



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2023 30 stp
Fakultetet for realfag og teknologi

Kartlegging av isbreoverflater i norske fjellområder ved bruk av pikselbasert klassifisering med laserdata

Glacier surface mapping in Norwegian mountain
regions through pixel-based classification using
laser data

Ole Markus Kolbustuen Moen
Geomatikk

Forord

Denne masteroppgaven markerer slutten på mine studier ved Norges miljø- og biovitenskapelige universitet (NMBU). Oppgaven ble skrevet ved Fakultet for realfag og teknologi våren 2023.

Først og fremst vil jeg rette en stor takk til min biveileder, Dagrun Aarsten, for hennes engasjement og støtte gjennom hele prosessen. Hennes faglige kompetanse og verdifulle innspill har vært uvurderlige. I tillegg ønsker jeg å takke hovedveileder, Ivar Maalen-Johansen, som har bidratt med veiledning når det har vært nødvendig.

Jeg vil også takke NVE og Liss Marie Andreassen for deres bidrag til oppgaven, inkludert tilgang til datasett og informasjon. Uten dette ville jeg ikke vært i stand til å gjennomføre denne studien.

Til slutt vil jeg gjerne takke mine medstudenter for vennskap og samarbeid gjennom noen uforglemmelige år på Ås. En spesiell takk går til min kjæreste for hennes tålmodighet og støtte gjennom dette siste halvåret. Jeg ønsker også å uttrykke stor takknemlighet til venner og familie for deres vedvarende støtte og oppmuntring under hele studietiden.

Ole Markus Kolbustuen Moen - Ås, 08.05.2023

Sammendrag

Isbreene våre er i stadig forandring og endrer seg i takt med miljø og klima, de kan fungere som klima-indikatorer, og historiske brelegdemålinger er naturlige måleparametere på tidligere klimaendringer (Leclercq et al., 2014; Nussbaumer et al., 2011). Nøyaktige kart over isbreer er avgjørende for massebalanseberegninger og endringsvurderinger ifølge Andreassen et al. (2020). Det er derfor betydningsfullt å utforske nye løsninger og metoder for forbedring og forenkling av målinger og overvåking av norske breer.

I 2019 ble overflatene av Hellstugubreen, Østre- og Vestre Memurubreen i Jotunheimen skannet med flybåren laserskanner av Terratec AS, på oppdrag fra NVE. Samtidig ble ortofoto over breene tatt. Dette har gitt muligheten til å studere metoder for bruk av flybåren laserdata til breovervåkning, og måle nøyaktighet til metodene sammenlignet med ortofoto.

I denne studien har det blitt utviklet og testet en pikselbasert klassifiseringsmetode basert på flybåren laserdata for å skille mellom bre, snø, mørk snø, berg og bresprekker på nevnte breer. Statistiske beregninger av høydeverdier og avstandskorrigerede intensitetsverdier ble brukt til å generere en TIFF-fil som ble brukt som treningsdatasett for fire forskjellige maskinlæringsalgoritmer: Support Vector Machine, Random Forest, k-nearest neighbor og Gaussian Mixture Model. Basert på den best presterende maskinlæringsalgoritmen har det blitt forsøkt å utlede brekant, som har blitt sammenlignet med ortofoto.

Forsøket viste at alle modellene ga gode resultater basert på valideringsparametere, hvorav SVM-modellen ga de beste resultatene med høyest presisjon og gjenkall på henholdsvis 0,8745 og 0,8780, i tillegg den høyeste F1-verdien på 0,8750. Det ble også gjennomført en visuell analyse, hvor det ble synlig at modellene i stor grad klarte å skille mellom ulike overflater på mindre komplekse områder av breen. I komplekse områder, spesielt ved bresprekker på is, bresprekker er i nærheten av berg og bre på berg oppstod det ofte forvekslinger.

Studien viste at laserdata har potensialet til å brukes i pikselbasert klassifisering av is, snø, mørk snø, berg og bresprekker på norske isbreer, og kan være et supplement til ortofoto når det gjelder å finne brekant. Totalt sett kan flybåren laserdata være med på å forbedre dagens måle- og overvåkningsmetoder på norske breer.

Abstract

Our glaciers are constantly changing and evolving in response to the environment and climate, and they can serve as climate indicators. Historical measurements of glacier length are natural parameters for assessing past climate changes (Leclercq et al., 2014; Nussbaumer et al., 2011). Accurate mapping of glaciers is crucial for mass balance calculations and assessments of glacier changes, according to Andreassen et al. (2020). Therefore, it is significant to explore new solutions and methods for improving and simplifying measurements and monitoring of Norwegian glaciers.

In 2019, the surfaces of Hellstugubreen, Østre- and Vestre Memurubreen in Jotunheimen were scanned with airborne laser scanner by Terratec AS on behalf of NVE. Orthophoto over the glaciers were also taken under the same mission. This has provided an opportunity to study methods for using airborne laser data for glacier monitoring, and measure the accuracy of the methods compared to orthophoto.

In this study, a pixel-based classification method based on airborne laser data was developed and tested to distinguish between glacier, snow, dark snow, rock, and crevasses on the glaciers. Statistical calculations of height values and distance-corrected intensity values were used to generate a TIFF file that was used as a training dataset for four different machine learning algorithms: Support Vector Machine, Random Forest, k-nearest neighbor, and Gaussian Mixture Model. Based on the best-performing machine learning algorithm, an attempt was made to derive glacier edge, which was compared with orthophoto.

The experiment showed that all models performed well based on validation parameters, with the SVM model giving the best results with the highest precision and recall at 0.8745 and 0.8780, and the highest F1 score of 0.8750. A visual analysis was also conducted, which showed that the models were able to distinguish between different surfaces to a large extent. In complex areas, especially near crevasses on ice, crevasses near rock and glacier on rock, there were often confusions.

The study showed that laser data has the potential to be used in pixel-based classification of ice, snow, dark snow, rock, and crevasses on Norwegian glaciers and can be a supplement to orthophoto in finding glacier edges. Overall, airborne laser data can contribute to improving current measurement and monitoring methods on Norwegian glaciers.

Innholdsfortegnelse

Forord.....	ii
Sammendrag	iii
Abstract	iv
1 Innledning.....	1
1.1 Bakgrunn	1
1.2 Problemstilling	2
1.3 Oppsett og struktur	2
1.4 Tidligere forskning	3
2 Teori.....	5
2.1 Isbreer.....	5
2.2 Lidar.....	5
2.3 Intensitet	6
2.4 Korrigerte intensitetsverdier	6
2.5 Flybåren laserskanning.....	6
2.6 Maskinlæring.....	7
2.6.1 Over- og undertilpasning.....	8
2.6.2 Random forest	8
2.6.3 Support vector machine.....	9
2.6.4 K-nearest neighbors.....	10
2.6.5 Gaussian Mixture model	10
2.7 Valideringsparametere.....	10
2.7.1 Forvirringsmatrise	10
2.7.2 Presisjon, gjenkall og F1- verdi.....	11
3 Material og metode.....	14
3.1 Studieområde.....	14
3.1.1 Datasett	15
3.1.2 Landskapet i Østre- Jotunheimen.....	16
3.1.3 Overflater på Hellstugubreen, Østre- og Vestre Memurubreen	17
3.2 Programvarer og moduler	21
3.2.1 Quick Terrain Modeler.....	21
3.2.2 Catalyst.....	21
3.2.3 QGIS.....	21

3.2.3.1	Dzetsaka Classification tool	21
3.2.3.2	GDAL Sieve	22
3.2.3.3	Raster til vektor	22
3.2.3.4	Gaussian smoothing.....	23
3.2.4	Python.....	24
3.2.5	Scikit- learn	24
3.2.6	Filformater.....	24
3.3	Preprosessering.....	25
3.3.1	Utklipp av breen	25
3.3.2	Generere rutenettstatistikk i QTM.....	26
3.3.3	Valg av variabler og statistikk.....	27
3.3.4	Konvertere og slå sammen data i Catalyst.....	29
3.3.5	Gaussian smoothing.....	31
3.3.6	Trening- og test sett.....	32
3.4	Veiledet klassifisering	33
3.4.1	Treningspunkter	33
3.4.2	Klassifisering	36
3.4.3	Testpunkter.....	36
3.5	Brekant utledet av pikselbasert klassifisering av laserdata.....	37
3.5.1	Sammenslå klasser tilknyttet breområde	37
3.5.2	GDAL Sieve	38
3.5.3	Raster til vektor	38
3.6	Test av algoritmene.....	39
3.6.1	Statistisk test av algoritmene	39
3.6.2	Visuell analyse av algoritmene	39
3.6.3	Test av utledning av brekant.....	39
4	Resultater	40
4.1.1	Statistisk test og valideringsparametere	40
4.1.2	Prosentvis fordeling i klassifisert areal.....	42
4.1.3	Resultater visuell analyse.....	43
4.1.4	Resultater brekant.....	47
5	Diskusjon.....	48
5.1	Analyse av resultatene	48

5.1.1	Statistisk test og valideringsparametere	48
5.1.2	Visuell analyse av klassifisering.....	49
5.1.3	Visuell analyse brekant	57
5.2	Usikkerhet knyttet til metodikk.....	59
5.2.1	Korrigert laserdata.....	59
5.2.2	Begrensninger med Dzetsaka algoritmen	60
5.2.3	Trening og testdata	60
5.2.4	Testpunkter.....	60
5.3	Tilgjengelig tidligere forskning.....	60
6	Konklusjon	61
6.1	Videre arbeid.....	61
	Bibliografi:.....	64

Figurer

1 Figuren viser tilfelle hvor utgående laserstråle fra skanneren går ut i vinkel, dette gjør avstanden laserstrålen går lenger, som medfører svakere signal tilbake til sensor. I tillegg vil innfallsvinkel påvirke refleksjonen.....	6
2 Figuren viser et tilfelle hvor skannevinkel er vinkelrett på terrenget. Signal er optimalt med tanke på vinkel mot underlaget, men høyden mot terrenget kan være avgjørende for mottatt signal	6
3 De viktigste komponentene i flybåren laserskanning.....	7
4 Overtilpasning (venstre), undertilpasning (høyre). Illustrasjon inspirert av (Bullinaria, 2015)	8
5 Random forest bruker en kombinasjon av flere beslutningstrær for å utføre prediksjonen. Illustrasjon hentet fra (Jibril et al, 2020).....	9
6 SVM-algoritmen finner den beste separasjonslinjen mellom forskjellige klasser av data ved å maksimere marginen mellom dem ved hjelp av støttevektorer. Illustrasjon inspirert av (Raschka & Mirjalili, 2019)	9
7 Radius KNN ser på alle eksempler i et gitt område rundt et testpunkt (rød stjerne) og tilordner klassen til testpunktet basert på flertallet av eksempler i dette området.	10
8 Forvirringsmatrisen sorterer forventede og predikerte verdier	11
9 oversikt over hva som regnes som SP, FP, SN, FN i et multiklasseproblem.	12
10 Studieområde - Kilde: norgebilder.no (Kartverket)	15
11 Studieområde - Kilde: norgebilder.no (Kartverket)	15
12 Studieområde Hellstugubreen samnt Østre- og Vestre Memurubreen i Jotunheimen – Kilde: norgebilder (Kartverket)	15
13 Topografisk kart over Hellstugubreen samt Østre- og Vestre Memurubreen. Kilde: norgebilder.no, Kartverket, Geovekst og kommuner	15
14 2D punktsky fra Hellstugubreen, Østre- og Vestre Memurubreen, vist i QTM, farget av intensitetsverdier	16
15 3D punktsky fra Hellstugubreen, Østre- og Vestre Memurubreen, vist i QTM, farget av intensitetsverdier	16
16 Typisk terrengformasjon for Jotunheimen. Breer i Norge kan periodevis ha store mengder snø, og måletidspunkt på året er viktig. Foto: Nils Haakensen (digitalmuseum).....	17
17 Overflatene som skal klassifiseres illustrert med ortofoto levert av NVE.	18
18 f) og g) viser forskjellige typer overflater av bre h) viser en overgang fra bre til berg som kan være utfordrende å registrere visuelt. - Vist med Ortofoto fra NVE.....	19
19 j), k) og l) viser forskjellige typer overflater av berg. - Vist med Ortofoto fra NVE.....	20
20 Før og etter Sieve algoritmen er kjørt med 3 i terskelverdi. Øyer med- og under 3 piksler blir slått sammen med omkringliggende rasterverdi.....	22
21 Fra raster(venstre) til vektor(høyre).....	22
22 a) viser bilde uten Gaussian smoothing, b) viser bilde med 3 i standardavvik, og c) viser bilde med 10 i standardavvik. Illustrasjon: IkamusumeFan (https://tinyurl.com/IkamusumeFan)	24
23 Flytskjema som beskriver gangen forarbeidene- og preprosessering av data	25
24 Punktsky fremstilt i QTM i 2D, før utklipp.....	26
25 Punktsky fremstilt i QTM i 2D, etter utklipp.....	26
26 Tverrsnitt av hellende fjell.....	27

27 Visualisering av tverrsnitt fra hellende berg.....	27
28 Eksempel på generert grid statistikk på variabel Z og statistikk helning som DEM i Quick terrain modeller	27
29 a) Visualisert rutenettstatistikk med variabel: Z og statistikk: helning	28
30 b) Visualisert rutenettstatistikk med variabel: Z og statistikk : andre grads helning	28
31 d) Visualisert rutenettstatistikk med variabel: intensitet, og statistikk: Gjennomsnitt	28
32 Histogram før etter skalering.....	30
33 Sammenslått TIFF med båndene med Z helning, Z andre grads helning og intensitet gjennomsnitt er illustrert med henholdsvis rød, grønn og blå i Catalyst.....	31
34 Utklipp fra den sammenslåtte filen før og etter Gaussian smoothing er påført	32
35 Trening- og testområde illustrert med visualisert laserdata	33
36 Prosentvis fordeling av treningspunkter.....	34
37 Røde punkter representerer treningsgrunnlag for bresprekker	35
38 Flytskjema i klassifiseringsprosessen.....	36
39 Testpunkter med sann verdi basert på ortofoto fra NVE	36
40 Prosentvis fordeling av testpunkter	37
41 Resultat av reklassifisering	38
42 Valideringsparametere for algoritmene	41
43 Forvirringsmatrise for SVM	41
44 Forvirringsmatrise for RF.....	41
45 Forvirringsmatrise for KNN.....	41
46 Forvirringsmatrise for GMM.....	41
47 Prosentvis fordelingen av klassifisert areal for de ulike algoritmene.....	42
48 Resultater SVM sammenlignet med ortofoto fra NVE.....	43
49 Resultater pikselbasert klassifisering RF sammenlignet med ortofoto fra NVE.....	44
50 Resultater pikselbasert klassifisering KNN sammenlignet med ortofoto fra NVE.....	45
51 Resultater pikselbasert klassifisering GMM sammenlignet med ortofoto fra NVE.....	46
52 Reklassifisert bilde med Sieve algoritmen. Terskelverdi er satt til 50 000 piksler	47
53 Resultatet etter raster til vektor algoritmen med ortofoto fra NVE i bakgrunn.....	47
54 Ortofoto levert av NVE.....	50
55 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på breflekk	50
56 Ortofoto levert av NVE.....	51
57 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på breflekk	51
58 Ortofoto levert av NVE.....	52

59 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på overflate med snø og mørk snø	52
60 Ortofoto levert av NVE.....	53
61 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM i komplekst område	53
62 Ortofoto levert av NVE.....	54
63 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM som illustrerer klassifisering av bresprekker	54
64 Ortofoto levert av NVE.....	55
65 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM i komplekst område	55
66 Ortofoto levert av NVE.....	56
67 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM som illustrerer overgang fra berg til breoverflater	56
68 a), b), c) og d) viser utklipp av vektor i nøyaktig klassifiserte områder	58
69 e), f), g) og h) illustrerer overflater på breen som har vært for komplekse for utledning av nøyaktig brekant.....	59

Tabeller og formler

Tabell 1 Brenavn og breID for breene (Andreassen, 2022; Andreassen et al., 2022; Breatlas, 2023). Posisjon hentet fra norgeskart.no, Kartverket	14
Tabell 2 Dataoversikt om Riegl VQ 1560i. (Riegl, 2019)	16
Tabell 3 Forklaring på statistikk utregninger i QTM. (Hentet fra hjelpesidene til Applied Imagery)	27
Tabell 4 Minimum og maksimumsverdi på rutenettstatistikken før konvertering.....	29
Tabell 5 Nye minimum og maksimum verdier for hvert bånd	30
Tabell 6 Den sammenslåtte TIFF filen og tilhørende farge i Catalyst.....	30
Tabell 7 Antall treningspunkter som blir brukt for å trene opp maskinlæringsmodellen	34
Tabell 8 Oversikt over antall tilfeldige testpunkter i område.....	37
Tabell 9 Prosentvis oversikt over klassifisert areal for hver modell.....	42
Formel 1 c: hastighet, t: tid, n: brytningsindeks.....	5
Formel 2 Presisjon for en klasse	12
Formel 3 Makrogjennomsnitt av presisjon	12
Formel 4 Gjenkall for en klasse.....	12
Formel 5 Makrogjennomsnitt av gjenkall	13
Formel 6 F1 verdi for hver klasse.....	13
Formel 7 Makrogjennomsnittet av F1 verdi.....	13
Formel 8 x, x: Koordinatene til pikselen, σ : Standardavviket for normalfordelingen	23

Forkortelser

CMYK	Cyan, Magenta, Yellow og Key
EUREF89	European Reference Frame 1989
FM	Forvirringsmatrise
FN	Falsk negativ
FOV	Field of view
IMU	Inertial measurement unit
KNN	k-Nearest Neighbors
Lidar	Light Detection and Ranging
MHz	Megahertz
NDH	Nasjonal detaljert høydemodell
NN1954	Normalnull 1954
QTM	Quick terrain modeller
RF	Random forest
RGB	Rød, grønn og blå
SN	Sann negativ
SOSI	Samordnet Opplegg for Stedfestet Informasjon
SP	Sann positiv
SVM	Support vector machine
UTM	Universal Transverse Mercator

1 Innledning

1.1 Bakgrunn

Norges isbreer dekker 0,7 % av landarealet, og har betydning for vannkraft, vannledelse og turisme i Norge (Andreassen et al., 2020). Isbreene våre er i stadig forandring og endrer seg i takt med miljø og klima, dette betyr at de kan fungere som klimaindikatorer, og historiske breleengdemålinger er naturlige måleparametere på tidligere klimaendringer (Leclercq et al., 2014; Nussbaumer et al., 2011).

Nøyaktige kart over isbreer er avgjørende for massebalanseberegninger og endringsvurderinger ifølge Andreassen et al. (2020). Det er derfor betydningsfullt å forske på nye løsninger og metoder for forbedring og forenkling av målinger og overvåking av norske breer. I en samtale relatert til denne oppgaven ga NVE uttrykk for interesse i å undersøke mulighetene for å bruke laserdata til å kartlegge breer. NVE er en norsk statlig organisasjon som overvåker vassdrag, energi, skred og flom. De måler hvert år norske isbreer og samler inn data om brevolum, brefrontposisjon og temperatur på norske breer. Dette gir viktig informasjon om endringer over tid og hvordan klimaendringer påvirker breene og vannressursene. Målingene inngår i FNs målenett for isbreer og brukes i rapportene til FNs klimapanel IPCC.

I denne oppgaven vil reflektansdata, inkludert rekkeviddekorrigerte amplitudeverdier fra Riegl VQ-1560i laserskanneren, samt høydeverdier fra samme sensor, bli anvendt i en pikselbasert klassifiseringsmetode for å identifisere og skille mellom forskjellige overflateklasser på Hellstugubreen, Østre- og Vestre Memurubreen. Videre vil det blir forsøkt å utlede brekant ut ifra dette. Det vil også bli utført en evaluering av metodene.

1.2 Problemstilling

I denne studien vil det bli undersøkt om det er mulig å utvikle en metode hvor intensitetsverdier og høydeverdier fra flybårne laserdata kan anvendes i pikselbasert klassifisering av ulike snøtyper, berg og bresprekker på norske isbreer, og videre utlede brekant ut ifra dette. Studien vil evaluere om denne metoden kan forbedre dagens overvåkings- og målemetoder

For å regne klassifiseringsdelen av metoden som anvendbar bør den klassifisere med høy nøyaktighet, ikke forveksle ulike overflater i for stor grad og den bør ha klare grenser mellom overflater. Når det kommer til utledning av brekant bør metoden gi et godt estimat som skal kunne være på like godt nivå som ved manuelt uttak av breomkrets fra ortofoto.

Problemstilling:

- Kan verdier innhentet fra flybåren laserskanning anvendes til pikselbasert klassifisering av overflatene is, snø, mørk snø, berg og bresprekker på norske isbreer?

Ut ifra denne problemstillingen dukker det opp to forskningsspørsmål som skal besvares:

- Kan laserdata være med på å forbedre dagens overvåkning- og målemetoder av norske isbreer?
- Er det mulig å avlede brekant ut ifra klassifiseringsresultatet?

1.3 Oppsett og struktur

Kapittel 1: Innledning

Dette kapitlet gir en innledende bakgrunn for oppgaven ved å presentere problemstillingen. I tillegg vil det bli gjennomgått tidligere forskning på feltet for å plassere oppgaven i en større kontekst.

Kapittel 2: Teori

Dette kapitlet gir en forklaring på den underliggende teorien og konseptene som er brukt i oppgaven. Det tar sikte på å gi en forståelse av de sentrale teoretiske begrepene som ligger til grunn.

Kapittel 3: Materialer og metode

Dette kapitlet forklarer fremgangsmåtene, program og materiale som er brukt i oppgaven. Det gir en detaljert beskrivelse av metodene som er brukt for å analysere og prosessere data, og hvilke materialer som er brukt i analysen. Kapitlet gir en forklaring på de viktigste tekniske aspektene ved oppgaven.

Kapittel 4: Resultater

Dette kapitlet presenterer resultatene som har kommet frem av metodene som er brukt i oppgaven.

Kapittel 5: Diskusjon

Dette kapitlet drøfter resultatene som er presentert i kapittel 4, samt de viktigste funnene. Videre diskuteres eventuelle begrensninger ved studien. Diskusjonskapitlet diskuterer eventuelle uavklarte spørsmål eller utfordringer som ble avdekket underveis i oppgaven, og peker på forbedringer som kan være aktuelle for å optimalisere resultatet.

Kapittel 6: Konklusjon og videre arbeid

Dette kapitlet presenterer svarene på oppgavens problemstilling og oppsummerer de viktigste funnene som er gjort. Det drøfter resultatene og hvor godt metodene har fungert. Kapitlet gir en oversikt over eventuelle videre forskning som kan undersøkes i fremtiden.

1.4 Tidligere forskning

Sindre Engh har i sin masteroppgave fra 2013 ved UiO, «Potensiale for anvendelse av Lidar-data i glasiologi» (Engh, 2013) forsket på automatisk kartlegging av bresprekker på Storsteinsfjellbreen i sammenheng med en geografisk analyse av område. Der var Engh innom mulige svakheter ved bruk av FLS for glasiologiske anvendelser på bakgrunn av punkttetthet. Videre testet Engh en korreksjonsmodell for intensitetsverdier, hvor det ble konkludert med at korrigeringsmodell gav gode resultater til segmentering av breoverflaten.

I artikkelen “Glacier Facies Mapping Using a Machine-Learning Algorithm: The Parlung Zangbo Basin Case Study” (Zhang et al., 2019) er det undersøkt bruk av laserdata for å klassifisere breområder i en del av Parlung Zangbo-bassenget i den sørøstlige delen av Tibetplatået. Der ble det konkludert med at laserdata kan være en effektiv måte å skille mellom forskjellige isdekke-typer, inkludert breer med stein og grus. I studien ble det utviklet en automatisk algoritme som benyttet seg av maskinlæringsmetoden Random Forest, i kombinasjon med analyse av topografiske og teksturbaserte egenskaper basert på flere digitale høydemodeller og Landsat-8 bilder. Resultatene fra studien viste at metoden var effektiv for å klassifisere alle terrengetyper i området, og at det var mulig å skille mellom de ulike isdekketyper med god presisjon.

I studien «Glacier surface feature detection and classification from airborne Lidar data» (Höfle et al., 2009) ble det utviklet en helautomatisk prosedyre for breoverflate klassifisering, direkte på Lidar-punktskyen. Der ble det oppnådd en nøyaktighet på klassifiseringen på over 90 %, som var høyere en nøyaktigheten til en tilsvarende ortofoto-segentering og klassifisering som ble gjort samtidig, som oppnådde en nøyaktighet på 86,7 %.

Kjøllmoen har i rapporten «Reanalysing a glacier mass balance measurement series - Langfjordjøkelen 2008-2018» (Kjøllmoen, 2019) sammenlignet geodetisk og glasiologiske måle-metoder. Resultatene viser at det ikke er noen signifikante avvik mellom glasiologisk og geodetisk metode for perioden 2009-2018 på Langfjordjøkelen.

2 Teori

2.1 Isbreer

Isbreer er komplekse og dynamiske systemer som består av ulike lag av snø og is. Isen blir dannet av nedbør i form av snø og rim (Liestøl), og øker eller minker over tid avhengig av klimaet (Andreassen, 2022). I tillegg til snø og is, finnes det ofte bergformasjoner i nærheten av isbreer som kan påvirke isbreen ved å gi støtte og struktur til isen.

Breenes størrelse og massebalanse påvirkes av klimatiske forhold som nedbør, temperatur og vind. Dette gjør isbreer til gode indikatorer på klimaendringer (Haeberli et al., 2007). Overvåkning av isbreer er viktig for å forstå endringer i massebalanse og struktur over tid. Dette kan gjøres ved hjelp av feltmålinger, satellittbilder, flyfoto og annen teknologi (Sørskaar et al.).

2.2 Lidar

Lidar, «Light Detection and Ranging» er en metode for å måle avstand fra et instrument som sender ut laserlys og til overflaten den treffer. Laseren forsterkes ved hjelp av stimulert emisjon og gir en svært intens, spesifikk bølgelengde og godt avgrenset stråle i en bestemt retning (Holtebekk & Ringvold, 2019).

Tidsintervallet mellom sending og mottak av lys som er sendt ut fra et laserinstrument kan bestemmes ved å benytte seg av lysets fundamentale egenskap som er konstant hastighet i vakuum. Siden lyset går igjennom atmosfæren må brytningsindeksen tas med i beregningen, som tar hensyn til luft-temperatur, trykk og fuktighet (Vosselman & Maas, 2010).

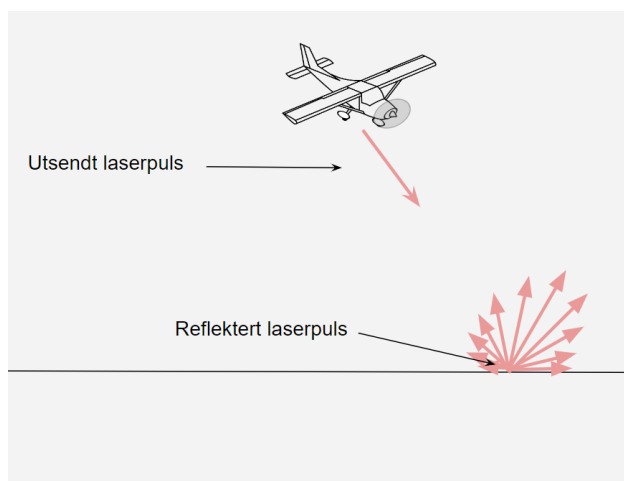
$$R = \frac{c \Delta t}{n \cdot 2}$$

Formel 1 c: hastighet, t: tid, n: brytningsindeks

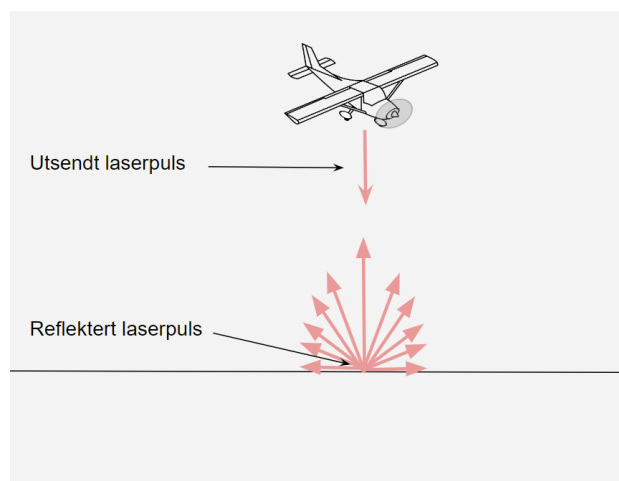
I tillegg til å måle avstanden fra sensoren til overflaten som laserpulsen treffer, kan en lidar-sensor også innhente intensitetsverdier, som gir informasjon om refleksjonen til overflaten som er truffet av laserstrålen. Intensitetsverdiene kan brukes til å differensiere mellom ulike materialer eller til å identifisere objekter i et område.

2.3 Intensitet

Intensitet fra laserinstrument kan defineres som et mål for styrken av signalet som returnerer til skanneren, ofte definert som amplitude eller refleksjon i sammenheng med laserskanning. Intensitetsverdiene som registreres avhenger av atmosfæriske forhold, innfallsvinkel til laserpulsene og avstand til overflaten (Aarsten, 2018). Dette betyr at intensitet ikke alltid er konsistent for like overflater, og korrigering av intensitetsverdier kan være nødvendig ved bruk i klassifisering.



1 Figuren viser tilfelle hvor utgående laserstråle fra skanneren går ut i vinkel, dette gjør avstanden laserstrålen går lenger, som medfører svakere signal tilbake til sensor. I tillegg vil innfallsvinkel påvirke refleksjonen.



2 Figuren viser et tilfelle hvor skannevinkel er vinkelrett på terrenget. Signal er optimalt med tanke på vinkel mot underlaget, men høyden mot terrenget kan være avgjørende for mottatt signal

2.4 Korrigerte intensitetsverdier

Reflektans er en benevnelse på avstandskorrigerte amplitudeverdier fra Riegl laserskannere, og er en moderne metode for å lagre intensitetsverdier. Dette gir homogene intensitetsverdier for like objekter uavhengig avstand fra skanner (Solli, 2019). Reflektansverdier kan være nyttige i pikselbasert klassifisering av overflater, som i denne oppgaven. Dette skyldes at ikke-avstandskorrigerte intensitetsverdier kan variere over en overflate avhengig av forskjellige faktorer, inkludert atmosfæriske forhold, avstand til overflaten og innfallsvinkel. Ved å bruke reflektansverdier, er det større sannsynlighet for å ha de samme verdiene på like overflatetyper, noe som resulterer i et bedre treningsgrunnlag for maskinlæringsmodeller.

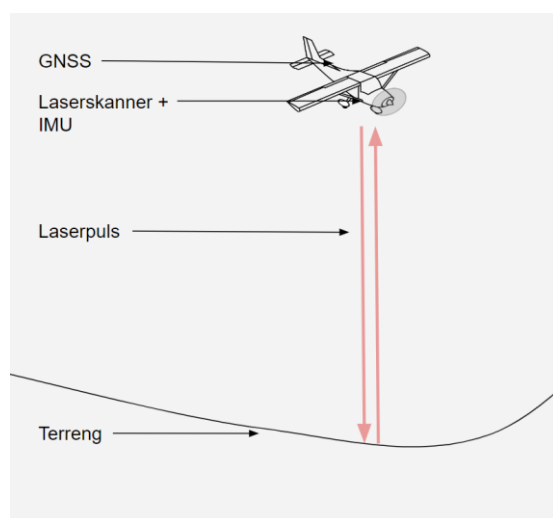
2.5 Flybåren laserskanning

Flybåren laserskanning er en teknologi som brukes til å samle inn detaljert tredimensjonal informasjon om overflater ved hjelp av en laser som er montert på et luftfartøy. Laseren sender korte pulser

nærinfrarødt lys ned mot bakken ved hjelp av motoriserte speil ifølge Dick (2020). Med utgangspunkt i tidsmålinger av laserpulser fra sensoren, og nøyaktig posisjon til flyet blir punktet hvor laserstrålen treffer en overflate registrert.

For en nøyaktig måling av treffpunktet på bakken, er systemet avhengig av en laserskanner som måler avstanden fra fartøyet til et bestemt punkt på bakken, vinkel på det motoriserte speilet, samt et system bestående av GNSS og IMU som kan måle den presise posisjonen til flyet (Vosselman & Maas, 2010). GNSS-komponenten på flyet måler posisjon med lav oppdateringsfrekvens, mens IMU'en registrerer akselerasjon og rotasjoner om flyets akser med høy oppdateringsfrekvens.

I følge SOSI- standard fra Geovekst og Kartverket er kravet på høy nøyaktighet på prosjekter i Norge for systematisk- horisontal og vertikal nøyaktighet henholdsvis 15- og 5 centimeter (Geovekst, 2022).



3 De viktigste komponentene i flybåren laserskanning

2.6 Maskinlæring

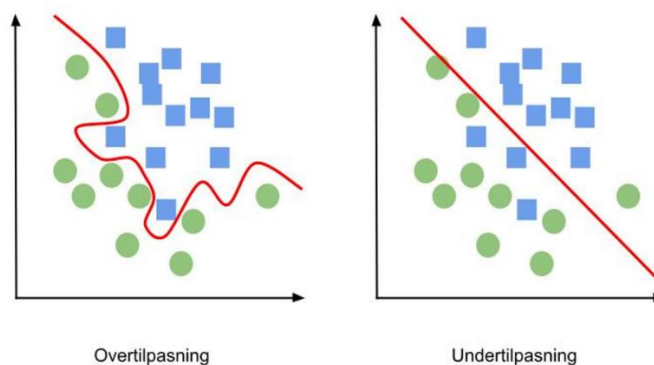
Maskinlæring kan defineres som datametoder som bruker erfaring til å gjøre forutsigelser. Erfaringen baseres på tidligere informasjon tilgjengelig, som kan være former for innhentet data fra virkeligheten (Mohri et al., 2018), og baseres på statistiske metoder for å finne mønstre i dataene (Tidemann).

Styrt pikselbasert klassifiseringer er en type maskinlæringsprosess der algoritmen lærer å kategorisere data ved hjelp av et treningsdatasett der innholdet i datasettet er merket med en bestemt kategori. Ved hjelp av dette treningsdatasettet trener en valgt algoritme en såkalt modell til å identifisere mønstre eller egenskaper som er assosiert med hver kategori i datasettet. Deretter kan modellen brukes til å forutsi kategoriene for nye, usette datasett (Mohri et al., 2018).

2.6.1 Over- og undertilpasning

Overtilpasning skjer når den opptrente maskinlæringsmodellen blir for tilpasset til treningsdataene. Dette betyr at modellen kan gi veldig nøyaktige resultater for treningsdataene, men når den støter på nye, ukjente data, kan den gi feilaktige resultater (Raschka & Mirjalili, 2019).

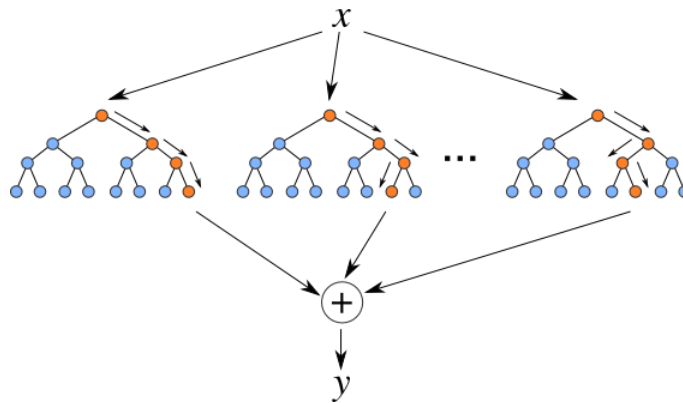
Undertilpasning skjer når modellen ikke har nok kapasitet til å fange opp de viktigste mønstrene i treningsdataene. Dette betyr at modellen ikke kan gi nøyaktige resultater for både treningsdata og ukjente data. Derfor er det viktig å finne en balanse mellom å trene en modell som passer godt til treningsdataene, samtidig som den kan generalisere og gi nøyaktige resultater for nye, ukjente data (Bonaccorso, 2017; Raschka & Mirjalili, 2019).



4 Overtilpasning (venstre), undertilpasning (høyre). Illustrasjon inspirert av (Bullinaria, 2015)

2.6.2 Random forest

Random forest (RF) er en beslutningstre-basert algoritme som er kjent for å være lett å skalere og bruke. Ideen bak RF er å trene en modell som gir god forutsigelse basert på flere beslutningstrær som hver tar en beslutning basert på tilfeldige utvalg fra datasettet, hvorav predikasjonen baseres på gjennomsnittet eller middelverdien av beslutningene (Pedregosa et al., 2011). Det endelige resultatet skal være en robust modell som har god generaliseringsytelse og er mindre utsatt for overtilpasning (Raschka, 2015).

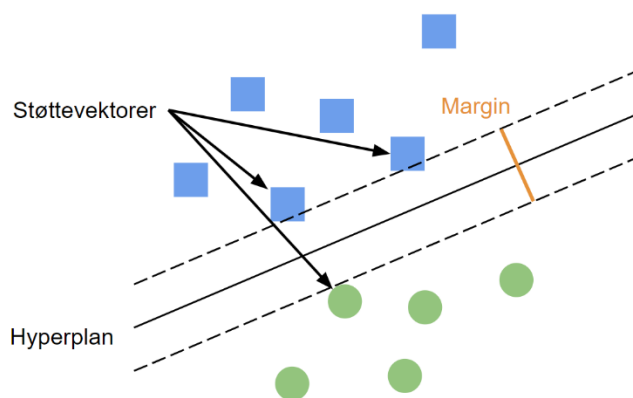


5 Random forest bruker en kombinasjon av flere beslutningstrær for å utføre prediksjonen. Illustrasjon hentet fra (Jibril et al., 2020)

2.6.3 Support vector machine

SVM er en maskinlæringsalgoritme som tar et sett med data som input og finner den beste måten å skille datapunktene fra forskjellige kategorier ved hjelp av en matematisk linje, denne linjen kalles et hyperplan. SVM beregner en margin som er avstanden mellom hyperplanet og de nærmeste datapunktene fra hver klasse for å optimalisere hyperplanet. Marginen og hyperplanet er et produkt av støttevektorer, som er datapunktene som er nærmest hyperplanet, eller avgjørelsesgrensen i SVM-algoritmen (Raschka, 2015).

SVM algoritmen kan utvides til å løse problemer med flere klasser. I denne oppgaven er det brukt standardmetoden for multiklasseproblem med SVM som kalles «one-vs-all». I denne tilnærmingen trener man en SVM-modell for hver klasse, og prediksjonen for et nytt datapunkt blir gjort ved å sammenligne resultatene av hver modell (Pedregosa et al., 2011).

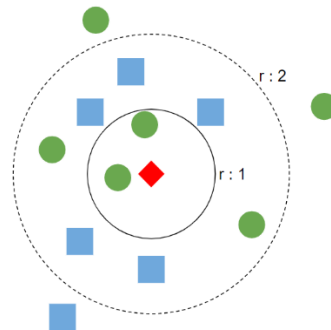


6 SVM-algoritmen finner den beste separasjonslinjen mellom forskjellige klasser av data ved å maksimere marginen mellom dem ved hjelp av støttevektorer. Illustrasjon inspirert av (Raschka & Mirjalili, 2019)

2.6.4 K-nearest neighbors

KNN (K-nearest neighbors) er en maskinlæringsalgoritme som klassifiserer et datapunkt basert på k nærmeste naboer i treningsdataene. KNN søker å finne de punktene treningsdataene som ligger nærmest datapunktet som skal klassifiseres, og bruker deres tilhørende klasser i en flertallsavgjørelse til å klassifisere det nye datapunktet, KNN antar at datapunktene som er nærmere hverandre, er mer sannsynlig å tilhøre samme klasse, og dermed gir algoritmen en prediksjon basert på denne antakelsen (Raschka, 2015).

Radiusbasert KNN, som er benyttet i klassifiseringen i denne oppgaven, er en type KNN-algoritme som brukes når datasettet ikke er jevnt fordelt. I stedet for å bruke et fast antall nærmeste naboer, bestemmes antall naboer basert på en fast radiusverdi som settes av brukeren (Pedregosa et al., 2011).



7 Radius KNN ser på alle eksempler i et gitt område rundt et testpunkt (rød stjerne) og tilordner klassen til testpunktet basert på flertallet av eksempler i dette området.

2.6.5 Gaussian Mixture model

En Gaussian mixture model (GMM) er en modell som brukes til å modellere sammensatte datasett ved hjelp av en kombinasjon av Gaussiske fordelinger. Modellen antar at datapunktene kommer fra en blanding av slike fordelinger, der hver fordeling tilsvarer en klynge eller klump av datapunkter. GMM kan brukes til å analysere data og finne mønstre i dem (Bishop & Nasrabadi, 2006).

2.7 Valideringsparametere

2.7.1 Forvirringsmatrise

En forvirringsmatrise er en tabell som brukes til å evaluere hvor godt en maskinlæringsmodell har prestert i å predikere verdier. Matrisen sammenligner sanne verdier mot de predikerte og viser hvor mange ganger modellen har klassifisert dataene riktig eller feil, og hva modellen har predikert om den tar feil. Typisk vil forvirringsmatrisen ha en rad for de sanne klassene og en kolonne for de predikerte

klassene. Oversikten over predikerte mot sanne verdier kan brukes til å beregne statistikk for evaluering av modellen. Presisjon, gjenkall og F1- verdi er ofte brukte statistiske parametere som avledes av forvirringsmatrisen for å evaluere en maskinlæringsmodell.

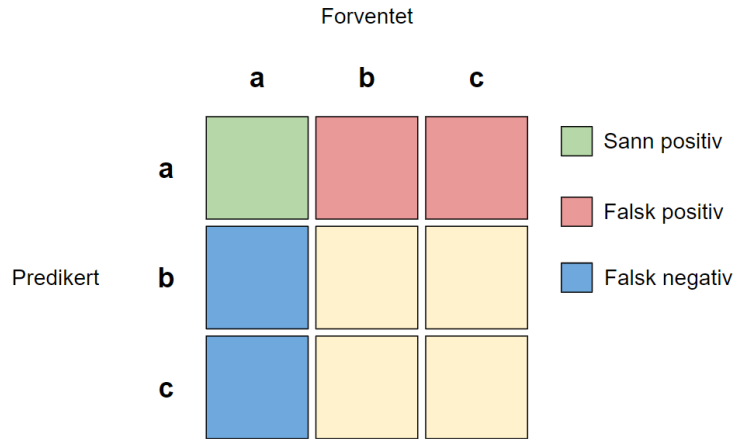
		Forventet			
		a	b	c	d
Predikert	a	23	4	1	7
	b	4	30	4	8
	c	5	3	38	5
	d	7	1	7	19

8 Forvirringsmatrisen sorterer forventede og predikerte verdier

2.7.2 Presisjon, gjenkall og F1- verdi

Presisjon, gjenkall og F1- verdi er valideringsparametere som ofte brukes i maskinlæring og er produkter av forvirringsmatrisen som kan brukes til å vurdere hvor godt en maskinlæringsmodell har prestert. Disse valideringsparameterne blir beregnet ut ifra «sann positiv», «falsk positiv», «sann negativ» og «falsk negativ». I et multiklasse problem summeres disse begrepene opp for hver klasse, for å deretter brukes i utregning av disse parameterne.

I denne oppgaven benyttes makro-gjennomsnitt for presisjon, gjenkall og F1- verdi. Makro-gjennomsnitt benyttes når alle klasser skal behandles likt, og for å evaluere den generelle ytelsen til klassifiseringen (Kumar, 2020).



9 oversikt over hva som regnes som SP, FP, SN, FN i et multiklasseproblem.

Presisjon er en parameter som måler hvor mange positive prediksjoner som faktisk tilhører en bestemt klasse. Det vil si at presisjon gir informasjon om modellens evne til å identifisere sanne positive eksempler og unngå falske positive klassifiseringer. Presisjon beregnes ved å dele antall sanne positive prediksjoner på summen av alle positive prediksjoner. Presisjon gir spesifikk informasjon om modellens evne til å klassifisere eksempler innenfor en bestemt klasse.

$$Presisjon_n = \frac{SP_n}{SP_n + FP_n}$$

Formel 2 Presisjon for en klasse

$$Presisjon_{makro} = \frac{SP_{avg}}{FP_{avg} + FN_{avg}}$$

Formel 3 Makrogjennomsnitt av presisjon

Gjenkall måler hvor godt modellen kan gjenkjenne alle positive tilfeller i datasettet, og beregnes ved å dele antall sanne positive prediksjoner på antall sanne positive og alle falske negative. Gjenkall er en viktig metrikk i situasjoner der det er viktig å unngå falske negative prediksjoner. Både presisjon og gjenkall gir en score mellom 0 og 1, hvor 1 er best.

$$Gjenkall_n = \frac{SP_n}{SP_n + FN_n}$$

Formel 4 Gjenkall for en klasse

$$Gjenkall_{makro} = \frac{SP_{avg}}{P_{avg} + FP_{avg}}$$

Formel 5 Makrogjennomsnitt av gjenkall

F1-verdi er en evaluasjonsmetode som tar hensyn til både presisjon og gjenkall for å gi en helhetlig vurdering av modellens ytelse, og gir en score mellom 0 og 1, der 1 er den beste ytelsen. F1-score er spesielt nyttig i tilfeller der både presisjon og gjenkall er like viktige for å unngå både falske positive og falske negative resultater. F1-verdi brukes for å vurdere modellens ytelse og balanse mellom presisjon og gjenkall (Raschka & Mirjalili, 2019).

$$f1\ verdi_n = 2 \times \frac{Presisjon_n \times Gjenkall_n}{Presisjon_n + Gjenkall_n}$$

Formel 6 F1 verdi for hver klasse

$$f1\ verdi_{makro} = 2 \times \frac{Presisjon_{avg} \times Gjenkall_{avg}}{Presisjon_{avg} + Gjenkall_{avg}}$$

Formel 7 Makrogjennomsnittet av F1 verdi

3 Material og metode

Metodekapittelet i denne oppgaven gir en detaljert beskrivelse av fremgangsmåten og verktøy som ble brukt for å analysere data. Kapittelet vil forklare hvilke programvarer som ble brukt for å teste og hente ut ønsket informasjon. I tillegg vil det bli gitt en oversikt over sammensetningen av dataene. Ved å gi en grundig beskrivelse av metodene som ble brukt, vil dette kapittelet gi en klar forståelse av hvordan dataene ble bearbeidet og analysert i denne oppgaven.

3.1 Studieområde

Studieområdet for denne oppgaven befinner seg i den østlige delen av Jotunheimen og består av tre breer: Hellstugubreen, Østre- og Vestre Memurubreen. Disse breene ligger i en høyde mellom 1700 og 2200 meter over havet, og dekker et samlet areal på 16.21 kvadratkilometer (Andreassen, 2022; Andreassen et al., 2022; Breatlas, 2023).

Brenavn	BreID	Posisjon (NØ, UTM33)	Areal (KM ²)
Hellstugubreen	2768	6842840.72, 151911.38	2.67
Østre Memurubreen	2769	6841592.51, 154492.44	5.38
Vestre Memurubreen	2772	6839979.34, 152366.24	8.16

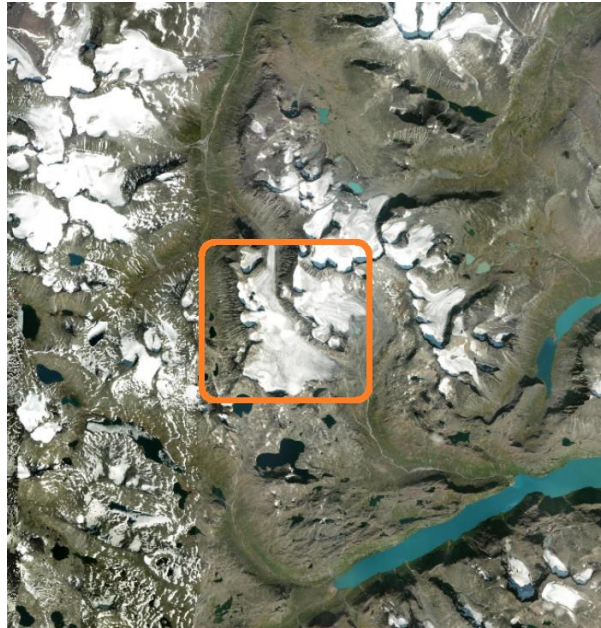
Tabell 1 Brenavn og breID for breene (Andreassen, 2022; Andreassen et al., 2022; Breatlas, 2023). Posisjon hentet fra norgeskart.no, Kartverket

Østre Memurubreen er den største av de tre breene, og ligger på østsiden av Memurutindene. Breen har et areal på 5.38 kvadratkilometer (27.08.2019). Vestre Memurubreen ligger på vestsiden av Memurutindene og har et areal på omtrent 8.16 kvadratkilometer (27.08.2019). Hellstugubreen ligger i en avstikker fra Memurutindene og har et areal på omtrent 2.67 kvadratkilometer (27.08.2019) (Andreassen, 2022; Andreassen et al., 2022; Breatlas, 2023).

Studieområdet har også betydning for turisme og friluftsliv, og er en populær destinasjon for fotturer og brevandring. Området er tilgjengelig via flere populære ruter og hytter, inkludert turisthyttene Memurubu og Spiterstulen.



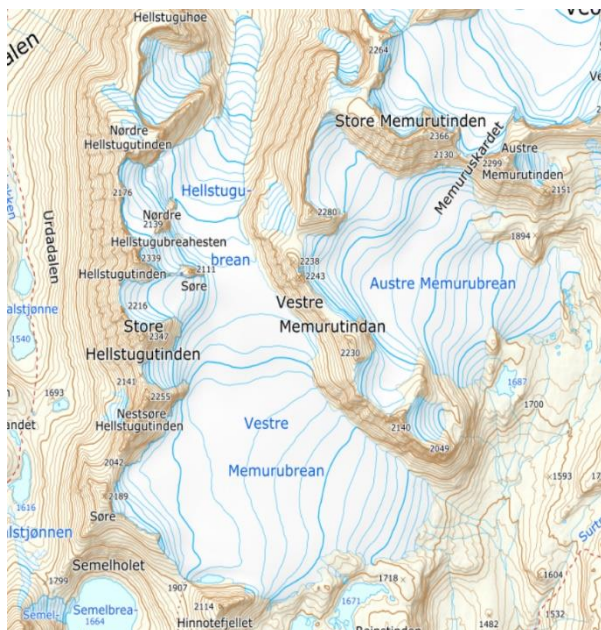
10 Studieområde - Kilde: norgebilder.no (Kartverket)



11 Studieområde - Kilde: norgebilder.no (Kartverket)



12 Studieområde Hellstugubreen samt Østre- og Vestre Memurubreen i Jotunheimen – Kilde: norgebilder (Kartverket)



13 Topografisk kart over Hellstugubreen samt Østre- og Vestre Memurubreen. Kilde: norgebilder.no, Kartverket, Geovekst og kommuner

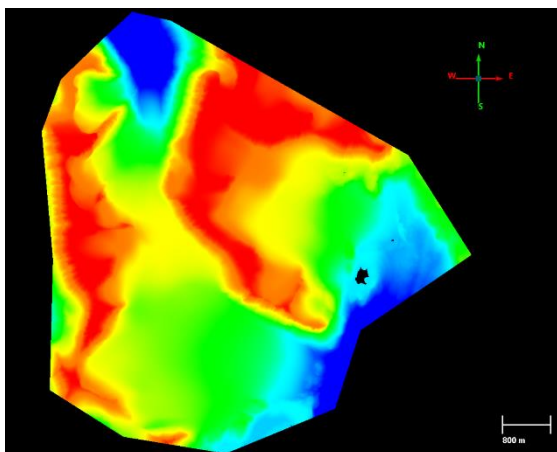
3.1.1 Datasett

TerraTec AS har utført laserskanning over område med NVE som oppdragsgiver. Prosjektet innebar bruk av en Riegl VQ-1560i laser sensor, Trimble Applanix IMU-57 treghetssensor, og Trimble Applanix AV-610 GNSS for å utføre laserskanningen. Datasettet ble prosessert ved hjelp av reflektans for å oppnå

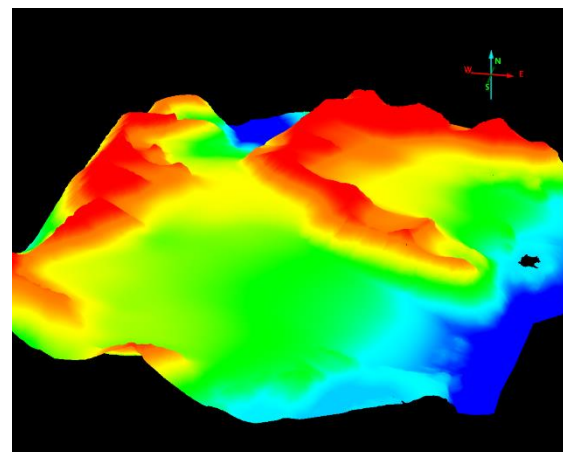
homogene intensitetsverdier for lignende overflater uavhengig av avstand fra skanneren. Flyhøyden var i gjennomsnitt ca. 2200 meter over terrenget, med en gjennomsnittlig punktetthet på 2 punkt per kvadratmeter. Horisontal referanse er Euref89, UTM-sone 32, med vertikal referanse NN 1954. I følge Terra-tec er det ikke gjort kontrollmålinger for å bekrefte nøyaktigheten, men basert på erfaring forventes en vertikal nøyaktighet på under 10 cm og en horisontal nøyaktighet som er 2-3 ganger dårligere (Solli, 2019).

Riegl VQ-1560i	
Spesifikasjon	Verdi
Laserteknologi	Pulsed Waveform Lidar
Pulser per sekund	Opp til 1,333,00
Nøyaktighet i avstandsmåling	+/- 5 mm
Nøyaktighet i vinkelmåling	+/- 0,015 grader
Pulser i luften samtidig	25

Tabell 2 Dataoversikt om Riegl VQ 1560i. (Riegl, 2019)



14 2D punktsky fra Hellstugubreen, Østre- og Vestre Memurubreen, vist i QTM, farget av intensitetsverdier



15 3D punktsky fra Hellstugubreen, Østre- og Vestre Memurubreen, vist i QTM, farget av intensitetsverdier

3.1.2 Landskapet i Østre- Jotunheimen

Østre Jotunheimen ligger i det sentrale Sør- Norge og omfatter et stort område med variert landskap og fjellformasjoner. Området består hovedsakelig av høye fjelltopper, isbreer, dype daler og innsjøer.

Fjellene i Østre Jotunheimen er preget av bratte tinder og skarpe egger, og noen av de høyeste toppene i Norge ligger her. Blant de mest kjente fjellene i området er Galdhøpiggen (2469 moh), Glittertind (2465 moh) og Besseggen (1743 moh). Området har mange store og små isbreer som har målinger helt tilbake

til 50- 60- tallet (Liestøl et al., 1962). Breenes størrelse og form endrer seg over tid, og noen av breene i Østre Jotunheimen har opplevd tilbakegang de siste tiårene. Dette har ført til at enkelte breer har krympet med flere hundre meter siden 1990-tallet ifølge breatlasen til NVE (Breatlas, 2023)



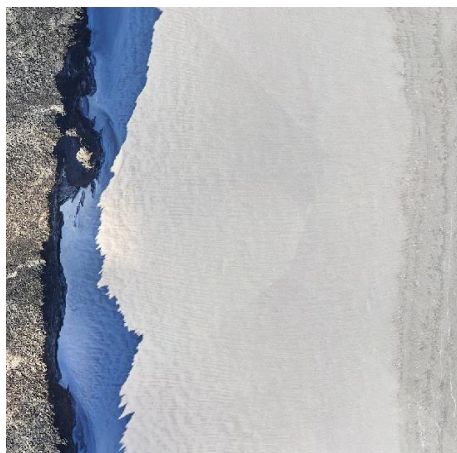
16 Typisk terrengformasjon for Jotunheimen. Breer i Norge kan periodevis ha store mengder snø, og måletidspunkt på året er viktig. Foto: Nils Haakensen (digitaltmuseum)

3.1.3 Overflater på Hellstugubreen, Østre- og Vestre Memurubreen

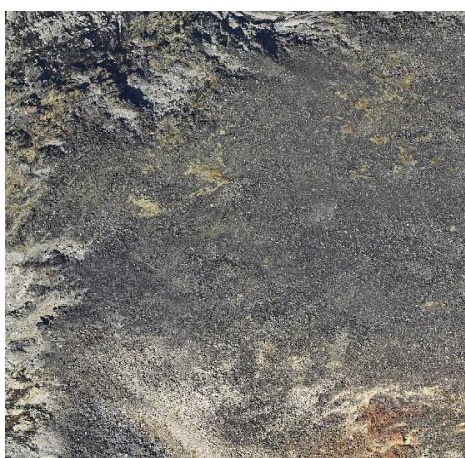
Det er mulig å kategorisere flere typer overflater på isbreene, hvorav fem hovedkategorier identifiseres som berg, isbre, snø, mørk snø og bresprekker. Selv om det kan eksistere underkategorier innenfor hver av disse overflatetyperne, er formålet med denne oppgaven å fokusere på nevnte fem overflatene.



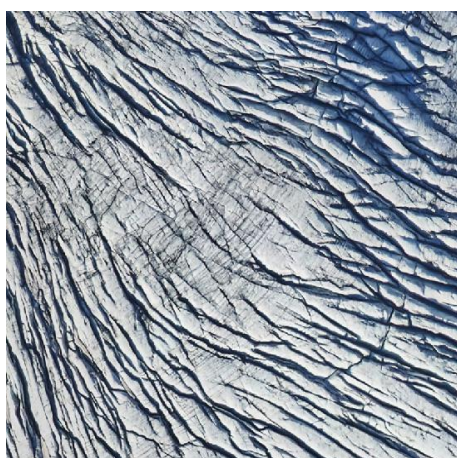
a) Bre



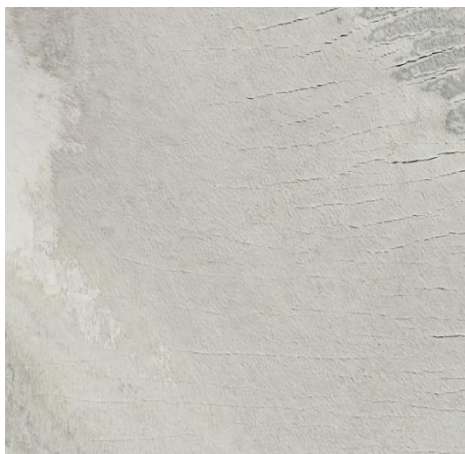
b) Snø



c) Berg



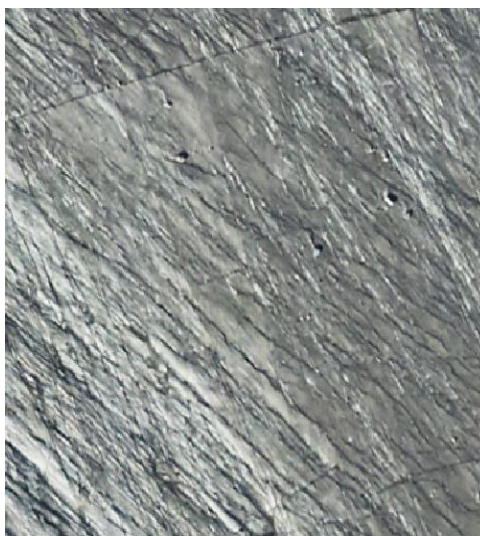
d) Bresprekker



e) Mørk snø

17 Overflatene som skal klassifiseres illustrert med ortofoto levert av NVE.

Isbreen kan også ha overflater med en mer kompleks sammensetning som består av en blanding av de nevnte overflatetyperne. For eksempel viser bilde f) og g) en overflate som består av en isbre under et lag med grus og bergkomponenter, mens bilde h) viser en diffus overgang fra isbre med bergkomponenter til bergoverflate. Bilde j), k) og l) illustrer de ulike bergoverflatene med ulike overflatefarger.



f)



g)



h)

18 f) og g) viser forskjellige typer overflater av bre h) viser en overgang fra bre til berg som kan være utfordrende å registrere visuelt. - Vist med Ortofoto fra NVE



j)



k)



l)

19 j), k) og l) viser forskjellige typer overflater av berg. - Vist med Ortofoto fra NVE

3.2 Programvarer og moduler

3.2.1 Quick Terrain Modeler

Quick Terrain Modeler (QTM) er en programvare fra Applied Imagery som brukes til å behandle store Lidar data, og blant annet produsere digitale terrengmodeller (DTM). Programvaren er utviklet for å håndtere store datamengder.

En av funksjonene til QTM er å produsere rutenettstatistikk som TIFF-filer, hvor hver piksel i TIFF-filen får en verdi basert på en statistisk utregning av en valgt variabel som kan brukes til å analysere og visualisere egenskaper fra laserdataene. Et eksempel er gjennomsnittlig intensitetsverdier innad i hver enkelt rute i rutenettet i en TIFF. QTM bruker en sammensetning av algoritmer for å oppnå dette, og brukeren kan justere parametere for å tilpasse analysen til spesifikke behov.

3.2.2 Catalyst

Catalyst fra PCI Geomatics er en programvare som brukes til å behandle og analysere ulike data, primært bildeanalyse av rasterdata. Den har funksjoner for å utføre raster manipulasjon, sammenslåing av rasterfiler, og visualisere rastere.

3.2.3 QGIS

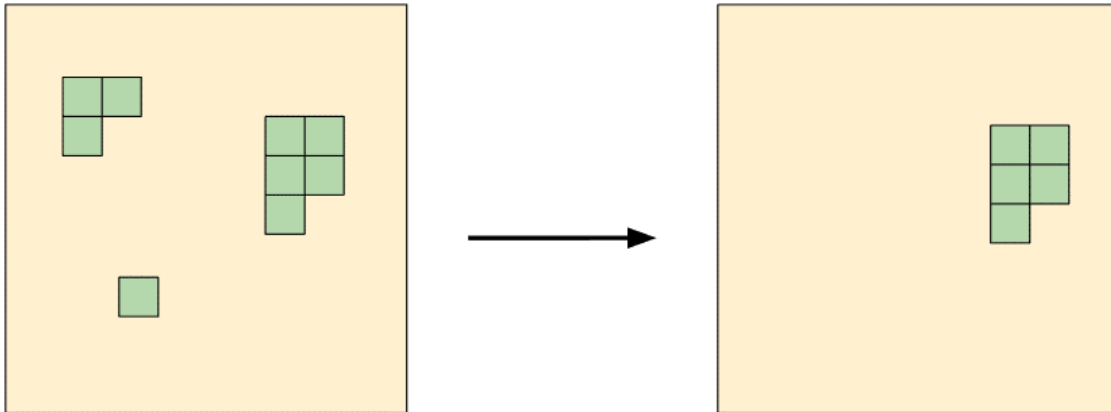
QGIS er en åpen kildekode geografisk informasjonssystem (GIS) som brukes til å visualisere, redigere, analysere og håndtere geografiske data. Programvaren støtter en rekke GIS formater, inkludert ESRI Shapefile, og GeoTIFF. En del av kjernen til QGIS er at det styres av bidragsyttere som er med på å utvikle funksjonaliteten i programmet ved blant annet utvikling av plugins. Plugins er en måte å legge inn tilleggsfunksjonalitet som noen har utviklet og lagt ut for nedlasting i QGIS (*QGIS - The Leading Open Source Desktop GIS*).

3.2.3.1 Dzetsaka Classification tool

Dzetsaka er en QGIS-plugin som kan gjennomføre en semi-automatisk klassifiseringsprosess. Den opprinnelige intensjonen med pluginen var å utføre pikselbasert klassifisering med en Gaussisk-miksmo-
dell, men i senere versjoner har muligheten for Random Forest, KNN og SVM blitt introdusert. Pluginen tar inn data i form av et raster i tillegg til polygoner og punkter for å trene en algoritme. Den trente algoritmen kan deretter brukes til å klassifisere nye områder basert på opplæring fra treningsdata. Pluginen er utviklet av Nicolas Karasiak og benytter Python og Scikit learn biblioteket.

3.2.3.2 GDAL Sieve

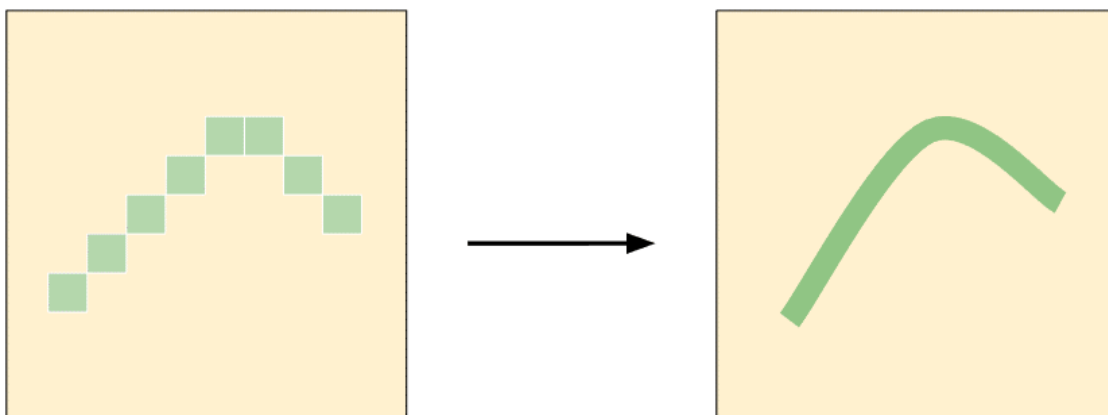
GDAL Sieve er et verktøy i QGIS som brukes til bildebehandling av rasterdatasett. Spesifikt brukes det til å fjerne små «raster-øyer» i et bilde som er mindre enn en gitt terskelstørrelse, og erstatte dem med pikselverdien til det største nabo-polygonet. Terskelstørrelsesparameteren angir størrelsen på de polygonene som skal fjernes.



20 Før og etter Sieve algoritmen er kjørt med 3 i terskelverdi. Øyer med- og under 3 piksler blir slått sammen med omkringliggende rasterverdi

3.2.3.3 Raster til vektor

«Raster til vektor» er en QGIS plugin som lar brukeren konvertere rasterdata til vektordata i form av polygoner. Dette verktøyet lar deg velge formatet på den resulterende vektoren, samt ulike innstillinger for å justere nøyaktigheten og detaljnivået i konverteringen. Når konverteringen er fullført, vil vektoren inneholde geometrier som representerer områder, linjer eller punkter basert på verdien av pikslene i det opprinnelige rasteret.



21 Fra raster(venstre) til vektor(høyre)

3.2.3.4 Gaussian smoothing

Gaussian smoothing er en vanlig bildefiltreringsteknikk som benyttes til å fjerne hørfrekvent støy og glatte ut bilder. Denne teknikken er ofte benyttet som en del av kantdeteksjonsteknikker.

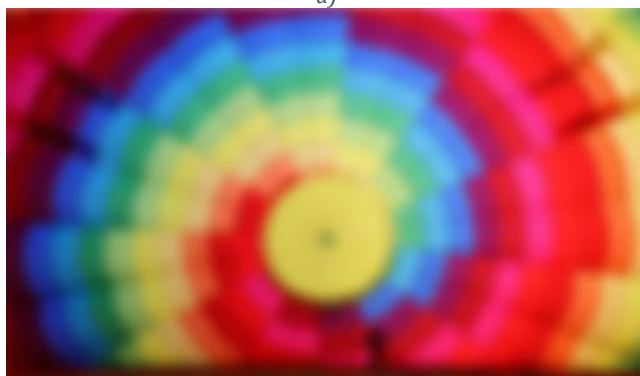
Teknikken fungerer ved at den beregner nye verdier basert på Gaussisk fordeling rundt hver piksel i et raster. Fordi den Gaussiske funksjonen gir høyere vekt til pikslene nærmere midten av filteret og lavere vekt til pikslene lenger unna vil de nære omkringliggende pikslene ha større påvirkning på verdien til de nye pikslene som blir enn de som ligger lenger unna. Dette resulterer i en jevnere og mer avrundet effekt på rasteret, som kan være nyttig for å begrense uønsket støy, og beholde de viktige egenskapene i dataene.

$$G(x, y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Formel 8 x, x: Koordinatene til pikselen, σ : Standardavviket for normalfordelingen



a)



b)



c)

22 a) viser bilde uten Gaussian smoothing, b) viser bilde med 3 i standardavvik, og c) viser bilde med 10 i standardavvik. Illustrasjon: IkamusumeFan (<https://tinyurl.com/IkamusumeFan>)

3.2.4 Python

Python er et høynivå programmeringsspråk, som betyr at det er enkelt å lese og skrive for mennesker, og det har en rekke innebygde funksjoner som gjør det raskt og enkelt å utvikle programmer. Språket har et stort utvalg av biblioteker som gjør det egnet for en rekke bruksområder, inkludert datavitenskap, maskinlæring, og vitenskapelig beregning.

3.2.5 Scikit-learn

Scikit-learn er et Python-bibliotek som brukes til maskinlæring og dataanalyse. Biblioteket inneholder en rekke modeller for å løse ulike typer maskinlæringsproblemer, inkludert lineær regresjon, logistisk regresjon, beslutningstre, k-nearest neighbors, random forest, support vector machines og mer (Pedregosa et al., 2011).

3.2.6 Filformater

TIFF - Tagged Image File Format er en filtype for digitale bilder som brukes til å lagre rastere. Filtypen kan utvides til GEOTIFF, og kan dermed inkludere kartprojeksjon, koordinatsystemer, ellipsoider og datum. I denne oppgaven blir GEOTIFF referert til som TIFF (Ruth et al., 1995; *TIFF Revision 6.0*, 1992).

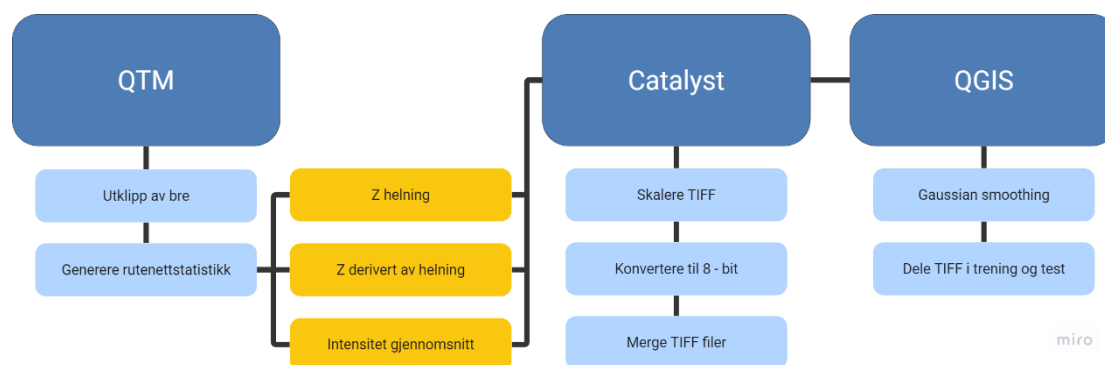
LAS - Lidar Binary er en filtype som brukes i geografiske informasjonssystemer (GIS) og andre programvareverktøy for å lagre og dele geografiske data som er samlet inn ved hjelp av Lidar-teknologi. LAS-filer lagrer disse dataene som en samling av punkter, hver med en geografisk posisjon og en rekke attributter. Attributtene kan blant annet omfatte høyde og intensitetsverdier for hvert punkt. Hensikten med LAS er å gi et åpent format som tillater forskjellige maskinvare- og programvareverktøy for å utveksle punktskydata i et felles format (*LAS Specification 1.4*, 2019).

CSV - CSV står for "Comma-Separated Values", og det er en filtype som brukes til å lagre data i tabellform. En CSV-fil består av tekstdata organisert i rader og kolonner, hvor hver verdi eller celle er separert med et komma. CSV-filer kan opprettes og leses av mange forskjellige programmer, og de brukes ofte til å importere og eksportere data mellom forskjellige programmer (Shafranovich, 2005).

3.3 Preprosessering

Dette kapittelet presenterer en beskrivelse av bearbeidings- og preprosesseringstrinnene som inngår i den overordnede metoden for pikselbasert klassifisering av is, snøtyper, berg og bresprekker på norske isbreer, samt beregning av breomkrets. Hensikten med disse trinnene er å optimalisere forutsetningene for å utvikle en effektiv metode som kan testes.

En betydelig del av maskinlæringsmetoder omfatter forarbeid og preprosessering av data. Det er sjeldent at rådata kommer i den nødvendige formen og strukturen for å oppnå optimal ytelse fra en læringsalgoritme. Derfor er forbehandling av dataene en av de mest avgjørende trinnene innen maskinlæring i følge Raschka og Mirjalili (2019).

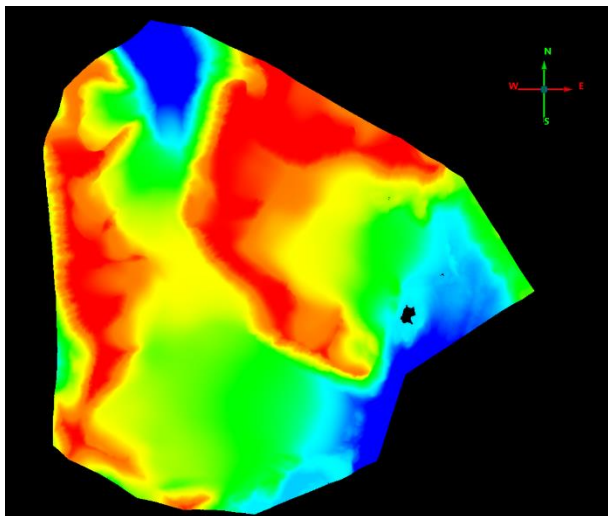


23 Flytskjema som beskriver gangen forarbeidene- og preprosessering av data

3.3.1 Utklipp av breen

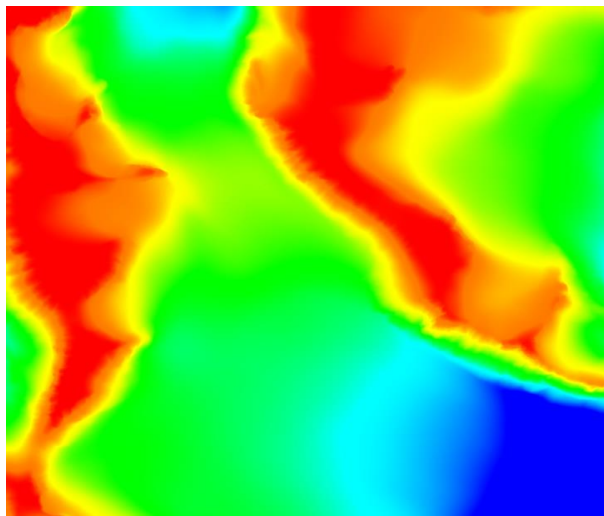
Det første trinnet i preprosessering er å avgjøre å begrense arbeidsområdet til en delvis utklipp av isbreen. Beslutningen om å gjøre dette var basert på flere faktorer. Område med laserdata fra denne breen er ikke helt rektangulær og vil ved produksjon av rutenettstatistikk i QTM bestå av store områder uten data. Ved å klippe et rektangel bestående av kun breen vil dette gjøre at videre arbeid og testing av metoden forenkles og effektiviseres, slik at det ikke trengs å ta hensyn til områder uten data. Samtidig vil utklippet fremdeles inneholde overflater som er representative for hele breen

Punktsky før utklipp:



24 Punktsky fremstilt i QTM i 2D, før utklipp

Punktsky etter utklipp:



25 Punktsky fremstilt i QTM i 2D, etter utklipp

3.3.2 Generere rutenettstatistikk i QTM

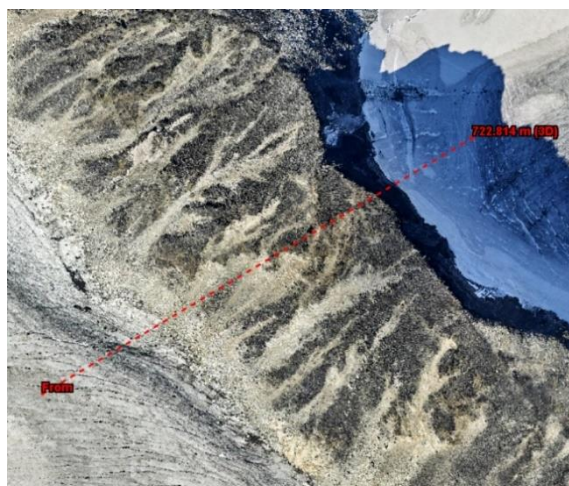
Neste steg i prosessen skjer i QTM, der rutenettstatistikk beregnes ved hjelp av ulike algoritmer i programmet, for å så eksporteres som TIFF. Rutenettstatistikk verktøyet er raskt og visuelt. Det gir brukeren muligheten til å generere TIFF'er med forskjellige sammensetninger av variabler og statistikk på punktdata fra datasettet. Verktøyet fungerer på den måten at brukeren laster inn en punktsky, velger en variabel som skal undersøkes og en statistisk beregning som skal utføres på denne variabelen. Utdata er en TIFF fil med pikselverdier i hver rute basert på statistikken som er påført variabelen.

Variablene som kan velges mellom er hovedsakelig Z (høyde), intensitet, antall punkter, punkttetthet, samt alle tilgjengelige punktattributter som kan lagres i en las-fil. Statistikk som kan beregnes for hver rutenettcelle inkluderer minimum, maksimum, gjennomsnitt, rekkevidde, avvik, helning, aspekt og andre ordens helning. Forklaring av aktuelle statistikk beregninger er presentert i tabell 3.

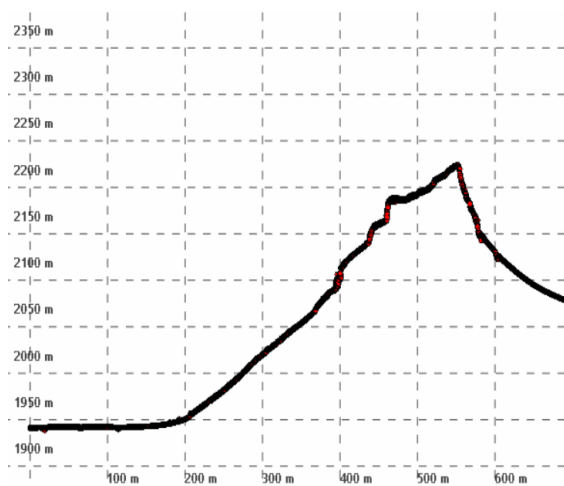
Statistikk	Forklaring
Gjennomsnitt	Summen av variabelverdiene i en rutenettcelle delt på antall punkter i den cellen.
Helning	Helning er en vektet gjennomsnittsorientering representert i grader opp fra horisontalplanet. For å beregne helningsvinkelen, beregnes først en gjennomsnittsverdi av variabelen (f.eks. Z) i hver celle. Deretter etableres en normalvektor til hver av de 8 "nabocellene" og en vektet gjennomsnittsnormal vektor for hver celle. Helning er den vertikale vinkelkomponenten av denne vektete gjennomsnittsnormalvektoren.

Andre ordens	Tilpasses en kvadratisk funksjon til det lokale terrenget i hver celle av rutenettet. Deretter finnes hellingen til tangenten til denne funksjonen i hvert punkt, og dette brukes til å beregne "andre ordens" verdien for hver celle.
--------------	--

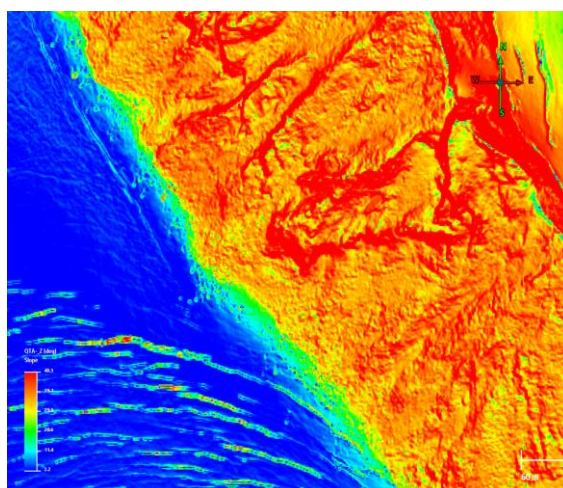
Tabell 3 Forklaring på statistikk utregninger i QTM. (Hentet fra hjelpesidene til Applied Imagery)



26 Tverrsnitt av hellende fjell



27 Visualisering av tverrsnitt fra hellende berg

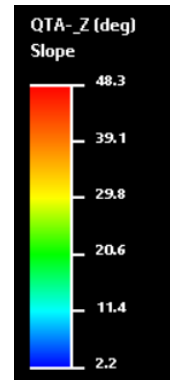
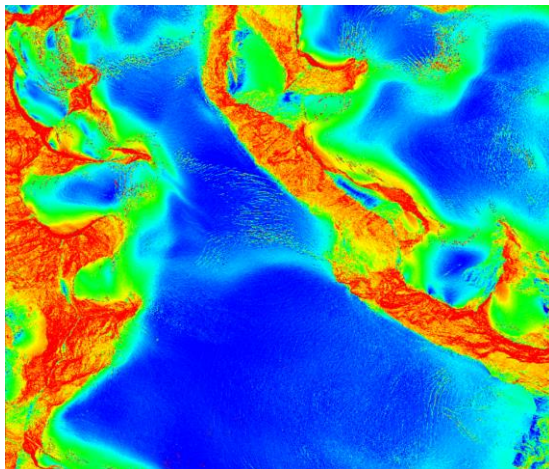


28 Eksempel på generert grid statistikk på variabel Z og statistikk helning som DEM i Quick terrain modeller

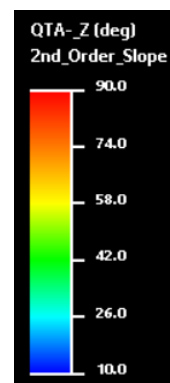
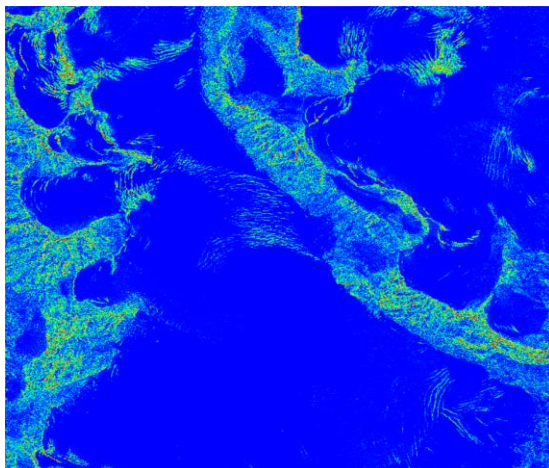
3.3.3 Valg av variabler og statistikk

For å oppnå optimal utnyttelse av algoritmene er det avgjørende å velge de riktige variablene og statistikkene. I denne delen av prosessen velges de variablene og statistikkene som gir mest mulig informasjon om de forskjellige egenskapene til breen. Det er tre variabler og tilhørende statistikker som anses som sannsynlige for et godt resultat med pikselbasert klassifisering av breen, og har blitt valgt i denne oppgaven. To av disse baserer seg på høydeverdier, med statistikk helning og andre grads helning. Den siste variabelen gir informasjon om gjennomsnittlig intensitetsverdi.

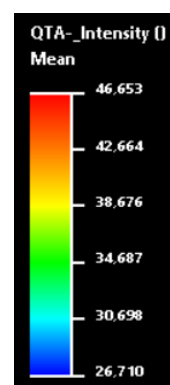
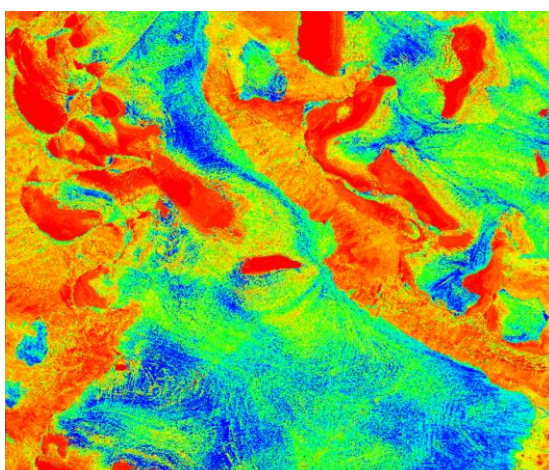
Variabler og tilhørende statistikker er utvalgt på bakgrunn av visuell og statistisk analyse i QTM, og rutenettstatistikkene som ble valgt skilte seg tydelig ut i å kunne differensiere ulike overflater på breen.



29 a) Visualisert rutenettstatistikk med variabel: Z og statistikk: helning



30 b) Visualisert rutenettstatistikk med variabel: Z og statistikk : andre grads helning



31 d) Visualisert rutenettstatistikk med variabel: intensitet, og statistikk: Gjennomsnitt

Variabel	Statistikk	Minimum	Maksimum
Z	Helning	0.00687147	89.9248
Z	Andre grads av helning	0	169.556
Intensitet	Gjennomsnitt	15378	65535

Tabell 4 Minimum og maksimumsverdi på rutenettstatistikken før konvertering.

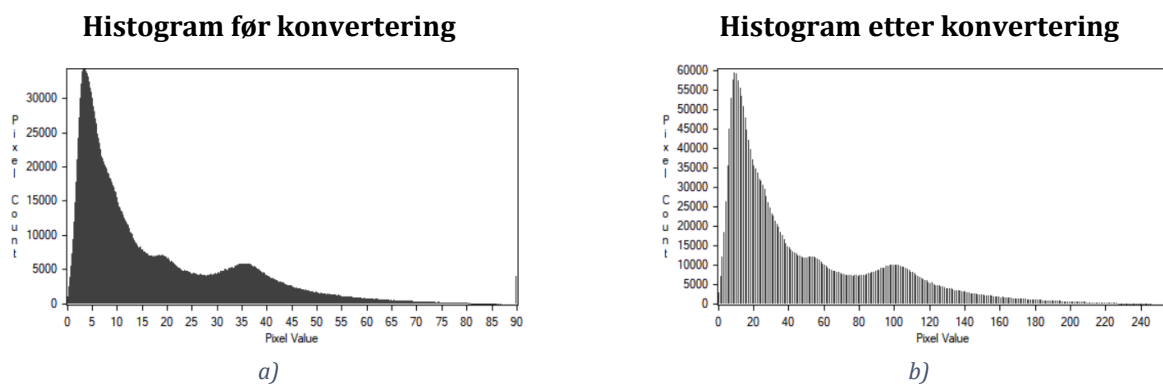
For å kunne gjennomføre ytterligere analyser av isbreen, konverteres og eksporteres rutenettstatistikken til en TIFF. Ved å konvertere rutenettstatistikken til en rasterfil, hvor hver piksel inneholder informasjon basert på rutenettstatistikens utdata, vil det være mulig å utføre videre analyser og beregninger.

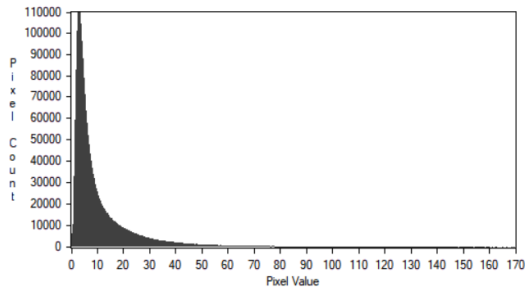
3.3.4 Konvertere og slå sammen data i Catalyst

For at informasjonen i hvert raster-bånd fra QTM skal veie likt i opplæringen av maskinlæringsmodeller, skal de skaleres. Hver TIFF-fil som kommer ut av QTM vil ha ulike størrelsesforhold og ulike verdier av intensitets- og Z-verdier. Intensitetsverdiene varierer mellom 0 og 65536, mens Z-verdiene varierer mellom fra minste verdi på 0 til høyeste verdi på 169.556. Dette kan føre til skjevhet i opplæringen av modellene, og påvirke resultatet.

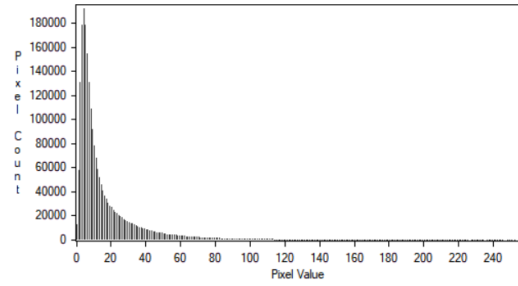
Dermed benyttes rasterkalkulatoren i Catalyst for å skalere dataene i hvert rasterbånd fra 16- til 8-bits format og dermed begrense verdien til en skala fra 0 til 255. Dette kan gi bedre datagrunnlag i klassifiseringen.

Til slutt blir de tre TIFF-filene slått sammen til en enkelt tre-båndes TIFF-fil ved hjelp av programvarens funksjon for å slå sammen filer, slik at de kan brukes i pikselbasert klassifisering.

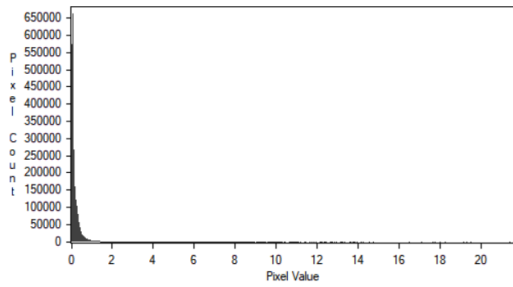




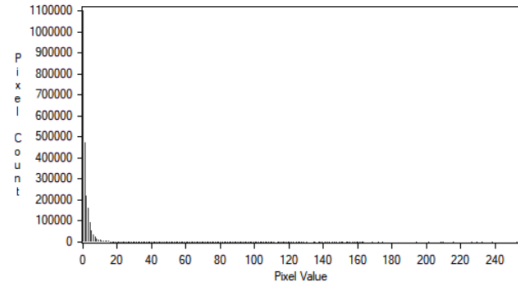
c)



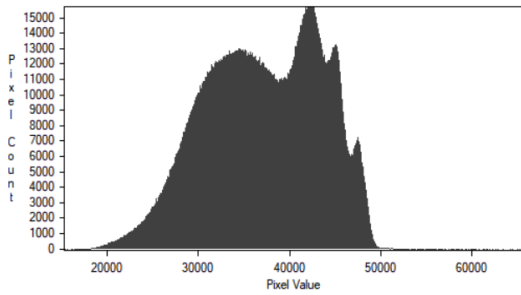
d)



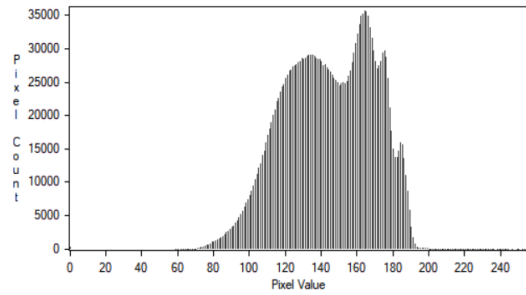
e)



f)



g)



h)

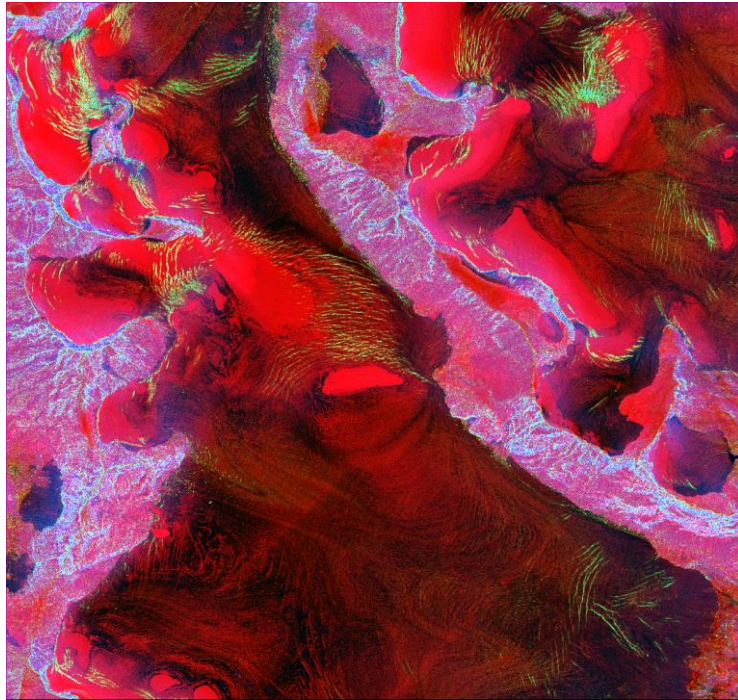
32 Histogram før etter skalering

Bånd	Variabel	Statistikk	Minimum	Maksimum
1	Z	Helning	0	235
2	Z	Andre grads helning	0	242
3	Intensitet	Gjennomsnitt	65	255

Tabell 5 Nye minimum og maksimum verdier for hvert bånd

Sammenslått TIFF		
Bånd	Variabel	Statistikk
Rød	Intensitet	Gjennomsnitt
Grønn	Z	Helning av andre grad
Blå	Z	Helning

Tabell 6 Den sammenslåtte TIFF filen og tilhørende farge i Catalystr



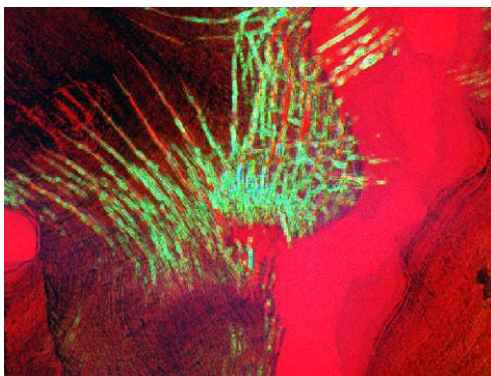
33 Sammenslått TIFF med båndene med Z helning, Z andre grads helning og intensitet gjennomsnitt er illustrert med henholdsvis rød, grønn og blå i Catalyst.

3.3.5 Gaussian smoothing

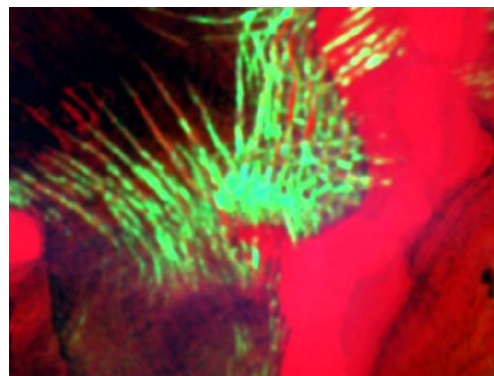
Gaussian smoothing blir påført bilde for å redusere høyfrekvent støy imens lave frekvenser bevares og glatte ut detaljer. Dette kan bidra til å redusere effekten av små feil eller unøyaktigheter i bildet som kan påvirke resultatene av klassifiseringsalgoritmen. Ved å glatte ut detaljer i bildet kan algoritmen fokusere på de større og mer signifikante egenskapene som er relevante for klassifiseringen. Det er imidlertid viktig å finne riktig balanse mellom glatting av bildet og bevarelsen av de viktige egenskapene som er nødvendige for å korrekt klassifisere bildene.

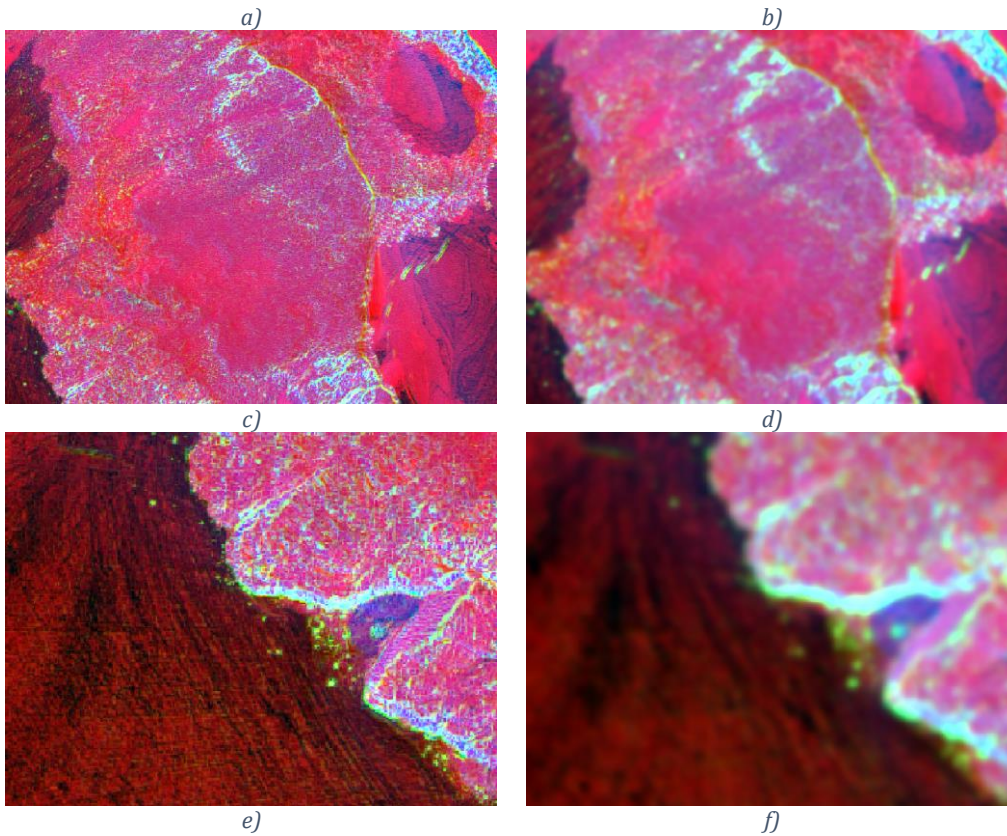
Valget av det endelige standardavviket ble bestemt gjennom en iterativ utforskning av hva som ga det beste resultatet intuitivt, og endte opp med å bli satt til 2.

Før Gaussian smoothing



Etter Gaussian smoothing



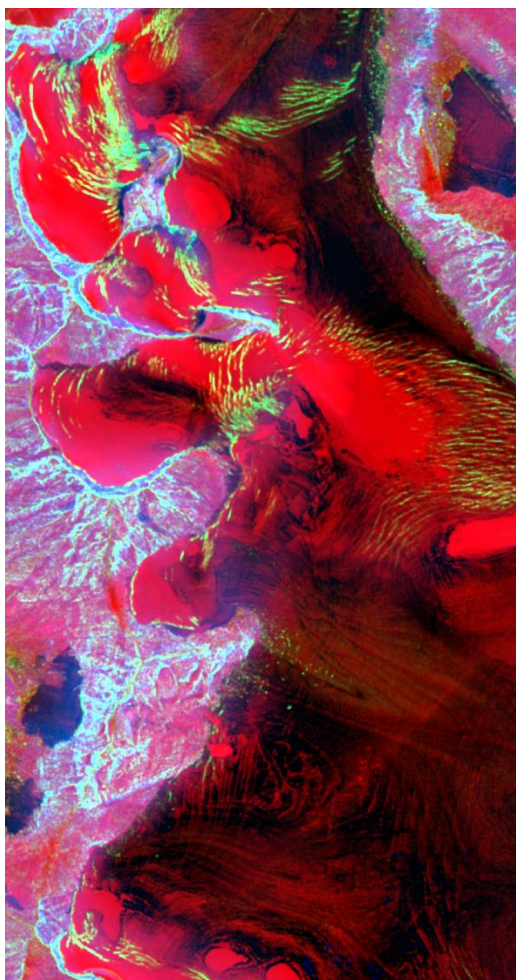


34 Utklipp fra den sammenslåtte filen før og etter Gaussian smoothing er påført

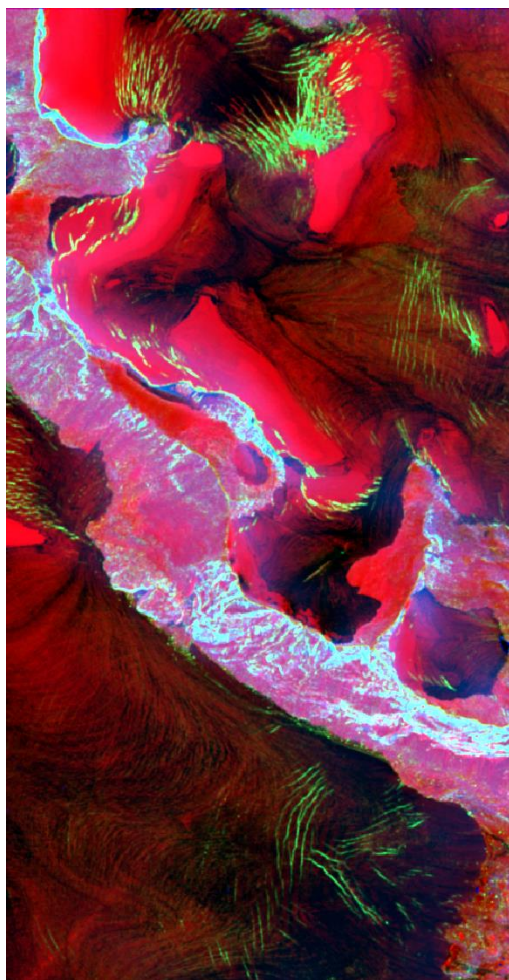
3.3.6 Trening- og test sett

For å gjennomføre trening og testing av klassifiseringsmodellene, blir området delt i to deler. Den høyre delen av bildet er satt av til trening, mens den venstre delen brukes til testing. Formålet med å dele datasettet inn i en treningsdel og en testdel er å evaluere ytelsen til klassifiseringsmodellen. Treningsdelen brukes til å lære modellen å klassifisere riktig, mens testdelen brukes til å evaluere modellens nøyaktighet og ytelse på nye og uprøvde data. Når en modell trenes på et datasett, vil den kunne utføre oppgaver på en nøyaktig måte på dette datasettet, men den vil sannsynligvis ikke fungere like bra på nye data hvis den kun er tilpasset det spesifikke datasettet. Derfor er det nødvendig å teste modellen på en uavhengig testdel for å evaluere dens evne til å generalisere til nye data. Resultatene fra testdelen vil gi en indikasjon på modellens ytelse og evne til å fungere på "virkelige" data.

Test



trening



35 Trening- og testområde illustrert med visualisert laserdata

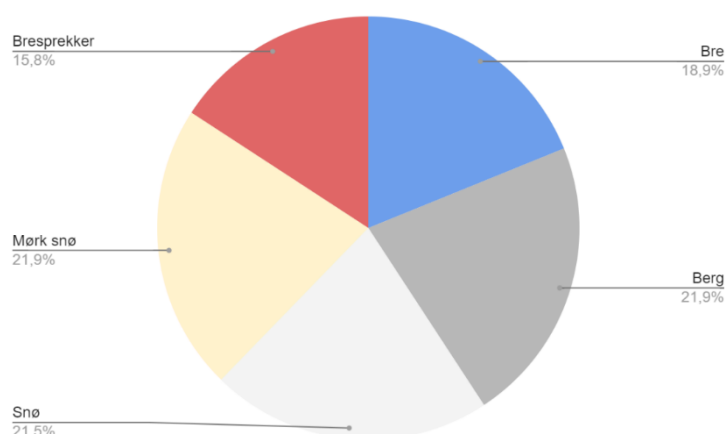
3.4 Veiledet klassifisering

3.4.1 Treningspunkter

Ved veiledet pikselbasert klassifisering er valget av treningsdata en viktig faktor som påvirker nøyaktigheten til modellen (Merry et al., 2023) . Treningspunkter har blitt valgt ut visuelt basert på ortofoto levert av NVE tatt samme tidspunkt som laserdataene er innhentet. Totalt antall treningspunkter er 986, og sprer seg utover treningsområdet med en prosentvis fordeling med 18.9% bre, 21.5% snø, 21.9% mørk snø, 21.9% berg og 15.8% bresprekker.

Klasse	Antall punkter
Bre	186
Berg	216
Snø	212
Mørk snø	216
Bresprekker	156
<u>Totalt</u>	<u>986</u>

Tabell 7 Antall treningspunkter som blir brukt for å trene opp maskinlæringsmodellen



36 Prosentvis fordeling av treningspunkter

Ved opprettelse av treningspunkter har det blitt tatt hensyn til følgende faktorer: Jevnt fordelte klasser i treningsdataene bidrar til å unngå skjevhet i modellens læring og kan øke nøyaktigheten i klassifiseringen. Ved å ha jevnt fordelte klasser vil modellen lære å skille mellom de ulike klassene på en mer balansert måte.

Tilstrekkelig med treningspunkter kan være avgjørende for å lære modellen å gjenkjenne og skille mellom ulike mønstre og egenskaper i bildene på nøyaktig måte. Hvis antall treningspunkter er for lavt, vil modellen ikke kunne lære tilstrekkelig om variasjonene og mønstrene som finnes i bildene, og vil dermed kunne gjøre feilaktige klassifiseringer når den møter nye bilder. På den annen side kan å bruke for mange treningspunkter føre til overtilpasning, og det er fare for at modellen ikke generalisere nok på usette data.

Det bør være variert utvalg av treningsdata fra hele treningsdatasettet. Ved å velge treningsdata fra et utvalg som inkluderer ulike terreng og områder, kan modellen lære å håndtere varierte situasjoner og

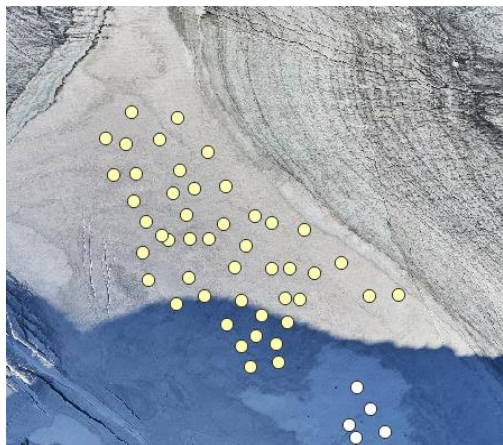
gjøre mer nøyaktige prediksjoner. Dette bidrar til å gjøre modellen mer allsidig og i stand til å generalisere bedre.



a) Blå punkter representerer treningsgrunnlag for berg



b) Hvite punkter representerer treningsgrunnlag for snø



c) Gule punkter representerer treningsgrunnlag for mørk snø



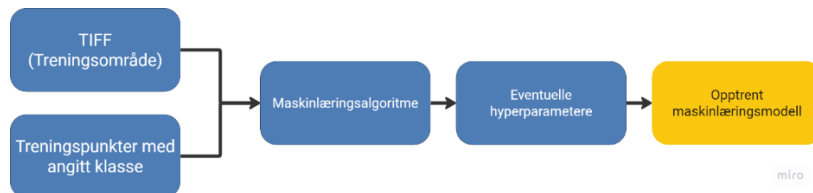
d) Svarte punkter representerer treningsgrunnlag for berg



37 Røde punkter representerer treningsgrunnlag for bresprekker

3.4.2 Klassifisering

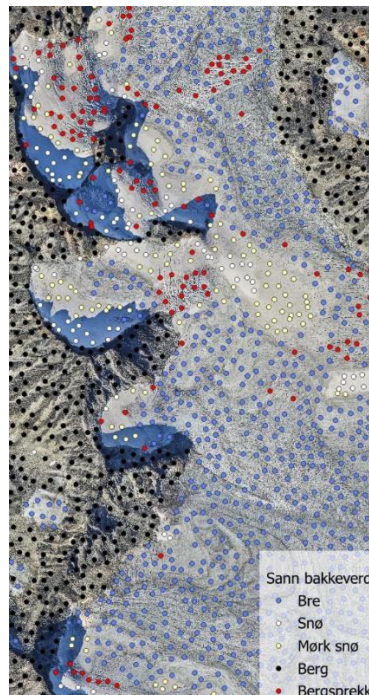
Med treningsdatasettet i form av TIFF og treningspunkter kan det gjennomføres trening av maskinlæringsmodeller som kan klassifisere usett datasett. Treningen av modellene utføres i QGIS ved hjelp av Dzetsaka-pluginnet utviklet av Nicolas Karasiak. Pluginnet tar treningspunkter som input, disse punktene er merket med en spesifikk klasse i form av heltall, i tillegg til et input-raster som modellen skal trenes på og et valg av algoritme. Utdata er en opptrent maskinlæringsmodell som kan benyttes til å klassifisere usett data.



38 Flytskjema i klassifiseringsprosessen

3.4.3 Testpunkter

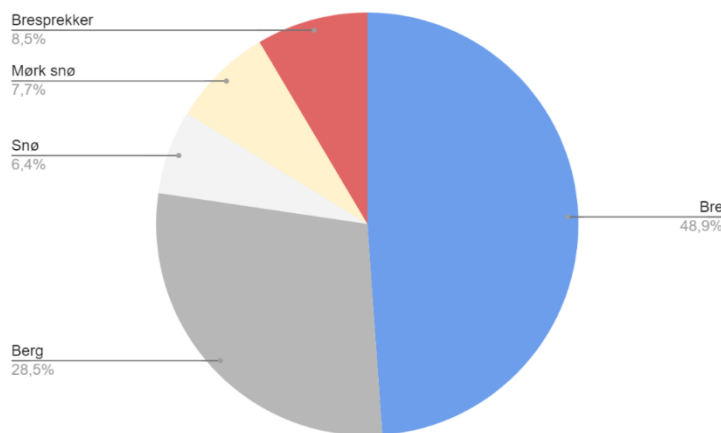
For å evaluere modellen blir det opprettet 1 500 punkter innenfor testområde med sann verdi basert på ortofoto fra NVE. Når de opptrente maskinlæringsmodellene har klassifisert testområdet, ekstraheres rasterverdier fra klassifiseringen, og sammenlignes med sann verdi. Ut ifra dette kan vi sammenligne de predikerte verdiene med sanne verdier, og beregne en forvirringsmatrise og tilhørende valideringsparametere i Python.



39 Testpunkter med sann verdi basert på ortofoto fra NVE

Klasse	Antall
Bre	733
Snø	96
Mørk snø	116
Berg	427
Bresprekker	128
Totalt	1 500

Tabell 8 Oversikt over antall tilfeldige testpunkter i område



40 Prosentvis fordeling av testpunkter

3.5 Brekant utledet av pikselbasert klassifisering av laserdata

Problemstillingen inkluderer evaluering av utledning av brekant basert på pikselbasert klassifisering av laserdata. Kvaliteten på utledning av brekant avhenger av resultatet av klassifiseringen, men det finnes metoder for å forbedre klassifiseringsresultatene i etterkant av klassifiseringen. Dette kan føre til økt nøyaktighet når det kommer til å identifisere brekant.

3.5.1 Sammenslå klasser tilknyttet breområde

For å utlede brekant ved hjelp av klassifiseringsresultatet, er det første trinnet å kombinere de ulike klassene som befinner seg på breen. Klassene for snø, mørk snø, bresprekker og selve breen kombineres til en enkelt klasse ved hjelp av rasterkalkulator i QGIS. Som et resultat av denne operasjonen vil pikslene som tilhører breområdene ha en tilhørende rasterverdi på 1, mens pikslene som tilhører bergområdene vil ha en verdi på 0. Dette gir et bilde med to rasterverdier som gir et klart skille mellom overflatene som befinner seg på breen, og det som utgjør bergoverflater rundt.



41 Resultat av reklassifisering

3.5.2 GDAL Sieve

I klassifiseringsteknikker som er brukt i denne oppgaven kan det oppstå støy i form av små klynger med piksler som er feilklassifisert. Dette kan føre til unøyaktigheter i overgangen fra raster til vektor, og det er derfor viktig å minimere støyen for å oppnå best mulig resultat ved utledning av brekant. For å redusere støyen i rasterbildet benyttes GDAL-verktøyet "Sieve", som fjerner uønskede pikseloverflater i rasterbildet som dukker opp innenfor en annen klasse. Dette resulterer i et renere bilde, som kan føre til en mer nøyaktig kartlegging av brekanten.

3.5.3 Raster til vektor

Det endelige reklassifiserte bildet, som har blitt etterprosessert brukes i en raster til vektor algoritme i QGIS. Algoritmen konverterer det reklassifiserte bildet til georefererte vektorer, og vil generere vektorer som befinner seg mellom to klasser, og i vårt tilfelle, mellom klassen som er klassifisert som bre, og klassen som er klassifisert som berg.

3.6 Test av algoritmene

3.6.1 Statistisk test av algoritmene

Under gjennomførelsen av den statistiske testen ble verdien fra rasteret for hvert testpunkt i det klassifiserte TIFF-bildet ekstrahert og sammenlignet med sann verdi i testpunktene. En forvirringsmatrise og tilhørende valideringsparametere ble deretter generert for å evaluere modellene mot hverandre. Forvirringsmatrisen, viste antall korrekte og feilaktige klassifiseringer av hvert testpunkt og hvilke klasser som ble feilklassifisert mot hverandre, og hvor ofte modellene treffer riktig klasse. Basert på dette ble flere evalueringsparametere av algoritmens ytelse beregnet, inkludert presisjon, gjenkall og F1-verdi, og disse vil bli diskutert.

3.6.2 Visuell analyse av algoritmene

I tillegg til å gjennomføre en statistisk test av algoritmene, utføres det også en visuell analyse for å kompensere for usikkerheten som oppstår som følge av ufullstendig dekning av testpunkter i testområdet. Den visuelle analysen sammenligner resultatet fra klassifisering for hver algoritme med ortofoto levert fra NVE. Ved å sammenligne resultatene fra algoritmene med ortofotoet kan man støtte den statistiske testen og oppdage eventuelle avvik og feil som ikke ble fanget opp. Dette kan bidra til å øke påliteligheten og nøyaktigheten av resultatene.

Det visuelle aspektet gir også en mer intuitiv forståelse av algoritmens ytelse og kan hjelpe til med å identifisere eventuelle mønstre eller trender i dataene som ikke nødvendigvis kommer frem i den statistiske testen alene. Samlet sett gir en kombinasjon av statistiske og visuelle tester et mer komplett bilde av algoritmenes ytelse og nøyaktighet.

3.6.3 Test av utledning av brekant

Test av utledning av brekant innebærer en sammenligning av ortofoto fra NVE, med den genererte brekanten utledet fra klassifiseringsresultatet fra den best presterende algoritmen. Ortofotoet gir et høyoppløselig bilde av området som blir undersøkt, og gir dermed en god referanse for å evaluere nøyaktigheten av den genererte brekanten. Sammenligningen gjøres ved å visuelt inspisere områdene der vektoren for brekanten er generert. Dette gir en intuitiv forståelse av hvor nøyaktig den genererte brekanten er, og hvor eventuelle feil eller unøyaktigheter oppstår.

4 Resultater

Dette kapittelet inneholder resultatene som er oppnådd ved å følge stegene som er presentert i metodekapittelet. Først presenteres resultatene oppnådd med RF, SVM, KNN og GMM fra den statistiske testen. Deretter presenteres resultater fra visuell analyse av klassifiseringen, og til slutt utledning av brekant. En nærmere analyse og diskusjon av alle resultatene kommer frem i kapittel 5.

4.1.1 Statistisk test og valideringsparametere

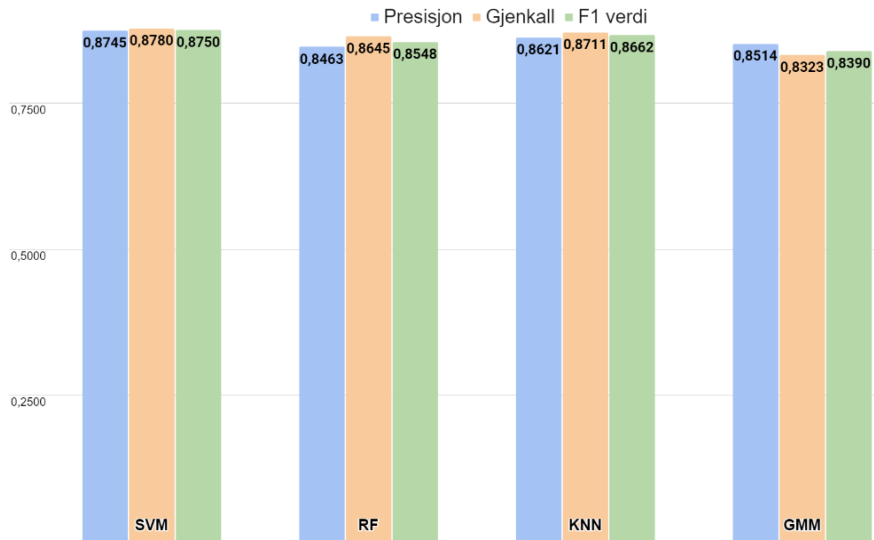
Maskinlæringsalgoritmene ble vurdert ved hjelp av forvirringsmatrise og valideringsparametere beregnet fra denne. Resultatene viser at SVM-algoritmen har høyest presisjon og gjenkall, med verdier på henholdsvis 0,8745 og 0,8780, i tillegg til den høyeste F1-verdien på 0,8750.

KNN- algoritmen har en presisjon på 0,8621, en gjenkall på 0,8711 og en F1-verdi på 0,8662. Dette indikerer at KNN hadde noe lavere ytelse enn SVM, men bedre enn RF og GMM.

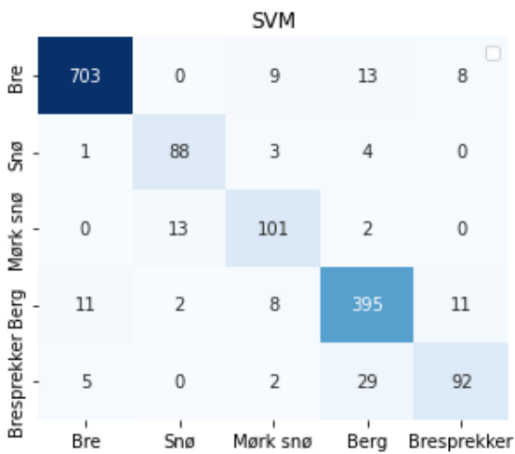
RF oppnådde en presisjon på 0,8463 og gjenkall på 0,8645, som resulterer i en F1-verdi på 0,8548. Dette indikerer at RF-modellen hadde noe lavere ytelse enn SVM og KNN-modellen, og høyere enn GMM.

GMM algoritmen har den laveste presisjonen og gjenkall, med henholdsvis 0,8514, og 0,8323, som resulterer i en F1-verdi på 0,8390. Dette indikerer at GMM algoritmen presterer lavest i den statistiske testen.

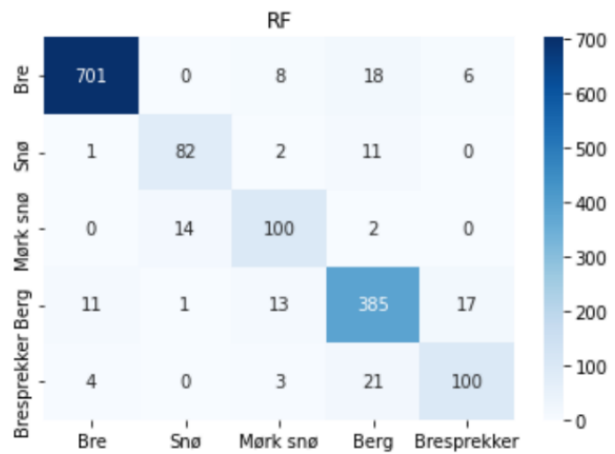
Basert på resultater fra den statistiske testen blir SVM valgt til å utlede brekant, ettersom denne modellen hadde høyest ytelse på datasettet.



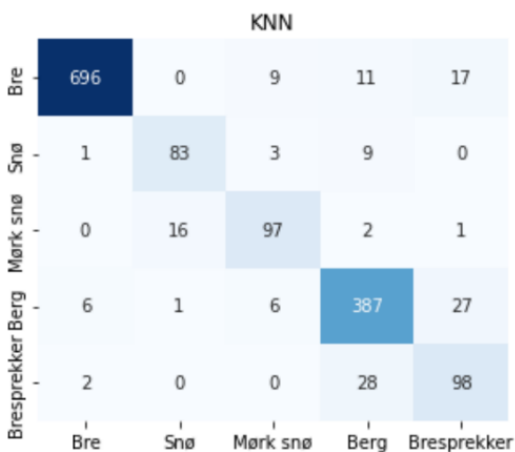
42 Valideringsparametere for algoritmene



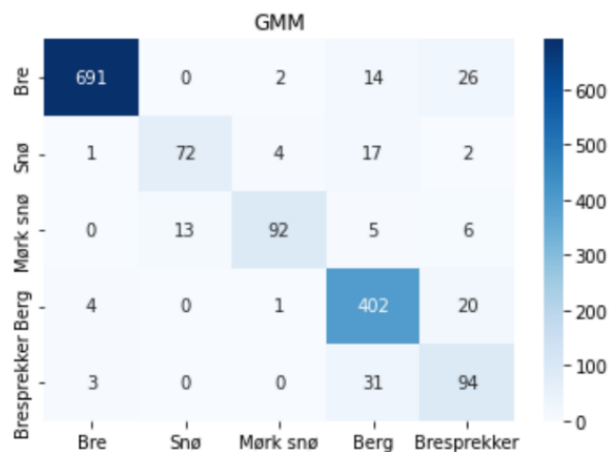
43 Forvirringsmatrise for SVM



44 Forvirringsmatrise for RF



45 Forvirringsmatrise for KNN



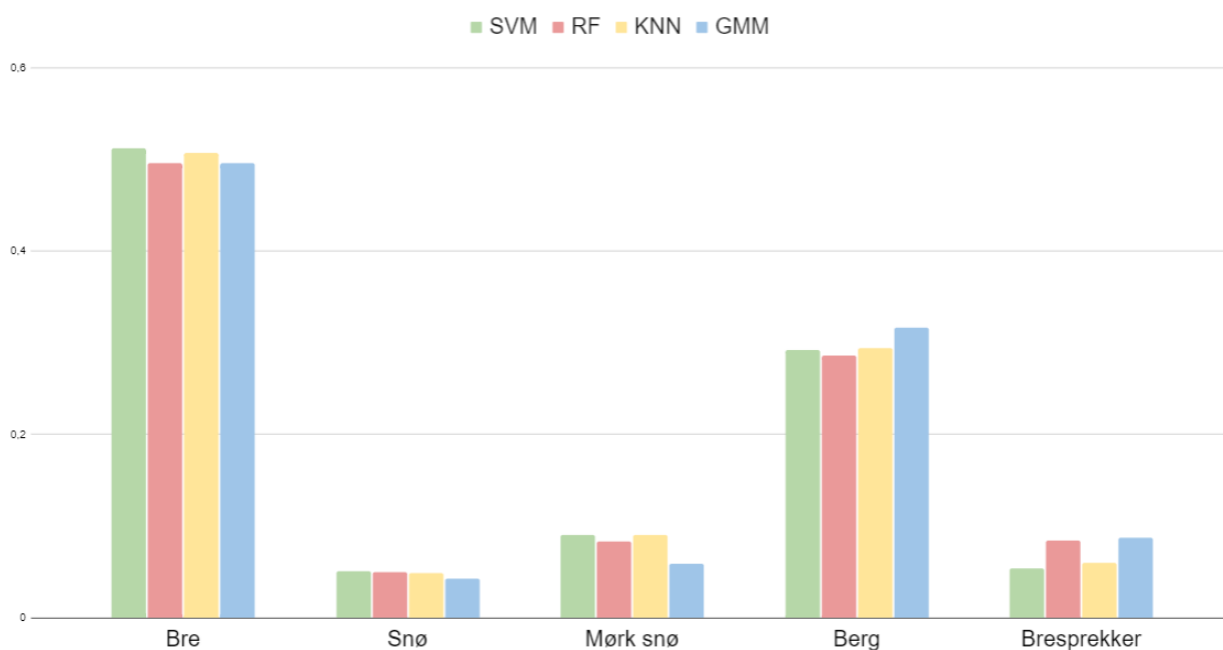
46 Forvirringsmatrise for GMM

4.1.2 Prosentvis fordeling i klassifisert areal

Resultatene presentert i Tabell 10 viser prosentfordelingen av klassifisert areal for hver av de fem kategoriene (bre, snø, mørk snø, berg og bresprekker) for hver av de fire maskinlæringsalgoritmene (GMM, KNN, RF og SVM). Selv om det er små forskjeller mellom algoritmene, er det to kategorier som skiller seg betydelig ut fra de andre. Hvor de største forskjellene kan sees på klassen bresprekker og mørk snø. I klassen mørk snø skiller det 3.3% mellom SVM og KNN. Videre ser vi at SVM klassifiserer mindre områder som mørk snø i forhold til de andre modellene, 3.16% mindre enn RF som klassifiserer mest areal som mørk snø.

	GMM	KNN	RF	SVM
Bre	51,22%	49,62%	50,69%	49,56%
Snø	5,10%	5,02%	4,88%	4,22%
Mørk snø	9,02%	8,32%	9,03%	5,87%
Berg	29,24%	28,60%	29,44%	31,60%
Bresprekker	5,41%	8,45%	5,97%	8,75%

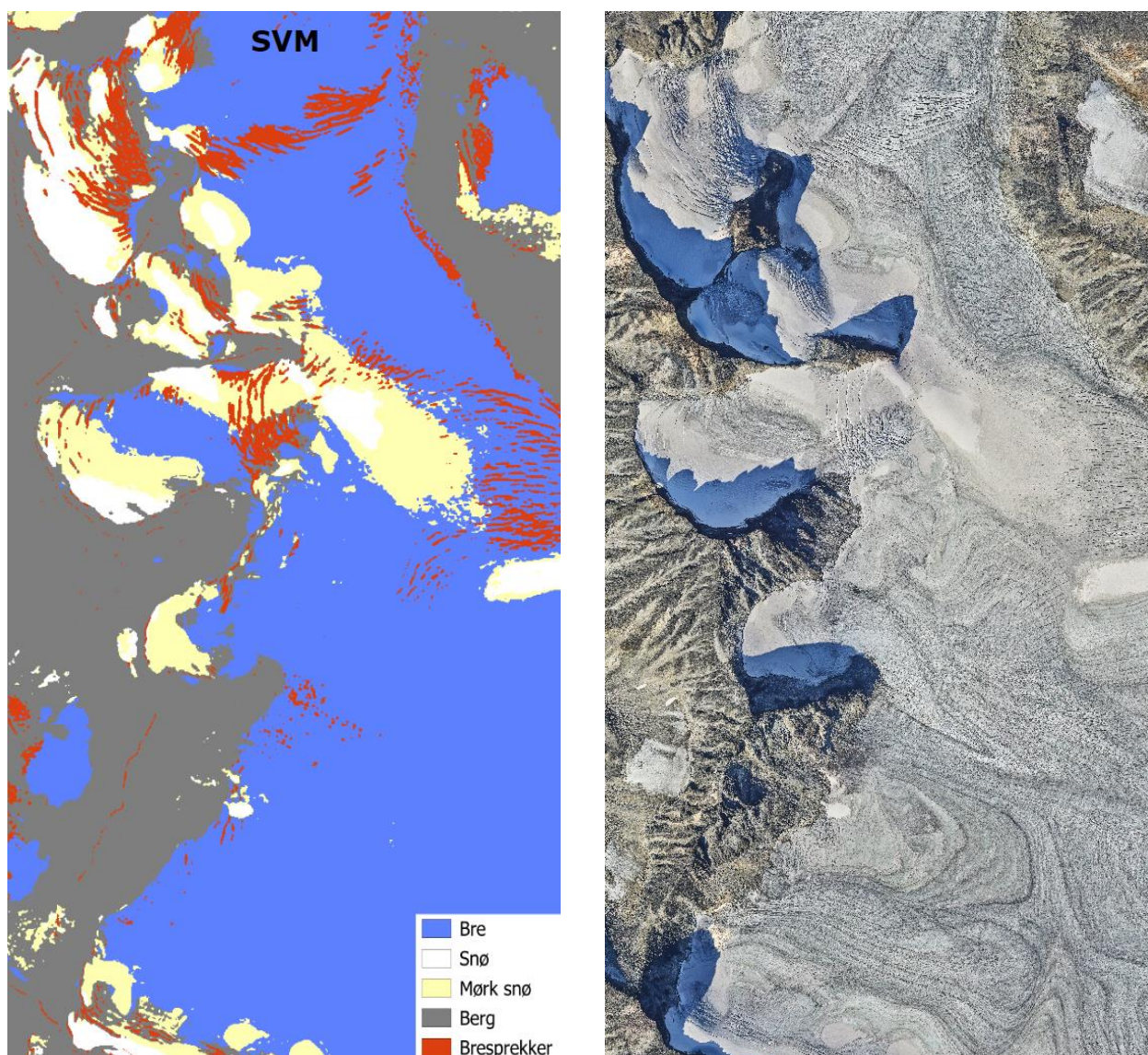
Tabell 9 Prosentvis oversikt over klassifisert areal for hver modell



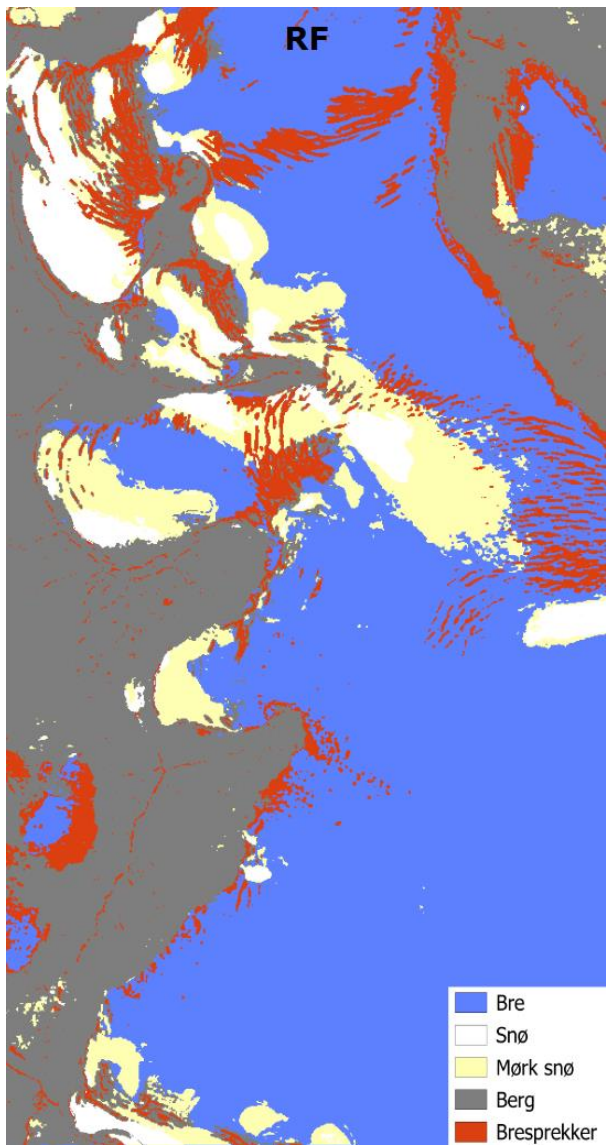
47 Prosentvis fordelingen av klassifisert areal for de ulike algoritmene

4.1.3 Resultater visuell analyse

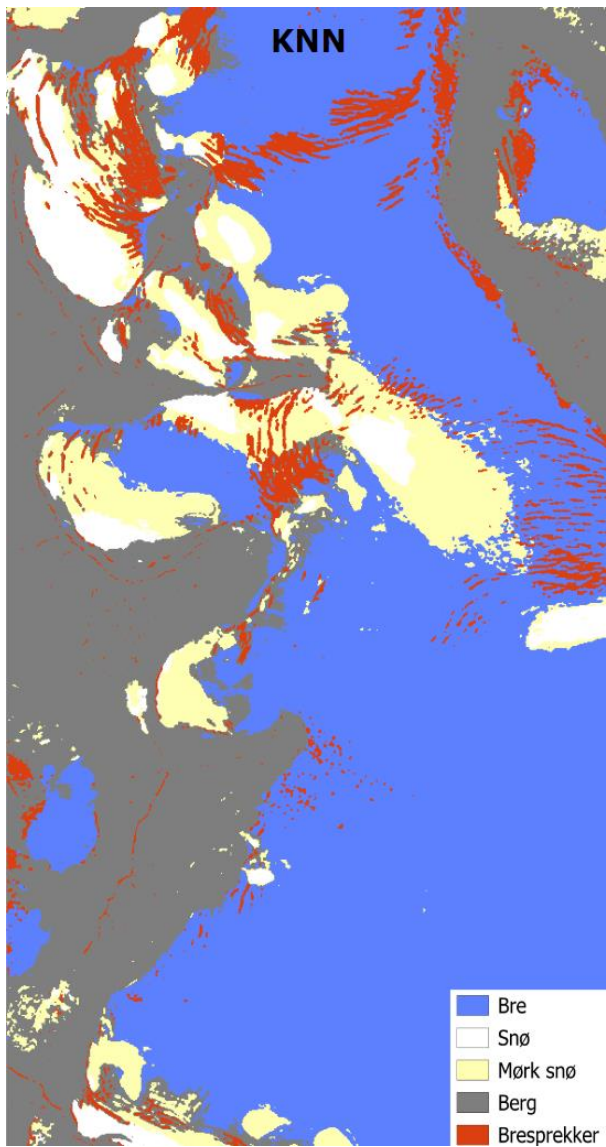
Figurene 47 til 50 presenterer resultatene av pikselbasert klassifisering med flybåren laserdata som datagrunnlag for opptrening av de fire maskinlæringsalgoritmene SVM, RF, KNN og GMM, sammenlignet med ortofoto fra NVE. Disse resultatene er produkter av metoden beskrevet i metodekapittelet, og vil bli nærmere analysert og diskutert i kapittel 5.



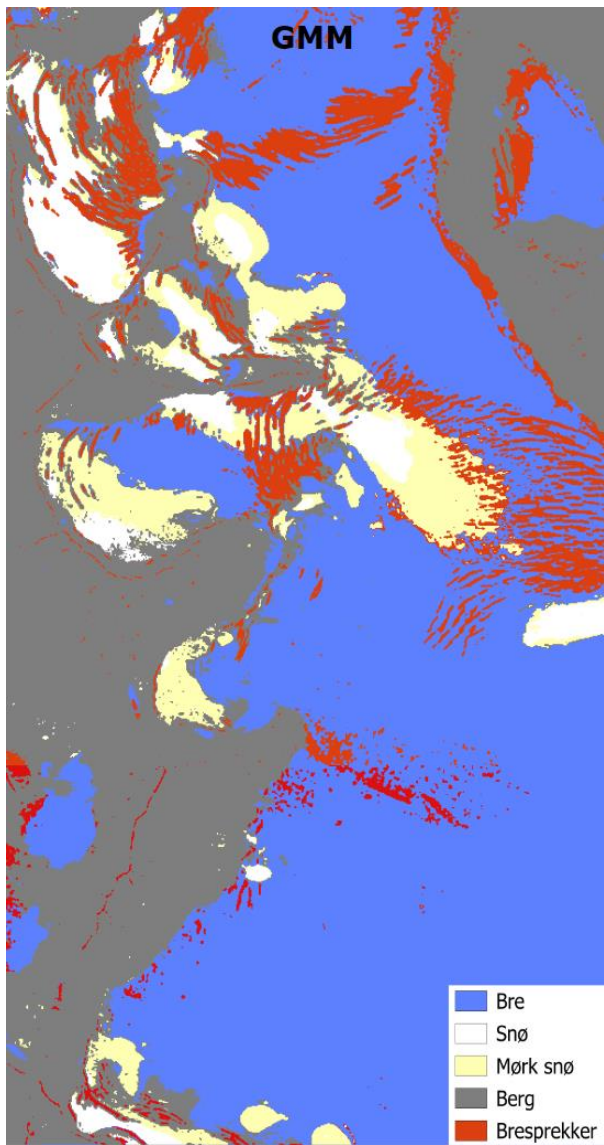
48 Resultater SVM sammenlignet med ortofoto fra NVE



49 Resultater pikselbasert klassifisering RF sammenlignet med ortofoto fra NVE



50 Resultater pikselbasert klassifisering KNN sammenlignet med ortofoto fra NVE



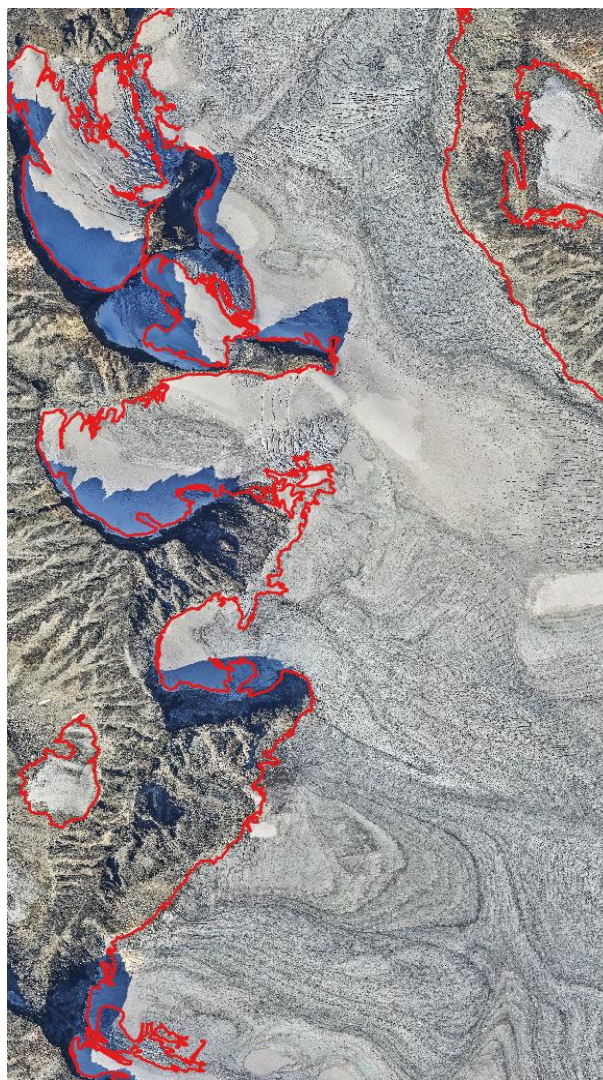
51 Resultater pikselbasert klassifisering GMM sammenlignet med ortofoto fra NVE

4.1.4 Resultater brekant

Figur 51 viser resultatet av reklassifisering og prosessering av pikselbasert klassifisering ved hjelp av SVM algoritmen, slik som beskrevet i metodekapittelet. Figur 52 viser resultatene av raster til vektor metoden som tidligere ble beskrevet i oppgaven, med ortofoto fra NVE som bakgrunn. Analyse av resultater og funn ved utledning av brekant blir nærmere sett og diskutert i kapittel 5.



52 Reklassifisert bilde med Sieve algoritmen. Terskelverdi er satt til 50 000 piksler



53 Resultatet etter raster til vektor algoritmen med ortofoto fra NVE i bakgrunn

5 Diskusjon

Dette kapittelet diskuterer resultatene som er oppnådd ved å følge prosessen i metodekapittelet, fra inspeksjon av data, preprosessering, klassifisering og testing. Målet med denne oppgaven var å utvikle og evaluere en metode for bruk av laserdata til pikselbasert klassifisering av is, ulike snøtyper, berg og bre-sprekker på norske isbreer, samt forsøke å utlede brekant basert på resultatene av klassifiseringen.

5.1 Analyse av resultatene

5.1.1 Statistisk test og valideringsparametere

Resultatene viser at SVM-algoritmen har oppnådd den høyeste presisjonsverdien på 0,8745, etterfulgt av KNN med en verdi på 0,8621, GMM med en presisjonsverdi på 0,8514 og RF med en presisjonsverdi på 0,8463. Når det kommer til gjenkall, var SVM og KNN også de to beste algoritmene med verdier på henholdsvis 0,8780 og 0,8711. Gjenkall verdien til RF og GMM var på henholdsvis 0,8645 og 0,8323. F1-verdiene, som er en kombinasjon av presisjon og gjenkalling, var også høyest for SVM (0,8750), etterfulgt av KNN (0,8662), RF (0,8548) og GMM (0,8390).

Generelt sett kan man si at resultatene er gode, da alle algoritmene har en F1-score på over 0,83, som indikerer god generell ytelse fra modellene. Modellene har også generelt god balanse mellom presisjon og gjenkall. SVM var imidlertid den modellen som ga best resultat på alle valideringsparameterne, noe som antyder at denne algoritmen kan være spesielt effektiv for klassifisering av isbreoverflater beskrevet i denne oppgaven. I tillegg til SVM, utmerker KNN og RF seg som bedre enn GMM, og F1-verdien til GMM er relativt lav sammenlignet med de andre algoritmene.

Til tross for at SVM har utmerket seg som den beste algoritmen basert på valideringsparameterne, bør det bemerkes at de andre modellene har prestert godt. Som nevnt viser alle modellene høye verdier på valideringsparametere, med ulike egenskaper til å predikere de forskjellige overflateklassene. Hvilket kan bety at datasettet modellene blir trent opp på inneholder mye informasjon som legger et godt grunnlag for opptrening av modeller, delvis uavhengig av algoritme.

I denne oppgaven ble det valgt å bruke makrogjennomsnitt for utregning av valideringsparametere fra forvirringsmatrisene. En av bakdelene med å bruke makrogjennomsnitt på et ubalansert testdatasett, som i denne oppgaven, er at det kan gi en overoptimistisk evaluering av algoritmens ytelse. Dette skyldes at makrogjennomsnitt likestiller ytelsen til hver klasse, uavhengig av om det er en stor eller liten andel av eksemplene i datasettet som tilhører den gitte klassen.

Hvis for eksempel et datasett har en stor andel av eksempler i klasse A og en liten andel i klasse B, vil makrogjennomsnittet gi like stor vekt til ytelsen for begge klassene. Dette betyr at selv om algoritmen presterer godt på klasse A, men dårlig på klasse B, vil makrogjennomsnittet gi en høyere ytelsesvurdering enn hva som faktisk er tilfelle. På en annen side er en fordel med å bruke makrogjennomsnitt at det gir lik vekt til hver klasse uavhengig av størrelsen på klassene. I denne sammenheng er det viktig fordi klassene med mindre testpunkter, som bresprekker og snøtypene er like viktige som de to andre. Det kan også leses ut ifra forvirringsmatrisen at selv om datasettet er skjevt, så er det ingen av klassene som peker seg ut med stor andel feil prediksjoner.

5.1.2 Visuell analyse av klassifisering

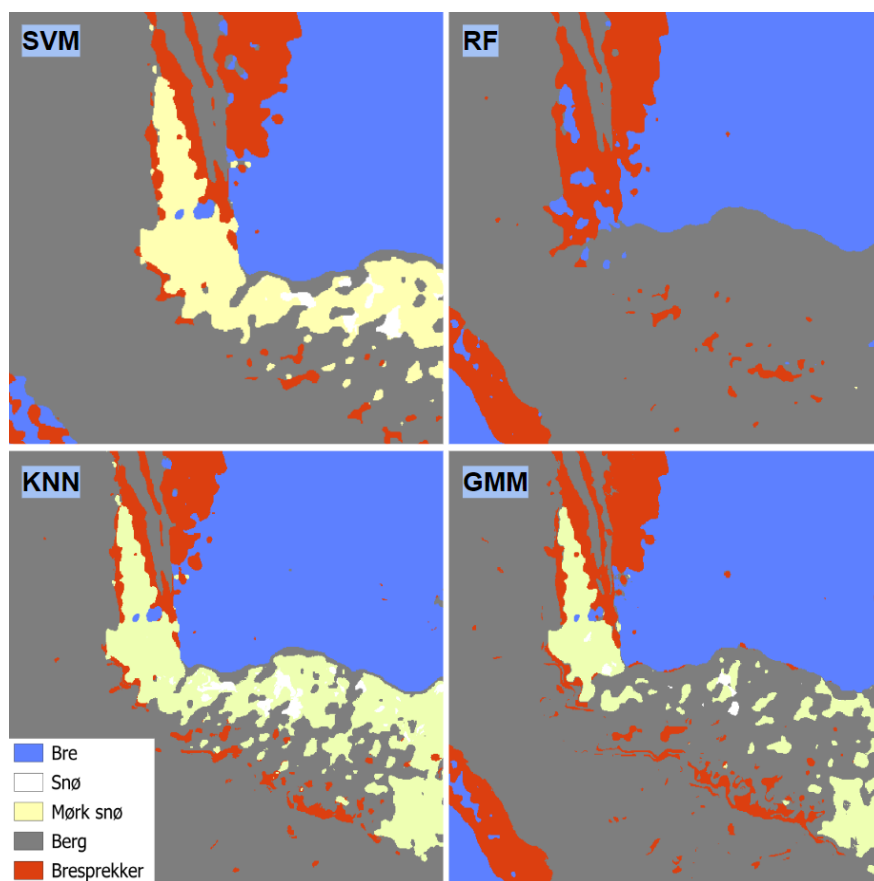
Den visuelle analysen viser at klassifiseringsmodellene presterer på et generelt høyt nivå, og de presterer ganske likt, med visse forskjeller på ulike overflater, i likhet med den statistiske testen. Det faktum at modellene har relativt lik ytelse kan tyde på at treningsdataene og den generelle metoden er av høy kvalitet. Dette kan bety at datasettet inneholder en tilstrekkelig mengde og variasjon av informasjon, som muliggjør god opptrening av modeller uavhengig av den spesifikke algoritmen som brukes. Det kan også indikere at den anvendte metoden for maskinlæring og klassifisering var effektiv og tilpasset datasettet på en egnet måte.

Selv om modellene presterte ganske likt finns det områder med avgjørende forskjeller visuelt, som kan bli trukket frem i ulike eksempler med utklipp fra klassifiseringsresultatene og ortofoto.

Eksempel 1 viser en breflekk hvor alle modellene har utfordringer med overgangen fra bre til berg, men at RF utpeker seg som den beste isolert sett i dette utklippet siden den ikke feilklassifiserte berg som mørk snø. Alle modellene mistolket overgangen fra berg til bre som bresprekker i venstre del av breflekken.

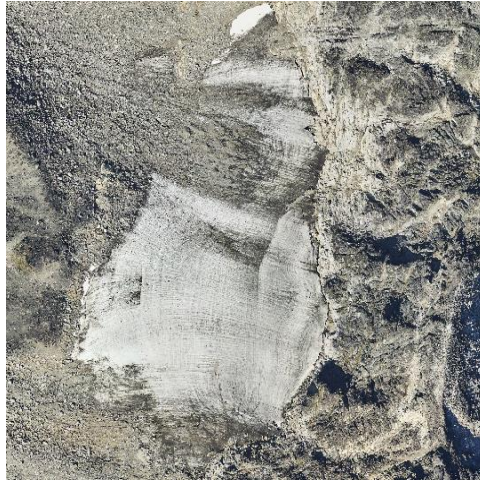


54 Ortofoto levert av NVE

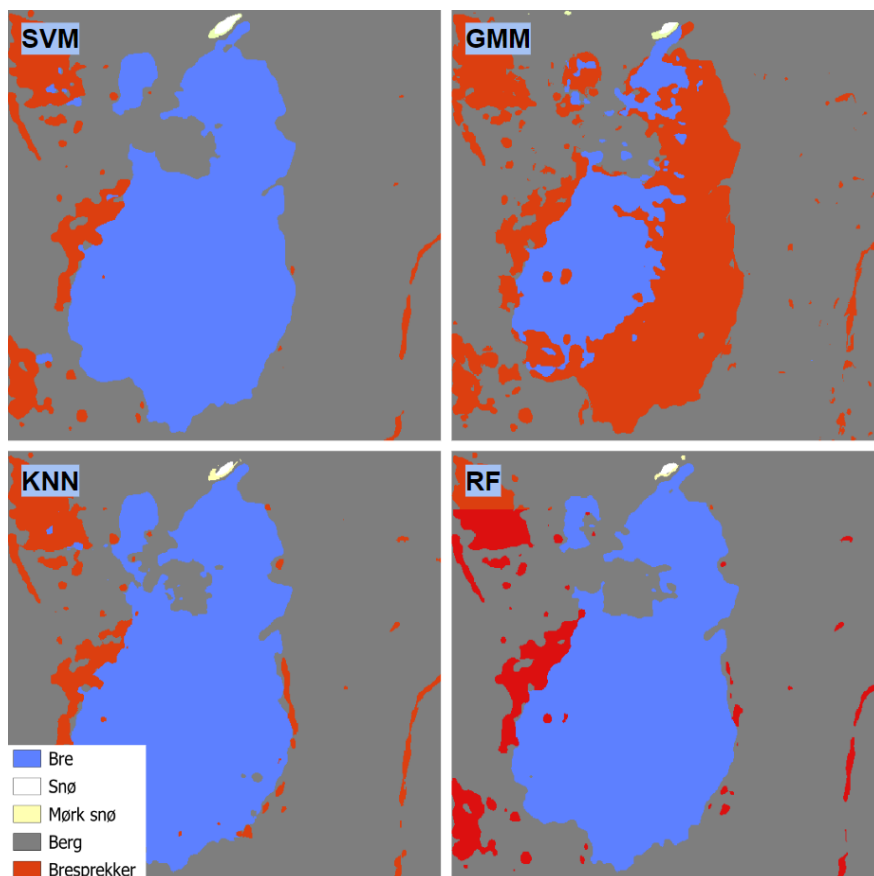


55 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på breflekk

I eksempel 2, på en annen breflekk over berg har GMM klassifisert høyre del av breflekken som bresprekk. Ut ifra ortofoto kan det se ut som at berget har en bratt helning ned mot bre, så det er naturlig å anta at GMM har klassifisert bresprekk på grunn av en rask høydeendring. Det ser ut til at SVM, RF og KNN har klassifisert riktig i overgang fra berg til breflekk, og ganske nøyaktig på fortsettelsen av breflekken nordover i utsnittet.



56 Ortofoto levert av NVE

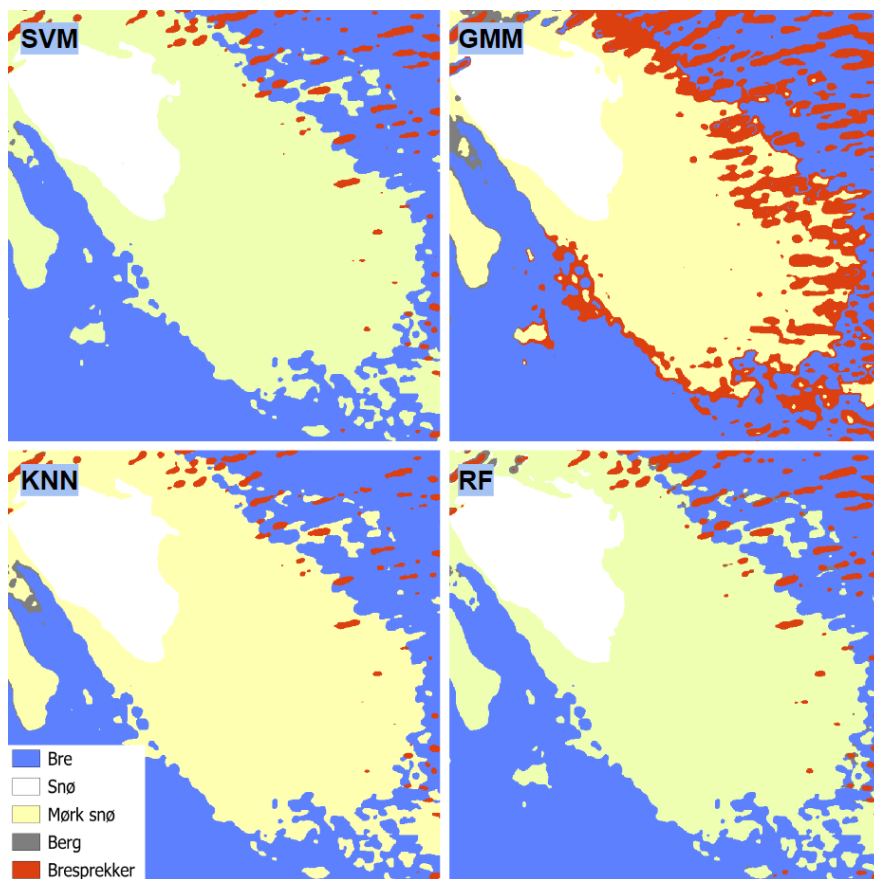


57 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på breflekk

I eksempel 3 vises et område med overgang fra snø til mørk snø. Alle modellene predikerer snøtypene godt utenfor komplekse områder hvor helningsgraden er for stor. I dette området ute på breen klassifiserer modellene godt. GMM predikerer større overflater med bresprekker i forhold til de andre algoritmene, ut ifra ortofoto er det vanskelig å tyde i hvilken grad dette er riktig eller feil.

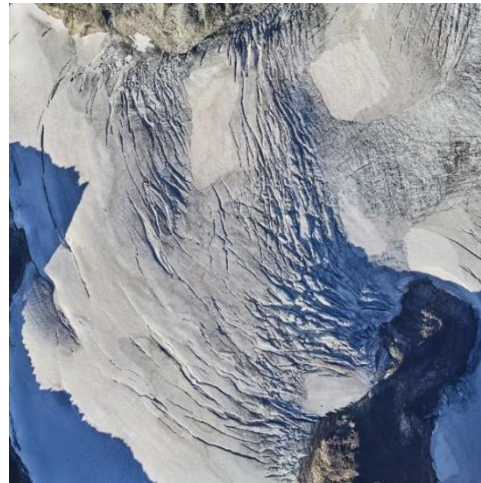


58 Ortofoto levert av NVE

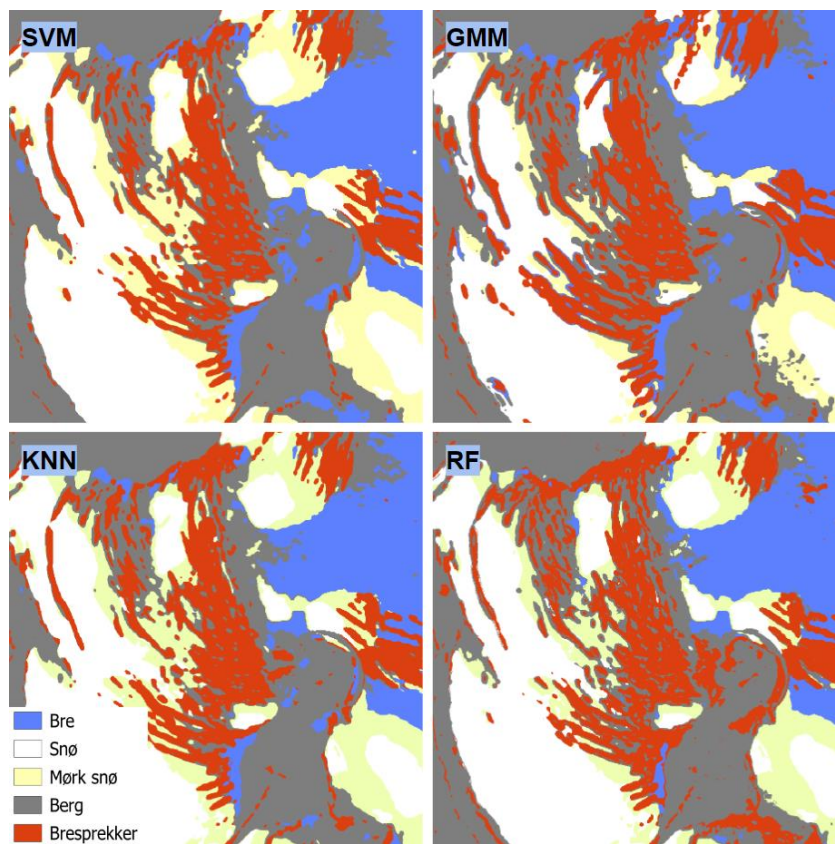


59 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM på overflate med snø og mørk snø

Eksempel 4 viser et annet interessant område i klassifiseringen med høyere grad av kompleksitet, i form av flere typer overflater på et mindre område, hvor alle modellene har forveksler mørk snø og bresprekker med berg på breen. Dette kan være på grunn av at høydeverdiene har mye å si for klassifiseringsresultatet. Dette er i likhet med bergtoppene, hvor områder med berg blir klassifisert som bresprekker. På disse overflatene kan man argumentere for at dette har med helning, og andre grads helning verdiene i opptrening av modellen å gjøre.



60 Ortofoto levert av NVE

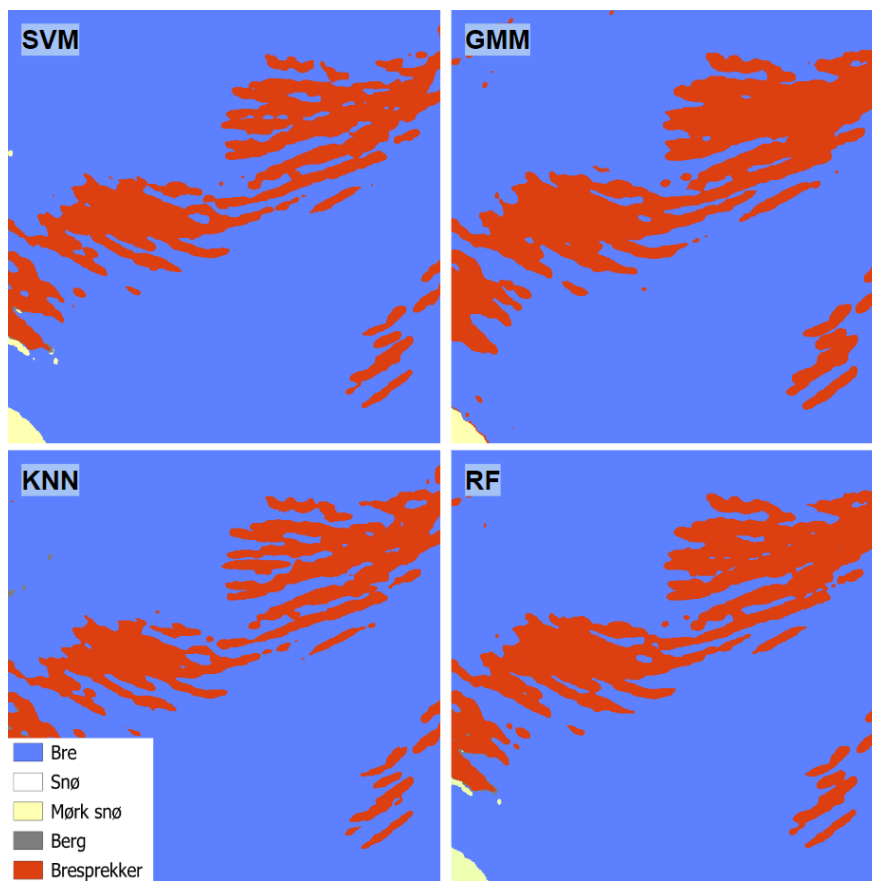


61 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM i komplekst område

Eksempel 5 er annet område som er interessant å trekke frem, som er overflater hvor det finns bresprekker. Algoritmene er gode på å plukke opp disse, i varierende grad, men visuelt ser det ut som de fleste bresprekkene er fanget opp av modellene.

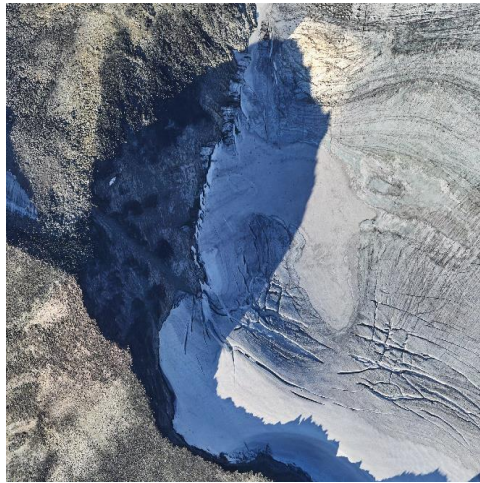


62 Ortofoto levert av NVE

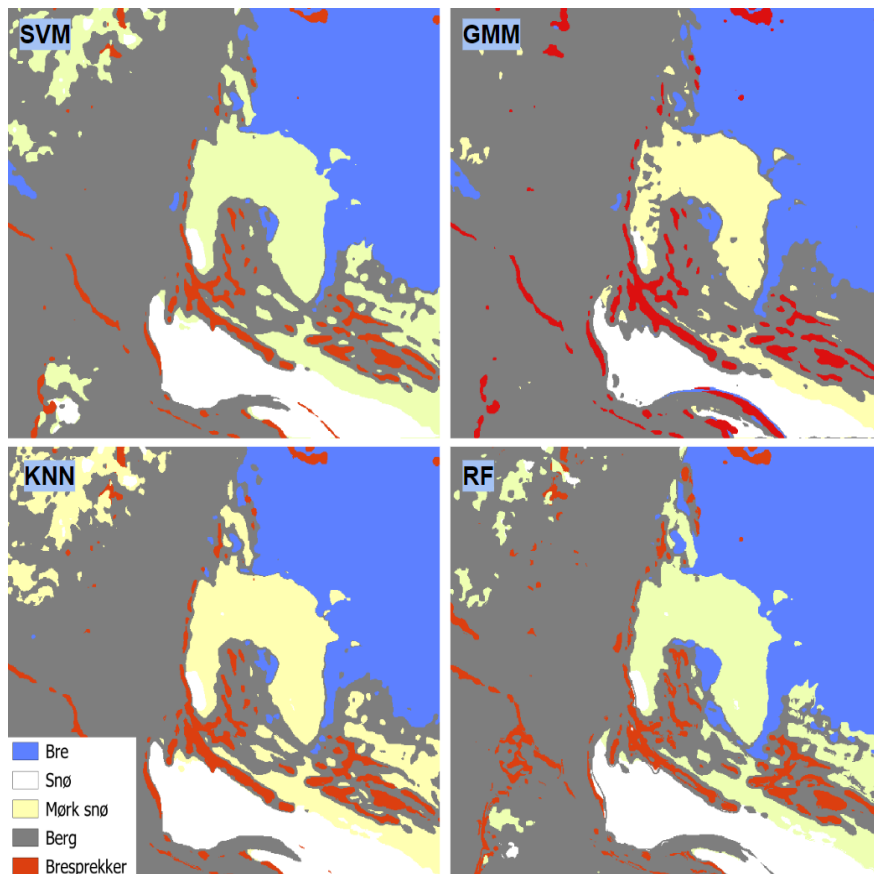


63 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM som illustrerer klassifisering av bresprekker

Eksempel 6 illustrerer et annet område med høy kompleksitet. Igjen er dette en utfordring for modellene, og områder med bresprekker blir ofte klassifisert som berg. Her synes også bergtoppene som har blitt klassifisert som bresprekker.

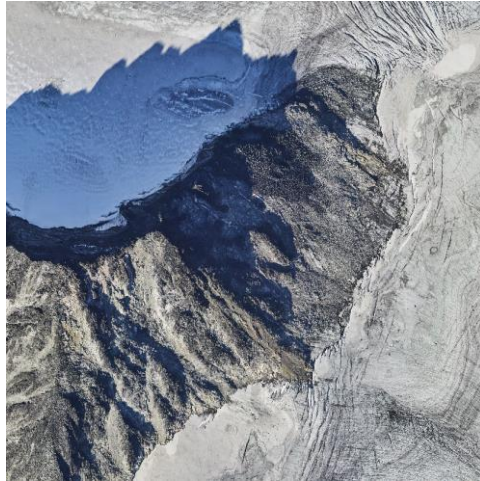


64 Ortofoto levert av NVE

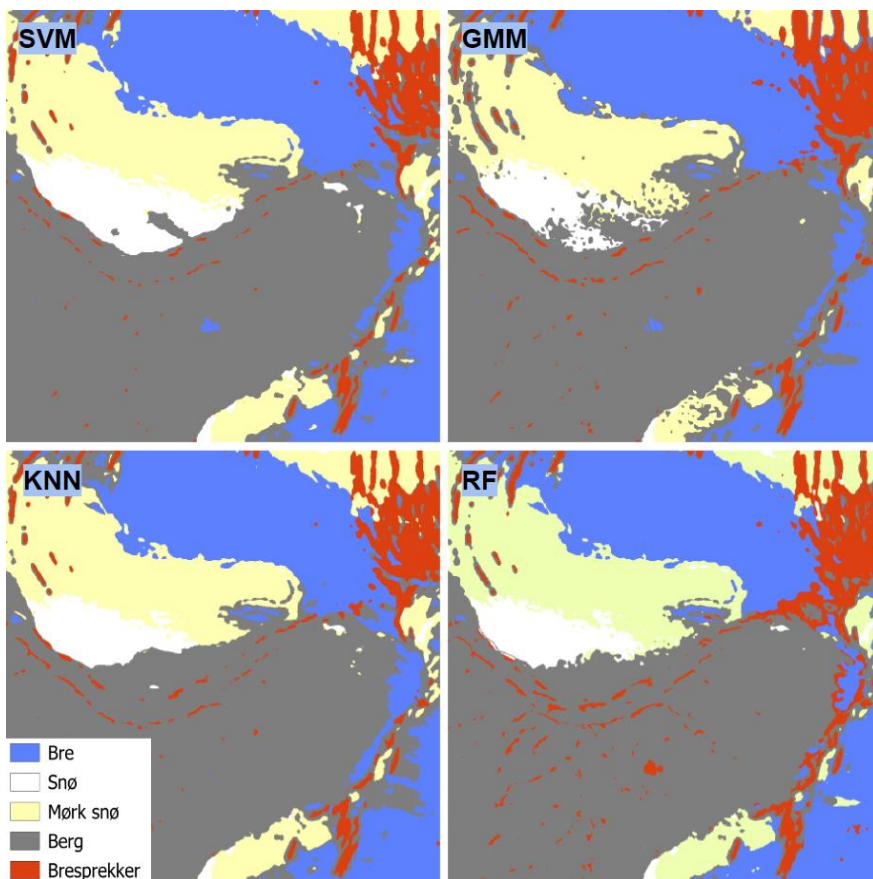


65 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM i komplekst område

Vi ser i eksempel 7 en variant som fremviser algoritmenes egenskaper til å fange opp overgangen fra berg til bre. Sammenlignet med ortofoto presterer SVM best, men predikerer noe feil. Visuelt sett kan vi se at de andre algoritmene har større problemer med å plukke opp overgangen med like stor nøyaktighet.



66 Ortofoto levert av NVE



67 Visuelt klassifiseringsresultat av SVM, RF, KNN og GMM som illustrerer overgang fra berg til breoverflater

5.1.3 Visuell analyse brekant

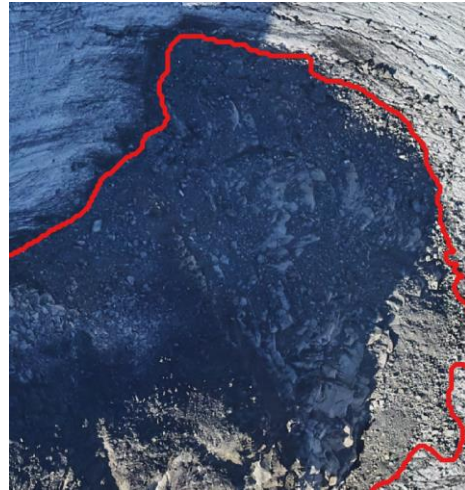
Resultatet av utledning av brekanten avhenger av klassifiseringsresultatet som danner grunnlaget for dataene. Generert vektor viser at klassifiseringen må være svært god for å oppnå en tilstrekkelig nøyaktighet på utledning av brekanten. Vi ser at når klassifiseringen er presis og reklassifiseringen har vært vellykket at deteksjonen av brekanten, visuelt sett, kan være svært nøyaktig.

Det kan dermed argumenteres både for og imot bruken av denne metoden for breovervåking. På den ene siden er ikke deteksjonen perfekt over hele breområdet, og den kan ikke benyttes i en helautomatisk metode, men på den annen side kan metoden være effektiv i områder hvor den fungerer godt, og kan være et verktøy for å supplere dagens metoder for å øke presisjon, og mulig forenkle noen prosesser.

Utklippene a), b), c) og d) illustrerer områder hvor metoden har lykket i utledning av brekanten, mens utklippene e), f), g) og h) viser overflater med høy kompleksitet på breen hvor klassifiseringsgrunnlaget har vært dårlig og reklassifiseringen ikke har vært vellykket, noe som fører til unøyaktig deteksjon av brekanten.



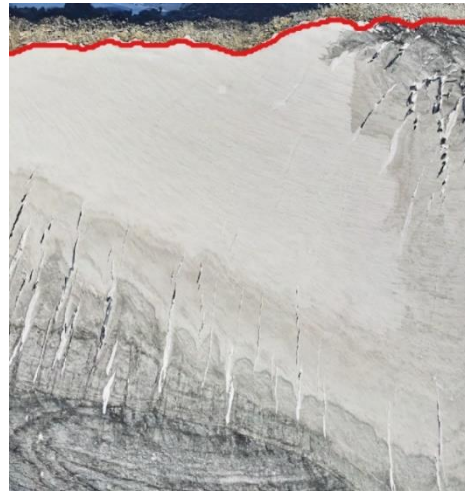
a)



b)

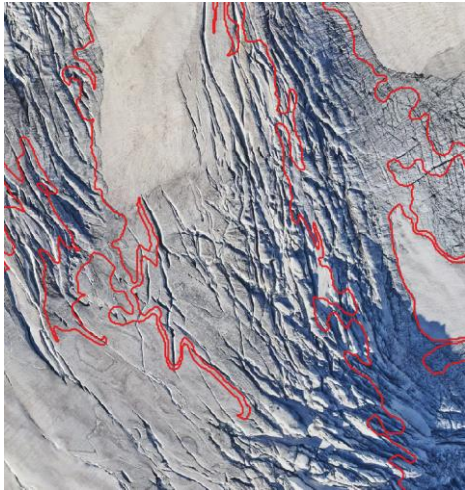


c)



d)

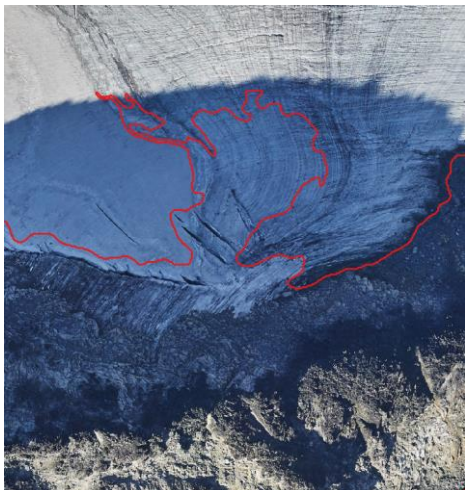
68 a), b, c) og d) viser utklipp av vektor i nøyaktig klassifiserte områder



e)



f)



g)



h)

69 e), f), g) og h) illustrerer overflater på breen som har vært for komplekse for utledning av nøyaktig brekant

5.2 Usikkerhet knyttet til metodikk

5.2.1 Korrigert laserdata

I denne oppgaven blir det som nevnt brukt et datasett produsert med reflektans, som gir homogene intensitetsverdier for like overflater uavhengig av avstanden fra laserskanneren. Selv om det ikke er gjennomført tester på datasett som ikke er korrigert, antas det at bruken av reflektans er avgjørende for resultatene som ble oppnådd i oppgaven. Det er imidlertid verdt å merke seg at datasettet ikke er korrigert for innfallsvinkel, noe som kunne ha bidratt til å forbedre resultatene ytterligere. Men det skal nevnes at innfallsvinkel har betydelig mindre og si for returnert signal enn avstand, og ikke nødvendigvis hadde endre klassifiseringsresultatet.

5.2.2 Begrensninger med Dzetsaka algoritmen

Forsøket med å utvikle en metode for benyttelse av laserdata og testing av disse dataene visuelt på et delt område, hvor halvparten ble brukt som test og halvparten som trening skaper begrensninger i Dzetsaka algoritmen. Det er en vanlig praksis å justere hyperparametere til algoritmene basert på trening og testing, for å oppnå optimal ytelse. Dzetsaka- pluginen gir mulighet for å justere disse parametrene, men det er ikke mulig å utføre dette i denne oppgaven på grunn av forsøkets utforming. Som et resultat kan ytelsen til algoritmen ha blitt begrenset. For å undersøke algoritmens ytelse grundigere kunne det blitt lagt opp til mulighet for tuning av hyperparametere, og mer testing.

5.2.3 Trening og testdata

I denne studien ble trening- og testdatasett delt opp fra samme grunnlag, dermed har ikke maskinlæringsalgoritmene fått prøvd seg på helt ukjent data. Dette kan være med på å svekke troverdigheten til resultatene. Når algoritmene blir klassifisert og testet på et datasett som ligner så mye på det det har blitt trent på, er det en risiko for overtilpasning. På den andre siden har antall treningspunkter vært relativt lavt i forhold til mengden piksler som eksisterer i rasterne, noe som minker sannsynligheten for at overtilpasning har oppstått. Det kunne vært hensiktsmessig å inkludere helt usette data i testen, men i denne studien var det ikke tilgang på flere datasett.

5.2.4 Testpunkter

Det ble opprettet 1 500 tilfeldige punkter for validering av resultatet fra klassifiseringene i denne oppgaven. Testens fordel er at den dekker et stort areal på breen, og er godt fordelt arealmessig, men bakdelen er relativt lavt antall piksler til validering. Man kan argumentere for at det statistiske grunnlaget er svakt, og en mer omfattende test ville vært mulig dersom det fantes fasitdata på overflatene. Fasitdata har ikke vært tilgjengelig for overflatene på breen, og sann verdi på testpunkter har blitt opprettet av forfatter basert på ortofoto fra NVE. Det ligger dermed en begrensning i grunn, hvor et enormt antall testpunkter hadde vært praktisk urimelig å opprette. Videre ble det gjort en vurdering hvor det ble konkludert at 1 500 testpunkter har vært tilstrekkelig for å trekke ut et bredt nok statistisk grunnlag for evaluering.

5.3 Tilgjengelig tidligere forskning

I denne oppgaven har det vært begrenset tilgang på forskning innenfor den spesifikke problemstillingen som er undersøkt. Dette har gjort det nødvendig å utforske og teste forskjellige metoder og tilnærminger for å oppnå best mulig resultat. En rekke metoder blitt prøvd og feilet før en passende tilnærming ble funnet. Dette har involvert en iterativ tilnærming hvor man har evaluert og justert parametrene og tilnærmingene underveis for å optimalisere resultatene.

6 Konklusjon

Formålet med oppgaven var å undersøke om det var mulig å utvikle en metode basert på flybårne laserdata, som kunne anvendes i pikselbasert klassifisering av is, snø, mørk snø, berg og bresprekker på norske isbreer. Videre ble det undersøkt om man kan utlede brekanten basert på denne klassifiseringen, og evaluere om metoden kunne forbedre dagens overvåkings- og målemetoder.

Resultatene viste at alle maskinlæringsalgoritmene - SVM, RF, KNN og GMM - presterte godt ifølge valideringsparameterne, og at SVM oppnådde de høyeste verdiene med en presisjon på 0,8745, en gjentakelse på 0,8780 og en F1-verdi på 0,8750.

De visuelle analysene viste at alle maskinlæringsmodellene presterer godt i områder med lav kompleksitet, men har problemer med å skille mellom ulike klasser på komplekse overflater, spesielt hvor berg og bresprekker møtes, samt på breflekker over berg. Metoden er også utsatt hvor endringen i høydeverdier i et område er like på tvers av klasser.

Selv om metoden ikke var i stand til å identifisere en sammenhengende vektor på overgangen mellom isbre og berg over hele isbreen, viste utledningen av isbrekanten fra SVM-klassifiseringen positive resultater. Denne tilnærmingen kan utgjøre et nyttig tillegg til eksisterende metoder for kartlegging av isbreer.

Studien konkluderer med at laserdata har potensial for pikselbasert klassifisering og måling av norske isbreer. Maskinlæringsmodellene GMM, RF, KNN og GMM kan klassifisere is, snø, mørk snø, berg og bresprekker på norske isbreer med god nøyaktighet, spesielt i mindre komplekse områder. Brekant kan utledes svært nøyaktig i områder hvor klassifiseringen har vært god, men metoden er ikke pålitelig nok til automatiske prosesser. Totalt sett kan metoden være et bidrag i å forbedre dagens overvåking- og målemetoder av norske isbreer.

6.1 Videre arbeid

Som nevnt i konklusjonen har laserdata et betydelig potensial for å klassifisere breområder, og det er muligheter for videre forskning på dette området. Her er noen forslag til videre arbeid:

- Justere og teste hyperparametere for maskinlæringsalgoritmer for å undersøke om resultatene kan forbedres ytterligere.

- Utvikle og teste avstandskorreksjon for eldre laserdatasett for å kunne sammenligne endringer i breen over tid.

A.1 Python kode

```
import pandas as pd
from sklearn.metrics import accuracy_score, precision_recall_fscore_support as score

# Definerer filnavnene for hver algoritme
algorithms = {'SVM': 'SVM1500.csv',
              'RF': 'RF1500.csv',
              'KNN': 'KNN1500.csv',
              'GMM': 'GMM1500.csv'}

# Oppretter en tom liste for å lagre resultatene for hver algoritme
results = []

# Itererer over algoritmene og deres respektive filer
for alg, csv_file in algorithms.items():
    # Leser inn CSV-filen ved hjelp av pandas
    df = pd.read_csv(csv_file, header=None)

    # Tildeler første kolonne til y_test og andre kolonne til predicted
    y_test = df.iloc[:, 0].values
    predicted = df.iloc[:, 1].values

    # Beregner macro precision, recall, fscore og support ved hjelp av sklearn.metrics
    precision_macro, recall_macro, fscore_macro, support_macro = score(y_test, predicted, average='macro')

    # Beregner nøyaktighet ved hjelp av sklearn.metrics
    accuracy = accuracy_score(y_test, predicted)

    # Legger til resultatene for hver algoritme til listen
    results.append({
        'Algorithm': alg,
        'Precision (Macro)': precision_macro,
        'Recall (Macro)': recall_macro,
        'F1-Score (Macro)': fscore_macro,
        'Support (Macro)': support_macro,
        'Accuracy': accuracy
    })

# Oppretter en Pandas DataFrame fra listen med resultater
df_results = pd.DataFrame(results)

# Lagrer DataFrame til en Excel-fil
df_results.to_excel('results_macro_average_and_accuracy.xlsx', index=False)
```

Bibliografi:

- Andreassen, L. M., Elvehøy, H., Kjøllmoen, B. & Belart, J. M. (2020). Glacier change in Norway since the 1960s—an overview of mass balance, area, length and surface elevation changes. *Journal of Glaciology*, 66 (256): 313-328.
- Andreassen, L. M. (2022). Breer og fonner i Norge. *Norges vassdrags- og energidirektorat*.
- Andreassen, L. M., Nagy, T., Kjøllmoen, B. & Leigh, J. R. (2022). An inventory of Norway's glaciers and ice-marginal lakes from 2018–19 Sentinel-2 data. *Journal of Glaciology*, 68 (272): 1085-1106. doi: 10.1017/jog.2022.20.
- Bishop, C. M. & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, b. 4: Springer.
- Bonaccorso, G. (2017). *Machine learning algorithms*: Packt Publishing Ltd.
- Breatlas, N. (2023). *Breatlas*. NVE.no. Tilgjengelig fra: <https://temakart.nve.no/tema/breatlas> (lest 09.05.23).
- Bullinaria, J. A. (2015). Bias and variance, under-fitting and over-fitting. *Neural Computation: Lecture*, 9: 16.
- Dick, Ø. B. (2020). *Laserskanning av terreng i Store norske leksikon*, 29. mars. digitaltmuseum. 22. desember 2018. Tilgjengelig fra: <https://digitaltmuseum.no/021018030852/hellstugubreen>.
- Engh, S. (2013). *Potensiale for anvendelse av LiDAR-data i glasiologi. Interpolasjon av DTM, korreksjon av intensitetsverdier og automatisk kartlegging av bresprekker*.
- Geovekst, K. (2022). *Produktspesifikasjon - Punktsky 1.0.3*.
- Haeberli, W., Hoelzle, M., Paul, F. & Zemp, M. (2007). Integrated monitoring of mountain glaciers as key indicators of global climate change: the European Alps. *Annals of glaciology*, 46: 150-160.
- Höfle, B., Sailer, R., Vetter, M., Rutzinger, M. & Pfeifer, N. (2009). *Glacier surface feature detection and classification from airborne LiDAR data*.
- Jibril, M., Algehyne, E., Sani & Usman, S. (2020). Predictive Supervised Machine Learning Models for Diabetes Mellitus. *SN Computer Science*, 1. doi: 10.1007/s42979-020-00250-8.
- Kjøllmoen, B. (2019). Reanalysing a glacier mass balance measurement series-Langfjordjøkelen 2008–2018. *NVE Report*: 28.
- Kumar, A. (2020). Micro-Average & Macro-Average Scoring Metrics-Python. *Data Analytics*, 4.
- LAS Specification 1.4*. (2019). The American Society for Photogrammetry & Remote Sensing. Tilgjengelig fra: https://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf.
- Leclercq, P., Oerlemans, J., Basagic, H., Bushueva, I., Cook, A. & Bris, R. (2014). A data set of worldwide glacier length fluctuations. *The Cryosphere*, 8. doi: 10.5194/tc-8-659-2014.
- Liestøl, O., Hoel, A. & Werenskiold, W. (1962). List of the areas and number of glaciers. *Glaciers and snowfields in Norway, Norsk Polarinstitutt Skrifter*, 114: 35-54.
- Liestøl, O. H., Jon Ove; Nesje, Atle; Mangerud, Jan. *Isbre i Store norske leksikon*. Tilgjengelig fra: <https://snl.no/isbre> (lest 31. mars).
- Merry, K., Bettinger, P., Crosby, M. & Boston, K. (2023). *Geographic Information System Skills for Foresters and Natural Resource Managers*. I: *Geographic Information System Skills for Foresters and Natural Resource Managers*, s. xv-xvii: Elsevier.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A. (2018). *Foundations of machine learning*: MIT press.

- Nussbaumer, S., Nesje, A. & Zumbühl, H. (2011). Historical glacier fluctuations of Jostedalsglaciären and Folgefonna (southern Norway) reassessed by new pictorial and written evidence. *Holocene*, 21: 455-471. doi: 10.1177/0959683610385728.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12 (null): 2825–2830.
- QGIS - *The Leading Open Source Desktop GIS*. Tilgjengelig fra: <https://qgis.org/en/site/about/index.html>.
- Raschka, S. (2015). *Python machine learning*: Packt publishing ltd.
- Raschka, S. & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*: Packt Publishing Ltd.
- Ruth, M., Ritter, N., Grissom, E., Borup, B., Galang, G., Haller, J., Stephenson, G., Covington, S., Nagy, T., Moyers, J., et al. (1995). *GeoTIFF format specification GeoTIFF revision 1.0*.
- Shafranovich, Y. (2005). *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. Tilgjengelig fra: <https://www.hjp.at/doc/rfc/rfc4180.html> (lest 10.05.23).
- Solli, P. (2019). *Rapport for luftbåren laserskanning - Hellstugubreen- Memurubreen*.
- Sørskaar, F., Andreassen, L. M. & Lytskjold, B. *Isbreene - en del av vår identitet. En viktig klimaindikator og energikilde*: NVE. Tilgjengelig fra: <https://storymaps.arcgis.com/stories/cb7a81c93c454338814b2729061a6d0d>.
- Tidemann, A., Elster, Anne Cathrine. (2022). *Maskinlæring i Store norske leksikon*. Tilgjengelig fra: <https://snl.no/maskinl%C3%A6ring> (lest 29. mars).
- TIFF Revision 6.0*. (1992). Aldus Corporation.
- Vosselman, G. & Maas, H.-G. (2010). *Airborne and Terrestrial Laser Scanning*.
- Zhang, J., Jia, L., Menenti, M. & Hu, G. (2019). Glacier Facies Mapping Using a Machine-Learning Algorithm: The Parlung Zangbo Basin Case Study. *Remote Sensing*, 11. doi: 10.3390/rs11040452.
- Aarsten, D. (2018). Analyse av fordeler og eventuelle begrensninger ved bruk av reflektansverdier for intensitet i punktsky til Nasjonal Detaljert Høydemodell (NDH)



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway