



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2021 30 stp
Fakultet for realfag og teknologi

Evaluering av maskinlæring brukt til endringsdeteksjon av arealtyper i AR5 basert på flybilder

Evaluation of Machine Learning Applied on Change
Detection of Land Types in AR5 Based on Aerial
Imagery

Nicolai Munsterhjelm
Datavitenskap

Sammen drag

Arealressurskartet AR5 ajourføres periodisk av Norsk institutt for bioøkonomi (NIBIO) ved bruk av flybilder. Maskinlæring kan øke kvaliteten på endringskartene og effektivisere ajourføringsprosessen. NIBIO har i løpet av det siste året jobbet med å utvikle og trene en maskinlæringsmodell, en siamesisk konvolusjonell variasjonsautokoder. Modellen har som formål å oppdage områder med endret arealtype i AR5 basert på RGB-flybilder.

Denne oppgaven tar for seg en statistisk evaluering av prediksjonene fra modellen. Den undersøker også forskjellen i klassifiseringsnøyaktighet mellom forskjellige areal typer og forskjellige geografiske områder. Kommunene med data i denne oppgaven hører til de tre regionene Romerike/Glåmdalen, Valdres og Jæren. Matthews korrelasjonskoeffisient (MCC) ble valgt som mål på klassifiseringsnøyaktighet fordi den er uavhengig av balanse i datasett, som gjør at MCC til ulike datasett kan sammenliknes direkte.

Resultatene viser at MCC for det samlede datasettet ligger på 0,104, som betyr at det er en svak korrelasjon mellom modellens prediksjoner og faktiske endringer, litt bedre enn tilfeldig gjetning. Modellen har altså lært seg enkelte sammenhenger som er nyttige for klassifiseringen, men når ikke opp til nøyaktigheten som forventes av en endringsdeteksjonsmodell som faktisk skal kunne tas i bruk.

MCC varierer en del mellom areal typene, hvor det er skog, fulldyrka jord, bebyggd og innmarksbeite som oppnår best resultat ($0,09 < MCC < 0,14$). Åpen fastmark, myr og overflatedyrka jord er areal typene som er mest utfordrende for modellen å predikere ($-0,01 < MCC < 0,03$). Flere konkrete områder med høy andel feilprediksjoner har blitt identifisert, blant annet vegetasjon i bebyggelse, flyplasser, hogstflater, golfbaner, snaumark i fjellet, tresatt myr og skyggelagte områder.

Modellen oppnådde bedre resultat i Romerike/Glåmdalen enn i Valdres og på Jæren. Dette kan delvis forklares av at en større andel av treningsdataene kommer fra Romerike/Glåmdalen. I tillegg er det en større andel åpen fastmark i Valdres og på Jæren, som reduserer klassifiseringsnøyaktigheten.

Flere mulige tiltak for å forbedre modellens klassifiseringsnøyaktighet foreslås: mer variasjon i treningsdata, vektning av vanskelige observasjoner, skyggekorreksjon,

kombinere flybilder med andre typer fjernmålingsdata, hyperparameteroptimalisering og eventuelt å implementere bildesegmenteringsalgoritmer.

Abstract

The land resource map AR5 is periodically updated by the Norwegian Institute of Bioeconomy Research (NIBIO) from aerial imagery. Machine learning can increase the quality of the change maps and streamline the updating process. NIBIO has during the last year developed and trained a machine learning model, a siamese convolutional variational autoencoder. The purpose of this model is to detect areas where the land type has changed in AR5, based on RGB aerial imagery.

This thesis performs a statistical evaluation of the model's predictions. It also examines the difference in classification performance between different land types and different geographic regions. The municipalities with data in this thesis are in the three regions Romerike/Glåmdalen, Valdres, and Jæren. Matthew's correlation coefficient (MCC) was chosen for classification performance metric because it is independent of balance in datasets, which means that MCC of different datasets can be directly compared.

The results show that MCC for the total dataset equals 0,104, which means that the model's predictions are weakly correlated to actual changes. In other words, the model has learned certain patterns that are useful for classification but does not reach the performance that is expected of a change detection model that is to be used.

MCC varies somewhat between land types, where woods, fully cultivated land, built-up area, and infield pastures achieve the best performance ($0,09 < \text{MCC} < 0,14$). Open land, bog, and surface cultivated land are the area types that are most difficult for the model to predict ($-0,01 < \text{MCC} < 0,03$). Several specific situations with a high proportion of wrong predictions were identified, including vegetated built-up areas, airports, felling areas, golf courses, open vegetated land in the mountains, wooded bogs, and shaded areas.

The model achieved better performance in Romerike/Glåmdalen than in Valdres and Jæren. This can be partly explained by the fact that more of the training data is from

Romerike/Glåmdalen. Furthermore, open land is more common in Valdres and Jæren, which reduces the classification performance.

Several measures to increase classification performance are proposed: more varied training data, weighting difficult observations, shadow correction, combining with other remote sensing data, hyperparameter optimization, and possibly implementing image segmentation algorithms.

Forord

Denne oppgaven avslutter min master i datavitenskap ved NMBU. Jeg er svært takknemlig for å ha fått muligheten til å skrive om et så spennende tema. Kartlegging av landskapet har alltid interessert meg, og etter hvert har jeg innsett de store samfunnsmessige verdiene dette arbeidet innebærer. I løpet av tiden siden jeg begynte å studere har det vært en voldsom utvikling innen maskinlæring. Vi er nok ennå bare i startgropa av denne utviklingen, og jeg gleder meg til å se hva vi kan oppnå med disse metodene i fremtiden.

Jeg vil rette en stor takk til mine veiledere, hovedveileder Ingunn Burud ved RealTek, NMBU, og biveileder Misganu Debella-Gilo ved NIBIO. Deres gode innspill og konstruktive tilbakemeldinger har vært av uvurderlig verdi under arbeidet med oppgaven. Videre vil jeg takke flere fra NIBIO. Takk til Åsmund Ertshus Mathisen for behjelpelighet med AR5-data og tilgang til WMS-tjenester. Takk til Jonathan Rizzi for informasjon og detaljer angående modell og data. Takk til Jostein Frydenlund for introduksjon til prosjektet.

Til slutt til jeg takke familie og samboer for all støtte og oppmuntring i løpet av arbeidet med oppgaven.

Holmestrand, 21. mai 2021

Nicolai Munsterhjelm

Figurliste

Figur 1. Figuren illustrerer et forenklet eksempel på et kunstig nevralt nettverk med tre skjulte lag. Antall lag og antall nevroner i hvert lag kan varieres.....	4
Figur 2. Illustrasjon av et typisk konvolusjonelt nevralt nettverk. Konvolusjoner og oppsamlinger fanger opp mønstre i “feature maps“. Figur fra (Wikimedia Commons bruker Aphex34, 2015).....	5
Figur 3. Illustrasjon av en autokoder. X er input-dataene, h er den komprimerte representasjonen, og X' er de dekomprimerte dataene. Figur fra (Massi, 2019).....	5
Figur 4. Illustrasjon av en variasjonsautokoder. Den komprimerte representasjonen består av et sett med normalfordelinger (μ), og standardavvik (σ).....	6
Figur 5. Viser oppsettet til en forvirringsmatrise. TP er sanne positive observasjoner, FN er falske negative observasjoner, FP er falske positive observasjoner, mens TN er sanne negative observasjoner.	7
Figur 6. Kartet over sørlige Norge viser hvilke kommuner som har blitt brukt til trening, og hvilke som ble brukt til evaluering. I sørvest har vi Jæren-kommunene, lengst nord er Valdres-kommunene, og klyngen lengst øst er Romerike/Glåmdalen-kommunene.....	12
Figur 7. Figuren viser hvordan kvadratiske områder med manglende prediksjoner (markert i svart) opptrer tilfeldig spredt rundt i enkelte kommuner. Eksempel fra Gjerdrum.....	15
Figur 8. Kartet til venstre viser et utsnitt av arealtypekartet, og til høyre er arealtypene som har blitt tilordnet rutenettet til prediksjonene.....	16
Figur 9. Viser forenklet hvordan modellen tar AR5-data og flybilledata som input, og beregner en sannsynlighet for endring basert på for lite de to stemmer overens med hverandre.....	19
Figur 10. Matthews korrelasjonskoeffisient til alle prediksjonene, fordelt på arealtype.....	21
Figur 11. Matthews korrelasjonskoeffisient til prediksjonene for alle arealtyper, fordelt på kommune.....	22
Figur 12. Matthews korrelasjonskoeffisient til prediksjonene for alle arealtyper, fordelt på region.....	22
Figur 13. To kirkegårder, venstre i Nes, og høyre i Ullensaker. Ruter med røde kanter markerer predikert endring.....	25
Figur 14. Eksempel på boligfelt med trær i Nes. Legg merke til skyggene på bakken mellom trærne. Ruter med røde kanter markerer predikert endring.....	26
Figur 15. Et utsnitt av Oslo Lufthavn (Gardermoen) i Ullensaker. Rødfargede ruter markerer predikert endring. Det svarte området mangler prediksjoner.....	27

Figur 16. Eksempel på grønn fulldyrka jord. Ruter med røde kanter markerer predikert endring.....	27
Figur 17. Viser tre eksempler på ulik representasjon av innmarksbeite fra Nord-Aurdal, med ulik grad av tredekke. Ruter med røde kanter markerer predikert endring. Blå linje markerer grense for innmarksbeite.....	28
Figur 18. Fire hogstflater, de to øverste fra Eidskog, og de to nederste fra Nes. Ruter med røde kanter markerer predikert endring.....	29
Figur 19. Viser et eksempel på snaumark i fjellet i Nord-Aurdal. Ruter med røde kanter markerer predikert endring. Blå linje markerer grense for åpen fastmark.	30
Figur 20. Eksempel på golfbane klassifisert som åpen fastmark i Nes. Ruter med røde kanter markerer predikert endring.....	31
Figur 21. Til venstre flybilde av et gårdstun i Sola, i midten arealtypene før endring, og til høyre arealtypene etter endring.....	32
Figur 22. To eksempler på tresatt myr, venstre fra Nes, og høyre fra Sør-Odal. Ruter med røde kanter markerer predikert endring. Blå linje markerer grensen til myra.....	32
Figur 23. Eksempel på skygger fra trær i Nes. Ruter med røde kanter markerer predikert endring.....	33
Figur 24. Viser et eksempel på skygge i flybildene, fra en dal i Gjesdal kommune. Ruter med røde kanter markerer predikert endring.....	34
Figur 25. Det skraverte området i Sola kommune har verdien skog i AR5, men mesteparten av dette området holder ikke kravet for skog (minst 6 trær per dekar).....	36
Figur 26. Illustrasjon av U-Net-arkitekturen. De blå boksene representerer «feature maps», og de grå pilene representerer sammenkoblingene fra krympende til utvidende del. Blå piler representerer 3x3 konvolusjoner, og røde piler representerer 2x2 maks-oppsamling. Figur fra (Ronneberger et al., 2015).	42
Figur 27. Matthews korrelasjonskoeffisient til prediksjonene i Gjerdrum kommune fordelt på arealtype. Det var ingen observasjoner med arealtypen overflatedyrka mark i Gjerdrum, og MCC er derfor udefinert for denne arealtypen.....	52
Figur 28. Matthews korrelasjonskoeffisient til prediksjonene i Ullensaker kommune fordelt på arealtype. Det var ingen negative observasjoner av overflatedyrka mark i Ullensaker, og MCC er derfor udefinert for denne arealtypen.	52
Figur 29. Matthews korrelasjonskoeffisient til prediksjonene i Nes kommune fordelt på arealtype.....	53

Figur 30. Matthews korrelasjonskoeffisient til prediksjonene i Sør-Odal kommune fordelt på arealtype. Det var ingen negative observasjoner av overflatedyrka mark i Sør-Odal, og MCC er derfor udefinert for denne arealtypen.	53
Figur 31. Matthews korrelasjonskoeffisient til prediksjonene i Eidskog kommune fordelt på arealtype.	54
Figur 32. Matthews korrelasjonskoeffisient til prediksjonene i Nord-Aurdal kommune fordelt på arealtype.	54
Figur 33. Matthews korrelasjonskoeffisient til prediksjonene i Etnedal kommune fordelt på arealtype.	55
Figur 34. Matthews korrelasjonskoeffisient til prediksjonene i Gjesdal kommune fordelt på arealtype.	55
Figur 35. Matthews korrelasjonskoeffisient til prediksjonene i Sola kommune fordelt på arealtype.	56
Figur 36. Matthews korrelasjonskoeffisient til prediksjonene i Randaberg kommune fordelt på arealtype.	56
Figur 37. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen bebygd, fordelt på kommune.	57
Figur 38. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen fulldyrka jord, fordelt på kommune.	57
Figur 39. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen overflatedyrka jord, fordelt på kommune.	58
Figur 40. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen innmarksbeite, fordelt på kommune.	58
Figur 41. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen skog, fordelt på kommune.	59
Figur 42. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen åpen fastmark, fordelt på kommune.	59
Figur 43. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen myr, fordelt på kommune.	60
Figur 44. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen bebygd, fordelt på region.	61
Figur 45. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen fulldyrka jord, fordelt på region.	61
Figur 46. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen overflatedyrka jord, fordelt på region.	61

Figur 47. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen innmarksbeite, fordelt på region.	61
Figur 48. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen skog, fordelt på region.	62
Figur 49. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen åpen fastmark, fordelt på region.	62
Figur 50. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen myr, fordelt på region.	62

Tabelliste

Tabell 1. Viser kommunene med data som brukes til trening, og til å evaluere modellens prediksjoner. Viser også hvilken region kommunene hører til.	11
Tabell 2. Viser verdiene av egenskapen "Arealtype" i AR5, sammen med arealtypenes kodeverdi og definisjon. Fargene i tabellen viser hvilken farge arealtypene visualiseres med i kartet.	13
Tabell 3. Viser Norge i Bilder-prosjekter til data i kommunene.	14
Tabell 4. Viser antallet prediksjoner av hver arealtype innen hver kommune. Prosentene viser hvor stor andel arealtypen utgjør innen kommunen. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle kommunene samlet.	17
Tabell 5. Viser antallet prediksjoner av hver arealtype innen hver region. Prosentene viser hvor stor andel arealtypen utgjør innen regionen. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle regionene samlet.	18
Tabell 6. Viser Matthews korrelasjonskoeffisient for alle arealtyper og alle kommuner. Gjerdrum har ingen observasjoner med arealtypen overflatedyrka jord, og Ullensaker og Sør-Odal har ingen negative observasjoner med denne arealtypen. MCC for overflatedyrka jord er derfor udefinert for disse kommunene. En MCC på 0 visualiseres med hvit bakgrunn, mens negative MCC er røde og positive MCC er blå, med sterkere farge jo mer positiv/negativ MCC er. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle kommunene samlet.	23
Tabell 7. Viser Matthews korrelasjonskoeffisient for alle arealtyper og alle regioner. En MCC på 0 visualiseres med hvit bakgrunn, mens negative MCC er røde og positive MCC er blå, med	

sterkere farge jo mer positiv/negativ MCC er. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle regionene samlet. 24

Ordliste

Engelsk	Norsk
Adaptive random search	Adaptivt tilfeldig søk
Artificial neural network	Kunstig nevralt nettverk
Autoencoder	Autokoder
Baseline	Grunnlinje
Categorical cross-entropy	Kategorisk kryssentropi
Classification	Klassifisering
Confusion matrix	Forvirringsmatrise
Convolutional neural network	Konvolusjonelt nevralt nettverk
Data augmentation	Dataøkning
Decoder	Dekoder
Encoder	Koder
False negative	Falsk negativ
False positive	Falsk positiv
Grid search	Rutenettsøk
Hidden layer	Skjult lag
Hyperparameter optimization / hyperparameter tuning	Hyperparameteroptimalisering
Image segmentation	Bildesegmentering
Informedness	Informerthet
Input layer	Inputlag
Learning rate	Læringsrate
Linear correlation correction	Lineær korrelasjonskorreksjon
Loss function	Tapsfunksjon
Markedness	Markertthet
Matthew's correlation coefficient	Matthews korrelasjonskoeffisient
Output layer	Outputlag
Pooling	Samling
Precision	Presisjon
Random search	Tilfeldig søk
Recall (Sensitivity)	Sensitivitet
Shadow detection	Skyggedeteksjon
Shadow removal / shadow compensation	Skyggefjerning

Siamese convolutional variational	Siamesisk konvolusjonell variasjonsautokoder
Statistical classification	Statistisk klassifisering
Trimmed categorical cross-entropy	Trimmet kategorisk kryssentropi
True negative	Sann negativ
True positive	Sann positiv
Up-convolution	Oppkonvolusjon
Variational autoencoder	Variasjonsautokoder

Innhold

Sammendrag	i
Abstract	ii
Forord	iv
Figurliste	v
Tabelliste	viii
Ordliste	ix
1 Introduksjon	1
2 Teori	3
2.1 Fjernmåling	3
2.2 Kunstige nevrane nettverk	3
2.3 Forvirringsmatrise og mål på statistisk klassifiseringsnøyaktighet	7
3 Metode og data	11
3.1 Prosjektområde	11
3.2 Data	12
3.2.1 Arealressurskartet AR5	12
3.2.2 Flybilde-data fra Norge i bilder	14
3.2.3 Prediksjoner	15
3.3 Beskrivelse av maskinlæringsmodellen	18
3.4 Programvare	20
4 Resultater	21
5 Diskusjon	25
5.1 Diskusjon av resultater fordelt på arealtype	25
5.1.1 Bebyggd	25
5.1.2 Fulldyrka jord	27
5.1.3 Overflatedyrka jord	28
5.1.4 Innmarksbeite	28
5.1.5 Skog	29

5.1.6	Åpen fastmark	30
5.1.7	Myr	32
5.1.8	Skygge	33
5.2	Diskusjon av resultater fordelt på geografisk område	34
5.3	Feilkilder	35
5.4	Hva kan gjøres for å forbedre modellen?	37
5.4.1	Større variasjon i treningsdataene	37
5.4.2	Vekting av observasjoner	37
5.4.3	Skyggekorreksjon	38
5.4.4	Muligheter for bruk av andre fjernmålingsdata	39
5.4.5	Hyperparameteroptimalisering	40
5.4.6	Bildesegmentering	41
6	Konklusjon	44
7	Referanser	46
8	Appendiks	52
8.1	Resultater for kommuner fordelt på arealtype	52
8.2	Resultater for arealtype fordelt på kommuner	57
8.3	Resultater for arealtype fordelt på region	61
		61

1 Introduksjon

AR5 står for arealressurskart tilpasset målestokk 1:5000, og er et detaljert nasjonalt kartdatasett og klassifikasjonssystem for arealressurser (Ahlstrøm et al., 2019; Bjørkelo et al., 2013). AR5 er produsert og ajourholdt av Norsk institutt for bioøkonomi (NIBIO). Alt areal i AR5 klassifiseres etter egenskapene arealtype, treslag, skogbonitet, og grunnforhold.

God informasjon om arealressursene er en forutsetning for gode beslutninger innen arealplanlegging, landbruksforvaltning og landbruksnæringen. Disse dataene spiller også en stor rolle innen landskapsforskning. Blant annet er tjenesten gårdskart, som gir arealstatistikk for landbrukseiendommer, basert på data fra AR5 (NIBIO, 2021a).

Arealressursene endrer seg hele tiden. Jordarealer blir nedbygd, tidligere skogområder dyrkes, beitemark gror igjen. Derfor er det viktig med kontinuerlig ajourhold av arealressurskartet. I dag utføres kontinuerlig ajourhold av arealressurskartet av kommunene (NIBIO, 2021b). De endringene som kommunene ikke registrerer, ajourføres av NIBIO periodisk med 5-8 års mellomrom (NIBIO, 2021c). Dette gjøres ved at en person manuelt ser gjennom arealressurskartet for en kommune, og sammenlikner med oppdaterte flybilder for å oppdage endringer. Denne oppgaven er en tidkrevende og kostbar prosess. Det er i tillegg mulighet for menneskelige feil.

Hvis en maskinlæringsmodell kan identifisere områder hvor det er sannsynlig at arealressurskartet må oppdateres, vil det effektivisere arbeidet betraktelig. Det kan også være at maskinlæringsmodellen kan finne områder med endret arealtype som ikke ville blitt oppdaget i en manuell prosess, og dermed øke kvaliteten på ajourholdet. Derfor kan bruk av maskinlæring gi oss bedre og mer oppdaterte kart.

NIBIO har siden april 2020 jobbet med å utvikle og trene en maskinlæringsmodell, en siamesisk konvolusjonell variasjonsautokoder. Denne har som formål å identifisere områder med sannsynlig endret arealtype i arealressurskartet AR5, basert på flybilder. Å predikere endringer basert på flybilder er ingen triviell oppgave. Arealtypene er heterogene i utseendet fra luften. Noen endringer er det ikke mulig å predikere med kun flybilder. En annen utfordring er at grensen mellom to arealtyper ikke alltid er tydelig definert, og kan i mange tilfeller også være vanskelig å bestemme for et menneske.

Denne oppgaven tar i hovedsak for seg evaluering av modellens statistiske klassifiseringsnøyaktighet, både fordelt på areal typer og på forskjellige kommuner og regioner. Basert på resultatene trekkes det noen konklusjoner om anbefalte tiltak som kan forbedre modellens klassifiseringsnøyaktighet.

Helt konkret så vil oppgaven forsøke å svare på følgende spørsmål:

- Hvordan er den statistiske klassifiseringsnøyaktigheten til den siamesiske konvolusjonelle variasjonsautokoderen (modellen) som skal predikere endringer i areal type i arealressurskartet AR5?
- Hvordan varierer den statistiske klassifiseringsnøyaktigheten mellom forskjellige areal typer?
- Hvordan varierer den statistiske klassifiseringsnøyaktigheten mellom forskjellige geografiske områder?
- Hvilke situasjoner er spesielt utfordrende for modellen?
- Hva kan gjøres for å forbedre modellens statistiske klassifiseringsnøyaktighet?

2 Teori

2.1 Fjernmåling

Fjernmåling er å måle objekter på jordas overflate ved hjelp av data fra luftfartøy eller satellitter (Schowengerdt, 2006). Det kan være ulike typer data, for eksempel radar, laser, synlig lys, lyd, eller annet. Ofte danner de innsamlede dataene bilder.

Fjernmåling kan deles opp i to deler: aktiv og passiv fjernmåling. Aktiv fjernmåling er når det sendes ut et signal som reflekteres og detekteres av en sensor. Passiv fjernmåling er når det bare er sollyset reflektert fra jordas overflate som detekteres av sensoren.

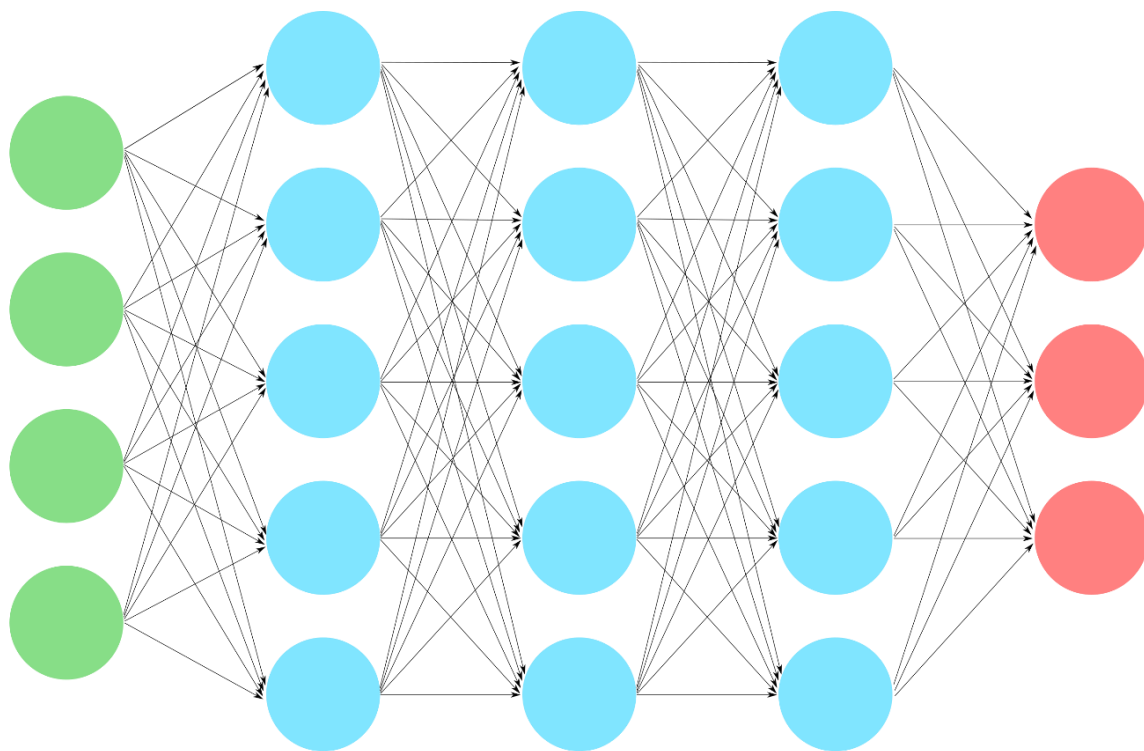
Innen fjernmåling er det ofte et mål å kategorisere objekter. Statistisk klassifisering (ofte omtalt som bare *klassifisering*) er en del av maskinlæring, som går ut på å klassifisere observasjoner i riktig kategori (Raschka & Mirjalili, 2017). Det kan for eksempel være å bestemme arealtypen til et område, eller å avgjøre om det har foregått en endring. Det sistnevnte kalles for endringsdeteksjon.

Endringsdeteksjon ble først utført ved manuelt konstruerte metoder på individuelle piksler, i hovedsak basert på spektralsignaturen til hver enkel piksel (Hussain et al., 2013). Deretter fulgte en utvikling av metoder som også tar hensyn til den romlige konteksten. I løpet av det siste tiåret har det vært en kraftig vekst av maskinlæring anvendt til endringsdeteksjon, spesielt kunstige nevralt nettverk (Audebert et al., 2017; Daudt et al., 2018a; Mou et al., 2019; Zhan et al., 2017).

2.2 Kunstige nevralt nettverk

Kunstige nevralt nettverk (ofte omtalt som bare *nevralt nettverk*) er en maskinlæringsmetode som er inspirert av biologisk nervevev fra hjernen (Raschka & Mirjalili, 2017). Et slikt nettverk består av flere lag av nevroner som er koblet sammen. Det første laget er inputlaget, det siste laget er outputlaget, og alle lagene mellom kalles skjulte lag. Nevronene i ett lag har bindinger til nevronene i neste lag, og hver binding har en vekt. Når nettverket mates med data, justeres vektene slik at outputen (prediksjonen) fra nettverket skal stemme best mulig med dataene.

Nettverket lærer seg sammenhengen mellom dataene og fasiten ved å prøve og feile. Ved bruk av en tapsfunksjon måler nettverket feilen den gjør, og justerer vektene for å minimere denne. Figur 1 viser en illustrasjon av et kunstig nevralt nettverk.



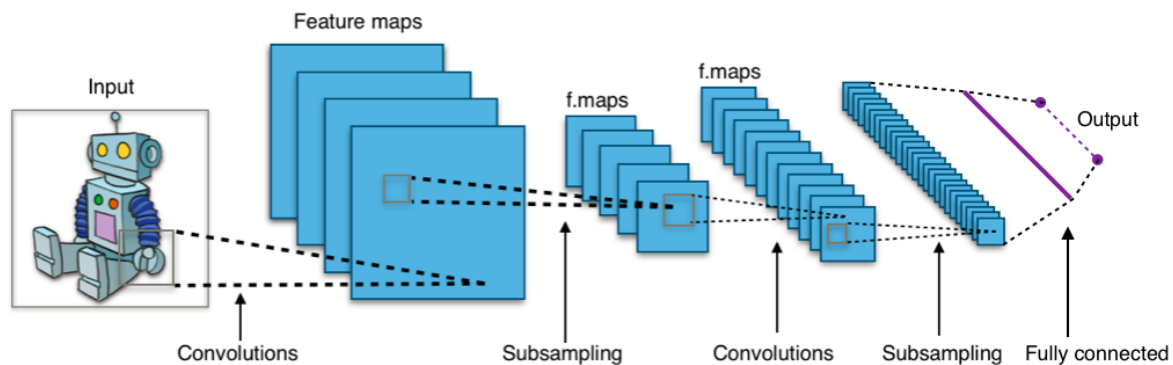
Inputlag

Skjulte lag

Outputlag

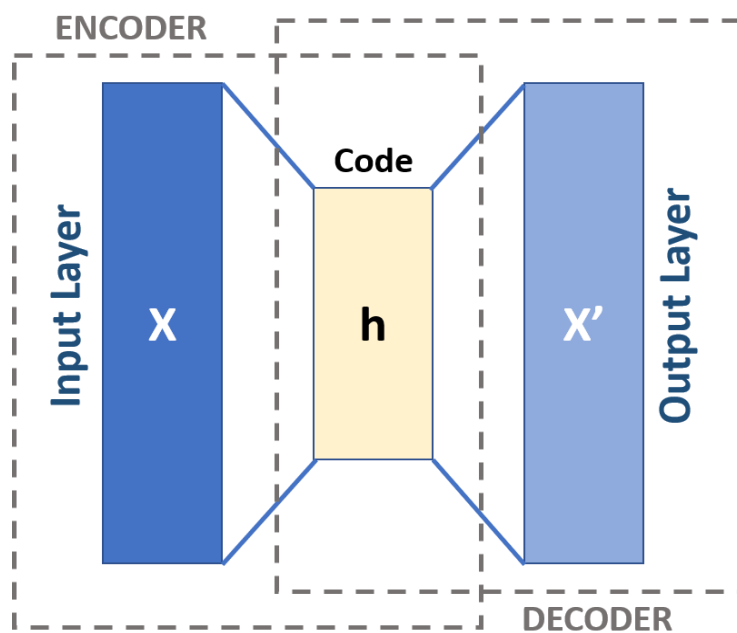
Figur 1. Figuren illustrerer et forenklet eksempel på et kunstig nevralt nettverk med tre skjulte lag. Antall lag og antall nevroner i hvert lag kan varieres.

Kunstige nevralt nettverk blir stadig brukt mer innen endringsdeteksjon, spesielt såkalte konvolusjonelle nevralt nettverk (Daudt et al., 2018a; Daudt et al., 2018b; Mou et al., 2019; Zhan et al., 2017). Konvolusjonelle nevralt nettverk er en type nevralt nettverk som er spesielt godt egnet til billedata (Raschka & Mirjalili, 2017). Dataene går gjennom en rekke konvolusjonsoperasjoner i nettverket. Det foregår ved at et konvolusjonsfilter går over hver piksel i bildet, og beregner en vektet sum basert på intensitetsverdiene til pikselens og nabopikslens intensitetsverdier. Når nettverket trenes, justeres vektene i konvolusjonsfiltrene slik at de fanger opp de relevante mønstrene i bildene. Mellom konvolusjonsoperasjonene utføres samlingsoperasjoner. Disse gjør at informasjonen fra konvolusjonene blir samlet opp, og den romlige informasjonen blir mindre. Til slutt ender man opp med ett tall, som i et klassifiseringsproblem indikerer hvilken kategori nettverket mener at bildet hører til. Figur 2 illustrerer strukturen til et konvolusjonelt nevralt nettverk. Figuren viser hvordan konvolusjonene fanger opp mønstre i “feature maps”, og samlingene sammenstiller mønstrene fra ulike deler av bildet.



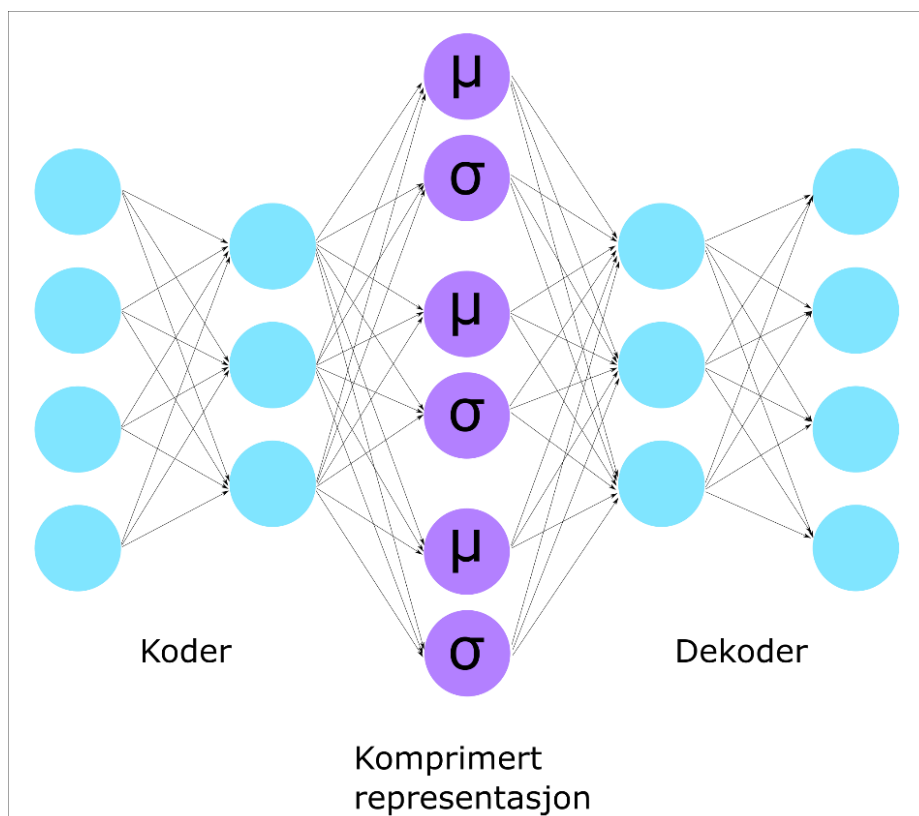
Figur 2. Illustrasjon av et typisk konvolusjonelt nevralt nettverk. Konvolusjoner og oppsamlinger fanger opp mønstre i "feature maps". Figur fra (Wikimedia Commons bruker Aphex34, 2015).

En autokoder er et kunstig nevralt nettverk som består av to deler: En koder, som komprimerer input-dataene, og en dekode, som dekomprimerer de komprimerte dataene. koderen trenes opp til å gradvis finne mer effektive måter å komprimere dataene, mens dekodeen gradvis blir bedre til å dekomprimere de komprimerte dataene. Autokoderen har flaskehals-form. Nettverket blir tynnere, med færre nevroner mot midten, som tvinger en enklere representasjon av dataene. Figur 3 illustrerer hvordan autokoderen er bygget opp av en koder og en dekode.



Figur 3. Illustrasjon av en autokoder. X er input-dataene, h er den komprimerte representasjonen, og X' er de dekomprimerte dataene. Figur fra (Massi, 2019).

En variasjonsautokoder er en type autokoder som kan brukes til å generere nye data. Den komprimerte representasjonen i en variasjonsautokoder er parameterne til flere sannsynlighetsfordelinger (An & Cho, 2015). Dette kan i utgangspunktet være hvilken som helst sannsynlighetsfordeling, men når vi har kontinuerlige data, for eksempel billedata, brukes gjerne normalfordelingen. Da består den komprimerte representasjonen av et sett med forventninger og standardavvik som representerer flere normalfordelinger (se Figur 4). Tilfeldige utvalg fra fordelingene sendes videre til dekoderen, som ender opp med en generert versjon av dataene basert på det tilfeldige utvalget.



Figur 4. Illustrasjon av en variasjonsautokoder. Den komprimerte representasjonen består av et sett med normalfordelinger (μ), og standardavvik (σ).

I tillegg til å generere nye data, kan en variasjonsautokoder brukes til endringsdeteksjon (An & Cho, 2015). Som nevnt er oppgaven til en variasjonsautokoder å lage en representasjon av fordelingen til dataene. Man kan da identifisere endringer ved å markere data som avviker betraktelig fra den representasjonen som autokoderen har lært seg.

En konvolusjonell variasjonsautokoder er en kombinasjon av en variasjonsautokoder og et konvolusjonelt nevralt nettverk (Chen et al., 2020). Denne typen variasjonsautokoder er derfor spesielt egnet til todimensjonale data, for eksempel bilder. Koderen ligner et vanlig konvolusjonelt nevralt nettverk, og består av en rekke konvolusjoner og samlinger. Dekoderen jobber omvendt av koderen, og gjør om den komprimerte representasjonen til et fullt bilde gjennom en rekke oppkonvolusjoner og oppsamlinger.

Siamesiske nevralt nettverk er to eller flere identiske nevralt nettverk som er koblet sammen ved nettverkenes output (Bromley et al., 1993). De brukes ofte til å bestemme i hvilken grad to observasjoner ligner på eller stemmer overens med hverandre.

2.3 Forvirringsmatrise og mål på statistisk klassifiseringsnøyaktighet

Det finnes mange ulike mål for å evaluere prediksjonene fra en klassifiseringsmodell. De aller fleste nøyaktighetsmål er basert på verdiene i en forvirringsmatrise. En forvirringsmatrise er en tabell som viser predikert klasse mot faktisk klasse (Chicco & Jurman, 2020; Luque et al., 2019). Figur 5 viser hvordan en forvirringsmatrise er strukturert.

		Predikert klasse	
		Positiv	Negativ
Faktisk klasse	Positiv	TP	FN
	Negativ	FP	TN

Figur 5. Viser oppsettet til en forvirringsmatrise. TP er sanne positive observasjoner, FN er falske negative observasjoner, FP er falske positive observasjoner, mens TN er sanne negative observasjoner.

I forvirringsmatrisa finnes fire verdier:

- Antallet faktisk positive observasjoner som også blir predikert positive, kalles sanne positive (TP).
- Antallet faktisk positive observasjoner som blir feilaktig predikert negative, kalles falske negative (FN).

- Antallet faktisk negative observasjoner som blir feilaktig predikert positive, kalles falske positive (FP).
- Antallet faktisk negative observasjoner som også blir predikert negative, kalles sanne negative (TN).

I et klassifiseringsproblem ønsker man naturligvis at den predikerte klassen skal stemme overens med den faktiske klassen for hver observasjon. Man vil altså maksimere TP og TN, og minimere FP og FN.

Den mest fullstendige måten å evaluere en maskinlæringsmodell på er å oppgi hele forvirringsmatrisa. Til tross for det er den ganske upraktisk å bruke, siden den ikke gir noen entydig verdi som er sammenliknbar. Derfor brukes gjerne verdiene fra forvirringsmatrisa til å beregne et klassifiseringsnøyaktighetsmål. Valget av nøyaktighetsmål avhenger i stor grad av om datasettet er balansert eller ikke, og på hvilken måte nøyaktighetsmålet representerer de ulike verdiene i forvirringsmatrisa (Luque et al., 2019).

Balanserte datasett er datasett der det er en relativt lik fordeling mellom positive og negative observasjoner. Nøyaktighet (engelsk: accuracy) er et mye brukt nøyaktighetsmål for balanserte datasett, og er definert som andelen prediksjoner som var riktige (Raschka & Mirjalili, 2017):

$$Nøyaktighet = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{antall riktige prediksjoner}}{\text{antall prediksjoner}} \quad (1)$$

Fordelen med nøyaktighet er at den er enkel å tolke. En nøyaktighet på 80 % betyr nettopp det at 80 % av prediksjonene var riktige.

Ubalanserte datasett er datasett hvor fordelingen mellom positive og negative observasjoner er ulik, noe som naturligvis ofte er tilfelle innen endringsdeteksjon. Da er det gjerne få områder med endringer (positive observasjoner), mens de fleste områder er uten endringer (negative observasjoner). I slike situasjoner er det veldig viktig å bruke et nøyaktighetsmål som tar høyde for ubalansen i datasettet.

Nøyaktighet gir et veldig dårlig bilde på hvor flink modellen er til å predikere i tilfeller med ubalanserte data. Dersom man har et datasett hvor 99 % av observasjonene er

negative, vil man oppnå 99 % nøyaktighet ved å bare predikere alle observasjonene negative. Dette er likevel sjelden et ønsket resultat, fordi man oftest også ønsker å ta hensyn til klassifiseringsnøyaktigheten til minoritetsklassen.

I stedet for nøyaktighet er presisjon og sensitivitet (engelsk: precision & recall) vanlige mål på klassifiseringsnøyaktighet i tilfeller med ubalanserte datasett. Presisjon måler kvaliteten på prediksjonene, og er definert som andelen positive prediksjoner som også var faktisk positive (Raschka & Mirjalili, 2017):

$$\textit{Presisjon} = \frac{TP}{TP + FP} \quad (2)$$

Sensitivitet måler hvor mange positive observasjoner som blir identifisert av modellen, og er definert som andelen faktisk positive som ble predikert positive (Raschka & Mirjalili, 2017):

$$\textit{Sensitivitet} = \frac{TP}{TP + FN} \quad (3)$$

I praksis brukes gjerne F1-verdien, som er det harmoniske gjennomsnittet av presisjon og sensitivitet (Raschka & Mirjalili, 2017):

$$F1 = 2 \times \frac{\textit{Presisjon} \times \textit{Sensitivitet}}{\textit{Presisjon} + \textit{Sensitivitet}} \quad (4)$$

Det er verdt å legge merke til at TN (sanne negative) ikke inngår i F1-verdien. Det gjør at F1-verdien er mest egnet i situasjoner der negative observasjoner ikke er like viktige som de positive. F1-verdien påvirkes også av andelen positive observasjoner i datasettet. Det betyr at F1-verdien til forskjellige datasett med ulik balanse ikke kan direkte sammenlignes.

Powers (2007), Chicco og Jurman (2020), og Yao og Shepperd (2020) argumenterer for at man kan unngå disse problemene til F1-verdien ved å bruke informerthet og markerthet i stedet for presisjon og sensitivitet, og Matthews korrelasjonskoeffisient (MCC) i stedet for F1-verdien. Disse nøyaktighetsmålene tar nemlig hensyn til både positive og negative observasjoner:

$$\textit{Informerthet} = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1 \quad (5)$$

$$\textit{Markerthet} = \frac{TP}{TP + FP} + \frac{TN}{TN + FN} - 1 \quad (6)$$

Både informerthet og markerthet går fra -1 til 1, hvor 0 er på nivå med tilfeldig gjetting. Det geometriske gjennomsnittet av informerthet og markerthet kalles Matthews korrelasjonskoeffisient (Chicco & Jurman, 2020; Powers, 2007):

$$\begin{aligned}
 MCC &= \pm \sqrt{\frac{\text{Informerthet} \times \text{Markerthet}}{TP \times TN - FP \times FN}} \\
 &= \frac{\pm \sqrt{\text{Informerthet} \times \text{Markerthet}}}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
 \end{aligned}
 \tag{7}$$

Matthews korrelasjonskoeffisient (MCC) er den binære korrelasjonen mellom de faktiske klassene og de predikerte klassene. Den tar på seg en verdi mellom -1 og 1, der -1 betyr at alle observasjonene ble feilaktig predikert, 1 betyr at alle prediksjonene ble riktig predikert, og 0 er på nivå med tilfeldig gjetting.

I motsetning til F1-verdien, så tar MCC hensyn til alle verdiene i forvirringsmatrisa, også sanne negative (TN). Den tar like mye hensyn til negative som positive observasjoner. Det gjør at MCC er uavhengig av balanse i datasettet. Dermed kan MCC til datasett med ulik balanse, for eksempel arealtyper innenfor en kommune, sammenliknes direkte. Grunnlinjeverdien til Matthews korrelasjonskoeffisient er 0, uavhengig av fordelingen i datasettet. Dette gjør også at man umiddelbart kan avgjøre om et klassifiseringsresultat er bedre eller verre enn tilfeldig gjetting, ved å se om MCC er positiv eller negativ.

Cohens Kappa er et klassifiseringsnøyaktighetsmål som har blitt brukt en del innen fjernmåling. Delgado og Tibau (2019) viser at Cohens Kappa er uegnet til bruk ved ubalanserte datasett, og at MCC derfor er et mer robust alternativ.

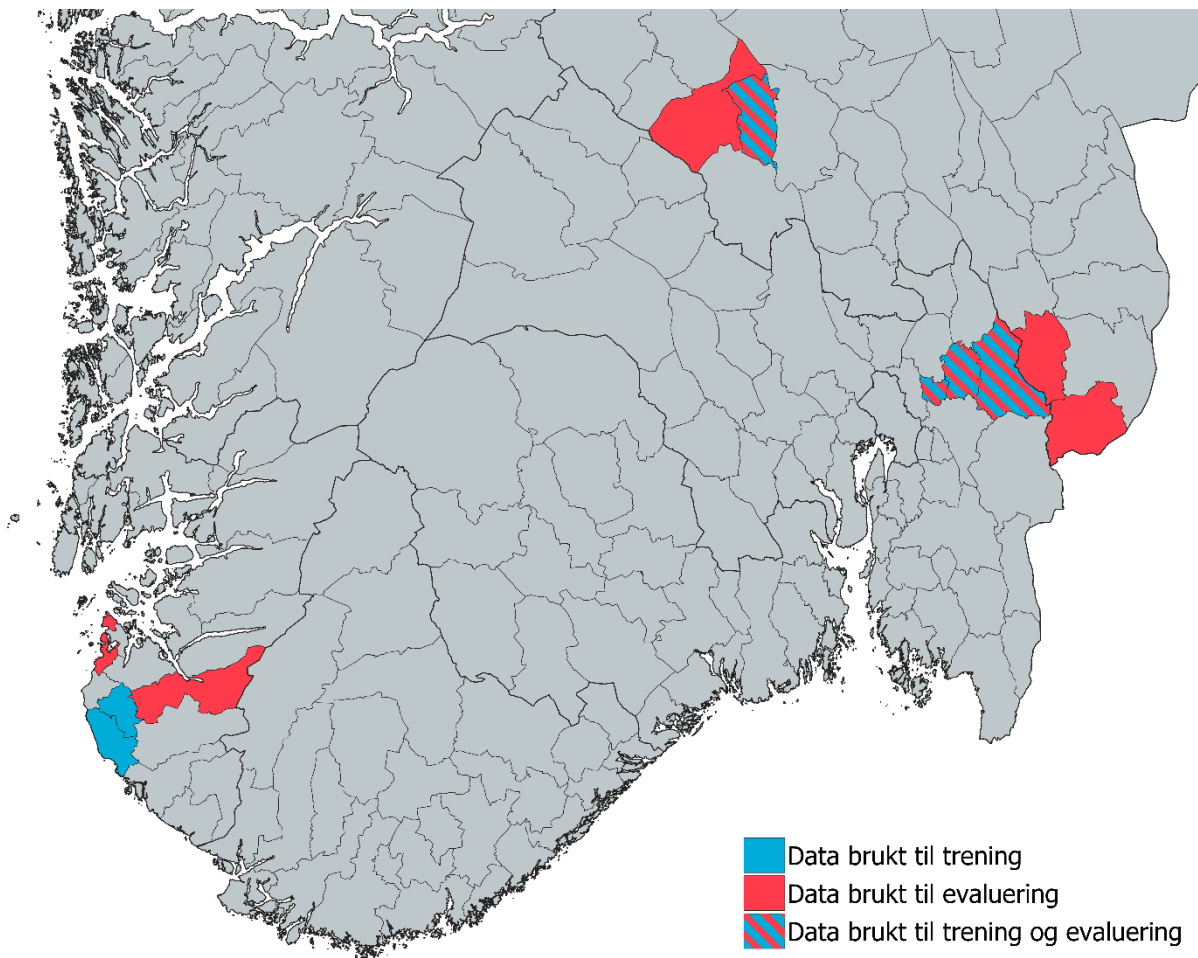
3 Metode og data

3.1 Prosjektområde

Dataene som er brukt til å trene og evaluere maskinlæringsmodellen kommer fra flere kommuner. Kommunene er gruppert i tre regioner: Romerike/Glåmdalen, Valdres og Jæren. Tabell 1 viser en oversikt over kommunene, og om data fra kommunen har blitt brukt til trening eller evaluering av modellen, eller begge. Tabellen viser også hvilket geografisk område hver kommune hører til. Figur 6 viser trenings- og evalueringskommunene i et kart.

Tabell 1. Viser kommunene med data som brukes til trening, og til å evaluere modellens prediksjoner. Viser også hvilken region kommunene hører til.

Kommune	Trening eller validering?	Region
Gjerdrum	Trening og evaluering	Romerike/Glåmdalen
Ullensaker	Trening og evaluering	
Nes	Trening og evaluering	
Sør-Odal	Evaluering	
Eidskog	Evaluering	
Nord-Aurdal	Evaluering	Valdres
Etnedal	Trening og evaluering	
Hå	Trening	Jæren
Time	Trening	
Gjesdal	Evaluering	
Sola	Evaluering	
Randaberg	Evaluering	



Figur 6. Kartet over sørlige Norge viser hvilke kommuner som har blitt brukt til trening, og hvilke som ble brukt til evaluering. I sørvest har vi Jæren-kommunene, lengst nord er Valdres-kommunene, og klyngen lengst øst er Romerike/Glåmdalen-kommunene.

3.2 Data

3.2.1 Arealressurskartet AR5

Hovedinndelingen i arealressurskartet AR5 er arealtype. Arealtypen bestemmes etter kriterier for vegetasjon og kulturpåvirkning (Bjørkelo et al., 2013). Det er den faktiske tilstanden, ikke bruken av arealene som avgjør arealtypen. Tabell 2 viser verdier av arealtype (Ahlstrøm et al., 2019):

Tabell 2. Viser verdiene av egenskapen "Arealtype" i AR5, sammen med arealtypenes kodeverdi og definisjon. Fargene i tabellen viser hvilken farge arealtypene visualiseres med i kartet.

Egenskapsverdi	Kodeverdi	Definisjon
Bebyggd	11	Areal som er utbygd eller i betydelig grad opparbeidet, samt tilstøtende arealer som i funksjon er nært knyttet til bebyggelsen.
Samferdsel	12	Areal som brukes til samferdsel.
Fulldyrka jord	21	Jordbruksareal som er dyrka til vanlig pløyedybde, og kan benyttes til åkervekster eller til eng, og som kan fornyes ved pløying.
Overflatedyrka jord	22	Jordbruksareal som for det meste er rydda og jevna i overflata, slik at maskinell høsting er mulig.
Innmarksbeite	23	Jordbruksareal som kan benyttes som beite, men som ikke kan høstes maskinelt. Minst 50 % av arealet skal være dekt av godkjente grasarter eller beitetålende urter.
Skog	30	Areal med minst 6 trær per dekar som er eller kan bli 5 meter høye, og disse bør være jevnt fordelt på arealet.
Åpen fastmark	50	Fastmark som ikke er jordbruksareal, skog, bebyggd eller samferdsel.
Myr	60	Areal med myrvegetasjon og minst 30 cm tykt torvlag.
Snøisbre	70	Blanding av isbre og snø som ikke smelter i løpet av sommeren.
Ferskvann	81	Ferskvann
Hav	82	Hav
Ikke kartlagt	99	Areal som har ukjent beskaffenhet.

Fargene i tabellen viser hvilken farge arealtypene visualiseres med i kartet. De samme fargene blir brukt til å representere arealtypene i kart og diagrammer i denne oppgaven.

Arealtypene samferdsel, ferskvann og hav er arealtyper som ikke blir bestemt fra flybilder, men fra andre offentlige kartdatasett. I kommunene med data i denne oppgaven var det ingen områder med arealtypen snøisbre, så denne arealtypen blir også sett bort fra. Områder uten arealtype, dvs. med arealtype "ikke kartlagt", ble også fjernet.

AR5-data tilhørende hver kommune ble hentet fra tidspunktet før flybildene ble tatt. I tillegg ble det brukt et endringsdatasett, som viste hvor det hadde blitt registrert endringer basert på de nye flybildene. Disse endringene er å betrakte som fasit til positive observasjoner.

3.2.2 Flybilde-data fra Norge i bilder

Flybilder fra Norge i Bilder ble sammen med AR5-dataene brukt til å trene modellen og for å predikere. Disse flybildene består av rød-grønn-blå (RGB)-bilder. Tabell 3 viser hvilke Norge i Bilder-prosjekter ortofotoene fra hver kommune kommer fra.

Tabell 3. Viser Norge i Bilder-prosjekter til data i kommunene.

Kommune	Prosjekt
Gjerdrum	Oslo-Østlandet 2016
Ullensaker	
Nes	
Sør-Odal	
Eidskog	
Nord-Aurdal	Østlandet 2016
Etnedal	
Gjesdal	Rogaland 2019
Sola	
Randaberg	

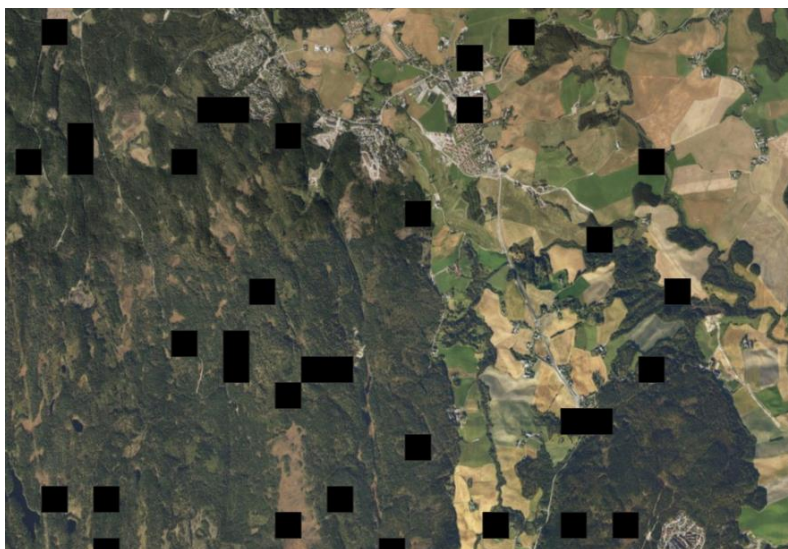
De tre prosjektene er en del av nasjonalt program for omløpsfotografering. Alle flybildene er 24-bits RGB-bilder, med pikselstørrelse (bakkeoppløsning, GSD) på 0,25 meter (Blom Norway, 2020; Terratec, 2016a; Terratec, 2016b).

Flybildene ble delt inn i 50x50 meter ruter før trening og prediksjon, og AR5-dataene ble delt inn i samme ruter. Hver av disse rutene er å anse som én observasjon.

3.2.3 Prediksjoner

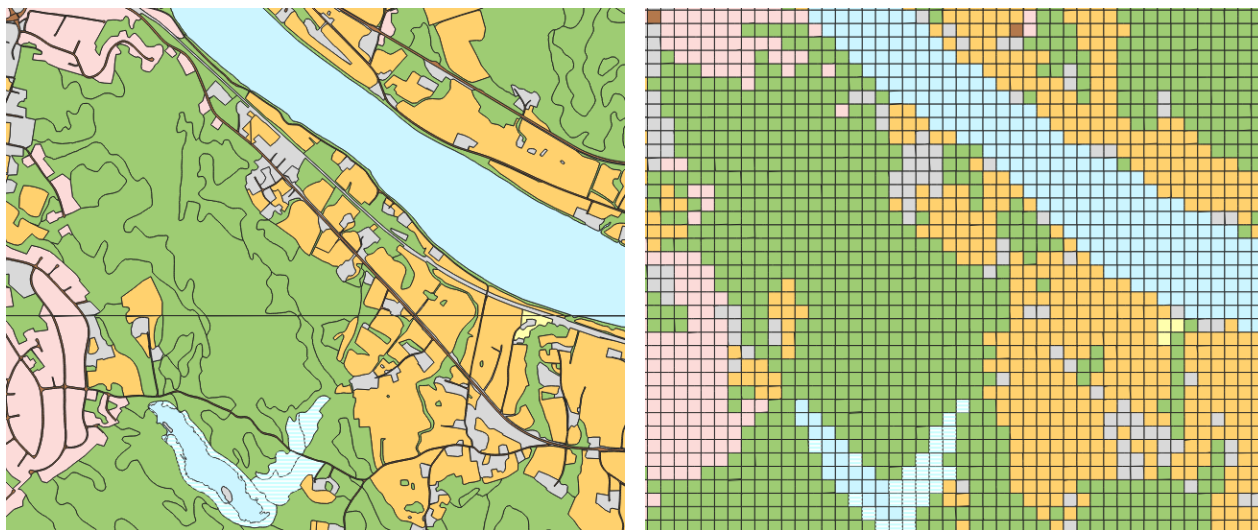
Prediksjonene består av ruter på 50x50 meter, tilsvarende rutene som ble brukt til trening. Hver av disse har en sannsynlighet for endret arealtype, bestemt av modellen, som ligger mellom 0 og 1. Ruter med sannsynlighet $\geq 0,5$ regnes som predikert endring, mens ruter med sannsynlighet $< 0,5$ regnes som predikert ingen endring.

Noen områder spredd tilfeldig rundt har manglende prediksjoner. Disse områdene er på størrelse 5x5 ruter, altså 250x250 meter. Det er ikke klart hva som kan være årsaken til disse områdene med manglende prediksjoner. De er uansett utelatt fra datasettet, og ikke tatt med i analysen i denne oppgaven. Figur 7 viser et eksempel på dette fenomenet fra Gjerdrum.



Figur 7. Figuren viser hvordan kvadratiske områder med manglende prediksjoner (markert i svart) opptrer tilfeldig spredt rundt i enkelte kommuner. Eksempel fra Gjerdrum.

Prediksjonene kommer fra modellen i rasterformat, i filformatet TIFF. De ble gjort om til vektorform for å kunne kobles til AR5-dataene. Hver rute i prediksjonene ble da tilordnet arealtypen som dekket mest av ruta. Figur 8 viser et utsnitt av AR5-kartet før og etter tilordning av arealtype til rutenettet. Ruter med ikke-relevant arealtype (se avsnitt 3.2.1) ble fjernet fra datasettet.



Figur 8. Kartet til venstre viser et utsnitt av arealtypekartet, og til høyre er arealtypene som har blitt tilordnet rutenettet til prediksjonene.

Deretter ble prediksjonene koblet til endringsdatasettet. Hver rute med område med endring større enn 50 m^2 ble ansett som rute med endring. Til slutt ble klassifiseringsnøyaktigheten beregnet for alle areal typer i alle kommunene.

Tabell 4 viser antallet observasjoner i prediksjonsdatasettet fordelt på kommune og arealtype. Tabell 5 viser samme informasjon fordelt på region i stedet for kommune.

Tabell 4. Viser antallet prediksjoner av hver arealtype innen hver kommune. Prosentene viser hvor stor andel arealtypen utgjør innen kommunen. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle kommunene samlet.

Kommune	Bebyggd	Fulldyrka jord	Overflate - dyrka jord	Innmarks- beite	Skog	Åpen fastmark	Myr	Totalt
Gjerdrum	565 (1,8 %)	10235 (32,0 %)	0 (0 %)	585 (1,8 %)	18135 (56,7 %)	769 (2,4 %)	1688 (5,3 %)	31977 (100 %)
Ullensaker	5710 (6,3 %)	30369 (33,7 %)	9 (0,01 %)	1990 (2,2 %)	44397 (49,2 %)	6533 (7,2 %)	1212 (1,3 %)	90220 (100 %)
Nes	2714 (1,2 %)	49465 (22,7 %)	10 (0,01 %)	2109 (1,0 %)	146768 (67,4 %)	3960 (1,8 %)	12644 (5,8 %)	217670 (100 %)
Sør-Odal	1074 (0,6 %)	26193 (13,9 %)	10 (0,01 %)	297 (0,2 %)	145827 (77,2 %)	3139 (1,7 %)	12361 (6,5 %)	188901 (100 %)
Eidskog	1473 (0,6 %)	15548 (6,5 %)	40 (0,02 %)	1012 (0,4 %)	199072 (83,1 %)	2557 (1,1 %)	19878 (8,3 %)	239580 (100 %)
Nord- Aurdal	2032 (0,6 %)	11836 (3,5 %)	516 (0,2 %)	3577 (1,1 %)	192649 (57,5 %)	74786 (22,3 %)	49686 (14,8 %)	335082 (100 %)
Etnedal	136 (0,1 %)	4862 (3,1 %)	369 (0,2 %)	2080 (1,3 %)	121276 (77,0 %)	11401 (7,2 %)	17397 (11,0 %)	157521 (100 %)
Gjesdal	1304 (0,7 %)	7345 (3,8 %)	100 (0,1 %)	15942 (8,2 %)	40622 (20,9 %)	125074 (64,3 %)	3918 (2,0 %)	194305 (100 %)
Sola	4003 (14,6 %)	14061 (51,3 %)	66 (0,2 %)	3005 (11,0 %)	1340 (4,9 %)	4663 (17,0 %)	276 (1,0 %)	27414 (100 %)
Randaberg	923 (9,6 %)	5673 (59,2 %)	117 (1,22 %)	767 (8,0 %)	513 (5,4 %)	1505 (15,7 %)	77 (0,8 %)	9575 (100 %)
Samlet	19934 (1,3 %)	175587 (11,8 %)	1237 (0,1 %)	31364 (2,1 %)	910599 (61,0 %)	234387 (15,7 %)	119137 (8,0 %)	1492245 (100 %)

Tabell 5. Viser antallet prediksjoner av hver arealtype innen hver region. Prosentene viser hvor stor andel arealtypen utgjør innen regionen. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle regionene samlet.

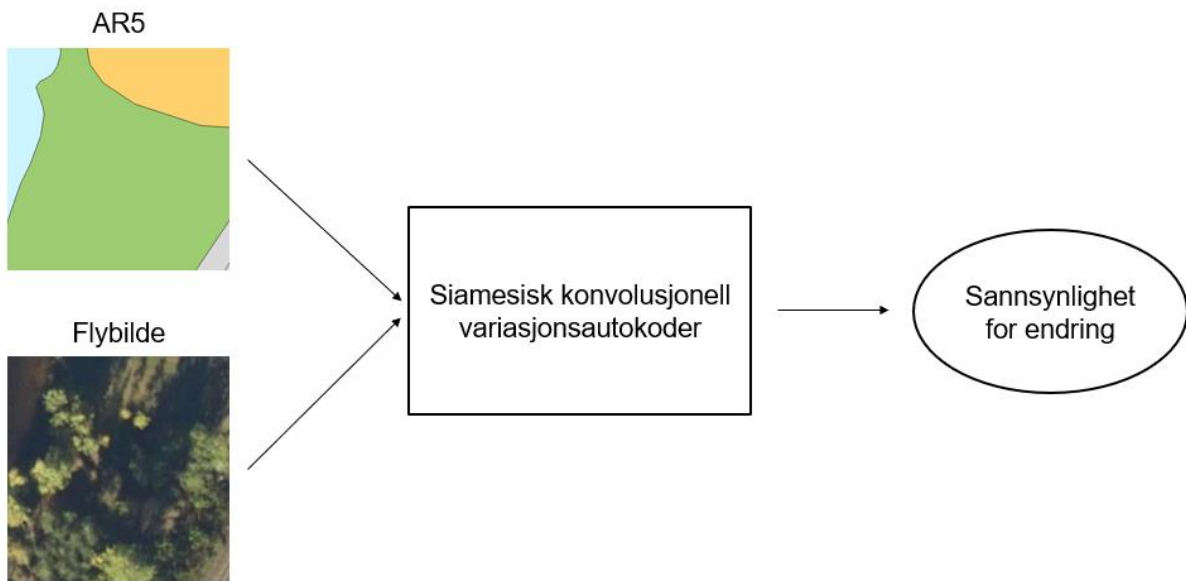
Region	Bebyggd	Fulldyrka jord	Overflate - dyrka jord	Innmarks-beite	Skog	Åpen fastmark	Myr	Totalt
Romerike/ Glåmdalen	11536 (1,5 %)	131810 (17,2 %)	69 (0,01 %)	5993 (0,8 %)	554199 (72,1 %)	16958 (2,2 %)	47783 (6,2 %)	768348 (100 %)
Valdres	2168 (0,4 %)	16698 (3,4 %)	885 (0,2 %)	5657 (1,1 %)	313925 (63,7 %)	86187 (17,5 %)	67083 (13,6 %)	492603 (100 %)
Jæren	6230 (2,7 %)	27079 (11,7 %)	283 (0,1 %)	19714 (8,5 %)	42475 (18,4 %)	131242 (56,7 %)	4271 (1,8 %)	231294 (100 %)
Samlet	19934 (1,3 %)	175587 (11,8 %)	1237 (0,1 %)	31364 (2,1 %)	910599 (61,0 %)	234387 (15,7 %)	119137 (8,0 %)	1492245 (100 %)

Siden oppgaven blant annet har som formål å sammenlikne modellens klassifiseringsnøyaktighet mellom ulike areal typer og geografiske områder, er det nødvendig å ta i bruk et klassifiseringsnøyaktighetsmål som kan direkte sammenliknes mellom forskjellige datasett uavhengig av datasettenes balanse. Matthews korrelasjonskoeffisient oppfyller disse kravene (forklart i detalj i avsnitt 2.3), og derfor brukes den til evaluering av modellens prediksjoner i denne oppgaven.

All kode skrevet i forbindelse med oppgaven finnes i følgende github-repository:
<https://github.com/Nicomunster/ending-ar5>

3.3 Beskrivelse av maskinlæringsmodellen

Maskinlæringsmodellen som er brukt er en siamesisk konvolusjonell variasjonsautokoder. Den tar et par av observasjoner som input: arealkartet til en rute, og flybildet til den samme ruta. Modellen trenes dermed til å lære sammenhengen mellom arealtypene fra AR5 og flybildene fra Norge i Bilder. Outputen fra nettverket er sannsynlighet for endring, som beregnes ut fra hvor mye modellen mener at de to inputene ikke stemmer overens med hverandre. Modellens kode ble utviklet av bedriften Epigram, mens trening og prediksjon ble utført av NIBIO.



Figur 9. Viser forenklet hvordan modellen tar AR5-data og flybilledata som input, og beregner en sannsynlighet for endring basert på for lite de to stemmer overens med hverandre.

De konvolusjonelle nettverkene er basert på ResNet50V2-arkitekturen (He et al., 2016). Denne består av 50 konvolusjonelle lag. I tillegg har det blitt lagt til ytterligere 3 konvolusjonslag og ett fullkoblet lag til slutt. Tapsfunksjonen som er brukt er kategorisk kryssentropi. Modellen ble trent i 300 epoker, med 100 steg i hver epoke. Under trening av modellen har observasjonene blitt vektet etter arealtypen, slik at modellen fokuserer på å lære like mye av hver arealtype, uavhengig arealtypenes fordeling i datasettet.

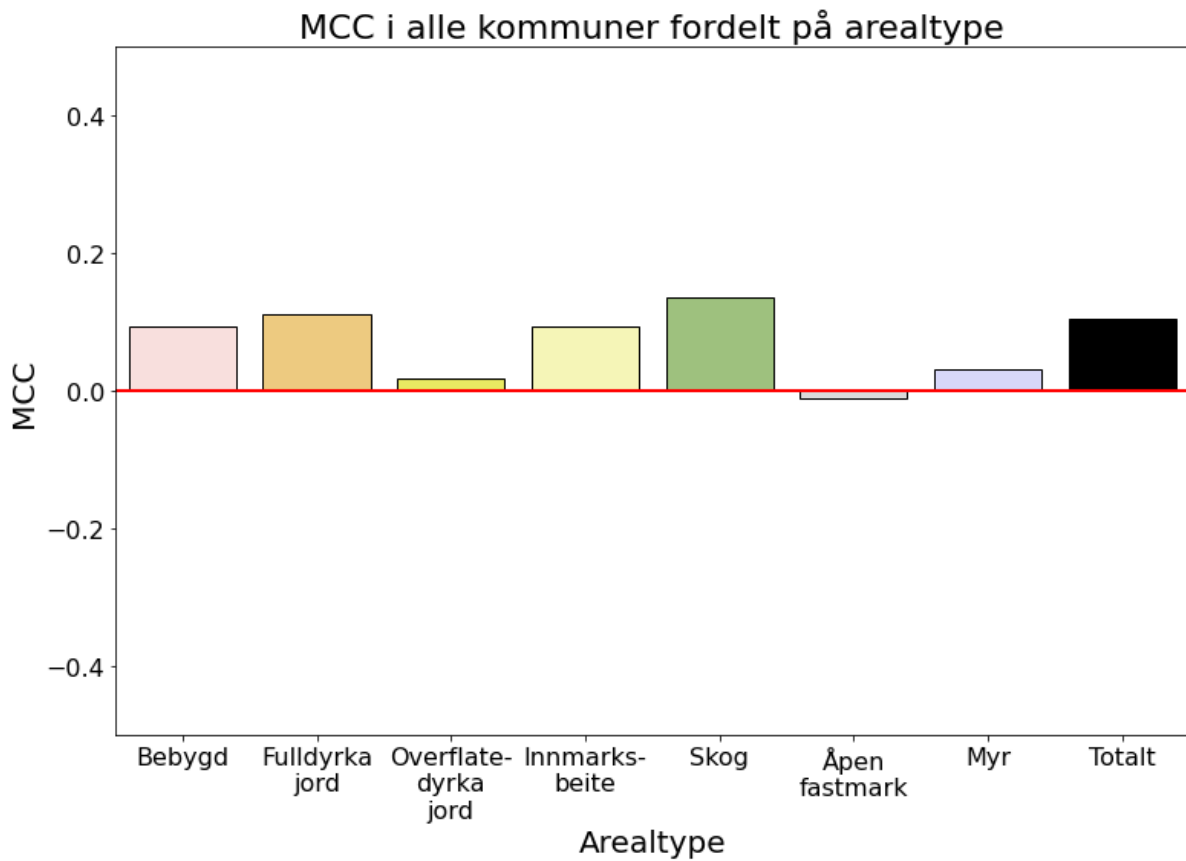
Det har også blitt utført dataøkning. Det vil si at modifiserte versjoner av observasjonene har blitt lagt til i treningen for å øke datamengden. Dette har som formål å hindre at modellen lærer seg sammenhenger som finnes i det opprinnelige som ikke er relevante for klassifiseringen, for eksempel skyggeretningen (Stivaktakis et al., 2019). Skyggeretningen er ikke relevant for klassifiseringen, og man kan da innføre rotasjon av bilder som dataøkning, for at modellen skal lære fra bilder med alle skyggeretninger. Dataøkning kan dermed bidra til å øke generaliseringsevnen til modellen. Dataøkningsoveroperasjonene som har blitt utført er rotasjon, skalering, oppløsningsendring, forskyvning, speiling, lysstyrkeendring og hvit støy.

3.4 Programvare

Til geografisk databehandling ble GIS-programvaren QGIS 3.10.14 og ArcGIS Pro 2.3.1 benyttet. Python 3.8 ble benyttet til annen databehandling, preprosessering og beregning av nøyaktighetsmål. Mer spesifikt ble biblioteket pandas brukt til å lagre og behandle data i DataFrames, scikit-learn til å beregne verdiene i forvirringsmatrisa, og matplotlib til å plote klassifiseringsnøyaktigheten.

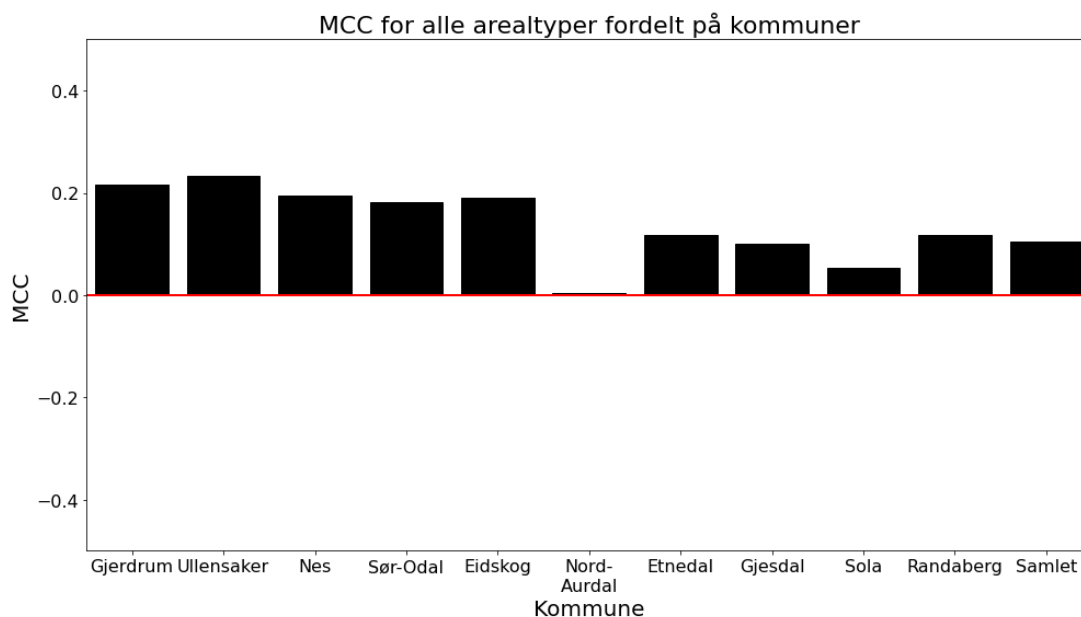
4 Resultater

Figur 10 viser Matthews korrelasjonskoeffisient til prediksjonene fra alle kommunene fordelt på arealtype. Tilsvarende diagram for hver kommune finnes i appendiks 8.1.



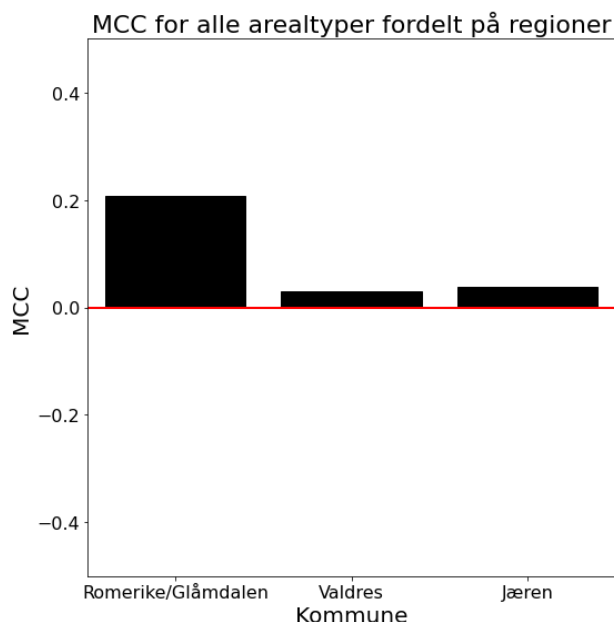
Figur 10. Matthews korrelasjonskoeffisient til alle prediksjonene, fordelt på arealtype.

Figur 11 viser Matthews korrelasjonskoeffisient til alle prediksjonene fra alle areal typer, fordelt på kommune. Tilsvarende diagram for hver areal type finnes i appendiks 8.2.



Figur 11. Matthews korrelasjonskoeffisient til prediksjonene for alle areal typer, fordelt på kommune.

Figur 12 viser Matthews korrelasjonskoeffisient til alle prediksjonene av alle areal typer, fordelt på region. Tilsvarende diagram for hver areal type finnes i appendiks 8.3.



Figur 12. Matthews korrelasjonskoeffisient til prediksjonene for alle areal typer, fordelt på region.

Tabell 6 viser Matthews korrelasjonskoeffisient fordelt på arealtyper og kommuner.

Tabell 7 er samme type tabell, men samlet i regionene i stedet for individuelle kommuner.

Tabell 6. Viser Matthews korrelasjonskoeffisient for alle arealtyper og alle kommuner. Gjerdrum har ingen observasjoner med arealtypen overflatedyrka jord, og Ullensaker og Sør-Odal har ingen negative observasjoner med denne arealtypen. MCC for overflatedyrka jord er derfor udefinert for disse kommunene. En MCC på 0 visualiseres med hvit bakgrunn, mens negative MCC er røde og positive MCC er blå, med sterkere farge jo mer positiv/negativ MCC er. "Totalt" representerer alle arealtypene samlet, mens "Samlet" representerer alle kommunene samlet.

Kommune	Bebyggd	Fulldyrka jord	Overflatedyrka jord	Innmarksbeite	Skog	Åpen fastmark	Myr	Totalt
Gjerdrum	0,041	0,105	-	-0,118	0,144	0,081	0,034	0,217
Ullensaker	0,070	0,156	-	-0,023	0,145	-0,175	-0,041	0,233
Nes	0,112	0,136	-0,167	-0,090	0,132	-0,153	0,031	0,194
Sør-Odal	0,146	0,106	-	-0,178	0,16	-0,100	0,004	0,182
Eidskog	0,157	0,186	-0,129	-0,091	0,132	-0,046	0,037	0,19
Nord-Aurdal	0,188	0,051	0,011	-0,098	0,012	-0,289	0,027	0,003
Etnedal	0,013	0,144	-0,040	-0,065	0,051	-0,342	0,077	0,116
Gjesdal	0,055	0,062	0,177	0,002	0,097	0,031	0,135	0,101
Sola	0,129	0,057	0,295	0,063	0,209	-0,119	0,162	0,054
Randaberg	0,280	0,117	0,022	0,059	0,256	-0,101	-0,174	0,117
Samlet	0,094	0,111	0,018	0,094	0,137	-0,010	0,031	0,104

Tabell 7. Viser Matthews korrelasjonskoeffisient for alle areal typer og alle regioner. En MCC på 0 visualiseres med hvit bakgrunn, mens negative MCC er røde og positive MCC er blå, med sterkere farge jo mer positiv/negativ MCC er. "Totalt" representerer alle areal typene samlet, mens "Samlet" representerer alle regionene samlet.

Kommune	Bebyggd	Fulldyrka jord	Overflate- dyrka jord	Innmarksbeite	Skog	Åpen fastmark	Myr	Totalt
Romerike/ Glåmdalen	0,079	0,142	-0,103	-0,082	0,154	-0,205	0,009	0,208
Valdres	0,177	0,082	0,010	-0,065	0,025	-0,301	0,041	0,030
Jæren	0,123	0,029	0,050	-0,001	0,097	0,008	0,127	0,038
Samlet	0,094	0,111	0,018	0,094	0,137	-0,010	0,031	0,104

5 Diskusjon

MCC til modellens samlede prediksjoner ligger på 0,104, altså en svak positiv korrelasjon mellom predikerte endringer og faktiske endringer. Det betyr at modellen har lært seg enkelte sammenhenger som er nyttige for klassifiseringen. Dette er likevel bare litt over nivå med tilfeldig gjetting. Modellen når derfor ikke opp til nøyaktigheten som kan forventes av en endringsdeteksjonsmodell som faktisk kan tas i bruk.

Modellens MCC spriker fra -0,342 til 0,295 avhengig av arealtype og kommune. Det er tydelig at det er forhold i dataene som påvirker modellens evne til å oppdage endringer og samtidig unngå falske positive.

5.1 Diskusjon av resultater fordelt på arealtype

5.1.1 Bebygd

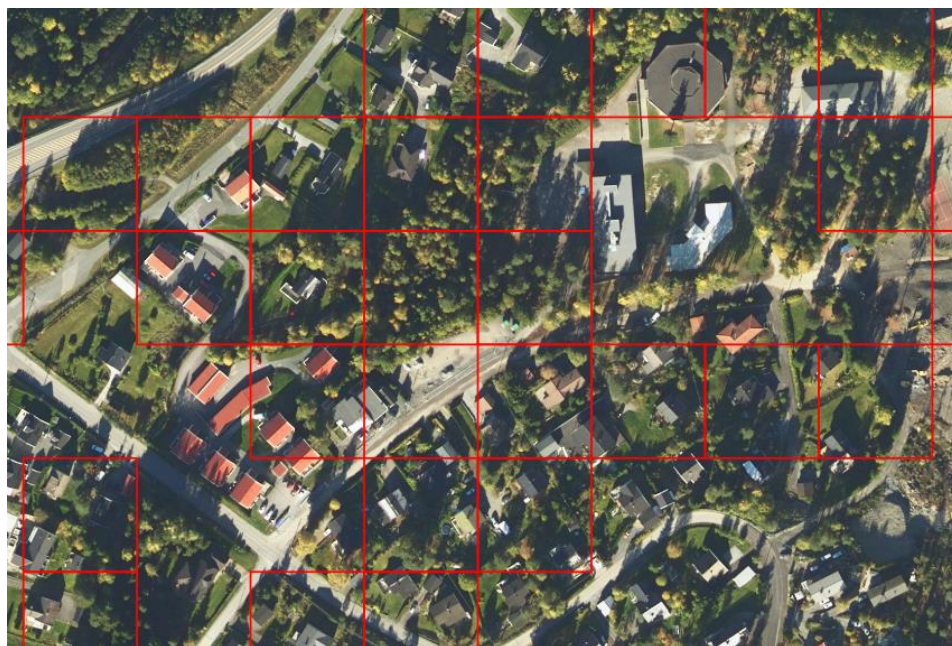
Bebygd har positiv MCC i alle kommunene, og samlet $MCC = 0,094$. Med en positiv MCC i alle kommuner betyr det at modellen har lært seg nyttige mønstre ved utseendet til bebygde områder, men det er likevel flere konkrete situasjoner der modellen har store utfordringer.

Kirkegårder er oftest grønne og dominert av gress. De klassifiseres likevel som bebygd i AR5. De er derfor lite typiske for arealtypen, og modellen er tilbøyelig til å predikere endring. Figur 13 viser et eksempel på dette.



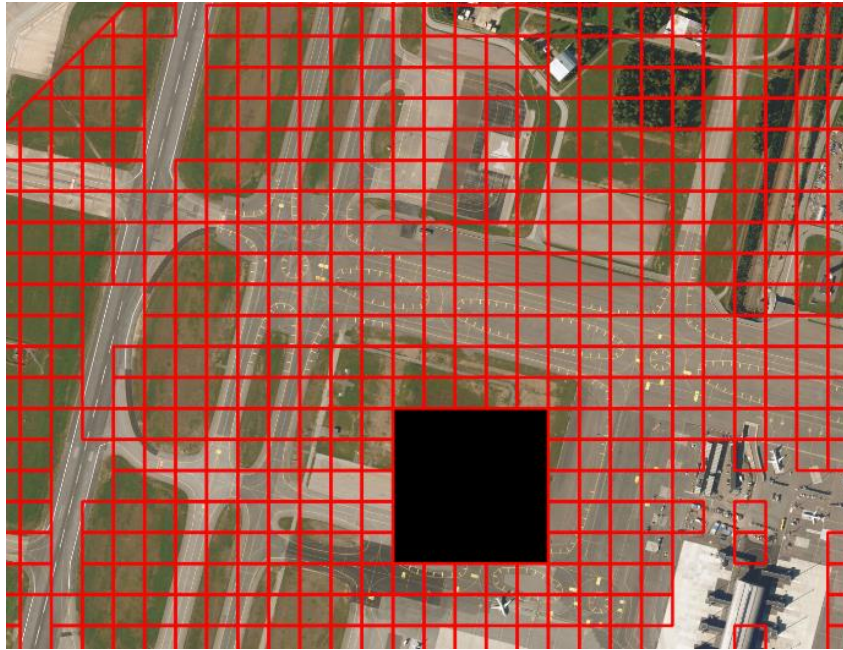
Figur 13. To kirkegårder, venstre i Nes, og høyre i Ullensaker. Ruter med røde kanter markerer predikert endring.

Tilstedeværelse av vegetasjon i boligområder generelt gjør modellen mer sannsynlig til å predikere endring. Dette er fordi disse områdene inneholder mindre av det som kjennetegner arealtypen bebygd, nemlig bebyggelse. I tillegg lager trærne skygger, som også er et problem for modellen. Trærne er oftere mer spredt enn i en skog, som gjør at det er større områder med skygge mellom trærne. Figur 14 viser eksempel på ruter i et boligområde med arealtype bebygd hvor det er feilaktig predikert endring. Her kan vi se mange trær og en del skygge fra disse trærne.



Figur 14. Eksempel på boligfelt med trær i Nes. Legg merke til skyggene på bakken mellom trærne. Ruter med røde kanter markerer predikert endring.

Rullebanene og bygningene på en flyplass klassifiseres som bebygd i AR5. Rullebanene har veldig liten variasjon i intensitet, og ligner generelt lite på annen bebyggelse. Figur 15 viser et eksempel fra Ullensaker.



Figur 15. Et utsnitt av Oslo Lufthavn (Gardermoen) i Ullensaker. Rødfargede ruter markerer predikert endring. Det svarte området mangler prediksjoner.

5.1.2 Fulldyrka jord

Fulldyrka jord har $MCC = 0,111$ samlet, og relativt stabilt rundt dette i alle kommunene. Innenfor arealtypen fulldyrka jord, har modellen i mange tilfeller problemer med å godta grønne jorder. Figur 16 viser et eksempel på dette.



Figur 16. Eksempel på grønn fulldyrka jord. Ruter med røde kanter markerer predikert endring.

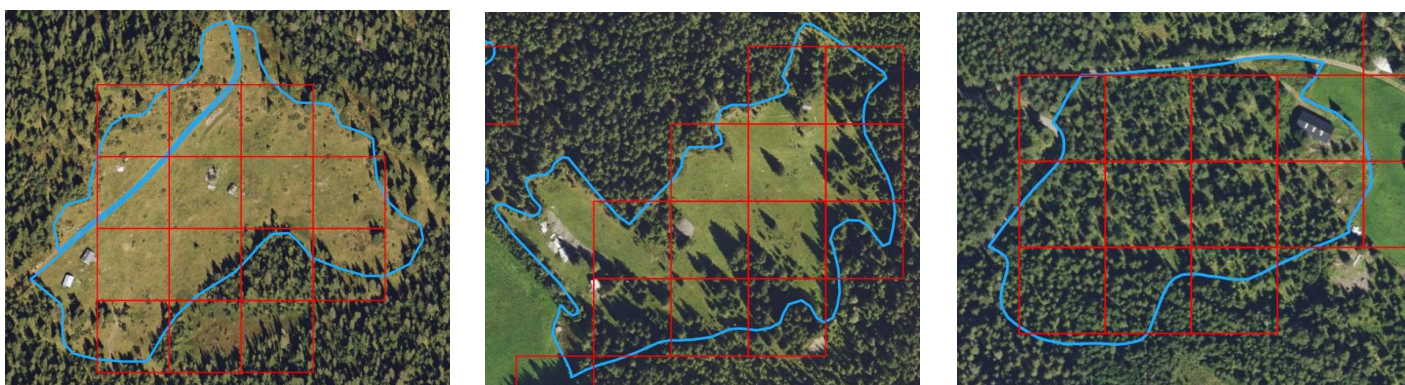
5.1.3 Overflatedyrka jord

Overflatedyrka jord har varierende MCC mellom kommuner, fra -0,167 i Nes til 0,295 i Sola. Dette kan til dels være fordi det er færre områder med disse arealtypen, og dermed blir det mer usikkerhet forbundet med den statistiske klassifiseringsnøyaktigheten i hver enkelt kommune. Dette gjelder spesielt for Romerike/Glåmdalen-kommunene, der det er svært få observasjoner av overflatedyrka jord i prediksjonsdatasettet (se Tabell 5).

5.1.4 Innmarksbeite

Innmarksbeite har i likhet med overflatedyrka jord relativt få observasjoner, dog ikke i like stor grad. Innmarksbeite har også varierende MCC mellom kommuner, men en samlet MCC på 0,094. MCC til innmarksbeite er tilnærmet 0 i Jæren, mens den er negativ i Romerike/Glåmdalen og Valdres.

Innmarksbeite kan være tresatt, så lenge arealet oppfyller definisjonen for innmarksbeite¹ (Ahlstrøm et al., 2019). Areal som oppfyller kravene til både innmarksbeite og skog skal ha arealtypen innmarksbeite. Det betyr at innmarksbeite kan ha et ganske ulikt utseende avhengig av i hvilken grad arealet er tresatt. Dette kan gjøre at modellen har vanskelig for å generalisere representasjonen av innmarksbeite. Figur 17 viser tre eksempler av innmarksbeite med ulik grad av tredekke.



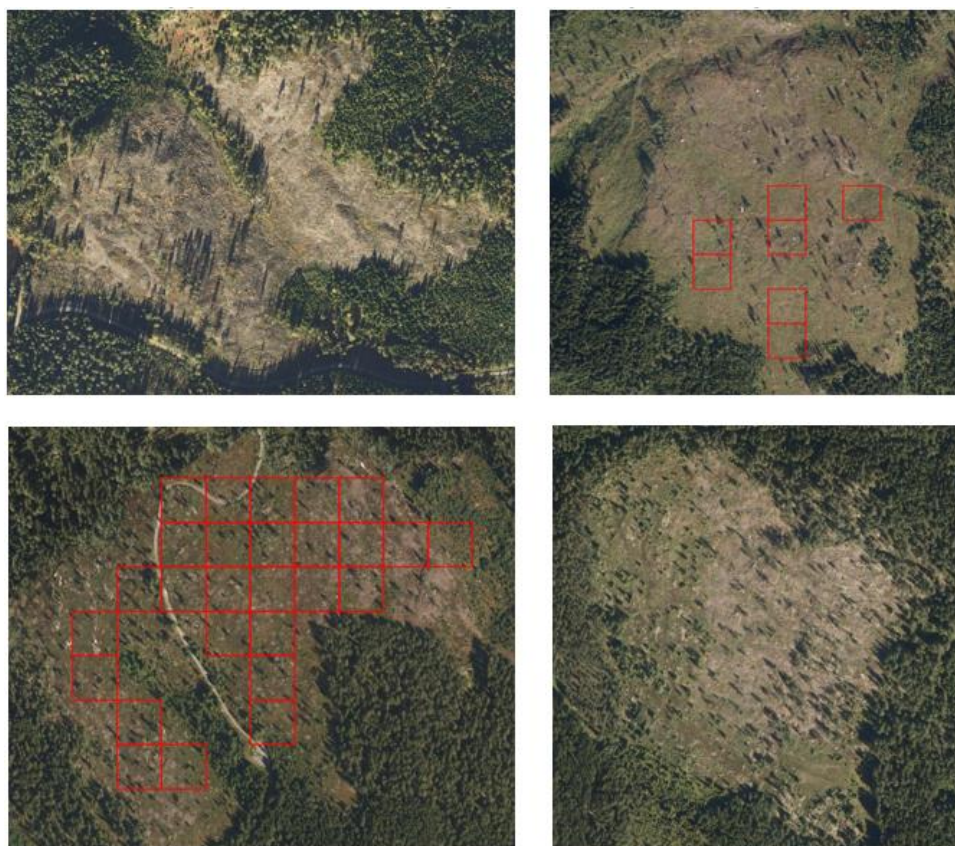
Figur 17. Viser tre eksempler på ulik representasjon av innmarksbeite fra Nord-Aurdal, med ulik grad av tredekke. Ruter med røde kanter markerer predikert endring. Blå linje markerer grense for innmarksbeite.

¹ Definisjon av innmarksbeite: "Jordbruksareal som kan benyttes som beite, men som ikke kan høstes maskinelt. Minst 50 % av arealet skal være dekt av godkjente grasarter eller beitetålende urter." (Ahlstrøm et al., 2019)

5.1.5 Skog

Skog er arealtypen som har best MCC, $MCC = 0,137$ samlet. Skog er den vanligste arealtypen, og da er det naturlig at modellen har lært seg denne arealtypen bedre. Det kan også tenkes at noe av den gode klassifiseringsnøyaktigheten kommer av at denne arealtypen inneholder en del «enkle» observasjoner, hvor det er tydelig at det er skog i bildet. Skog er en relativt homogen arealtype i utseende fra luften, både når det kommer til farge og tekstur. Arealtypen kjennetegnes av grønne områder med litt skygge. Unntakene er hogstflater og ledningsgater.

Hogstflater skiller seg fra vanlig skog i fraværet av trær. De klassifiseres likevel som skog i AR5 (Ahlstrøm et al., 2019). Hvor stort problem dette er for modellen, varierer ganske mye. Figur 18 viser fire hogstflater, de to øverste i Eidskog, og de to nederste i Nes. Vi kan se at i begge disse kommunene har vi tilfeller av hogstflate der modellen feilaktig predikerer endring, og tilfeller der modellen med rette ikke predikerer endring. Det viser at modellen til en viss grad har lært seg at utseendet til en hogstflate er et gyldig utseende for arealtypen skog. Likevel varierer også hogstflater i utseende. Det kan se ut til at det er de mørkere hogstflatene modellen har størst utfordringer med.



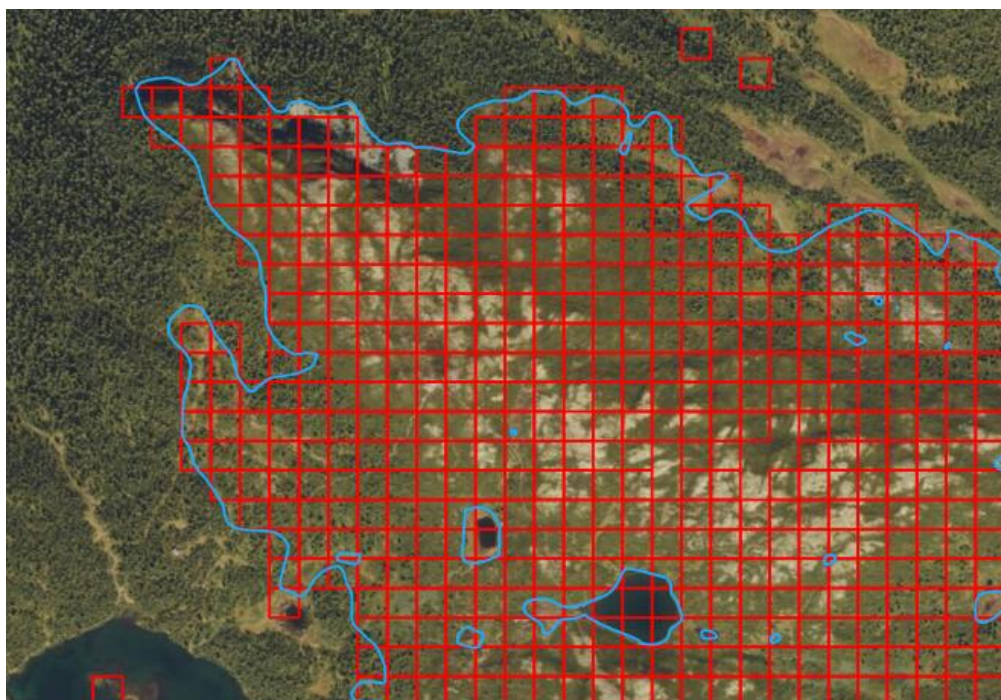
Figur 18. Fire hogstflater, de to øverste fra Eidskog, og de to nederste fra Nes. Ruter med røde kanter markerer predikert endring.

Ledningsgater gjennom skog skiller seg ut fra skog ved fravær av trær, i likhet med hogstflater. Dette ser likevel ikke ut til å være noe problem for modellen. Det kan være fordi ledningsgatene er relativt smale, og at dette da ikke gir nok utslag til at modellen predikerer endring.

5.1.6 Åpen fastmark

Åpen fastmark er arealtypen med lavest MCC, $MCC = -0,01$ samlet, og negativ MCC i alle kommuner unntatt Gjerdrum og Gjesdal. Åpen fastmark er svært heterogen arealtype. I og med at definisjonen på åpen fastmark er "Fastmark som ikke er jordbruksareal, skog, bebygd eller samferdsel" (Ahlstrøm et al., 2019), så er det åpenbart at denne arealtypen inneholder arealer av et vidt spekter i utseende. Dette gjør det vanskelig for en maskinlæringsmodell å generalisere utseendet til denne arealtypen i en felles representasjon.

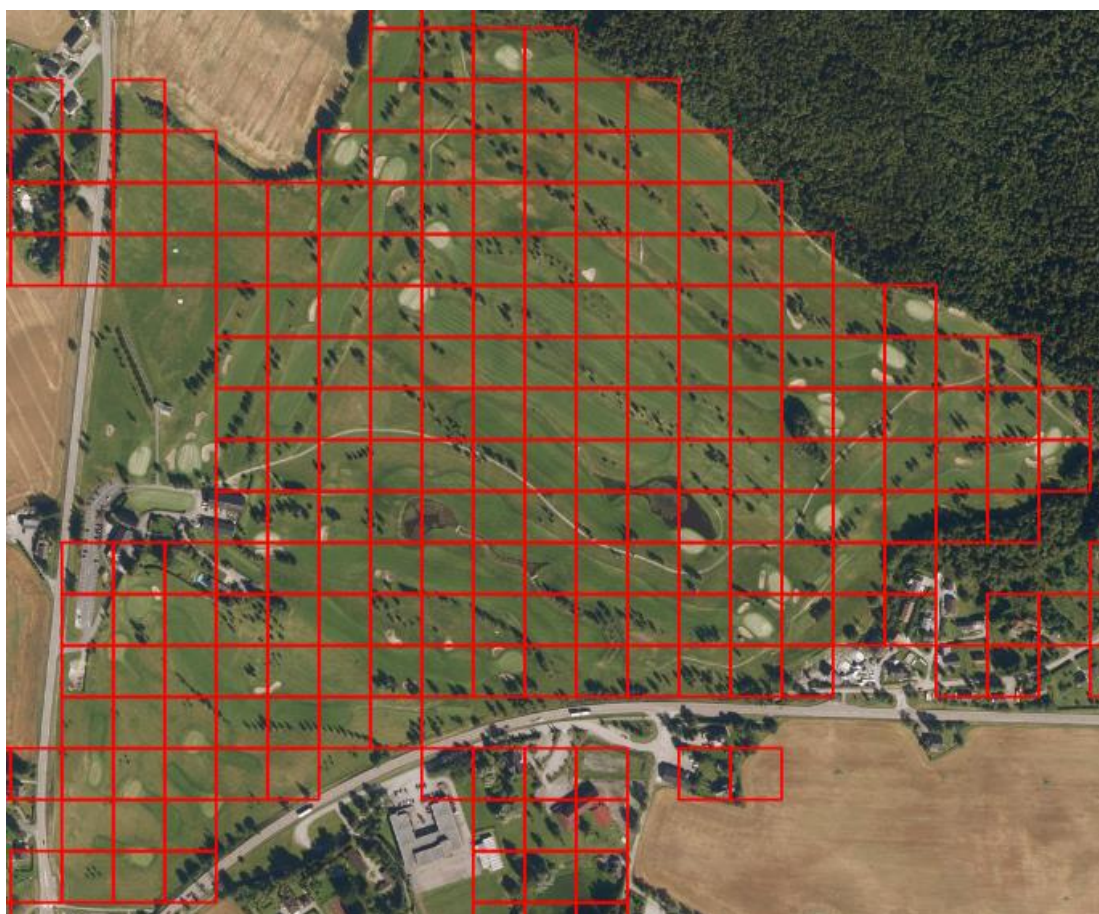
Spesielt dårlig er det i Valdres, med $MCC = -0,289$ i Nord-Aurdal og $-0,342$ i Etnedal. Dette er mye på grunn av snaumark i fjellet. Snaumark er fastmark med naturlig vegetasjonsdekke som ikke er skog (NIBIO, 2017). Snaumark klassifiseres som åpen fastmark i AR5. Snaumark er ofte grønt i ortofoto, i motsetning til mange andre typer åpen fastmark. I tillegg er det i noen tilfeller i fjellet også spredt tredekke. Dette gjør at modellen kan predikere avvik i disse områdene. Figur 19 viser et eksempel på



Figur 19. Viser et eksempel på snaumark i fjellet i Nord-Aurdal. Ruter med røde kanter markerer predikert endring. Blå linje markerer grense for åpen fastmark.

snaumark i fjellet. Legg merke til at nesten alle rutene innenfor åpen fastmark har predikert endring, selv om arealtypen er riktig.

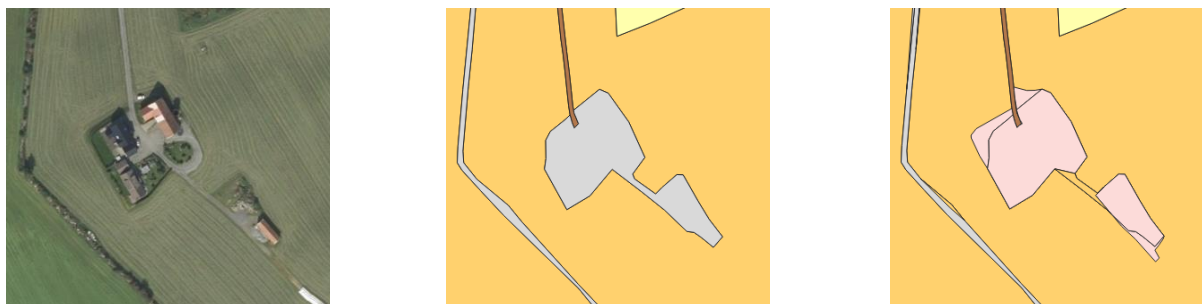
Et annet problemområde innen åpen fastmark er golfbaner. En golfbane der det kun er gjort mindre inngrep i jorda skal klassifiseres som jordbruksareal etter vanlige kriterier (se Tabell 2). Hvis det er gjort større inngrep i jorda skal golfbanen klassifiseres som åpen fastmark. Disse skiller seg ut fra andre områder med arealtypen åpen fastmark, og modellen har derfor tilbøyelighet til å predikere endring her. Et eksempel på en slik golfbane kan ses i Figur 20.



Figur 20. Eksempel på golfbane klassifisert som åpen fastmark i Nes. Ruter med røde kanter markerer predikert endring.

Før 2013 var regelen at alle gårdstun skulle klassifiseres som åpen fastmark, uavhengig av utbyggingsgrad og opparbeiding (Bjørkelo et al., 2013). Dette ble imidlertid endret for å få en mer ensartet klassifisering, der utbygde og betydelig opparbeidede arealer bare forekommer i arealtypen bebygd, og ikke i åpen fastmark. Siden modellen er trent på arealtypene før endringene, så har modellen lært at gårdstun er en helt vanlig representasjon av åpen fastmark. Dermed svekker det resultatene for åpen fastmark

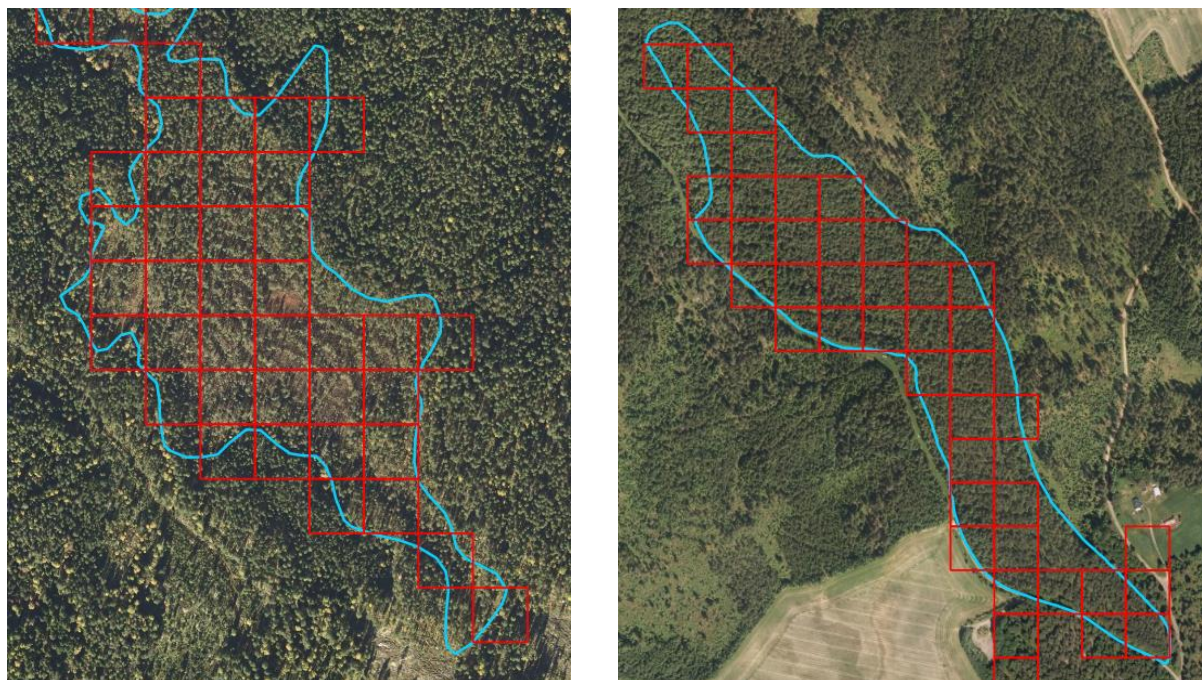
betydelig. Etter en visuell vurdering ser dette ut til å være en stor grunn til den dårlige klassifiseringsnøyaktigheten på åpen fastmark. Figur 21 viser et eksempel på et gårdstun, og hvordan arealtypene har blitt endret.



Figur 21. Til venstre flybilde av et gårdstun i Sola, i midten arealtypene før endring, og til høyre arealtypene etter endring.

5.1.7 Myr

MCC til myr ligger for alle østlandskommunene like rundt 0, og samlet for myr ligger den på 0,031, som er så vidt over nivå med tilfeldig gjetning. Det store problemet for modellen når det kommer til myr, er når myra er tresatt. Da synes bare trærne fra overflaten, som avviker fra utseendet til en typisk ikke-tresatt myr. Det er i de aller



Figur 22. To eksempler på tresatt myr, venstre fra Nes, og høyre fra Sør-Odal. Ruter med røde kanter markerer predikert endring. Blå linje markerer grensen til myra.

fleste tilfeller umulig å se fra flybildene om det finnes myr under tredekket. Figur 22 viser to eksempler på tresatt myr.

5.1.8 Skygge

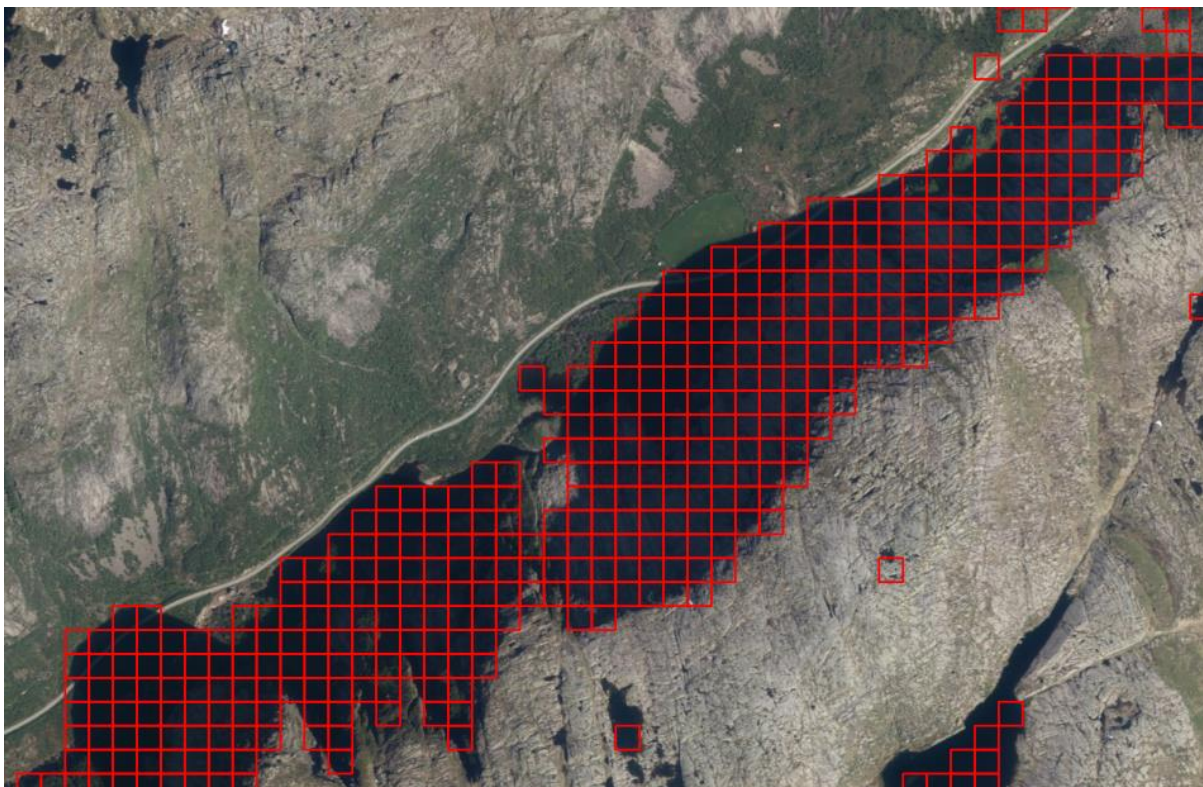
Skygge er et velkjent problem innen passiv fjernmåling (Mostafa, 2017; Storey et al., 2017). Skyggelagte områder opptrer nesten helt svarte på ortofotoene, og ligner derfor ikke på noen arealtype. Modellen har derfor stor tilbøyelighet til å registrere avvik på ruter med skygge uavhengig av arealtype, selv om arealtypen egentlig ikke er endret.

De fleste kommunene i denne oppgaven har relativt flatt terreng, og derfor ikke mange store skyggeområder. Likevel er det mange små områder spredt rundt omkring med skygger fra trær og bygninger, og modellen har stor tilbøyelighet til å predikere endringer i disse områdene. Figur 23 viser et eksempel hvor skygger fra trær dekker deler av et jorde.



Figur 23. Eksempel på skygger fra trær i Nes. Ruter med røde kanter markerer predikert endring.

Effekten av skygger er spesielt tydelig i områder med dype daler og høye fjell, der det i større grad opptrer store områder med skygge i ortofotoene. Den kommunen i oppgaven med størst tilstedeværelse av skygge er Gjesdal. Figur 24 viser en skyggelagt dal i Gjesdal.



Figur 24. Viser et eksempel på skygge i flybildene, fra en dal i Gjesdal kommune. Ruter med røde kanter markerer predikert endring.

5.2 Diskusjon av resultater fordelt på geografisk område

MCC til prediksjonene for alle arealtyper i Romerike/Glåmdalen er 0,208, mens i Valdres og på Jæren ligger den omtrent på nivå med tilfeldig gjetning, med MCC på henholdsvis 0,030 og 0,038. Det er antakeligvis flere grunner til at modellen gjør det bedre på prediksjonene fra Romerike/Glåmdalen. En mulig forklaring er at det er mest treningsdata fra denne regionen. Treningsdata fra tre kommuner på Romerike ble brukt til trening av modellen, mot en kommune i Valdres og to på Jæren. Som en følge har modellen da lært fra flest eksempler fra denne regionen. Dette kan da ha gjort at modellen har lært seg Romerike/Glåmdalen-representasjonen av arealtypene bedre.

En annen mulig årsak er at kommunene i Romerike/Glåmdalen har betraktelig mindre andel åpen fastmark enn kommunene i Valdres og på Jæren (se Tabell 5). I og med at åpen fastmark er arealtypen modellen har størst problemer med å predikere, er det naturlig at kommunene med større andel åpen fastmark har lavere klassifiseringsnøyaktighet.

I tillegg er det store områder med snaumark i fjellet (forklart i avsnitt 5.1.6) i Valdres. Dette har antakeligvis en ganske stor effekt på klassifiseringsnøyaktigheten i, noe vi

kan se i MCC for åpen fastmark i Valdres, som ligger på $-0,301$. Nøyaktig hvor stor effekt dette har på den totale klassifiseringsnøyaktigheten i Valdres har ikke blitt undersøkt. På samme måte er den en del store områder med skygge i Gjesdal på Jæren. Dette har antakeligvis også en negativ effekt på klassifiseringsnøyaktigheten her, men vanskeligere å anslå fordi skygge ikke opptrer på én enkelt arealtype.

5.3 Feilkilder

Det er viktig å påpeke at MCC-verdiene som er beregnet i denne oppgaven er statistiske estimater. Det betyr at verdiene er anslag på klassifiseringsnøyaktigheten, basert på utvalgte data. Presisjonen til statistiske estimater avhenger av mengden data brukt til å estimere. Jo mer data, desto bedre nøyaktighet på estimatene. Derfor er det grunn til å være skeptisk til estimatene for overflatedyrka jord, bebygd og innmarksbeite, som hver består av henholdsvis $0,1\%$, $1,3\%$ og $2,1\%$ av observasjonene i prediksjonsdatasettet. Spesielt estimatene for disse arealtypene innen kommuner, hvor antallet observasjoner av disse arealtypene er enda mindre.

Kvaliteten på grunnlagsdataene påvirker også estimatenes nøyaktighet. Intet kart er perfekt, og noen steder vil AR5 være feil eller upresist. Siden hovedfokuset til AR5 er landbruk og skog, vil kartet være mindre nøyaktig for andre areal typer enn disse. Det generelle kravet til stedfestingsnøyaktighet for AR5-grenser som er synlige i terrenget er 2 meter (Ahlstrøm et al., 2019). Mange AR5-grenser representerer imidlertid gradvise overganger, og i disse tilfellene kan avviket mellom flere mulige gyldige grenser langt overstige 2 meter.

Det gjelder ikke bare at grensene til polygonene kan ha lav nøyaktighet, men også feilklassifiserte areal typer. Feilklassifiserte areal typer kan enten komme av at de har blitt registrert feil, eller at endringer ikke har blitt oppdaget i ajourføringsprosessen. Figur 25 viser et eksempel på et slikt tilfelle. Når det er feil i datagrunnlaget vil det gå ut over klassifiseringsnøyaktigheten til modellen når den har predikert avvik med rette. Dette gjør altså at resultatene vil vise en noe lavere klassifiseringsnøyaktighet enn det som egentlig er tilfelle. Det kan også gjøre at modellen har vanskeligere for å lære når den lærer av feilaktige data. Det er ikke gjort noen omfattende undersøkelse av

kvaliteten på de nåværende AR5-dataene, og det er derfor vanskelig å anslå hvor stor effekt det har på modellen.



Figur 25. Det skraverte området i Sola kommune har verdien skog i AR5, men mesteparten av dette området holder ikke kravet for skog (minst 6 trær per dekar).

Manglende prediksjoner (forklart i delkapittel 3.2.3) gir oss mindre data, som svekker presisjonen til den estimerte klassifiseringsnøyaktigheten. Områdene med manglende prediksjoner ser ut til å være tilfeldig spredd rundt, og de vil derfor ikke systematisk forverre eller forbedre klassifiseringsnøyaktigheten.

Ved tildeling av kun én arealtype til hver rute antas det at det alltid er den største arealtypen innenfor hver type som modellen mener har endret seg, men dette er ikke nødvendigvis alltid tilfelle.

Data fra kommunene Gjerdrum, Ullensaker, Nes og Etnedal ble brukt til trening av modellen i tillegg til evaluering. Bruk av treningsdata til å evaluere en modell frarådes generelt, fordi det ofte gir et urealistisk bilde på hvor bra modellen faktisk predikerer på nye data (Raschka & Mirjalili, 2017). Det er likevel bare enkelte av rutene i hver kommune som har blitt tilfeldig valgt ut for å bli brukt i treningen. Det betyr at majoriteten av rutene fra disse kommunene ikke har blitt brukt til trening, og effekten av dette på evalueringen blir da minimal. Figur 11 viser også at disse kommunenes MCC ligger omtrent på nivå med østlandskommunene som ikke ble brukt til trening.

5.4 Hva kan gjøres for å forbedre modellen?

Det er flere ting som bør vurderes når det kommer til videreutvikling av modellen. Når det kommer til kunstige nevralt nettverk er det ofte en stor grad av prøving og feiling som må til for å komme fram til hva som gir gode resultater. Derfor er det sannsynlig at noen, men ikke alle disse tiltakene vil ha positiv effekt på klassifiseringsnøyaktigheten.

5.4.1 Større variasjon i treningsdataene

Et viktig punkt er å sørge for at treningsdataene er representative for hver arealtype. Det betyr å sørge for at treningsdataene er mer varierte. Det vil gjøre at modellen lærer seg alle forskjellige representasjoner av hver arealtype, for eksempel at myrer faktisk kan ha tredekke. Dette kan oppnås ved å klargjøre data fra andre kommuner, helst fra andre landsdeler med annet landskap, for at modellen skal kunne generalisere best mulig.

5.4.2 Vekting av observasjoner

I tillegg til å skaffe mere treningsdata er det mulighet for å gi større vekt til de typene observasjoner som modellen har vanskelig for å predikere riktig. Da vil modellen endre seg mer for å klare å predikere de observasjonene som den sliter med, og endre seg mindre for de observasjonene som den klarer fint allerede (Gómez-Verdejo et al., 2006). I treningsdataene er allerede observasjoner fra sjeldnere areal typer vektet høyere, men dette bør også gjøres for de utfordrende områdene som er nevnt i avsnitt 5.1.

Utfordringen er å skaffe data om hvor disse utfordrende områdene befinner seg.

Tresatt myr er registrert i AR5 ved egenskapen "treslag", og kan da vektet direkte fra dataene. Datasettet FKB Arealbruk inneholder informasjon om flere av disse områdene, blant annet golfbaner og kirkegårder (Kartverket, 2018). Hogstflater er tidspunkt-avhengige, og det finnes derfor ingen kartdata for disse. De må derfor enten markeres manuelt eller ved bruk av gjenkjenningsalgoritmer dersom de skal vektet.

For å unngå problemet med at modellen lærer seg representasjonen fra en region bedre enn andre (forklart i avsnitt 5.2) kan det vektet basert på region. Da vil observasjoner fra regioner med få eksempler bli viktigere, og modellen vil lære like mye fra hver region.

Et alternativ til vekting er oversampling og undersampling (Maxwell & Warner, 2018). Oversampling er å duplisere vanskelige observasjoner, mens undersampling er å fjerne unødvendige (lette) observasjoner. Formålet er i hovedsak det samme som vekting,

altså å tvinge modellen til å lære mer fra de vanskelige observasjonene. Effekten kan likevel variere i praksis. Fordelen med undersampling er at treningen da vil gå raskere, siden treningsdatasettet er mindre. På den andre siden er det sjans for at man taper nyttig informasjon i observasjonene man fjerner. Ved bruk av oversampling i stedet unngår man å fjerne observasjoner, men øker samtidig datamengden. Dette vil da føre til lengre treningstid, avhengig av tilgjengelig beregningskraft. Effekten av oversampling avhenger også av kvaliteten på de opprinnelige dataene. Hvis observasjonene som dupliseres ikke inneholder noen nyttig eller generaliserbar informasjon, vil man ikke tjene noe på å oversample dem.

5.4.3 Skyggekorreksjon

Som nevnt i avsnitt 5.1.8 er skygge svært problematisk for modellen, spesielt i områder med mye fjell og daler. For å fjerne de negative effektene av skyggene, er det mulig å utføre skyggekorreksjon på flybildene. Det betyr at bildene modifiseres slik at landskapet ser ut slik det ville sett ut uten skygger. Skyggekorreksjon består av to deler: skyggedeteksjon og skyggefjerning (Mostafa, 2017). Først må skyggelagte områder identifiseres (skyggedeteksjon), og så må skyggeeffekten fjernes fra disse områdene (skyggefjerning).

Ifølge Movia et al. (2016) kan skyggedeteksjon deles inn i to typer: modellbaserte metoder og egenskapsbaserte metoder. Modellbaserte metoder for skyggedeteksjon utnytter de geometriske egenskapene til skyggene. Disse trenger informasjon om høyde, posisjon og orientering av landskapet og sensoren for å konstruere en geometrisk modell for skyggedeteksjon. Selv om disse metodene i teorien stiller sterkt, så foreligger sjelden de nøyaktige geometriske parameterne som trengs for å konstruere en god modell (Luo et al., 2020).

Egenskapsbaserte metoder for skyggedeteksjon baserer seg derimot på egenskaper ved skygger utledet fra bildet, som spektral og romlig informasjon. Her har tradisjonelt terskelbaserte metoder vært mye brukt (Chung et al., 2008; Dare, 2005; Tsai, 2006). Nylig har konvolusjonelle nevrale nettverk vist svært gode resultater innen skyggedeteksjon. Luo et al. (2020) foreslår en struktur de kaller DSSDNet. Denne trener de opp på et åpent tilgjengelig skyggedatasett, og viser god nøyaktighet både på dette datasettet og andre datasett som den ikke har blitt trent på. DSSDNet viser seg å ha utfordringer med å detektere skygger i flybilder med oppløsning på 7 cm. Dette er likevel ikke noe problem, for i Norge i Bilder-omløpsfotograferingen er oppløsningen 25

cm. Det anbefales derfor at DSSDNet eller et lignende konvolusjonelt nevralt nettverk tas i bruk for skyggedeteksjon på flybildene.

Når skyggelagte områder er detektert, må en fjerne skyggeeffekten fra disse områdene. En tradisjonelt mye brukt metode for skyggefjerning er lineær korrelasjonskorreksjon (Liu & Yamazaki, 2012; Lorenzi et al., 2012; Sarabandi et al., 2004). Denne metoden antar at det er en lineær sammenheng mellom intensitetsverdiene til pikslene i et skyggelagt område og det tilsvarende ikke-skyggelagte området, og transformerer skygge-pikslene ved bruk av denne sammenhengen. Ulempen med denne metoden er at det tilsvarende ikke-skyggelagte området må defineres manuelt, som fort kan bli en tidkrevende prosess.

Luo et al. (2019) bruker en romlig adaptiv teknikk basert på vektet total variasjon i bildet for skyggefjerning. Denne metoden viser gode resultater, men er likevel bare anvendt på relativt små skygger i urbane områder. I Norge har vi mye naturlige skygger fra fjell og daler, og disse dekker gjerne mye større områder enn urbane skygger. Det er derfor usikkert om metoden vil fungere like godt på de større skyggelagte områdene i Norge i Bilder-flybildene, men det er uansett en mulighet å prøve ut.

5.4.4 Muligheter for bruk av andre fjernmålingsdata

Rødt, grønt og blått lys dekker bare en liten del av det elektromagnetiske spekteret, og RGB-flybilder alene inneholder dermed relativt lite informasjon om objektene på jordas overflate sammenliknet med andre typer fjernmålingsdata (Ashapure et al., 2019; Marcial-Pablo et al., 2019). Derfor kan det være mye å tjene å kombinere RGB-flybildene med ytterligere data.

Spesielt nærinfrarødt lys inneholder informasjon som er nyttig for klassifisering av vegetasjon (Knipling, 1970; Townshend & Justice, 1986). Siden vegetasjon er et viktig kjennetegn for flere arealtyper, er det sannsynlig at nær-infrarøde data vil være spesielt nyttig for disse.

Data fra satellitten Sentinel-2 har allerede vist seg nyttig for klassifisering av landbrukstyper (Debella-Gilo & Gjertsen, 2021). Denne satellitten tilbyr fritt tilgjengelige nærinfrarøde bilder med 5-dagersintervall (ESA, 2021). En ulempe med satellittbildene er at den romlige oppløsningen til nær-infrarøde data fra Sentinel-2 er 10 meter, med det ganske mye lavere enn flybildenes 0,25 meter. En annen ulempe med satellittbildene er at det ofte er skyer i bildene, som gir en del manglende data.

Til tross for dette er det godt mulig at disse dataene kan tilføye nyttig informasjon for endringsdeteksjonen.

Det kan også i teorien være mulig å ta i bruk laserdata. Bjerke (2011) skrev en masteroppgave om mulighetene for ajourhold av AR5 basert på laserdata. Resultatene fra oppgaven viser at laserdataene er nyttige spesielt for identifisering av skog og bebyggelse. Denne oppgaven gjør likevel bare en undersøkelse i et begrenset område i Ås kommune, og det er ikke sikkert om dette kan generaliseres. Likevel er det ikke utenkelig at det kan være store fordeler å kombinere flybilder med laserdata, spesielt med tanke på den store utviklingen som har skjedd innen maskinlæring siden 2011.

Den store utfordringen med laserdata er uansett å få skaffet relevante data for ethvert område til riktig tid. Dersom laserdataene skal være nyttige for endringsdeteksjon, er det en viktig forutsetning at dataene er fra samme tidspunkt som flybildene. I praksis er nok ikke dette realistisk, ettersom laserskanning er mye dyrere enn flybilder, og ikke utføres på langt nær så ofte. Likevel vil prosjektet Nasjonal Detaljert Høydemodell være fullført i 2022, og vil da gi landsdekkende detaljerte laserdata (Kartverket, 2021). Disse vil likevel gå relativt fort ut på dato i en endringsdeteksjonssammenheng, og vil antakeligvis kun være nyttige om de tas i bruk med flybilder fra samme tidspunkt.

5.4.5 Hyperparameteroptimalisering

Hyperparameteroptimalisering går ut på å justere algoritmens parametere for å oppnå bedre klassifiseringsnøyaktighet (Raschka & Mirjalili, 2017). Dette er i stor grad en iterativ prosess der det er vanskelig å vite hvilke verdier av parameterne som vil gi best nøyaktighet på forhånd (Feurer & Hutter, 2019; Snoek et al., 2012). I tillegg må modellen trenes på nytt med nye verdier av hyperparameterne for å kunne vurdere om de gir bedre klassifiseringsnøyaktighet. Dette gjør igjen at hyperparameteroptimalisering fort kan bli en tidkrevende prosess. Med store datasett kan treningen av en maskinlæringsmodell ta flere dager, avhengig av tilgjengelig beregningskraft. Derfor er det mye å tjene på å bruke en effektiv metode for å finne de beste verdiene av hyperparameterne.

Rutenettsøk er den mest elementære hyperparameteroptimaliseringsmetoden. Den går ut på at brukeren spesifiserer et sett med verdier for hver hyperparameter som skal optimaliseres, og at modellen da trenes og evalueres for alle kombinasjoner av verdier for hyperparameterne (Feurer & Hutter, 2019). Ulempen med denne metoden er at tiden det tar øker eksponentielt med antall hyperparametere som skal optimaliseres. Et

alternativ til rutenettsøk er tilfeldig søk, som i stedet velger tilfeldige kombinasjoner av verdier for hyperparameterne. Denne metoden fungerer i praksis bedre enn rutenettsøk når noen hyperparametere er mye viktigere enn andre, som ofte er tilfelle (Bergstra & Bengio, 2012). Ulempen med tilfeldig søk er at den utnytter informasjonen fra tidligere resultater når den skal velge neste verdier av hyperparametere som skal evalueres. Geitle og Olsson (2019) foreslår en algoritme som de kaller adaptivt tilfeldig søk (adaptive random search). Denne metoden er mer effektiv enn vanlig tilfeldig søk fordi den tar hensyn til tidligere resultater, og antar at de beste resultatene ligger i nærheten av disse. Slik finner den mye raskere fram til optimale verdier av hyperparameterne. Adaptivt tilfeldig søk kan parallelliseres, som gir enda raskere utføringstid.

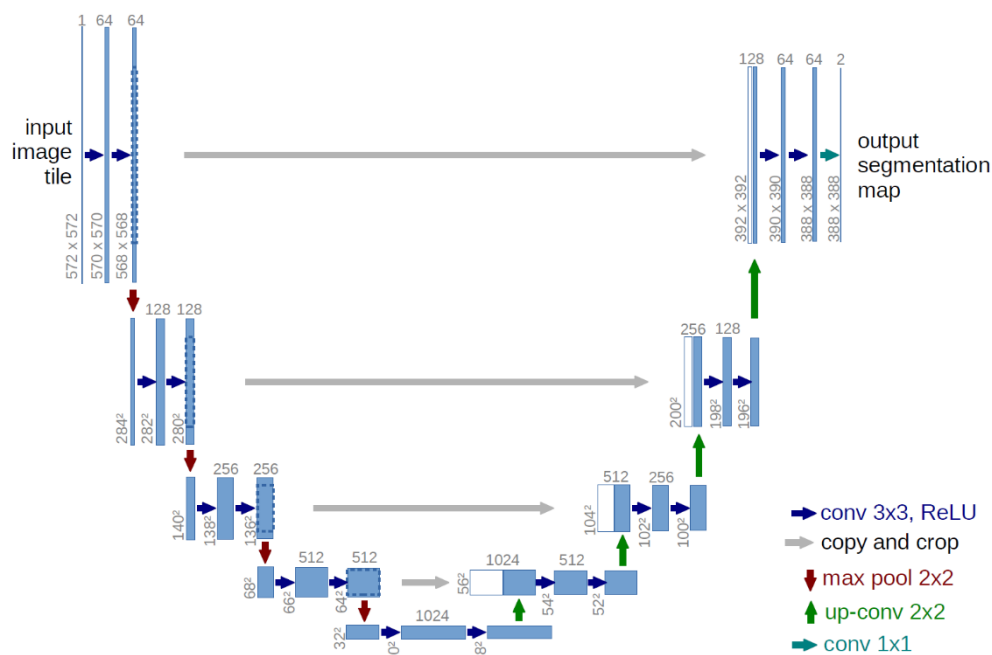
For et nevralt nettverk er det mange hyperparametere som kan optimaliseres. Noen går på strukturen til nettverket, for eksempel antall skjulte lag i koderen og dekodeeren og antall nevroner i hvert lag (Diaz et al., 2017). I en autokoder styrer disse forholdene sammen kompresjonsfaktoren til den komprimerte representasjonen. I tillegg kan antallet konvolusjoner og størrelsen til konvolusjonsfiltrene optimaliseres. Læringsraten til nettverket bør også justeres. En for lav læringsrate gjør at modellen lærer veldig tregt, mens en for høy læringsrate kan gjøre læringen veldig ustabil (Yu & Chen, 1997). I tillegg kan man justere antall epoker, som er hvor lenge modellen skal trenes. Dersom modellen trenes for få epoker, er det mulig at man stopper den før den er ferdig med å lære. Hvis man trener i for mange epoker, er det fare for at modellen tilpasser seg for godt til treningsdata, og ikke klarer å generalisere til nye data (Liu et al., 2008).

Siden arealtypene i AR5 i noen tilfeller er feilklassifisert, kan det være nyttig å ta i bruk en tapsfunksjon som tar høyde for dette. Rusiecki (2019) introduserer trimmet kategorisk kryssentropi, som viser seg å være en mer robust versjon av kategorisk kryssentropi når man har en andel feilkategoriserte data.

5.4.6 Bildesegmentering

Dersom ingen av de ovenfornevnte forslagene gir ønsket klassifiseringsnøyaktighet, er det en mulighet å prøve andre klassifiseringsalgoritmer. En særlig ulempe med den nåværende rutebaserte tilnærmingen er at man ikke får vite hvor i ruta modellen mener at det har skjedd en endring. Et godt alternativ er bildesegmenteringsalgoritmer, som detekterer endring på pikselnivå i stedet for å predikere endring i store ruter. Disse gir mye høyere oppløsning på prediksjonene, fordi

hver eneste piksel blir predikert. For eksempel bruker Daudt et al. (2018a) siamesiske konvolusjonelle nevralt nettverk for pikselvis endringsdeteksjon med suksess. Disse er basert på U-net-strukturen, fra Ronneberger et al. (2015). U-Net er en type konvolusjonelt nevralt nettverk som består av to deler: en krympende del, og en utvidende del (se Figur 26). Den krympende delen består av konvolusjoner og samlingsoperasjoner, mens den utvidende delen består av oppkonvolusjoner og oppsamlinger. Sann sett ligner denne arkitekturen på en konvolusjonell autokoder, hvor den krympende delen er koderen og den utvidende delen er dekoderen. Men U-Net har sammenkoblinger som overfører romlig informasjon fra den krympende til den utvidende delen. Formålet med dette er å kombinere den abstrakte og lite lokaliserte informasjonen i den utvidende delen med de romlige detaljene fra den krympende delen av nettverket. U-Net har vist seg å være eksepsjonelt godt egnet til å løse bilde-segneringsproblemer, både innen fjernmåling og i andre fagfelt.



Figur 26. Illustrasjon av U-Net-arkitekturen. De blå boksene representerer «feature maps», og de grå pilene representerer sammenkoblingene fra krympende til utvidende del. Blå piler representerer 3x3 konvolusjoner, og røde piler representerer 2x2 maks-opsamling. Figur fra (Ronneberger et al., 2015).

Selv om Daudt et al. (2018a) bruker siamesiske konvolusjonelle nevralt nettverk basert på U-net-strukturen til å sammenligne par av flybilder, står det ingenting i veien for at metoden skal kunne anvendes på arealkart og flybilde. Zhan et al. (2017), Jiang et al. (2020) og Chen et al. (2021) demonstrerer videre effektiviteten av siamesiske nevralt

nettverk for bildesegmentering til endringsdeteksjon, dog med varierende struktur på nettverkene.

En ytterligere fordel er at prediksjonene fra bildesegmenteringsalgoritmer egner seg bedre til å evaluering. Ved at hver piksel er en prediksjon, unngår man problemet med å måtte tildele kun én arealtype til en rute som egentlig inneholder flere areal typer. Dette vil selvfølgelig fortsatt være tilfelle i enkelte piksler på grensen mellom areal typene, men dette vil gjelde en mye mindre andel av observasjonene. Dermed vil estimatene på klassifiseringsnøyaktigheten styrkes ytterligere ved bruk av bildesegmenteringsalgoritmer.

6 Konklusjon

I denne oppgaven har den statistiske klassifiseringsnøyaktigheten til prediksjonene fra NIBIOs siamesiske konvolusjonelle variasjonsautokoder til endringsdeteksjon blitt undersøkt. Samlet ligger Matthews korrelasjonskoeffisient (MCC) på 0,104, som betyr at modellens prediksjoner er svakt korrelert med de faktiske endringene. Modellen har derfor lært seg enkelte sammenhenger som er nyttige for endringsdeteksjonen. Dette er likevel ikke særlig mye bedre enn tilfeldig gjetning ($MCC=0$). Modellens prediksjoner når derfor ikke opp til nivået som kan forventes av en modell som faktisk skal tas i bruk og være nyttig.

For de fleste arealtyper og kommuner ligger MCC mellom -0,1 og 0,3. Arealtypene skog, fulldyrka jord, bebygd og innmarksbeite har positiv klassifiseringsnøyaktighet ($0,09 < MCC < 0,14$, svak positiv korrelasjon). Overflatedyrka jord, innmarksbeite og myr er omtrent på nivå med tilfeldig gjetning ($MCC \approx 0$, ingen korrelasjon), mens modellen systematisk feilpredikerer åpen fastmark ($MCC < 0$, svak negativ korrelasjon). Mange konkrete områder har blitt identifisert som særlig utfordrende for modellen, blant annet vegetasjon i bebyggelse, flyplasser, hogstflater, golfbaner, snaumark i fjellet, tresatt myr og områder med skygge. Alle disse har til felles at de avviker betydelig fra det typiske utseendet til arealtypen de tilhører.

Klassifiseringsnøyaktigheten viser seg å være bedre i Romerike/Glåmdalen enn i Valdres og på Jæren. En større andel av treningsdataene kommer fra Romerike/Glåmdalen enn de to andre regionene. Dette kan ha ført til at modellen har lært mer herfra, og mindre fra de andre regionene. Det har også sammenheng med fordelingen av arealtyper i regionene. I Valdres og på Jæren er det større forekomst av åpen fastmark, som reduserer klassifiseringsnøyaktigheten i disse kommunene. I tillegg er det mye snaumark i fjellet i Valdres, og en del skygge på Jæren, som også bidrar negativt i de regionene.

Det er viktig å påpeke at resultatene bare er estimater på klassifiseringsnøyaktigheten. Arealtypene overflatedyrka jord, bebygd og innmarksbeite har færrest observasjoner, og følgelig vil usikkerheten ved estimatene for disse arealtypene være størst. AR5-dataene inneholder feil både når det kommer til feilklassifiserte arealtyper og unøyaktige grenser. I prediksjonsdataene er det også tilfeller av områder med manglende prediksjoner. Dette kan medvirke til at usikkerheten til estimatene øker ytterligere.

Flere ulike tiltak bør prøves for å forbedre modellens klassifiseringsnøyaktighet. Større variasjon i treningsdataene kan gjøre modellen mer robust og bedre til å gjenkjenne ulike utseender til areal typer. Det bør innhentes data fra ulike deler av landet.

For å få modellen til å lære seg de identifiserte problemområdene, bør disse observasjonene vektet tyngre under treningen. Undersampling og/eller oversampling kan vurderes som et alternativ til vektning. Skyggekorreksjon bør utføres for å redusere effekten av skyggelagte områder.

Det bør også vurderes å kombinere RGB-flybildene med andre fjernmålingsdata. Spesielt nærinfrarøde satellittbilder antas å ha en positiv effekt på endringsdeteksjonen. Laserdata kan også vurderes, men begrenses av tilgjengeligheten av dataene fra samme tidspunkt som flybildene.

Justering av modellens struktur og parametere bør gjøres kontinuerlig for å øke klassifiseringsnøyaktigheten. I tillegg er det mulighet for å prøve ut bilde-segmenteringsalgoritmer i stedet for den rutebaserte metoden. Disse har fordelen av å gi en mye høyere oppløsning på prediksjonene.

Et viktig steg videre er å trene og evaluere modellen på flere kommuner i andre deler av Norge. Med det kan vi være ett skritt nærmere et mer oppdatert arealressurskart av høyere kvalitet.

7 Referanser

- Ahlstrøm, A. P., Bjørkelo, K. & Fadnes, K. (2019). *AR5 Klassifikasjonssystem. Klassifisering av arealressurser*. NIBIO BOK: Norsk institutt for bioøkonomi.
- An, J. & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2 (1).
- Ashapure, A., Jung, J., Chang, A., Oh, S., Maeda, M. & Landivar, J. (2019). A Comparative Study of RGB and Multispectral Sensor-Based Cotton Canopy Cover Modelling Using Multi-Temporal UAS Data. *Remote Sensing*, 11 (23). doi: 10.3390/rs11232757.
- Audebert, N., Le Saux, B. & Lefèvre, S. (2017). Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sensing*, 9 (4). doi: 10.3390/rs9040368.
- Bergstra, J. & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *The Journal of Machine Learning Research*, 13: 281-305.
- Bjerke, H. (2011). *Bruk av laserdata ved ajourhold av arealressurskartet AR5*. Masteroppgave. Ås: UMB.
- Bjørkelo, K., Bjørken, A. & Frydenlund, J. (2013). *AR5 Klassifikasjonssystem: Endring av arealtypen åpen fastmark og bebyggd*: Norsk institutt for skog og landskap.
- Blom Norway. (2020). Omløpsfotografering Rogaland 2019.
- Bromley, J., Bentz, J., Bottou, L., Guyon, I., Lecun, Y., Moore, C., Sackinger, E. & Shah, R. (1993). Signature Verification using a "Siamese" Time Delay Neural Network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7: 25. doi: 10.1142/S0218001493000339.
- Chen, J., Yuan, Z., Peng, J., Chen, L., Huang, H., Zhu, J., Liu, Y. & Li, H. (2021). DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 1194-1206. doi: 10.1109/JSTARS.2020.3037893.
- Chen, T., Liu, X., Xia, B., Wang, W. & Lai, Y. (2020). Unsupervised Anomaly Detection of Industrial Robots Using Sliding-Window Convolutional Variational Autoencoder. *IEEE Access*, 8: 47072-47081. doi: 10.1109/ACCESS.2020.2977892.
- Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21 (1): 6-6. doi: 10.1186/s12864-019-6413-7.

- Chung, K.-L., Lin, Y.-R. & Huang, Y.-H. (2008). Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE Transactions on Geoscience and Remote sensing*, 47 (2): 671-682.
- Dare, P. M. (2005). Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71 (2): 169-177.
- Daudt, R. C., Saux, B. L. & Boulch, A. (2018a, 7-10 Oct. 2018). *Fully Convolutional Siamese Networks for Change Detection*. 2018 25th IEEE International Conference on Image Processing (ICIP).
- Daudt, R. C., Saux, B. L., Boulch, A. & Gousseau, Y. (2018b, 22-27 July 2018). *Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks*. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.
- Debella-Gilo, M. & Gjertsen, A. K. (2021). Mapping Seasonal Agricultural Land Use Types Using Deep Learning on Sentinel-2 Image Time Series. *Remote Sensing*, 13 (2). doi: 10.3390/rs13020289.
- Delgado, R. & Tibau, X.-A. (2019). Why Cohen's Kappa should be avoided as performance measure in classification. *PloS one*, 14 (9): e0222916-e0222916. doi: 10.1371/journal.pone.0222916.
- Diaz, G. I., Fokoue-Nkoutche, A., Nannicini, G. & Samulowitz, H. (2017). An effective algorithm for hyperparameter optimization of neural networks. *IBM Journal of Research and Development*, 61 (4/5): 9:1-9:11. doi: 10.1147/JRD.2017.2709578.
- ESA. (2021). *Sentinel-2 Products Specification Document*. Tilgjengelig fra: <https://sentinel.esa.int/documents/247904/685211/Sentinel-2-Products-Specification-Document> (lest 17.05.2021).
- Feurer, M. & Hutter, F. (2019). Hyperparameter Optimization. I: Hutter, F., Kotthoff, L. & Vanschoren, J. (red.) *Automated Machine Learning: Methods, Systems, Challenges*, s. 3-33. Cham: Springer International Publishing.
- Geitle, M. & Olsson, R. (2019, 2019//). *A New Baseline for Automated Hyperparameter Optimization*. Machine Learning, Optimization, and Data Science, Cham: Springer International Publishing.
- Gómez-Verdejo, V., Ortega-Moral, M., Arenas-García, J. & Figueiras-Vidal, A. R. (2006). Boosting by weighting critical and erroneous samples. *Neurocomputing*, 69 (7): 679-685. doi: 10.1016/j.neucom.2005.12.011.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016, 2016//). *Identity Mappings in Deep Residual Networks*. Computer Vision – ECCV 2016, Cham: Springer International Publishing.
- Hussain, M., Chen, D., Cheng, A., Wei, H. & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches.

- ISPRS Journal of Photogrammetry and Remote Sensing*, 80: 91-106. doi: <https://doi.org/10.1016/j.isprsjprs.2013.03.006>.
- Jiang, H., Hu, X., Li, K., Zhang, J., Gong, J. & Zhang, M. (2020). PGA-SiamNet: Pyramid Feature-Based Attention-Guided Siamese Network for Remote Sensing Orthoimagery Building Change Detection. *Remote Sensing*, 12 (3). doi: 10.3390/rs12030484.
- Kartverket. (2018). *Produktspesifikasjon FKB-Arealbruk 4.61*. Tilgjengelig fra: http://sosi.geonorge.no/Produktspesifikasjoner/Produktspesifikasjon_FKB-Arealbruk_4.61.pdf. (lest 17.05.2021)
- Kartverket. (2021). *Nasjonal detaljert høydemodell*. Tilgjengelig fra: <https://www.kartverket.no/geodataarbeid/nasjonal-detaljert-hoydemodell> (lest 17.05.2021).
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1 (3): 155-159. doi: 10.1016/S0034-4257(70)80021-9.
- Liu, W. & Yamazaki, F. (2012). Object-Based Shadow Extraction and Correction of High-Resolution Optical Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5 (4): 1296-1302. doi: 10.1109/JSTARS.2012.2189558.
- Liu, Y., Starzyk, J. & Zhu, Z. (2008). Optimized Approximation Algorithm in Neural Networks Without Overfitting. *Neural Networks, IEEE Transactions on*, 19: 983-995. doi: 10.1109/TNN.2007.915114.
- Lorenzi, L., Melgani, F. & Mercier, G. (2012). A Complete Processing Chain for Shadow Detection and Reconstruction in VHR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (9): 3440-3452. doi: 10.1109/TGRS.2012.2183876.
- Luo, S., Shen, H., Li, H. & Chen, Y. (2019). Shadow removal based on separated illumination correction for urban aerial remote sensing images. *Signal Processing*, 165: 197-208. doi: 10.1016/j.sigpro.2019.06.039.
- Luo, S., Li, H. & Shen, H. (2020). Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167: 443-457. doi: 10.1016/j.isprsjprs.2020.07.016.
- Luque, A., Carrasco, A., Martín, A. & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91: 216-231. doi: 10.1016/j.patcog.2019.02.023.
- Marcial-Pablo, M. d. J., Gonzalez-Sanchez, A., Jimenez-Jimenez, S. I., Ontiveros-Capurata, R. E. & Ojeda-Bustamante, W. (2019). Estimation of vegetation

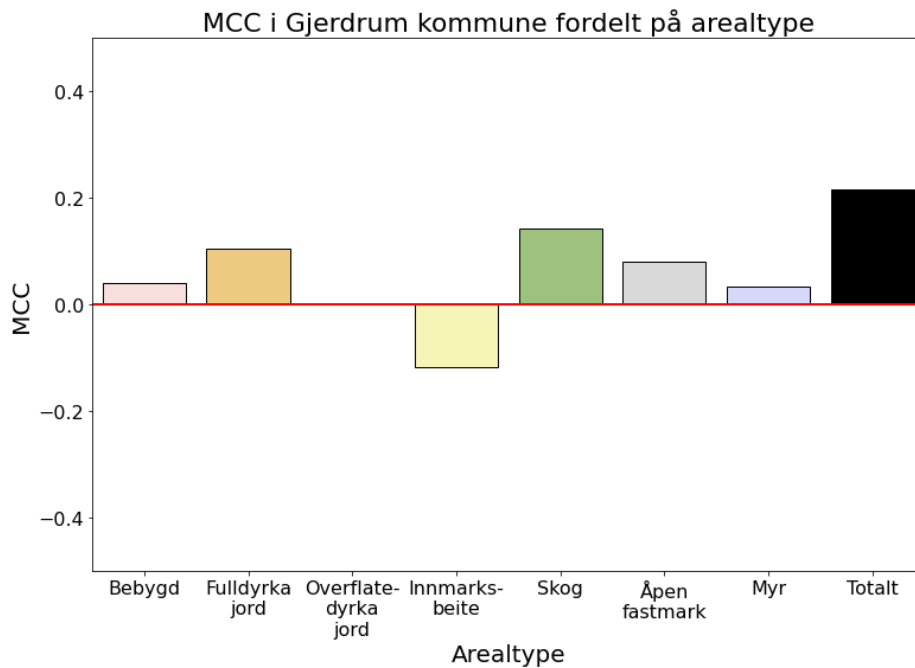
- fraction using RGB and multispectral images from UAV. *International Journal of Remote Sensing*, 40 (2): 420-438. doi: 10.1080/01431161.2018.1528017.
- Massi, M. (2019). *Autoencoder schema*.
- Maxwell, A. E. & Warner, T. A. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39 (9). doi: 10.1080/01431161.2018.1433343.
- Mostafa, Y. (2017). A Review on Various Shadow Detection and Compensation Techniques in Remote Sensing Images. *Canadian Journal of Remote Sensing*, 43 (6): 545-562. doi: 10.1080/07038992.2017.1384310.
- Mou, L., Bruzzone, L. & Zhu, X. X. (2019). Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (2): 924-935. doi: 10.1109/TGRS.2018.2863224.
- Movia, A., Beinat, A. & Crosilla, F. (2016). Shadow detection and removal in RGB VHR images for land use unsupervised classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119: 485-495. doi: 10.1016/j.isprsjprs.2016.05.004.
- NIBIO. (2017). *Snaumark*. Tilgjengelig fra: <https://www.nibio.no/tema/jord/arealressurser/ar250/snaumark> (lest 25.03.2021).
- NIBIO. (2021a). *Gårdskart*. Tilgjengelig fra: <https://www.nibio.no/tema/jord/arealressurser/arealressurskart-ar5/g%C3%A5rdskart> (lest 06.05.2021).
- NIBIO. (2021b). *Kontinuerlig ajourhold*. Tilgjengelig fra: <https://www.nibio.no/tema/jord/arealressurser/arealressurskart-ar5/kontinuerlig-ajourhold> (lest 18.04.2021).
- NIBIO. (2021c). *Periodisk ajourhold*. Tilgjengelig fra: <https://www.nibio.no/tema/jord/arealressurser/arealressurskart-ar5/periodisk-ajourhold> (lest 18.04.2021).
- Powers, D. M. W. (2007). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation*: School of Informatics and Engineering, Flinders University of South Australia
- Raschka, S. & Mirjalili, V. (2017). *Python Machine Learning*. Second utg. Birmingham, UK: Packt Publishing.
- Ronneberger, O., Fischer, P. & Brox, T. (2015, 2015//). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham: Springer International Publishing.

- Rusiecki, A. (2019). Trimmed categorical cross-entropy for deep learning with label noise. *Electronics Letters*, 55 (6): 319-320. Tilgjengelig fra: <https://digital-library.theiet.org/content/journals/10.1049/el.2018.7980>.
- Sarabandi, P., Yamazaki, F., Matsuoka, M. & Kiremidjian, A. (2004, 20-24 Sept. 2004). *Shadow detection and radiometric restoration in satellite high resolution images*. IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium.
- Schowengerdt, R. A. (2006). *Remote Sensing: Models and Methods for Image Processing*. Elsevier Science.
- Snoek, J., Larochelle, H. & Adams, R. P. (2012). *Practical Bayesian optimization of machine learning algorithms*. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2, Lake Tahoe, Nevada, s. 2951–2959: Curran Associates Inc.
- Stivaktakis, R., Tsagkatakis, G. & Tsakalides, P. (2019). Deep Learning for Multilabel Land Cover Scene Categorization Using Data Augmentation. *IEEE Geoscience and Remote Sensing Letters*, 16 (7): 1031-1035. doi: 10.1109/LGRS.2019.2893306.
- Storey, E. A., Stow, D. A., Coulter, L. L. & Chen, C. (2017). Detecting shadows in multi-temporal aerial imagery to support near-real-time change detection. *GIScience & Remote Sensing*, 54 (4): 453-470. doi: 10.1080/15481603.2017.1279729.
- Terratec. (2016a). *Ortofotorapport Omløpsfotografering Oslo 2016*.
- Terratec. (2016b). *Ortofotorapport Omløpsfotografering Østlandet 2016*.
- Townshend, J. R. G. & Justice, C. O. (1986). Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, 7 (11): 1435-1445. doi: 10.1080/01431168608948946.
- Tsai, V. J. (2006). A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE transactions on geoscience and remote sensing*, 44 (6): 1661-1671.
- Wikimedia Commons bruker Aphex34. (2015). *Typical CNN architecture*.
- Yao, J. & Shepperd, M. (2020). *Assessing software defect prediction performance: why using the Matthews correlation coefficient matters*. Proceedings of the Evaluation and Assessment in Software Engineering, Trondheim, Norway, s. 120–129: Association for Computing Machinery.
- Yu, X.-H. & Chen, G.-A. (1997). Efficient Backpropagation Learning Using Optimal Learning Rate and Momentum. *Neural Networks*, 10 (3): 517-527. doi: 10.1016/S0893-6080(96)00102-5.
- Zhan, Y., Fu, K., Yan, M., Sun, X., Wang, H. & Qiu, X. (2017). Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images. *IEEE*

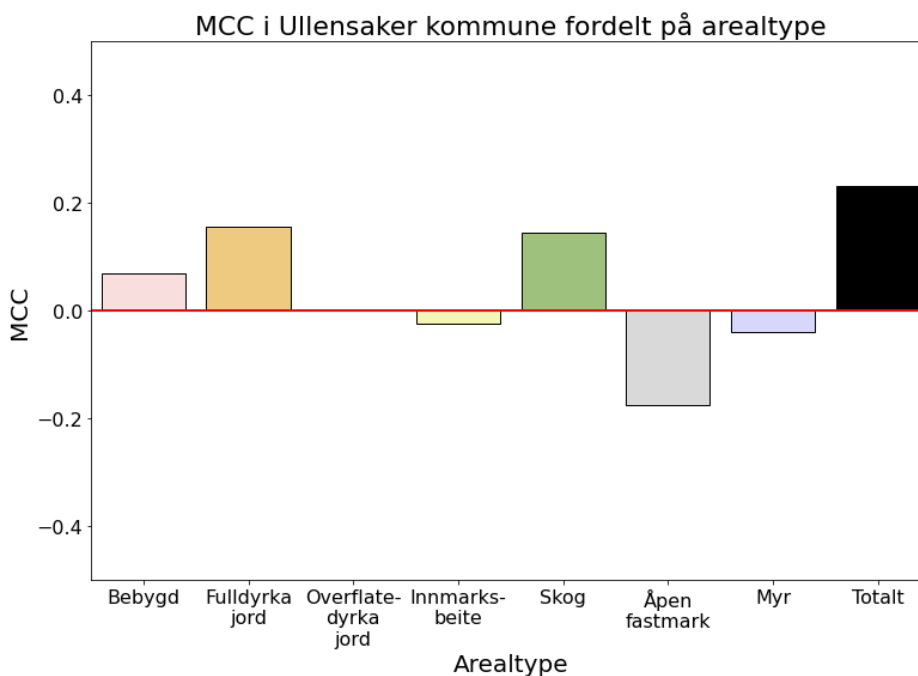
Geoscience and Remote Sensing Letters, 14 (10): 1845-1849. doi:
10.1109/LGRS.2017.2738149.

8 Appendiks

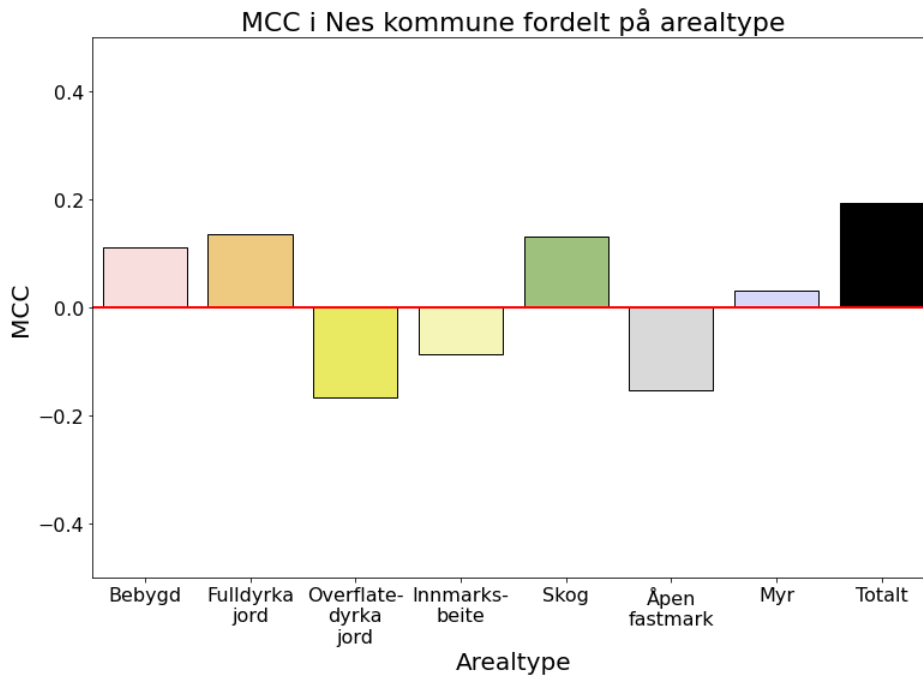
8.1 Resultater for kommuner fordelt på arealtype



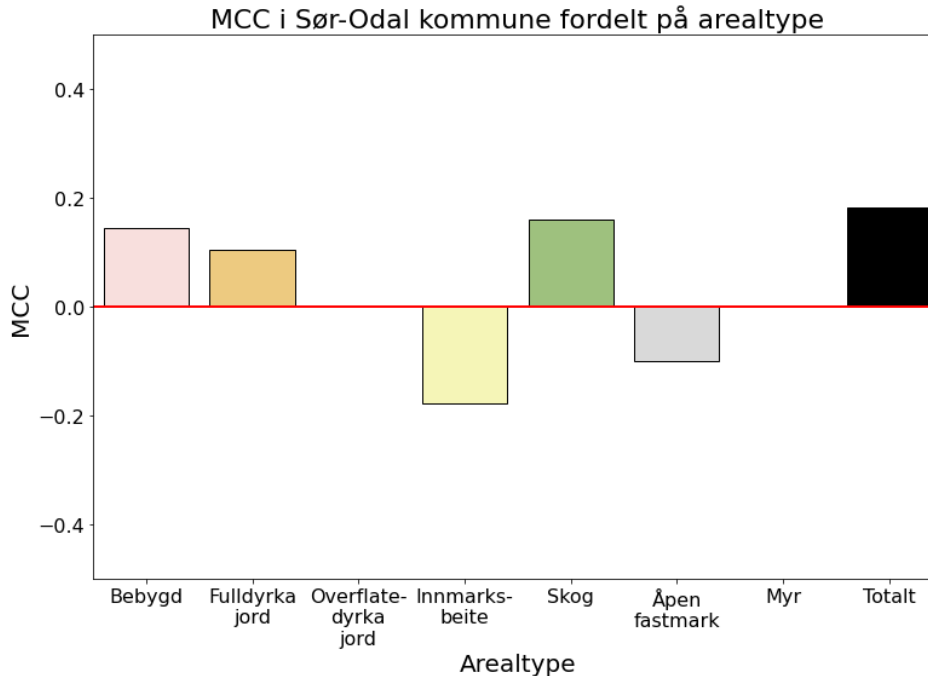
Figur 27. Matthews korrelasjonskoeffisient til prediksjonene i Gjerdrum kommune fordelt på arealtype. Det var ingen observasjoner med arealtypen overflatedyrka mark i Gjerdrum, og MCC er derfor udefinert for denne arealtypen.



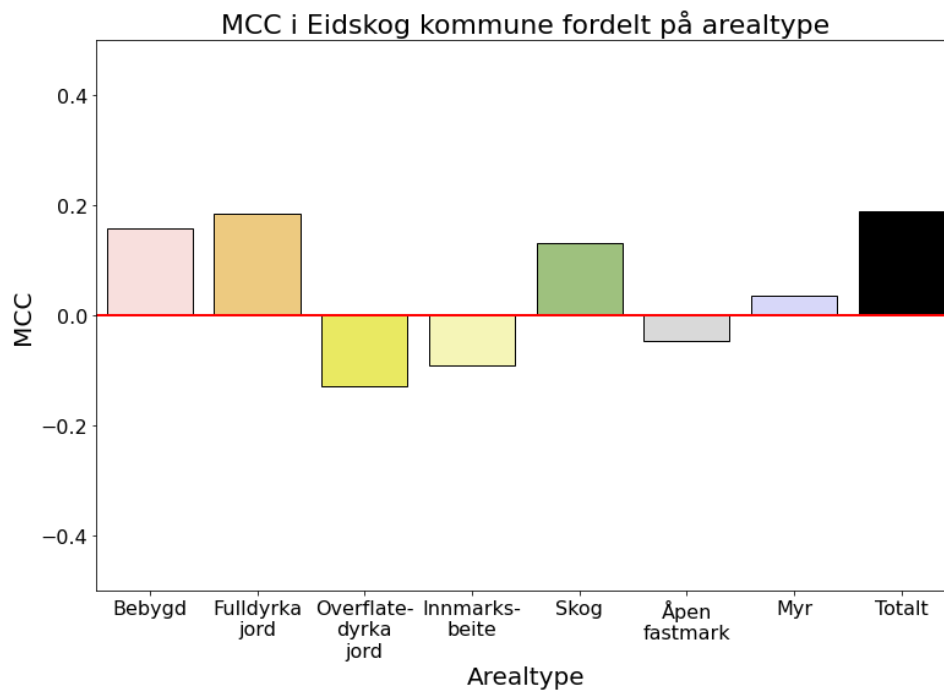
Figur 28. Matthews korrelasjonskoeffisient til prediksjonene i Ullensaker kommune fordelt på arealtype. Det var ingen negative observasjoner av overflatedyrka mark i Ullensaker, og MCC er derfor udefinert for denne arealtypen.



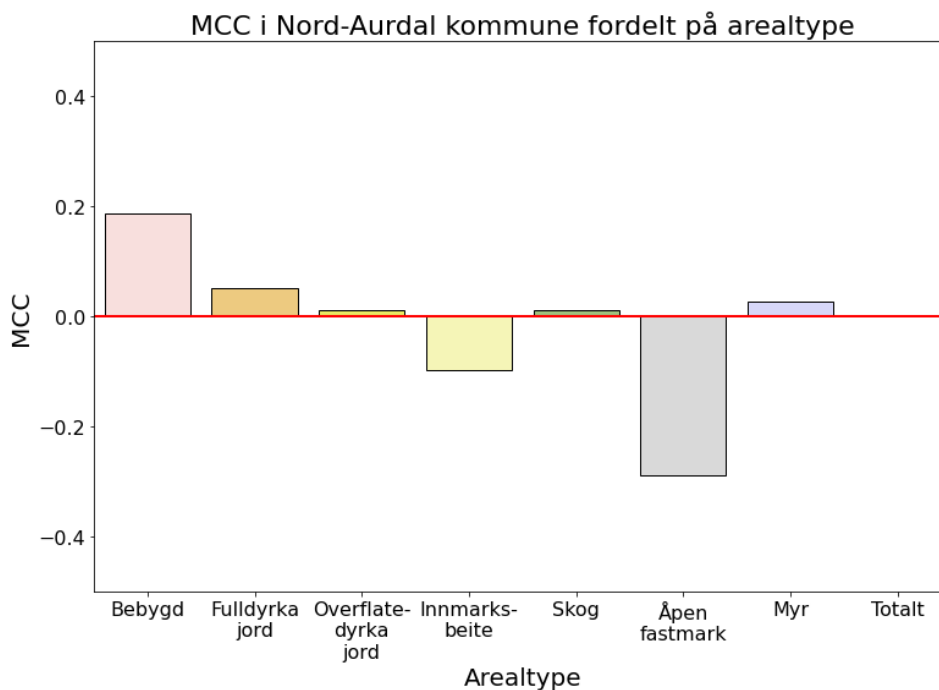
Figur 29. Matthews korrelasjonskoeffisient til prediksjonene i Nes kommune fordelt på arealtype.



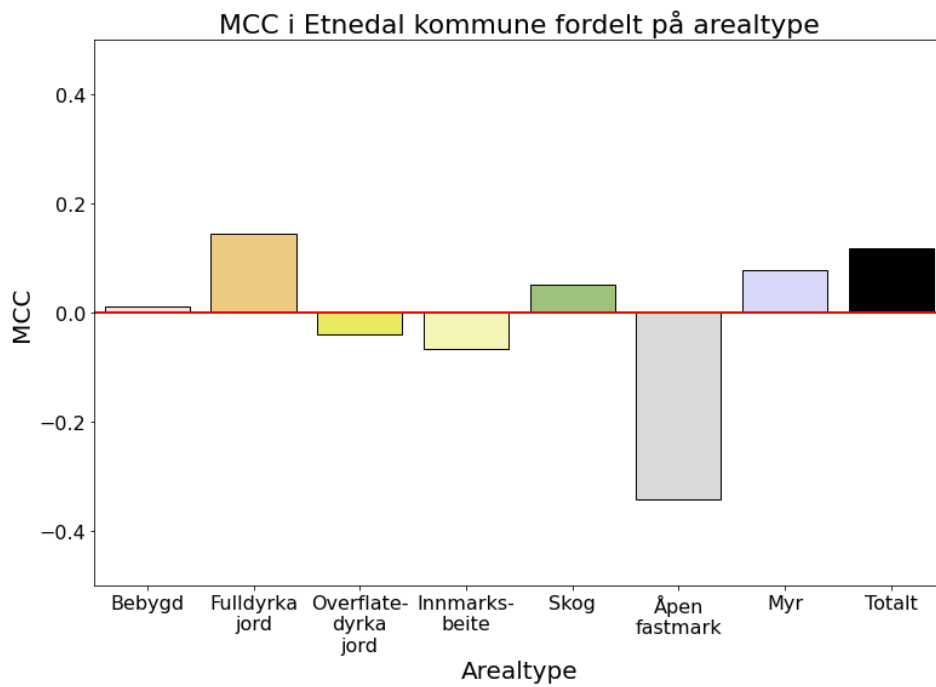
Figur 30. Matthews korrelasjonskoeffisient til prediksjonene i Sør-Odal kommune fordelt på arealtype. Det var ingen negative observasjoner av overflatedyrka mark i Sør-Odal, og MCC er derfor udefinert for denne arealtypen.



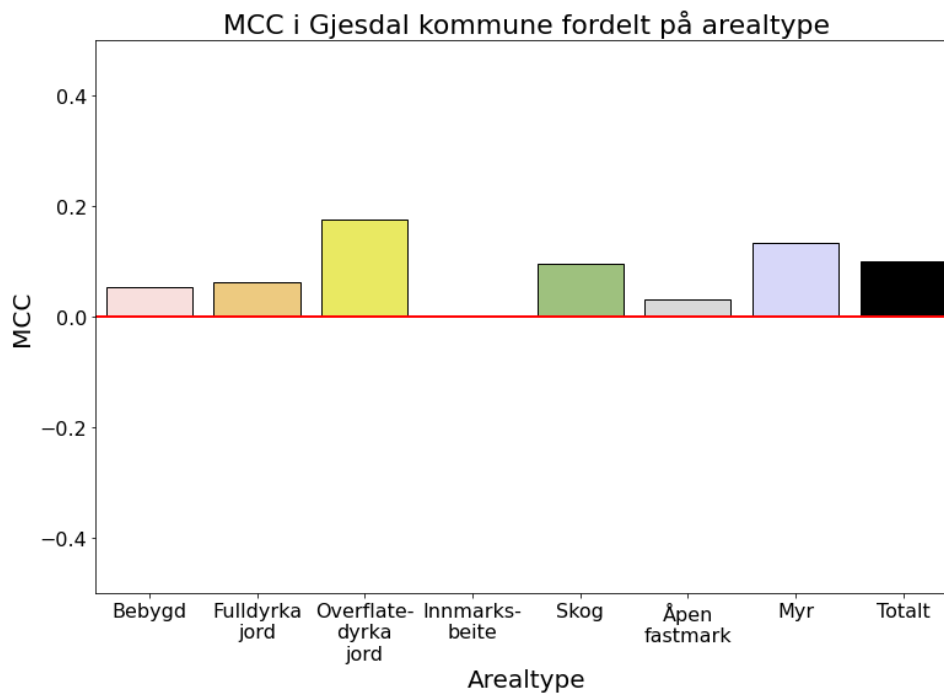
Figur 31. Matthews korrelasjonskoeffisient til prediksjonene i Eidskog kommune fordelt på arealtype.



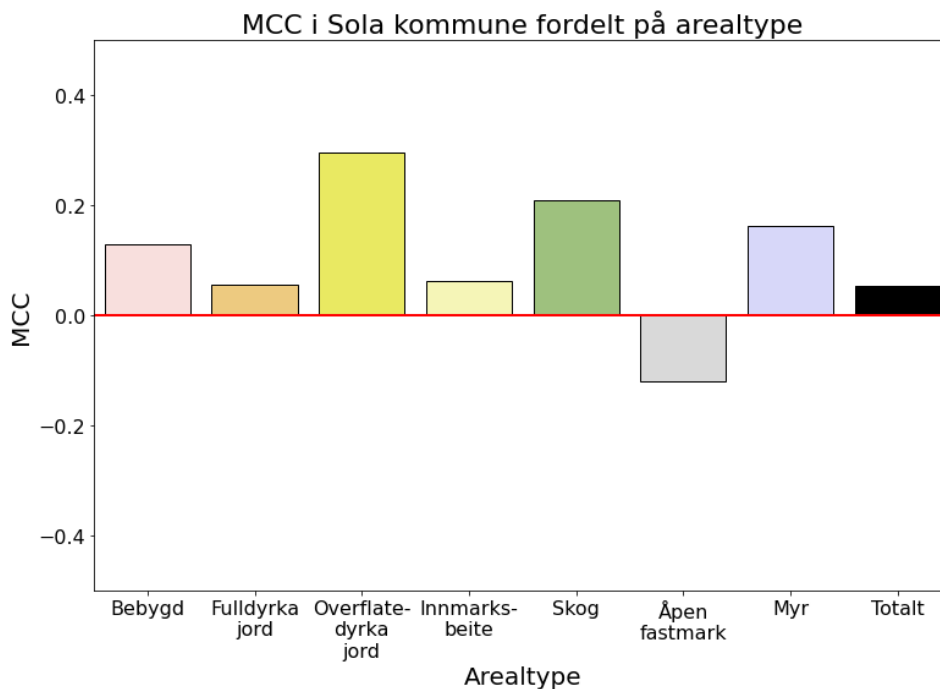
Figur 32. Matthews korrelasjonskoeffisient til prediksjonene i Nord-Aurdal kommune fordelt på arealtype.



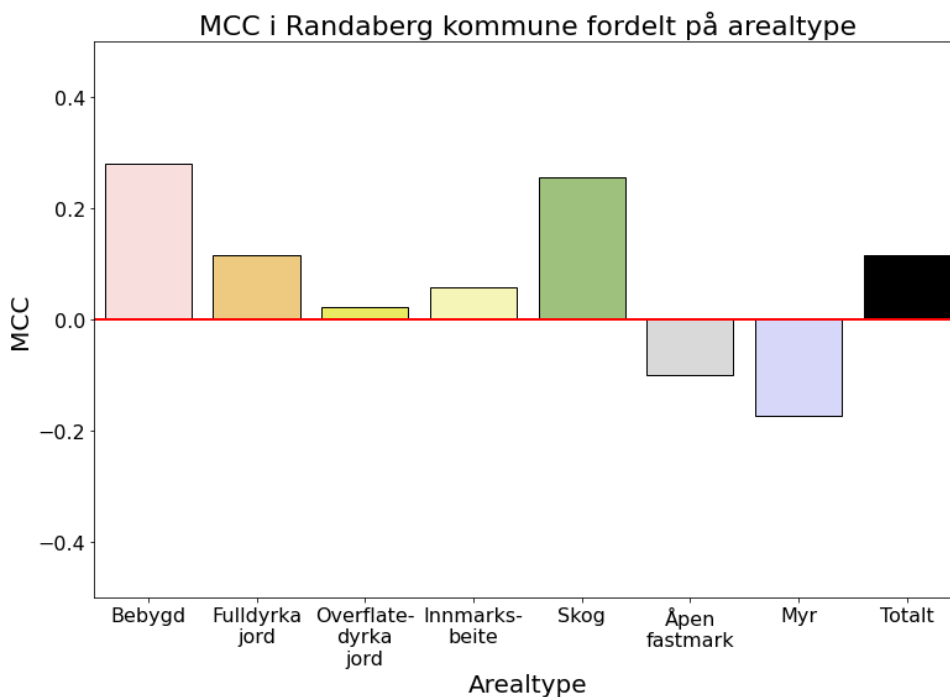
Figur 33. Matthews korrelasjonskoeffisient til prediksjonene i Etnedal kommune fordelt på arealtype.



Figur 34. Matthews korrelasjonskoeffisient til prediksjonene i Gjesdal kommune fordelt på arealtype.

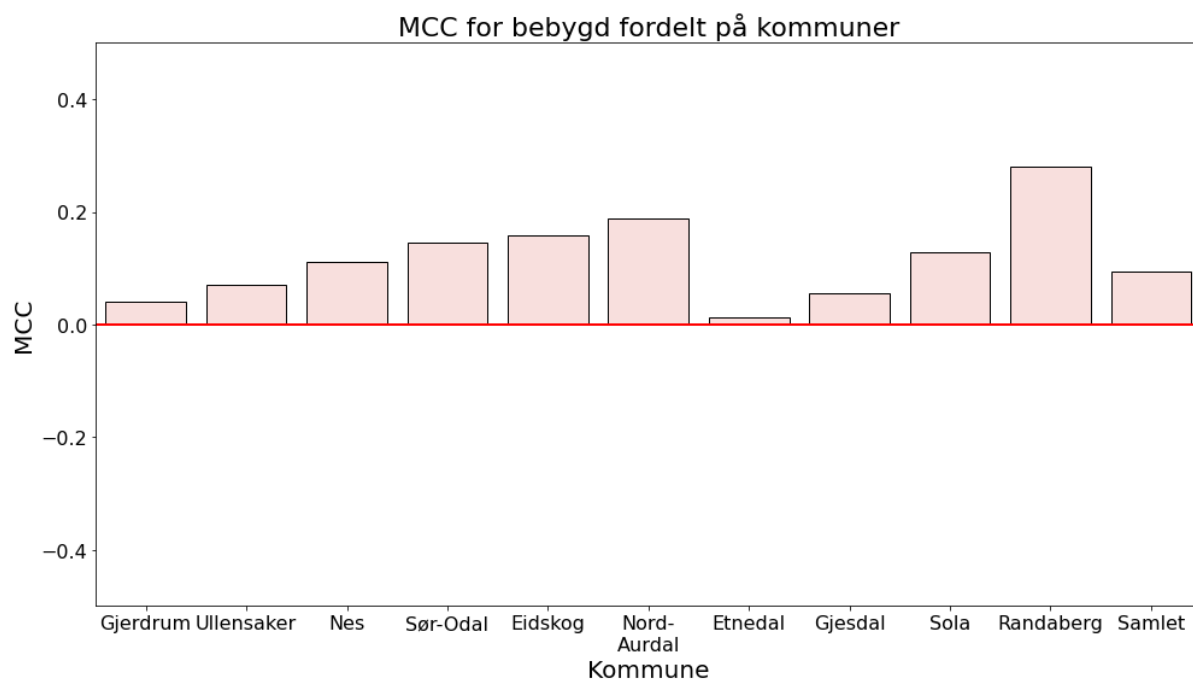


Figur 35. Matthews korrelasjonskoeffisient til prediksjonene i Sola kommune fordelt på arealtype.

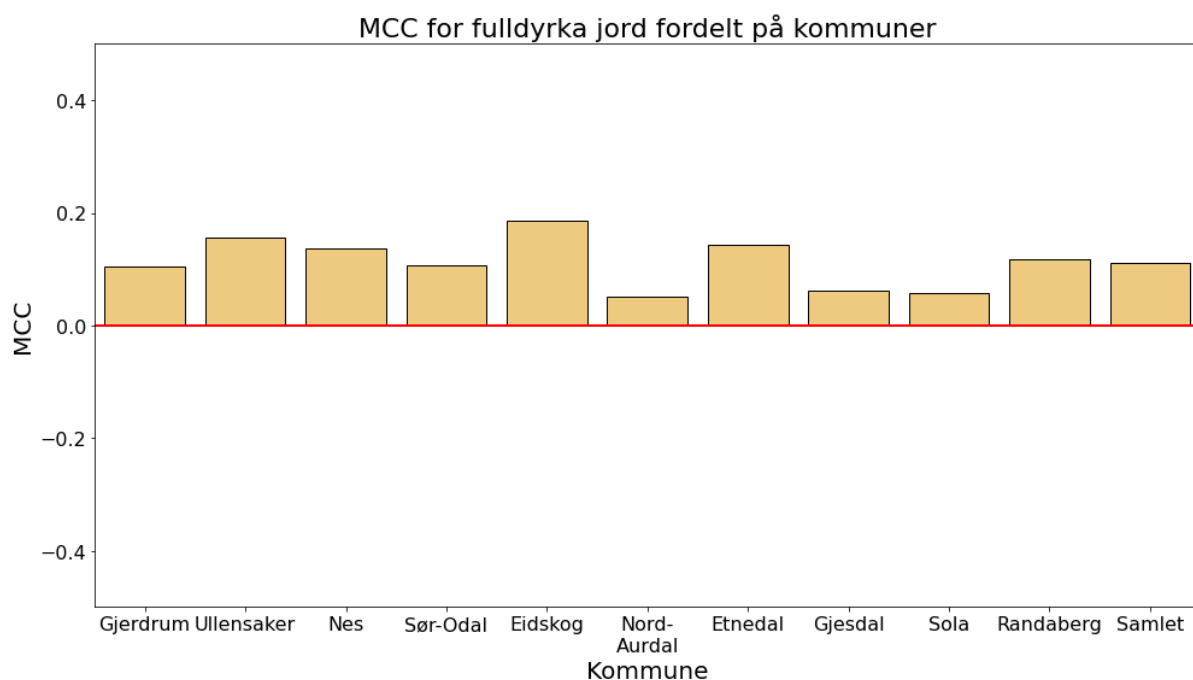


Figur 36. Matthews korrelasjonskoeffisient til prediksjonene i Randaberg kommune fordelt på arealtype.

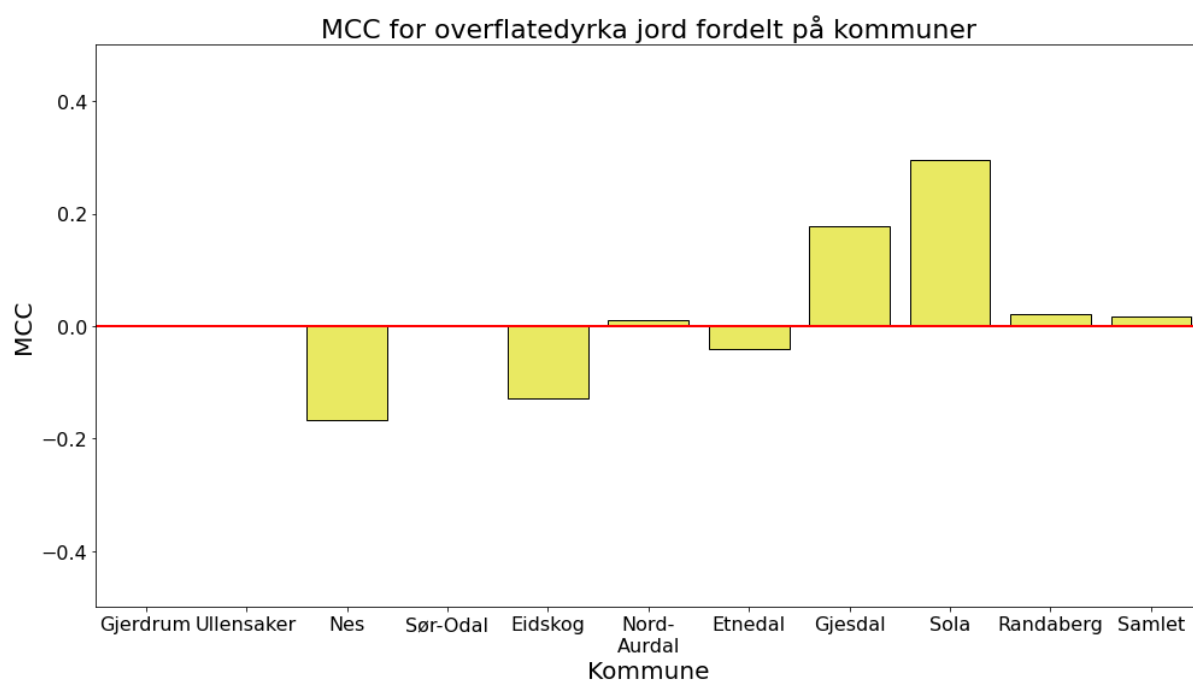
8.2 Resultater for arealtype fordelt på kommuner



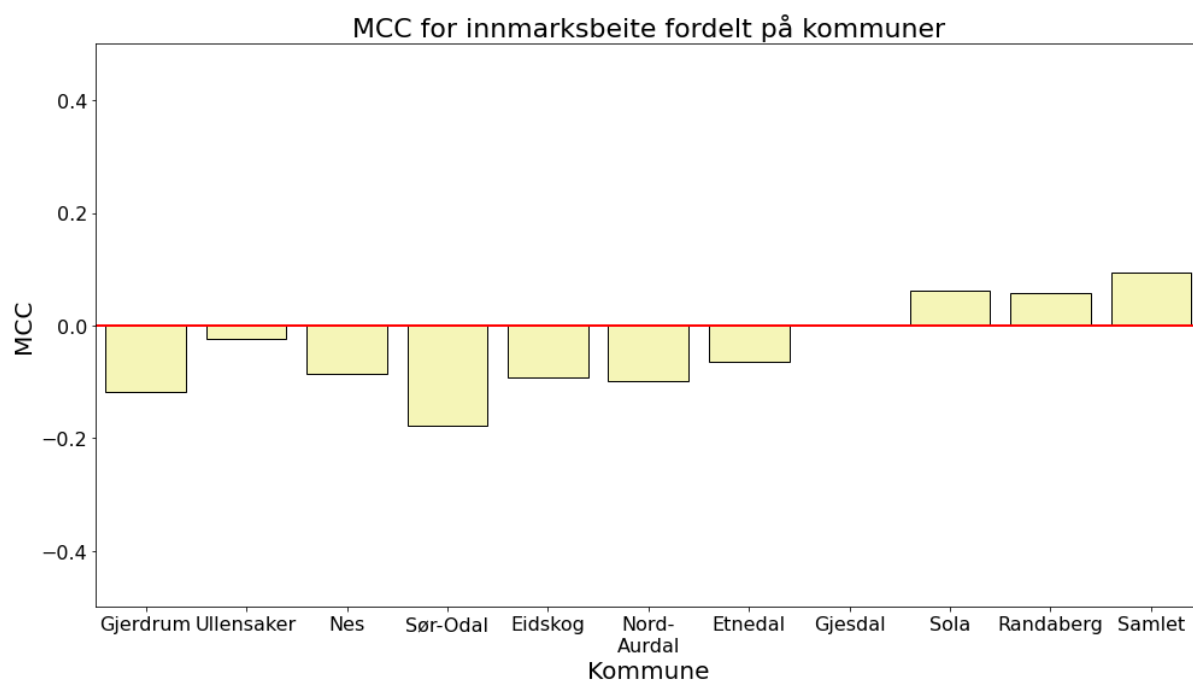
Figur 37. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen bebygd, fordelt på kommune.



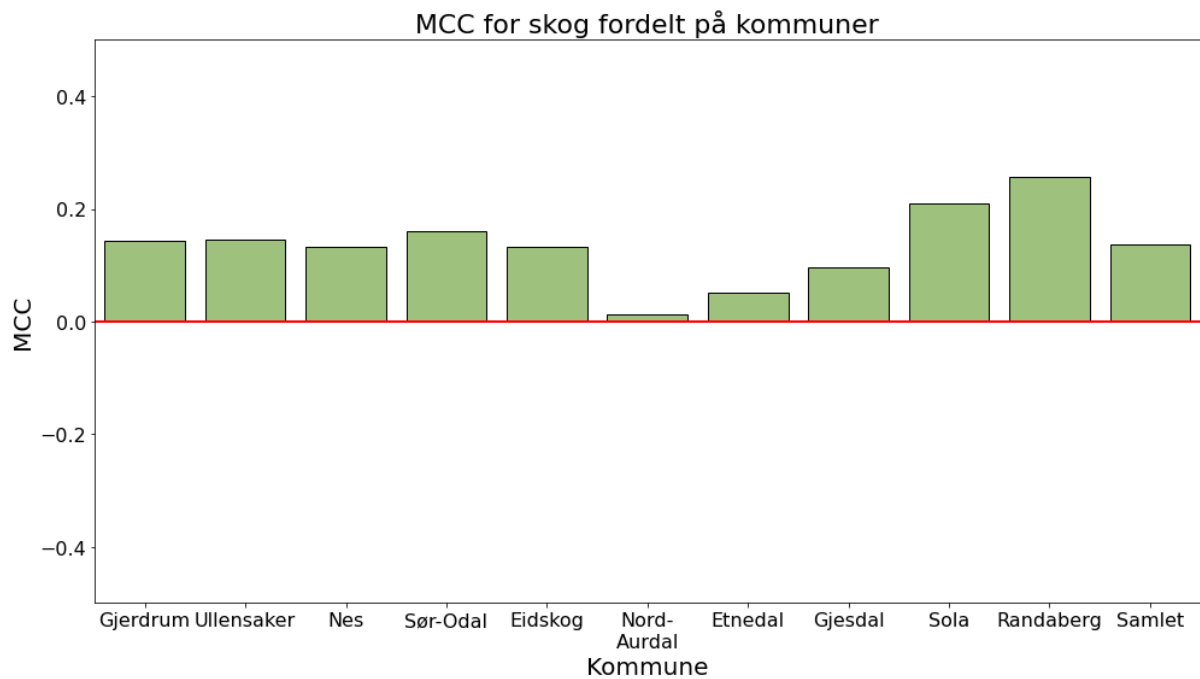
Figur 38. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen fulldyrka jord, fordelt på kommune.



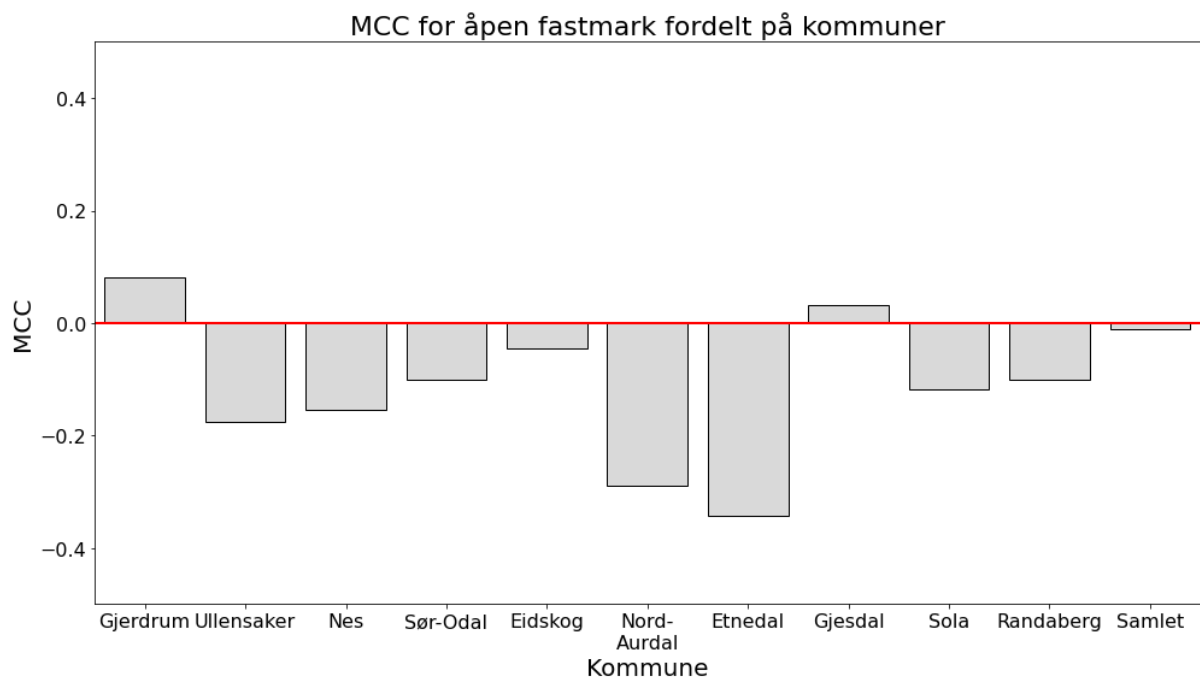
Figur 39. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen overflatedyrka jord, fordelt på kommune.



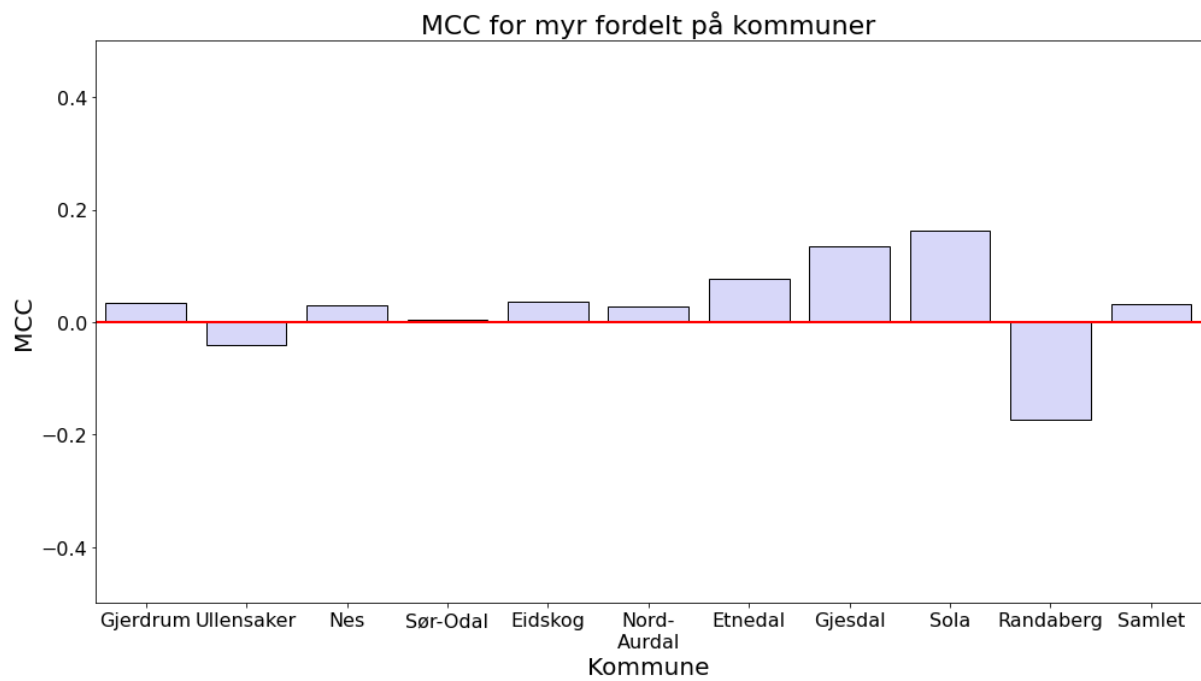
Figur 40. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen innmarksbeite, fordelt på kommune.



Figur 41. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen skog, fordelt på kommune.

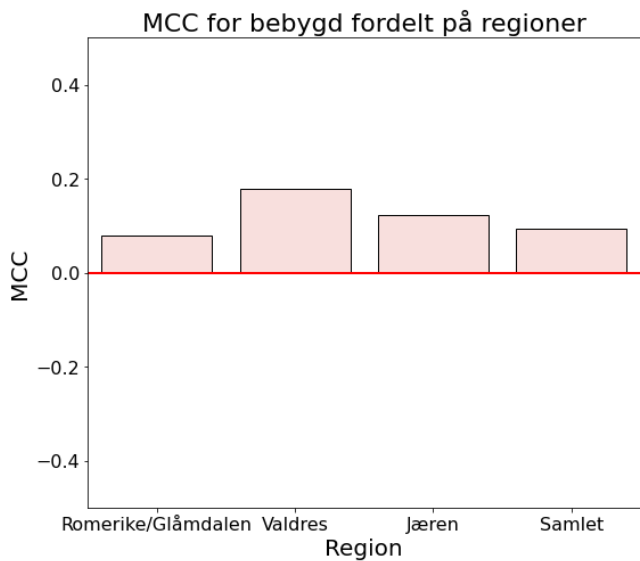


Figur 42. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen åpen fastmark, fordelt på kommune.

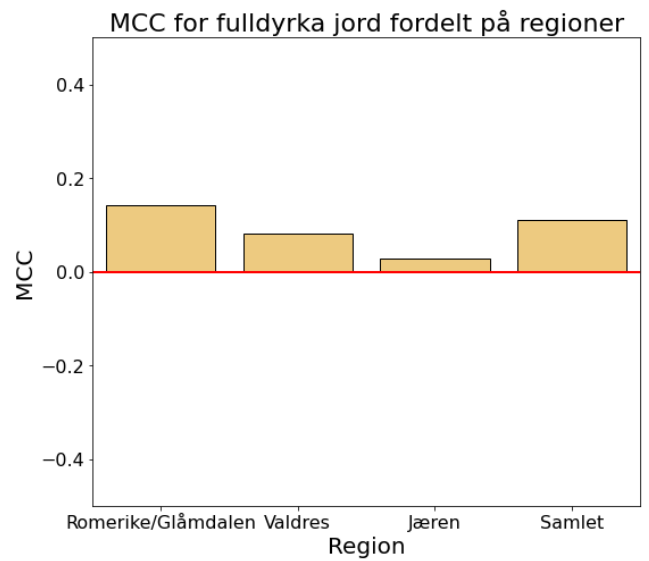


Figur 43. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen myr, fordelt på kommune.

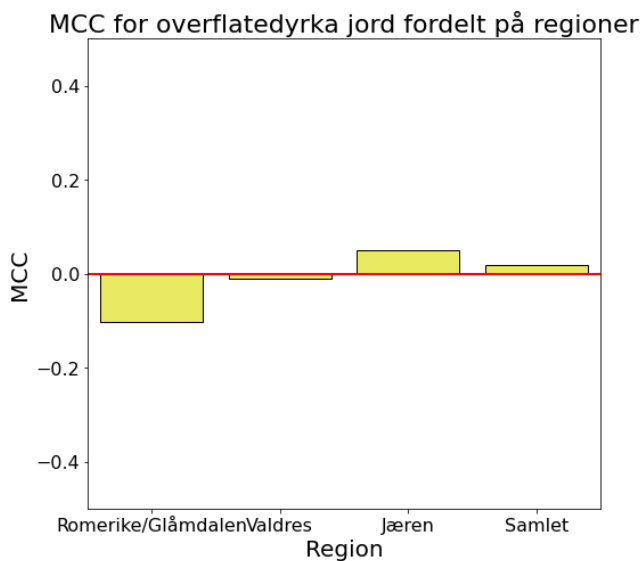
8.3 Resultater for arealtype fordelt på region



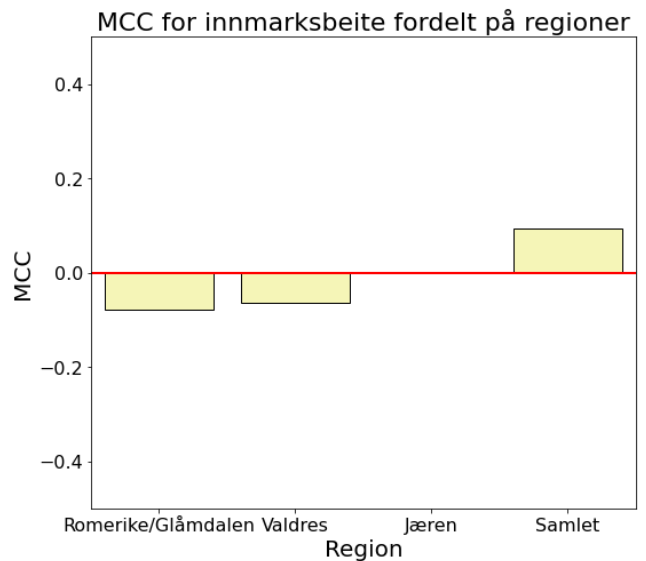
Figur 44. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen bebygd, fordelt på region.



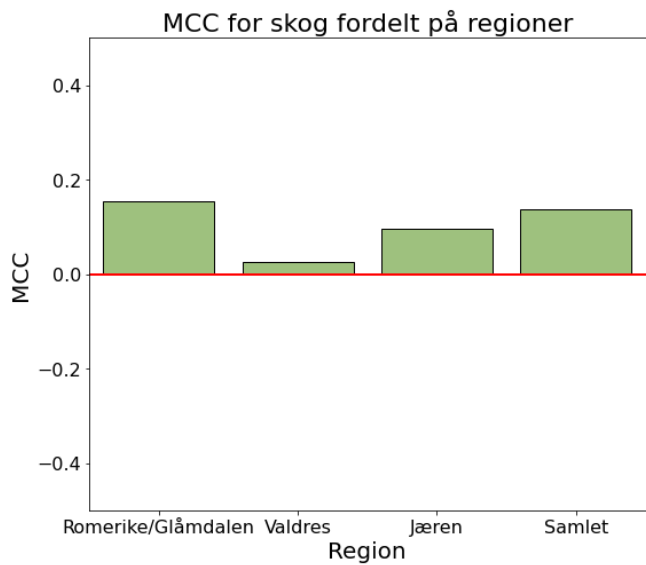
Figur 45. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen fulldyrka jord, fordelt på region.



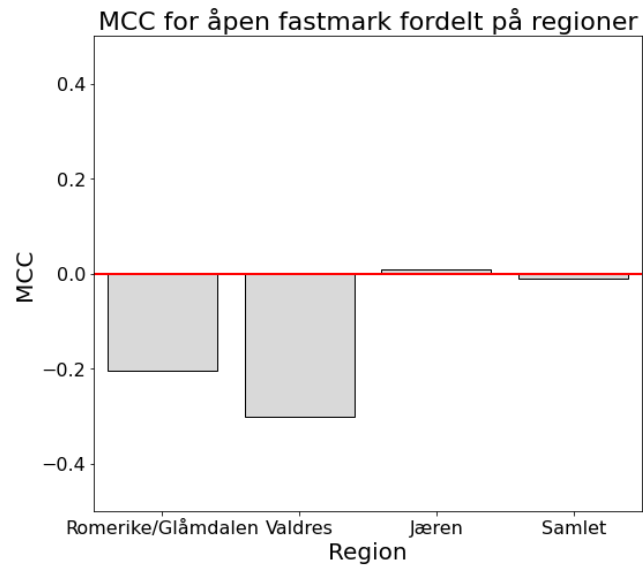
Figur 46. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen overflatedyrka jord, fordelt på region.



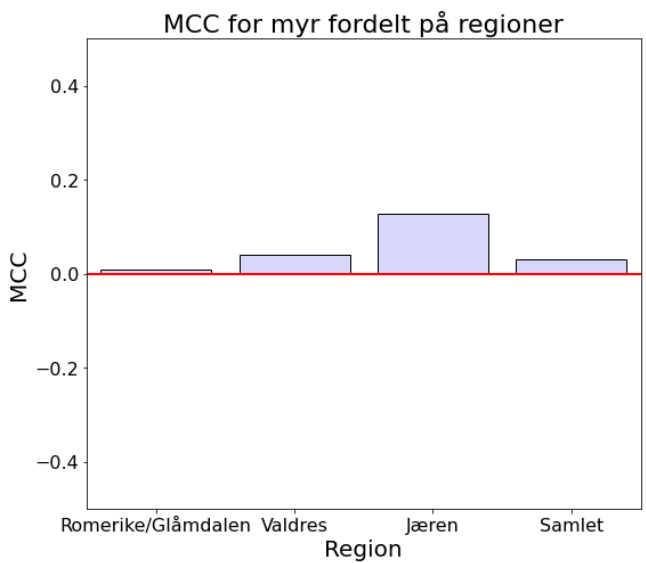
Figur 47. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen innmarksbeite, fordelt på region.



Figur 44. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen skog, fordelt på region.



Figur 46. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen åpen fastmark, fordelt på region.



Figur 45. Matthews korrelasjonskoeffisient til prediksjonene med arealtypen myr, fordelt på region.



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

Postboks 5003
NO-1432 Ås
Norway