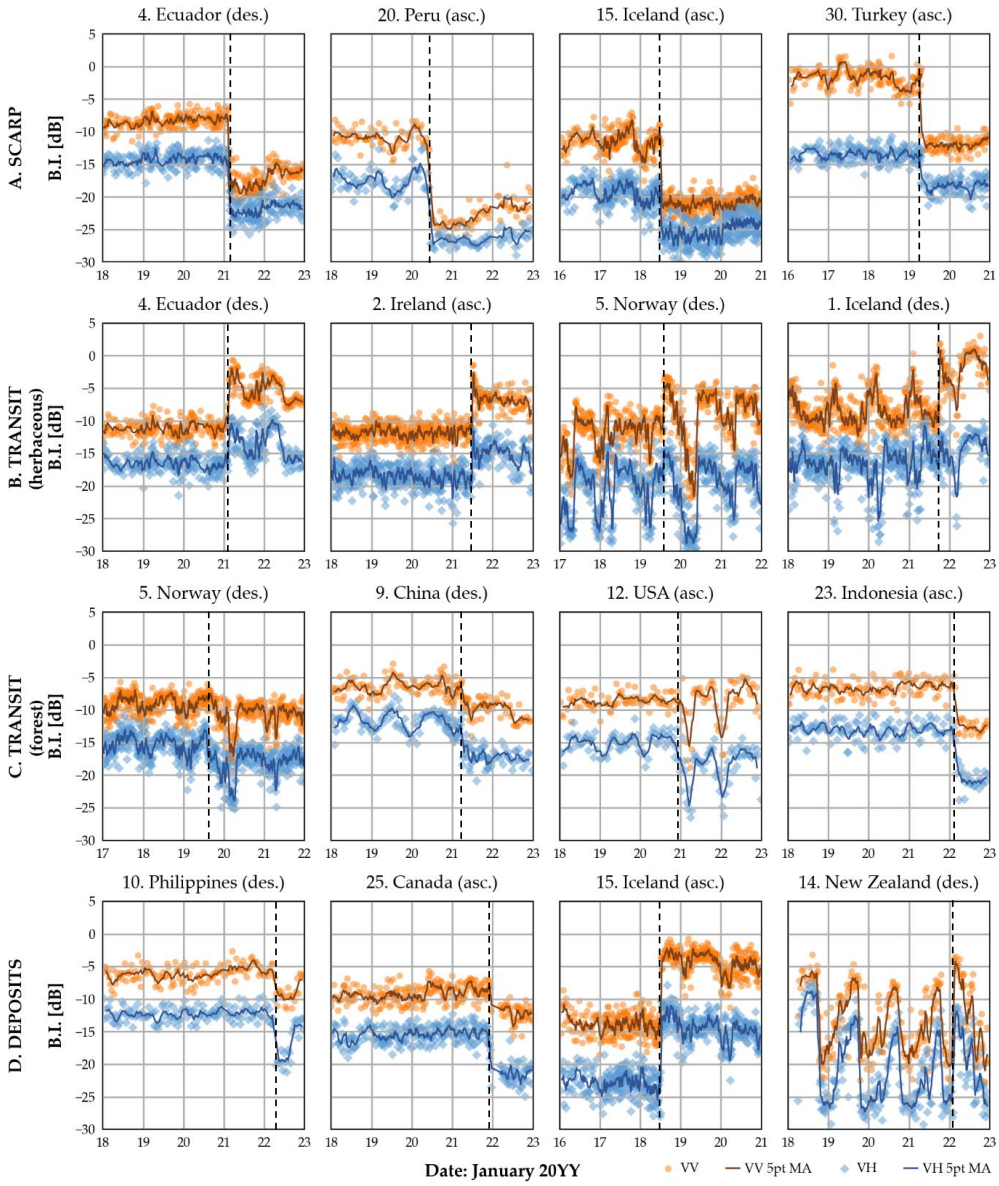


Figure 7 Five-year time-series plots of mean backscatter intensity (B.I.) in VV and VH polarisation, sampled from a 30 x 30 m patch within the landslide body. The location of the sample patches are shown in Figure 7. Landslides can be observed by an abrupt change in the seasonal cycle of backscatter intensity. The black dashed lines show when the landslide occurred. Event dates for each case are provided in Figure 3.

5.2 Factors controlling the expression of landslides in SAR data

The factors identified in the theory section as being relevant to the visibility and expression of landslides in SAR backscatter intensity change images included: terrain elevation and geometric distortions, local incidence angle (LIA), wavelength, land cover types, seasonal variations and water content. With the exception of wavelength, which was fixed in this study (C-band), the role of each of the factors is

considered based on the results of systematic mapping of the landslides along with pixel-based statistical analyses.

5.2.1 Large scale terrain features and geometric distortions

Geometric distortions were considered to have been a major limitation to landslide visibility in six of the cases (including 7. Vanuatu, 8. Brazil, 14. New Zealand, 26. USA, 27. Burundi, and 28. Australia). This was particularly problematic if only one orbit pass was available. In the Burundi case, although two orbit passes were available, a significant area was affected by foreshortening in both the ascending and descending images (see Figure 8), which appears as stretched pixels when corrected and distorts the landslide signatures significantly. This is problematic for areas with steep narrow valleys. Shadow zones were not as problematic as expected for large scale terrain features (e.g. the 4000 m high mountains in Case 14 from New Zealand, see Figure 7. Most of the shadow zones were related to the presence of medium scale topographic features, such as cliffs.

5.2.2 Local incidence angle

The local incidence angle can have a strong impact on the expression of landslides in SAR data, as illustrated in Figure 6. In particular, this is observed for landslides with significant scarps and those that occur in forested areas where the appearance of the landslide varies significantly depending on the sensor direction. Some of the landslides triggered by the earthquake in Hokkaido, Japan, shown in Figure 9, show how strongly the expression of landslides can vary depending on the orientation of the landslide surface.

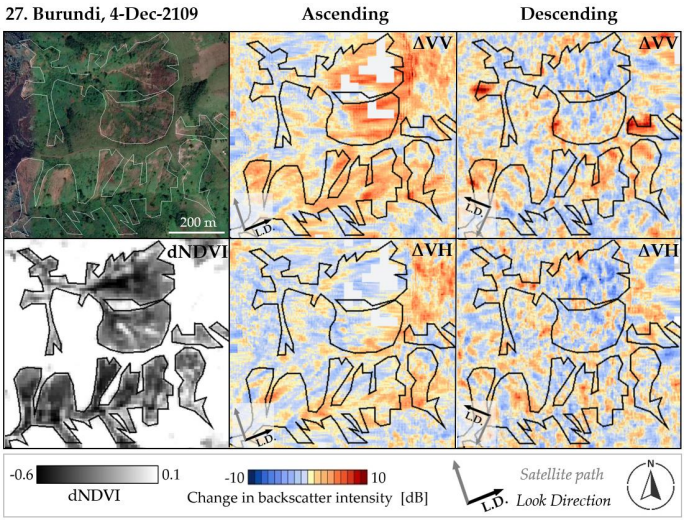


Figure 8 Landslide visibility was reduced in the Burundi case due to geometric distortions, as well as unfavorable mixed vegetation conditions. Black areas show shadow distortion, while the stretched pixels are presumably affected by foreshortening. Outlines were drawn based on the Sentinel-2 dNDVI image.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

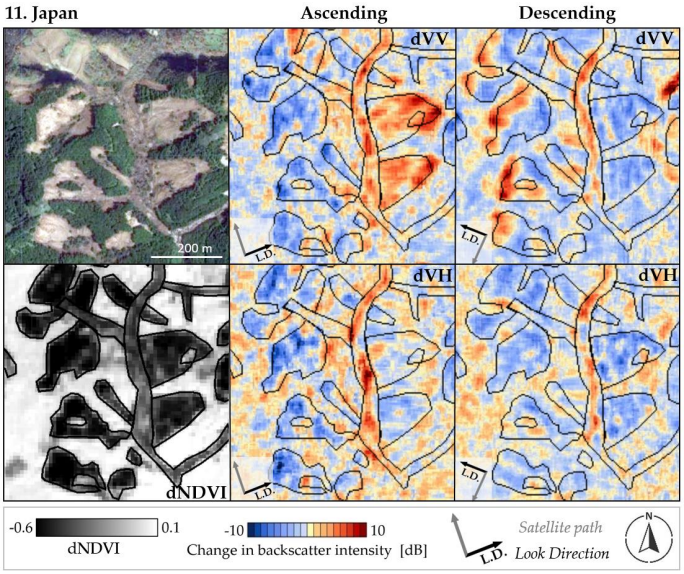


Figure 9 Landslides in forested area from Hokkaido, Japan (case 11) showing strong variation in expression in backscatter intensity change images depending on the orientation of the landslide surface. Outlines were drawn based on the Sentinel-2 dNDVI image.

The landslide scarps to the west of the central valley are mainly sloping down towards the east, in a similar direction to the look angle of the sensor on the ascending path (high LIA). In both the dVV and dVH ascending images, these western landslides show decreased backscatter intensity at the location of the back scarps. While the landslides to the east of the central valley facing approximately normal to the sensor (low LIA) show a strong increase in the dVV image at the location of the back scarp and a moderate increase in the mid slope. By contrast, in ascending dVH these eastern landslides show moderately decreased backscatter intensity in both the mid slope and towards the back scarps. A similar, but reversed pattern can be observed in the descending images. The local incidence angle appeared to be less relevant for landslides that occurred in herbaceous vegetation and for deposit zones. In Figure 9, the deposit zone in the centre of the valley shows an increase in all of the change images.

The 2D-histograms in Figure 10 show the spread of pixel values from within the mapped landslides depending on the LIA, and separated according to land cover type, landslide morphometric feature and polarisation. Although, there were relatively few data points in the scarp category, some strongly increased backscatter intensity values are observed at low LIA within forested examples, both for VV and VH polarisations. This fits with the observations from the Japanese case study shown in Figure 9. In the non-vegetated transit zone, and herbaceous deposits plots, there appears to be a weak trend of increasing dVV and dVH values with increasing LIA for the middle LIA values. However, there are a lack of points sampled at low and high LIAs in these categories. For the remaining plots there does not appear to be any clear relationships between the change in backscatter intensity values and the LIA.

5.2.3 Changes between ground cover types

The change in ground cover and associated scattering mechanisms depends both on the pre-event land cover, as well as the post-event texture of the ground produced by the landslide. From the 2D-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

histograms in Figure 10, it can clearly be seen that deposits tend to produce increased back scatter intensity in all land cover classes, although the difference is greater in herbaceous and non-vegetated classes. While from the transit plots in forested areas, that landslides generally produce a decrease in backscatter intensity.

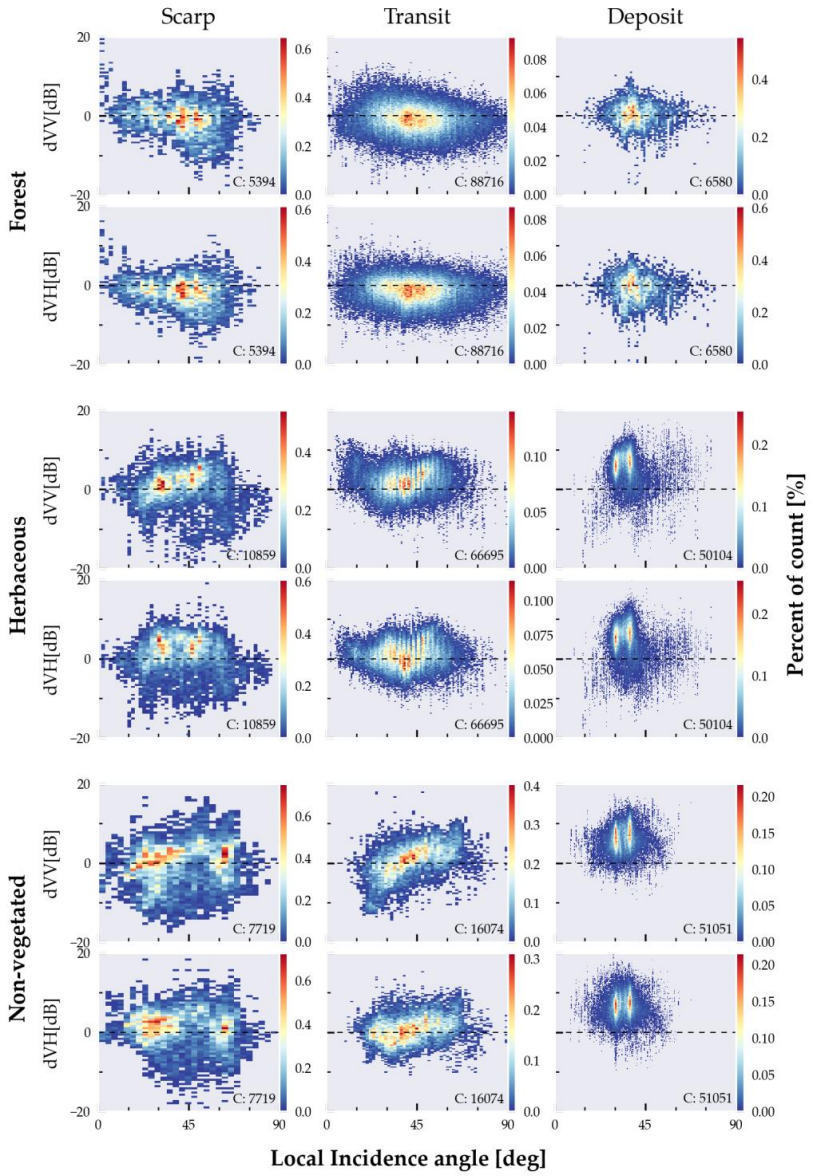


Figure 10 2D Histograms of data points sampled from within the landslides polygons (columns). Separated according to land cover class, and polarisation. Shows the distribution of the pixel values LIA vs change in backscatter intensity. In the herbaceous category, land cover classes including sparse vegetation, pasture, grass, wetland, scrub and herbaceous were combined. For non-vegetated urban, artificial water bodies, bare rocks, and

glacier and snow were combined. C: indicates the pixel count. Note that the LIA is calculated from the pre-event DEM.

Figure 11 shows increasing backscatter intensity of deposits from four cases. Post-event deposits show variation depending on the material, variation between the medians of 4 dB. The backscatter intensity increases with increasing material size (10. mainly fines, 5. mixed fine and coarse, 9. mixed with vegetation debris, to 11. rock and boulders). Note that the deposits in cases 5, 10, and 11, occurred on mainly flat or shallowly sloping areas, while those in 14 are from a slope with a lower incidence angle, therefore the values may be slightly higher due to this.

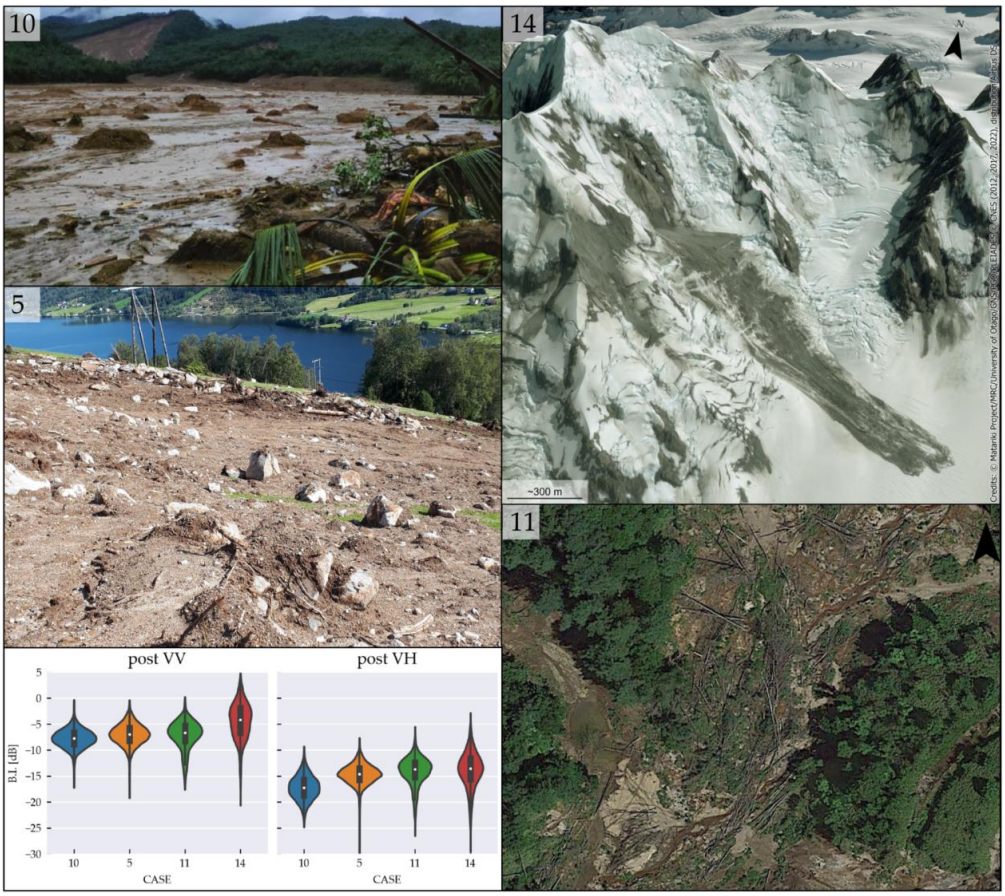


Figure 11 Different deposit material types, and the distribution of the pixel values from the four different case studies shown. These include: Smooth, flat deposits from mudflow near Baybay City, Philippines (10) (source: Philippines Coast Guard, AP). A mixture of soil and rocks in debris flow deposit in Vassenden, Norway (5). Rock avalanche deposits at Mt Tasman, New Zealand (14) (© Matariki Project/MRC/University of Otago/GNS/PGO/PLEIADES © CNES (2022), distribution Airbus DS.). (11) Forest debris from landslides in Hokkaido, Japan (© Maxar (2018)). Bottom left: Violin plots of post-event backscatter intensity distribution from the different deposit material types. The upper and lower quartiles and the median are indicated by the black box plot with a central white dot.

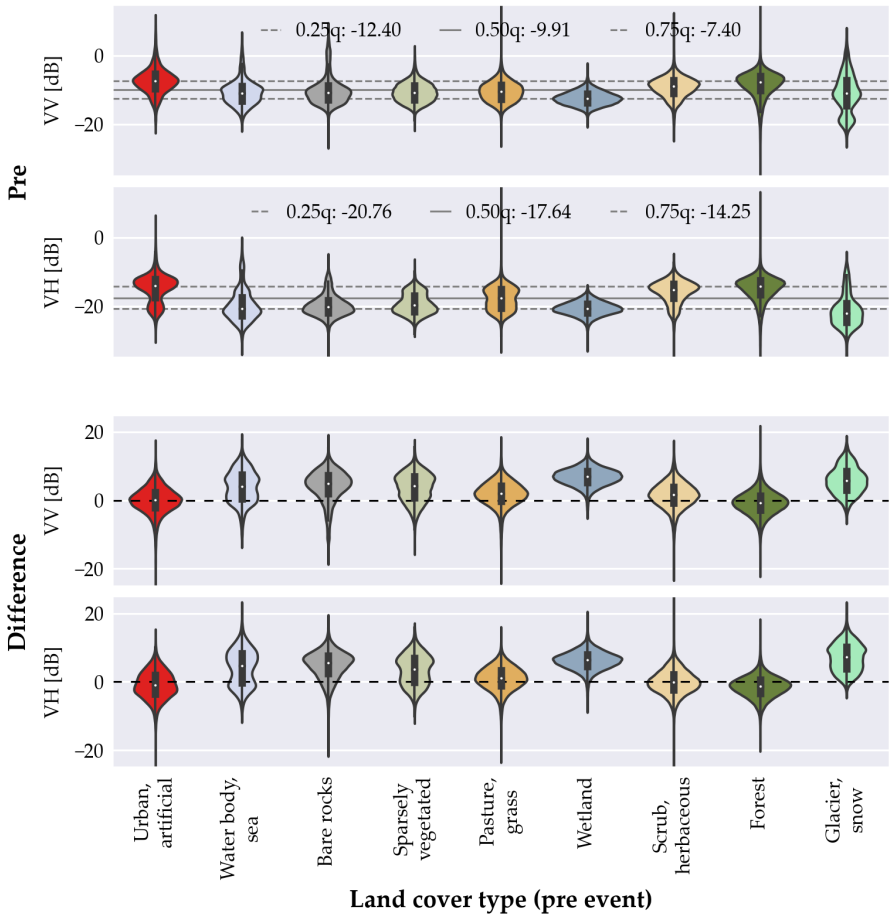


Figure 12 Distribution of sampled pixel values within mapped landslides from the pre-event and difference images, shown as violin plots with box-plots and kernel distribution. The upper and lower quartiles and the median are indicated by the black box plot with a central white dot. These are separated according to pre-event land cover types and polarisation.

Figure 12 shows the distribution of pixel values within the mapped landslide polygons according to land cover class, for pre-event and the change in backscatter intensity. The pre-event backscatter intensity values have a combined median of -9.9 and -17.6 dB for VV and VH respectively. Urban and forest land cover classes have slightly higher initial values relative to the median in VV and VH polarisations, while wetland and water bodies have slightly lower values in both. In VH non-vegetated land cover types including bare rocks and snow also show values lower than the median.

The differences in backscatter intensity are mostly positive, except for pixels within the forest class which show a mean decrease of -2 dB in both VV and VH polarisation. The strongest increases are observed for landslides that occurred in wetland and snow with average difference values of ~4 dB. Slightly negative changes are also seen for scrub and urban land cover classes in VH polarisation.

5.2.4 Seasonal variations and water content

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

The time-series plots in Figure 7 show variation in the seasonal cycles of backscatter intensity, in which reflect the local climate and vegetation conditions. In general, for vegetated areas the backscatter intensity is higher in summer, and lower in winter. While in tropical areas, the backscatter intensity is relatively constant through the year, and in sub-polar or alpine regions winter snowfalls produce strong decreases (in the order of 10 to 20 dB) in intensity. In most cases, the change due to landslides appears unambiguously and a clear difference in the seasonal cycle is observable. However, in cases with seasonal snow cover (e.g. 5B. Norway, 1B. Iceland and 14D. New Zealand) the changes due to landslides are of a lower magnitude than the change due to snow cover change, and therefore less clear.

Most of the inter-annual variability in the pre- and post-event seasonal cycles is likely caused by fluctuations in moisture content. Although we don't not have access to soil moisture measurements, some trends can be inferred. The effect of moisture content is seen most clearly in case 14 from a high alpine environment in New Zealand (Mt Tasman, approx. 2,500 m a.s.l.). Here, in spring at the onset of snow-melt, the backscatter intensity drops abruptly, with a magnitude of approximately 15 dB. In several cases, in the transit and deposit zones, it is seen that the following the initial change in backscatter intensity, there is a period over several weeks where the backscatter intensity decreases slightly. This is seen in cases 4, 2, and 5 in the herbaceous transit zone, and in 23 in the forested transit zone. We did not have access to soil moisture data, but it is possible that this pattern is due to decreasing moisture content after the rainfall triggered landslides. In contrast, the values increase in the deposit zone in case 15 in Iceland.

6. Discussion

The results showed that landslides produce predictable patterns in SAR backscatter change images related to morphology, land cover, and sensor and ground surface orientation. The trends identified have not previously been documented in relation to landslides. Here, we consider the physical basis for the identified trends in relation to the change in scattering mechanism, compare these with similar trends described in literature from other phenomena, and discuss limitations of the approach and possible future research directions.

6.1 Trends

In order to improve understanding of the physical-basis for the identified trends, we present the generalised conceptual model shown in Figure 13, relating the change in scattering behaviors to the observed change in backscatter intensity. This is described as follows, referring to the numbers illustrated:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2023 doi:10.20944/preprints202302.0390.v1

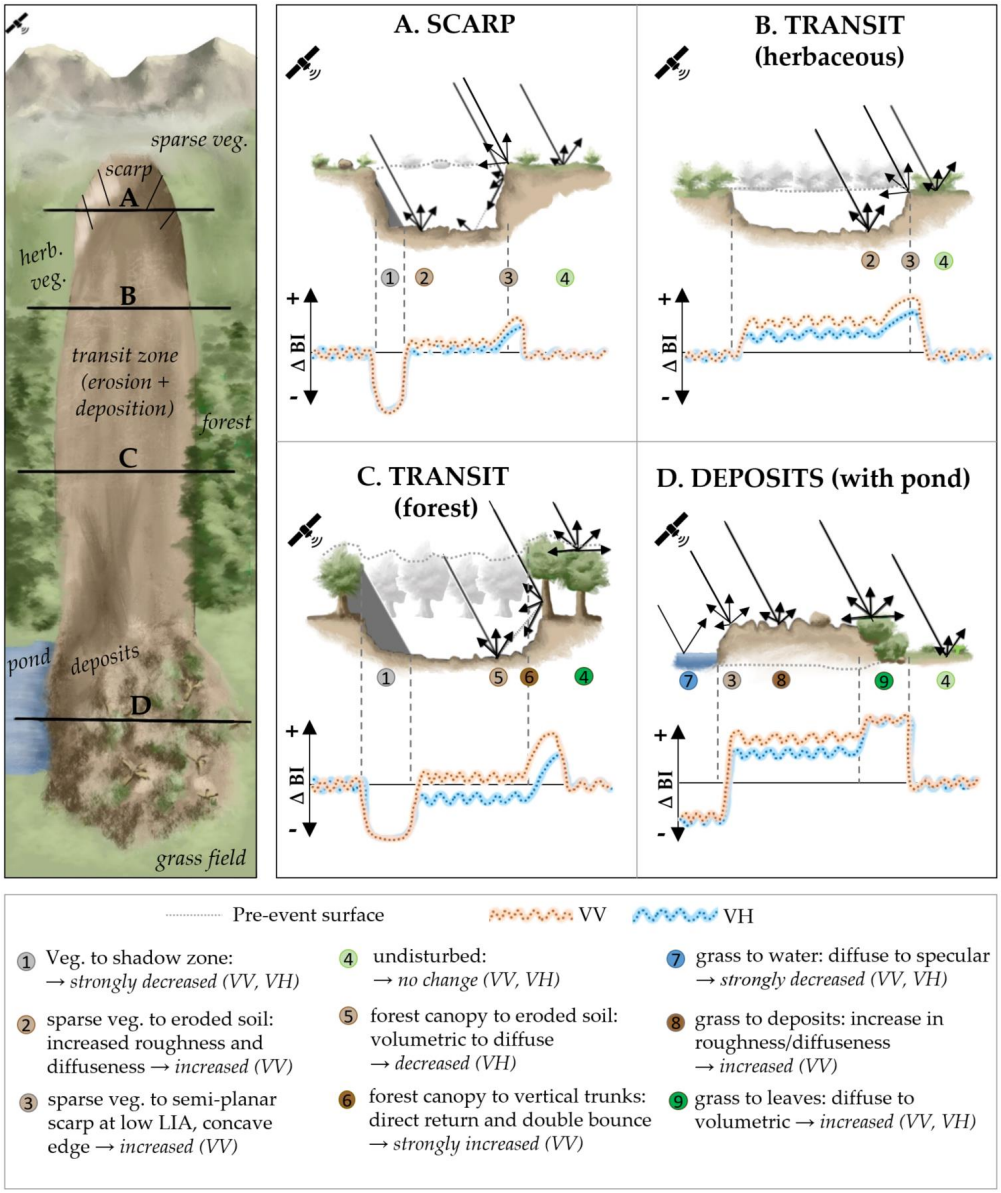


Figure 13 Conceptual model of a landslide and the relative changes in backscatter intensity (B.I.), depending on landslide morphology, and the change in ground-cover types and their associated scattering mechanisms.

A. SCARP

The sudden change in topography, usually expressed by a steep surface, results in a strongly decreased intensity for scarps facing away from the sensor (1). This is attributed to radar shadow, which occurs when the slope angle is steeper than the radar incident angle. The base of the scarp beyond the shadow zone, may show increased intensity due to an increase in surface roughness (2) as in the conceptual model. However, this will vary depending on the specific pre-event land cover and post-event surface

roughness and orientation. For scarps facing towards the sensor (3), if the LIA is decreased more energy will be reflected back towards the sensor, resulting in increased intensity. In addition, the concave top edge of the scarp will also give strongly increased reflectivity relative to a flat pre-event surface, producing strongly increased values on the outer edge of the scarp. These patterns are seen most clearly in VV, due to the greater sensitivity to changes in surface roughness. Similar changes have been documented in relation to changes in topography of a volcanic crater following an explosive eruption (Arnold et al., 2018).

B. TRANSIT (herbaceous)

The trend of predominantly increased backscatter intensity in these areas, is due to an increase in surface roughness of the landslide surface compared to a weathered pre-event surface (2). This results in increased diffusivity, and stronger reflection back towards the sensor. Minor scarps and concave features within transit zones of larger landslide are identifiable based on the same principles as described for the main scarp. There are numerous examples of agricultural studies relating increasing roughness of non-vegetated surfaces to increasing backscatter intensity (Baghdadi et al., 2018; Beaudoin et al., 1990)(Baghdadi et al., 2018; Beaudoin et al., 1990; Tempfli et al., 2009).

C. TRANSIT (forested)

The pattern illustrated for landslide transit zones in forest is very similar to the pattern described by (Hoekman et al., 2020) from drainage canals that are constructed within rainforests prior to deforestation. The changes relate to (1) radar shadow on the near edge, (5) change from forest to bare soil which produces a decrease in VH due to reduction of volumetric scattering, and possibly slight increase in VV, depending on the roughness of the surface, and (6) increased backscatter intensity on the far edge, due to a change from forest to a new near-vertical surface of the scarp and tree-trunks, which produce direct and double bounce scattering, increasing the energy returned to the sensor.

D. DEPOSITS (with ponding)

The example illustrated shows firstly a strong decrease caused by a new pond, related to a change from grass to water (7) which has a very low intensity due to specular reflection. From the case studies, we observed that new lakes or ponds caused by landslide dams, showed stronger signals than the landslides themselves and are very easy to detect in the change imagery. Although, they can be difficult to distinguish from scarps without contextual information, as can be seen in the Ecuador case, where there first seemed to be three large scarps, however one of them was actually a lake. Detecting newly formed lakes is important in disaster response, as these may occur in unpopulated areas, however can pose serious threats to people downstream if the landslide-dam bursts suddenly (Dellow et al., 2017). The signal is the same as can be used to detect flooding (Shen et al., 2019). Conversely, changes from ground to water result in strongly increased backscatter intensity which was observable in cases 10, 15 and 20 from the Philippines, Iceland and Peru.

The landslide deposits themselves were most frequently observed in change images by increased backscatter intensity (8) as is illustrated in the conceptual mode. This is due to increased surface roughness, and possibly also the presence of concave structures for landslides with a large volume, both of which produce increased diffusivity. However, as shown in Figure 6, 7 and 10, cases that

appeared to have smooth deposits formed by fine materials (mud or silt) produce lower backscatter intensity post-event, and this may result in decreased backscatter intensity in VH change images. Such deposits can indicate that the sediments were deposited by still or slow-moving water, related to blocked drainage.

6.2 Controlling factors

Factors that were expected to control the visibility and expression of landslides in backscatter intensity change images included; geometric distortions, LIA, ground cover, seasonal variations and water content. Wavelength is also considered relevant, however was not investigated in this study. Overall, we saw significantly stronger trends in the time-series data from individual cases, than from the means of the pixels within the mapped polygons. This is due to the selection of points within the change images for extracting the time-series data that showed strong changes. While, the sample pixel values within the mapped polygons show more variation due to irregular surface texture and topography and speckle noise.

The overall detection rate was 87% (n. test cases = 26/30) in this study. This is similar to that reported by Mondini et. al (2019) of 83% (n. test cases = 27/32), yet differs significantly from the rate reported by Lindsay et al., (2022) of less than 10% (n. test cases = 9/120). The difference between these reported rates lies in the method of case study selection. In the two former, the examples selected came mostly from news reports and are biased towards larger, more catastrophic events, whereas in Lindsay et. al., (2022), the reported rate was in comparison to a set of previously mapped landslides from a single case study and included mainly smaller landslides.

The influence of the LIA was clearly observable when looking at the differences in the landslide expression between individual cases. This was particularly important in the case where landslides produced an abrupt change in surface height, whether due to removal of materials in the scarp area, or removal of trees in the transit or deposition zones. The pattern of shadows from steep or vertical surfaces facing away from the sensor, and bright edges facing towards the sensor is highly dependent on the sensor look direction (ascending or descending orbit path). The trends were not clear in the 2D histogram plots in this study, however the LIA has been shown to affect the backscatter intensity to different degrees, in experimental data from other fields where the sampled areas were more homogenous. Using post-event DEMs to estimate the LIA, would probably give clearer trends especially for the scarp. However, this data is not readily available in most cases (Dabiri et al., 2020).

The change in the ground cover type and associated scattering mechanisms was very important in determining the expression of the landslides. In the change images, time-series data, and statistical analyses, it was clear that landslides in forest tend to produce an overall decrease in backscatter intensity, while those in herbaceous and non-vegetated areas tend to produce an increase. In addition, changes to water bodies were very clear in the change images. Differences in the post-event backscatter intensity sampled within deposit zones, were also observed depending on the material size (fine, or coarse).

Changes due to landslides in time-series data were clearer for cases with tropical and temperate climates where the vegetation showed limited seasonal variation. Cases with changes in snow cover showed abrupt changes in backscatter intensity values that were not related to landslides. It was observable that the timing of landslides in relation to the underlying seasonal cycle, either enhanced

(i.e. 14. New Zealand), or reduced (i.e. 5. Norway) the change in backscatter intensity. These variations in seasonal conditions are important to be aware of if designing an operational landslide detection system based on time-series data. The effect of changes in water content were not clear.

6.3 Limitations and future research directions

The cases that were undetectable in this study (Table 1, set 1) included two rock fall events, one slush flow, and one case with mixed vegetation type. In the rock fall events, it was determined that the change to the ground surface texture was not significant, or in the case where deposition occurred on a road, not permanent. The slush flow was not detectable due to changes to the surrounding area due to snow-melt. While for the mixed vegetation case where deposition occurred without removing the pre-existing trees, the change produced by the landslide did not significantly alter the backscatter response, which included both diffuse and volumetric scatterers, before and after the event. In the cases where landslides were only partially visible or detectable only with a-priori knowledge (set 2), the main limiting factors included; small or narrow landslide geometry, geometric distortions that obscured or warped the landslide signatures, snow- melt in the surrounding area, deposition in an urban area, and only single orbit pass available. Large scale topography, in particular high mountains, was not as problematic as expected.

The dataset produced in this study provides a diverse training dataset that can be used for developing generalised automatic detection models. Previous studies have shown that locally-trained deep-learning models such as U-Net can detect landslides with good accuracy from Sentinel-1 images due to their ability to differentiate random speckle noise from clusters of pixels related to changes to the ground surface (Ganerød et al., 2023). The results of this study can be used to improve the design of deep-learning models through understanding how to ensure representative training cases and relevant input data are included. A challenge for such models, will be to differentiate signals of non-landslide related vegetation loss from landslides, as is an ongoing problem for landslide detection using optical data (Prakash et al., 2021). More investigations could also be done using this dataset for improving understanding of how land cover affects the generation of post-event DEMs (Dabiri et al., 2020). Then automated estimation of the volume of material would also be possible. This would enable disaster responders to estimate the magnitude of the problem, and send appropriate resources.

Experience from testing these methods on recent events, including landslides triggered by floods and cyclones in New Zealand (10 January 2023, 27 January 2023), and the Turkey earthquakes (6 February 2023), has shown that large landslides and landslide dams can be detected from single post-event images. However. the greatest limitation for applying these methods in disaster situations is the lack of daily images freely available (revisit frequency in Europe is 1-3 days, in equatorial regions it can be up to 12 days). Commercial high resolution SAR data is currently available at a daily frequency. In order to be prepared to respond quickly, one should have agreements set up in advance to access to satellite images from as many sources as possible, including commercial providers (e.g. ALOS, ICEYE, Spacety which may provide data for research and non-commercial activities), and have code prepared to quickly process the images. Research should be conducted into combining imagery from multiple sensors with varying wavelengths, view angles and resolutions, to be prepared to use whatever information is available.

The frequency of freely available SAR images will double with the launch of the NISAR (NASA-ISRO SAR) satellites in 2024. However, it will take some time to build up a new set of landslides observations in the X- and L-band. Furthermore, it is likely that Sentinel-1 coherence data will be added to the GEE catalogue (Gorelick. N. (2022). *pers comm.*). This will facilitate global studies of how coherence data and backscatter data can be used together to improve landslide detection. Automatic detection and consequence analyses (for instance highlighting the intersection of automatically detected landslides with buildings, or infrastructure) could also be useful for improving situational awareness during disaster response.

7. Conclusions

When cloud-cover inhibits the use of optical imagery during disaster response, SAR satellite imagery could be used to provide crucial information on the location, and timing of landslides. Through systematic and quantitative analysis of over 1000 landslides from 30 diverse global case studies, we found that landslides produce predictable patterns of increased and decreased backscatter intensity. The change in intensity relates to morphology, and changes in topography, land cover, and surface roughness and their associated scattering properties. The findings were summarised in a physically, and empirically based conceptual model which can facilitate interpretation of landslides in SAR backscatter change images. These patterns have not been documented in relation to landslides previously, however have similarities to signals produced by features related to volcanic eruptions, agriculture, deforestation, canal building and floods. Using freely available Sentinel-1 images and Google Earth Engine, the radar change images can be rapidly produced for a new location, when new satellite images become available online. The trends observed in this study may improve utilisation of satellite radar images for rapid detection of landslides and other natural hazards, and facilitate development of multi-sensor automatic detection models and continuous monitoring systems. This can improve situational awareness in disaster response and potentially improve outcomes for the affected populations.

Data availability: The codes and dataset will be made available when published.

Description of author's responsibilities: Erin Lindsay performed the following roles: conceptualisation, data curation, formal analysis, investigation, methodology, software, validation, visualisation, writing – original draft, and writing – review & editing. Graziella Devoli contributed with the roles: visualisation, review & editing and supervision. Johannes Reiche performed review & editing, and methodology. Steinar Nordal performed review & editing, supervision, and funding acquisition. Regula Frauenfelder contributed with methodology, review & editing and visualization. Alexandra Jarna performed data curation, investigation and visualisation. Lars-Christian Tokle assisted with, methodology, software, validation and visualisation.

Funding: This research was funded by the Research Council of Norway, through the research project *SFI Klima 2050* [grant number 237859].

Acknowledgements: Data provided by the European Space Agency and Planet under project ID: 192991 - *Optical satellite data for landslide detection using dNDVI method*. Project supported by ESA Network of Resources Initiative. The authors gratefully acknowledge the time, materials, and efforts contributed by the following people: Jørn Emil Gaarder (Klima 2050, NTNU) for illustrations, Angel Valdiviezo A. (Escuela Superior Politécnica del Litoral), Oddur Sigurdson and Tómas Jóhannesson (Icelandic Meteorological Office), Kevin Dockery (former Irish Garda), Sigurd Nerhus and Denise Ruther

(Western Norway University of Applied Sciences), Gylfi Gylfason (Just Icelandic), Bo Liu (Southwest Jiaotong University), Margaret Darrow (University of Alaska Fairbanks), Pascal Sirguey (Mountain Research Centre, Aotearoa/New Zealand) for kindly providing photos. Kejie Chen (Southern University of Science and Technology), for discussions of case study #13. Corey Scheip (BGC) for recommending case studies. Forrest Williams (Alaska Space Facility) and Eirik Malnes (NORCE) for discussing the interpretation of edges. Al Handwerger (JPL Laboratory, NASA), for feedback on SAR image processing method. Erlend Andenaes, Ivan Depina, Ola Fredin, and Tore Kvande (NTNU) for discussion of results and support.

Appendix 1: Links for landslide case study reports Accessed: 11-Dec-2022.

Location	Link
1. Iceland	https://blogs.agu.org/landslideblog/2021/10/08/multiple-landslides-in-thingeyjarsveit-and-in-kinnarfjoll-in-iceland/
2. Ireland	https://blogs.agu.org/landslideblog/2021/07/23/benbrack-1/
3. New Zealand	https://blogs.agu.org/landslideblog/2022/04/21/wairoa-1-2/
4. Ecuador	https://blogs.agu.org/landslideblog/2021/02/16/chunchi-a/
5. Norway	https://blogs.agu.org/landslideblog/2019/08/01/sogn-og-fjordane-1/
6. Sth. Africa	https://blogs.agu.org/landslideblog/2022/04/22/durban-1/
7. Vanuatu	https://hazmapper.org/2020/05/20/cyclone-harold-defoliation-and-mass-wasting-in-vanuatu/
8. Brazil	https://blogs.agu.org/landslideblog/2021/05/17/the-17-18-december-2020-landslide-disaster-in-presidente-getulio-southern-brazil/
9. China	https://blogs.agu.org/landslideblog/2021/10/28/the-21-july-2020-shaziba-landslide-at-mazhe-village-in-enshi-china/
10. Philippines	https://blogs.agu.org/landslideblog/2022/04/14/three-very-large-landslides-triggered-by-tropical-storm-megi-agaton/
11. Japan	https://link.springer.com/article/10.1007/s10346-019-01206-7
12. USA	https://blogs.agu.org/landslideblog/2022/08/03/haines/
13. China	https://blogs.agu.org/landslideblog/2021/10/26/the-5-april-2021-tiejiangwan-landslide-in-sichuan-province-china/
14. N. Zealand	https://www.otago.ac.nz/surveying/potree/pub/mrc/projects/matariki/changing-landscape
15. Iceland	https://blogs.agu.org/landslideblog/2018/07/26/fagraskogarfjall-landslide/
16. India	https://link.springer.com/article/10.1007/s10346-021-01802-6
17. India	https://link.springer.com/article/10.1007/s10346-021-01802-6
18. Norway	https://www.regobs.no/Registration/193067
19. India	https://link.springer.com/article/10.1007/s10346-020-01598-x
20. Peru	https://blogs.agu.org/landslideblog/2020/06/30/achoma-landslide-1/
21. Kygysztan	https://earthobservatory.nasa.gov/images/90255/landslide-in-southern-kyrgyzstan
22. Italy	https://blogs.agu.org/landslideblog/2019/11/25/savona-landslide-1/
23. Indonesia	https://blogs.agu.org/landslideblog/2022/03/29/mount-talakmau-1/
24. Brazil	https://blogs.agu.org/landslideblog/2022/05/31/recife-1/
25. Canada	https://blogs.agu.org/landslideblog/2021/11/16/bc-1/
26. USA	https://twitter.com/bclemms/status/1452333926949822468?lang=en
27. Burundi	https://hazmapper.org/2020/04/27/mass-wasting-in-burundi-december-2019/
28. Australia	https://blogs.agu.org/landslideblog/2022/03/11/main-arm-1/
29. Indonesia	https://link.springer.com/article/10.1007/s10346-021-01700-x
30. Turkey	https://blogs.agu.org/landslideblog/2019/05/17/ordu-1/

 doi:10.20944/preprints202302.0390.v1

References:

Arnold, D.W.D., Biggs, J., Wadge, G., Mothes, P., 2018. Using satellite radar amplitude imaging for monitoring syn-eruptive changes in surface morphology at an ice-capped stratovolcano. Remote Sens. Environ. 209, 480–488. https://doi.org/https://doi.org/10.1016/j.rse.2018.02.040

ASF, 2022. Introduction to SAR [WWW Document]. ASF HyP3 User Guid. URL https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar/ (accessed 10.27.22).

Baghdadi, N., El Hajj, M., Choker, M., Zribi, M., Bazzi, H., Vaudour, E., Gilliot, J.-M., Ebengo, D.M., 2018. Potential of Sentinel-1 Images for Estimating the Soil Roughness over Bare Agricultural Soils. Water 10. https://doi.org/10.3390/w10020131

Ban, Y., Zhang, P., Nascetti, A., Bevington, A.R., Wulder, M.A., 2020. Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning. Sci. Rep. 10, 1–15.

Beaudoin, A., Le Toan, T., Gwyn, Q.H.J., 1990. SAR observations and modeling of the C-band backscatter variability due to multiscale geometry and soil moisture. IEEE Trans. Geosci. Remote Sens. 28, 886–895. https://doi.org/10.1109/36.58978

Bouvet, A., Mermoz, S., Ballère, M., Koleck, T., Le Toan, T., 2018. Use of the SAR Shadowing Effect for Deforestation Detection with Sentinel-1 Time Series. Remote Sens. 10. https://doi.org/10.3390/rs10081250

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. Routledge.

Burrows, K., Marc, O., Remy, D., 2022. Using Sentinel-1 radar amplitude time series to constrain the timings of individual landslides: a step towards understanding the controls on monsoon-triggered landsliding. Nat. Hazards Earth Syst. Sci. 22, 2637–2653.

Casagli, N., Guzzetti, F., Jaboyedoff, M., Nadim, F., Petley, D.N., 2017. Hydrological risk: landslides. Underst. Disaster Risk Hazard Relat. Risk Issues-section II 209–218.

Chorlton, L.B., 2007. Generalized geology of the world: bedrock domains and major faults in GIS format (WMS). Geol. Surv. Canada, Open File 5529.

Cigna, F., Bateson, L.B., Jordan, C.J., Dashwood, C., 2014. Simulating SAR geometric distortions and predicting Persistent Scatterer densities for ERS-1/2 and ENVISAT C-band SAR and InSAR applications: Nationwide feasibility assessment to monitor the landmass of Great Britain with SAR imagery. Remote Sens. Environ. 152, 441–466. https://doi.org/https://doi.org/10.1016/j.rse.2014.06.025

Copernicus, 2019. Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3 [WWW Document]. Earth Engine Dev. URL https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_Landcover_100m_Proba-V-C3_Global?hl=en

Dabiri, Z., Hölbling, D., Abad, L., Helgason, J.K., Sæmundsson, Þ., Tiede, D., 2020. Assessment of Landslide-Induced Geomorphological Changes in Hítardalur Valley, Iceland, Using Sentinel-1 and Sentinel-2 Data. Appl. Sci. 10. https://doi.org/10.3390/app10175848

Dave Petley, 2022. The Landslide Blog [WWW Document]. URL https://blogs.agu.org/landslideblog

Dellow, S., Massey, C., Cox, S., 2017. Response and initial risk management of landslide dams caused by the 14 November 2016 Kaikoura earthquake, South Island, New Zealand, in: Proceedings of the 20th NZGS Geotechnical Symposium, Napier, New Zealand.

EEA/Copernicus, 2012. Copernicus CORINE Land Cover [WWW Document]. Earth Engine Dev. URL https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_CORINE_V20_100m

Froude, M.J., Petley, D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. 18, 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Fung, A.K., Li, Z., Chen, K.-S., 1992. Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sens. 30, 356–369.

Ganerød, A.J., Lindsay, E., Fredin, O., Myrvoll, T.-A., Nordal, S., Rød, J.K., 2023. Globally-vs Locally-trained Machine Learning Models for Land-slide Detection: A Case Study of a Glacial Landscape.

Gariano, S.L., Guzzetti, F., 2016. Landslides in a changing climate. Earth-Science Rev. https://doi.org/10.1016/j.earscirev.2016.08.011

GFDRR, 2023. Global Landslide Hazard Map [WWW Document]. World Bank. URL https://datacatalog.worldbank.org/search/dataset/0037584

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. Remote Sens. 11. https://doi.org/10.3390/rs11020196

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.-T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Rev. 112, 42–66. https://doi.org/10.1016/J.EARSCIREV.2012.02.001

Handwerger, A.L., Huang, M.-H., Jones, S.Y., Amatya, P., Kerner, H.R., Kirschbaum, D.B., 2022. Generating landslide density heatmaps for rapid detection using open-access satellite radar data in Google Earth Engine. Nat. Hazards Earth Syst. Sci. 22, 753–773.

Hanssen-Bauer, I., Drange, H., Førland, E.J., Roald, L.A., Børsheim, K.Y., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S., Sorteberg, A., others, 2009. Climate in Norway 2100. Backgr. Inf. to NOU Clim. Adapt. (In Nor. Klima i Norge 2100. Bakgrunnsmateriale til NOU Klim. Oslo Nor. klimasenter.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. A J. R. Meteorol. Soc. 25, 1965–1978. https://doi.org/10.1002/joc.1276.

Hoekman, D., Kooij, B., Quiñones, M., Vellekoop, S., Carolita, I., Budhiman, S., Arief, R., Roswintiarti, O., 2020. Wide-Area Near-Real-Time Monitoring of Tropical Forest Degradation and Deforestation Using Sentinel-1. Remote Sens. 12. https://doi.org/10.3390/rs12193263

Hungr, O., Leroueil, S., Picarelli, L., 2014. The Varnes classification of landslide types, an update. Landslides 11, 167–194. https://doi.org/10.1007/s10346-013-0436-y

Katiyar, V., Tamkuan, N., Nagai, M., 2021. Near-real-time flood mapping using off-the-shelf models with SAR imagery and deep learning. Remote Sens. 13, 2334.

Kedia, T., Ratcliff, J., O'Connor, M., Oluic, S., Rose, M., Freeman, J., Rainwater-Lovett, K., 2022. Technologies Enabling Situational Awareness During Disaster Response: A Systematic Review. Disaster Med. Public Health Prep. 16, 341–359. https://doi.org/10.1017/dmp.2020.196

Kellndorfer, J., Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., 2019. Using SAR data for mapping deforestation and forest degradation. SAR Handbook. Compr. Methodol. For. Monit. Biomass Estim. ServirGlobal Hunstville, AL, USA 65–79.

Kjekstad, O., Highland, L., 2009. Economic and Social Impacts of Landslides, in: Landslides – Disaster Risk Reduction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 573–587. https://doi.org/10.1007/978-3-540-69970-5_30

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated.

Lacroix, P., Bièvre, G., Pathier, E., Kniess, U., Jongmans, D., 2018. Use of Sentinel-2 images for the detection of precursory motions before landslide failures. Remote Sens. Environ. 215, 507–516.

Lindsay, E., Frauenfelder, R., Rüther, D., Nava, L., Rubensdotter, L., Strout, J., Nordal, S., 2022. Multi-Temporal Satellite Image Composites in Google Earth Engine for Improved Landslide Visibility: A Case Study of a Glacial Landscape. Remote Sens. 14. https://doi.org/10.3390/rs14102301

Meyer, F., 2019. Spaceborne Synthetic Aperture Radar: Principles, Data Access, and Basic Processing Techniques, in: SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. NASA.

Mondini, A., Santangelo, M., Rocchetti, M., Rossetto, E., Manconi, A., Monserrat, O., 2019. Sentinel-1 SAR Amplitude Imagery for Rapid Landslide Detection. Remote Sens. 11, 760. https://doi.org/10.3390/rs11070760

Mondini, A.C., Guzzetti, F., Chang, K.-T., Monserrat, O., Martha, T.R., Manconi, A., 2021. Landslide failures detection and mapping using Synthetic Aperture Radar: Past, present and future. Earth-Science Rev. 216, 103574. https://doi.org/https://doi.org/10.1016/j.earscirev.2021.103574

Nava, L., Monserrat, O., Catani, F., 2022. Improving Landslide Detection on SAR Data Through Deep Learning. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3127073

Prakash, N., Manconi, A., Loew, S., 2021. A new strategy to map landslides with a generalized convolutional neural network. Sci. Rep. 11, 9722. https://doi.org/10.1038/s41598-021-89015-8

Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.-E., Odongo-Braun, C., Vollrath, A., Weisse, M.J., Stolle, F., Pickens, A., others, 2021. Forest disturbance alerts for the Congo Basin using Sentinel-1. Environ. Res. Lett. 16, 24005.

Santangelo, M., Cardinali, M., Bucci, F., Fiorucci, F., Mondini, A.C., 2022. Exploring event landslide mapping using Sentinel-1 SAR backscatter products. Geomorphology 397, 108021. https://doi.org/https://doi.org/10.1016/j.geomorph.2021.108021

Shen, X., Wang, D., Mao, K., Anagnostou, E., Hong, Y., 2019. Inundation Extent Mapping by Synthetic Aperture Radar: A Review. Remote Sens. 11. https://doi.org/10.3390/rs11070879

Sudmanns, M., Tiede, D., Augustin, H., Lang, S., 2020. Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. Int. J. Digit. Earth 13, 768–784.

Tempfli, K., Huurneman, Gc., Bakker, Wh., Janssen, L.L.F., Feringa, W.F., Gieske, A.S.M., Grabmaier, K.A., Hecker, C.A., Horn, J.A., Kerle, N., others, 2009. Principles of remote sensing: an introductory textbook. International Institute for Geo-Information Science and Earth Observation.

Ulaby, F., Dobson, M.C., Álvarez-Pérez, J.L., 2019. Handbook of radar scattering statistics for terrain. Artech House.

van Natijne, A.L., Bogaard, T.A., van Leijen, F.J., Hanssen, R.F., Lindenbergh, R.C., 2022. World-wide InSAR sensitivity index for landslide deformation tracking. Int. J. Appl. Earth Obs. Geoinf. 111, 102829. https://doi.org/https://doi.org/10.1016/j.jag.2022.102829

Vollrath, A., Mullissa, A., Reiche, J., 2020. Angular-Based Radiometric Slope Correction for Sentinel-1 on Google Earth Engine. Remote Sens. 12. https://doi.org/10.3390/rs12111867

Waksom, M., 2022a. seaborn.violinplot [WWW Document]. Seaborne. URL https://seaborn.pydata.org/generated/seaborn.violinplot.html?highlight=violin&fbclid=IwAR1DDTMdluEMZNbvKMfFiUO0jhRCyWcmYBCqvBUBPopybyNPHRTtGJIRIYI#seaborn.violinplot

Waksom, M., 2022b. seaborn.histplot [WWW Document]. Seaborne. URL https://seaborn.pydata.org/generated/seaborn.histplot.html?highlight=histplot&fbclid=IwAR3jDVV1F0iX5S8ATZhxqnXhgAa22vMzFx3GkANQKFc1xYuicPVdQk9BmXM#seaborn.histplot

Williams, J.G., Rosser, N.J., Kincey, M.E., Benjamin, J., Oven, K.J., Densmore, A.L., Milledge, D.G., Robinson, T.R., Jordan, C.A., Dijkstra, T.A., 2018. Satellite-based emergency mapping using optical imagery: experience and reflections from the 2015 Nepal earthquakes. Nat. hazards earth Syst. Sci. 18, 185–205.

Zhang, P., Ban, Y., Nascetti, A., 2021. Learning U-Net without forgetting for near real-time wildfire monitoring by the fusion of SAR and optical time series. Remote Sens. Environ. 261, 112467.

Article V.

Automating global landslide detecting with heterogenous ensemble deep-learning classification

Alexandra Jarna Ganerød [1,2*†], Gabriele Franch [3†], Erin Lindsay [4], Martina Calovi [1]

[1] NTNU, Department of Geography; alexandra.jarna@ntnu.no, martina.calovi@ntnu.no, jan.rod@ntnu.no
[2] NGU, Geological Survey of Norway; alexandra.jarna@ngu.no
[3] FBK, Fondazione Bruno Kessler, Trento, Italy franch@fbk.eu
[4] NTNU, Department of Civil and Environmental Engineering; erin.lindsay@ntnu.no
† These authors contributed equally to this work

Automating global landslide detection with heterogeneous ensemble deep-learning classification

Alexandra Jarna Ganerød [1,2*†], Gabriele Franch [3†], Erin Lindsay [4], Martina Calovi [1]

[1] NTNU, Department of Geography, Trondheim, Norway; alexandra.jarna@ntnu.no, martina.calovi@ntnu.no
[2] NGU, Geological Survey of Norway, Trondheim, Norway; alexandra.jarna@ngu.no
[3] FBK, Fondazione Bruno Kessler, Trento, Italy; franch@fbk.eu
[4] NTNU, Department of Civil and Environmental Engineering, Trondheim, Norway; erin.lindsay@ntnu.no

* Correspondence: alexandra.jarna@ntnu.no; Tel: +47 4512 6688
† Author 1 and Author 2 contributed equally to this work

Keywords: Deep learning, Segmentation model, Ensemble model, Rapid detection, Global landslides

Highlights:

- **Building a heterogeneous ensemble of models substantially improves the prediction performance.**
- **Ensemble learning methods reduced the uncertainty of the single models' results.**
- **The best result is achieved with using S1 & S2 only data (F1 = 0,66)**
- **Biggest improvement Setting 2 only S2 bands ensemble model increase from 0.52 to 0.61.**
- **Possibility of creating a monitoring system based only on Setting 2 (dNDVI & land cover classification) – notice of possible change in the area.**

ABSTRACT

With changing climatic conditions, we are already seeing an increase in extreme weather events and their secondary consequences, including landslides. Landslides threaten infrastructure, including roads, railways, buildings, and human life. Hazard-based spatial planning and early warning systems are cost-effective strategies to reduce the risk to society from landslides. However, these both rely on data from previous landslide events, which is often scarce. Many deep learning (DL) models have recently been applied for landside mapping using medium- to high-resolution satellite images as input. However, they often suffer from sensitivity problems, overfitting, and low mapping accuracy. This study addresses some of these limitations by using a diverse global landslide dataset, using different segmentation models, such as Unet, Linknet, PSP-Net, PAN, and DeepLab and based on their performances, building an ensemble model. The ensemble model achieved the highest F1-score (0.69) when combining both Sentinel-1 and Sentinel-2 bands, with the highest average improvement of 6.87 % when the ensemble size was 20. On the other hand, Sentinel-2 bands only performed very well, with an F1 score of 0.61 when the ensemble size is 20 with an improvement of 14.59 % when the ensemble size is 20. This result shows considerable potential in building a robust and reliable monitoring system based on changes in vegetation index dNDVI only.

Introduction

Landslides occur daily around the world and can cause significant damage to infrastructure and property. Human life and health can also be at stake. In recent years, their occurrence and associated costs have increased, given the increase in extreme precipitation events and unregulated urban expansion in landslide-prone areas (Froude and Petley, 2018; Gariano and Guzzetti, 2016; Hanssen-Bauer et al., 2009). However, landslides are frequently underrepresented in global natural catastrophe databases due to a lack of reporting or because insurance claims due to landslide damages may be categorised under the primary triggering event, such as earthquakes or extreme weather events such as hurricanes or floods (Kirschbaum et al., 2010).

The scarcity of data on previous landslide events is a major limiting factor in developing mitigating strategies for future risks posed by landslides (Guzzetti et al., 2012). Two cost-effective methods for mitigating landslide risk to society include hazard-based spatial planning of new developments and early warning systems (Piciullo et al., 2018; Shano et al., 2020). However, these both rely on data from previous landslide events. Recent technological developments offer great potential to improve landslide detection through automatic landslide monitoring systems using satellite imagery and deep learning-based image classification algorithms (Mondini et al., 2021; Tehrani et al., 2021). Since 2017, there has been a significant increase in research into this topic (Mondini et al., 2021).

Two types of satellite images are mainly used for automatic landslide detection. These are optical images, from which vegetation indices (commonly, the normalised difference vegetation index, NDVI) can be derived, and synthetic aperture radar (SAR) images. Change detection methods are often used, either from single pre- and post-event images or composites produced from multitemporal image stacks, that help reduce noise from seasonal changes in vegetation or snow, clouds, or speckle (Lindsay et al., 2022). Landslides are most clearly visible in optical change images; however, SAR imagery has the advantage of being cloud penetrating and, therefore, can detect landslides more rapidly in case of scarcity of cloud-free images. With optical images, it is common to see landslides based on a loss of vegetation, which produces a negative NDVI value (Lin et al., 2005). For the changes in ground surface texture and the associated scattering behaviour that is detected by using cross- or co-polarised backscatter intensity images, we are using SAR images (Lindsay et al., 2022). A detailed explanation of the physical mechanisms that control landslide expression in SAR backscatter change images is available (Lindsay et al., 2022). Geographical Information Systems (GIS) have helped with data preparation, but manual mapping remains a time-consuming and subjective practice. Therefore, there is considerable potential in using automated machine- and deep-learning techniques to increase data collection and make landslide mapping practice more efficient (Bai et al., 2022; Ganerød et al., 2023; Ghorbanzadeh et al., 2019; Nava et al., 2022; Prakash et al., 2021).

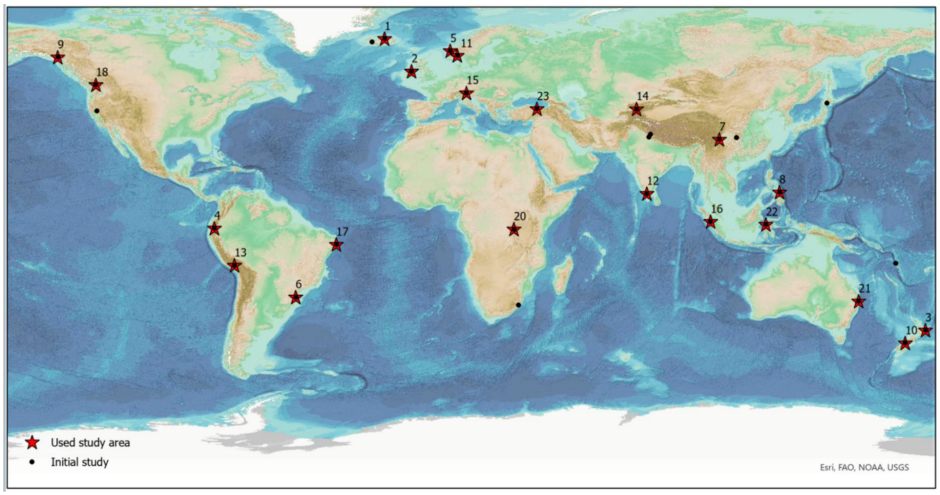
Computer-based methods are rapidly emerging as powerful, efficient, and viable tools able to work with high-dimensional datasets (Chen and Lin, 2014; Dargan et al., 2020; Zhu et al., 2017). These methods can thus reduce costly and time-consuming manual labour both in the field and during post-digital interpretation. Applying computer-based methods to the landslide susceptibility maps can play a crucial role in determining the most vulnerable and prone areas for landslides (Shano et al., 2020; Tehrani et al., 2022; Wei et al., 2022; Youssef and Pourghasemi, 2021). The well-mapped landslides are the primary key to further developing a reliable monitoring system. And since landslide hazard mapping relies on knowledge gained from previous events (Band et al., 2020; Mondini et al., 2021; Prakash et al., 2021), the availability of high-quality data can help in developing strategies that aim to reduce landslide hazards and consequently, the risk to society.

The literature shows that qualitative and quantitative methods can be applied to detect landslides (Saadatkhah et al., 2014). Data-driven machine-learning techniques are widely used nowadays, but they still present some weaknesses that affect the predictive performances of single models (Lv et al., 2022). The scarcity of training data gives typical weaknesses. Most DL models require a large amount of diverse training data. This issue can cause the possibility of missing the best-fit function during the training process (Blum, A., Wang, 2011; Thomas G. Dietterich, 1997; Wang et al., 2011). However, considering ensemble learning in landslide applications, the majority of the studies are conducted for landslide susceptibility mapping (Bai et al., 2022; Hong et al., 2020; Lin et al., 2022; Liu et al., 2021; Lv et al., 2022; Merghadi et al., 2020; Saha et al., 2021; Setargie et al., 2023; Tong et al., 2023; Wei et al., 2022) rather than landslide detection or prediction.

With this study, we compare different segmentation models using PyTorch, in combination with two different learning rates (0,001 and 0,01) and five loss functions to test how the prediction and statistics change using different settings and Fully Convolutional Neural Networks (FCNN). Out of 30 well-mapped study areas, (Lindsay et al., 2023) presented as potential case studies, 21 representing the descending orbit were selected for our study. Limitations are connected to the landslides on east-facing slopes that are more likely to be detectable in descending SAR data due to increased sampling density on slopes facing towards the sensor and geometric distortions. In contrast, the total number of case studies with descending obit was higher. Furthermore, we evaluate whether the resulting predictions vary according to the different settings: (1) only Sentinel-1, (2) only Sentinel-2, (3) combinations of Sentinel-1 and Sentinel-2 and lastly, (4) all the bands by calculating precision, recall, F1-score on the test dataset. Based on all combinations, the different ensemble learning settings are assessed to calculate the performance of each ensemble model based on the 10, 20 and 40 best-performing single models on validation data. The goal of this study is threefold: first, to find the setting with the best performance and least data amount. Second, to create an ensemble model, compare its performance and evaluate its average improvement to the single models. Third, assess the resulting performances and discuss their potential use.

Case studies

This study uses the global landslide datasets created by Lindsay et al. (2023b), where landslides have been manually mapped mainly through dNDVI images derived from Sentinel-2. This dataset included 30 case studies worldwide, characterised by varying terrain types, orientation and size, ground cover, climate zones, geological materials, and failure mechanisms. These 30 sites have been mainly identified thanks to the reports titled ‘The Landslide Blog’ (https://blogs.agu.org/landslideblog/), news reports, or journal articles found through an extensive online search. The locations and dates of these mapped landslide events are shown in Figure 1. In contrast, the specific properties of each event, together with the local environmental conditions, are shown in Table 1 (Lindsay et al., 2023).



1. Iceland	2021-10-06	(-17.5558, 65.8674)
2. Ireland	2021-07-04	(-7.85187, 54.15745)
3. New Zealand	2022-03-27	(177.4694, -38.9681)
4. Ecuador	2021-02-12	(-78.93655, -2.2895)
5. Norway	2019-07-30	(6.11699, 61.51734)
8. Brazil	2020-04-06	(-49.60777, -27.0774)
9. China	2021-04-05	(103.15055, 29.66474)
10. Phillipines	2022-04-10	(124.85286, 10.67658)
12. USA	2019-12-02	(-135.41194, 59.22035)
14. New Zealand	2022-03-27	(170.17485, -43.57436)
15. Iceland	2022-02-02	(-22.109811, 64.787941)
19. India	2020-08-06	(77.01188, 10.166)
20. Peru	2020-06-16	(-71.685998, -15.660252)
21. Kyrgzstan	2017-04-30	(73.290135, 40.587722)
22. Italy	2018-11-02	(11.972334, 46.34617)
23. Indonesia	2022-02-25	(99.99442, 0.11412)
24. Brazil	2022-05-26	(-34.958299, -8.120723)
25. Canada	2021-11-14	(-121.63403, 49.307154)
26. USA	2021-09-24	(-121.38802, 39.86029)
28. Australia	2022-02-28	(153.39578, -28.51734)
29. Indonesia	2018-09-28	(119.91418, -0.93951)

Figure 1. Lindsay et al., 2023 case study locations and event dates (Lindsay et al., 2023). The case studies where landslides were not visible or did not have descending data have been removed—base map: Landslide Hazard Map (GFDRR, 2023).

Table 1. Lindsay et al., 2023: 21 used case studies. **Type** refers to the different types of landslides, specifically: 1. Rockfall, 14. Clay/silt compound slide, 22. Debris flow, 23. Mudflow, 25. Debris avalanche, 26. Earthflow, 27. T refers to the trigger: R – rainfall, E- earthquake, S – snowmelt, U - unknown. Size presents length and width in kilometres, Rainfall in millimetres per year and occurrence date.

Location	Type	T	Size L x W [km]	Rainfall [mm/yr]	Date
1. Iceland*	22	R	0.8 x 0.1	672	06.10.2021
2. Ireland	27	R	0.58 x 0.7	1358	04.07.2021
3. N. Zealand	13	R	0.13 x 0.05	1508	27.03.2022
4. Ecuador	26	R	1.5 x 1.5	918	12.02.2021
5. Norway*	22 25	R	0.11 x 0.03	2285	30.07.2019
8. Brazil	22	R	1.6 x 0.02	1547	06.04.2020
9. China	13-22	R	1.34 x 0.92	1297	05.04.2021
10. Philippines	23	R	2.1 x 0.7	2915	10.04.2022
12. USA	25	R	1.7 x 0.18	1282	02.12.2019
14. N. Zealand	18	R	1.8 x 0.28	4222	27.03.2022
15. Iceland	18-25	R	2.4 x 1.7	829	02.02.2022
19. India	22	R	1.2 x 0.12	2848	06.08.2020
20. Peru	26	U	0.6 x 1	506	16.06.2020
21. Kyrgyzstan	14-26	RS	5 x 0.6	394	30.04.2017
22. Italy	22	R	0.35 x 0.07	886	02.11.2018
23. Indonesia	22	E	6 x 0.3	2775	25.02.2022
24. Brazil	13	R	0.06 x 0.03	1678	26.05.2022
25. Canada	18	R	0.85 x 0.32	1712	14.11.2021
26. USA	1	U	0.09 x 0.06	1560	24.09.2021
28. Australia	22	R	0.8 x 0.04	2031	28.02.2022
29. Indonesia	16	E	2.1 x 1.1	1534	28.09.2018

In this study, we included only the case studies with Sentinel-1 – 1 descending orbit images to have the highest amount of use cases with the same parameters (21 cases). Only the visible landslides in both Sentinel-1 and Sentinel-2 bands (Figure 2) have been included for the final evaluation of each single model. We

classified them as follows: The Norway case (#5) was reserved as a test and used to evaluate the area. In contrast, the remaining 20 cases have been randomly split into 16 locations for model training and 4 locations for model validation. This split was performed only once and has been used for all the experiments (fixed random seed). Figure 2 shows the landslides mapped in this area's Norwegian (#5) evaluation set.

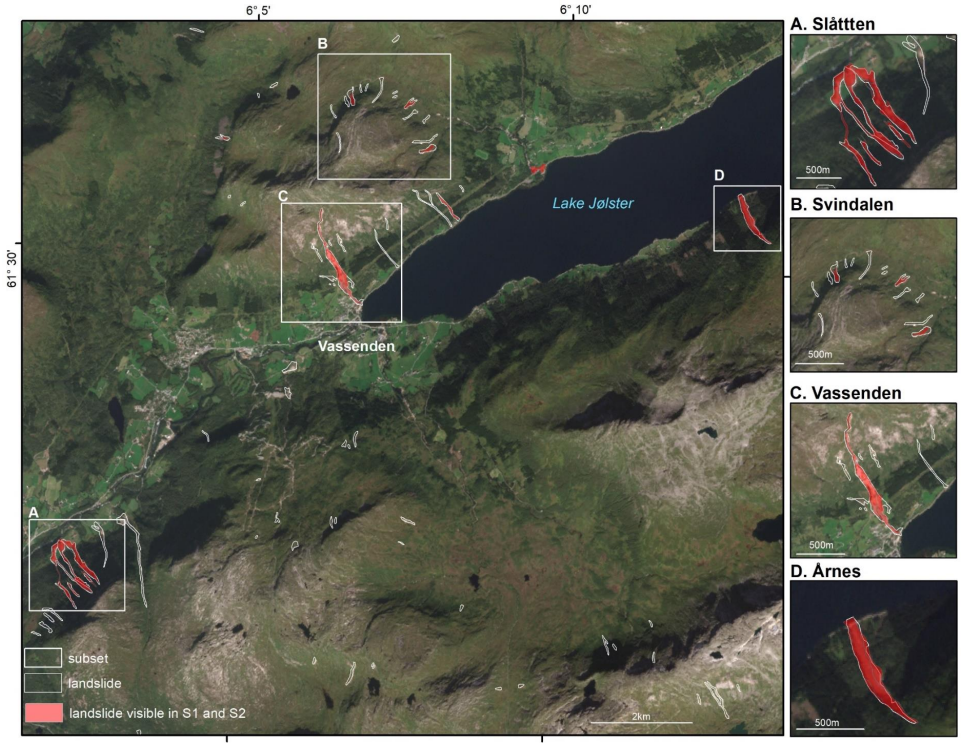


Figure 2: Norwegian case study area #5. Norway is the test area, showing four subsets with ground truth landslide outlines in white. However, only red-defined regions have been used for test evaluation because they are visible in both Sentinel-1 and Sentinel-2 layers.

Looking at the size of the case study areas, eight case studies show less than 0.1 km^2, and five were between 0.1 and 1 km^2. The remaining eight events are between 1 and 5 km^2 (Table 1). The mean annual rainfall varied from 394 mm/yr in Kyrgyzstan (#21) to 4.222 mm/yr in New Zealand (#14). The land cover types, identified using the Copernicus Global Land Cover classification map, included 7 cases of forest and 6 of herbaceous, cropland, or shrubs. For most of our case studies, the main trigger events that initiated the landslide are rainfall (16 events), earthquakes (2 events), a combination of rainfall and snowmelt (1 event only), and in 2 cases, the triggers are unknown (see Table 1).

Methods – data and computing

3.1 Input data – bands

This section describes the input data and the pre-processing, visualisation, and interpretation methods, summarised by Lindsay et al. (2023). The Sentinel-1 SAR change images and time series were prepared for each case study using Google Earth Engine (GEE) (Gorelick et al., 2017). The date of occurrence is crucial because it defines the correct ranges for the pre-and post-event image collections. The coordinates of the approximate event location have been used as filter conditions to produce pre- and post-event image stacks of both Sentinel-1 GRD (Ground Range Detected) products and Sentinel-2 (Level 2A) images. Periods of one, two, or 12 months before or after the event date were used, ensuring comparison between temporally homogenous periods, depending on the local image acquisition frequency and seasonality. The date ranges and coordinate locations for each case study are shown in Figure 1. Pre- and post-event image composites were produced from the image stacks for a 4 km2 area about the defined point for all the cases. The Sentinel-1 images, retrieved from GEE, are available pre-processed (calibrated and ortho-corrected) at 10 m resolution. A terrain correction (Vollrath et al., 2020) was applied to each image in the stack using either a volumetric or a surface model, depending on the land cover type. The terrain correction projects the side-looking images onto the terrain while masking out pixels with shadow, and layover corrects over- and under-brightening. The 30 m resolution Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), available within GEE, was used for this. However, local 10 m DEMs were utilised for the Icelandic case (link: ÍslandsDEM v1.0 10m - awesome-gee-community-catalog, accessed: 11 December 2022) and for Norwegian patients (https://hoydedata.no/LaserInnsyn2/, accessed: 11 December 2022).

Furthermore, Sentinel-1 composites were created by taking the median of the terrain-corrected image collections and changing images (post minus pre-event image composite). For Sentinel-2, an NDVI band was added to each image in the pre-and post-event stacks; then, a greenest-pixel composite was created (maximum NDVI) using the quality-mosaic tool. We produced an NDVI change image (dNDVI) as a final product by subtracting the pre- from the post-event composite. The dNDVI images are the only Sentinel-2-band included in the dataset (Figure 3).

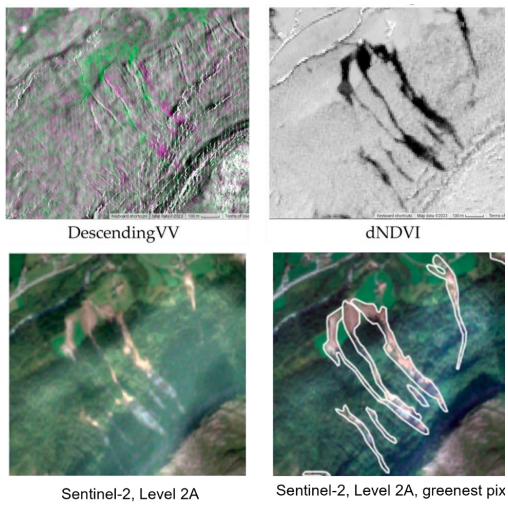


Figure 3: Example of the landslide showing case study #5 (test case study as a subset) in (a) multi-temporal VV-polarized SAR backscatter descending intensity change images (green indicates a backscatter intensity increase; purple indicates a decrease); (b) Differenced Normalised Difference Vegetation Index (dNDVI); (c) Sentinel- Level 2A, with atmospheric correction applied to the Level-1C TOA image. (d) cloud-filtered, greenest-pixel composite produced from Sentinel-2 Level2A images (white outlines were mapped from the Sentinel-2 dNDVI image).

Table 2 Bands description. Acronyms: pre- pre-event, post – post-event image. V – vertical polarisation, H – horizontal polarisation; Digital Elevation Model (DEM)

Band		Resolution [m]	Origin	Description
0	**preVV**	10	Sentinel-1 + DEM	Pre-event VV polarised backscatter intensity. VV is good for showing the difference in surface roughness.
1	**postVV**	10	Sentinel-1 + DEM	Post-event VV polarised backscatter intensity.
2	**dVV**	10	Sentinel-1 + DEM	PostVV minus PreVV
3	**preVH**	10	Sentinel-1 + DEM	Pre-event VH polarised backscatter intensity. VH is more sensitive to changes in biomass.
4	**postVH**	10	Sentinel-1 + DEM	Post-event VH polarised backscatter intensity. VH is more sensitive to changes in biomass.
5	**dVH**	10	Sentinel-1 + DEM	PostVH minus PreVH
6	**Layover**	10	Sentinel-1 + DEM	Mask of layover distortion areas, calculated from (Vollrath et al., 2020) terrain correction. It should match the NoData areas in the VV and VH bands.
7	**Shadow**	10	Sentinel-1 + DEM	Mask of shadow distortion areas, calculated from the Vollrath et al. (2020) terrain correction. It should match the NoData areas in the VV and VH bands.
8	**liaDeg**	10	Sentinel-1 + DEM	The local incidence angle (LIA) is in degrees. The angle describes the orientation of the ground surface relative to the sensor's line of sight (Lindsay et al., 2023).
9	**Elevation**	30	DEM	The SRTM global 30 m resolution DEM was used for case studies below 60 deg latitude. Above 60 deg, other local DEMs were used – Icelandic 10 m and Norway 10 m resolution.
10	**Slope**	30	DEM	Derived from elevation
11	**KG_climate**			Köppen-Geiger climate map with a spatial resolution of 5 arc minutes for 1986-2010. http://koeppen-geiger.vu-wien.ac.at/
12	**Popatpv_tree height**	30	LiDar	A 30-m spatial resolution global forest canopy height map - through the integration of the Global Ecosystem Dynamics Investigation (GEDI) LIDAR forest structure measurements and Landsat analysis-ready data time series (Landsat ARD) (Potapov et al., 2021).
13	**CART_LC_classified**	10	Sentinel-1 and Sentinel-2	Land cover classification. This was prepared for each case study individually, based on Sentinel-1 and -2 images from pre-event, using a CART machine-learning algorithm in Google Earth Engine.
14	**dNDVI**	10	Sentinel-2	Change in Normalised Difference Vegetation Index (NDVI)

The Local Incidence Angle (LIA) (liaDeg in Table 2) was included to investigate the factors that influence the expression of landslides in the change images. The LIA is particularly important when there has been a height change due to erosion of soil or rock material or the removal of trees (O'Grady et al., 2013).

Here, we used the manually mapped landslide polygons for each case study, exporting the shapefiles along with the Geotiff raster images for each case study, and finally extracting and plotting the pixel values from within the polygons.

For land cover classification (CART_LC_classified), locally trained land cover maps were produced. A machine-learning-based land cover classification was performed for each case study using the ee. smile.CART algorithm (Classification and Regression Tree) in GEE (Breiman et al., 2017). These were trained using a Sentinel-2 image with minimal cloud cover from before the event and the DEM and pre-event Sentinel-1 composite images (VV and VH polarisation) (Lindsay et al., 2023). The classifier was trained by manually selecting points within the following classes: urban, artificial; forest, trees, scrub, herbaceous; pasture, grass; sparsely vegetated; water body, sea; wetland; bare rocks; and glacier and snow (Lindsay et al., 2023).

Finally, for each case study, the following were exported: 1) landslide polygons (reclassified as a binary geotiff), (2) geotiff raster images, including the pre-processed Sentinel-1 bands from only descending mode, were used for training. The pixel values within each of the landslide polygons were extracted to produce a dataset consisting of approximately 300,000 pixels. As a result, from the initial 30 case studies, only 21 have been included in this study.

3.2 Proposed settings

All the presented bands are in logical content and calculated with the use of DEM, Sentinel-1, or Sentinel-2 data. Here we decided to apply our combination of segmentation models, loss functions, and learning rate (Section 3.3) as follows (Table 3): Setting 1 is proposed to use nine bands created with the use of only Sentinel-1 data, while Setting 2 presents two bands of multi-temporal dataset containing only Sentinel-2 data. Setting 3 combines both Sentinel-1 and Sentinel-2 bands and Setting 4 includes all 15 multi-temporal bands.

Table 3. Proposed and tested settings 1-4 in this study.

	Setting	Bands
1	only Sentinel-1	0,1,2,3,4,5,6,7,8
2	only Sentinel-2	13,14
3	Sentinel-1 and Sentinel-2	0,1,2,3,4,5,6,7,8,13,14
4	all bands	0,1,2,3….14

3.3 Segmentation models and loss functions

Image segmentation is a crucial task in computer vision and image processing. It has various applications, including scene understanding, medical image analysis, robotic perception, video surveillance, augmented reality, and image compression (Blaschke, 2010; Blaschke et al., 2014). The widespread success of deep learning (DL) has led to the development of image segmentation approaches that utilise DL models (Al-Obeidat et al., 2016; Minaee et al., 2022) in many fields of application. In our study, we build upon these foundations. Specifically, we relied on the models provided by the open-source library segmentation models PyTorch (SMP - https://smp.readthedocs.io/), which implements nine different segmentation models (many of which have shown state-of-the-art results when presented). The aim was to evaluate them both in single model configuration and ensemble settings to analyse the performances for the geospatial dimension in landslide mapping with high imbalance in the dataset (Table 4).

Loss functions guided deep learning model training to optimise model parameters (weights). The loss function compared the ground truth and predicted output values, and the model weights were optimised to minimise the loss values between the predicted and target outputs. In our work, we trained all nine models on five different loss functions among all losses provided by the SMP library: BCE, Dice, Focal, Jacckard, and Lovasz Loss (Table 5). The loss functions were chosen as these are known to be a good fit for semantic segmentation problems (Jadon, 2020).

The last configuration we explore for model training is the learning rate parameter, where we test both 0.01 and 0.001 as possible values. To recap, we trained nine deep learning model architectures (Table 4) in combination with five loss functions (Table 5): BCELoss, DiceLoss, FocalLoss, JaccardLoss, LovaszLoss, and two learning rates of 0.01 and 0.001, which in total resulted in 9x5x2 = 90 combinations (training sessions) for each Settings 1-4 (Table 3). The best-performing models on the validation set among these 90 were further used to form the multi-model ensembles.

Table 4. Segmentation models. All the models are Fully Convolutional Neural Networks (FCNN) with the same supported metadata: classified tiles and task: pixel classification.

Segmentation Models	Stands for	Categories (Minaee et al., 2022)	Specialty	Source
Unet	U shape	Encoder-decoder (3)	works with fewer training images	(Ronneberger et al., 2015)
Unet++	U shape ++	Nested encoder-decoder (3)	more complex decoder	(Zhou et al., 2018)
MA-Net	Multi-scale Attention Net	Multiscale and Pyramid Network-Based Model (4)	Two blocks: position-wise and multi-scale fusion attention block	(Fan et al., 2020)
Linknet	LinkNet	Encoder-decoder (3)	uses sum for fusing decoder blocks	(Chaurasia and Culurciello, 2018)
FPN	Feature Pyramid Network	Multiscale and Pyramid Network-Based Model (4)	without the need to compute image pyramids - suitable for small objects	(Li *et al.*, **2019**)
PSP-Net	Pyramid Scene Parsing Network	Multiscale and Pyramid Network-Based Model (4)	not suitable for small objects-based context aggregation.	(Zhao et al., 2017)
PAN	Pyramid Attention Network	The Regional CNN (R-CNN) (5)	data-to-text generation	(Jiang et al., 2020)
DeepLabV3	DeepLab version 3	Dilated Convolutional Model (6)	uses dilated convolutions	(Chen et al., 2018a)
DeepLabV3+	DeepLab version 3+	Dilated Convolutional Model (6)	adds a simple yet effective decoder module	(Chen et al., 2018b)

Table 5. Loss functions used in this study all support binary, multiclass, and multilabel. Here, we describe their specifics.

Loss function	Specifics
BCELoss	creates a criterion that measures the Binary Cross Entropy between the target and the input probabilities.
DiceLoss	a standard metric for pixel segmentation that can also be modified to act as a loss function
FocalLoss	applies a modulating term to the cross-entropy loss in order to focus learning on complex misclassified examples. It is a dynamically scaled cross-entropy loss, where the scaling factor decays to zero as confidence in the correct class increases.
JaccardLoss	similar to the Dice metric and is calculated as the ratio between the overlap of the positive instances between two sets and their mutual combined values
LovaszLoss	It is designed to optimise the Intersection over Union score for semantic segmentation, particularly for multi-class instances.

3.4 Training, validation and model selection

Data augmentation (Ghasemi et al., 2022) is a technique for artificially increasing the training set by creating modified copies of a dataset from the same training data. Given the relatively small size of our training set (16 images of different sizes), we employ multiple augmentation options to increase the data diversity as much as possible. Among many augmentation techniques, in our experiment, we choose to employ both “smart” random cropping with a resolution of 256x256 pixels and random 90-degree rotation. The “smart” random cropping selects a random area of the image to contain at least one positive pixel (a pixel marked as a landslide); this is done to reduce the data imbalance and guarantee that every input image always contains both labels.

We employ an early stopping technique for all training sessions to avoid overfitting. Specifically, we train the model for a fixed number of steps (1000 iterations with batch size 4, forming 1 training epoch). Then, we evaluate the classification performance on the validation set (4 images) by computing the F1 score metric (see section 3.5 for the F1 score definition). When the F1 score no longer improves for 50 epochs, we stop training and keep the models’ weights with the best F1 score on the validation set as the final model for testing.

3.5 Evaluation and ensemble model

Model evaluation was performed quantitatively by computing the confusion matrix between the label and the model’s classification on the Norwegian test set (Figure 2). Table 6 shows the equations for the performance metrics (precision, recall, and F1-score, all ranging from 0 to 1) computed from confusion, starting from true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values. We used the F1 score as the main performance metric given its suitability in highlighting classification performance differences on datasets with unbalanced classes (with 0 indicating a model with no skill and 1 indicating a perfect skill).

Table 6. Equations for performance evaluation metrics from confusion matrix values.

Metric	Formula
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2TP}{2TP + FP + FN}$

All metrics are summarised in the performance diagram (Roebber, 2009) and presented in Figure 5, which shows precision, recall, F1 score and frequency bias in a single plot, highlighting both the overall performance and if the model or ensemble is overpredicting or underpredicting respect to the ground truth.

As introduced in section 3.3, we trained nine segmentation architectures with five loss functions and two learning rates on four different settings, creating a total of 360 models (9x5x2x4) with other statistical proprieties. In statistics and machine learning, ensemble methods are commonly used to combine multiple model outputs to obtain better predictive performance than could be obtained from any of the constituent learning algorithms on their own (Band et al., 2020; Yang et al., 2022). We built and tested ensembles of different sizes for all four settings by averaging the best-performing models' classification results on the validation set (Table 7). Figure 4 graphically summarises the ensemble prediction process:

- The K best models (2 <= K <= 90) with the highest score on the validation set are collected to form the ensemble.
- Input data is processed individually by each model in the ensemble, generating K output classification maps.
- The pixel average of the K classification maps is computed as the final classification output.

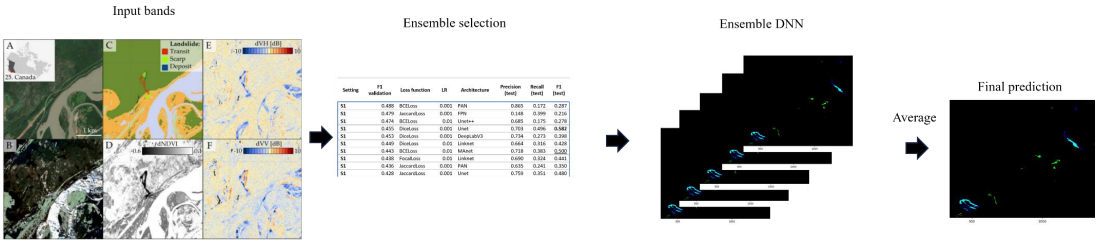


Figure 4: Workflow of compiling the final prediction using ensemble techniques. The best-performing models' output is averaged to obtain the final prediction.

4. Results

In this section, we analyse the outcomes of all the four settings, both on a single model perspective and with ensembles of different sizes focused on the automated detection of landslides using an ensemble global deep-learning approach.

4.1 Single model performance

Table 7 summarises the characteristics (model architecture, loss function, learning rate) of the ten best-performing combinations of each Setting 1-4. On average, lower scores on the validation translate into lower performances on the test set, indicating a positive correlation between training/validation and test performance. Thus, the models were correctly trained to generalise without overfitting. The worst performing setting is Sentinel 1 (S1), where the best validation F1 scores range between 0.428 and 0.488, with a corresponding range in test performance between 0.216 and 0.582, with 5 out of 10 models with a test score below 0.4. Sentinel 2 (S2) shows substantially improved validation scores, from 0.637 to 0.694 and consistently more stable test performances, from 0.42 to 0.56. The best performances overall are given by the settings using both Sentinel 1 and Sentinel 2 bands (S1+S2) with a very consistent high validation scores (0.741 to 0.751) and test performance (from 0.581 to 0.695). Counterintuitively, the additions of other bands on top of S1 and S2 in the last setting (all) result in a slight degradation in performance in validation (0.706 to 0.727) and mixed results in test (0.591 to 0.687), hinting that the additional bands may be of limited or detrimental utility for landslide identification.

Interestingly, in the best-performing setting, S1+S2, 7 out of 10 models use the Unet++ architecture. All the other locations instead show a varied mix of model, loss and learning rate combination in the top 10, highlighting that most architectures were able to deliver good performances given the proper configuration.

Table 7: Ten best-performing segmentation models for each setting based on validation F1 score and their corresponding performance metrics on the test set. In **bold,** the best performance on the test set underlined the worst performance.

Setting	F1 validation	Loss function	Learning Rate	Architecture	Precision (test)	Recall (test)	F1 (test)
S1	0.488	BCELoss	0.001	PAN	0.865	0.172	0.287
S1	0.479	JaccardLoss	0.001	FPN	0.148	0.399	0.216
S1	0.474	BCELoss	0.01	Unet++	0.685	0.175	0.278
S1	0.455	DiceLoss	0.001	Unet	0.703	0.496	**0.582**
S1	0.453	DiceLoss	0.001	DeepLabV3	0.734	0.273	0.398
S1	0.449	DiceLoss	0.01	Linknet	0.664	0.316	0.428
S1	0.443	BCELoss	0.01	MAnet	0.718	0.383	0.500
S1	0.438	FocalLoss	0.01	Linknet	0.690	0.324	0.441
S1	0.436	JaccardLoss	0.001	PAN	0.635	0.241	0.350
S1	0.428	JaccardLoss	0.001	Unet	0.759	0.351	0.480
S2	0.694	FocalLoss	0.01	Linknet	0.516	0.534	0.525
S2	0.691	BCELoss	0.01	Unet++	0.322	0.673	0.435
S2	0.687	JaccardLoss	0.001	PAN	0.553	0.510	0.531
S2	0.682	BCELoss	0.01	Linknet	0.355	0.514	0.420
S2	0.668	DiceLoss	0.01	Unet++	0.297	0.747	0.425
S2	0.645	JaccardLoss	0.01	DeepLabV3	0.516	0.490	0.503
S2	0.643	LovaszLoss	0.001	DeepLabV3+	0.495	0.646	**0.560**
S2	0.641	JaccardLoss	0.001	Unet	0.342	0.679	0.455
S2	0.640	FocalLoss	0.01	MAnet	0.375	0.730	0.496
S2	0.637	BCELoss	0.01	Unet	0.352	0.690	0.467
S1+S2	0.751	FocalLoss	0.01	Unet++	0.708	0.589	0.643
S1+S2	0.750	FocalLoss	0.001	Unet++	0.647	0.598	0.621
S1+S2	0.748	JaccardLoss	0.01	Unet++	0.791	0.620	**0.695**
S1+S2	0.747	DiceLoss	0.001	Linknet	0.806	0.589	0.681
S1+S2	0.746	JaccardLoss	0.001	FPN	0.709	0.492	0.581
S1+S2	0.746	DiceLoss	0.001	Unet++	0.636	0.616	0.626
S1+S2	0.745	JaccardLoss	0.001	Unet++	0.678	0.624	0.650
S1+S2	0.744	BCELoss	0.01	Unet++	0.716	0.546	0.619
S1+S2	0.742	DiceLoss	0.01	Unet++	0.648	0.661	0.655
S1+S2	0.741	BCELoss	0.01	Unet	0.584	0.612	0.597
all	0.727	DiceLoss	0.001	Unet	0.710	0.599	0.650
all	0.725	DiceLoss	0.001	Unet++	0.742	0.639	**0.687**
all	0.720	JaccardLoss	0.001	PAN	0.657	0.575	0.613
all	0.719	JaccardLoss	0.01	FPN	0.684	0.620	0.650
all	0.719	JaccardLoss	0.001	Unet	0.847	0.550	0.667
all	0.715	JaccardLoss	0.001	Linknet	0.812	0.556	0.660
all	0.713	DiceLoss	0.001	FPN	0.629	0.556	0.591
all	0.712	DiceLoss	0.01	Unet	0.788	0.568	0.660
all	0.708	FocalLoss	0.001	Unet	0.731	0.582	0.648
all	0.706	JaccardLoss	0.01	Unet++	0.770	0.565	0.652

4.2 Ensemble Performance

Figure 5 and Table 8 summarise the performance improvement on the test set of building an ensemble of different sizes by averaging the predictions of the best-performing validation models. All four settings showed a sensible improvement in F1 skill when compared to the single best validation model. The skill improvement was more pronounced for the sets that performed lower in the single model configurations and shows a consistent pattern where most of the skill improvement was reached when considering ensemble sizes of up to 20 ensemble members. With the exclusion of the S1 setting, the use of ensemble sizes larger than 20 members determines a shift toward underestimation.

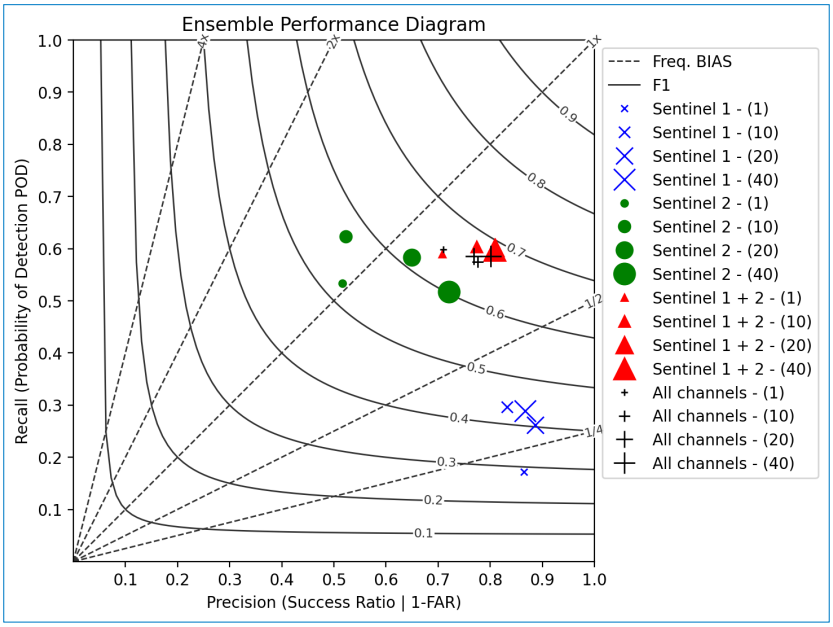


Figure 5: Performance diagram showing performance scores for all four settings on the test set. The model with the best validation score is compared with the ensemble average of the top 10, 20 and 40 models. Horizontal and vertical axes show precision and recall, respectively, while the curved lines represent the isoline of the F1 score. Dashed lines represent the frequency bias: if the marker is on the middle diagonal, the model outputs the same number of landslide pixels as the ground truth, if above the model is overestimating, if below, the model is underestimating. The closer the marker is to the top right corner, the better the classification performance (higher F1 score).

Table 8: Summarizing the ensemble performance diagram showing the F1 score for single model performance and performances of best 10 (F1 ens (10)), 20 (F1 ens (20)), and 40 (F1 ens (40)). The best-performing configuration is shown in bold, underlined with the second best, while in parenthesis, the % of improvement of the ensemble is compared to the single best model.

	Performance on test set (Improvement %)			
Setting	**F1 single**	**F1 ens (10)**	**F1 ens (20)**	**F1 ens (40)**
1) S1	0.29	**0.44 (+34%)**	0.40 (+29%)	0.43 (+34%)
2) S2	0.52	0.57 (+8%)	**0.61 (+15%)**	0.60 (+13 %)
3) S1+S2	0.64	0.68 (+5%)	**0.69 (+7%)**	0.69 (+6 %)
4) all	0.65	0.66 (+2%)	0.66 (+2%)	**0.68 (+4 %)**

Among the diverse configurations, our findings confirm that in ensemble settings, the most promising outcome was achieved by utilising both Sentinel-1 and Sentinel-2 bands (Figure 5 and Table 8). The statistics showed mostly differences in the overall performance using the different data settings when using only Sentinel-1 bands, where the overall F1 model performance based on the validation dataset shows the best score to be 0.44 for an ensemble model size of 10. On the other hand, Setting 2 with only two bands, achieved a high performance of 0.61 and an improvement of the ensemble model of 14.59% when the size is 20 models. For the combination of bands based on Sentinel-1 and Sentinel-2, the highest score achieved is equal to 0.69 for both ensemble sizes of 20 and 40. These observations are shown in Figure 6. The highest score was 0.68 for ensemble size 40 when using all the bands. This can also be observed when comparing the predictions to the reference datasets (Tables 8 and 9).

4.3 Performance Enhancement through Ensemble Models and Optimal Ensemble Size

Notably, the ensemble model employing only Sentinel 2 bands displayed the highest increase in performance using ensemble size 20. This means the viability of developing a monitoring system primarily based on Sentinel 2 bands, enabling the quick identification of potential changes in susceptible regions. Our analysis indicates that the most favourable results were attained when employing an ensemble size between 10 and 20 best-performing single models for Setting 2 and 4 (only S2 and all bands). However, we achieved the best performance for Settings 1 and 3 (only S1 and a combination of S1 and S2) when the ensemble size was between 5 and 10 single models. Beyond this threshold, a discernible shift toward underestimation became evident, suggesting a potential trade-off between ensemble size and accuracy. Figure 6 synthesises our findings, showing that after a certain size, adding more members does not improve, or is detrimental to, the overall skill due to the poor performance of the newly added models. As such, we suggest considering multi-model ensembles formed with up to 20 members.

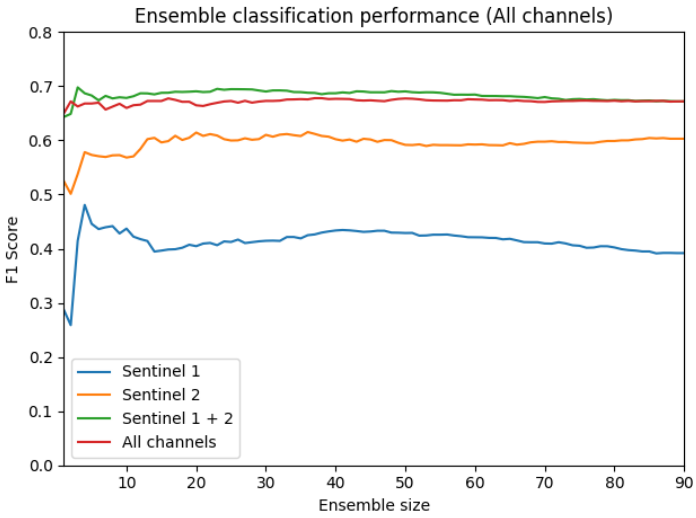
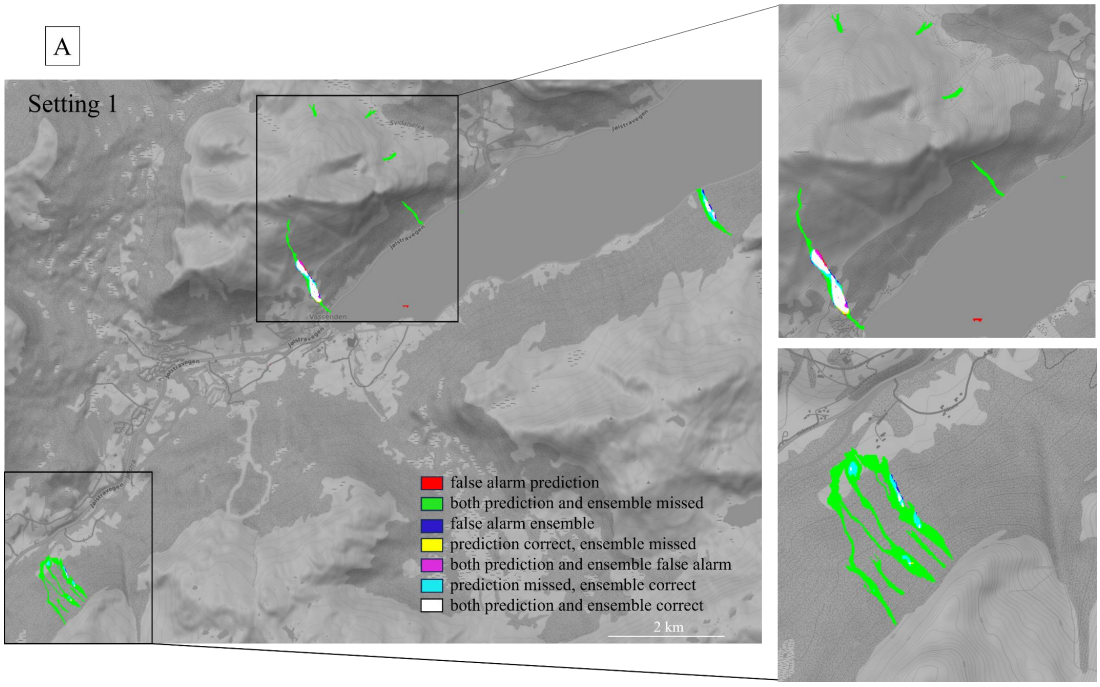


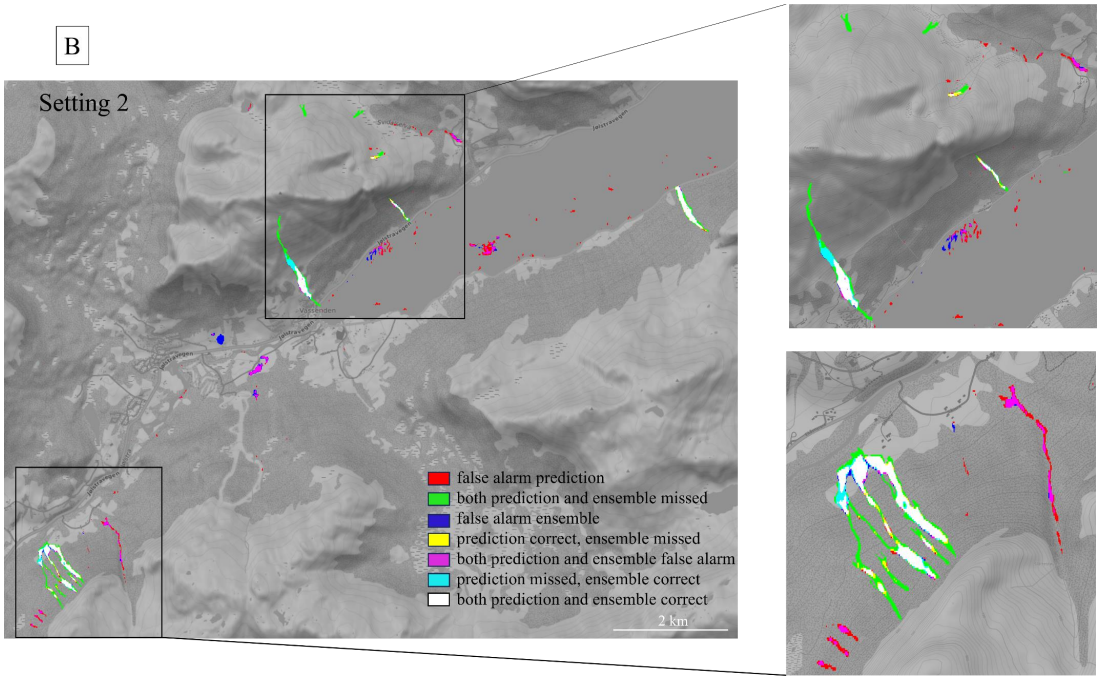
Figure 6: Plot showing the performance of all four settings when using an increasing number of models to build the ensemble one.

Lastly, we present the results as a set of four maps, presenting all four Settings. We created those maps based on the predictions generated through the use of the ensemble size 20 for all Settings. The respective performance information shown in Figure 7A with the colour cyan, where the prediction of a single model is missed. Still, the ensemble model of size 20 predicted the landslide correctly, and white colour represents the areas where both prediction and ensemble model predicted the landslide correctly. For Setting 1 (S1), we can see that only one of the large landslides was mapped correctly (represented in white on the map).

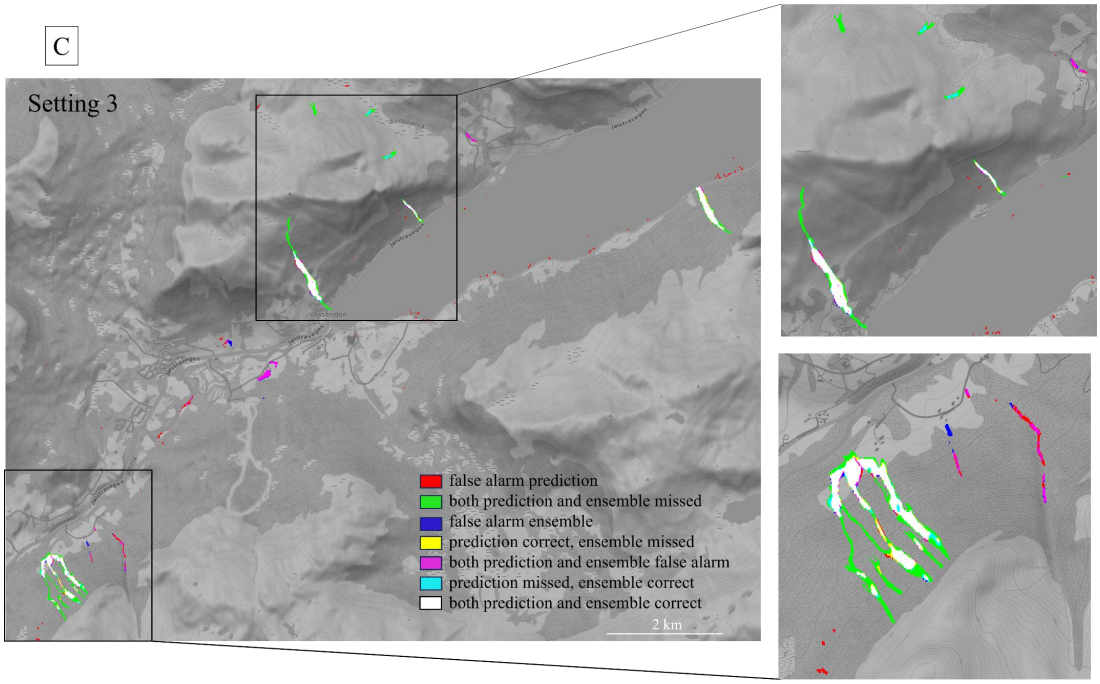


However, when applying Setting 2 (S2) (see Figure 7B), we can clearly see that most of the large landslides were detected by both the single and the ensemble model. The important and interesting fact in this case is that when we are looking at the test area (#5) shown in the Figure 2, many of the landslides represented in red and magenta correspond to the existing landslides. However, these were not used to evaluate the test scores because of the lack of visibility of these landslides within the Sentinel-1 bands. There are also false positives in the water, which do not correspond to landslides (red colour). Viewing the original dNDVI map, one can see that there can be areas of water that have a similar signature to landslides (negative NDVI), and the models here were not

able to capture the difference. Map B confirmed that these landslides are only visible in Sentinel-2 bands, and in fact this setting was able to predict them. Also, the improvement of ensemble models – highlighted with in cyan, can be clearly seen in this figure both on the overview map and close-ups.



Setting 3 (S3) (Figure 7C) confirms that the best performance is achieved with the combination of the Sentinel-1 and Sentinel-2 bands, where both prediction of the single model and ensemble model predicted correctly many more the landslides.



A similar phenomenon as in S2 can be seen in the map presenting Setting 4 (S4). The Sentinel-2 bands were present in training for S3 and S4 (Figure 7C, D). However, much less of red and pink was predicted compared to S2, where Sentinel-2 data stand-alone without any disturbance of other bands.

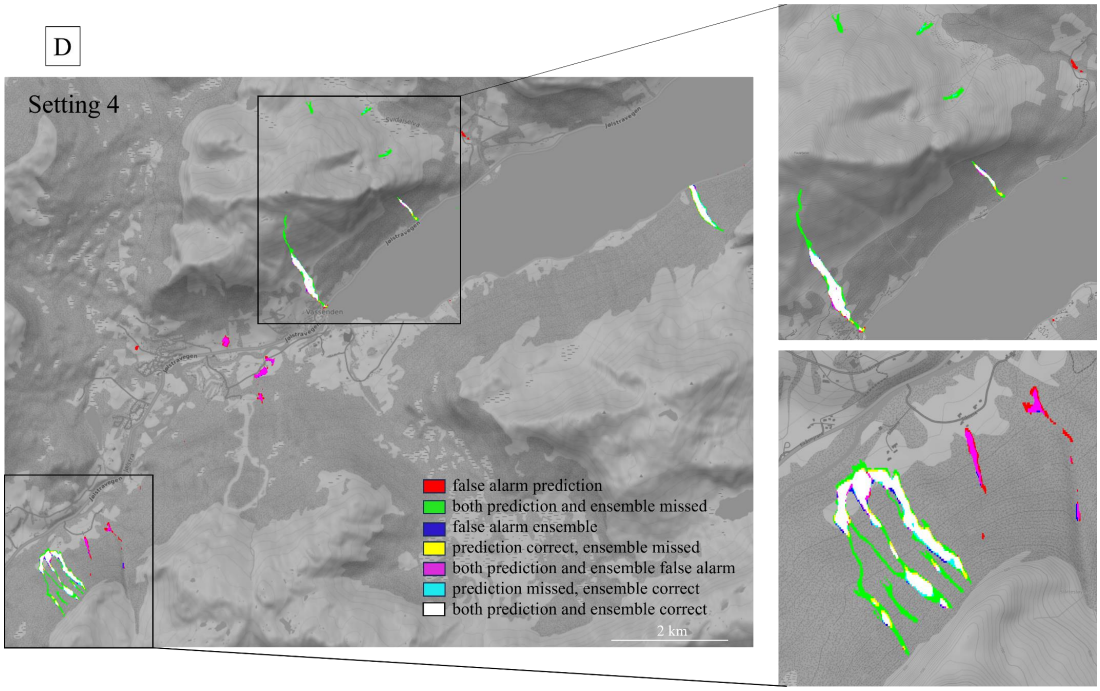


Figure 7: Test area showing the entire test area and two close ups to the landslide areas. For all Settings, the ensemble model is based on the size of 20 ensemble models. A) Setting 1 (Sentinel-1 only); B) Setting 2 (Sentinel-2 only); C) Setting 3 (combination of Sentinel-1 and Sentinel-2) and D) all the bands. (Colours meanings: red = false alarm prediction, green = both prediction and ensemble missed, blue = false alarm ensemble, yellow = prediction correct ensemble missed, magenta = both prediction and ensemble false alarm, cyan = prediction missed, ensemble correct, white = both prediction and ensemble correct)

5. Discussion

This study presents an automated approach for a globally trained deep-learning model for detecting landslides and testing the performance of four different settings based on Sentinel-1, Sentinel-2 and terrain data. The overarching objective of this study is to investigate the functionalities of varying segmentation models and evaluate which are the most suitable for automatic landslide detection. The results could potentially be used to develop a landslide monitoring system able to detect landslide and mitigate their hazards and risks to society. A lack of landslide data is a crucial limitation for the development of mitigation measures, including hazard mapping and early warning systems. The landslide spatial distribution and the local hazard conditions can be easily identified by using an automating detection system based on freely available images. Using a globally trained ensemble model approach helps overcome the lack of available training data. The one used in this study utilises a dataset that was originally designed to include an environmentally diverse range of case studies. The objective is to develop a model that could be used to predict landslides occurrences in a diverse range of environmental conditions. The case studies used for this project belong to a complete collection of different type of landslides occurred across the world over a period that goes from September 2018 to May 2022.

The performances of 10, 20 and 40 ensemble models were compared and presented. All the models' predictions performances were evaluated for recall, precision and F1 score, that set the choice of the ideal predictors (Ganerød et al., 2023; Ji et al., 2020; Liu et al., 2022).

The computational requirements for training the samples and running the model using deep learning techniques are not in the capacity of every computer. We tested the possibilities of a local machine with a highly powerful GPU (Nvidia RTX 4090 GPU (24GB VRAM)) running all the 360 single models and all the ensemble models. In total, four days (corresponding to 96 hours) were necessary to run the 360 models, averaging 16 minutes to train one single model. The training settings and the correct data loading enabled the GPU to its fully capabilities. Afterwards, the interference time to produce the prediction for the ensemble model the size 10, 20 and 40 on the test area was almost instantaneous. Therefore, a powerful GPU on the locally based computer proved to be working efficiently for this study. Furthermore, our study has unveiled novel insights into the effective detection and classification of landslides, significantly contributing to enhance the geohazard management research field. This discussion section presents our findings, the broader context of our research, and the potentials for further directions and work.

5.1 Combining SAR and optical data for a continuous monitoring system

The primary objective of our research was to advance the capabilities of landslide detection through the implementation of ensemble global deep-learning models. Our results indicate that the integration of diverse datasets, particularly the combination of Sentinel-1 and Sentinel-2 bands, achieved the highest

values of F1 and reliability scores in landslide prediction. These results support the findings of a similar study that focuses on illegal logging detection (Luigi and Guzzetti, 2016). The highest performance, achieved by combining Sentinel-1 and Sentinel-2 data, was presented previously, where Sentinel-1 data improved overall landslide prediction. The locally trained deep learning models, such as the Unet, can accurately detect landslides from Sentinel-1 images. This is because these models can distinguish random speckle noise from clusters of pixels that indicate changes in the ground surface (Ganerød et al., 2023). This finding has significant implications into the rapid automatic detection field and can potentially be used as a part of a continuous monitoring system. When used in combination with Sentinel-2, the Sentinel-1 data helps to reduce false positives related to water bodies that occur when using Sentinel-2 data only. Using only Sentinel-2 data returns an overall higher detection rate. However, the key limitation in using it for a monitoring system is the necessary availability of cloud-free images, which can take several months in the worst cases. Another challenge that these models will face is related to the ability of distinguishing between signals of vegetation loss unrelated to landslides and those caused by landslides (Lindsay et al., 2023). This issue has persisted in landslide detection using optical data, as mentioned by Prakash et al. (2021). For a continuous monitoring system, we recommend to further develop a SAR-based detection model, which can be updated when cloud-free optical images become available.

5.2 Sentinel-2 for Landslide Inventory Production

One of the objectives of this study was to find the minimal amount of band with the highest performance. Therefore, we further evaluated combination of only Sentinel-2 bands in Setting-2 of the pre- from the post-event composite images. This approach can potentially be used for continuous landslide monitoring or as a tool for producing landslide inventories after a known rain event in the affected area (Lindsay et al., 2020). We believe that the dNDVI method is helpful in collecting more information about the size and location of landslides. However, it is not advised to use it in areas that covered by snow due to the lack of vegetation and changes in snow cover, which can lead to wrong predictions. There is a need for more investigation in differentiating landslides from changes in the snow. This suggests that inventories of landslides could be improved, which is important for defining thresholds and analysing the risks associated with them. (Ghorbanzadeh et al., 2019; Lindsay et al., 2020)

5.3 Optimal Ensemble Size

The testing of ensemble size's impact on performance revealed an intricate trade-off between the performance and the computational efficiency. While larger ensemble sizes demonstrated improved performance up to a certain point, a noticeable shift towards underestimation occurred as the ensemble size exceeded 20 models. These findings highlight the significance of carefully optimising ensemble size to strike a balance between precision and resource consumption.

5.4 Implications for Geohazard Management and Future Directions

Building on the current study, several future research can emerge. Exploring the integration of other remote sensing data sources and fine-tuning ensemble parameters could lead to even more accurate prediction models. Additionally, investigating the scalability and applicability of our findings to different geographic regions could further validate the robustness of the proposed approach. Based on our overall findings, we can recommend the implementation of a single setting and model architecture combination. Specifically, we suggest training the Unet++ with S1+S2 bands, given its consistent performance and robustness concerning parameter changes (loss function, learning rate).

6. Conclusions

Developments in computing, deep learning algorithms, and the increased availability of high-quality and free satellite data can potentially automate many mapping problems in Earth sciences. Choosing the most efficient machine learning algorithm is necessary to reduce inconsistencies in landslide detection. Usually, the main objective is to identify the optimal model based on its predictive capabilities. The primary aim of this study was to estimate the optimal model with maximum predictive ability, considering the limited availability of historical landslide data. We applied four different settings, all based on either Sentinel-1, Sentinel-2 and terrain model bands and their combinations, with nine segmentation models, two learning rates and five loss functions. In total, we generated 360 prediction models. The proposed framework is based on Python and uses PyTorch utilities. We used 21 case studies globally spread around the world. Our deep learning approach proved to delineate landslides with high performance in all the mentioned cases. However, the best-performing setting was based on a combination of Sentinel-1 and Sentinel-2 bands with the F1 score equal to 0.69. We achieved a promising performance of 0.6 when using Sentinel-2 bands only and applying the change in vegetation index (pre-and post-images) with the size of ensemble model 20. We observed that the Sentinel-2 only based ensemble model could delineate many more landslides and pick up those impossible to map using Sentinel-1 bands only. However, the main drawbacks of this approach were that in the test case results, it wrongly predicted landslides in the water, and it is necessary to wait for cloud-free images, which can limit the use for rapid detection. These findings help to inform the development of continuous monitoring systems and approaches for producing precise landslide inventories after widespread triggering events.

To summarise, our study underscores the potential of ensemble global deep-learning models in automating landslide detection. The integration of various data sources and the identification of context-specific applicability presents the potential for innovative geohazard management strategies. The outcomes also

emphasise the significance of ensemble size, with an optimal range of 10 to 20 models providing the best compromise between calculated scores and computational efficiency for Settings 2 and 4 (only Sentinel-2 and all bands). However, we achieved the best performance for Settings 1 and 3 (only Sentinel-1 and a combination of Ssentinel-1 and Sentinel-2) when the ensemble size was between 5 and 10 single models. The insights are working towards developing robust early warning systems to minimise landslide hazards and their impact on society.

Code availability section

The source code is available for download here: GitHub

Program language: Python

Software required: data preparation (ArcGIS Pro/QGIS)

References

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in Al-Obeidat, F., Feltrin, L., Marir, F., 2016. Cloud-based Lineament Extraction of Topographic Lineaments from NASA Shuttle Radar Topography Mission Data. Procedia Computer Science 83, 1250–1255. https://doi.org/10.1016/j.procs.2016.04.260

Bai, L., Li, W., Xu, Q., Peng, W., Chen, K., Duan, Z., Lu, H., 2022. Multispectral U-Net : A Semantic Segmentation Model Using Multispectral Bands Fusion Mechanism for Landslide Detection 73–76.

Band, S.S., Janizadeh, S., Pal, S.C., Saha, A., Chakrabortty, R., Shokri, M., Mosavi, A., 2020. Novel ensemble approach of deep learning neural network (Dlnn) model and particle swarm optimization (pso) algorithm for prediction of gully erosion susceptibility. Sensors (Switzerland) 20, 1–28. https://doi.org/10.3390/s20195609

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87, 180–191. https://doi.org/10.1016/J.ISPRSJPRS.2013.09.014

Blum, A., Wang, L., 2011. Machine Learning Theory, Sci 1541-1672.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification And Regression Trees. Routledge. https://doi.org/10.1201/9781315139470

Chaurasia, A., Culurciello, E., 2018. LinkNet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE Visual Communications and Image Processing, VCIP 2017 2018-Janua, 1–4. https://doi.org/10.1109/VCIP.2017.8305148

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11211 LNCS, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

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

Dargan, S., Kumar, M., Ayyagari, M.R., Kumar, G., 2020. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. Archives of Computational Methods in Engineering 27, 1071–1092. https://doi.org/10.1007/s11831-019-09344-w

Fan, T., Wang, G., Li, Y., Wang, H., 2020. Ma-net: A multi-scale attention network for liver and tumor segmentation. IEEE Access 8, 179656–179665. https://doi.org/10.1109/ACCESS.2020.3025372

Ganerød, A.J., Lindsay, E., Fredin, O., Myrvoll, T.A., Nordal, S., Rød, J.K., 2023. Globally vs. Locally Trained Machine Learning Models for Landslide Detection: A Case Study of a Glacial Landscape. Remote Sensing 15. https://doi.org/10.3390/rs15040895

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing 11. https://doi.org/10.3390/rs11020196

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.-T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Reviews 112, 42–66. https://doi.org/10.1016/J.EARSCIREV.2012.02.001

Hong, H., Liu, J., Zhu, A.X., 2020. Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Science of the Total Environment 718, 137231. https://doi.org/10.1016/j.scitotenv.2020.137231

Jadon, S., 2020. A survey of loss functions for semantic segmentation. 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020. https://doi.org/10.1109/CIBCB48159.2020.9277638

Ji, S., Yu, D., Shen, C., Li, W., Xu, Q., 2020. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides 17, 1337–1352. https://doi.org/10.1007/s10346-020-01353-2

Jiang, N., Chen, J., Zhou, R.G., Wu, C., Chen, H., Zheng, J., Wan, T., 2020. PAN: Pipeline assisted neural networks model for data-to-text generation in social internet of things. Information Sciences 530, 167–179. https://doi.org/10.1016/j.ins.2020.03.080

Kirschbaum, D.B., Adler, R., Hong, Y., Hill, S., Lerner-Lam, A., 2010. A global landslide catalog for hazard applications: method, results, and limitations. Natural Hazards 52, 561–575. https://doi.org/10.1007/s11069-009-9401-4

Li, X., Lai, T., Wang, S., Chen, Q., Yang, C., Chen, R., 2019. Weighted feature pyramid networks for object detection. Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking,

ISPA/BDCloud/SustainCom/SocialCom 2019 1500–1504. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217

Lin, Q., Steger, S., Pittore, M., Zhang, J., Wang, L., Jiang, T., Wang, Y., 2022. Evaluation of potential changes in landslide susceptibility and landslide occurrence frequency in China under climate change. Science of the Total Environment 850, 158049. https://doi.org/10.1016/j.scitotenv.2022.158049

Lin, W.T., Chou, W.C., Lin, C.Y., Huang, P.H., Tsai, J.S., 2005. Vegetation recovery monitoring and assessment at landslides caused by earthquake in Central Taiwan. Forest Ecology and Management. https://doi.org/10.1016/j.foreco.2005.02.026

Lindsay, E., Frauenfelder, R., Rüther, D., Nava, L., Rubensdotter, L., Strout, J., Nordal, S., 2022. Multi-Temporal Satellite Image Composites in Google Earth Engine for Improved Landslide Visibility: A Case Study of a Glacial Landscape. Remote Sensing 14. https://doi.org/10.3390/rs14102301

Lindsay, E., Garnerød, A.J., Devoli, G., Reiche, J., Nordal, S., Frauenfelder, R., Tokle, L.-C., 2023. Understanding Landslide Expression in SAR Backscatter Data: A Global Study 1–32. https://doi.org/10.20944/preprints202302.0390.v1

Lindsay, E., Rouault, C., Fjeld, M., Nordal, S., 2020. Potential of dNDVI-method for landslide detection in Norway. https://doi.org/https://doi.org/10.5194/egusphere-egu2020-4328

Liu, B., Zhang, B., Feng, H., Wu, S., Yang, J., Zou, Y., Siddique, K.H.M., 2022. Ephemeral gully recognition and accuracy evaluation using deep learning in the hilly and gully region of the Loess Plateau in China. International Soil and Water Conservation Research 10, 371–381. https://doi.org/10.1016/j.iswcr.2021.10.004

Liu, Z., Gilbert, G., Cepeda, J.M., Lysdahl, A.O.K., Piciullo, L., Hefre, H., Lacasse, S., 2021. Modelling of shallow landslides with machine learning algorithms. Geoscience Frontiers 12, 385–393. https://doi.org/10.1016/j.gsf.2020.04.014

Luigi, S., Guzzetti, F., 2016. Earth-Science Reviews Landslides in a changing climate. Earth Science Reviews 162, 227–252. https://doi.org/10.1016/j.earscirev.2016.08.011

Lv, L., Chen, T., Dou, J., Plaza, A., 2022. A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. International Journal of Applied Earth Observation and Geoinformation 108, 102713. https://doi.org/10.1016/j.jag.2022.102713

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image Segmentation Using Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Mondini, A.C., Guzzetti, F., Chang, K.-T., Monserrat, O., Martha, T.R., Manconi, A., 2021. Landslide failures detection and mapping using Synthetic Aperture Radar: Past, present and future. Earth-Science Reviews 216, 103574. https://doi.org/10.1016/j.earscirev.2021.103574

Nava, L., Bhuyan, K., Meena, S.R., Monserrat, O., Catani, F., 2022. Assessment of deep learning based landslide detection and mapping performances with backscatter SAR data. https://doi.org/10.5194/egusphere-egu22-10072

O'Grady, D., Leblanc, M., Gillieson, D., 2013. Relationship of local incidence angle with satellite radar backscatter for different surface conditions. International Journal of Applied Earth Observation and Geoinformation 24, 42–53. https://doi.org/10.1016/j.jag.2013.02.005

Piciullo, L., Calvello, M., Cepeda, J.M., 2018. Territorial early warning systems for rainfall-induced landslides. Earth-Science Reviews. https://doi.org/10.1016/j.earscirev.2018.02.013

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., Armston, J., Dubayah, R., Blair, J.B., Hofton, M., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment 253, 112165. https://doi.org/10.1016/j.rse.2020.112165

Prakash, N., Manconi, A., Loew, S., 2021. A new strategy to map landslides with a generalized convolutional neural network. Scientific Reports 11, 1–15. https://doi.org/10.1038/s41598-021-89015-8

Roebber, P.J., 2009. Visualizing multiple measures of forecast quality. Weather and Forecasting 24, 601–608. https://doi.org/10.1175/2008WAF2222159.1

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Saadatkhah, N., Kassim, A., Lee, L.M., 2014. Qualitative and quantitative landslide susceptibility assessments in Hulu Kelang area, Malaysia. Electronic Journal of Geotechnical Engineering 19 C, 545–563.

Saha, S., Arabameri, A., Saha, A., Blaschke, T., Ngo, P.T.T., Nhu, V.H., Band, S.S., 2021. Prediction of landslide susceptibility in Rudraprayag, India using novel ensemble of conditional probability and boosted regression tree-based on cross-validation method. Science of the Total Environment 764, 142928. https://doi.org/10.1016/j.scitotenv.2020.142928

Setargie, T.A., Tsunekawa, A., Haregeweyn, N., Tsubo, M., Fenta, A.A., Berihun, M.L., Sultan, D., Yibeltal, M., Ebabu, K., Nzioki, B., Meshesha, T.M., 2023. Random Forest–based gully erosion susceptibility assessment across different agro-ecologies of the Upper Blue Nile basin, Ethiopia. Geomorphology 431. https://doi.org/10.1016/j.geomorph.2023.108671

Shano, L., Raghuvanshi, T.K., Meten, M., 2020. Landslide susceptibility evaluation and hazard zonation techniques – a review. Geoenvironmental Disasters 7, 18. https://doi.org/10.1186/s40677-020-00152-0

Tehrani, F.S., Calvello, M., Liu, Z., Zhang, L., Lacasse, S., 2022. Machine learning and landslide studies: recent advances and applications. Natural Hazards 114, 1197–1245. https://doi.org/10.1007/s11069-022-05423-7

Tehrani, F.S., Santinelli, G., Herrera Herrera, M., 2021. Multi-Regional landslide detection using combined unsupervised and supervised machine learning. Geomatics, Natural Hazards and Risk 12, 1015–1038. https://doi.org/10.1080/19475705.2021.1912196

Thomas G. Dietterich, 1997. Machine learning research. AI Magazine 18, 97–136. https://doi.org/10.1145/1056743.1056744

Tong, Z. ling, Guan, Q. tao, Arabameri, A., Loche, M., Scaringi, G., 2023. Application of novel ensemble models to improve landslide susceptibility mapping reliability. Bulletin of Engineering Geology and the Environment 82, 309. https://doi.org/10.1007/s10064-023-03328-8

Vollrath, A., Mullissa, A., Reiche, J., 2020. Angular-based radiometric slope correction for Sentinel-1 on google earth engine. Remote Sensing 12, 1–14. https://doi.org/10.3390/rs12111867

Wang, G., Hao, J., Ma, J., Jiang, H., 2011. A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications 38, 223–230. https://doi.org/10.1016/j.eswa.2010.06.048

Wei, R., Ye, C., Sui, T., Ge, Y., Li, Y., Li, J., 2022. Combining spatial response features and machine learning classifiers for landslide susceptibility mapping. International Journal of Applied Earth Observation and Geoinformation 107, 102681. https://doi.org/10.1016/j.jag.2022.102681

Yang, Z., Li, L., Xu, X., Kailkhura, B., Xie, T., Li, B., 2022. on the Certified Robustness for Ensemble Models and Beyond. ICLR 2022 - 10th International Conference on Learning Representations.

Youssef, A.M., Pourghasemi, H.R., 2021. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. Geoscience Frontiers 12, 639–655. https://doi.org/10.1016/j.gsf.2020.05.010

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Janua, 6230–6239. https://doi.org/10.1109/CVPR.2017.660

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation BT - Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Miccai 11045, 3–11. https://doi.org/10.1007/978-3-030-00889-5

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 Geoscience and Remote Sensing Magazine 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307

Article VI.

Where are the ravines? A Case Study of Gully Landscapes in Norway Using Deep Learning

Alexandra Jarna Ganerød[1,2†], Mikis van Boeckel[2*†], Inger-Lise Solberg[2]

1 Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway; alexandra.jarna@ntnu.no
2 Geological Survey of Norway (NGU), 7040 Trondheim, Norway
† These authors contributed equally to this work

Where are the ravines? A Case Study of Gully Landscapes in Norway Using Deep Learning

Alexandra Jarna Ganerød[1,2], Mikis van Boeckel[1], and Inger-Lise Solberg[1]

[1]Geological Survey of Norway (NGU), 7040 Trondheim, Norway
[2]Department of Geography, Norwegian University of Science and Technology, 7049 Trondheim, Norway

August 25, 2023

Abstract

Gullies and ravines are common landforms in raised marine fine-grained deposits in Norway. Gullies in marine clay are significant landforms indicative of soil erosion, natural hazards and are of high conservation value. As a result of the substantial impact of human intervention over the past century, marine clay gullies are now red-listed. To monitor the condition of these landforms we need to improve our understanding of their spatial extent, complexity, and morphology. We explore the applicability of automated approaches that uses a methodology of combining deep learning (DL), fully convolutional neural networks (FCNN), and a U-Net model with ArcPy libraries and ground truth data to derive a high-resolution map of gullies in raised marine fine-grained deposits. Predictors used comprise solely terrain derivatives to broaden the usage of the pre-trained model to other regions. Our best model achieved a precision score of 0.82 and a recall of 0.75. We find that our pre-trained model can successfully predict gullies in blind-test areas. The model performs better in regions with similar geological settings, scoring a length-weighted overlap of >72% with reference datasets. We also find that the model's applicability increases when we post-process the predictions by eliminating noise, especially by using the predictions derived from ensembled models. We, therefore, conclude that the pre-trained models can effectively be used to supplement the geomorphological mapping of marine clay gullies in Norway. The outcome of this research contributes towards mapping the spatial extent and condition of red-listed landforms in Norway, as well as the development of monitoring systems for future landscape change.

Keywords: gullies, ravines, landforms, marine clay, deep learning, U-net

1. INTRODUCTION

During and after the rapid deglaciation of the Scandinavian ice sheet, the retreating outlet glaciers fed fjords and sea inlets with large quantities of fine-grained glaciomarine and marine sediments. As a result of the ongoing post-glacial isostatic rebound after deglaciation, these marine deposits, commonly consisting of clay stratified with silt and sand, gradually emerged above sea level (Reite *et al.* ,**1999** ; Hansen *et al.* , **2007**). During regression, these emerging flats of marine deposits tend to develop into a characteristic landscape called marine clay landscapes (Erikstad,**1992**) or gully landscapes (Bergqvist, **1990** ; Hamre*et al.* , **2021**). The gully landscapes in Norway have a rough topography mainly controlled by gully erosion, incision by rivers, and landsliding in quick clay.

Characteristics of marine clay landscapes are gullies and ravines. These narrow, often v- or u-shaped landforms have steep sides and head scarps incised into unconsolidated material (Higgins and Coates,**1990**). Erosion along these gullies involves the removal of sediment due to concentrated flow converging towards lower points of the watershed and is often associated with groundwater seepage and shallow sliding (Bridge, **2003**). With the narrow channels' increasing size, depth, and branching, gullies gradually transition into ravines. In this paper, we do not differentiate between gullies and ravines. We use both terms interchangeably. Gullies and ravines may have permanent or intermittent flowing streams which control the local drainage network and influence the direction of groundwater flow.

Gullies in marine fine-grained deposits are of high conservation value due to the marine clay's high nutrient content and moisture-holding capacity (Erikstad, **1992** ; Hamre *et al.* , **2021**). Networks of the gully and ravine systems are essential wildlife corridors (Blindheim and Abel, **2002**) and facilitate large diversity of habitat types (Jansson, U. & Høitomt, **2013** ; Blindheim *et al.* , **2018**). Agricultural policies of levelling and ploughing, along with urban development, caused the gully landscape to be subjected to substantial landscape change over the past century (Erikstad, **1992** ; Hamre *et al.* , **2021**), which resulted in the red-listing of the landform marine clay gully (Erikstad *et al.* , **2018**).

As part of developing a nationwide conservation plan for preserving the gully landscape, there is a need to map and monitor the change and condition of marine clay gullies. Moreover, since many quick-clay landslides are initiated in ravines, an overview of marine clay gullies' spatial extent and development contributes to improving hazard assessment. The establishment of monitoring systems with repeated mapping and comparison of time series, areas of comprehensive vertical erosion and migration of ravines may attract attention to hazard mitigation (Ryan *et al.* , **2022**) and soil erosion (Kværnø*et al.* , **2020** ; Barneveld *et al.* , **2022**).

Earlier studies on the delineation and condition of marine clay gullies in Norway have relied on manual mapping of aerial images (Hamre *et al.* , **2021**) or high-resolution terrain models and surficial geological maps (Christoffersen *et al.* , **2021** ; van Boeckel *et al.* , **2022**). Approaches to automated delineations of gullies outside of Norway have also developed rapidly but have mostly focused on gully erosion susceptibility and comparison of different machine-learning algorithms (Gayen *et al.* ,**2019** ; Arabameri *et al.* , **2020** , **2022** ; Band *et al.* , **2020** ; Chen *et al.* , **2021** ; Mohebzadeh *et al.* , **2022** ; Chuma *et al.* ,**2023** ; Setargie *et al.* , **2023**). Setargie *et al.* (**2023**) used a Random Forest-based approach in Ethiopia, combining 74 manually mapped gullies with 20 predictors. The predictors used in this study included: elevation, slope, Topographic Positioning Index (TPI), Topographic Ruggedness Index (TRI), profile curvature, convergence index, soil type, and distance from streams. Band et al. (**2020**) applied a deep learning approach using 132 gully erosion locations with 13 independent variables, comprising: lithology, rainfall, Stream Power Index (SPI), and Topographic Wetness Index (TWI) and terrain derivatives, similar to the study from Setargie *et al.* (**2023**). Liu *et al.* (**2022**) also tested the applicability of automated approaches in new blind-test areas by applying U-net for image segmentation using satellite (QuickBird-2, Pleiades: worldView-03) with obtained UAV image data (worldView2 and PHANTOM 4 RTK). Here, the authors successfully used vector lines of gullies as ground truth data but provided limited information on the accuracy of delineation of the depressions. On the other hand, Arabameri*et al.* (**2021**) and Roy & Saha (**2022**) applied ensemble models with conventional machine learning algorithms. They used topographical and hydrological gully erosion conditioning factors such as rainfall, distance from the river, surface runoff, length of overland flow, and topographical wetness index. Although the studies mentioned above succeeded in identifying and delineating the gullies, little is known about automatic differentiations of gullies impacted by human interventions, such as agricultural levelling, and how these predictions can further be used for geomorphological mapping routines.

In this study, we address these issues by (1) exploring automated differentiation of intact and impacted ravines and gullies, (2) choosing the minimal amount and best suitable predictors for delineating ravines and gullies, (3) testing the applicability of pre-trained models to blind-test areas on similar and different geological settings, and (4) discussing the usage of deep learning models for geomorphological mapping. We do so by assessing the automatic differentiation of gullies with high precision, using U-net architecture and Fully Convolutional Neural Networks (FCNN). The selected study areas are in Romerike and in Trøndelag (Figure 1 and 2), where intact and impacted ravine systems are found in different geological and geomorphological settings. We evaluate the predictions statistically by (a) calculating precision, recall, and F1-score with ground truth data and (b) by comparing the predictions to a reference dataset comprising vector lines (Christoffersen *et al.* , **2021** ; van Boeckel *et al.* , **2022**).

2. STUDY AREAS

The study areas represent different marine clay landscapes in two regions: in Romerike, South-East Norway, divided into Romerike North and Romerike South (Figure 1b), and in Mid-Norway, comprising of Byneset, Orkdal, and Stadsbygd (Figure 2). All study areas are located below the marine limit, representing a modelled elevation of the highest relative sea level after deglaciation (Høgaas *et al.* , **2022**). The marine limit varies throughout the country; in South-East Norway, the marine limit reaches up to 220 m a.s.l., and in Mid-Norway, up to 190 m a.s.l. (NGU, **2023**). The vast majority of the raised marine fine-grained deposits, hosting the marine clay gullies, are found below this limit.

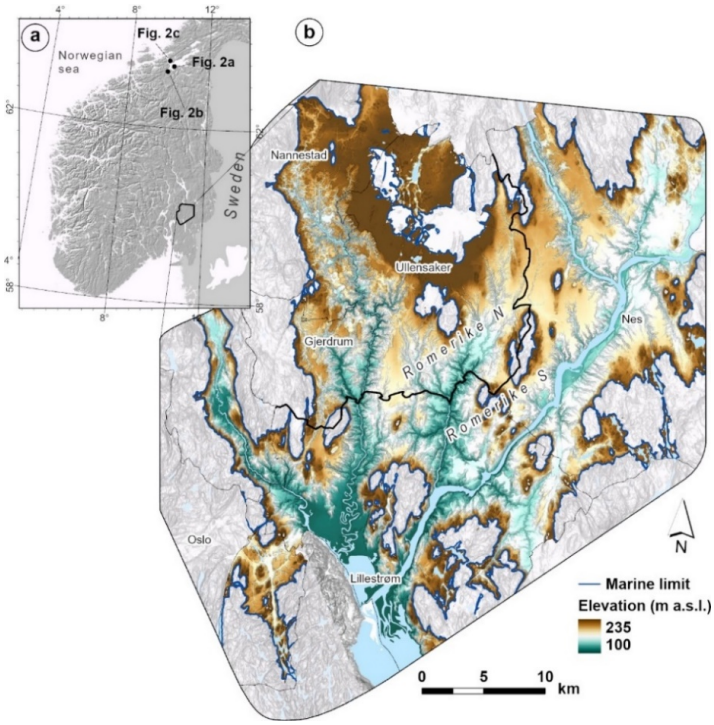
All the study areas comprise large raised marine clay deposits, reflecting a near horizontal surface of the old seabed before the inception of gully and river erosion. This near horizontal surface can be regarded as a reference surface for estimating erosion depth in gullies and differs in elevation above sea level for each study area. The base level of erosion along gullies is therefore located at different elevations, but is also controlled by bedrock, rivers, or the sea.

Romerike and Byneset are characterized by bedrock hills protruding through large flats of fine-grained marine deposits (Figures 1b and 2a). These deposits have widely been incised by large, often dendritic networks of ravines or substantial quick-clay landsliding. The extensive networks of ravines vary in length from nearly a kilometre up to 10 km and depths between a few meters up to 40 meters.

Romerike has two large rivers, Glomma and Vorma. Along these river systems, adjacent to the fluvial bars and plains, the ravines are relatively parallel and straight, oriented perpendicular to the main river, with less dendritic branching. Romerike also has four other smaller river systems, and large ravine networks are connected to three of them.

Orkdal is a relatively narrow fjord valley with steep bedrock sides. Along the valley runs a large meandering river, Orkla, with fluvial plains and terraces in the valley bottom. The marine clay deposits are exposed at higher elevations in the valley sides but lies stratigraphically beneath the fluvial deposits (Figure 2b). The gullies in Orkdal are steep, closely spaced, and oriented perpendicular to the main river. Here, the gullies are relatively short, with only a few branched into networks and longer than one kilometer.

Stadsbygd has one mainstream in relatively flat-lying marine deposits confined by bedrock and with an outlet in the sea. The stream has only a few attached short and shallow gullies with depths of less than 10 m (Figure 2b).



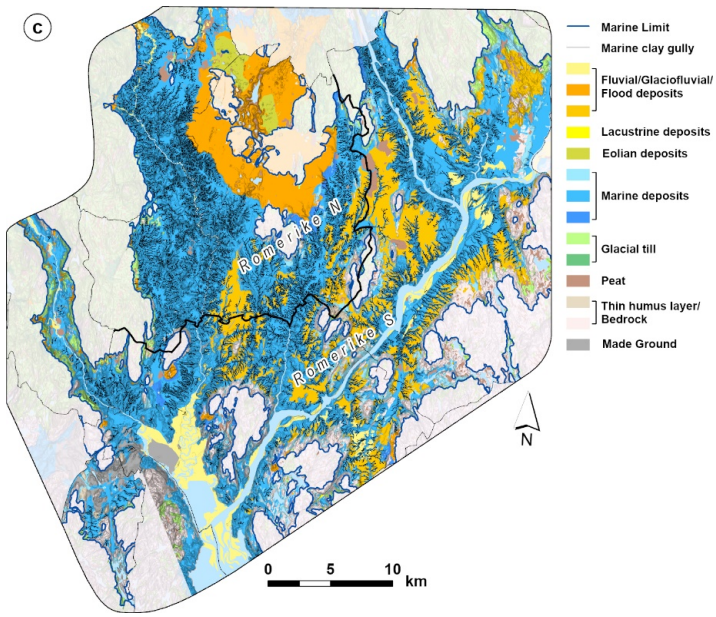


Figure 1. (a) Overview of the two study areas south of Norway: Trøndelag and Romerike. (b) Elevation below the marine limit in Romerike, in shaded relief, using a slope map from Kartverket. (c) Surficial geology map of Romerike (NGU, **2023**). The blue line marks the marine limit, and the thick black line marks the division between Romerike North and Romerike South. The marine clay gullies are mapped as vector lines by van Boeckel et al. (**2022**).

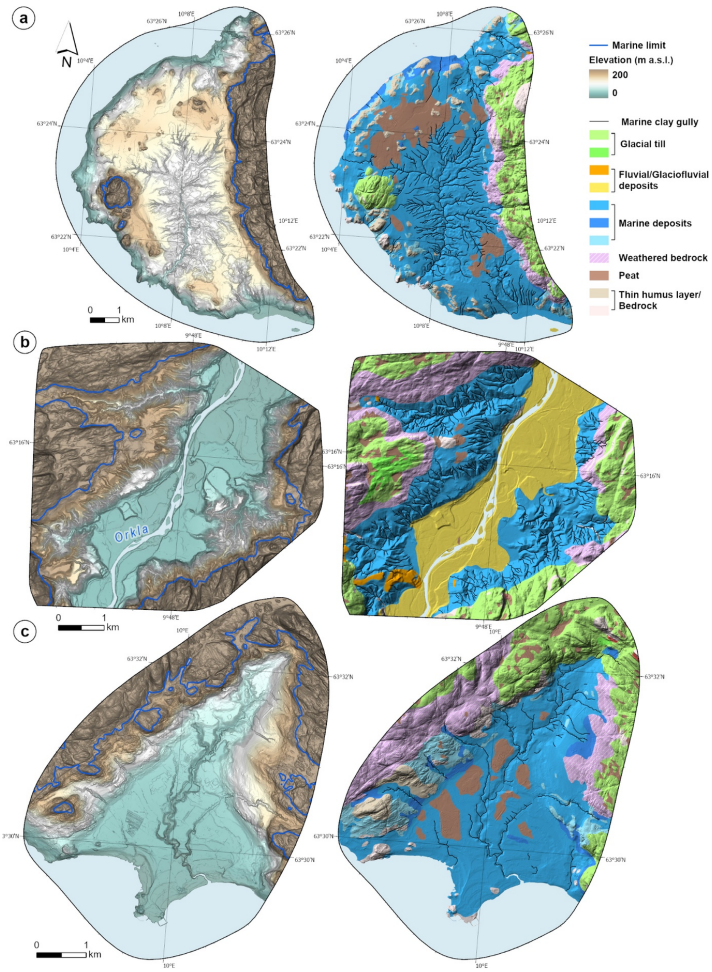


Figure 2. Study areas in Trøndelag: (a) Byneset, (b) Orkdal, and (c) Stadsbygd. The panels to the left show the elevation in shaded relief using a slope map from Kartverket. The blue line indicates the marine limit. The panels to the right show surficial geology maps (NGU,**2023**) and the references dataset of marine clay gullies as black lines (NGU 2023; Chistoffersen et al., 2021).

3. METHODS

We aimed to train our models to identify and delineate gullies using a minimum number of predictors variables, comprising solely terrain derivatives, and ground truth data from Romerike S. We then assessed the performance of our models to sampled ground truth data for evaluation and compared our predictions to reference datasets, see Figure 3 and Table 1.

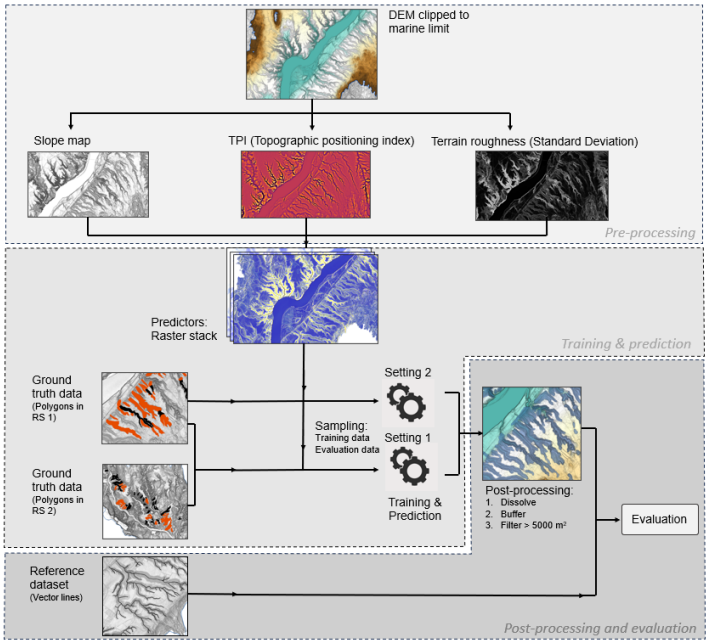


Figure 3. The terrain derivatives: Slope, Topographic positioning index (TPI), and Terrain Roughness (TR), are derived from a clipped DEM to the marine limit and stacked into a composite band. The predictor raster stack was then sampled with ground truth data from both RS 1 and RS 2 (Setting 1) and ground truth data from only RS 1 (Setting 2) to train and evaluate the models Setting 1 and Setting 2 (Figure 1), respectively. The model's product was then post-processed to better match manually mapped ravines and to remove noise.

Table 1. Input data used in this study comprising terrain derivates, ground truth and reference data.

Input data	Source	Data type	Study area	Purpose
Terrain derivatives	Slope, Topographic positioning index (TPI), and Terrain Roughness (TR), derived from DEM	Raster	All	Predictor variables used for training the model and predicting in all study areas.
Ground truth data	Manually mapped sharp and smooth ravines	Vector polygons	Romerike South	Used for training the model.
Reference data	Manually mapped ravines	Vector lines	All	Used to evaluate the performance of the predictions in blind-test areas.

'

3.1. Data input and preparation

3.1.1. Predictor variables – terrain derivatives from DEM

We calculated the terrain derivatives from a high-resolution Digital Elevation Model (DEM), derived from Light Detection And Ranging (LiDAR), accessed January 2023 at https://hoydedata.no. The DEM is clipped to the area below the marine limit (Høgaas *et al.* , **2022**) and has a spatial resolution of the DEM 1 m with a vertical accuracy of ˜0.1 m (Terratec, **2022**). The terrain derivatives comprised of Slope, Topographic Positioning Index (TPI), and Terrain Roughness (TR) were stacked into one composite band; accordingly, see Figure 3. We used a moving window of 100 m for calculating TPI with the tool "DiffFromMeanElev" using WhiteboxTools (Lindsay, **2014**) and a moving window of 5 m for calculating the TR using the standard deviation of the surrounding the topography (Grohmann *et al.* , **2009**).

Initial test runs also included additional categorical data, such as land-use maps (AR5) (Ahlstrøm et al., **2019**), surficial deposit maps in 1:50 000 (NGU, **2023**), and continuous elevation data. Due to unsuccessful predictions of gullies of the pre-trained model for the blind-test areas, and because the scope of this study was to apply the pre-trained model to other regions, we decided to drop these predictor variables for further analysis and will not be presented in our results section.

3.1.2. Data preparation - training and evaluation

Marine clay gullies, used for training the model, were digitized manually from LiDAR data (Light Detection and Ranging) and orthophotos provided by Kartverket (the Norwegian Mapping Authority), field surveys, and existing Quaternary geological maps (NGU, **2023**). The landforms were mapped on a scale of 1: 20 000 or better, with a minimum size of 2 500 m^2. The shape of many individual gullies, or sections of more extensive ravine networks, are often partly impacted by human activity, often as a result of agricultural levelling and ploughing, and filling of construction material (Erikstad,**1992** ; van Boeckel *et al.* , **2022**). Since the morphology of impacted, often smoothened, ravines differ to such an extent from the unimpacted, steep, and V-shaped ravines, we divided the ground truth dataset into two categories: sharp and smoothened ravines. Because the morphology of ravines also differs from near parallel ravines to dendritic ravine systems, we depicted subsets of training data that covered both types of ravine networks in the regions RS 1 and RS 2 in Romerike South (Figures 3 and 4). In order to test the amount of training data needed to predict ravines, we used two different settings: Setting 1, which included training data of whole Romerike South, including ground truth data from RS 1 and RS 2, and Setting 2, which only used training data from RS 1 (Figure 3 and 4). In total, 186 smooth and 147 sharp (Table 3) ravines randomly spread with different sizes were used for training.

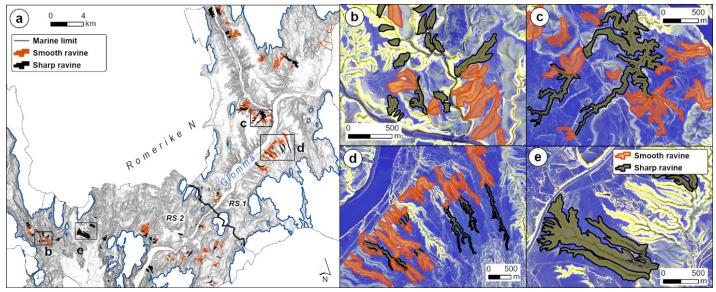


Figure 4. (a) Overview of ground truth data used for training the model, where the thick black line indicates the division between the RS 1 and RS 2 used for sampling the training data. Due to their morphological differences, ravines depicted for training are classified as smooth (orange) and sharp (black). (b-e) Detailed areas of composite bands (RGB = TR, Slope, TPI) with ground truth data of both RS 1 and RS 2. (b and c) Subsets of dendritic ravine systems. (d and e) Subsets of near-parallel ravines perpendicular to the incised river systems.

3.2. Method and evaluation

3.2.1. Semi-automated mapping

We used Convolutional Neural Network (CNN) to recognize ravines using training data consisting of predictor variables (slope, TPI, and TR) and ground truth data. Models based on CNN can distinguish patterns by applying a series of mathematical operations and perform exceptionally well in applications that deal with image data (Chen *et al.* ,**2019** ; Zaidi *et al.* , **2022**). Convolutions are matrix calculations based on a moving window, usually using 3x3 cells, to compile geospatial information into classified tiles (Albawi *et al.* , **2018**). The spatial dimension of the classified tiles is crucial; therefore, we apply U-net (Ronneberger *et al.* ,**2015**) and CNN architecture for semantic segmentation and pixel-based classification (Zhang *et al.* , **2016** ; Prakash*et al.* , **2021**). U-Net is known for being fast, effective, and precise in segmentation, recognizing objects based on local information in the ground truth (Leng *et al.* , **2019**). This approach requires two types of data sources for training: ground truth data, with vector-based manually mapped features aimed to be predicted, and predictor layers used for recognizing these features (Nodjoumi *et al.* , **2023**). To train robust models (Shelhamer *et al.* , **2017** ; Ye *et al.* ,**2017**), we used training datasets for Setting 1 and Setting 2 containing both a raster stack of predictor variables (composite bands) and vector-based classified ground truth for the corresponding areas, see also Figure 3. After testing the training data with different numbers of classified tiles, we found that 20 000 randomly generated samples, exported as classified tiles, performed the best to train the model. The randomly generated samples were exported as classified tiles using an Image Analyst license ('Export Training Data for Deep Learning') from ArcGIS Pro (ArcGIS Pro, **2022**). The most suitable classified tile size in our case was 256x256 pixels, and in order to have 50 % overlap in each sample tile when creating the following image chips, the stride, which describes the distance of movement in the x- and y-direction, was set to be 128x128 pixels. The entire process of training, evaluating, and exporting the model was conducted using Jupyter Notebook and ArcPy libraries. During the training process, an input image in the form of classified tiles flows through the CNN network that recognizes it with a set of trainable kernels, resulting in a group of feature maps (Liu, **2018**). The trained model was saved as a 'Deep Learning Package' ('.dpk' format), which is the standard format used to deploy deep learning models on the ArcGIS Pro platform and can be used further as a pre-trained model (Miranda & Von Zuben, **2015** ; Ma & Mei, **2021** ; ESRI,**2023**). The trained models were then used to predict ravines in the other study areas for the blind test (Figures 1b and 2).

3.2.2. Evaluation

The resulting predictions of smooth and sharp ravines were evaluated quantitatively by comparing pixels of the sampled ground truth data of ravines to the automated ravine predictions of the same areas. We calculated metric precision, recall, and F1-score metrics to evaluate the performance of the two proposed models. Precision is a measure of how many of the positive predictions are made correctly (true positives) (Table 2a), while recall is a measure of how many of the positive cases were correctly predicted, over all the positive cases in the data. F1-Score is a measure combining both precision and recall. A satisfactory F1 score means that there are low false positives and low false negatives. An F1 score is considered solid with a value close to 1 (Table 2b) (Lipton *et al.* , **2014**).

Table 2. (a) Explained values of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values (Skaik,**2008** ; Safari *et al.* , **2015**) (b) to evaluate performance metrics precision, recall, and F1-score that were calculated based on (a).

(a)

Prediction	Actual value	Type	Explanation
1	1	True Positive (TP)	Predicted Positive and was Positive

Prediction	Actual value	Type	Explanation
0	0	True Negative (TN)	Predicted Negative and was Negative
1	0	False Positive (FP)	Predicted Positive but was Negative
0	1	False Negative (FN)	Predicted Negative but was Positive

(b)

Metric	Formula
Precision	$\frac{\mathrm{TP}}{TP+FP}$
Recall	$\frac{\mathrm{TP}}{TP+FN}$
F1-score	$\frac{2TP}{2TP+FP+FN}$

When trained properly, pre-trained models can be used for similar problems in similar settings to save time and reduce the need for more ground truth data (Ma *et al.* , **2021** ; Tehrani *et al.* , **2022**). For this reason, we test our pre-trained models' applicability to the four blind-test areas: Romerike N, Byneset, Orkdal, and Stadsbygd (Figures 1 and 2). Here, we regard Romerike N and Byneset to have a similar geological setting with large networks of dendritic ravines. In Orkdal and Stadsbygd, the geological setting is different, with shorter, often shallower, and more individual gullies. We compare the predictions in all study areas with a reference dataset comprising manually mapped gullies as vector lines from the Norwegian geological survey (Christoffersen *et al.* , **2021** ; van Boeckel*et al.* , **2022** ; NGU, **2023**) (Figure 1c and 2). First, we post-processed the predicted delineations to remove noise, which therefore readily can be incorporated into manual geomorphological mapping routines. The post-processing comprised 1) transforming the pixels into vector shapes, 2) buffering and dissolving the vector shapes with a 5 m radius, and 3) applying a filter by removing polygons smaller than 5000 m^2. Then, we compared the post-processed predictions using an overlay and intersect analysis to calculate length-weighted overlap and coverage of intersecting ravines. The latter represented the relative surface area of post-processed predictions intersecting with the reference dataset. We regard the intersecting predictions as true positives, which we can then use as a first-order indication of the agreement between post-processed predictions and the reference dataset. The length-weighted overlap represented the relative length of the vector lines overlapping with the post-processed predictions. The cumulative lines that did not overlap with the post-processed predictions can be regarded as an indicator of false negatives (Figure 5). We note that comparing the reference dataset with post-processed predictions does not give any information about the accuracy of the delineation of the landforms, which was done visually.

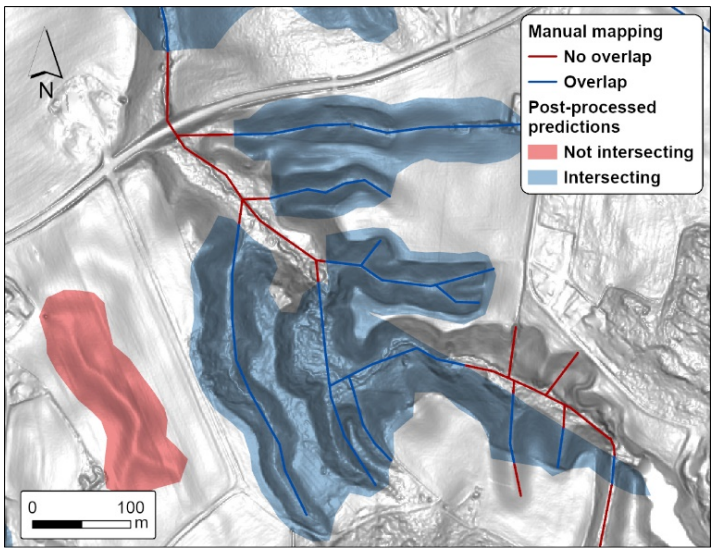


Figure 5. Comparison of the reference dataset as vector lines (Christoffersen *et al.* , **2021** ; van Boeckel et al.,**2022**) with the post-processed predictions as polygons. Blue polygons are intersecting with the vector lines, and regarded as true positives, while red polygons are not overlapping and regarded as false positives. The cumulative area of the blue polygons concerning the total area represents the coverage of intersecting ravines. The red lines do not overlap with the predicted ravines, were not picked up by the model, and were regarded as false negatives. The cumulative length of the blue lines concerning the total length represents the length-weighted overlap.

4. RESULTS

In this section, we present the performance of the U-net model in Romerike S by applying two data settings (Setting 1 and Setting 2) using Jupyter Notebook and ArcPy libraries environment. The quantitative evaluation of the sampled pixels between ground truth data and predictions are presented in Table 3. Statistically, we can see higher precision (0.75-0.82), recall (0.69-0.73), and F1 score (0.72-0.74) for sharp ravines. On the other hand, smooth ravines have a tendency to achieve lower scores for precision (0.70-0.72), recall (0.66-0.72), and 0.68-0.72 for the F1 score (Table 3). Our results show that by using the same amount of ground truth data (20 000) but from a more extensive and more diverse study area, the performance of Setting 1 (Figure 1) only increased slightly for sharp ravines but decreased for smooth ravines, with a F1 score of +0.02 and –0.04, respectively.

Even though the statistics show minor differences in the overall performance using the different data settings, visual inspection reveals that the different models pick up different sections along the same ravines. This can also be observed when comparing the predictions to the reference datasets. When combining the predictions of Setting 1 and Setting 2, the length-weighted overlap and coverage of intersecting ravines, scoring 91.1 % and 87.4 %, are slightly higher compared to the values by only using Setting 2, scoring 86.7 % and 82.8 %, respectively (Table 4).

Table 3. The statistics presented show the performances when applying training in two different settings: Setting 1 with training data from RS 1 and RS 2 and Setting 2 with training data only from RS 1 (Figure

1). GTPs = ground truth polygons.

	Setting 1	Setting 1	Setting 2	Setting 2
GTP's	smooth ravines 186	sharp ravines 147	smooth ravines 95	sharp ravines 47
Precision	0.70	0.75	0.72	0.82
Recall	0.66	0.73	0.72	0.69
F1	0.68	0.74	0.72	0.72

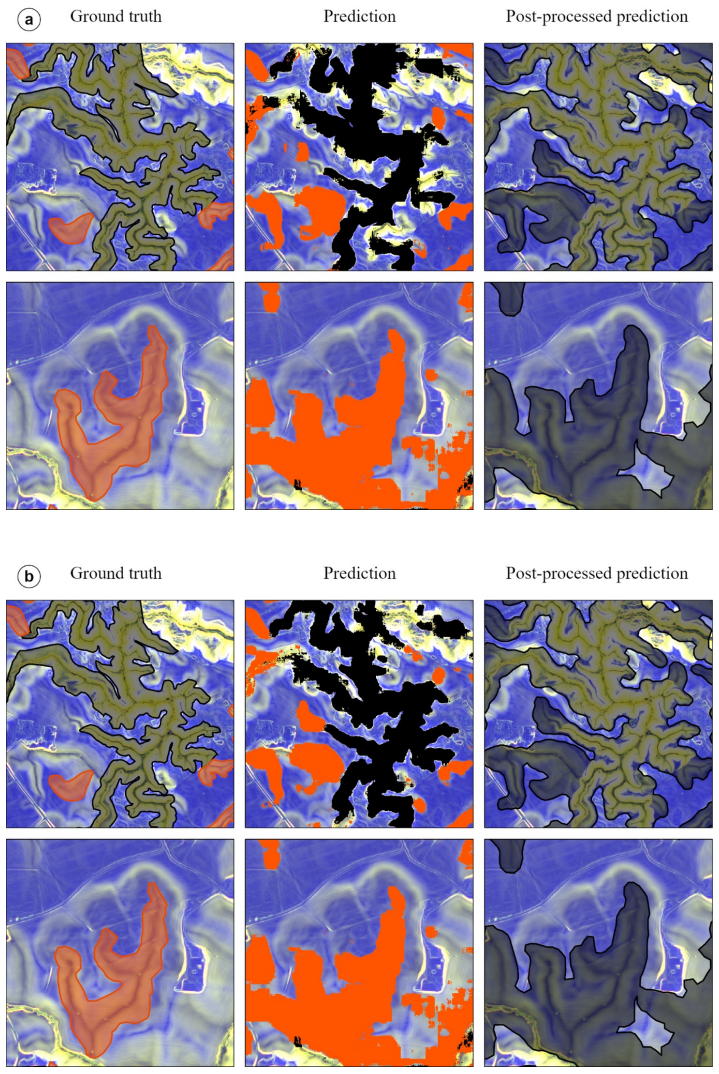


Figure 6. Detailed area comparison of ground truth data, predictions, and final post-processed predictions for (a) Setting 1 and (b) Setting 2. The background image is the predictor raster stack (RGB = TR, Slope, TPI) with sharp (black) and smooth (orange) ravines.

Table 4. The comparison of the reference dataset and post-processed predictions for Setting 2 and Setting 1 & 2 (the combined products of both Setting 1 and Setting 2).

	Setting 2			Setting 1 & 2		
	Predictions (nr.)	Coverage of intersecting ravines	Length-weighted overlap	Predictions (nr.)	Coverage of intersecting ravines	Length-weighted overlap
Romerike S	2117	82.8 %	86.7 %	1855	87.4 %	91.1 %
Romerike N	1245	94.7 %	76.6 %	954	94.1 %	86.1 %
Byneset	175	91.7 %	64.5 %	160	91.4 %	72.7 %
Orkdal	120	94.9 %	55.1 %	105	94.1 %	67.2 %
Stadsbygd	59	71.2 %	38.4 %	72	80.2 %	56.1 %

The next step was to compare the predictions of the pre-trained models to the reference datasets in Romerike S and the four blind-test areas: Romerike N, Byneset, Orkdal, and Stadsbygd (Figures 1 and 2 and Table 4). We applied both pre-trained models (Setting 1 and Setting 2) to all the study areas and post-processed the predicted pixels, explained in section 3.2.2. The first category of blind-test areas, Byneset and Romerike N, with a similar geological setting as Romerike S, show very promising results scoring 91.7 % and 94.7 % for coverage of intersecting ravines and 65.5 % and 76.6 % for length-weighted overlap, respectively. These values increased slightly when combining the post-processed predictions of Setting 1 and 2 (Table 4). The blind-test areas Stadsbygd and Orkdal scored poorly in length-weighted overlap with 38.4 % and 55.1 %, respectively, when only using the post-processed products of Setting 2. The low length-weighted overlap values indicate that the pre-trained models did not pick up many vector lines from the reference dataset. The coverage of intersecting ravines scored relatively high (> 71 %) for all the blind-test areas, which indicates that the post-processed predictions of the pre-trained models largely managed to successfully identify the ravines.

4. DISCUSSION

4.1 Automated differentiation of intact and impacted ravines and gullies

The delineation of landforms is the fundamental process of mapping the spatial extent and condition of landscape change, which conventionally is performed manually using high-resolution optical remote-sensing images or LiDAR data. Our results show that using only three terrain derivatives and ˜150 manually mapped gullies (Setting 2), the U-net model successfully predicted and differentiated intact sharp ravines from impacted smooth ravines. Quantitative pixel evaluation of sampled ground truth data revealed that doubling the ground truth data (Setting 1) only slightly improved the F1 score for sharp ravines (+0,02) but decreased for smooth ravines (-0,04). Overall, both automated identification models revealed promising results in differentiating ravines impacted by agricultural levelling from intact ravines, as also seen in Roy & Saha (**2022**).

4.2 The minimal amount and best suitable predictors

For the blind-test areas, the best predictions were achieved using predictors indicative of relative elevation, e.g., the terrain derivatives slope, TR, and TPI, as opposed to absolute elevation. We explain this by the fact that the ravines are found in raised marine fine-grained deposits at varying elevations between the study

areas (Figures 1 and 2). Because the model was trained on ground truth data located at elevations between ˜100 and 180 m a.s.l., the pre-trained model was unable to detect ravines at lower elevations. We, therefore, stress that using elevation data as a predictor should be used with caution when predicting blind-test areas.

Unlike similar studies delineating ravines with the usage of over a dozen independent predictors (Band *et al.* , **2020** ; Setargie *et al.* , **2023**), we show that a promising delineation of ravines can be achieved by only using three predictor layers derived from high-resolution elevation data. We recognize that adding additional terrain derivatives as predictors and ground truth data could potentially improve our model. However, we argue that having few predictors makes our approach more accessible and applicable in other areas for future mapping.

4.3 The applicability of pre-trained models to blind-test areas

The robustness of a model increases when successful predictions are not limited to trained areas but also manage to predict in blind-test areas for other regions (Sarker, **2021**). As we do not have ground truth data of our blind-test areas, we rely on our evaluation-based comparing the reference dataset with post-processed predictions. Our results show that the post-processed predictions broadly intersect with the reference dataset ($> 71.2\%$). If we only regard the blind-test areas with similar geological settings, namely Romerike N and Byneset, the coverage of intersecting ravines increases to > 91.7 %, indicating that the model manages to accurately identify ravines. Similarly, Romerike N and Byneset score significantly higher in length-weighted overlap compared to Orkdal and Stadsbygd, reflecting that large stretches of the reference dataset overlap with the pre-trained models. The difference in geological setting can explain the discrepancy of lower length-weighted overlapping values for Orkdal and Stadsbygd. In these areas, the ravines are much shorter and less branched into networks compared to ravines used in the ground truth dataset. Future incentives to train the model specifically for these settings or to include them in the training dataset might increase the model's performance.

We noticed that the delineation of the predictions was improved by using the combined predictions of Setting 1 and 2. This improved performance is also reflected by higher length-weighted overlap values for all blind-test areas, increasing the length-weighted overlap by at least +8.2 %. Similar to the studies of Arabameri *et al.*(**2021**) and Roy & Saha (**2022**), which used ensemble models for forecasting areas vulnerable to gully erosion, our findings confirms that the combined products of the pre-trained models increase the overall delineation of the ravines.

4.4 Applicability of DL in geomorphological mapping

One of the advantages of using automated approaches compared to manual mapping is that the automatic delineation of landforms can be evaluated quantitatively against ground truth data. For example, predictions can be evaluated by positively identified pixels (e.g., Setargie *et al.* , **2023**) and positively identified vector lines (e.g., Band*et al.* , (**2020**) and Band *et al.* , (**2020**)). Even though quantitative evaluations can give satisfactory results, there is little information about the correctness of the delineation of the predicted ravines. A ravine can, for example, be identified with a pixel accuracy of 75 %, but this does not necessarily mean that the outer extent of the predicted ravine corresponds to the actual landform. As manual mapping routines often involves the delineation of individual landform, having an inaccurate outer delineation still requires substantial adjustments to be implemented to satisfy the prerequisites for usage in geomorphological maps. We found that post-processing the pixel-based predictions into coherent polygons and reducing noise with a minimum size filter significantly increased the applicability of the product for mapping routines, see also Figure 6. We point out that the post-processing of prediction delineations should be considered when implementing automated approaches in manual mapping routines.

Conclusion

Development in computing, deep learning algorithms, and increased availability of high-resolution and free data have the potential to automate many mapping problems in Earth sciences. However, its application in differentiating and delineating landforms using deep learning techniques has not yet been investigated thoroughly. We contribute by exploring the automated differentiation of intact (sharp) and impacted (smooth) ravines and gullies with high precision using the combination of the deep learning FCNN model with only three terrain derivatives (slope, roughness, and TPI). Our best model achieved a precision score of 0.82, recall of 0.74 against ground truth data, and 91.1 % length-weighted overlap with reference datasets comprising vector lines. Our pre-trained models successfully predicted ravines in blind-test areas, scoring $>$ 72 % length-weighted overlap for regions with similar geological settings. This study shows that the post-processing ensemble models increase the applicability of Deep learning models in geomorphological mapping routines. The outcome of this research contributes towards mapping the spatial extent and condition of red-listed landforms in Norway, as well as the development of monitoring systems for future landscape change.

Code availability section

The source code is available for download here: https://github.com/alexandra-jarna/Ravines-Norway.

Program language: Python

Software required: data preparation (ArcGIS Pro/QGIS)

Author Contributions

Conceptualization, A1 and A2.; methodology, A1 and A2.; software, A1 and A2.; validation, A1 and A2; formal analysis, A1 and A2.; investigation, A1 and A2.; resources, A1, A2 and A3; data curation, A1 and A2; writing—original draft preparation, A1, A2, A3; visualization, A1 and A2; All authors have read and agreed to the published version of the manuscript

Acknowledgments

We are grateful to all those with whom we have had the pleasure to work during this and other related projects connected to the topic. Thank you to Gabriela Spakman-Tánásescu for introducing ArcGIS Pro and deep learning possibilities.

References

Ahlstrøm et al (**2019**) *AR5 Klassifikasjonssystem Klassifisering av arealressurser* .

Albawi, S. *et al.* (**2018**) 'Understanding of a convolutional neural network', *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017* , 2018-Janua(April 2018), pp. 1–6. Available at: https://doi.org/10.1109/ICEngTechnol.2017.8308186.

Arabameri, A. *et al.* (**2020**) 'Comparison of machine learning models for gully erosion susceptibility mapping',*Geoscience Frontiers* , 11(5), pp. 1609–1620. Available at: https://doi.org/10.1016/j.gsf.2019.11.009.

Arabameri, A. *et al.* (**2021**) 'Perdition of gully erosion susceptibility mapping using novel ensemble machine learning algorithms', *Geomatics, Natural Hazards and Risk* , 12(1), pp. 469–498. Available at: https://doi.org/10.1080/19475705.2021.1880977.

Arabameri, A. *et al.* (**2022**) 'Drought risk assessment: integrating meteorological, hydrological, agricultural and socio-economic factors using ensemble models and geospatial techniques',*Geocarto International* , 37(21), pp. 6087–6115. Available at: https://doi.org/10.1080/10106049.2021.1926558.

ArcGIS Pro (**2022**) *Export Training Data For Deep Learning (Image Analyst)* . Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/export-training-data-for-deep-learning.htm.

Band, S.S. *et al.* (**2020**) 'Novel ensemble approach of deep learning neural network (Dlnn) model and particle swarm optimization (pso) algorithm for prediction of gully erosion susceptibility', *Sensors (Switzerland)* , 20(19), pp. 1–28. Available at: https://doi.org/10.3390/s20195609.

Barneveld, R.J. *et al.* (**2022**) 'Estimating Ephemeral Gully Erosion Rates in a Norwegian Agricultural Catchment Using Low-Altitude Uav Imagery', *SSRN Electronic Journal* , (1), pp. 1–34. Available at: https://doi.org/10.2139/ssrn.4085344.

Bergqvist, E. (**1990**) 'Terrace-and-gully landscapes in southern and central Sweden', *U. Universitet.* [Preprint]. Available at: https://doi.org/UNGI report 77.

Blindheim, T. *et al.* (**2018**) *Kartlegging av arter i raviner i Skedsmo kommune 2017 [Mapping of species in gullies in Skedsmo municipality 2017]* .

Blindheim, T. and Abel, K. (**2002**) . *Vilt i Skedsmo kommune [Wildlife in Skedsmo municipality]* . Siste Sjan.

van Boeckel, M. Christoffersen, M. *et al.* (**2022**)*Kartlegging av rødlistede landformer, resultater fra kartlegging i 2022* . Available at: https://doi.org/NGU rapport 2022.028.

Bridge, J.S. (**2003**) 'Rivers and floodplains: forms, processes, and sedimentary record', *Journal of Quaternary Science* , 19(6), pp. 618–619. Available at: https://doi.org/10.1002/jqs.856.

Chen, W. *et al.* (**2019**) 'Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection', *Remote Sensing* , 195(October), p. 104777. Available at: https://doi.org/10.3390/rs11020196.

Chen, W. *et al.* (**2021**) 'Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility', *Journal of Environmental Management* , 284. Available at: https://doi.org/10.1016/j.jenvman.2021.112015.

Christoffersen, M. *et al.* (**2021**) *Kartlegging av rødlistede landformer: videreføring av pilotprosjekt 2019* , *NGU rapport nr. 2021.001* .

Chuma, G.B. *et al.* (**2023**) 'Gully erosion susceptibility mapping using four machine learning methods in Luzinzi watershed, eastern Democratic Republic of Congo', *Physics and Chemistry of the Earth* , 129. Available at: https://doi.org/10.1016/j.pce.2022.103295.

Erikstad, L. (**1992**) 'Recent changes in the landscápe of the marine clays, Østfold, southeast Norway', *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography* , 46(1), pp. 19–28. Available at: https://doi.org/10.1080/00291959208552279.

Erikstad, L. *et al.* (**2018**) *Landformer. Norsk rødlista for naturtyper 2018* . Available at: https://www.artsdatabanken.no/Pages/259126.

ESRI (**2023**) *Pretrained Deep Learning Models* . Available at: https://www.esri.com/en-us/arcgis/deep-learning-models.

Gayen, A. *et al.* (**2019**) 'Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms', *Science of the Total Environment* , 668, pp. 124–138. Available at: https://doi.org/10.1016/j.scitotenv.2019.02.436.

Grohmann, C.H. *et al.* (**2009**) 'Surface Roughness of Topography : A Multi-Scale Analysis of Landform Elements in Midland Valley, Scotland', *Proceedings of Geomorphometry 2009* , pp. 140–148. Available at: https://doi.org/citeulike-article-id:8857982.

Hamre, L.N. *et al.* (**2021**) 'Paradise lost — transformation of the gully landscape in South-East Norway',*Landscape Research* , 46(3), pp. 377–389. Available at: https://doi.org/10.1080/01426397.2020.1847263.

Hansen, L. *et al.* (**2007**) 'Stratigraphic evaluation of a Holocene clay-slide in Northern Norway', *Landslides* , 4(3), pp. 233–244. Available at: https://doi.org/10.1007/s10346-006-0078-4.

Higgins, C.G. and Coates, D.R. (**1990**) *Groundwater Geomorphology: The Role of Subsurface Water in Earth-surface Processes and Landforms* . Geological Society of America (Geological Society of America Special Paper). Available at: https://books.google.no/books?id=4jFmf4cGyRYC.

Hogaas, F. *et al.* (**2022**) 'Datasett for registrering av marin grense (MG) i Norge'. NGU-Rapport nr.: 2022.005, 35.

Jansson, U. & Hoitomt, T. (**2013**) *Ravinekartlegging i Nannestad kommune 2012 [Mapping of gullies in Nannestad municipality 2012]* .

Kvaerno, S. *et al.* (**2020**) 'Tiltak mot erosjon pa jordbruksareal', *4* , 6(38). Available at: https://doi.org/NIBIO POP. 2020, 6 (38).

Leng, J. *et al.* (**2019**) 'Context-Aware U-Net for Biomedical Image Segmentation', *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018* , (February), pp. 2535–2538. Available at: https://doi.org/10.1109/BIBM.2018.8621512.

Lindsay, J.B. (**2014**) 'The Whitebox Geospatial Analysis Tools project and open-access GIS', in *Proceedings of the GIS research UK 22nd annual conference* . Glasgow, UK, pp. 16–18.

Lipton, Z.C. *et al.* (**2014**) 'Optimal Thresholding of Classifiers to Maximize F1 Measure', in T. Calders et al. (eds)*Machine Learning and Knowledge Discovery in Databases* . Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 225–239.

Liu, B. *et al.* (**2022**) 'Ephemeral gully recognition and accuracy evaluation using deep learning in the hilly and gully region of the Loess Plateau in China', *International Soil and Water Conservation Research* , 10(3), pp. 371–381. Available at: https://doi.org/10.1016/j.iswcr.2021.10.004.

Liu, Y.H. (**2018**) 'Feature Extraction and Image Recognition with Convolutional Neural Networks', *Journal of Physics: Conference Series* , 1087(6). Available at: https://doi.org/10.1088/1742-6596/1087/6/062032.

Ma, Z. and Mei, G. (**2021**) 'Deep learning for geological hazards analysis: Data, models, applications, and opportunities',*Earth-Science Reviews* , 223, p. 103858. Available at: https://doi.org/10.1016/j.earscirev.2021.103858.

Miranda, C.S. and Von Zuben, F.J. (**2015**) 'Reducing the Training Time of Neural Networks by Partitioning', pp. 1–10. Available at: http://arxiv.org/abs/1511.02954.

Mohebzadeh, H. *et al.* (**2022**) 'Machine Learning Techniques for Gully Erosion Susceptibility Mapping: A Review',*Geosciences (Switzerland)* . MDPI. Available at: https://doi.org/10.3390/geosciences12120429.

NGU (**2023**) *Quaternary geological map from Geological Survey of Norway* . Available at: https://geo.ngu.no/kart/losmasse_mobil.

Nodjoumi, G. *et al.* (**2023**) 'DeepLandforms: A Deep Learning Computer Vision Toolset Applied to a Prime Use Case for Mapping Planetary Skylights', *Earth and Space Science* , 10(1). Available at: https://doi.org/10.1029/2022EA002278.

Prakash, N. *et al.* (**2021**) 'A new strategy to map landslides with a generalized convolutional neural network',*Scientific Reports* , 11(1), pp. 1–15. Available at: https://doi.org/10.1038/s41598-021-89015-8.

Reite, A.J. *et al.* (**1999**) 'Trondheim fra istid til natid – landskapshistorie og losmasser', *Norges geologiske undersokelse Grasteinen* , p. 40 p.

Ronneberger, O. *et al.* (**2015**) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in N. Navab et al. (eds)*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* . Cham: Springer International Publishing, pp. 234–241.

Roy, J. and Saha, S. (**2022**) 'Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach', *Artificial Intelligence in Geosciences* , 3(March), pp. 28–45. Available at: https://doi.org/10.1016/j.aiig.2022.07.001.

Ryan, I. *et al.* (**2022**) *Pa trygg grunn — Bedre handtering av kvikkleirerisiko* . Available at: https://www.regjeringen.no/no/dokumenter/nou-2022-3/id2905694/?ch=1.

Safari, S. *et al.* (**2015**) 'Evidence Based Emergency Medicine Part 2: Positive and negative predictive values of diagnostic tests.', *Emergency (Tehran, Iran)* , 3(3), pp. 87–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26495390%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4608

Sarker, I.H. (**2021**) 'Machine Learning: Algorithms, Real-World Applications and Research Directions', *SN Computer Science* , 2(3), pp. 1–21. Available at: https://doi.org/10.1007/s42979-021-00592-x.

Setargie, T.A. *et al.* (**2023**) 'Random Forest–based gully erosion susceptibility assessment across different agro-ecologies of the Upper Blue Nile basin, Ethiopia', *Geomorphology* , 431. Available at: https://doi.org/10.1016/j.geomorph.2023.108671.

Shelhamer, E. *et al.* (**2017**) 'Fully Convolutional Networks for Semantic Segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 39(4), pp. 640–651. Available at: https://doi.org/10.1109/TPAMI.2016.2572683.

Skaik, Y. (**2008**) 'Understanding and using sensitivity, specificity and predictive values', *Indian Journal of Ophthalmology* , 56(4), p. 341. Available at: https://doi.org/10.4103/0301-4738.41424.

Tehrani, F.S. *et al.* (**2022**) 'Machine learning and landslide studies: recent advances and applications', *Natural Hazards* , 114(2), pp. 1197–1245. Available at: https://doi.org/10.1007/s11069-022-05423-7.

Terratec (**2022**) *Viken Laser 2022. Laserskanning rapport* .

Ye, J. *et al.* (**2017**) 'Deep Learning Hierarchical Representations for Image Steganalysis', 12(11), pp. 2545–2557.

Zaidi, S.S.A. *et al.* (**2022**) 'A survey of modern deep learning based object detection models', *Digital Signal Processing: A Review Journal* , 126, p. 103514. Available at: https://doi.org/10.1016/j.dsp.2022.103514.

Zhang, Liangpei *et al.* (**2016**) 'Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art', *IEEE Geoscience and Remote Sensing Magazine* , 4(2), pp. 22–40. Available at: https://doi.org/10.1109/MGRS.2016.2540798.

BIO ALEXANDRA JARNA GANEROD

Alexandra Jarna Ganerød is a Ph.D. student at the Department of Geography, The Norwegian University of Science and Technology (NTNU).

Ganerød focuses on exploring different applications using the machine, intense learning for the automated classification of bioclastic sediments or coral reeves in the seafloor, bedrock, lineaments, wetlands, landslides, and ravines by bridging the gap between geoscientists and data scientists.

Ganerød studied Geomatics, geography, and cartography at the Technical University of Ostrava in the Czech Republic and three years of Geomatics and geology at NTNU before she started to work as a GIS specialist at Geological Survey Norway (NGU) in Trondheim.

BIO MIKIS VAN BOECKEL

Mikis van Boeckel is a researcher in the section of Quaternary geology at the *Norwegian Geological Survey* (NGU) in Trondheim.

Van Boeckel uses geomorphology, sedimentology, bedrock geology, and GIS to study landscape change and erosional processes during the Quaternary. He is interested in glacial erosion, landforms, and conservation.

Van Boeckel received a master's degree in Quaternary geology at UiT, the Arctic University of Norway, and a bachelor's in Earth Science at Utrecht University. Earlier, he worked as a GIS specialist at *Provincie Noord-Holland* with nature planning and as a research assistant in glacial erosion at *Stockholm University* .

BIO INGER-LISE SOLBERG

Inger-Lise Solberg is a research scientist at the section Geohazards and Earth Observations at the Geological Survey of Norway (NGU). She has a PhD in Engineering Geology from Norwegian University of Science and Technology (NTNU).

Solberg works with mapping and research related to quick-clay landslides. She combines geophysical, geological, geomorphological and geotechnical information to get a good overview of the ground conditions. In addition, she is works with databases for ground investigations.

ISBN 978-82-326-7724-5 (printed ver.)
ISBN 978-82-326-7723-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology