



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2024 30 stp

Fakultet for realfag og teknologi

Klassifisering av mineralene alunitt og goethitt med tre sensorer, HySpex, Sentinel-2 og EMIT, med varierende geometrisk og spektral oppløsning

Classification of the minerals alunite and goethite using three sensors, HySpex, Sentinel-2, and EMIT, with varying geometric and spectral resolution.

Torje Fostås & Astrid Ingeborg Skiaker

Geomatikk

Forord

Denne oppgaven markerer slutten på 5 fantastiske år på Norges Miljø og Biovitenskapelige Universitet.

En spesiell takk går til vår hovedveileder, Ivar Maalen-Johansen. Din kunnskap og ro har vært avgjørende for oppgaven. Kontoret ditt har gjennom 5 år alltid vært åpent og du bryr deg virkelig om studentene dine. Geomatikk sin hærfører for et godt miljø! En stor takk rettes også til vår biveileder, Dagrun Aarsten. Ditt engasjement og nysgjerrighet for faget er virkelig inspirerende, og har drevet oppgaven i spennede retninger ingen hadde forutsett. En takk rettes også til Berit F. Fostås for korrekturlesing.

Gutta som vi har tilbragt uendelig mange timer sammen med på geomatikk-rommet fortjener også å nevnes. Deres evne til å effektivt veksle mellom kødding og seriøs jobbing har ikke bare gjort studiehverdagen lettere, men også gjort hele studietiden til en bedre opplevelse.

Til slutt vil vi takke NMBU, foreningslivet, NMBUI og Studentsamfunnet for en begivenhetsrik studietid og de livslange vennskapene som har blitt til her. Åsbobla vil aldri bli glemt!

Ås, mai 2024

Astrid Ingeborg Skiaker og Torje Fostås

Sammendrag

Fjernmåling er en banebrytende metode for deteksjon av mineraler, og muliggjør identifisering av mineraler med høy presisjon (van der Meer mfl., 2012; Okada, 2022, s. 429). Nevada har en rik geologi, og er derfor et ideelt område for å gjennomføre mineralanalyser med fjernmåling (Tripathi og Garg, 2021). Datagrunnlaget for analysen er hentet fra fjellområdet Cuprite som ligger i Nevada, USA.

I denne studien utføres en analyse ved hjelp av tre ulike sensorer: HySpex, Sentinel-2 og EMIT. Studien undersøker hvordan variasjoner i geometrisk og spektral oppløsning påvirker evnen til å oppdage mineralene alunitt og goethitt. Tre sentrale forskningsspørsmål adresseres. Det første setter søkelys på hvordan sensorenes ulike oppløsninger påvirker kvaliteten på mineraldeteksjonen. Det andre spørsmålet handler om muligheten til å oppdage alunitt og goethitt fra satellittbilder. Det tredje spørsmålet omhandler hvor godt den utviklede klassifiseringsmetoden presterer over et større geografisk område. Forskningsmetodikken inkluderer bruk av prinsipalkomponentanalyse (PCA) for å fremheve forskjeller, etterfulgt av mineralindekser. Til slutt gjennomføres en Random Forest-klassifisering for å identifisere de spesifikke mineralene.

Hovedfunnene indikerer at HySpex-sensoren oppnår høy nøyaktighet i klassifiseringen, mens både EMIT og Sentinel-2 viser betydelig svakere resultater. Dette er i tråd med forventningene på grunn av deres lavere spektrale og geometriske oppløsning. Interessant nok er resultatene mellom de to sistnevnte sensorene overraskende jevne, selv om EMIT har betydelig dårligere geometrisk oppløsning per piksel, med 60 meter mot Sentinel-2s 10 meter. Resultatene fra EMIT er imidlertid ikke helt pålitelige, da det er få testpunkter på grunn av det begrensede studieområdet. Sentinel-2 presterer generelt bedre, men EMIT utmerker seg med sin overlegne spektrale oppløsning, noe som er avgjørende for nøyaktig mineralidentifikasjon. Studien underbygger derfor betydningen av spektral fremfor geometrisk oppløsning i mineraldeteksjon.

Abstract

Remote sensing technology is a groundbreaking tool for detecting minerals, allowing identification and characterization with high precision (van der Meer *et al.*, 2012; Okada (2022), s. 429). Nevada has a rich geology, making it an ideal place for mineral analysis using remote sensing (Tripathi and Garg, 2021). The data for the analysis comes from the Cuprite mountain area in Nevada, USA.

This study conducts an analysis using three different sensors: HySpex, Sentinel-2, and EMIT. It explores how variations in geometric and spectral resolution affect the ability to detect the minerals alunite and goethite. The study addresses three key research questions. The first focuses on how different sensor resolutions impact the quality of mineral detection. The second question investigates the ability to detect alunite and goethite from satellite imagery. The third explores the potential of applying the developed classification method to larger geographical areas. The research methodology uses principal component analysis (PCA) to highlight differences, followed by mineral indices. Finally, a Random Forest classification is used to identify specific minerals.

The main findings indicate that the HySpex sensor achieves high accuracy in classification, while both EMIT and Sentinel-2 show significantly weaker results. This is expected due to their lower spectral and geometric resolution. Interestingly, the results between EMIT and Sentinel-2 are surprisingly similar, even though EMIT has much lower geometric resolution per pixel, at 60 meters compared to Sentinel-2's 10 meters. However, the results from EMIT are not entirely reliable because there are too few test points due to the limited study area. Sentinel-2 generally performs better, but EMIT stands out for its superior spectral resolution, which is crucial for accurate mineral identification. The study thus emphasizes the importance of spectral over geometric resolution in mineral detection.

Innholdsfortegnelse

| | | |
|----------|------------------------------------------------------------------|-----------|
| 1 | Introduksjon | 1 |
| 1.1 | Bakgrunn | 1 |
| 1.2 | Problemstilling | 2 |
| 1.2.1 | Forskningsspørsmål | 2 |
| 1.3 | Tidligere forskning | 2 |
| 2 | Teoridel | 4 |
| 2.1 | Fjernmålingsteknologi | 4 |
| 2.1.1 | Hyperspektral bildebehandling | 4 |
| 2.1.2 | Spektralsignatur | 5 |
| 2.1.3 | Radians og reflektans | 6 |
| 2.2 | Mineraler | 6 |
| 2.3 | Maskinlæring | 7 |
| 2.3.1 | Ikke styrt klassifisering | 7 |
| 2.3.2 | Styrt klassifisering | 8 |
| 2.3.3 | Validering av maskinlæringsmetoden | 9 |
| 3 | Metode | 12 |
| 3.1 | Data | 12 |
| 3.1.1 | Studieområde | 12 |
| 3.1.2 | Sensorer | 13 |
| 3.2 | Programvarer | 16 |
| 3.2.1 | Python | 16 |
| 3.2.2 | Catalyst Professional Focus | 16 |
| 3.2.3 | QGIS | 16 |
| 3.2.4 | Kunstig intelligens | 17 |
| 3.2.5 | Filformater | 17 |
| 3.2.6 | Preprosessering av data | 18 |
| 3.3 | Analysens fremgangsmåte | 22 |
| 3.4 | PCA | 23 |
| 3.5 | Klassifisering 1: alunitt | 24 |
| 3.5.1 | Indeks for alunitt | 24 |
| 3.5.2 | Generering av treningsdata | 26 |
| 3.5.3 | Treningsdata for Sentinel-2 og EMIT | 28 |
| 3.5.4 | Treningsområde og testområde | 30 |
| 3.5.5 | Gjennomføring av klassifisering 1: alunitt | 30 |
| 3.6 | Klassifisering 2: alunitt og goethitt | 30 |
| 3.6.1 | Indeks for goethitt | 31 |
| 3.6.2 | Nye treningspiksler for goethitt | 34 |
| 3.6.3 | Visuell fordeling basert på spektralsignatur | 35 |
| 3.6.4 | Sammenslåing av goethitt-klasser | 38 |
| 3.6.5 | Gjennomføring av klassifisering 2: alunitt og goethitt | 38 |
| 3.7 | Klassifisering 3: utvidelse av området | 39 |
| 4 | Resultater | 40 |
| 4.1 | Klassifisering 1: alunitt | 40 |

| | | |
|----------|---------------------------------------------------------|-----------|
| 4.1.1 | Valideringsparametere | 40 |
| 4.1.2 | Prosentvis fordeling | 42 |
| 4.1.3 | Visuelle resultater | 42 |
| 4.2 | Klassifisering 2: alunitt og goethitt | 45 |
| 4.2.1 | Valideringsparametere | 45 |
| 4.2.2 | Prosentvis fordeling | 47 |
| 4.2.3 | Visuelle resultater | 48 |
| 4.3 | Klassifisering 3: utvidet område | 51 |
| 4.3.1 | Visuelle resultater | 51 |
| 5 | Diskusjon | 52 |
| 5.1 | Diskusjon av data | 52 |
| 5.1.1 | Data fra EMIT | 52 |
| 5.1.2 | Mineraldeteksjon på et begrenset studieområde | 52 |
| 5.2 | Diskusjon av metode | 52 |
| 5.2.1 | Valg av maskinlæringsalgoritme | 52 |
| 5.2.2 | Test og treningsdata | 53 |
| 5.2.3 | Begrensninger med goethittindeksen | 55 |
| 5.2.4 | Fasit til verifisering | 56 |
| 5.3 | Diskusjon av resultatene | 56 |
| 5.3.1 | Valideringsparametere | 56 |
| 5.3.2 | Klassifisering 1: alunitt | 57 |
| 5.3.3 | Klassifisering 2: alunitt og goethitt | 58 |
| 5.3.4 | Klassifisering 3: utvidelse av området | 59 |
| 6 | Konklusjon | 63 |
| 6.1 | Forslag til videre arbeid | 65 |

Figurer

| | | |
|------|-----------------------------------------------------------------------------------------------|----|
| 2.1 | Multi- og hyperspektrale bilder | 4 |
| 2.2 | Spektralsignatur | 5 |
| 2.3 | Visuell fremvisning av mineralene | 7 |
| 2.4 | Random Forest | 8 |
| 2.5 | Trening- og testdata | 9 |
| 2.6 | Over- og undertilpasning | 10 |
| 2.7 | Forvirringsmatrise | 10 |
| 3.2 | Studieområdet HySpex | 13 |
| 3.3 | Studieområde Sentinel-2 | 14 |
| 3.4 | Studieområde EMIT | 15 |
| 3.5 | Flytdiagram HySpex | 19 |
| 3.6 | Flytdiagram Sentinel-2 | 20 |
| 3.7 | Flytdiagram EMIT | 21 |
| 3.8 | Flytdiagram fremgangsmåte | 22 |
| 3.9 | PCA med HySpex | 23 |
| 3.10 | PCA med Sentinel-2 | 24 |
| 3.11 | Alunitt-index, rå | 25 |
| 3.12 | Alunitt-index, terskelverdi | 26 |
| 3.13 | Spektralsignatur alunitt | 27 |
| 3.14 | Treningspikslers klassifisering 1 | 28 |
| 3.15 | Treningspikslers klassifisering 1: Sentinel-2 | 29 |
| 3.16 | Treningspikslers klassifisering 1: EMIT | 29 |
| 3.17 | Goethittindeks 1 | 32 |
| 3.18 | Karakteristiske topper og bunner goethitt | 33 |
| 3.19 | Goethittindeks 2 | 34 |
| 3.20 | Treningspunkter klassifisering 2 | 35 |
| 3.21 | Goethitt lab og fjernmåling, hele spekteret | 36 |
| 3.22 | Goethitt lab og fjernmåling, 0.2 μm - 1.0 μm | 36 |
| 3.23 | Høyst sannsynlig goethitt lab og fjernmåling, hele spekteret | 36 |
| 3.24 | Høyst sannsynlig goethitt lab og fjernmåling, 0.2 μm - 1.0 μm | 36 |

| | | |
|------|-----------------------------------------------------------------------------------------|----|
| 3.25 | Sannsynlig goethitt lab og fjernmåling, hele spekteret | 37 |
| 3.26 | Sannsynlig goethitt lab og fjernmåling, 0.2 μm - 1.0 μm | 37 |
| 3.27 | Utvidelse av området | 39 |
| 4.1 | Forvirringsmatriser fra klassifisering 1 | 41 |
| 4.2 | Valideringsparametere for alunitt-klassifisering | 41 |
| 4.3 | Klassifisering 1 med HySpex | 43 |
| 4.4 | Klassifisering 1 med Sentinel-2 | 44 |
| 4.5 | Klassifisering 1 med EMIT | 45 |
| 4.6 | Forvirringsmatriser fra klassifisering 2 | 46 |
| 4.7 | Valideringsparametere for klassifisering 2 | 46 |
| 4.8 | Klassifisering 2 med HySpex | 48 |
| 4.9 | Klassifisering 2 med Sentinel-2 | 49 |
| 4.10 | Klassifisering 2 med EMIT | 50 |
| 4.11 | Klassifisering 3 med Sentinel-2 | 51 |
| 4.12 | Klassifisering 3 med EMIT | 51 |
| 5.1 | Utvidet område med Sentinel-2 for alunitt | 60 |
| 5.2 | Utvidet område med Sentinel-2 for goethitt | 61 |
| 5.3 | Utvidet område med EMIT for alunitt | 61 |
| 5.4 | Utvidet område med EMIT for goethitt | 62 |
| 6.1 | Sammenligning av sensorenes ytelse | 64 |

Tabeller

| | | |
|-----|---------------------------------------------------------|----|
| 3.1 | Sammenligning av sensorene | 15 |
| 3.2 | Test- og treningspunkter for klassifisering 1 | 30 |
| 3.3 | Fordeling av piksler til klassifisering 2 | 37 |
| 3.4 | Endelige treningspiksler klassifisering 2 | 38 |
| 3.5 | Test- og treningspiksler for klassifisering 2 | 38 |
| 4.1 | Prosentvis fordeling klassifisering 1 | 42 |
| 4.2 | Prosentvis fordeling klassifisering 2 | 47 |

Ordforklaring

| | |
|--------------------|-------------------------------------------------|
| AIS | Airborne Imaging Spectrometer |
| AVIRIS | Airborne visible/infrared imaging spectrometer |
| EMIT | Earth Surface Mineral Dust Source Investigation |
| FOV | Field of View |
| GMM | Gaussian Mixture Model |
| KNN | K-Nearest Neighbour |
| Metrikker | Målingsstandarder |
| NIR | Nærinfrarødt |
| Open source | Åpen kildekode |
| PCA | Principal Component Analysis |
| RF | Random Forest |
| SWIR | Kortbølget infrarødt |
| SVM | Support Vector Machine |

1 Introduksjon

1.1 Bakgrunn

Bruk av multi- og hyperspektrale sensorer for geologisk kartlegging og mineralidentifikasjon har blitt et avgjørende verktøy for å utforske og forstå jordens overflate (van der Meer mfl., 2012). Med avanserte teknologiske fremskritt er det nå mulig å kartlegge og identifisere mineralforekomster med imponerende presisjon og nøyaktighet (Okada, 2022, s. 429). Oppgaven anvender maskinlæringsalgoritmen Random Forest (RF) for å identifisere mineralene alunitt og goethitt i et fjellområde i Nevada.

Nevada kjennetegnes ved rik geologi og varierte mineralressurser, og utgjør en av de mest kjente områdene for geologiske analyser med fjernmåling (Tripathi og Garg, 2021). Denne analysen fokuserer på å detektere mineralene alunitt og goethitt. Mineralene har tydelige spektrale signaturer som gjør dem lett gjenkjennelige og forskjellige fra sine omgivelser (Peyghambari og Zhang, 2021, s. 14). Valget av disse spesifikke mineralene er dermed strategisk for å tilrettelegge for en analyse fra et geomatikk-perspektiv, fremfor et geologi-perspektiv.

Studiens hensikt er å utvikle en metode som kan identifisere mineraler ved å analysere deres spektrale signaturer gjennom bruk av data fra multi- og hyperspektrale sensorer montert på droner og satellitter. Dataen som benyttes er innsamlet fra tre forskjellige sensorer. Den første sensoren er fra HySpex og har høy geometrisk oppløsning og høy spektral oppløsning. Sensor nummer to er satellitten Sentinel-2, og har moderat geometrisk oppløsning og lav spektral oppløsning. Den siste sensoren er satellitten EMIT med lav geometrisk oppløsning og høy spektral oppløsning. Denne masteroppgaven fokuserer konsekvent på bruk av open source-programvare. Med unntak av bildet fra HySpex er all data og programvare brukt gratis og tilgjengelig for alle.

I første fase av analysen gjennomføres en Principal Component Analysis (PCA), for å fremheve og visualisere spektrale forskjeller i studieområdet. Deretter anvendes en indeksmetode for å identifisere mineralet alunitt. Basert på mineralets spektrale signatur og tilhørende indeks utvikles treningspikslar. Så iverksettes RF-klassifisering over datasettene fra de tre ulike sensorene, med en påfølgende prosess for å validere klassifiseringenes og sensorenes nøyaktighet. Denne metodikken gjentas for et annet mineral, goethitt. Avslutningsvis testes algoritmen over et utvidet geografisk område for å få et inntrykk av dens robusthet og presisjon i en større skala.

1.2 Problemstilling

Problemstillingen i denne oppgaven er som følger:

Klassifisering av mineralene alunitt og goethitt med tre sensorer, HySpex, Sentinel-2 og EMIT, med varierende geometrisk og spektral oppløsning

1.2.1 Forskningsspørsmål

Med bakgrunn i problemstillingen er følgende forskningsspørsmål formulert for videre arbeid:

1. *Hvordan påvirker de varierende geometriske og spektrale oppløsningene til sensorene kvaliteten på deteksjonen av mineraler?*
2. *Er det mulig å detektere alunitt og goethitt med fjernmålingsdata fra satellitt?*
3. *Hvor godt presterer klassifiseringsmetoden over et større geografisk område?*

1.3 Tidligere forskning

Området Cuprite i Nevada, har tradisjonelt vært ansett som et sentralt område for bruk av fjernmålingsteknologier for å kartlegge geologiske endringer og identifisere mineraler (Tripathi og Garg, 2021). Området er godt dokumentert gjennom en rekke rapporter, fra både nyere tid og tidligere perioder.

Det første kjente tilfellet av mineraldeteksjon i Cuprite-distriktet ved bruk av fjernmålingsteknologi fant sted i 1983, hvor Airborne Imaging Spectrometer (AIS) ble anvendt. Denne hendelsen var banebrytende, ettersom den muliggjorde detaljert kartlegging av mineralforekomster fra luften, og regnes som en milepæl innenfor geologisk kartlegging fra fjernmåling (Goetz og Srivastava, 1985).

Artikkelen «Mapping Advanced Argillic Alteration at Cuprite, Nevada Using Imaging Spectroscopy» av Swayze mfl. (2014) omhandler analyse og kartlegging av mineralforekomster i Cuprite, Nevada, ved bruk av spektral bildeteknologi (AVIRIS). Studien presenterer to detaljerte kart: et kart over jernholdige mineraler og et kart over leire, sulfater og karbonater. Det første kartet, som omhandler jernholdige mineraler, bruker spekteret fra 400 nm til 1300 nm mikrometer og viser tilstedeværelsen av mineralet goethitt. Det andre kartet dekker spekteret fra 1300 nm til 2500 nm mikrometer og viser distribusjonen av mineralet alunitt. Disse kartene fungerer som sammenligningsgrunnlag i analysen

på de områdene bildet fra HySpex ikke dekker.

Bahr (2022a) fra HySpex har i nyere tid gjennomført en analyse i Cuprite, med hyperspektrale bilder fra drone. Dataen fra dette oppdraget brukes som grunnlag for analysen. Rapporten benytter programmet ENVI som kan kobles opp mot det spektrale biblioteket til USGS (Kokaly mfl., 2017). Denne metoden sammenligner spektralsignaturer med høy presisjon.

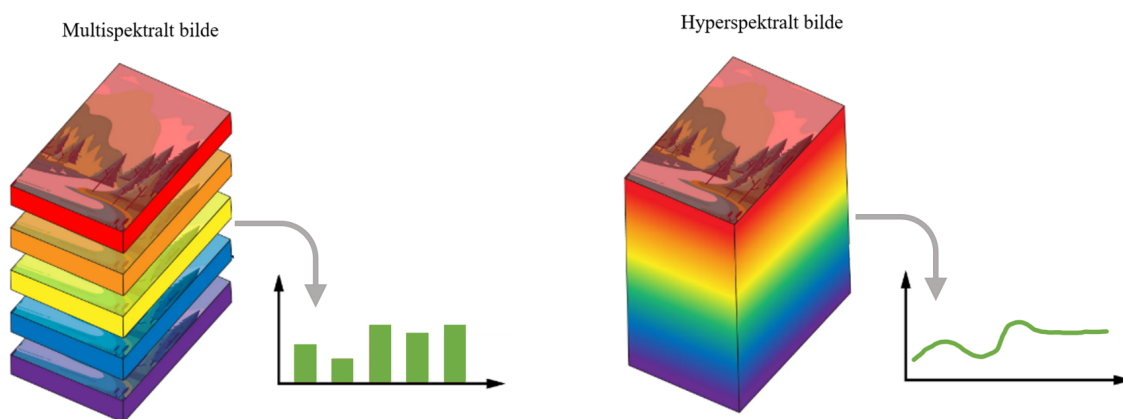
2 Teoridel

2.1 Fjernmålingsteknologi

Fjernanalyse omhandler å skaffe informasjon om jordens overflate uten å være i direkte kontakt med den. Fjernanalyse bygger på prinsippet om deteksjon av energi reflektert fra jordoverflaten. Fjernmålingsteknologi kan kategoriseres i aktive og passive sensorer. En aktiv sensor, slik som en laser, sender ut et signal og måler energien som reflekteres tilbake. I motsetning, fanger en passiv sensor opp reflektert energi fra naturlige eller kunstige lyskilder (Lied og Birkeland, 2022). Denne studien benytter seg av tre ulike passive fjernmålingssensorer, hvorav to er fra satellitter og én fra drone.

2.1.1 Hyperspektral bildebehandling

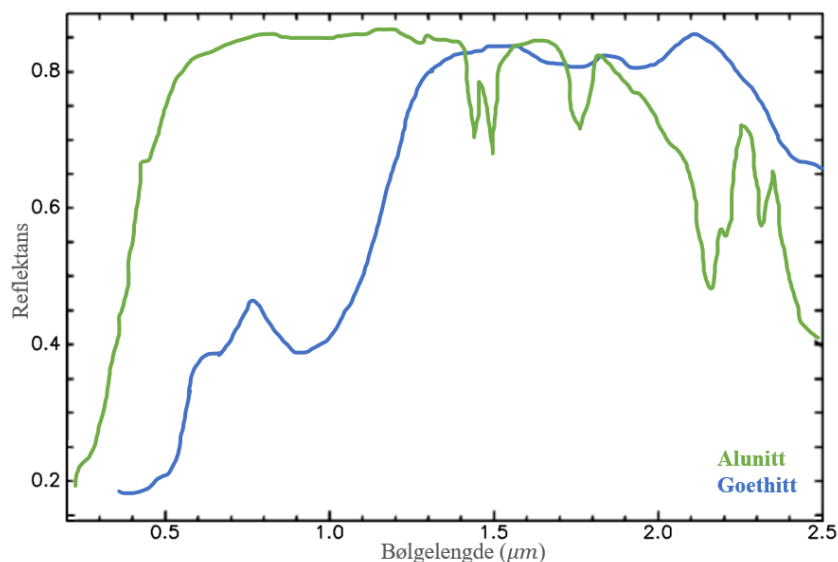
I denne analysen gjøres det en sammenligning av multispektrale og hyperspektrale bilder. Multispektrale bilder består typisk av et mindre antall bånd, vanligvis fra 3 til 10. Hyperspektrale bilder inneholder derimot et mye større antall bånd, ofte fra hundrevis til tusenvis. Multispektrale bånd fanger opp et bredere spekter av bølgelengder innenfor hvert bånd, mens hyperspektrale bånd er smalere. Dette fører til finere spektral oppløsning, noe som visualiseres i Figur 2.1. Dette gjør hyperspektrale bilder i stand til å differensiere subtile forskjeller i det elektromagnetiske spekteret (GISGeography, 2023).



Figur 2.1: Illustrasjonen viser et multispektralt bilde til venstre og et hyperspektralt bilde til høyre med tilsvarende spektralsignatur. Figur inspirert av EdmundOptics (2024)

2.1.2 Spektralsignatur

Forskjellige materialer har forskjellige spektralsignaturer, derfor kan spektralsignaturen brukes til å identifisere ulikheter. Elektromagnetisk stråling i kontakt med et materiale vil til en viss grad absorbere, reflektere eller emittere stråling av ulik bølgelengde. Utfallet av dette er en spektralsignatur, og kalles ofte stoffets «fingeravtrykk» (Carrion mfl., 2021). Figur 2.2 viser spektralsignaturen til de to aktuelle mineralene i denne analysen, alunitt og goethitt.



Figur 2.2: Illustrasjonen viser spektralsignaturen fra 0 μm til 2.5 μm til alunitt og goethitt fra Kokaly mfl. (2017). Alunitt er visualisert i grønt, mens goethitt er visualisert i blått. Y-aksen representerer reflektans mens x-aksen representerer bølgelengden.

2.1.2.1 Absorpsjon i atmosfæren

Atmosfærisk absorpsjon er prosessen der atmosfæren beholder en del av den innkommende solstrålingen eller jordoverflatens utstråling, i stedet for å la den passere gjennom eller reflekteres tilbake til verdensrommet (Birkeland, 2022).

Atmosfærisk absorpsjon påvirker nøyaktigheten i fjernmåling av mineraler, spesielt innenfor bølgelengdeområdet til NIR og SWIR. Disse båndene, som er avgjørende for å fange opp de unike spektralsignaturene til mineraler, er særlig sårbare for absorpsjon (GISGeography, 2024b). Derfor er det enklere å tolke mineralenes spektrale signaturer nøyaktig i et laboratoriemiljø sammenlignet med data innhentet fra fjernmåling, hvor atmosfæriske forstyrrelser påvirker målingene.

2.1.2.2 Spektralsignatur fra fjernmåling

En utfordring ved avlesning av spektralsignaturen fra fjernmåling er at alle piksler er av blandet innhold, uansett størrelse. Dette vil i stor grad påvirke spektralsignaturen, da den alltid vil ha en påvirkning sammensatt av forskjellige mineraler, støv og lysforhold.

2.1.3 Radians og reflektans

Reflektans beskriver materialets naturlige evne til å reflektere og absorbere ulike bølglengder av lys, representert på en skala fra 0 til 1 eller fra 0 til 100 prosent. Radians er signalet som måles av sensoren i øyeblikket, og avhenger av både belysningens intensitet og objektets refleksjonsegenskaper (Aarsten, 2022). Radians på den ene siden avhenger av flere eksterne faktorer inkludert atmosfæriske forhold, mens reflektans på den andre siden er en iboende egenskap ved objektet (NV5, 2013).

2.2 Mineraler

I denne analysen fokuseres det på å detektere mineralene alunitt og goethitt. Disse er valgt på grunn av tidligere forskning som har identifisert dem i studieområdet (Swayze mfl., 2014). I tillegg har disse mineralene svært forskjellige spektralsignaturer, noe som gjør dem enkle å skille fra hverandre.

Alunitt er et mineral som dannes fra vulkansk svovelsyredamp (Editors of Encyclopaedia Britannica, 2021). Det fremtrer vanligvis i hvitt, grått eller rødt, og består kjemisk av aluminium, kalium og sulfat (Bryhni, 2020a). Alunitt har en veldig gjenkjennelig spektralsignatur vist i Figur 2.2, spesielt i kortbølget infrarødt (SWIR) område (Kokaly mfl., 2017). I Figur 2.3 blir mineralet alunitt visualisert til venstre.

Goethitt er et vanlig jernmineral, og utgjør hovedbestanddelen i jernrust. Fargen varierer fra gulbrun til rød (Editors of Encyclopaedia Britannica, 2019). Mineralet oppstår primært gjennom oksidasjon og forvitring av andre jernrike mineraler (Bryhni, 2020b). I Figur 2.3 blir mineralet goethitt visualisert til høyre.



Figur 2.3: Figuren viser hvordan mineralet alunitt til venstre ser ut, og hvordan mineralet goethitt til høyre ser ut. Bildet av alunitt er hentet fra (Le Comptoir Geologique, u.d) og bildet av goethitt er hentet fra (Astro Gallery of Gems, 2024).

2.3 Maskinlæring

Maskinlæring er en type teknologi som etterligner menneskelig læring ved å bruke data fra omgivelsene (El Naqa og Murphy, 2015). Istedenfor å følge strengt definerte regler, kan maskinlæringsalgoritmer identifisere mønstre i data og bruke disse til å ta beslutninger eller generere prediksjoner (Sharifani og Amini, 2023).

2.3.1 Ikke styrt klassifisering

I ikke styrt klassifisering utforsker maskinen datasettet på egenhånd, uten å ha noen forhåndsdefinerte mål eller veiledning om hva den skal lete etter. Istedenfor å ha et mål å predikere eller klassifisere, søker maskinen etter naturlige strukturer eller mønstre i dataene (Banoula, 2023). Et eksempel på en ikke styrt klassifisering som benyttes i denne analysen er en Principal Component Analysis.

2.3.1.1 PCA

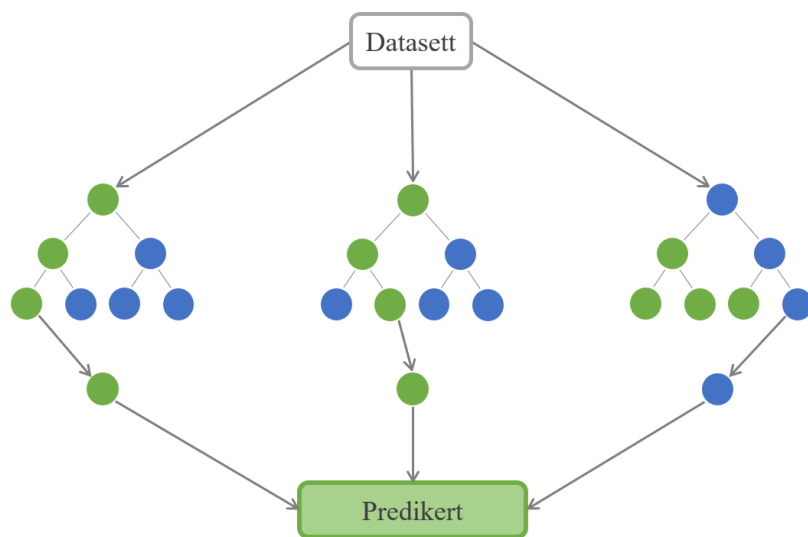
Prinsipalkomponentanalyse (PCA) er en teknikk som gjør det lettere å forstå store datasett. Metoden finner de viktigste mønstrene ved å se på forskjeller i dataene, og deretter redusere antall dimensjoner uten at viktig informasjon går tapt. Dette gjør det enklere å jobbe med og forstå store mengder informasjon (Shani og Binieli, 2019).

2.3.2 Styrte klassifisering

I styrte klassifisering blir maskinen trent med veiledning, der den lærer fra et datasett som inneholder allerede kjente resultater. Denne metoden gjør det mulig for maskinen å bruke en modell som kan forutsi fremtidige utfall basert på denne tidligere kunnskapen (Banoula, 2023). Et eksempel med styrte klassifisering som benyttes i denne analysen er Random Forest.

2.3.2.1 Random Forest

Random Forest er en populær klassifiseringsalgoritme som benytter flere beslutningstrær samtidig, vist i Figur 2.4. Metoden fungerer ved å trene hvert tre på forskjellige deler av datasettet, og deretter kombinere resultatene for å danne en endelig beslutning. En av fordelene med Random Forest er dens evne til å levere nøyaktige resultater samtidig som den er mindre utsatt for overtilpasning sammenlignet med enkelte andre algoritmer (Schonlau og Zou, 2020). Dette skyldes dens evne til å håndtere ulike deler av datasettet uavhengig av om dataene er skalert eller ikke. Random Forest viser seg også å være mer robust mot støy og variasjoner i treningsdataene, noe som gjør den til et pålitelig verktøy for klassifisering i ulike scenarier (Belgiu og Drăguț, 2016).



Figur 2.4: Forenklet illustrasjon av maskinlæringsalgoritmen Random Forest. Figuren viser hvordan datasettet fordeles og behandles gjennom flere beslutningstrær. Hver gren representerer et valg, som leder til det endelige predikerte resultatet nederst i figuren. Figur inspirert av Schonlau og Zou (2020).

2.3.3 Validering av maskinlæringsmetoden

Validering av maskinlæringsmetoder er essensielt for å sikre påliteligheten til modellene (Polyzotis mfl., 2019).

2.3.3.1 Test- og treningsdata

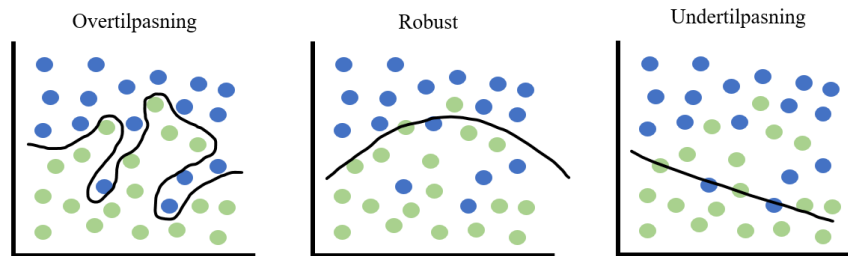
Test- og treningsdata er essensielt i maskinlæring. Treningsdata brukes for å lære opp maskinlæringsmodeller til å gjenkjenne mønstre og ta beslutninger. Det brukes vanligvis mer treningsdata enn testdata, vist i Figur 2.5, for å sikre at modellen er godt forberedt på forskjellige situasjoner. Etter trening testes modellen med testdata, som er ny og uavhengig data modellen ikke har sett før. Dette sjekker om modellen kan bruke det den har lært på nye situasjoner. Testdataene må være representative for den faktiske dataen og store nok til å gi pålitelige resultater (Barkved, 2022).



Figur 2.5: Grafisk fremstilling av hvordan datasettet er delt inn i trening- og testdata. De mørkegrønne blokkene representerer treningsdata og er majoriteten av dataen. En mindre del av dataen er satt til test av modellen, og er visualisert i lys grønn.

2.3.3.2 Over- og undertilpasning

Overtilpasning og undertilpasning, som vist i Figur 2.6, er kjente begreper i maskinlæring, og sier noe om hvor god modellen er. Overtilpasning skjer når modellen er veldig kompleks og tilpasses treningsdataene for nøye. Dette gjør at modellen ikke kan forutsi nye data nøyaktig. Underpasning skjer når modellen er for enkel og ikke fanger opp mønstrene i dataene skikkelig. Dette gjør at modellen ikke fungerer godt på hverken trenings- eller testdata (Koehrsen, 2018).



Figur 2.6: Visuell sammenligning av overtilpasning, robust tilpasning og undertilpasning i modeller. Til venstre viser diagrammet overtilpasning, hvor modellen (svart kurve) følger treningsdataene (blå og grønne prikker) for nøye og fanger opp støy. I midten illustreres en robust tilpasning med en balansert modell. Til høyre er et eksempel på undertilpasning, hvor modellen ikke klarer å fange opp den underliggende trenden i dataene. Figuren er inspirert av MathWorks (u.d).

2.3.3.3 Forvirringsmatrise

Forvirringsmatrisen er en tabell som brukes til å definere ytelsen av til en klassifiseringsalgoritme, ved å vise sammenhenger mellom de forventede og predikerte klassifiseringene utført av algoritmen (Grandini mfl., 2020, s. 2). Ut fra forvirringsmatrisen kan man hente verdier for sanne positive (TP), falske positive (FP), falske negative (FN) og sanne negative (TN). Figur 2.7 viser forholdet mellom forventede og predikerte verdier.

| | | Forventet | |
|-----------|---|-----------|----|
| | | a | b |
| Predikert | a | TP | FP |
| | b | FN | TN |

Figur 2.7: Figuren illustrerer en forvirringsmatrise som viser forholdet mellom forventet og predikerte verdier.

2.3.3.4 Valideringsparametere

Presisjon, sensitivitet og F1-verdi er viktig for å evaluere klassifiseringsmodellens ytelse, og kalles valideringsparametere. Disse metrikkene er avgjørende for å forstå og forbedre modellens effektivitet i ulike scenarier (Grandini mfl., 2020).

I denne analysen er det benyttet et makrogjennomsnitt av alle valideringsparameterene. Dette gjennomføres ved å regne ut valideringsparameterene for hver klasse først, deretter ta gjennomsnittet av alle sammen. Dette brukes for å behandle alle klasser likt, uavhengig av klassens størrelse, og er nyttig når det arbeides med datasett med ubalanserte klasser, slik som i denne analysen (Sefidian, 2022).

- Formel 2.1 viser presisjon og fokuserer på korrekte positive prediksjoner ved å beregne andelen ekte positive tilfeller blant alle positive prediksjoner (Grandini mfl., 2020, s. 7).
- Formel 2.2 viser sensitivitet og måler hvor godt modellen finner alle faktiske positive tilfeller ved å beregne andelen ekte positive tilfeller blant alle faktiske positive tilfeller (Grandini mfl., 2020, s. 7).
- Formel 2.3 viser F1-verdi og balanserer presisjon og sensitivitet for å gi et helhetlig bilde av modellens klassifiseringsevne. Det beskrives vanligvis som det harmoniske gjennomsnittet av de to. Harmonisk gjennomsnitt er en annen måte å beregne gjennomsnitt på og anses ofte som mer egnet for forholdstall (Kanstrén, 2020, Grandini mfl., 2020, s. 7).

$$\mathbf{Presisjon}_{\text{makro}} = \frac{TP_{\text{gjennomsnitt}}}{TP_{\text{gjennomsnitt}} + FP_{\text{gjennomsnitt}}} \quad (2.1)$$

Formel 2.1 Formel for makrogjennomsnitt av presisjon

$$\mathbf{Sensitivitet}_{\text{makro}} = \frac{TP_{\text{gjennomsnitt}}}{TP_{\text{gjennomsnitt}} + FN_{\text{gjennomsnitt}}} \quad (2.2)$$

Formel 2.2 Formel for makrogjennomsnitt av sensitivitet

$$\mathbf{F1-verdi}_{\text{makro}} = 2 \cdot \frac{\text{Presisjon}_{\text{makro}} \times \text{Sensitivitet}_{\text{makro}}}{\text{Presisjon}_{\text{makro}} + \text{Sensitivitet}_{\text{makro}}} \quad (2.3)$$

Formel 2.3: Formel for makrogjennomsnitt av F1-verdi

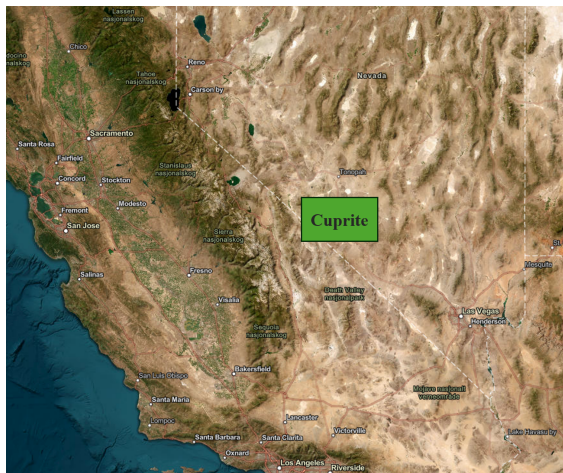
3 Metode

Metodekapitlet gir en detaljert gjennomgang av studieområdet, datasettene og programvarene som er benyttet. Oppbygging og fremgangsmåte for anvendelse og testing av maskinlæringsmodellen blir også beskrevet. Kapitlet gir et solid grunnlag for å forstå hvordan oppgaven ble gjennomført og resultatene ble oppnådd.

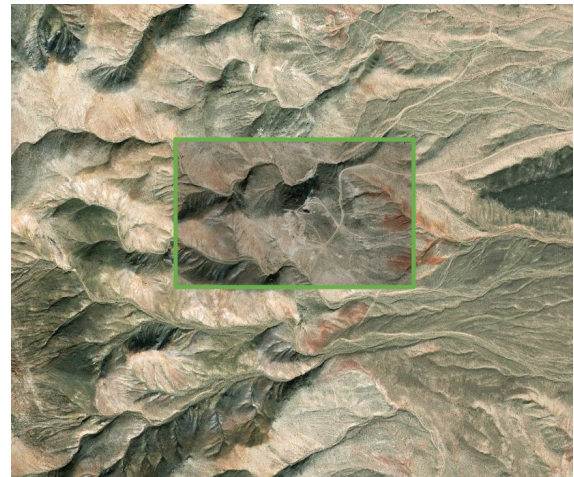
3.1 Data

3.1.1 Studieområde

Studieområdet ligger i Cuprite-fjellene i Nevada og dekker et område på omtrent 300 meter x 600 meter, tilsvarende 0,18 kvadratkilometer. Cuprite er spesielt godt egnet for forskning med fjernmåling på grunn av sin tydelige soneinndeling og unike mineralsammensetning (Hill, 2013). Figur 3.1a og Figur 3.1b viser studieområdet.



(a) Figuren viser oversiktsbilde av studieområdet. Studieområdet, markert med et grønt polygon. Bildet er hentet fra plattformen ArcGIS Online.

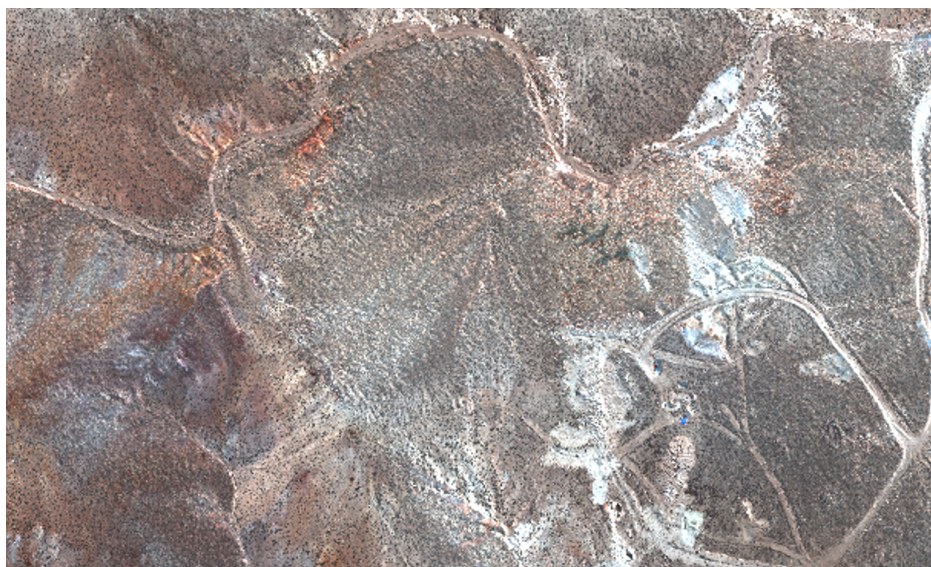


(b) Figuren viser studieområdets størrelse (300m x 600m), visualisert med et grønt omriss. Bildet er hentet fra plattformen ArcGIS Online.

3.1.2 Sensorer

3.1.2.1 HySpex Mjølner VS-620

Bildene fra HySpex er tatt med Mjølner VS-620-sensoren fra en drone, som har 490 bånd i spekteret 400 nm til 2500 nm (HySpex, 2019). Datasettet fra HySpex er samlet inn 25. februar 2020. HySpex leverer avansert hyperspektral avbildningsteknologi (NEO, u.d), og tillater nøyaktig analyse av materialer og objekter basert på deres spektrale signaturer. Figur 3.2 viser bildet som ble mottatt fra HySpex, og som danner grunnlaget for analysen.



Figur 3.2: Figuren viser bildet av studieområde tatt med sensoren fra HySpex. Bildet er visualisert i RGB, med rødt bånd: 640 nm, grønt bånd: 550 nm og blått bånd: 463 nm.

3.1.2.2 Sentinel-2 Copernicus

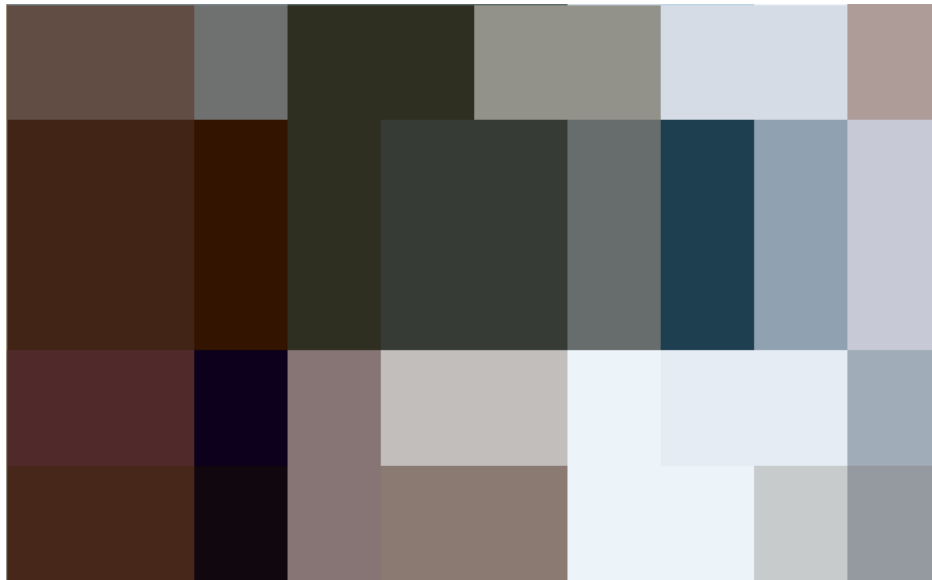
I denne analysen blir det brukt satellittbilder fra Sentinel-2, som leverer multispektrale bilder fordelt på 13 spektrale bånd, med en geometrisk oppløsning som varierer fra 10 til 60 meter (European Space Agency, 2023c). Sentinel-2-banene er solsynkrone, noe som betyr at de opprettholder en jevn sollysvinkel over jordoverflaten. Dette reduserer skyggeeffekter og sikrer påliteligheten til tidsseriedata, som er essensielt for nøyaktige vurderinger over tid (European Space Agency, 2023b). Dataene, presentert i Figur 3.3, er bevisst hentet ut til den datoen som var nærmest mulig HySpex-bildet. Sentinel-2-bildet som er brukt i analysen er datert til 23. februar 2020, kun to dager etter innsamlingen av HySpex-bildet.



Figur 3.3: Figuren viser bildet av studieområde tatt med satellitten Sentinel-2. Bildet er visualisert i RGB, med rødt bånd: 665 nm, grønt bånd: 560 nm og blått bånd: 490 nm.

3.1.2.3 EMIT NASA

Den andre satellitten brukt i denne analysen er EMIT, forkortelse for Earth Surface Mineral Dust Source Investigation, lansert av NASA sommeren 2022. Hensikten med EMIT er å kartlegge mineralstøv, som kan ha betydelige miljøeffekter. Sterke vinder kan bære mineralstøv fra tørre og ørkenaktige områder over store distanser. Avhengig av partiklenes farge og sammensetning, kan de enten absorbere sollys og varme opp jordoverflaten, eller reflektere sollys og virke kjølede (NASA, 2023a). EMIT leverer hyperspektrale bilder med en oppløsning på 60 meter fra en høyde på 400 km, og dekker bølgelengder fra 380 nm til 2500 nm fordelt på 285 bånd (NASA, 2023a). Dataene, som er vist i Figur 3.4, ble samlet inn den 4. august 2023, over tre år etter de andre sensorene, da EMIT ikke var i drift før 2022.



Figur 3.4: Figuren viser bildet av studieområde tatt med satellitten EMIT. Bildet er visualisert i RGB, med rødt bånd: 664 nm, grønt bånd: 559 nm og blått bånd: 492 nm.

3.1.2.4 Sammenligning av datasettene

Tabell 3.1 viser en sammenligning de 3 sensorene som er benyttet i analysen. Spektral oppløsning og geometrisk oppløsning er de spesifikasjonene som spiller størst rolle i denne analysen. Spektral oppløsning er sensor-spesifikk, mens geometrisk oppløsning er oppdragsspesifikk.

Tabell 3.1: Sammenligning av spesifikasjonene til datasettene fra sensorene HySpex Mjølner (HySpex, 2019), Sentinel-2 (European Space Agency, 2023a) og EMIT (NASA, 2023b)

| Spesifikasjoner | HySpex Mjølner | Sentinel-2 | EMIT |
|--------------------------------|----------------|-------------|-------------|
| Spektral oppløsning | 490 | 13 | 285 |
| Geometrisk oppløsning | 0.18 m | 10/20/60 m | 60 m |
| Radiometrisk oppløsning | 16 bit | 16 bit | 32 bit |
| Spektralbredde | 400-2500 nm | 443-2190 nm | 380-2500 nm |
| Flyhøyde | fra 20 m | 786 km | 370-460 km |
| FOV | 20° | 20° | 11° |
| Swath width | 72 m | 290 km | 75 km |

3.2 Programvarer

3.2.1 Python

Python er et «open source» programmeringsspråk, brukt av utviklere verden over. Språket er enkelt og brukervennlig, samtidig kan det brukes til å utvikle komplekse programmer. Derfor er Python populært for en stor målgruppe. Språket støtter anvendelser som automatisering, dataanalyse, maskinlæring og webutvikling. Python har en omfattende samling av biblioteker og innebygde funksjoner, noe som gjør verktøyet både robust og godt egnet for å håndtere geografiske data (Nätt, 2023).

3.2.2 Catalyst Professional Focus

Catalyst Professional Focus, fra PCI Geomatics (tidligere kjent som Geomatica), er en brukervennlig programvare for geografisk informasjon. Programvaren inneholder verktøy for blandt annet fjernmåling, digital fotogrammetri, bildeanalyse, kartproduksjon og mosaikklegging (Catalyst, 2024; PCI Geomatics, 2013).

3.2.3 QGIS

QGIS er et åpent og gratis geografisk informasjonssystem som brukes til å visualisere, redigere og analysere geografiske data. Programvaren tilbyr omfattende verktøy for kartvisning, geoprosessering og analyse (Dawson, 2024). Med støtte for plugins kan brukerne utvide funksjonaliteten, noe som gjør QGIS populært både blant nybegynnere og profesjonelle innen geovitenskap og kartografi.

3.2.3.1 Dzetsaka - Klassifiseringsverktøy

«Dzetsaka» er et klassifiseringsplugin for QGIS utviklet av Karasiak (2016). Pluginet er kjent for sin hastighet, enkelhet i bruk og robuste funksjonalitet. Verktøyet trenger rasterdata som grunnlag og en shapefil som treningsdata for å utføre klassifiseringene. Verktøyet støtter 4 maskinlæringsmetoder, Gaussian Mixture Model, Random Forest, K-Nearest Neighbour og Support Vector Machine (Karasiak, 2016; Pinasu, 2020). Pluginet installeres enkelt i QGIS fra plugin-menyen, men for å kunne bruke alle algoritmene i Dzetsaka avhenger det av at tilleggspakker er installert, som Scikit-learn.

3.2.4 Kunstig intelligens

KI, representert ved språkmodellen ChatGPT 3.5 (OpenAI, 2023), har forbedret effektiviteten og nøyaktigheten i oppgaven på følgende måter:

1. Kodeoptimalisering: ChatGPT har identifisert og rettet flere feil i programmeringskoden, noe som har resultert i mer effektive og feilfrie skript.
2. Tekstforbedring: ChatGPT har gitt forslag til strukturelle og språklige endringer for å forbedre tekstkvaliteten.
3. Dokumentformatering: KI har hjulpet med å organisere elementer i dokumentet, ved å blant annet å gi instruksjoner for formatering av figurer og tabeller i Latex.

3.2.5 Filformater

BSQ står for «band sequential» og er formatet den hyperspektrale dataen fra HySpex leveres i. BSQ er ikke i seg selv et bildeforamt, men en ordning som enkelt organiserer båndene i bilder. I BSQ-filer organiseres billedata slik at informasjonen til hvert bånd lagres etter hverandre. Det betyr at pikseldataene for ett bånd lagres først, deretter for det neste båndet, og så videre. Denne organiseringen gjør det lettere å jobbe med dataene og gir enkel tilgang til hele bildet for hvert bånd. BSQ-filer er binære og trenger en tilhørende ASCII-filheader for riktig tolkning. Headeren inneholder ekstra informasjon om bildet, som antall rader og kolonner, om det er et fargekart, og geografiske koordinater (ArcMap, 2021).

BIL står for «band interleaved by line» og er formatet som dataen fra EMIT blir transformert til. Dette formatet deler noen likheter med BSQ. Det organiserer data i bånd, men i stedet for å lagre all informasjonen for et gitt bånd etter hverandre, lagres dataene for hver linje i bildet først for ett bånd, deretter neste bånd, og så videre. Dette formatet er også binært og krever en tilsvarende ASCII-filheader for å tolkes riktig. Headeren inneholder ekstra informasjon om bildet, slik som BSQ, og gir enkel tilgang til pikselverdiene for hvert bånd (ArcMap, 2021).

HDR er kalt «header»-filer, og er ASCII-filheaderen. Denne filtypen er en form for oppslagsbok som forteller datamaskinen hvordan den skal tolke de tilhørende bildefilene, enten de er i BSQ- eller BIL-format. Headeren inneholder viktig informasjon som antall bånd, antall rader og kolonner i bildet, og annen metadata som er avgjørende for riktig tolkning og behandling av dataene (ArcMap, 2021).

NetCDF4 er filformatet som brukes til å lagre dataen fra EMIT Earthdata. All informasjon som er nødvendig å forstå for å jobbe med dataene er organisert hierarkisk inne i filen. Dette inkluderer grupper, variabler og metadata som er innlemmet direkte i filen. Datasettene inneholder ofte store mengder data (Gerdoo, 2022).

PIX er et format som er skapt av PCI Geomatics. Formatet brukes til å lagre geospasiale datasett som kan variere i størrelse. Filene inneholder flere bildekanaler som sammen utgjør den romlige informasjonen (FileInfo, 2024).

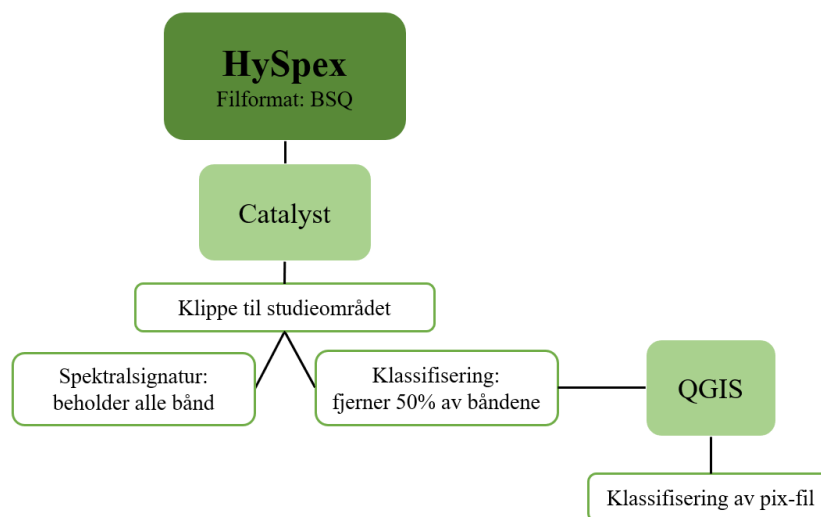
TIFF er formatet dataen fra Sentinel-2 er lastet ned som. En stor fordel med TIFF er det at ingen bildeinformasjon går tapt under komprimeringen. Dette gjør det mulig å opprettholde høy bildekvalitet samtidig som filstørrelsen reduseres. GeoTIFF-bildefiler, som også bruker TIF-filtypen, inkluderer GPS-koordinater som metadata i filen ved hjelp av TIFF-formatets utvidede funksjoner (Fisher, 2022).

Shapefil er et format som lagrer posisjonen, geometrien og attributtene til punkt-, linje- og polygonfunksjoner. Formatet er et vektor-databasert filformat som ofte brukes til geospasial analyse. Treningsdataen i klassifiseringen er i Shapefil-formatet (Precisely, u.d).

3.2.6 Preprosessering av data

3.2.6.1 HySpex

Selv om bildet fra HySpex allerede var korrigert for atmosfæriske forstyrrelser (Bahr, 2022a, s. 18), var det nødvendig med ytterligere prosessering før klassifisering med bildet kunne gjennomføres. Først importeres bildet til Focus, selve grunnpakken av Catalyst-programvaren. Bildet klippes så til et mindre område for å unngå tilfeller av «døde» piksler innenfor studieområdet, som videre kunne påvirke nøyaktigheten av analysen. På grunn av den store størrelsen på HySpex-datasettet (5 GB), forenkles bildet ved å fjerne annethvert bånd. Dette gjorde håndteringen av dataen mulig, samtidig som om viktig informasjon ble beholdt for en pålitelig klassifisering over hele spekteret. Figur 3.5 illustrerer prosessen fra mottakelsen av bildet fra HySpex til det var klart for klassifisering i QGIS.

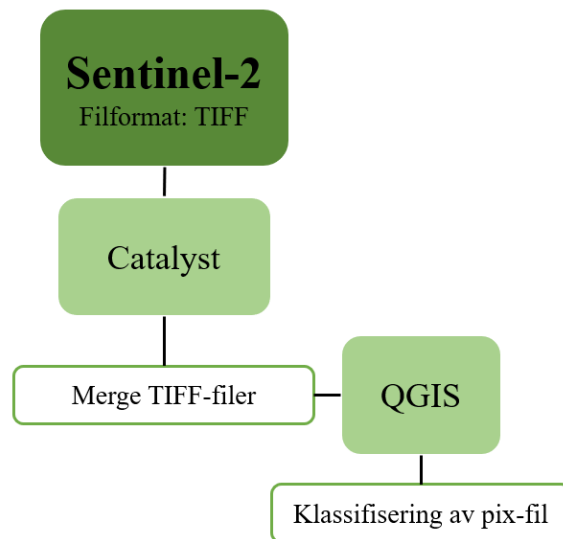


Figur 3.5: Figuren viser flytskjema av preprosesseringen som ble gjort med dataen fra HySpex før klassifisering. Programvarer er vist med grønn boks. Beskrivelse av hva som skjer i hver programvare vises med hvit boks med grønt omriss.

3.2.6.2 Sentinel-2

Bildene fra Sentinel-2 ble hentet fra LandViewer-portalen, som er en del av Earth Observation System (EOS) underlagt NASA. Dette er en åpen kilde hvor det kun kreves registrering som bruker for å laste ned bilder. Sentinel-2-bildene har en oppløsning på 10 meter for båndene B02, B03, B04 og B08, som tilsvarer blå, grønn, rød og nærinfrarød. Videre har de en oppløsning på 20 meter for båndene B05, B06, B07, B8A, B11 og B12, henholdsvis VNIR- og SWIR-båndene (GISGeography, 2024a).

For å unngå å jobbe med flere separate bilder, slås bildene sammen i Catalyst ved hjelp av «Data Merge»-verktøyet. Dette verktøyet tillater brukeren å velge ut bånd fra to eller flere kilder og sette dem sammen til ett nytt bilde. Bildet med høyest geometrisk oppløsning velges som grunnlag for å beholde en oppløsning på 10 meter. Ved en oppløsning på 10 meter blir båndene med en oppløsning på 20 meter splittet opp i fire nye piksler med en kunstig oppløsning på 10 meter, mens verdiene til de opprinnelige båndene blir bevart. I denne prosessen blir det opprinnelige TIFF-bildet transformert til PIX-format og deretter overført til QGIS for å starte analysen. Figur 3.6 viser hva som er gjort med bildet fra Sentinel-2 fra nedlasting til det var klart for å brukes i klassifisering.



Figur 3.6: Figuren viser flytskjema av preprosesseringen som ble gjort med dataen fra Sentinel-2 før klassifisering. Programvarer er vist med grønn boks. Beskrivelse av hva som skjer i hver programvare vises med hvit boks med grønt omriss.

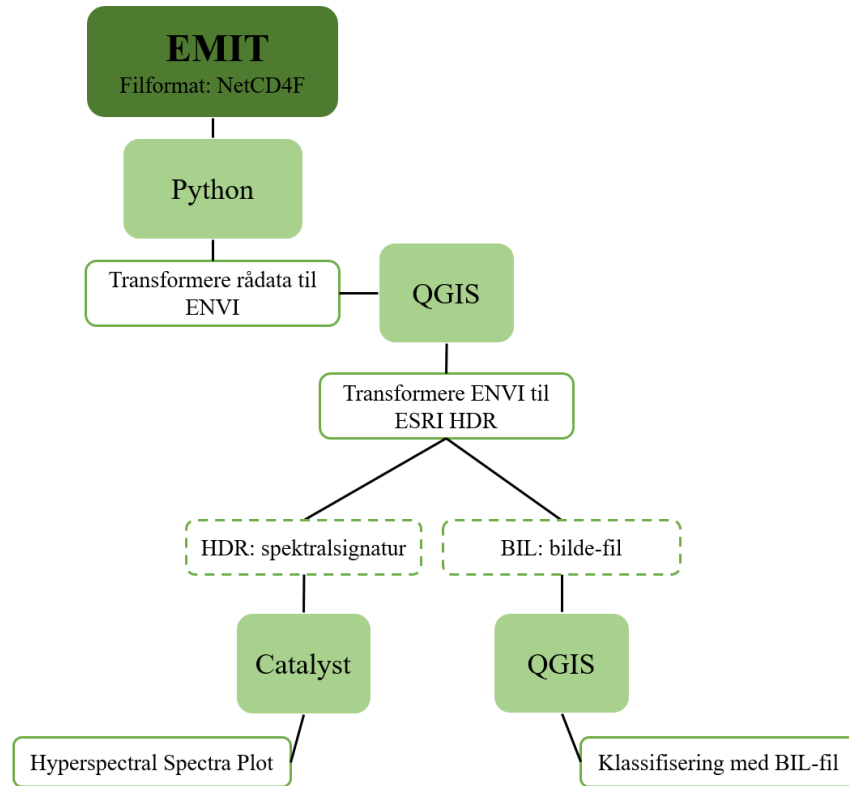
3.2.6.3 EMIT

Som ved nedlasting av Sentinel-2-data, tilbyr EMIT en egen portal kalt «Earthdata Search», som også er underlagt NASA. Brukerne har friheten til å definere et interesseområde ved å tegne et polygon eller markere et punkt av interesse. Deretter kan parametere som skydekke og tidspunkt justeres for å sikre best mulig datakvalitet og relevans. Etterpå kan brukerne enten laste ned bildet lokalt i NetCDF4-format eller kopiere en unik lenke til det aktuelle bildet.

For å behandle bildet i de aktuelle programvarene for denne analysen, benyttes et Github-repository utviklet av NASA. For å konvertere den nedlastede bildefilen fra «Earthdata Search» til et lesbart format, brukes «How to Convert to ENVI» (NASA, 2023d). Siden ENVI-formatet ikke direkte støttes av Catalyst-programvaren, kreves det en konvertering før bildet kan importeres. Den resulterende ENVI-filen åpnes deretter i QGIS, hvor den konverteres til «ESRI HDR»-formatet. Denne konverteringen resulterer i en BIL-fil og en HDR-fil, som begge er lesbare. BIL-filen brukes til klassifisering i QGIS, mens HDR-filen brukes til å avlese spektralsignaturene til pikslene i Catalyst.

I repositoryet finnes det et verktøy, «Exploring EMIT L2A Reflectance» (NASA, 2023c), som lar brukeren lese inn den aktuelle bildefilen, og analysere denne ved hjelp av interaktive plott. Her kan brukeren selv legge til punkter og få skrevet ut spektralsignaturen. Dette er en god løsning for å

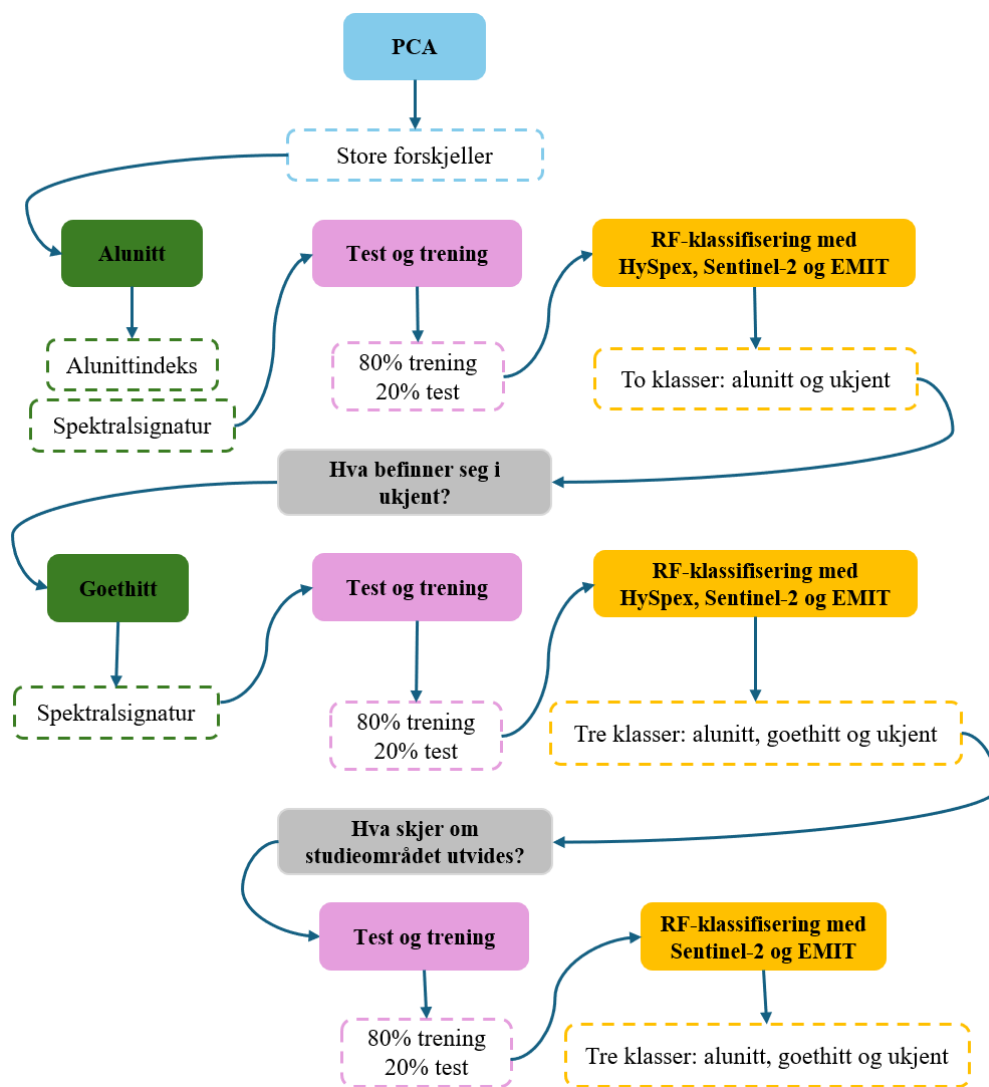
skaffe et overblikk av spektralsignaturen, men til en dypere analyse ble det erfart at det var mer nøyaktig å ta i bruk QGIS og Catalyst. Hva som ble gjort med dataen fra EMIT før den var klar for klassifisering i QGIS vises i Figur 3.7.



Figur 3.7: Figuren viser flytskjema av preprosesseringen som ble gjort med dataen fra EMIT før klassifisering. Programvarer er vist med grønn boks. Beskrivelse av hva som skjer i hver programvare vises med hvit boks med grønt omriss.

3.3 Analysens fremgangsmåte

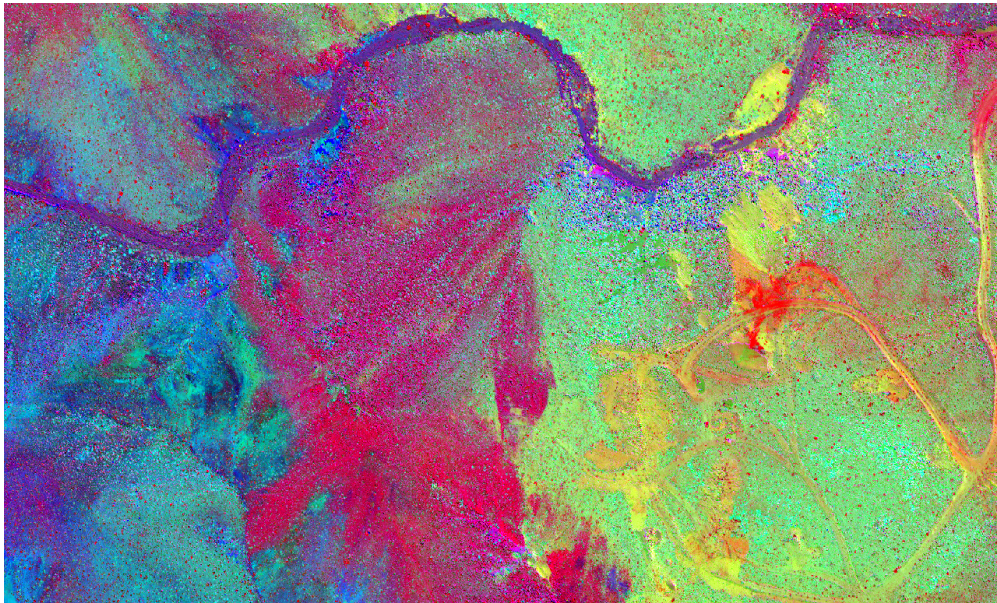
Figur 3.8 illustrerer en forenklet fremgangsmåte for databehandlingen som har blitt anvendt i analysen. Trening og testing av en Random Forest-klassifiseringsmodell utgjør kjernen i prosedyren. Pilene viser rekkefølgen på de ulike fargekoordinerte trinnene i prosessen. Blå boks illustrerer det første trinnet i analysen, en PCA. De grønne boksene illustrerer hvilke mineraler som letes etter. De rosa boksene illustrerer test- og treningsdata. De gule boksene illustrerer selve klassifiseringen. De grå boksene representerer sentrale spørsmål som leder analysen fremover, og viser hvordan prosessen har utviklet seg.



Figur 3.8: Figuren illustrerer analysens fremgangsmåte med et flytdiagram.

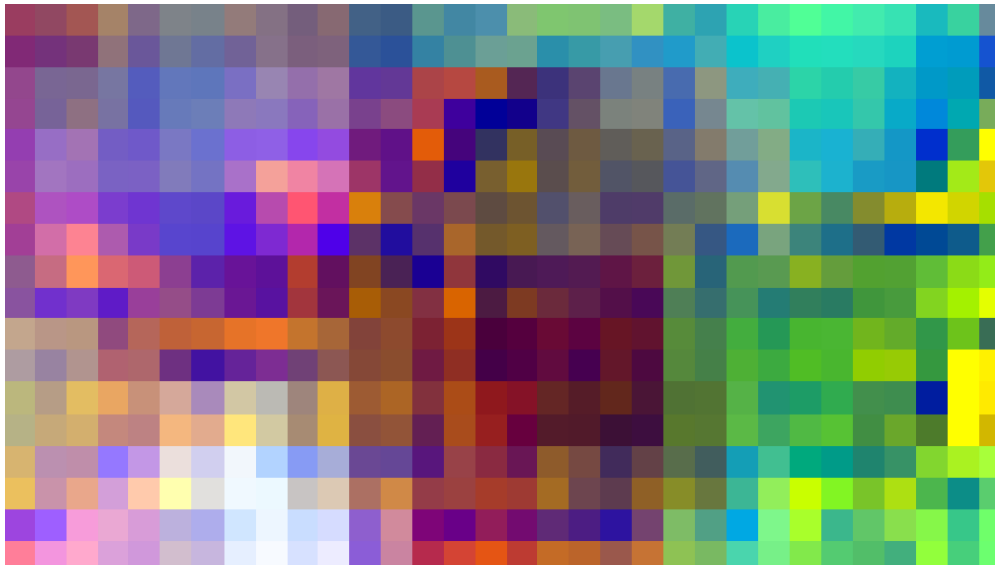
3.4 PCA

Etter å ha mottatt den høyoppløselige dataen fra HySpex over studieområdet, startet analysen med å identifisere forskjellene. En PCA ble utført for å fremheve disse forskjellene. Figur 3.9 visualiserer prinsipalkomponent 2, 3 og 4, der mesteparten av den nyttige informasjonen observeres. Den første prinsipalkomponenten visualiserer ofte lysstyrke og skyggeeffekter, og ble derfor ikke vist for å unngå forstyrrelser. Resultatene avslørte betydelige variasjoner i området, og det var særlig det fremtredende rosa/røde område midt i Figur 3.9 som vekket interesse.



Figur 3.9: Figuren viser en PCA med HySpex, med prinsipalkomponent 2, 3 og 4

En PCA utføres også med Sentinel-2 for å vurdere om de samme forskjellene som ble observert med HySpex, også var synlige her. Som forventet viser Figur 3.10 resultatene representert med andre farger enn i Figur 3.9. Det tydelig å se at lignende mønstre og trender er gjenkjennelige i begge datasettene.



Figur 3.10: Figuren viser en PCA med Sentinel-2, med prinsipalkomponent 2, 3 og 4

3.5 Klassifisering 1: alunitt

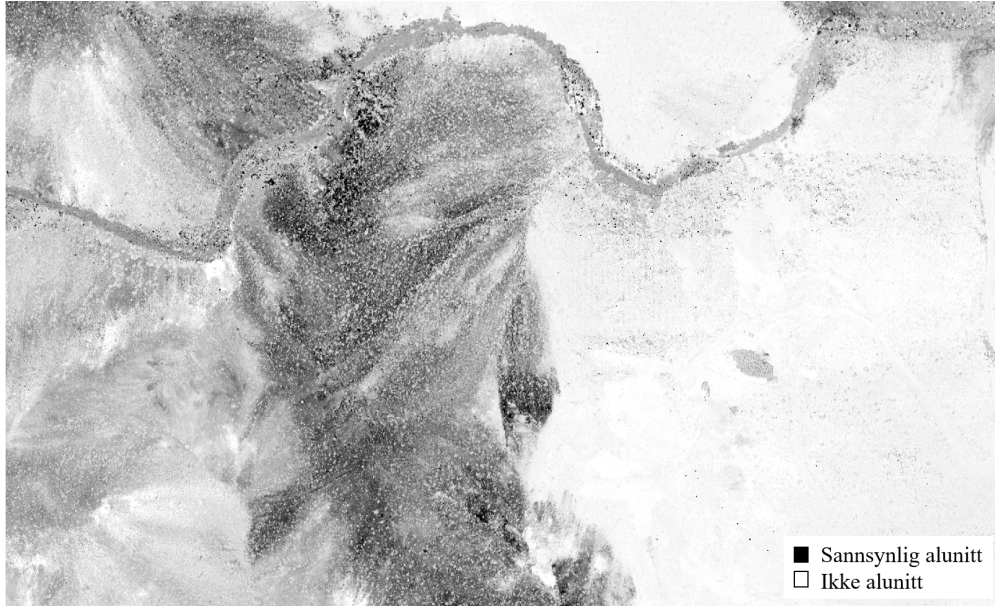
Via litteratursøk (Bahr, 2022b) viste det seg at det røde området fra PCAen trolig var mineralet alunitt. Hensikten med klassifisering 1 ble derfor å klassifisere alunitt.

3.5.1 Indeks for alunitt

Alunittindeksen er benyttet for å detektere mineralet alunitt. Indeksen er utformet av Ninomiya (2003), for å fremheve og detektere tilstedeværelsen av mineralet i hyperspektrale bilder. Spektralbåndene som benyttes i alunitt-indeksen velges ut i Raster Calculator-funksjonen i Catalyst. Indeksen er optimalisert for å avsløre karakteristiske absorpsjonsmønstre assosiert med alunitt. Indeksen vises i Formel 3.1, og er definert som refleksjon ved 2260 nm dividert med 2165 nm, ganget med refleksjon ved 2260 nm dividert med 2330 nm (Ninomiya, 2003).

$$\text{Alunittindeks} = \left(\frac{2260\text{nm}}{2165\text{nm}} \right) \times \left(\frac{2260\text{nm}}{2330\text{nm}} \right) \quad (\text{Ninomiya, 2003}) \quad (3.1)$$

I dette tilfellet er det brukt HySpex bånd 440, 421, 440 og 453, som korresponderer med de aktuelle bølgelengdene. Figur 3.11 viser bildet fra HySpex, etter indeksen var gjennomført. I dette bildet er hver enkelt piksel tilegnet en verdi ut ifra reflektansen til de utvalgte båndene brukt i indeksen.



Figur 3.11: Figuren viser resultatet fra alunitt-indeksen, presentert i en råfil, med varierende verdier. Høye verdier indikerer en større sannsynlighet for forekomst av mineralet alunitt. I visualiseringen representeres den høyeste verdien med svart, mens den laveste verdien vises som hvitt.

Figur 3.12 viser bildet fra HySpex etter en terskelverdi var satt for å justere funn fra indeksen. Terskelverdien ble satt ved hjelp av en visuell analyse av fremtredende piksler fra Figur 3.11. Med fremtredende piksler menes de mørkeste pikslene. Spektralsignaturene til disse pikslene ble evaluert opp mot spektralsignaturen til alunitt fremstilt i lab. Ut ifra denne analysen dannet verdiene til pikslene en øvre terskelverdi på 2.633 og en nedre terskelverdi på 1.395. Piksler med en verdi innenfor denne terskelen anses da som piksler med innhold av alunitt. Med denne terskelverdien ble også ekstremverdier og støy fjernet. Dette er en pekepinn for videre analyse.



Figur 3.12: Figuren viser resultatene fra alunitt-indeksen etter en satt terskelverdi. Pikslene som er visualisert i svart er piksler med verdier mellom en øvre terskelverdi på 2.633 og en nedre terskelverdi på 1.395. Piksler utenfor dette intervallet vises i hvitt.

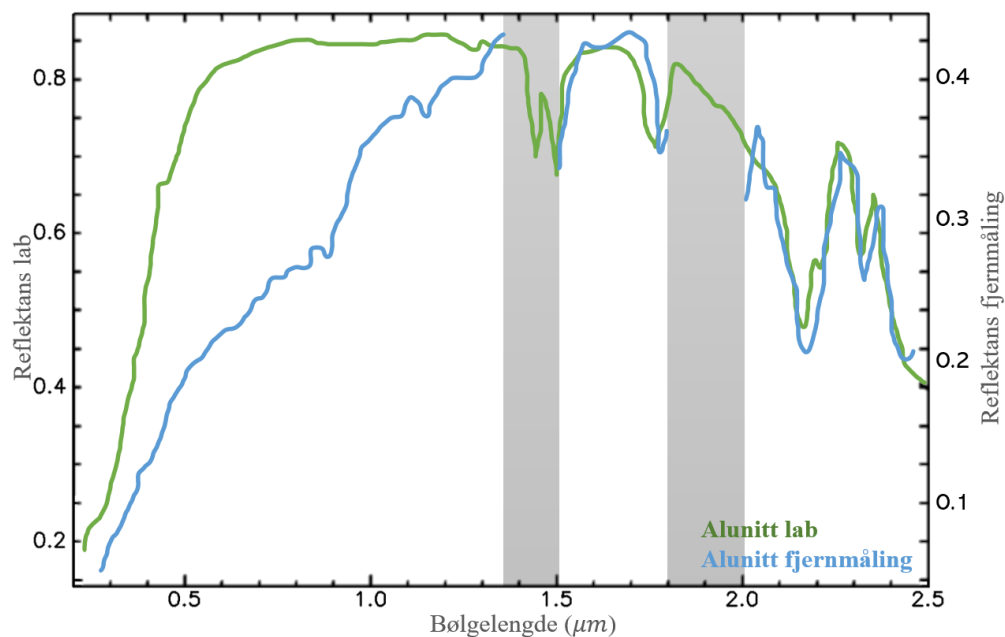
3.5.2 Generering av treningsdata

For å generere treningsdata ble alunitt-indeksen brukt for å identifisere mulige funnområder. Deretter verifiseres spektralsignaturene til pikslene i disse områdene og klassifiseres som enten «alunitt» eller «ukjent».

Focus har et verktøy kalt «Hyperspectral Spectra Plot» som tillater brukeren å visualisere spektralsignaturen til en valgt piksel. Imidlertid har ikke Focus en gunstig funksjon for å huske hvilken piksel som har en bestemt signatur. Selv om det er mulig å markere piksler i Focus, kan disse være krevende å eksportere som et shapefil-lag. Derfor utføres den videre analysen i QGIS. Basert på informasjonen fra spektralsignaturene, tildeles hver piksel en klasse i et nytt lag i QGIS.

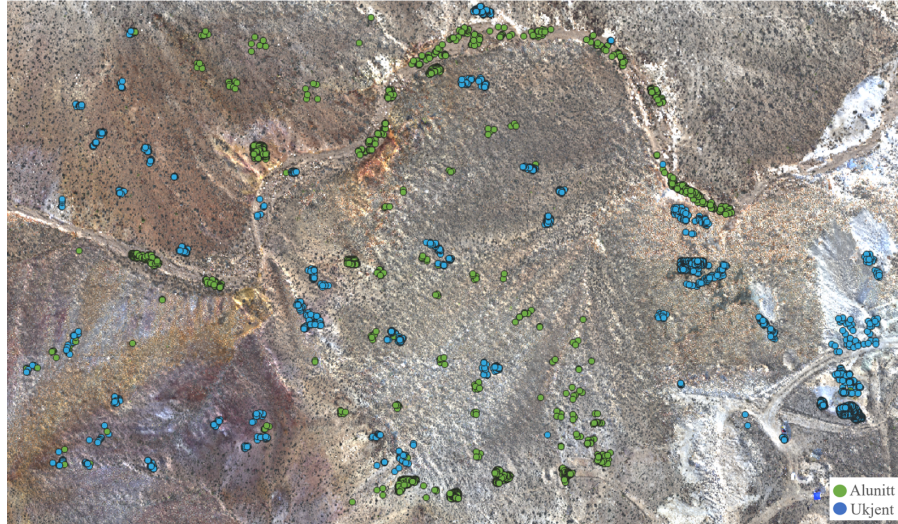
I valideringen av spektralsignaturen er det hovedsakelig toppen og bunnen innenfor intervallet 2000 nm til 2500 nm som er interessante, da disse absorpsjonsmønstrene er karakteristiske for alunitt (Kokaly mfl., 2017). Signaturen som brukes i valideringer er produsert i et laboratorium. Signaturen fremstilles ved at en preparert, homogen flate belyses av en kunstig lyskilde, og ikke av sollys, som er tilfellet ved fjernmåling. Dette betyr at signaturen avviker noe fra den som observeres i felt, da eksponering av sollys påvirker styrken på reflektansen. Av den grunn er det de karakteristiske toppene og bunnene til signaturen som er avgjørende for å sikre en nøyaktig identifikasjon av mineralet. Figur

3.13 viser signaturen til alunitt fra lab i grønn og et eksempel fra fjernmåling med HySpex fra denne analysen i blått. I Figur 3.13 kan de karakteristiske dippene for alunitt observeres. Disse er ved 2166 nm og 2321 nm. Basert på denne informasjonen blir pikselen klassifisert som alunitt, selv om de to signaturene avviker fra hverandre.



Figur 3.13: Figuren viser spektralsignaturen til alunitt fra lab og alunitt fra fjernmåling, vist i fullt spekter fra 0 nm til 2500 nm. Spektralsignaturen til alunitt lab hentet fra Kokaly mfl. (2017) og vises i grønt. Spektralsignaturen til alunitt fra fjernmåling er hentet fra HySpex-bildet i denne analysen. Y-aksen til venstre er reflektansen fra lab, mens y-aksen til høyre er reflektansen fra fjernmåling.

Figur 3.14 viser treningspikslene brukt i klassifisering 1. De grønne punktene representerer klassen «alunitt» mens de blå punktene representerer klassen «ukjent». Disse punktene er spredt jevnt over området for en robust klassifisering. Ved å dekke ulike terrengetyper og geologiske formasjoner, blir treningsdataene mer pålitelige. Dette gir bedre forståelse av alunittforekomster og reduserer risikoen for feil (Dubrov, 2023).



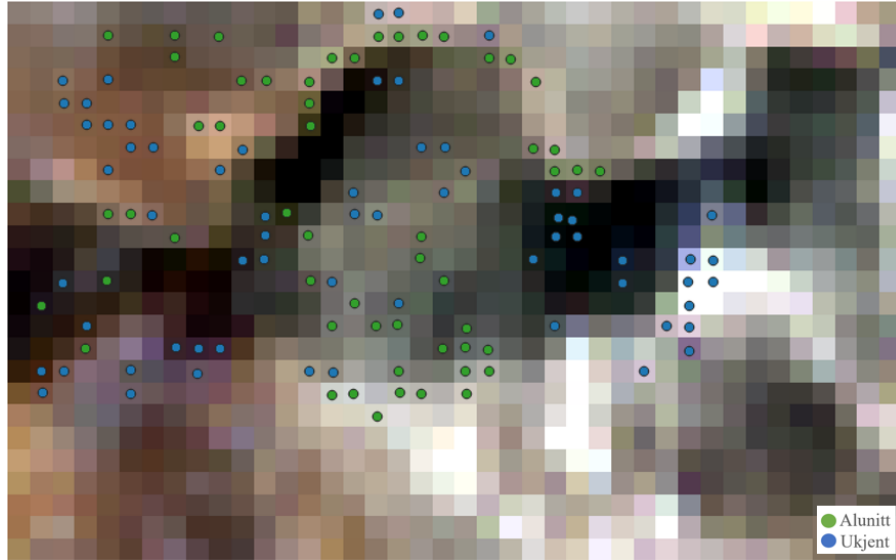
Figur 3.14: Figuren viser pikslene som er brukt for klassifiseringen av alunitt med HySpex og hvordan klassene er geografisk distribuert over terrenget. Alunitt er visualisert i grønt og de ukjente pikslene er visualisert i blått.

3.5.3 Treningsdata for Sentinel-2 og EMIT

Siden bildene fra Sentinel-2 og EMIT har betydelig færre piksler enn HySpex-bildet, var det nødvendig å utvikle nye trenings- og testsett for å gjennomføre klassifisering med EMIT og Sentinel-2.

3.5.3.1 Treningsdata for Sentinel-2

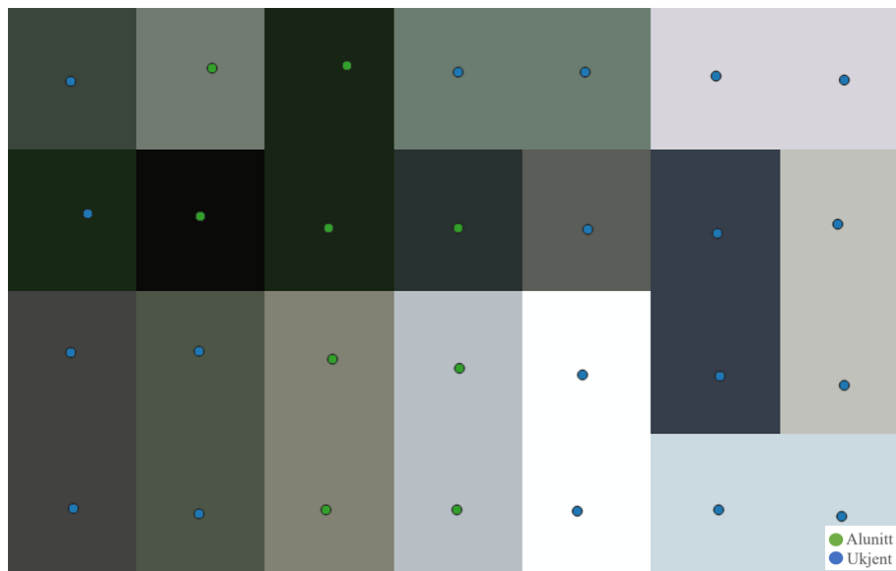
Sentinel-2 har en lav spektral oppløsning, noe som gjør det svært vanskelig å detektere markante topper og bunner i den spektrale signaturen. For å håndtere denne utfordringen, ble antallet treningspiksler fra HySpex-datasettet justert slik at det korresponderte med antallet piksler i det lavoppløste bildet fra Sentinel-2. Utgangspunktet var HySpex-pikslene for å skape nye treningspiksler. For å verifisere at pikselen var rett, ble spektralsignaturen fra bånd 12 fra Sentinel-2 (2190 nm) brukt. Dette er båndet hvor alunitt viser den sterkeste absorpsjonen. Om det ikke samsvarte, ble pikselen satt til klassen "ukjent". Figur 3.15 viser pikslene som ble benyttet i klassifiseringen med Sentinel-2.



Figur 3.15: Figuren viser pikslene som ble brukt i klassifiseringen av alunitt med Sentinel-2 og hvordan klassene er geografisk distribuert over terrenget. Alunitt er visualisert i grønt og de ukjente pikslene er visualisert i blått.

3.5.3.2 Treningsdata for EMIT

EMIT har en god spektral oppløsning, og genereres med samme fremgangsmåte som brukt ved HySpex-bildet, forklart i Kapittel 3.5.2. Figur 3.16 viser pikslene som ble benyttet under klassifisering 1 med EMIT.



Figur 3.16: Figuren viser pikslene som ble brukt i klassifiseringen av alunitt med EMIT og hvordan klassene er geografisk distribuert over terrenget. Alunitt er visualisert i grønt og de ukjente pikslene er visualisert i blått.

3.5.4 Treningsområde og testområde

For å sikre en robust klassifisering, har datasettet blitt delt opp i to distinkte sett: et treningspunktsett og et testpunktsett. Treningspunktsettet består av 80% av de tilgjengelige pikslene og er anvendt for å trene klassifiseringsmodellen. De resterende 20% av pikslene er tilfeldig valgt ut til å utgjøre testpunktsettet, og brukes til å evaluere modellen. Denne tilfeldige utvalgsprosessen bidrar til å sikre at modellen ikke bare tilpasses spesifikt til treningsdataen, men også har kapasiteten til å prestere på upåvirket data.

3.5.5 Gjennomføring av klassifisering 1: alunitt

Klassifiseringen av mineralet alunitt ble utført ved bruk av Dzetsaka - Klassifiseringsverktøy i QGIS, forklart i Kapittel 3.2.3.1. Detaljer rundt antallet trenings- og testpunkter som ble anvendt i denne prosessen, er oppgitt i Tabell 3.2. Klassifiseringen ble gjennomført med data fra alle tre sensorene.

Tabell 3.2: Tabellen viser antall piksler brukt i klassifiseringen av alunitt. Av totalt 4684 piksler, ble 3750 brukt for trening (80%) og 934 for testing (20%)

| | Treningsdata 80% | Testdata 20% |
|--------------|------------------|--------------|
| Alunitt | 1446 | 360 |
| Ukjent | 2304 | 574 |
| Total | 3750 | 934 |

3.6 Klassifisering 2: alunitt og goethitt

Etter en vellykket klassifisering av mineralet alunitt, ble det bestemt å styrke modellen ved å utforske potensialet for å identifisere et ytterligere mineral. Klassifiseringen avdekket betydelige områder innenfor klassen «ukjent», og det ble besluttet å utforske hvilke mineraler denne klassen kunne inneholde.

Inspirert av tidligere forskning på området, omtalt i Kapittel 1.3, falt valget på mineralet goethitt. Dette valget var motivert av to hovedgrunner: For det første indikerte tidligere studier tilstedeværelsen av goethitt i området. For det andre, goethitts kjemiske og fysiske egenskaper er markant forskjellige fra alunitt, noe som resulterer i svært forskjellige spektralsignaturer, slik at de enkelt kan skilles fra hverandre og omgivelsene (Kokaly mfl., 2017).

3.6.1 Indeks for goethitt

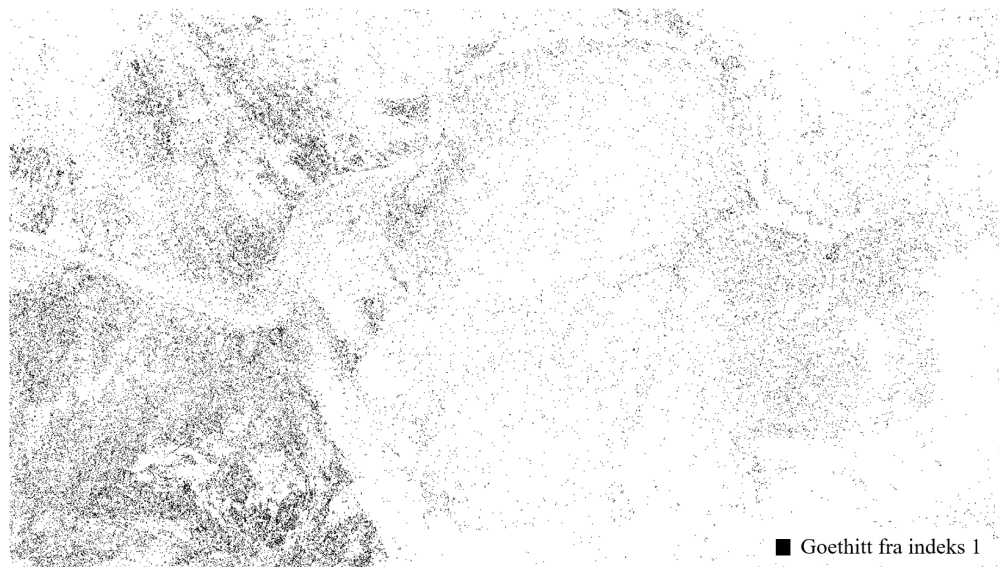
En indeks for goethitt ble ikke funnet gjennom litteratursøk, slik som for alunitt. Derfor ble det besluttet å utvikle indeksen selv. To forskjellige indekser ble utviklet for å identifisere goethitt.

3.6.1.1 Goethittindeks 1: utviklet fra andre indekser

Den første indeksen er basert på eksisterende indekser for jern (Rowan og Mars, 2003) og for hydroksylgrupper (Henrich mfl., 2011), ettersom goethitt består av akkurat disse to komponentene. Selv om den selvlagde indeksen ikke kan anses som like pålitelig som den som brukes for alunitt, fordi den ikke er eksternt validert, gir den et utgangspunkt for å starte søket etter goethitt. Dette forenklet prosessen med å finne treningspunkter. Formel 3.2 viser indeksen, og Figur 3.17 viser resultatet.

$$\text{Goethittindeks 1} = \frac{660\text{nm} - 560\text{nm}}{660\text{nm} + 560\text{nm}} * \frac{1650\text{nm} - 810\text{nm}}{1650\text{nm} + 810\text{nm}} \quad (3.2)$$

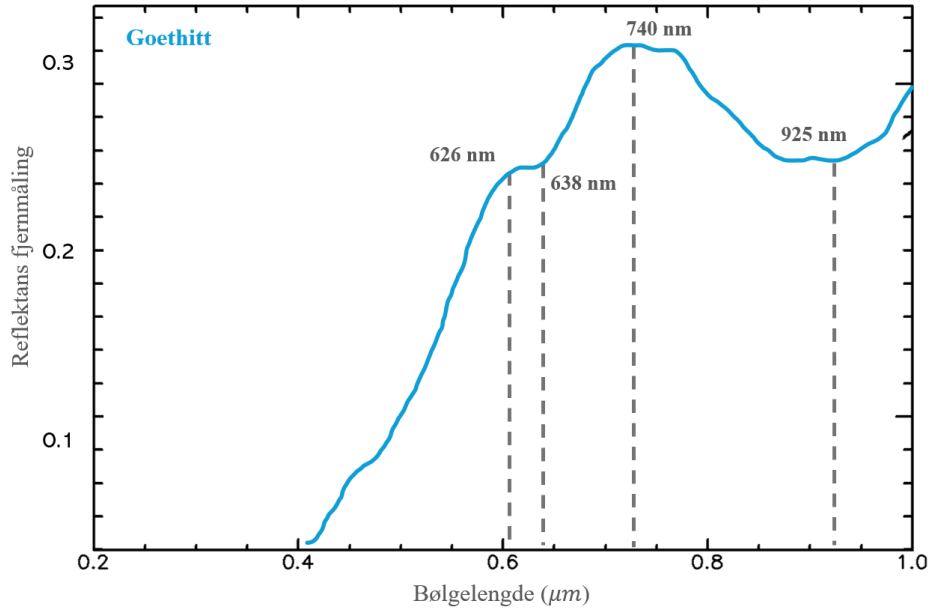
I dette tilfellet brukes HySpex bånd 87, 52, 290 og 138, som korresponderer med de aktuelle bølgelengdene. Figur 3.17 viser bildet fra HySpex etter en terskelverdi var satt for å justere funn fra indeksen. Terskelverdien ble satt ved hjelp av en visuell analyse av fremtredende piksler fra resultatet av indeksen. Spektralsignaturene til disse pikslene ble evaluert opp mot spektralsignaturen til goethitt fremstilt i lab. Ut ifra denne analysen dannet verdiene til pikslene en øvre terskelverdi på 1 og en nedre terskelverdi på 0.3. Piksler med en verdi innenfor denne terskelen anses som piksler med innhold av goethitt. Med denne terskelverdien ble også ekstremverdier og støy fjernet. Dette er en pekepinn for videre analyse.



Figur 3.17: Figuren viser resultatene fra den selvgenererte goethitt-indeksen etter en satt terskelverdi. Pikslene som er visualisert i svart er piksler med verdier mellom en øvre terskelverdi på 1 og en nedre terskelverdi på 0.3. Piksler utenfor dette intervallet vises i hvitt.

3.6.1.2 Goethittindeks 2: utviklet fra topp og bunnpunkter i spektralsignaturen

Den andre indeksen ble utviklet ved å analysere spektralsignaturen for goethitt. Spektralsignaturen som ble analysert var den fra fjernmåling, og ikke fra lab, slik at indeksen var mest optimalisert for de virkelige forholdene i fra fjernmåling. Båndene som korresponderer med toppene og bunnen i signaturen ble identifisert, og deretter brukt til å utvikle en spesifikk indeks som utnytter disse karakteristiske egenskapene. Goethitt har en karakteristisk topp ved 740 nm og bunn ved 925 nm, slik som Figur 3.18 viser. Figur 3.18 viser også det karakteristiske intervallet mellom cirka 626 nm og 638 nm der signaturen er flat (Kokaly mfl., 2017).



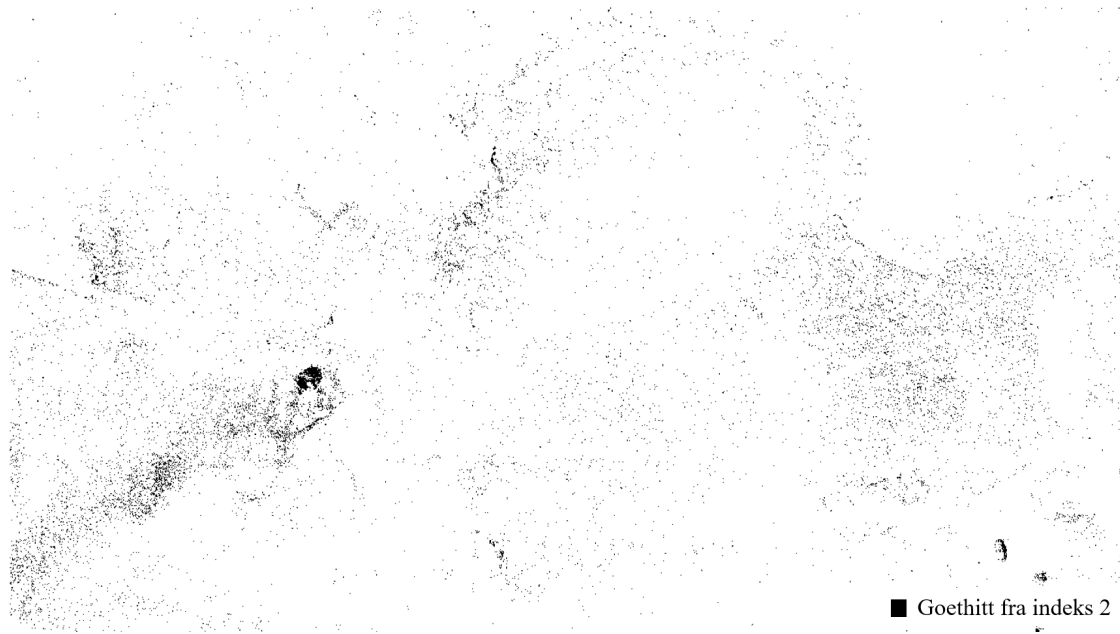
Figur 3.18: Figuren illustrerer karakteristiske topper og bunner for goethitt. 626 nm og 638 nm er der signaturen er flat. 740 nm og 925 nm viser der signaturen har topp- og bunnpunkt. Spektralsignaturen er hentet fra en eksempelpiksel fra HySpex-bildet.

Formel 3.3 viser indeksen. I første element av indeksen brukes bølgelengdene der signaturen er flat, som vil si 626 nm og 638 nm. I andre element av indeksen brukes topp- og bunnpunktet, som tilsvarer 740 nm og 925 nm.

$$\text{Goethittindeks 2} = \frac{626\text{nm} - 638\text{nm}}{626\text{nm} + 638\text{nm}} * \frac{740\text{nm} - 925\text{nm}}{740\text{nm} + 925\text{nm}} \quad (3.3)$$

I dette tilfellet ble HySpex-båndene 75, 79, 114 og 178 brukt, som tilsvarer de relevante bølgelengdene. Figur 3.19 viser HySpex-bildet etter at en terskelverdi settes for å tilpasse funn fra indeksen. Terskelverdien etableres gjennom en visuell analyse av fremtredende piksler fra indekseresultatet. Spektralsignaturene til disse pikslene sammenlignes med spektralsignaturen til goethitt fremstilt i lab. Basert på denne analysen dannet pikslene en øvre terskelverdi på -0.001 og en nedre terskelverdi på -0.5. Piksler med verdier innenfor denne terskelen ble vurdert å inneholde goethitt. Ved å bruke denne terskelverdien ble også ekstremverdier og støy fjernet, noe som gir en indikasjon for videre analyse.

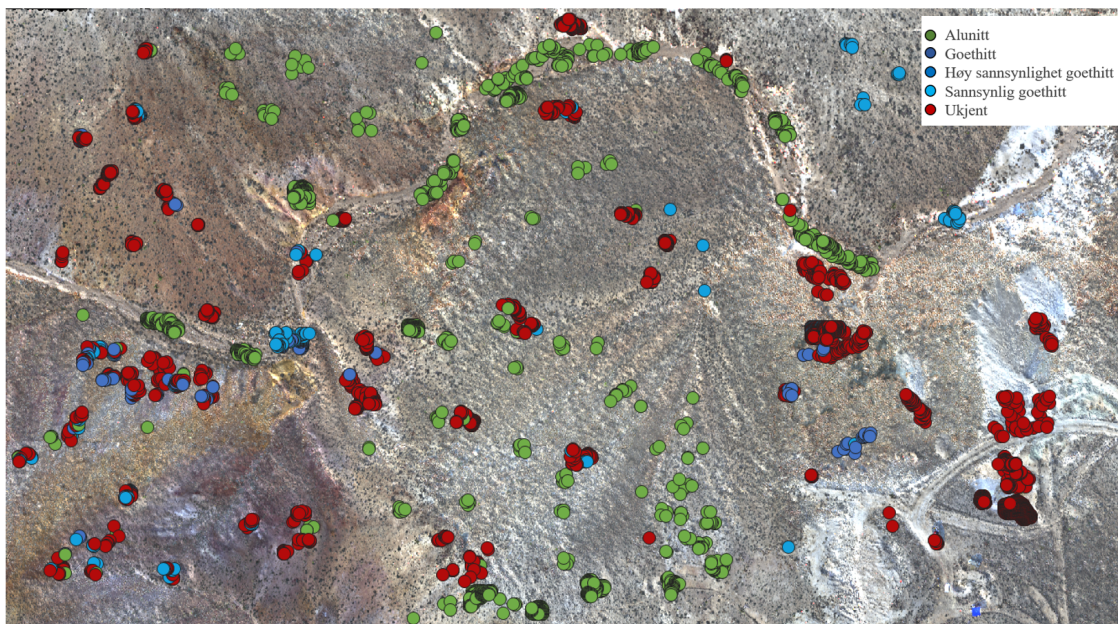
Resultatet fra indeksen er vist i Figur 3.19. Dette er indeksen etter en terskel er blitt lagt på. Goethittindeks nummer 2 er den indeksen som har de mest nøyaktige resultatene og er den som dannet grunnlaget for hvor treningsdataen for goethitt skulle plasseres.



Figur 3.19: Figuren viser resultatene fra den selvgenererte goethitt-indeksen fra spektralsignaturen, etter en satt terskelverdi. Pikslene som er visualisert i svart er piksler med verdier mellom en øvre terskelverdi på -0.001 og en nedre terskelverdi på -0.5 . Piksler utenfor dette intervallet vises i hvitt.

3.6.2 Nye treningspiksler for goethitt

For å kunne klassifisere et nytt mineral, ble det opprettet nye treningspiksler. Spektralsignaturen til goethitt er ikke like unik som alunitt sin, da den har færre signifikante topper og bunner. Treningspikslene for goethitt ble av den grunn delt inn i tre forskjellige klasser etter hvor god signaturen var. I Figur 3.20 vises pikslene for alunitt, ukjent og de tre goethittklassene. De mørkeblå punktene representerer de beste pikslene og klassen kalles «Goethitt». De blå punktene tilhører klassen «Høyst sannsynlig goethitt» og representerer de nest beste pikslene. De lyseblå punktene er klassen «Sannsynlig goethitt» og er de mest usikre pikslene.



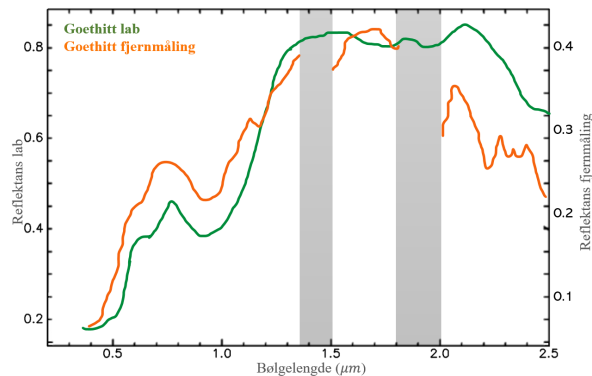
Figur 3.20: Figuren viser de nye treningspikslene som benyttes i klassifisering 2. Grønne punkter representerer «Alunitt», røde punkter indikerer «Ukjent», mørkeblå punkter indikerer «Goethitt» som de mest pålitelige deteksjonene, mens mellomblå punkter angir «Høyst sannsynlighet for goethitt» og lyseblå punkter viser til «Sannsynlig goethitt» som de mest usikre deteksjonene.

3.6.3 Visuell fordeling basert på spektralsignatur

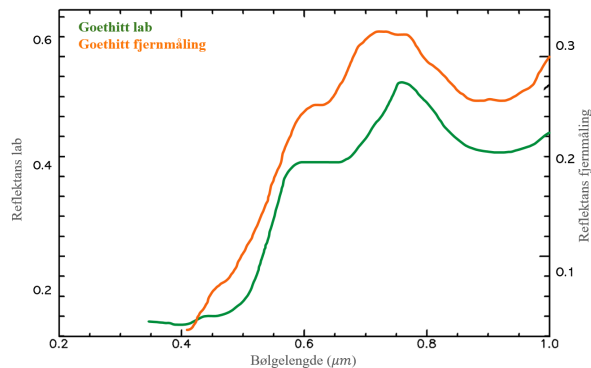
Figurene 3.21, 3.22, 3.23, 3.24, 3.25 og 3.26 viser spektralsignaturen til goethitt fra lab og fjernmåling med forskjellig nøyaktighet. Y-aksen til venstre gjelder reflektansen fra lab, mens y-aksen til høyre gjelder reflektansen fra fjernmåling. På grunn av variasjoner i lysforhold vil reflektansen være forskjellig mellom de to målingene, men figurene er satt til samme skala for å fokusere på mønstrene i signaturene, da disse er viktigere enn reflektansens intensitet.

Klasse 1: Goethitt

Spektralsignaturen til goethitt fra Kokaly mfl. (2017) er visualisert i grønt, mens spektralsignaturen til klassen «Goethitt» identifisert i analysen er fremstilt i oransje i. Den oransje linjen viser en utvalgt piksel fra de pikslene som er klassifisert i den mest nøyaktige klassen, ettersom disse pikslene følger spektralsignaturen fra lab best.



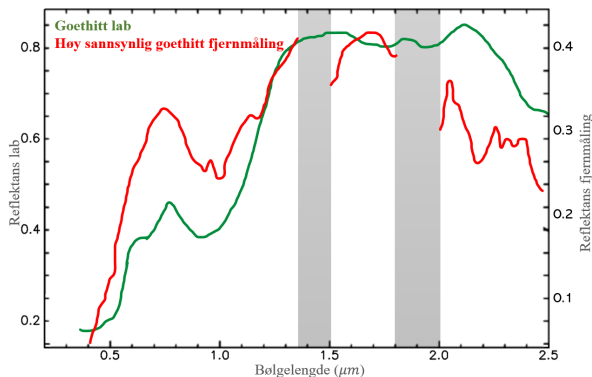
Figur 3.21: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og fra fjernmåling vist i oransje. Figuren viser hele spekteret fra 0.0 μm - 2.5 μm .



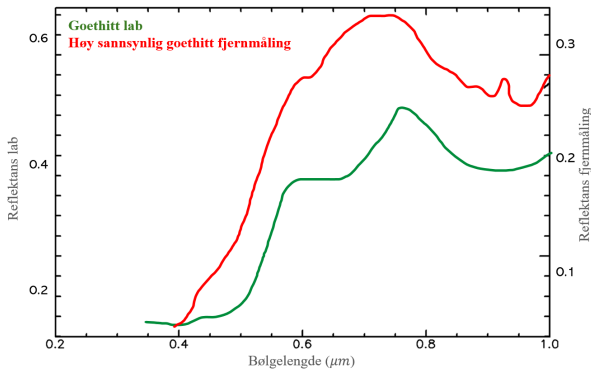
Figur 3.22: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og goethitt fra fjernmåling vist i oransje. Figuren viser kun deler av spekteret fra 0.2 μm - 1 μm .

Klasse 2: Høyst sannsynlig goethitt

Spektralsignaturen til goethitt fra Kokaly mfl. (2017) er visualisert i grønt, mens spektralsignaturen til klassen «Høyst sannsynlig goethitt» fra analysen er visualisert i rødt. Den røde linjen viser en utvalgt piksel fra de pikslene som er klassifisert i den nest mest nøyaktige klassen, ettersom disse pikslene avviker mer fra lab-signaturen enn den beste klassen.



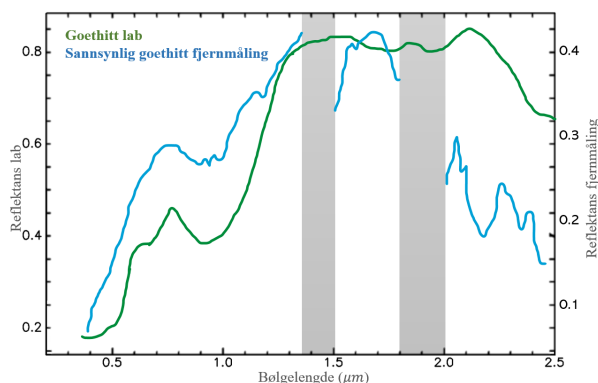
Figur 3.23: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og høyst sannsynlig goethitt fra fjernmåling vist i rød. Figuren viser hele spekteret fra 0.0 μm - 2.5 μm .



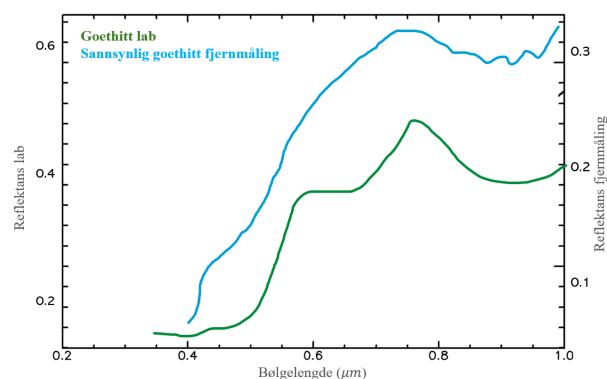
Figur 3.24: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og høyst sannsynlig goethitt fra fjernmåling vist i rød. Figuren viser kun deler av spekteret fra 0.2 μm - 1.0 μm .

Klasse 3: Sannsynlig goethitt

Spektralsignaturen til goethitt fra Kokaly mfl. (2017) er visualisert i grønt, mens spektralsignaturen til klassen «Sannsynlig goethitt» fra analysen er visualisert i blått. Den blå linjen viser en utvalgt piksel fra de pikslene som er klassifisert i den minst nøyaktige klassen, men likevel har trender som følger deler av spektralsignaturen fra lab.



Figur 3.25: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og sannsynlig goethitt fra fjernmåling vist i oransje. Figuren viser hele spekteret fra 0.0 μm - 2.5 μm .



Figur 3.26: Figuren viser mineralet goethitt fra lab (Kokaly mfl., 2017) vist i grønt og sannsynlig goethitt fra fjernmåling vist i oransje. Figuren viser kun deler av spekteret fra 0.2 μm - 1.0 μm .

Fordeling av nye treningspikslar

Tabell 3.3 viser fordeling av de nye treningspikslene for klassifisering 2.

Tabell 3.3: Tabellen viser fordelingen av pikslene i klassene alunitt, ukjent og de tre goethittklassene, som benyttes i klassifisering 2.

| Klasse | Antall |
|------------------------------|-------------|
| Goethitt | 45 |
| Høyst sannsynlighet goethitt | 192 |
| Sannsynlig goethitt | 362 |
| Alunitt | 1804 |
| Ukjent | 2878 |
| Totalt | 5281 |

3.6.4 Sammenslåing av goethitt-klasser

I Tabell 3.3 vises 599 goethitt-piksler fordelt over tre klasser. Det ble testet 3 ulike klassifiseringskombinasjoner. 1: klassifisering med hver klasse separat, 2: alle 3 klassene slått sammen og 3: fjerning av den dårligste klassen samt kombinerings av de to beste klassene. Det viste det seg at kombinasjon 3 gav de mest pålitelige resultatene. I videre klassifiseringer for deteksjon av goethitt og alunitt, ble derfor bare de to beste klassene brukt for å representere goethitt, noe som omfatter 237 piksler, slik Tabell 3.4 viser. Siden det ikke finnes en «fasit» for goethitt-klassen falt valget på å heller ta med for få enn for mange piksler. Dette valget ble gjort for å sikre presisjon ved å unngå inkludering av usikre data.

Tabell 3.4: Tabellen viser den endelige treningsdataen som ble benyttet i klassifisering 2. Klassene for alunitt og ukjent holder seg stabile, men klassen for goethitt er slått sammen av de to beste klassene.

| Klasse | Antall |
|------------------------|-------------|
| Slått sammen: Goethitt | 237 |
| Alunitt | 1804 |
| Ukjent | 2878 |
| Totalt | 4919 |

3.6.5 Gjennomføring av klassifisering 2: alunitt og goethitt

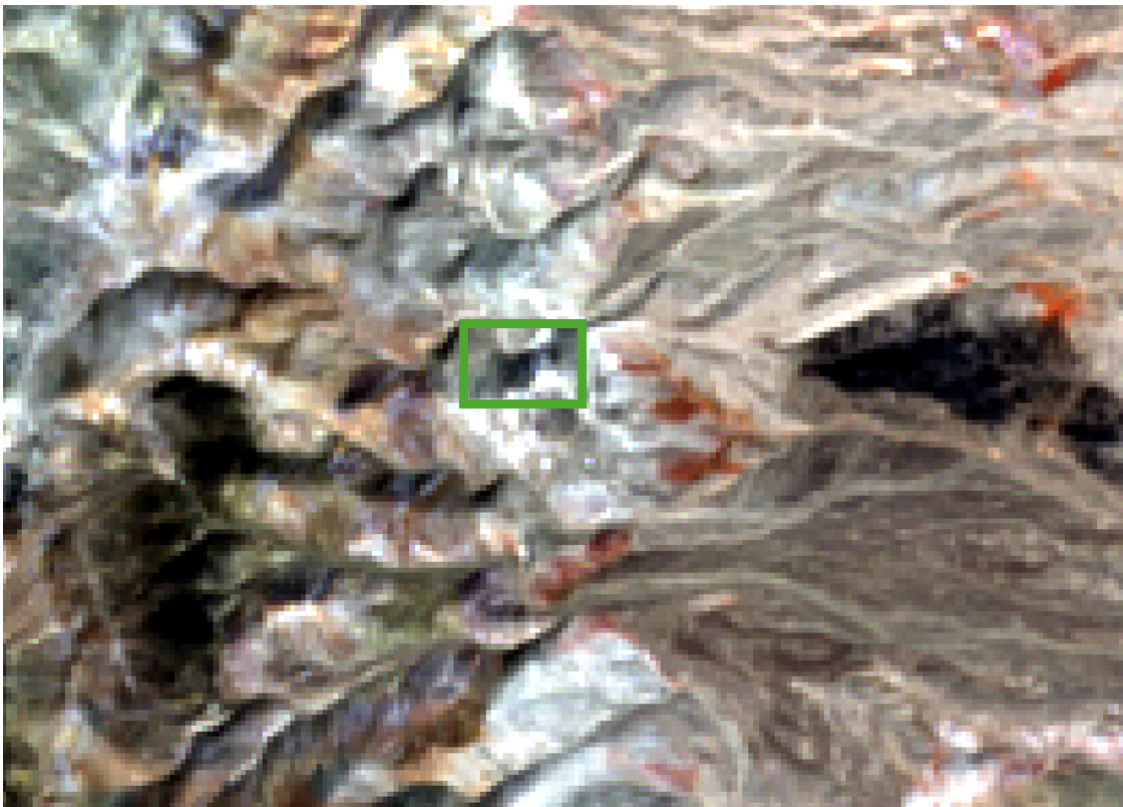
Klassifisering 2, av alunitt og goethitt ble også gjennomført i QGIS, slik det er beskrevet i Kapittel 3.2.3.1. Denne klassifiseringen baserte seg på de nye treningspikslene vist i Tabell 3.4, fordelt i testdata og treningsdata slik som vist i Figur 3.5. Klassifiseringen ble utført med data fra alle tre sensorene.

Tabell 3.5: Tabellen viser antall piksler brukt i klassifiseringen av alunitt og goethitt. Av totalt 4919 piksler, ble 3935 brukt for trening (80%) og 984 for testing (20%)

| | Treningsdata 80% | Testdata 20% |
|--------------|------------------|--------------|
| Goethitt | 185 | 50 |
| Alunitt | 1446 | 360 |
| Ukjent | 2304 | 574 |
| Total | 3935 | 984 |

3.7 Klassifisering 3: utvidelse av området

Det var ønskelig å teste modellens effektivitet i et større område. Imidlertid begrenser mangelen på tilgjengelige HySpex-data utenfor det opprinnelige studieområdet dette, da det ikke er tilgang til høyoppløselige sammenligningsdata. Til tross for denne begrensningen, er det planlagt å utføre en evaluering ved hjelp av data fra EMIT og Sentinel-2. De samme treningsdatasettene vil bli benyttet, men studieområdet vil utvides til omtrent 30 kvadratkilometer. Resultatene vil bli sammenlignet med tidligere forskning fra området (Swayze mfl., 2014). Figur 3.27 viser det utvidede studieområde med data fra Sentinel-2. Det grønne polygonet i midten viser det opprinnelige studieområdet.



Figur 3.27: Figuren viser bilde av det utvidede området, med en størrelse på omtrent 5 km x 6 km. Det grønne polygonet i midten viser det opprinnelige studieområdet.

4 Resultater

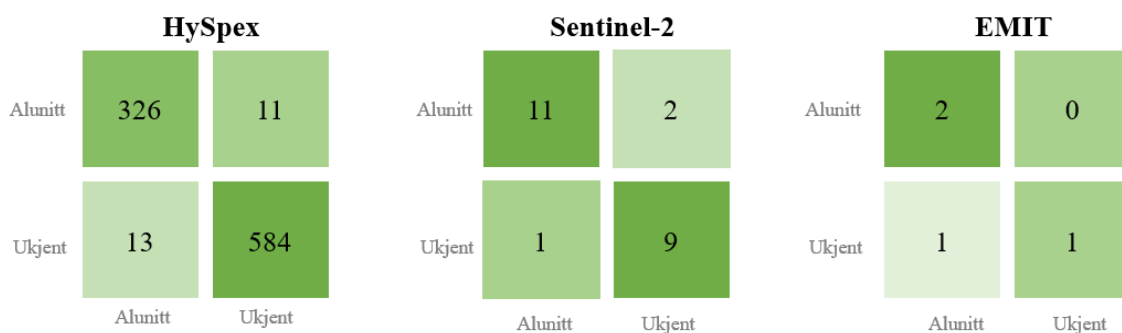
Dette kapitlet presenterer resultatene fra de utførte klassifiseringene og de tilhørende valideringsparameterene. Første klassifisering fokuserer på deteksjon av alunitt og tar i bruk data fra alle tre sensorer. Den andre klassifiseringen tar sikte på å identifisere både alunitt og goethitt, og bruker også alle tre sensorer. For den tredje klassifiseringen, som undersøker tilstedeværelsen av alunitt og goethitt over et utvidet område, anvendes data fra Sentinel-2 og EMIT, ettersom HySpex-dataene er begrenset til et mindre område.

Evalueringen av klassifiseringene er basert på forvirringsmatriser med tilhørende valideringsparametere for hver sensor. Disse verdiene er beregnet ut ifra en analyse der 20% av treningsdatasettet er benyttet som et testsett for å validere modellenes nøyaktighet.

4.1 Klassifisering 1: alunitt

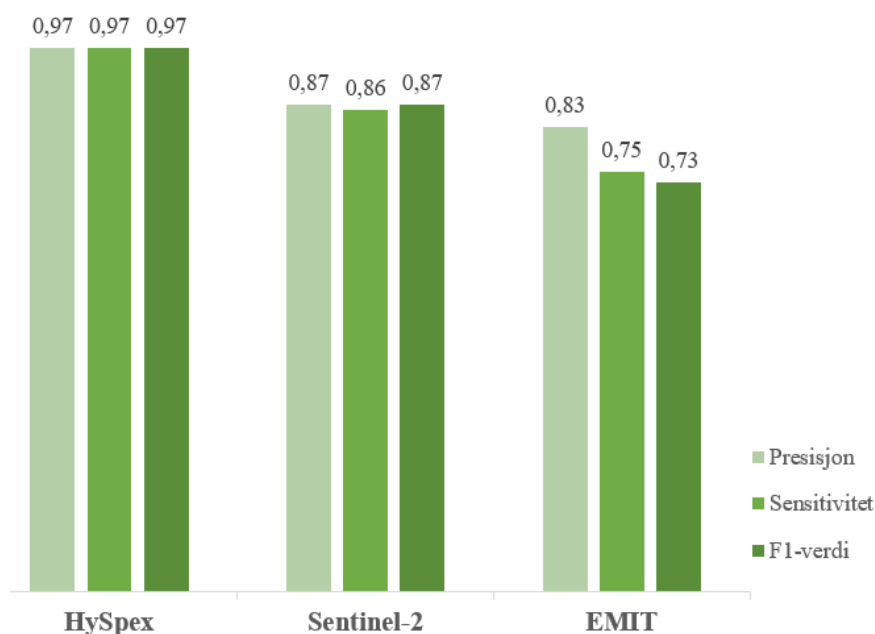
4.1.1 Valideringsparametere

Resultatene fra HySpex-sensoren indikerte høy evne til å identifisere alunitt, med 326 sanne positive og 584 sanne negative, og bare 11 falske positive og 13 falske negative. For Sentinel-2 viser resultatene også en sterk klassifiseringskapasitet med 11 sanne positive og 9 sanne negative, men med flere feil. Sentinel-2 klassifiserte 1 falsk negativ og 2 falske positive. EMIT-sensoren med kun 4 testpikslar viste 2 sanne positive, 1 sann negativ og 1 falsk negativ. Uavhengig av fordeling av testpikslar for EMIT er 4 pikslar for få til å gi en pålitelig klassifisering. Forvirringsmatrisene fra klassifisering 1 vises i Figur 4.1.



Figur 4.1: Figuren illustrerer forvirringsmatrisene for klassifisering 1 med alle tre sensorene. Matrisene viser resultatene av klassifiseringen ved å vise antall korrekte og feilklassifiserte observasjoner mellom de to klassene: alunitt og ukjent

Figur 4.2 viser valideringsparameterene fra klassifisering 1. HySpex-sensoren viser konsekvent høye verdier av presisjon, sensitivitet, og F1-verdi, der alle målinger er på 0.97. Dette tyder på en meget nøyaktig klassifisering av alunitt. Sentinel-2 viser litt lavere, men fortsatt sterke valideringsparametere med en sensitivitet på 0.86, og både presisjon og F1-verdi på 0.87. EMIT-sensoren viser en presisjon på 0.83 og sensitivitet på 0.75. Dette resulterer i en F1-verdi på 0.73, hvilket indikerer at selv om EMIT kan ha en god presisjon, er dens evne til å gjenkjenne alle alunitt-tilfeller lavere.



Figur 4.2: Valideringsparameterene for alunitt-klassifisering for alle tre sensorene. Presisjon vist i lys grønn, sensitivitet vist grønn og F1-verdi vist i mørk grønn.

4.1.2 Prosentvis fordeling

Tabell 4.1 presenterer fordelingen av piksler av de to forskjellige klassene som er identifisert av sensorene. HySpex har kategorisert 28,7 % av pikslene som alunitt. Sentinel-2 har en betydelig høyere andel alunitt-klassifiseringer på 49,5 %, mens EMIT fant at 42,9 % av pikslene representerer alunitt. De øvrige pikslene fra hver sensor er klassifisert som «Ukjent». Denne store forskjellen kan skyldes at sensorene med færre piksler kan gi større prosentvis variasjon og mindre nøyaktighet, da hver piksel utgjør flere prosent.

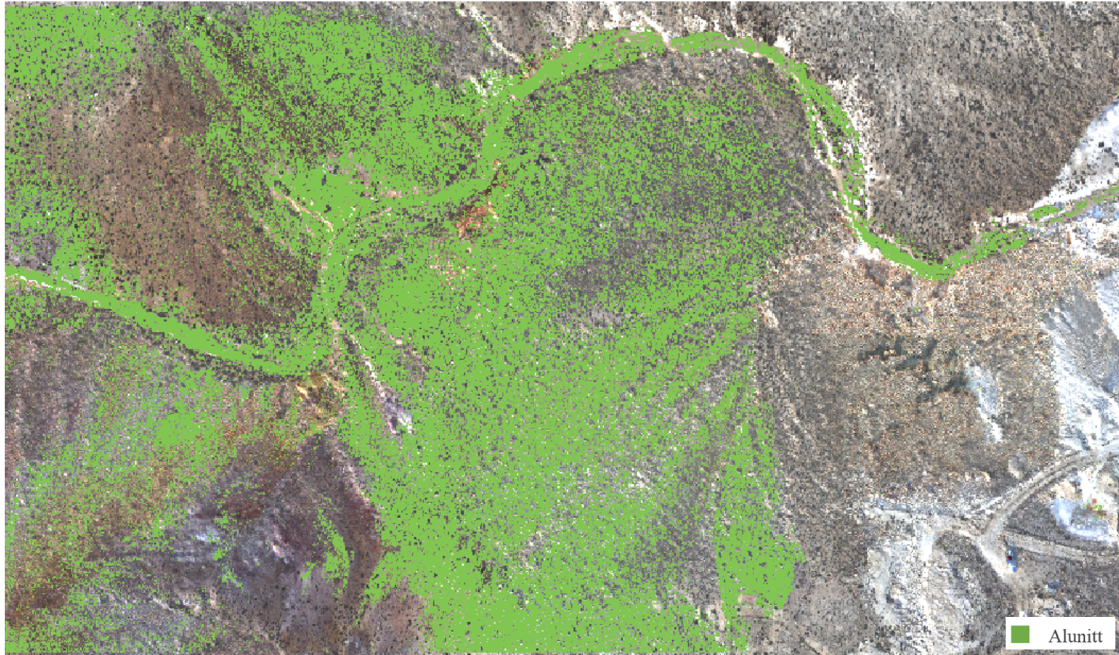
Tabell 4.1: Tabellen viser antall piksler i hver klasse fra alle 3 sensorene, og til slutt den prosentvise fordelingen av de ulike klassene for klassifisering 1.

| Sensor | Alunitt | Ukjent | Total | Prosent alunitt |
|------------|---------|---------|---------|-----------------|
| HySpex | 502616 | 1247384 | 1750000 | 28.7 % |
| Sentinel-2 | 301 | 307 | 608 | 49.5 % |
| EMIT | 12 | 16 | 28 | 42.9 % |

4.1.3 Visuelle resultater

4.1.3.1 HySpex

I Figur 4.3 ser man den totale klassiferingen av mineralet alunitt med HySpex-bildet som grunnlag. Resultatet samsvarer godt med alunitt-indeksen. Det nye laget som blir klassifisert har god dekning på de samme områdene, og unnviker å klassifisere piksler der det er kjent fra indeksen og spektralsignaturen at det ikke er forekomst.



Figur 4.3: Bildet viser klassifisering av alunitt med HySpex. Funn av alunitt er illustrert med fargen grønn.

4.1.3.2 Sentinel-2

I Figur 4.4 vises den totale klassifiseringen av mineralet alunitt med Sentinel-2-bildet som grunnlag. Som bakgrunn er den klassifiserte alunitten basert på HySpex-data visualisert. Dette gir en referanse for å vurdere nøyaktigheten av klassifiseringen med Sentinel-2-data. Det er viktig å merke seg at Sentinel-2 har lavere oppløsning og færre bånd sammenlignet med HySpex, noe som fører til betydelig større unøyaktighet i resultatene.

Den endelige klassifiseringen viser en tendens til overklassifisering, der flere piksler enn nødvendig klassifiseres som alunitt. Imidlertid oppnår klassifiseringen treffsikkerhet på visse områder i bildet; der det er bekreftet tilstedeværelse av alunitt fra HySpex-dataene, blir det også klassifisert som alunitt ved hjelp av Sentinel-data. På samme måte observeres områder der det ikke klassifiseres som alunitt, noe som stemmer godt overens med grunnlaget.

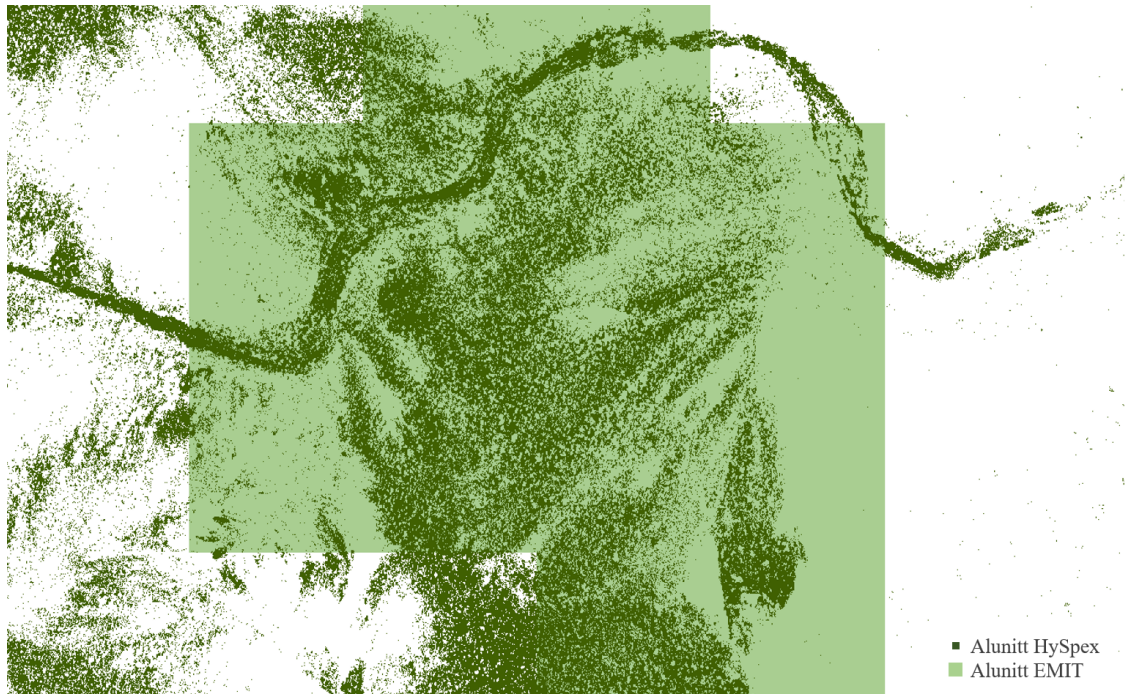


Figur 4.4: Bildet viser klassifisering av alunitt med Sentinel-2. Funn av alunitt er illustrert med fargen grønn. Klassifiseringen med HySpex ligger i bunn for sammenligning, og er illustrert med fargen mørk grønn.

4.1.3.3 EMIT

I Figur 4.5 vises den totale klassifiseringen av mineralet alunitt basert på EMIT-bildet. De mørke pikslene som utgjør bakgrunnen i Figur 4.5 er alunitt-klassifiseringen gjort med HySpex. Dette utgjør en referanse for hvor klassene alunitt og ukjent skal befinne seg. Selv med lav oppløsning, klarer EMITs 285 bånd å skille ulike mineraler basert på deres spektrale egenskaper, spesielt i de høyere båndene. Dette er sentralt for deteksjonen av alunitt, som har sine tydelige kjennetegn fra 2.0 - 2.5 μm .

Til tross for bruk av grovkornede piksler, viser EMIT-klassifiseringen i Figur 4.5 en tydelig tendens til å identifisere «alunitt» i områder med høyere konsentrasjon og identifiserer «ukjent» i områder med lavere forekomst av mineralet.

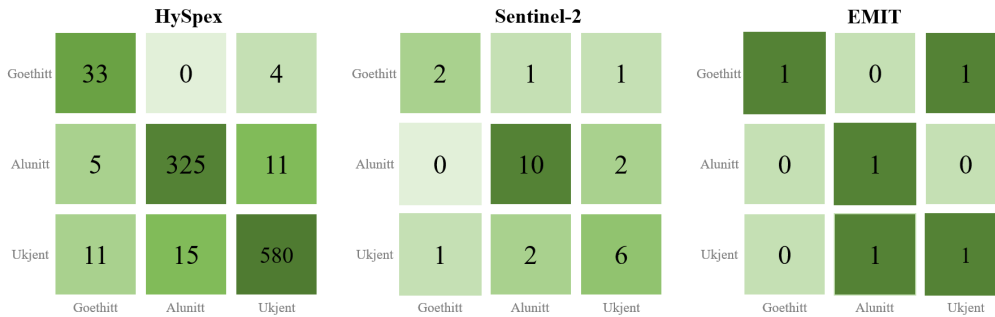


Figur 4.5: Bildet viser klassifisering av alunitt med EMIT. Funn av alunitt er illustrert med fargen grønn. Klassifiseringen med HySpex ligger i bunn for sammenligning, og er illustrert med fargen mørk grønn.

4.2 Klassifisering 2: alunitt og goethitt

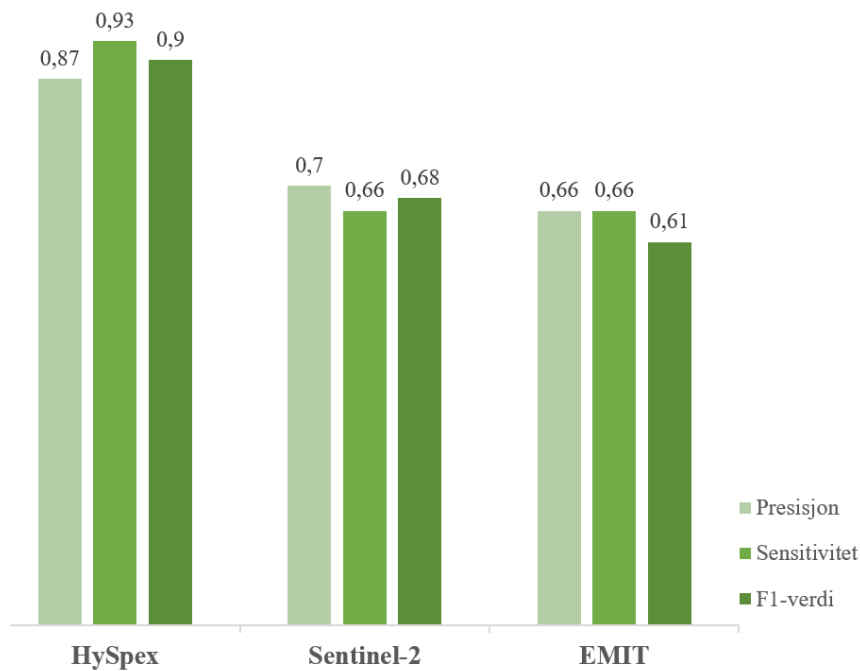
4.2.1 Valideringsparametere

HySpex-sensoren viste høy nøyaktighet i å identifisere alunitt og goethitt, med 938 av 984 piksler klassifisert riktig, noe som betyr at bare 46 piksler ble feilklassifisert. Derimot leverte Sentinel-2 svakere resultater, med 91 av 142 piksler riktig klassifisert og 51 piksler feil. De aller fleste av feilene kom fra de 22 pikslene som feilaktig ble identifisert som alunitt, selv om de var i klassen for ukjent. EMIT-sensoren hadde bare fem piksler å klassifisere, hvorav tre var riktige og to feil, noe som indikerer en lav nøyaktighet. Igjen fører de få testpunktene for EMIT til at klassifiseringen ikke er pålitelig. Forvirringsmatrisene til klassifisering 2 vises i Figur 4.6.



Figur 4.6: Figuren illustrerer forvirringsmatrisene for klassifisering 2 med alle tre sensorene. Matrisene demonstrerer resultatene av klassifiseringen ved å vise antall korrekte og feilklassifiserte observasjoner mellom de tre klassene: alunitt, goethitt og ukjent

Figur 4.7 viser valideringsparameterene for klassifisering 2. Det er heller ikke her overraskende at HySpex-sensoren fremviser konsistent høye verdier, med presisjon, sensitivitet, og F1-verdier som alle er 0,87 eller høyere. Dette tyder på en nøyaktig klassifisering. Sentinel-2 viser en god del lavere valideringsverdier, med sensitivitet og F1-verdi fra 0,66 til 0,68. Selv om presisjonen har en verdi på 0,7, er valideringsparameterene for Sentinel-2 betydelig lavere enn for HySpex. EMIT-sensoren viser både presisjon og sensitivitet på 0,66, og F1-verdi på 0,61. Dette indikerer at det er større rom for forbedring i dens evne til å gjenkjenne alle tilfeller av alunitt og goethitt.



Figur 4.7: Valideringsparametere for klassifisering 2 for alle tre sensorene. Presisjon vist i lys grønn, sensitivitet vist grønn og F1-verdi vist i mørk grønn.

4.2.2 Prosentvis fordeling

Tabell 4.2 viser antall piksler som er klassifisert i hvilke klasser, med de tre ulike sensorene. HySpex-sensoren har identifisert 3,3 % av pikslene som goethitt og 28,8 % som alunitt. For Sentinel-2 er prosentene henholdsvis 1,8 % for goethitt og 62,3 % for alunitt, mens EMIT viser 14,2 % for goethitt og 42,9 % for alunitt. De resterende pikslene er klassifisert som «ukjent». Prosentandelen for alunitt i denne klassifiseringen er nær identisk med prosentandel for alunitt i den første klassifiseringen. Også her kan den store forskjellen skyldes at sensorer med færre piksler kan føre til større variasjoner i prosent, fordi hver piksel har større påvirkning på resultatet.

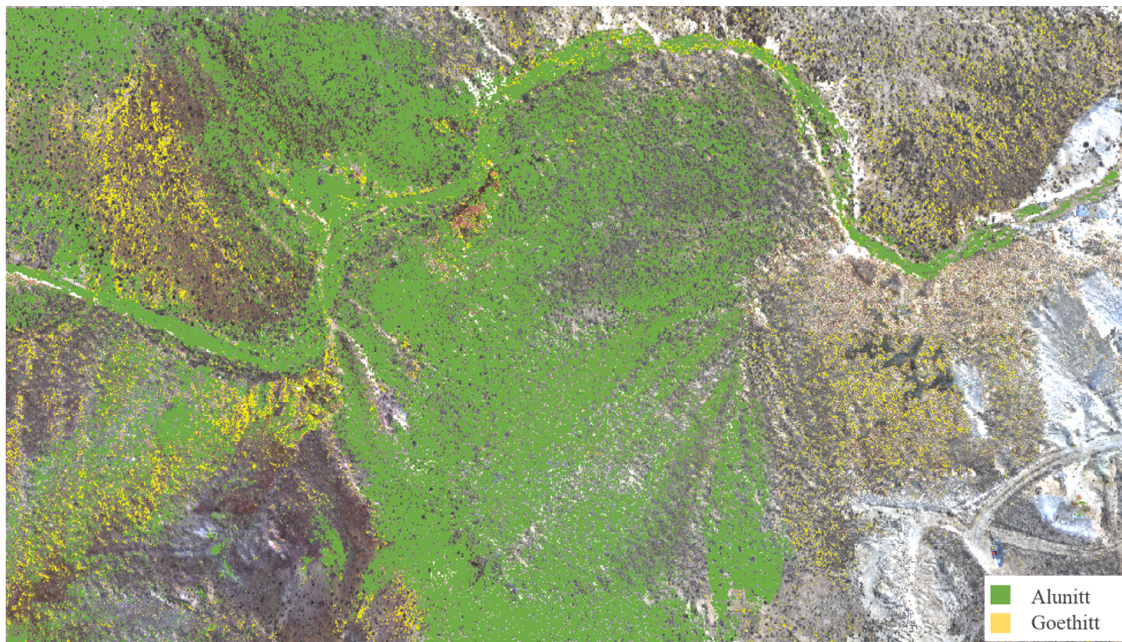
Tabell 4.2: Tabellen viser antall piksler i hver klasse fra alle 3 sensorene, og til slutt den prosentvise fordelingen av de ulike klassene for klassifisering 2.

| Sensor | Goethitt | Alunitt | Ukjent | Total | Prosent goethitt | Prosent alunitt |
|------------|----------|---------|---------|---------|------------------|-----------------|
| HySpex | 57136 | 504036 | 1188828 | 1750000 | 3.3 % | 28.8 % |
| Sentinel-2 | 11 | 379 | 218 | 608 | 1.8 % | 62.3 % |
| EMIT | 4 | 12 | 12 | 28 | 14.2 % | 42.9 % |

4.2.3 Visuelle resultater

4.2.3.1 HySpex

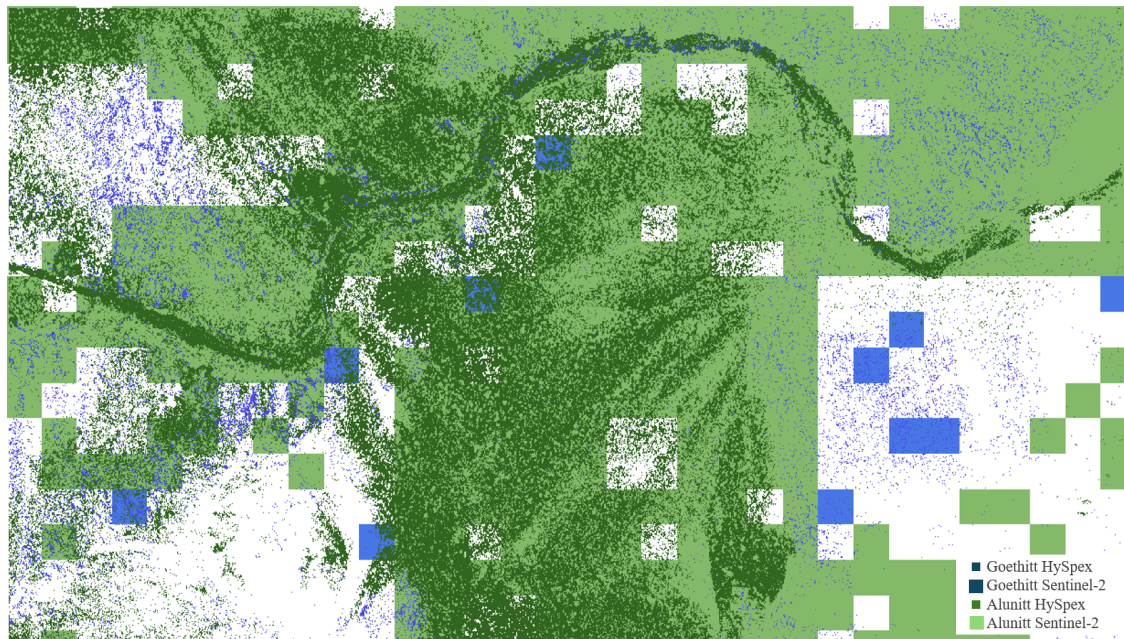
Figur 4.8 viser HySpex-sensorens klassifisering av alunitt og goethitt, omtalt i Kapittel 3.6. Disse resultatene bekrefter tidligere funn for alunitt, og viser at goethitt hovedsakelig forekommer i mindre områder, i tråd med forventningene.



Figur 4.8: Bildet viser klassifisering av alunitt og goethitt med HySpex. Funn av alunitt er illustrert med fargen grønn, og funn av goethitt er illustrert med fargen gul.

4.2.3.2 Sentinel-2

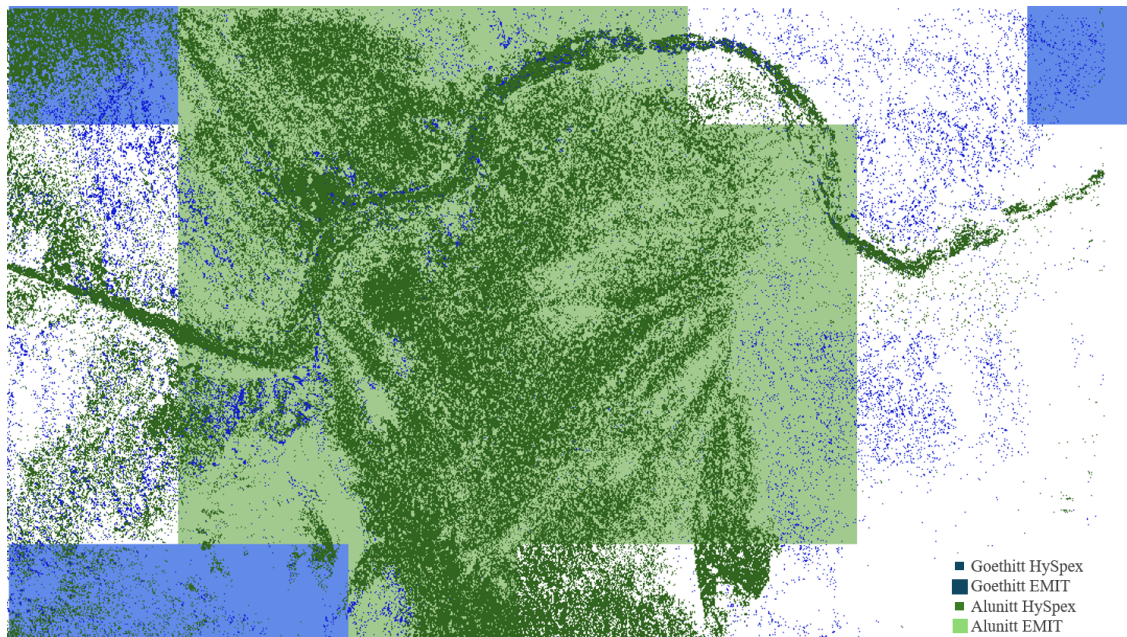
Klassifisering med Sentinel-2 av alunitt og goethitt er illustrert i Figur 4.9. HySpex-klassifisering brukes som bakgrunn for å enklere kunne vurdere nøyaktigheten til klassifiseringen. På grunn av Sentinel-2s grovere oppløsning og færre spektralbånd, er klassifiseringen vesentlig dårligere enn for HySpex. En tendens av overklassifisering av alunitt og underklassifisering av goethitt kan observeres.



Figur 4.9: Bildet viser klassifisering av alunitt og goethitt med Sentinel-2. Funn av alunitt er illustrert med fargen grønn, og funn av alunitt er visualisert med fargen blå. Klassifiseringen med HySpex ligger i bunn for sammenligning, og er illustrert med fargen mørk grønn, og mørk blå.

4.2.3.3 EMIT

Klassifisering av alunitt og goethitt med EMIT presenteres i Figur 4.10, og visualiseres også med HySpex-klassifisering som bakgrunns referanse. EMIT utnytter sin høye spektraloppløsning til tross for lavere pikseloppløsning. Dette gjør det mulig å identifisere store områder med alunitt nøyaktig. Klassifisering av goethitt viser seg å være mindre vellykket.



Figur 4.10: Bildet viser klassifisering av alunitt og goethitt med EMIT. Funn av alunitt er illustrert med fargen grønn, og funn av alunitt er visualisert med fargen blå. Klassifiseringen med HySpex ligger i bunn for sammenligning, og er illustrert med fargen mørk grønn, og mørk blå.

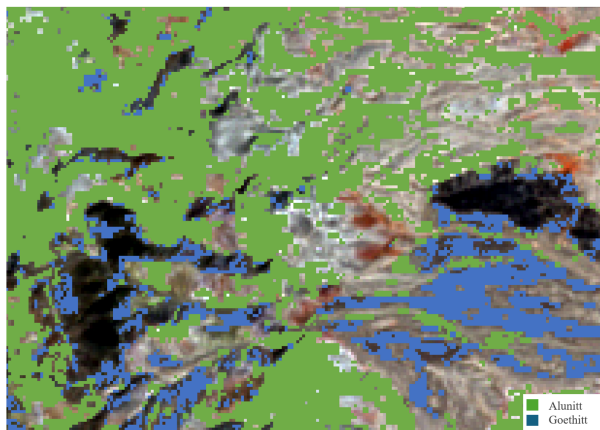
4.3 Klassifisering 3: utvidet område

For å teste modellen ytterligere er det blitt gjennomført klassifisering med Sentinel-2 og EMIT på et utvidet område. Siden det ikke finnes tilgjengelige HySpex-data for dette området, mangler det en direkte referanse for sammenligning. Valideringsparameterene vil være identiske som for det første studieområdet, og er derfor ikke inkludert her. Fokuset vil ligge på de visuelle resultatene.

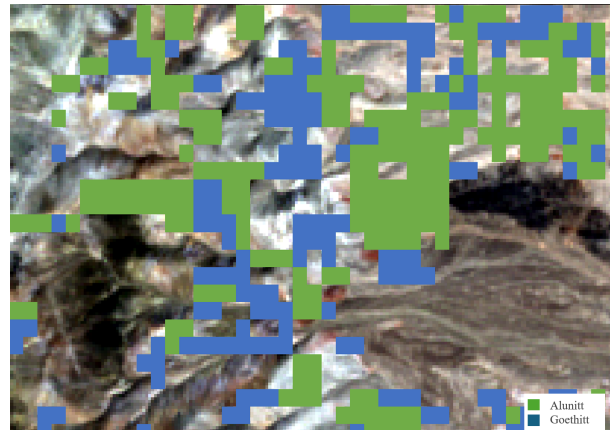
4.3.1 Visuelle resultater

4.3.1.1 Sentinel-2 og EMIT

Figur 4.11 viser klassifiseringen av alunitt og goethitt med Sentinel-2 over et større område. Figur 4.12 viser det samme, men med EMIT. En diskusjon av resultatet og sammenligning med tidligere forskning på dette temaet vil bli diskutert i Kapittel 5.3.4.



Figur 4.11: Figuren viser resultater fra klassifisering 3 av et utvidet område med Sentinel-2. Alunitt er illustrert i fargen grønn og goethitt i fargen blå.



Figur 4.12: Figuren viser resultater fra klassifisering 3 av et utvidet område med EMIT. Alunitt er illustrert i fargen grønn og goethitt i fargen blå.

5 Diskusjon

Diskusjonskapitlet går nøye gjennom datagrunnlaget og metodene som brukes for å nå resultatene. Det ser også på svakheter i dataen og metoden, og vurderer hvordan disse kan påvirke påliteligheten av resultatene. Kapitlet er delt inn i tre deler: Del 1 ser på dataene, del 2 tar for seg metoden brukt i analysen, og del 3 vurderer de oppnådde resultatene.

5.1 Diskusjon av data

5.1.1 Data fra EMIT

Bildene samlet inn med EMIT-sensoren ble tatt tre år etter bildene fra de to andre sensorene, noe som kan ha påvirket datasammenligningen. I løpet av disse årene kan landskapet ha forandret seg på grunn av naturlige prosesser eller menneskelige inngrep, og disse endringene kan ha en effekt på studieområdet. Avvik mellom datasettene kan svekke den samlede nøyaktigheten av klassifiseringen.

5.1.2 Mineraldeteksjon på et begrenset studieområde

Studieområdet som er benyttet i analysen er lite, likevel er det for HySpex tilstrekkelig stort til å oppnå en nøyaktig klassifisering. For Sentinel-2 og EMIT derimot, påvirker områdets størrelse klassifiseringens nøyaktighet betydelig. Den store pikselstørrelsen fører til et begrenset antall test- og treningspiksler, noe som svekker påliteligheten av analysen. Dette er spesielt gjeldende for EMIT, hvor det kun finnes fem testpunkter.

5.2 Diskusjon av metode

5.2.1 Valg av maskinlæringsalgoritme

For analysen ble Random Forest-algoritmen valgt som klassifiseringsmetode. Denne algoritmen ble foretrukket fordi den effektivt håndterer store og komplekse datasett. Dette var avgjørende da det originale HySpex-bildet på 5 GB var for stort til å kunne gjennomføre klassifisering. En reduksjon av datamengden til 1.2 GB gjorde klassifisering med Random Forest gjennomførbar (Machine Learning Models, 2024).

Support Vector Machine (SVM) ble vurdert som mindre egnet fordi denne algoritmen krever betydelige ressurser og tid for å konstruere hyperplan når datamengden er stor. K-Nearest Neighbour

(KNN) ble ikke valgt på grunn av lavere nøyaktighet og effektivitet i store datasett, i tillegg til sårbarhet for støy. Gaussian Mixture Model (GMM) er ofte mindre egnet for komplekse datasett på grunn av antakelsen om en Gaussisk fordeling. Random Forest ble derfor vurdert som den beste løsningen, gitt det begrensede treningsdatagrunnlaget og behovet for effektiv klassifisering (Sarker, 2021, s. 6-7).

5.2.1.1 Svakheter med Dzetsaka klassifiseringsverktøy

Klassifiseringsverktøyet som brukes i denne oppgaven er en QGIS-plugin, med enkelte begrensninger. Det er ikke mulig å justere hyperparameterene, noe som ville vært mulig om klassifiseringen ble utført i for eksempel Python. Fri endring av hyperparametere ville ha forberedt modellens evne til å tilpasse seg og håndtere ny, usett data bedre (Probst mfl., 2024, s. 16). Selv om dokumentasjonen antyder at det er mulig å gjøre endringer, krever dette tiltak utenfor QGIS-plattformen. En nyttig forbedring ville vært å legge til muligheten for direkte justering av hyperparametere direkte i QGIS-grensesnittet.

5.2.2 Test og treningsdata

Valg av trenings- og testdata er avgjørende for modellens nøyaktighet. I denne studien ble dataene fordelt med 80 % til trening og 20 % til testing, en vanlig praksis som balanserer behovet for tilstrekkelig treningsdata med evnen til å validere modellen. For EMIT kunne det ha vært fordelaktig å bruke en fordeling med for eksempel 70% til trening og 30% til test. Da hadde det vært flere testpunkter, som hadde gitt en bedre og mer robust evaluering.

Testpikslene ble valgt ut helt tilfeldig fra datasettet. Dette er gjort for å unngå at testdataene blir skjevt fordelt geografisk, noe som kunne påvirket resultatene. Et tilfeldig utvalg av testpiksler sikrer at testdataene er representative for hele datasettet, noe som gir en pålitelig evaluering av modellen. Testdataene ble imidlertid hentet fra samme område som treningsdataene, fordi det kun var ett bilde tilgjengelig fra HySpex. Dermed var det ikke mulig å teste modellen på helt usett data, noe som kan føre til overtilpasning og mindre nøyaktige resultater når modellen brukes på nye data (Karasiak mfl., 2021).

5.2.2.1 Treningspiksler basert på visuell analyse

Treningspikslene ble generert gjennom visuell analyse av spektralsignaturen til hvert enkelt piksel. Denne tilnærmingen er sårbar for feilklassifisering, da menneskelige faktorer som presisjon og fokus

kan påvirke nøyaktigheten til treningspikslene. Dette ble spesielt tydelig i tilfellet med goethitt, hvor treningsdata ble generert uten støtte fra en dokumentert indeks eller andre analyser. I tillegg er spektralsignaturen til goethitt mindre unik enn for alunitt, noe som gjør den vanskeligere å identifisere.

Et tiltak som ble gjort for å håndtere denne utfordringen, var å kategorisere goethitt i tre forskjellige klasser basert på deres kvalitet. Pikslene ble rangert etter hvor «gode» de var, og den klassen som ble ansett som den svakeste, ble fjernet fra den endelige klassifiseringen. Denne tilnærmingen ble valgt for å øke presisjonen på treningspikslene og dermed forbedre nøyaktigheten til den endelige klassifiseringen. Ved å eliminere de svakere treningspikslene ble risikoen for feilklassifisering redusert, noe som bidro til mer pålitelige resultater.

5.2.2.2 Sammenligning av spektralsignatur fra lab og fjernmåling

Treningsdataen i analysen er basert på spektralsignaturen fra fjernmåling sammenlignet med spektralsignaturen til mineralene fra lab. Dette fører til usikkerheter. Flere faktorer påvirker gjennomsnittsreflektansen i en fjernmålt piksel sammenlignet med en laboratorieprøve. For eksempel kan lys og skygge samt tekstureffekter endre hvordan en piksel ser ut fra luften. For eksempel vil en haug med hvit grus/stein sett ovenfra virke grå, på grunn av skygge og tekstur. I lab kan man undersøke «rene piksler», hvor alt materialet i pikselen er det samme. I fjernmåling derimot vil en piksel alltid inneholde en blanding av forskjellige materialer. Derfor vil spektralsignaturen fra fjernmåling alltid være mindre nøyaktig enn fra lab, selv med god spektral og geometrisk oppløsning.

5.2.2.3 Ubalansert data

I klassifiseringene som ble gjennomført i denne analysen var fordeling av treningspiksler mellom de tre klassene ujevnt fordelt. Dette kan ha en påvirkning på modellens evne til å klassifisere korrekt, spesielt for den minste klassen, «goethitt». Modellen kan dermed favorisere de overrepresenterte klassene «alunitt» og «ukjent», noe som resulterer i lavere presisjon for den underrepresenterte klassen (Chawla, 2003).

For å håndtere denne ubalansen, kunne en justering av hyperparameterene vært en løsning. Dette er tidligere nevnt i Kapittel 5.2.1.1, som en av svakhetene ved å bruke Dzetsaka-pluginet, da det ikke er mulighet for å «vekte» klassene. Ved dette menes å tilegne underrepresenterte klasser en høyere vekt, slik at modellen legger mer vekt på nøyaktigheten av klassifiseringen av disse klassene (More og Rana, 2017).

Ved å bruke makrogjennomsnitt i beregningen av valideringsparameterene, som beskrevet i Kapittel 2.3.3.4, tas det hensyn til den minste klassen under valideringen. Dette diskuteres videre i Kapittel 5.3.1.1.

5.2.2.4 Test- og treningsdata for Sentinel-2 og EMIT

Generering av treningspiksler for Sentinel-2 ble utført på en annen og mer unøyaktig måte enn for HySpex og EMIT, og er forklart i Kapittel 3.5.3. Den lave spektrale oppløsningen til Sentinel-2 gir ikke den visuelle analysen en god nok evne til å beslutte hvilken klasse hver piksel skal klassifiseres som. Derfor ble treningspikslene fra HySpex brukt som grunnlag for nye treningspikslene for Sentinel-2. I de aller fleste tilfeller var punktene innenfor en piksel homogent, noe som betyr at alle treningspikslene fra HySpex var i samme klasse innenfor et Sentinel-2 piksel. Spektralsignaturen til Sentinel-2 ble benyttet som verifisering, de få gangene det var en blanding av treningspunkter fra klassene inne i pikselen. Siden spektralsignaturen til Sentinel-2 er grov, påvirker det punktets nøyaktighet, som igjen påvirker klassifiseringen negativt.

Treningspikslene for EMIT ble generert på samme måte som for HySpex. På grunn av EMITs oppløsning på 60 meter, inneholder hver piksel flere ulike mineraler. Alle piksler for Sentinel-2 og EMIT, vil i mye større grad en for HySpex, være en blanding av ulike mineraler i ulike konsentrasjoner i tillegg til de mineralene som ønskes å detektere. Dette fører til en svak og unøyaktig spektralsignatur. Når spektralsignaturen er uklar, blir det dårligere test- og treningspunkter, noe som gjør klassifiseringen mindre pålitelig.

5.2.3 Begrensninger med goethittindeksen

I analysen ble det utviklet to nye goethittindekser, da en eksisterende indeks ikke ble funnet via litteratursøk. Utgangspunktet for indeks 1 var den kjemiske formelen for goethitt, $\text{Fe}^{3+}\text{O}(\text{OH})$. Eksisterende indekser for både jernionet og hydroksylgruppen ble kombinert for å danne den nye indeksen. Den andre indeksen ble laget ved å analysere toppene og bunnene i goethitts

spektralsignatur, for så å utvikle en formel som kunne fange opp akkurat disse.

Selv om det å lage egne indekser primært var et eksperiment, viste det seg at en betydelig del av pikslene klassifisert som goethitt ved hjelp av disse indeksene stemte overens med spektralsignaturen til goethitt. Dette indikerer at selv om indeksene fortsatt er i en startfase, har de potensiale for videre bruk og utvikling.

5.2.4 Fasit til verifisering

For mineralet «alunitt» innenfor studieområdet finnes det en nyere rapport fra HySpex (Bahr, 2022a, s. 90) som kan brukes som sammenligningsgrunnlag. Imidlertid mangler det tilsvarende referanser for mineralet «goethitt» og det utvidede studieområdet. Å ha en fasit for sammenligning er en viktig del av verifiseringsprosessen, men denne muligheten finnes ikke her, noe som gjør det vanskelig å vurdere klassifiseringens nøyaktighet.

Tidligere publikasjoner fra sent på 1990-tallet og tidlig på 2000-tallet har dokumentert forekomster av både alunitt og goethitt i studieområdet (Swayze mfl., 2014). Selv om disse tidligere funnene samsvarer med mineralfunnene i denne analysen, varierer resultatene mellom publikasjonene, og det kan være utfordrende å fastslå hvor nøyaktige de er.

Siden mye kan ha endret seg siden disse eldre publikasjonene, kan resultatene fra den visuelle analysen samsvare uten nødvendigvis å gjenspeile virkeligheten nøyaktig. Det er derfor viktig å ta potensielle endringer i betraktning når man vurderer resultatene.

5.3 Diskusjon av resultatene

5.3.1 Valideringsparametere

Resultatene i Kapittel 4 viser at klassifiseringen med data fra HySpex har oppnådd de helt klart høyeste verdiene på valideringsparameterene. F1-verdien er 0.97 i klassifisering 1 og 0.90 i klassifisering 2. Sentinel-2 presterer generelt bedre enn EMIT. I klassifisering 1 har Sentinel-2 en F1-verdi på 0.87, mens EMIT har en F1-verdi på 0.73. Klassifisering 2 var mindre vellykket, men forholdet mellom sensorene holder seg relativt stabilt. Sentinel-2 har en F1-verdi på 0.68, mens EMIT har en F1-verdi på 0.61.

Selv om Sentinel-2, med sin oppløsning på 10 meter, generelt sett yter bedre, viser EMIT seg også å være konkurransedyktig til tross for sin betydelig dårligere oppløsning på 60 meter. Fordelen

med EMIT ligger i dens overlegne spektrale oppløsning, noe som understreker hvor viktig spektrale egenskaper er for å oppnå gode resultater. Denne evnen til å fange opp detaljerte spektralsignaturer gjør at EMIT fortsatt hevder seg godt, selv om Sentinel-2 generelt sett leverer bedre. Denne observasjonen fremhever betydningen av spektral oppløsning i mineraldeteksjon.

Studieområdet viser seg å være mindre egnet for EMIT, hovedsakelig på grunn av dets begrensede størrelse som resulterer i et lite antall tilgjengelige piksler for trening og testing av modellen. Med kun fem piksler til validering, blir hver enkelt piksel avgjørende for modellens ytelse. En enkelt piksel kan dramatisk endre F1-verdien, fra å indikere en dårlig til en god modell. På grunn av det svært begrensede datagrunnlaget kan man derfor ikke stole på valideringsparameterene for å vurdere hvor god modellen er. Dette gjør det utfordrende å trekke pålitelige konklusjoner med EMIT som datagrunnlag.

5.3.1.1 Utregning av valideringsparametere med makrogjennomsnitt

Makrogjennomsnitt ble valgt for å beregne valideringsparametere på grunn av betydelige størrelsesforskjeller mellom klassene. Dette sikrer at alle klasser blir behandlet likt, noe som er viktig når størrelsen på klassene varierer kraftig. Makrogjennomsnittet bidrar til et mer balansert bilde av algoritmens ytelse ved å beregne et gjennomsnitt av prestasjonene for hver klasse (Sefidian, 2022).

Selv om makrogjennomsnitt kan synes å være en ideell løsning, er det viktig å være klar over at det kan føre til en skjev oppfatning av algoritmens reelle effektivitet. Det kan skjule svake resultater i mindre klasser ved å blande disse med sterkere resultater fra større klasser. I forvirringsmatrisen i Figur 4.6 er ubalansen tydelig for sensorene HySpex og Sentinel-2, mens den er mindre synlig for EMIT på grunn av færre testpiksler. Bruk av makrogjennomsnitt er nødvendig for å sikre en rettfærdig vurdering av ytelsen, selv om det kan lede til en overvurdering av effektiviteten (Sefidian, 2022).

5.3.2 Klassifisering 1: alunitt

I klassifiseringen 1 med HySpex, vist i Figur 4.3, observeres det at klassifiseringen er vellykket. Resultatene presentert i Kapittel 4.1.1 viser at nøyaktigheten er overraskende høy for både Sentinel-2 og EMIT, med tanke på sensorenes begrensninger i spektral og geometrisk oppløsning. Trendene i begge klassifiseringene, sammenlignet med HySpex-klassifiseringene, er gode. Det oppdages alunitt i de samme områdene der det er store forekomster, noe som indikerer at andelen av alunitt innenfor

pikslene skal være betydelig høy for at pikslene skal bli klassifisert som «alunitt» og ikke som «ukjent».

Det er en god overensstemmelse mellom klassifiseringen og forekomstene av ukjente piksler. Det observeres at Sentinel-2 har en større avviksrate når det gjelder feilklassifiserte piksler sammenlignet med EMIT. Hovedgrunnen til dette antas å være den spektrale oppløsningen, som er vesentlig forskjellig for de to sensorene. Alunitt detekteres i SWIR-området, der mineralet absorberes tydelig fra 1300 nm til 2500 nm. EMIT har bedre dekning enn Sentinel-2 innenfor dette intervallet, og har derfor en mindre feilmargin av piksler som er klassifisert feil. Til tross for EMITs svakere geometriske oppløsning, har EMIT en helt annen evne til å detektere forskjeller i den spektrale signaturen innenfor SWIR-båndene enn hva Sentinel-2 har. Som nevnt tidligere har EMIT alt for få testpunkter. Av den grunn er ikke valideringsparameterene til EMIT pålitelige, selv om punktene er klassifisert riktig.

5.3.3 Klassifisering 2: alunitt og goethitt

I den andre klassifiseringsrunden ble treningsdatasettet utvidet med de to beste goethitt-klassene, i tillegg til klassene «alunitt» og «ukjent». Den dårligst presterende goethitt-klassen ble utelatt, en beslutning som er nærmere beskrevet i Kapittel 3.6.4.

I klassifiseringen med HySpex, vist i Figur 4.8, observeres det at klassifiseringen er vellykket. Den spektrale og geometriske oppløsningen til HySpex-bildet, gjør det mulig å klassifisere de to mineralene ut ifra treningsdataen som er blitt generert. Dette verifiseres i Figur 4.7, der valideringsparameterene fortsatt presterer godt. Valideringsparameterene er blitt noe svakere sammenlignet med klassifiseringen av kun alunitt, i Figur 4.2.

Goethitt detekteres best i spektralområdet fra 400 nm til 1300 nm. Sentinel-2 har hele 8 av 10 bånd innenfor dette intervallet. På bakgrunn av dette gir det Sentinel-2 bedre forutsetninger for deteksjon av mineralet goethitt enn alunitt. EMIT har en eksepsjonell god spektral oppløsning og har hele 121 bånd innenfor dette området. Imidlertid, på grunn av sin kjente svake geometriske oppløsning, er deteksjonsevnen betydelig svekket. Klassen «goethitt» er mindre enn klassene «alunitt» og «ukjent», noe som har resultert i en underklassifisering av goethitt og overklassifisering av alunitt.

På grunn av grovere spektral og geometrisk oppløsning i Sentinel-2 og EMIT, kan klassifiseringen snarere betraktes som en indikasjon på potensielle mineralforekomster enn en direkte deteksjon.

Likevel kan begge sensorene være nyttige i tidlig fase for å gi kostnadsbesparende og tidseffektive indikasjoner på interessante områder for videre analyse.

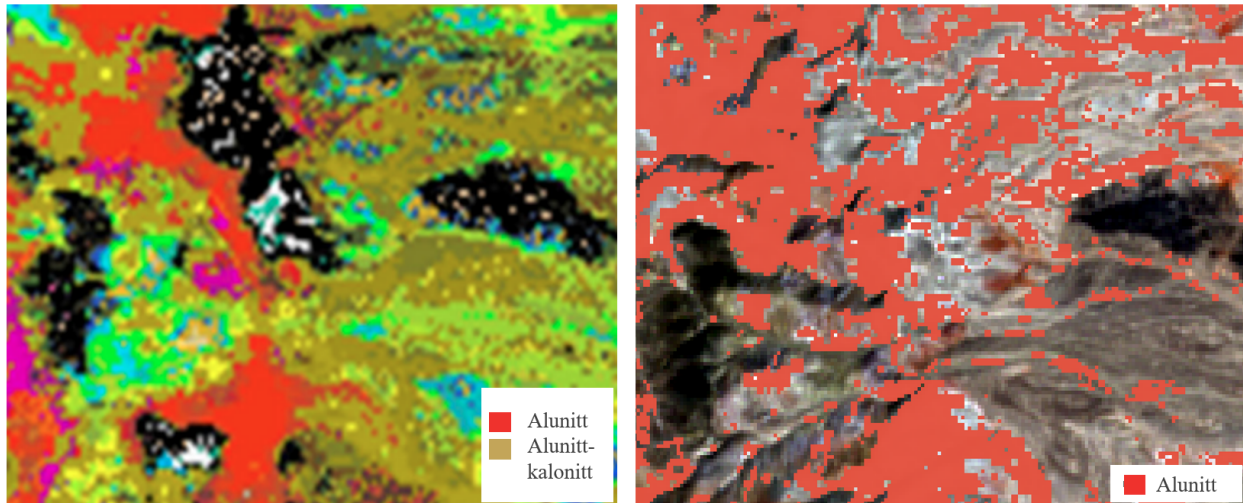
5.3.4 Klassifisering 3: utvidelse av området

Resultatene presentert i Kapittel 4.3 viser klassifisering av alunitt og goethitt med Sentinel-2 og EMIT. Test- og treningsdataene er de samme som i tidligere klassifiseringer, noe som betyr at valideringsparameterene også forblir uendrede. Disse parameterene gir derfor ingen innsikt i hvor effektiv klassifiseringen er på det utvidede området, og er utelatt fra resultatene. En betydelig utfordring ved å bruke test- og treningsdata fra et begrenset geografisk område er at lokale forhold kan variere sterkt når dette området utvides. Spesielt kan faktorer som skygger, påvirke klassifiseringens nøyaktighet.

En annen utfordring er det benyttede sammenligningsgrunnlaget, diskutert i Kapittel 5.2.4, som er basert på tidligere forskning på det utvidede området. Denne forskningen er av eldre dato, og presenterer resultater som varierer mye, slik at de gir et begrenset grunnlag for pålitelige sammenligninger. Til tross for disse begrensningene er det gjort forsøk på å sammenligne tilnærmingen med eksisterende data så godt som mulig i Figur 5.1, Figur 5.2, Figur 5.3 og Figur 5.4.

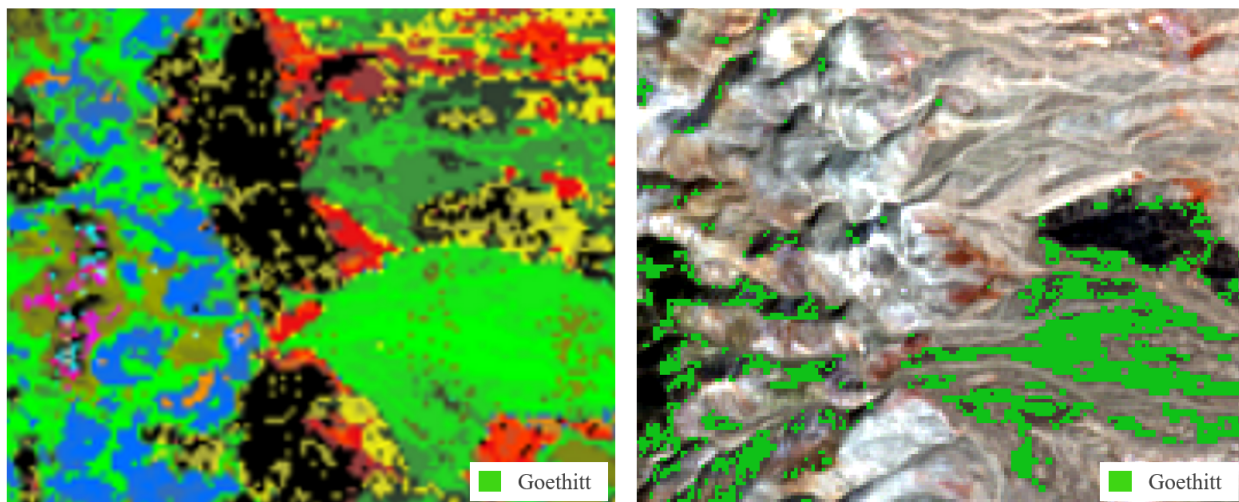
I de fire følgende figurene vises tidligere klassifisering gjort av Swayze mfl. (2014) i bildet til venstre, mens klassifisering gjort i denne analysen vises til høyre. Bildene fra denne analysen er de samme som ble presentert i Kapittel 4, men nå er de visualisert med kun et mineral om gangen. Dette er fordi Swayze mfl. (2014) har presentert to kart - et med jernholdige mineraler (Swayze mfl., 2014, s. 17) og et med leire, sulfater og karbonater (Swayze mfl., 2014, s. 18). Sammenligningen blir enklere ved å kun visualisere alunitt ved siden av kartet med leire, sulfater og karbonater og kun visualisere goethitt ved siden av kartet med jernholdige mineraler.

I Figur 5.1 inkluderes både alunitt og alunitt-kalonitt i tegnforklaringen. Dette valget er basert på at deres spektralsignaturer er svært like. For vår relativt grove klassifisering er det derfor mer presist å inkludere begge. Det er tydelige likhetstrekk i tendensene som observeres i begge bildene, og klassifiseringen med Sentinel-2 ser vellykket ut.



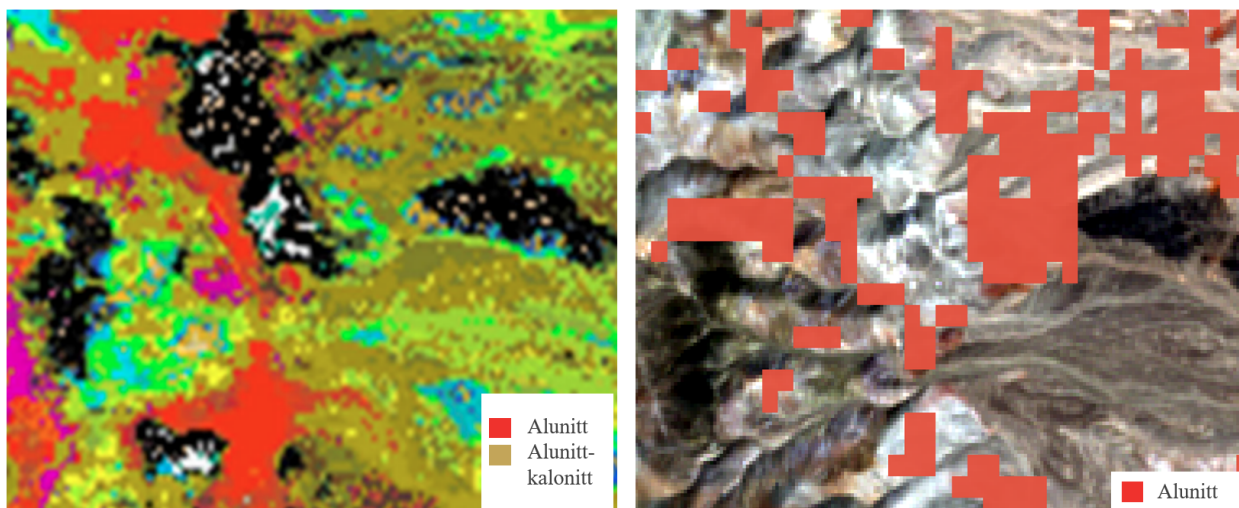
Figur 5.1: Figuren viser to utklipp fra samme studieområdet for alunitt-funn med Sentinel-2. Bildet til venstre er fra Swayze mfl. (2014) (s. 18) og bildet til høyre er fra denne analysen. I bildet til venstre er både alunitt og alunitt-kalonitt tatt med, i henholdsvis fargene rød og mørk gul. I bildet til høyre er alunitt representert i fargen rød.

I Figur 5.2 er det likhetstrekk i tendensene som observeres. De samme områdene med funn er identifisert i begge kartene, men kartet til høyre viser en mindre omfattende klassifisering. Dette skyldes, som nevnt i Kapittel 5.2.2.1, at den laveste kvalitetsklassen av goethitt ble fjernet for å unngå overklassifisering. Resultatet er en tendens til underklassifisering, noe som er tydelig i dette tilfellet.



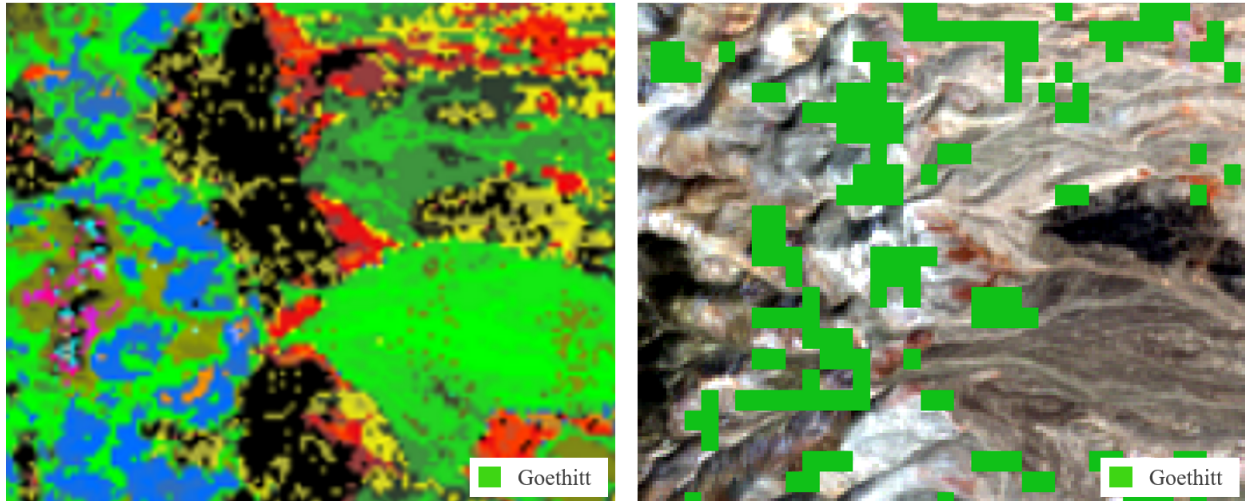
Figur 5.2: Figuren viser to utklipp fra samme studieområdet for goethitt-funn med Sentinel-2. Bildet til venstre er fra Swayze mfl. (2014) (s. 17) og bildet til høyre er fra denne analysen. I bildet til venstre er goethitt representert med fargen grønn. I bildet til høyre er goethitt også representert i fargen grønn.

Figur 5.3 sammenligner tidligere klassifisering av alunitt med klassifisering utført i denne analysen ved bruk av EMIT. Klassifiseringen viser tegn på underklassifisering, og det er utfordrende å identifisere tydelige likhetstrekk. Et positivt aspekt er imidlertid at de fleste områdene som er klassifisert som ukjente, er konsekvent klassifisert slik i begge tilfeller.



Figur 5.3: Figuren viser to utklipp fra samme studieområdet for alunitt-funn med EMIT. Bildet til venstre er fra Swayze mfl. (2014) (s. 18) og bildet til høyre er fra denne analysen. I bildet til venstre er både alunitt og alunitt-kalonitt tatt med, i henholdsvis fargene rød og mørk gul. I bildet til høyre er alunitt representert i fargen rød.

I Figur 5.4 presenteres den minst vellykkede klassifiseringen utført i hele analysen. Det er ingen tydelige tegn til likhetstrekk mellom klassifiseringene. Goethittmineralet krever flere trening- og testpunkt for å kunne klassifiseres nøyaktig.



Figur 5.4: Figuren viser to utklipp fra samme studieområdet for goethitt-funn med EMIT. Bildet til venstre er fra Swayze mfl. (2014) (s. 17) og bildet til høyre er fra denne analysen. I bildet til venstre er goethitt representert med fargen grønn. I bildet til høyre er goethitt også representert i fargen grønn.

Det er ikke umiddelbart tydelig at det er fellestrekk mellom klassifisering av utvidede områder med EMIT og Sentinel-2. Tidligere forskning presenterer også forskjellige resultater, noe som gjør det utfordrende å utføre en grundig validering.

6 Konklusjon

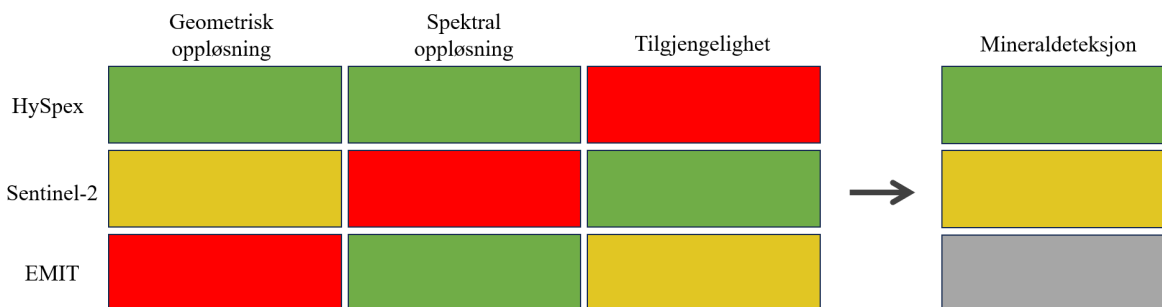
Formålet med analysen var å utforske muligheten for å klassifisere mineralene alunitt og goethitt ved hjelp av tre ulike sensorer, med hver sine særegne geometriske og spektrale oppløsninger. Forskningsspørsmålene fokuserte på hvordan disse varierende oppløsningene påvirker kvaliteten på mineraldeteksjonen, om det er mulig å detektere alunitt og goethitt med fjernmålingsdata fra satellitt, og hvor godt den utviklede klassifiseringsmetoden presterer over et større geografisk område.

Det er allment kjent at det er mulig å detektere mineraler med HySpex-teknologi, noe som også konkluderes med i denne analysen. Muligheten for å oppnå tilsvarende resultater fra satellitt er imidlertid mer usikkert. Resultatene fra Sentinel-2 viser at det er mulig å indikere mineraler med satellitt. For EMIT-satellitten er det vanskeligere å trekke konklusjoner om mineraldeteksjon, siden det begrensede studieområdet gir for lite test- og treningsdata. Påliteligheten i klassifiseringen av større områder er vanskelig å bedømme på grunn av mangel på sammenligningsgrunnlag. Likevel er det tydelige likheter i kartleggingen av større områder med Sentinel-2, noe som støtter konklusjonen om at denne satellitten kan indikere mineralforekomster.

HySpex-sensoren oppnådde de høyeste F1-verdiene med 0.97 i klassifisering 1 og 0.90 i klassifisering 2, noe som indikerer en svært høy grad av nøyaktighet. I kontrast viste Sentinel-2 vesentlig lavere F1-verdier på 0.81 og 0.67, mens EMIT hadde de laveste F1-verdiene på 0.73 og 0.61. Selv om EMITs større piksler resulterte i en lavere geometrisk oppløsning og dermed en svekket evne til å konkurrere med Sentinel-2, gjorde dens overlegne spektraloppløsning at den nesten nådde nivåene til Sentinel-2 i nøyaktighet. Dette understreker hvor kritisk de kortbølgede infrarøde båndene, som EMIT håndterer bedre enn Sentinel-2, er for nøyaktig mineraldeteksjon. EMIT viser særlig styrke i deteksjonen av alunitt på grunn av dets fremragende spektrale oppløsning. Alunitt har markante absorpsjonsbånd mellom 1300 nm og 2500 nm, en region hvor EMITs kapasitet til å skille mellom fine spektrale detaljer gir en betydelig fordel. Hadde analysen fokusert på noe annet enn mineraler, for eksempel vegetasjon, ville sannsynligvis gapet i nøyaktighet mellom Sentinel og EMIT vært større.

Figur 6.1 viser en forenklet sammenligning av sensorenes evne til å detektere mineraler, basert på geometrisk oppløsning, spektral oppløsning og hvor tilgjengelige dataen er. Figuren bruker fargekoding med fargene grønn, gul og rød, som henholdsvis betyr høy, moderat og lav. Mineraldeteksjon med HySpex kan gjennomføres med høy nøyaktighet. Mineraldeteksjon med Sentinel-2 er mulig, men den svekkede spektrale oppløsningen gjør det utfordrende uten støtte fra annen sensor. Fargen grå er også

benyttet for mineraldeteksjon med EMIT. Dette grunnet at det ikke er tilstrekkelig datagrunnlag for EMIT i denne analysen, og resultatene er dermed ikke er pålitelige. Likevel har vi grunnet den gode spektrale oppløsningen tro på at EMIT kan brukes i mineraldeteksjon, så lenge det finnes store nok områder med et konsentrert mineral.



Figur 6.1: Sammenligning av sensorenes evne til å detektere mineraler. Grønn betyr høy, gul betyr moderat og rød betyr lav. Grå er også benyttet der det ikke er nok datagrunnlag for å uttale seg.

Når det gjelder klassifisering 3 av et større geografisk område er det enda en usikkerhetsfaktor som spiller inn. På grunn av mangel på sammenligningsgrunnlag og varierende resultater i tidligere forskning er det utfordrende å trekke distinkte konklusjoner. Visuell sammenligning med et større område og eldre forskning indikerer at Sentinel-2 kan avsløre trender som ligner på tidligere observasjoner. Det er vanskelig å se de samme trendene med EMIT på grunn av dens begrensninger i geometrisk oppløsning. Denne usikkerheten krever ytterligere studier, med for eksempel HySpex-data i det utvidede området, for å bekrefte metodens effektivitet i større skala.

Når effektiviteten til klassifiseringsmodellen skal vurderes, er det viktig å huske på at den ikke trenger å være perfekt for å være nyttig. For eksempel kan en F1-verdi på 0.70 være tilstrekkelig for å identifisere store områder med forekomster av de mineralene det søkes etter. I praksis, når målet er å finne nye steder å utvinne mineraler, er det ofte mer relevant å oppdage steder hvor det er store konsentrasjoner av mineralet. I slike tilfeller kan sensorer med grovere piksler være gode nok. Det første trinnet er å få en indikasjon på potensielle funnsteder; deretter kan man reise ut til disse områdene for å verifisere om dataene fra fjernmålingen stemmer overens med den faktiske situasjonen. Denne tilnærmingen balanserer kostnadseffektivitet og praktisk anvendelighet, og sikrer at ressursene brukes der de gir mest verdi.

6.1 Forslag til videre arbeid

Følgende forslag til videre arbeid ville ha gjort både metoden og resultatene i analysen bedre.

- 1. Innhenting av utvidede datasett fra HySpex:** Det hadde vært gunstig å skaffe data fra HySpex for et større område. Dette vil gjøre det mulig å teste hvor godt klassifiseringen fungerer over et bredere område og gi et klarere bilde av hvor nyttig modellen er generelt.
- 2. Videreutvikling av mineralindeksene for goethitt:** For å styrke grunnlaget for treningspiksler, foreslås det å videreutvikle mineralindeksen for goethitt, slik at denne kan bli mer presis. Dette vil gi et klarere utgangspunkt for klassifisering og forbedre presisjonen for funn av mineralforekomster.
- 3. Anvendelse av en alternativ maskinlæringsmodell:** For å forbedre nøyaktigheten i prediksjonene kunne det ha blitt benyttet en mer spesialisert maskinlæringsmodell der parametere kan justeres. Dette vil muliggjøre en mer finjustert tilnærming og potensielt gitt bedre resultater.
- 4. Spectral unmixing av store piksler:** For å forbedre nøyaktigheten i analysen ville det vært fordelaktig å gjennomføre en grundig spektral unmixing av pikslene fra EMIT. EMITs store geografiske piksler fører til at alle pikslene er blandingspiksler, men den høye spektrale oppløsningen tillater likevel en spektral unmixing (Lanaras mfl., 2017, s. 2). Dette kan brukes til å estimere prosentandeler av alunitt, goethitt og ukjente materialer. Resultatet vil gi en mer detaljert forståelse av stoffenes fordeling og sammensetning i bildene.

Referanser

- ArcMap. (2021). BIL, BIP, and BSQ raster files [Hentet: 11.04.24]. <https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/bil-bip-and-bsq-raster-files.htm>
- Astro Gallery of Gems. (2024). Goethite from Wulashan mining district, Baotou City, Inner Mongolia, China [Hentet: 02.05.24]. <https://astrogallery.com/collections/minerals/products/rounded-crystal-aggregates-of-brown-goethite-on-pyrolusite-matrix-with-minor-crystals>
- Bahr, T. (2022a). ENVI Advanced Spectral Analytics Part II: Case Study Cuprite Mining District. *HySpex*.
- Bahr, T. (2022b). ENVI Advanced Spectral Analytics Part II: Case Study Cuprite Mining District. *L3Harris*.
- Banoula, M. (2023). Supervised and Unsupervised Learning in Machine Learning [Hentet: 17.04.24]. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-and-unsupervised-learning>
- Barkved, K. (2022). What is Training Data? [Hentet: 29.04.24]. *The Difference Between Training Data vs. Test Data in Machine Learning*. <https://obviously.ai/post/the-difference-between-training-data-vs-test-data-in-machine-learning>
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Birkeland, R. (2022). Atmosfærisk vindauge. *Store norske leksikon*. https://snl.no/atmosf%C3%A6risk_vindauge
- Bryhni, I. (2020a). Alunitt [Hentet: 05.04.24]. *Store norske leksikon*. <https://snl.no/alunitt>
- Bryhni, I. (2020b). Goethitt [Hentet: 05.04.24]. *Store norske leksikon*. <https://snl.no/goethitt>
- Carrion, G., Oxoli, A. B., Biraghi, G., Bratic, Polinelli, T. H., & Vavassori. (2021). Principles of image analysis [Hentet: 11.04.24]. *What is a spectral signature*. https://gis4schools.readthedocs.io/en/latest/part3/rs_imageanalysis.html
- Catalyst. (2024). About CATALYST [Hentet: 29.04.24]. *Catalyst Earth*. <https://catalyst.earth/about/>
- Chawla, N. V. (2003). C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure [Hentet: 06.05.24]. *Proceedings of the ICML*, 3, 66.
- Dawson, N. (2024). QGIS [Hentet: 11.04.24]. *GitHub*. <https://github.com/qgis/QGIS>

- Dubrov, V. (2023). Understanding Machine Learning Robustness: Why It Matters and How It Affects Your Models [Hentet: 16.04.24]. *Medium*. <https://medium.com/@slavadubrov/understanding-machine-learning-robustness-why-it-matters-and-how-it-affects-your-models-5e2cb5838dab>
- Editors of Encyclopaedia Britannica. (2019). Goethite [Hentet: 02.05.24]. <https://www.britannica.com/science/goethite>
- Editors of Encyclopaedia Britannica. (2021). Alunite [Hentet: 02.05.24]. <https://www.britannica.com/science/alunite>
- EdmundOptics. (2024). Hyperspectral and Multispectral Imaging [Hentet: 03.04.24]. *Edmund Optics*. <https://www.edmundoptics.com/knowledge-center/application-notes/imaging/hyperspectral-and-multispectral-imaging/>
- El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? I *Machine Learning in Radiation Oncology* (s. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1
- European Space Agency. (2023a). S2 Mission [Hentet: 03.05.24]. *SentiWiki*. <https://sentiwiki.copernicus.eu/web/s2-mission>
- European Space Agency. (2023b). Satellite Description [Hentet: 03.05.24]. *SentiWiki*. <https://sentiwiki.copernicus.eu/web/s2-mission/#S2Mission-SatelliteDescriptionS2-Mission-Satellite-Descriptiontrue>
- European Space Agency. (2023c). Spatial Resolution [Hentet: 03.05.24]. *SentiWiki*. <https://sentiwiki.copernicus.eu/web/s2-mission/#S2Mission-SpatialResolutionS2-Mission-Spatial-Resolutiontrue>
- FileInfo. (2024). .PIX File Extension [Hentet: 19.04.24]. *PCI Geomatics*. <https://fileinfo.com/extension/pix>
- Fisher, T. (2022). *What Are TIF and TIFF Files?* [Hentet: 09.04.24]. <https://www.lifewire.com/tiff-file-2622393>
- Gerdo. (2022). NetCDF4 for Dummies [Hentet: 18.04.24]. *Medium*. <https://medium.com/@gerdo/netcdf4-for-dummies-7a77ca5e083f>
- GISGeography. (2023). What are the Differences Between Multispectral and Hyperspectral Imagery? [Hentet: 26.03.24]. *Multispectral vs Hyperspectral Imagery Explained*. <https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>
- GISGeography. (2024a). Sentinel 2 Bands [Hentet: 26.03.24]. *Sentinel 2 Bands and Combinations*. <https://www.gisgeography.com/sentinel-2-bands-combinations/>

- GISGeography. (2024b). Spectral Signature Cheatsheet in Remote Sensing [Hentet: 26.03.24]. <https://gisgeography.com/spectral-signature/>
- Goetz, A. F. H., & Srivastava, V. (1985). Mineralogical Mapping in the Cuprite Mining District, Nevada. *Proceedings of the Airborne Imaging Spectrometer Data Analysis Workshop*, N86–11619.
- Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: an Overview [Hentet: 08.04.24]. *Cornell University*. <https://arxiv.org/abs/2008.05756>
- Henrich, V., Krauss, G., Götze, C., & Sandow, C. (2011). Index: Ferric Oxides [Hentet: 03.05.24]. *Index DataBase*. <https://www.indexdatabase.de/db/i-single.php?id=20>
- Hill, B. (2013). Integrating remote sensing techniques at Cuprite, Nevada: AVIRIS, Thematic Mapper, and field spectroscopy [Hentet: 22.04.24]. *NTRS - NASA Technical Reports Server*. <https://ntrs.nasa.gov/citations/19940012228>
- HySpex. (2019). HySpex MJOLNIR VS-620 [Hentet: 11.03.24]. *Norsk Elektro Optikk AS*. https://is.earsel.org/workshop/11-IS-Brno2019/wp-content/uploads/2018/08/HySpex_Mjолnir_VS-620v2.pdf
- Kanstrén, T. (2020). A Look at Precision, Recall, and F1-Score [Hentet: 24.04.24]. *Towards Data Science*. <https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec>
- Karasiak, N., Dejoux, J.-F., Monteil, C., & Sheeren, D. (2021). Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, 111(7). <https://doi.org/10.1007/s10994-021-05972-1>
- Karasiak, N. (2016). Dzetsaka Qgis Classification plugin [Hentet: 18.04.24]. <https://github.com/nkarasiak/dzetsaka>
- Koehrsen, W. (2018). Overfitting vs. Underfitting: A Complete Example [Hentet: 11.04.24]. *Towards Data Science*. <https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765>
- Kokaly, R. F., Clark, R. N., Swayze, G. A., Livo, K. E., Hoefen, T. M., Pearson, N. C., Wise, R. A., Benzel, W. M., Lowers, H. A., Driscoll, R. L., & Klein, A. J. (2017). *USGS Spectral Library Version 7*. <https://doi.org/10.3133/ds1035>
- Lanaras, C., Baltasvias, E., & Schindler, K. (2017). Hyperspectral Super-Resolution with Spectral Unmixing Constraints. *Remote Sensing*, 9(11). <https://doi.org/10.3390/rs9111196>

- Le Comptoir Geologique. (u.d). Alunite - Encyclopedia [Hentet: 07.05.24]. <https://le-comptoir-geologique.com/alunite-encyclopedia.html>
- Lied, F., & Birkeland, R. (2022). Fjernmåling [Hentet: 05.04.24]. *Store norske leksikon*. <https://snl.no/fjernm%C3%A5ling>
- Machine Learning Models. (2024). Pros and Cons of Random Forest Algorithm for ML Classification [Hentet: 26.04.24]. <https://machinelearningmodels.org/pros-and-cons-of-random-forest-algorithm-for-ml-classification/>
- MathWorks. (u.d). Overfitting vs. Underfitting [Hentet: 30.04.24]. <https://se.mathworks.com/discovery/overfitting.html>
- More, A. S., & Rana, D. P. (2017). Review of random forest classification techniques to resolve data imbalance, 72–78. <https://doi.org/10.1109/ICISIM.2017.8122151>
- NASA. (2023a). EMIT - science - objectives. *California Institute of Technology*. <https://earth.jpl.nasa.gov/emit/science/objectives/>
- NASA. (2023b). EMIT - science - specifications. *California Institute of Technology*. <https://earth.jpl.nasa.gov/emit/instrument/specifications/>
- NASA. (2023c). Exploring EMIT L2A Reflectance [Hentet: 18.03.24]. *EMIT-Data-Resources*. https://github.com/nasa/EMIT-Data-Resources/blob/main/python/tutorials/Exploring_EMIT_L2A_Reflectance.ipynb
- NASA. (2023d). How to Convert to ENVI [Hentet: 18.03.24]. *EMIT-Data-Resources*. https://github.com/nasa/EMIT-Data-Resources/blob/main/python/how-tos/How_to_Convert_to_ENVI.ipynb
- NEO. (u.d). HySpex [Hentet: 12.04.24]. *Norsk Elektro Optikk AS*. <https://www.neo.no/products/hyspex>
- Ninomiya, Y. (2003). A stabilized vegetation index and several mineralogic indices defined for ASTER VNIR and SWIR data. *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. <https://doi.org/10.1109/igarss.2003.1294172>
- NV5. (2013). Radiance vs. Reflectance [Hentet: 12.04.24]. *NV5 GEOSPATIAL SOFTWARE*. <https://www.nv5geospatialsoftware.com/Support/Maintenance-Detail/ArtMID/13350/ArticleID/19247/3377>
- Nätt, T. H. (2023). Python (programmeringsspråk) [Hentet: 14.03.24]. *Store Norske Leksikon*. https://snl.no/Python_-_programmeringsspr%C3%A5k

- Okada, K. (2022). Breakthrough technologies for mineral exploration. *Mineral Economics*, 35(3–4), 429–454. <https://doi.org/10.1007/s13563-022-00317-3>
- OpenAI. (2023). ChatGPT: Optimizing Language Models for Dialogue [Hentet: 13.05.24]. <https://openai.com/chatgpt/>
- PCI Geomatics. (2013). *Geomatica Focus User Guide* [Hentet: 30.04.24]. <https://is.muni.cz/el/1431/jaro2016/Z8108/um/61822414/55340959/focus.pdf>
- Peyghambari, S., & Zhang, Y. (2021). Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *Journal of Applied Remote Sensing*, 15(03). <https://doi.org/10.1117/1.jrs.15.031501>
- Pinasu, D. (2020). Classification Using dzetsaka Plugin in QGIS [Hentet: 18.04.24]. *Medium*. <https://medium.com/@st120929/classification-using-dzetsaka-plugin-in-qgis-90b93d1784c6>
- Polyzotis, N., Zinkevich, M., Roy, S., Breck, E., & Whang, S. (2019). Data validation for machine learning. *Proceedings of machine learning and systems*, 1, 334–347.
- Precisely. (u.d). Shapefile [Hentet: 09.04.24]. *precisely*. <https://www.precisely.com/glossary/shapefile>
- Probst, P., Wright, M., & Boulesteix, A.-L. (2024). Hyperparameters and Tuning Strategies for Random Forest [Hentet: 15.04.24]. *arXiv preprint arXiv*. <https://arxiv.org/abs/1804.03515>
- Rowan, L. C., & Mars, J. C. (2003). Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sensing of Environment*, 84(3), 350–366. [https://doi.org/10.1016/S0034-4257\(02\)00127-X](https://doi.org/10.1016/S0034-4257(02)00127-X)
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00592-x>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal: Promoting communications on statistics and Stata*, 20(1), 3–29. <https://doi.org/10.1177/1536867x20909688>
- Sefidian, A. M. (2022). Understanding Micro, Macro, and Weighted Averages for Scikit-Learn metrics in multi-class classification with example [Hentet: 11.05.24]. <https://iamirmasoud.com/2022/06/19/understanding-micro-macro-and-weighted-averages-for-scikit-learn-metrics-in-multi-class-classification-with-example/>
- Shani, C., & Binieli, M. (2019). An overview of Principal Component Analysis [Hentet: 04.05.24]. *freeCodeCamp*. <https://www.freecodecamp.org/news/an-overview-of-principal-component-analysis-6340e3bc4073/>

- Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(07), 3897–3904. <https://ssrn.com/abstract=4458723>
- Swayze, G. A., Clark, R. N., Goetz, A. F. H., Livo, K. E., Breit, G. N., Kruse, F. A., Sutley, S. J., Snee, L. W., Lowers, H. A., Post, J. L., Stoffregen, R. E., & Ashley, R. P. (2014). Mapping Advanced Argillic Alteration at Cuprite, Nevada, Using Imaging Spectroscopy. *Economic Geology*, 109(5), 1179–1221. <https://doi.org/10.2113/econgeo.109.5.1179>
- Tripathi, P., & Garg, R. D. (2021). Revisiting the Geology of Cuprite Hills from Nevada Using Desis and Prisma Hyperspectral Imageries. *Astrophysics Data System*. <https://ui.adsabs.harvard.edu/abs/2021AGUFMGC15B0682T/abstract>
- van der Meer, F. D., van der Werff, H. M., van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., van der Meijde, M., Carranza, E. J. M., Smeth, J. B. d., & Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 14(1), 112–128. <https://doi.org/10.1016/j.jag.2011.08.002>
- Aarsten, D. (2022). Forelesning 3 - Plattformer for fjernmåling, radiometrisk bildekorreksjon og geometrisk bildekorreksjon [Forelesning i GMBB300, Norges miljø- og biovitenskapelige universitet, 21. sep 2022].



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

Postboks 5003
NO-1432 Ås
Norway