

Digitalt Skovkort – Dokumentation for arealkortet

Styrelsen for Grøn Arealomlægning og Vandmiljø

2025 - Marts

Udgiver: Styrelsen for Grøn Arealomlægning og Vandmiljø

Redaktion:
Enhederne Skov & Digitalisering

ISBN: 978-87-85306-98-2

Indhold

1.	Formål	6
1.1	Behov for et arealkort over de danske skove	6
1.2	Formål for og anvendelse af arealkortet	7
1.2.1	Understøttelse af myndighedernes dataindsamling vedr. skov	7
1.2.2	Internationale rapporter	7
2.	Afgrænsning	8
2.1	Skovdefinitionen	8
2.2	Skovrejsning og arealer uden træbevoksning	8
2.3	Lysåbne arealer	8
2.4	Andet træbevokset areal	8
2.5	Træer uden for skovene	9
2.6	Fredskov	9
3.	Grundlag for AI-Model	10
3.1	Remote Sensing til Arealforvaltning	10
3.2	Geografisk opdeling af kortlag anvendt til maskinlæringsmodellen	10
3.2.1	Geospatielt indeks.	10
3.3	Datagrundlag	11
3.3.1	Grunddata - Sentinel 2	11
3.3.2	Afledte data-features – Sentinel 2	11
3.3.3	Filtrering af data – Sentinel 2	13
3.3.4	Sentinel-2 Kompositstrategi	13
3.3.5	Tidsseriedata – Sentinel 2	14
3.3.6	Fænologi-data	14
3.3.7	Opsummering	14
3.4	Model-metode	14
3.5	Forklarbarhed og Gennemsigtighed	16
3.5.1	Feature Importance i LightGBM	16
3.5.2	SHAP Værdier	17
3.6	Træning af maskinlæringsmodel	17
3.7	Træningsdata og metode for udvælgelse af træningsdata	17
3.7.1	Træpixels	17
3.7.2	Ikke-træ pixels	19
3.7.3	Afbalancering mellem positiv og negative træningspunkter	19
4.	For- og efterbehandling	20
4.1	Afskæringslag før modellen prædikerer træ-dække	20
4.1.1	Landbrugsarealer	20
4.1.2	GeoDanmark - befæstede arealer	21
4.1.3	Byzoner og sommerhusområder	21
4.2	Efterbehandling	22
	Efter maskinlæringsmodellen har fundet alle pixels med trædække, inden for det afgrænsede område, skal disse pixels omdannes til polygoner som lever op til skovdefinitionens afgrænsninger.	22
4.2.1	Fra pixel til polygon	22

4.2.2	Veje	22
5.	Resultater og kvalitetssikring	24
5.1	Resultater fra maskinevaluering	24
5.1.1	Confusion Matrix	24
5.2	Sammenligning med Referencekort	25
5.2.1	Corine	25
5.2.2	Basemap	26
5.2.3	Generel Statistik	26
5.2.4	Unikke Polygoner	28
5.2.5	Regionale Forskelle	30
5.3	Faglig evaluering	31
5.3.1	Trædækkegenkendelse	31
5.3.2	For- og efterbehandling	32
5.4	Fremtidig udvikling	33
	Bilag 1. Supplerende data	33
Bilag 1.1	Afgrødekoder til udlægning af ikke-træ pixels.	33
Bilag 1.2	Habitatnaturtyper til ikke-træer pixels.	37
6.	Bibliography	40
Bilag 1.3	Litteraturliste.	42

Figur-oversigt:

Figur 3.1	Illustration af et beslutningstræ (What is a decision tree?, 2024)	15
Figur 3.2:	<i>LightGBM – "Leaf-wise tree growth"</i>	16
Figur 3.3:	<i>De 10x10km kvadrater hvori der er udlagt polygoner for træ-dække og prikker for ikke-trædække</i>	18
Figur 3.4:	<i>Mosaics grid_polyfill funktion</i>	19
Figur 4.1:	<i>Grafisk oversigt over dannelse af færdigt skovkort.</i>	20
Figur 4.2:	<i>Illustration af metoder til efterbehandling af prædiktionskortlag. Dette er et eksempel, og vejene stemmer ikke overens, med hvilke veje der bliver brugt til arealkortet.</i>	23
Figur 5.1	De individuelle polygoner fra CORINE datasættet indenfor de 4 udvalgte kategorier samles til sammenhængende skovpolygoner	25
Figur 5.2	Totalt skovareal (hektar) af CORINE, Basemap og SGAV Skovkort	26
Figur 5.3:	Fordeling af størrelserne på de enkelte skov polygoner.	27
Figur 5.4:	Kumulativ fordeling af de enkelte polygon arealer	27
Figur 5.5:	Det største skov polygon i SGAVs Skovkort (Blå), CORINE (Grøn) og Basemap (Orange)	28

Figur 5.6: Antal hektar ét kort (venstre) kan bidrage med til et andet kort (bunden), samt hvor stor en procentdel af referencekortet (bunden) dette areal udgør 29

Figur 5.7: Skovpolygoner fra Basemap der ikke har noget overlap med SGAVs skovkort (venstre) og SGAVs afskæringslag (højre). Områder der forudsiges i er grønne, og ellers røde. 30

Figur 5.8: Regionalt landkort farvelagt efter Jaccard indeks mellem SGAV og CORINE 30

Figur 5.9: Regionalt landkort farvelagt efter Jaccard index mellem SGAV og CORINE 31

1. Formål

1.1 Behov for et arealkort over de danske skove

Der er behov for et årligt opdateret arealkort over de danske skove, da der indtil nu ikke har eksisteret et samlet entydigt skovkort i Danmark, men derimod flere forskellige kort og statistiske opgørelser. De forskellige kort og data har arbejdet med forskellige metoder og definitioner, og resulterede er derfor ikke en entydig opgørelse af skovenes udbredelse og areal i Danmark. Indtil nu har der været følgende opgørelser:

- National Forest Inventory (NFI). Den stikprøvebaserede dataindsamling, som danner grundlag for den danske skovstatistik, som udarbejdes af Styrelsen for Grøn Arealomlægning og Vandmiljø (SGAV). Denne statistik anvender Food and Agriculture Organization of the United Nations' (FAO's) internationale definition for skov (se afsnit 2.1). Da NFI'en er baseret på stikprøveflader indeholder den ikke en arealbaseret kortlægning (SGAV, 2025).
- Basemap. Kortet er udarbejdet af Aarhus Universitet (AU) for Danmarks Statistik og er et arealanvendelseskort over Danmark. Dette er grundlag for et grønt nationalregnskab, hvor skovudbredelse også fremgår. Basemap anvender ikke FAO's definition på skov og indeholder således træbevoksede arealer som ikke er skov, samt arealer, der falder inden for FAO's skovdefinition, men som ikke er træbevoksede (fx hjælpearealer som brandlinjer, læggepladser mv i skov). AU beskriver deres metode som in situ observationer og luftfotofortolkninger (Levin, 2024).
- GeoDanmark. Et arealbaseret kortlag fra Klimadatastyrelsen (KDS), der er grunddatamyndighed. Kortene viser træbevokset område, og har en bredere fortolkning end FAOs definition for skov. Datasættet vedligeholdes systematisk ud fra årlige luftfotos om foråret inden løvspring med en 5-årig cyklus for totalajourføring i Danmark (GeoDanmark, 2025).
- Fredskovsarealet. Et kortlag, som omfatter skovarealer, der er noteret som fredskov i matriklen. Dette kort dækker alene de fredskovspligtige skovarealer i Danmark. Der indgår i kortlaget nogle lysåbne naturområder, som ikke er skov. Kortlaget vedligeholdes gennem sagsbehandling og ved fredskovspåbyggelser, jf. skovlovens regler og er grundlag for skovlovens administration. Det anslås, at kortlaget dækker ca. 70 % af det danske skovareal (SGAV, 2024).
- Coordination of Information on the Environment (CORINE) Land Cover. Et produkt fra EU/Det Europæiske Miljøagentur (EEA) Copernicus Land Monitoring Service, som kortlægger landuse på tværs af Europa. Det landfaste areal i Europa er inddelt i 44 tematiske kategorier og indeholder blandt andet information om områder med nåle- og løvtræer. CORINE kortet opdateres hvert 6. år, og det nyeste kort er fra 2018. Kortet for 2024 forventes udgivet i Q1 af 2026. (Europa-Kommissionen, 2024).
- Copernicus Land Monitoring Service (CLMS) – Skov og trædækkkort. Et andet produkt fra EU/EEA der kortlægger trædække og skov. Det seneste er fra 2018 og opdateres hvert 3. år. EEA har planer om at opdatere årligt, men det er stadig usikkert, hvornår de kommer op på den opdateringsfrekvens og hvor lang tid der er mellem år og udgivelse (EEA, 2025).

1.2 Formål for og anvendelse af arealkortet

Arealkortet skal fremover danne grundlag for opgørelsen af Danmarks officielle skovareal og rammen for statistisk overvågning af de danske skove, herunder understøtte det politiske beslutningsgrundlag for skov- og naturområdet. Arealkortet er første trin i projektet Digitalt Skovkort, der senere bliver udbygget med nål/løv genkendelse og biomasse. Denne dokumentation omhandler arealkortet.

1.2.1 Understøttelse af myndighedernes dataindsamling vedr. skov

- Styrelsen for Grøn Arealomlægning og Vandmiljø: Digitalt Skovkort vil understøtte Den Nationale Skovovervågning. Derudover kan Digitalt Skovkort understøtte EU-forordningen om afskovningsfri forsyningskæder (EUDR). Desuden forventes Digitalt Skovkort også at kunne understøtte naturdirektiverne og EU's naturgenopretningsforordning. Der er fra EU-kommisionens side stillet forslag om en ny EU-forordning om skovovervågning, som forventes at stille specifikke krav til indsamling af data om skovene, herunder kortlægning.
- Energistyrelsen: Digitalt Skovkort kan understøtte håndhævelse af VEIII-direktivet om skærpede krav til bæredygtig biomasse til energi fra skov. Direktivet indeholder bl.a. krav om beskyttelse af visse typer skove, krav om gentilplantning, begrænsninger på afdriftsstørrelser mv.
- Klima-, Energi- og Forsyningsministeriet (KEFM): Digitalt Skovkort kan bidrage til klimareporteringer (Kyotoprotokollen og LULUCF-rapportering). Skovarealet indgår desuden i den årlige nationale klimafremskrivning, som har ophæng i klimaloven.

Digitalt Skovkort vil kunne indgå i andet kommende understøttende, rapporterende og planlæggende myndighedsarbejde.

1.2.2 Internationale rapporteringer

Kortets areal er grundlag for internationale rapporteringer. Disse omfatter bl.a. FN hvor FAO laver en global skovopgørelse (Forest Resource Assessment), det europæiske samarbejde Forest Europe samt EUROSTAT, hvor en ny forordning om miljøøkonomiske regnskaber stiller krav om, at der opstilles nationale skovregnskaber.

2. Afgrænsning

Til afgrænsning af arealkortet er brugt forskellige definitioner og kortlag, som hjælper og præciserer skovkortlægningen.

2.1 Skovdefinitionen

Arealkortet er baseret på den internationale skovdefinition fra FAO, som er FN's fødevarer og landbrugsorganisation. I en dansk sammenhæng lyder definitionen:

Skov: Et areal større end 0,5 hektar med en minimumsbredde på 20 meter, bevokset med træer højere end 5 meter og et kronedække på mere end 10 pct. eller med træer, der potentielt er i stand til at nå disse værdier på voksestedet. Definitionen inkluderer midlertidigt ubevoksede arealer og hjælpearealer nødvendige for skovdriften, men ikke arealer domineret af landbrugs- eller bymæssig anvendelse, herunder sommerhusområder (Nord-Larsen et al., 2022, s. 5).

Juletræer, uanset om de dyrkes inde i skoven (omkranset af skov) eller på markjord, indgår i ovenstående definition. Det skyldes at træerne potentielt kan blive mere end 5 meter høje. Frugtplantager indgår ikke i definitionen, da de opfattes som en landbrugsafgrøde.

Herunder uddybes hvordan SGAV har implementeret skovdefinitionen i Digitalt Skovkort.

2.2 Skovrejsning og arealer uden træbevoksning

FAOs skovdefinition omfatter midlertidigt ubevoksede arealer (fx afdriftsarealer, der endnu ikke er tilplantet) samt hjælpearealer i skovdriften (fx brandbælter og læggepladser). Modellen er ikke trænet til at kunne finde disse arealer, ligesom den heller ikke er trænet til at kunne finde nyligt tilplantede arealer (fx skovrejsning på landbrugsjord). Disse arealer vil typisk ikke fremgå af skovkortet. Nytilplantede arealer kommer med i et opdateret Digitalt Skovkort efterhånden, som træerne vokser til.

2.3 Lysåbne arealer

Mindre lysåbne arealer er en del af FAO's skovdefinition og inkluderes derfor i skovkortet, hvorimod større lysåbne arealer (>500 m²) ikke inkluderes. Grænsen for lysåbne arealer med en anden arealanvendelse end skov er 500m². Huller i skovens kronetag kan opstå af mange forskellige årsager. Hugst af et enkelt stort bøgetræ kan lave et hul i kronetaget på op i mod 650 m², men det ændrer ikke ved at det fortsat er skov (Nord-Larsen, Skov eller ej, 2024).

2.4 Andet træbevokset areal

I en dansk sammenhæng omfatter *andet træbevokset areal* bl.a. tilgroede heder, enge og moser. For at være omfattet af definitionen skal disse arealer leve op til de samme arealkrav som gælder for skov. Heder, enge og moser, som er større en 0,5 ha er i hovedreglen kortlagt som § 3 natur efter naturbeskyttelsesloven.

FAOs definition af *Other wooded land (OWL)* på dansk *andet træbevokset areal* lyder:

Andet træbevokset areal: Arealer med samme arealkrav som for skovdefinitionen, men et kronedække på 5-10 pct. af træer højere end 5 meter eller træer, som på voksestedet potentielt er i stand til at nå disse værdier, eller arealer med et kronedække af træ- eller buskarter, større end 10 pct., der ikke er i stand til at nå en højde på mere end 5 meter på voksestedet (Nord-Larsen et al., 2022, s. 5).

Andet træbevokset areal er opgjort til ca. 45.000 ha svarende til ca. 1% af landets areal, men regnes ikke som en del af skovarealet (Nord-Larsen et al., 2022)., og indgår derfor ikke i Digitalt Skovkort.

2.5 Træer uden for skovene

Digitalt Skovkort kortlægger og identificerer, hvor der er skov, jf. FAO's skovdefinition, med de afgrænsninger der er nævnt herover. Træer uden for skovene i f.eks. sommerhusområder, haver, parker, bytræer, mindre beplantninger eller læhegn indgår således ikke i det digitale skovkort. FAO benytter termen *Other land with tree cover* til at beskrive denne type arealer.

2.6 Fredskov

Arealer, som er tinglyst som fredskov, indgår i prædiktionslaget. Det betyder at arealkortets maskinlæringsmodel kigger på fredsskovsarealer og prædiktere hvorvidt der er trædække eller ej. Da fredskovsarealet indeholder større lysåbne arealer, vil ikke alle fredskovsarealer være omfattet af det færdige Digitale Skovkort. For Fredskov gælder samme forhold som resten af arealkortet fra afsnit 2.2-2.4.

Der er i udarbejdelse af arealkortet anvendt et fredskovslag, som er hentet hos Klimadatastyrelsen (Klimadatastyrelsen, 2024).

3. Grundlag for AI-Model

Dette afsnit har til formål at beskrive hvilke data features der bliver brugt, samt redegøre for hvilken metode og model der anvendes til landsdækkende prædiktioner af træ/ikke-trædække.

3.1 Remote Sensing til Arealforvaltning

I det digitale skovkort er der anvendt remote sensing data og machine learning. Denne kombination af teknologier er blevet anvendt i mange år, og er længe blevet benyttet ifm National Forest Inventory i både Canada (Gillis, Omule, & Brierly, 2005), (Matasci, et al., 2018) og Laos (Vicharnakorn, Shrestha, Nagai, & Salam, 2014).

I SGAV's digitale skovkort anvendes data features fra remote sensing til træning af maskinlæringsmodel og efterfølgende prædiktion af trædække/ikke-trædække. Det er en velbeskrevet metode (Chergui et al., 2023), (Fassnacht et al., 2016), (Mellor et al., 2012), og bliver her brugt til at forudsige sandsynligheden for trædække baseret på open source remote sensing data og træningsdata med kendte områder for træer og ikke-træer.

SGAV's AI-model, der kortlægger sandsynligheden for særligt værdifuld §25-skov, bruger samme metode, hvilket effektiviserer feltarbejdet betydeligt (Miljøstyrelsen, 2022).

3.2 Geografisk opdeling af kortlag anvendt til maskinlæringsmodellen

For at anvende maskinlærning til kortlægning af træ/ikke-trædække er det nødvendigt at inddele det totale landfaste areal af Danmark i kvadrater/pixels, som hver især skal klassificeres. I dette projekt er valgt en opløsning på 10m x 10m, da Sentinel 2 data grundlaget er i denne opløsning. Til klassificeringen indgår der tre kategorier:

Ikke-relevant – er pixel, som via allerede registrerede anvendelse, kan kategoriseres som ikke-relevant for skovkortet, se afsnit 4.1.

Trædækket – kategorisering af en enkel 10x10m pixel som værende primært dækket af trækrone

Ikke-trædække – kategorisering af en enkel 10x10m pixel som værende primært uden trækrone

3.2.1 Geospatielt indeks.

Til opdelingen af Danmark i kvadrater af 10m x 10m til træning af maskinlæringsmodellen og efterfølgende prædiktion af træer og ikke-træer anvendes et geospatielt indeks.

Hver 10m x 10m kvadrat bliver tilskrevet et unikt id, og dette id anvendes på tværs af data features, til at samle alle relevante beskrivende variable for netop dette kvadrat.

Disse variable anvendes i træning af modellen til at lære datamønstre for henholdsvis træ og ikke-træ-kategorierne, og i kortlægningen til at klassificere kvadratet som træ eller ikke-træ baseret på de lærte mønstre.

3.3 Datagrundlag

Ved hjælp af satellitbaserede observationer er der opbygget et datagrundlag af features, der giver en detaljeret forståelse af relevant vegetation. I det følgende beskrives både de afledte data features og grundfeatures baseret på Sentinel 2 satellit-programmet.

Sentinel-2 er en EU-initieret rummission til at observere Jorden, der leverer højopløst optisk billedmateriale af Jordens landoverflade og kystvande. Missionen består af to identiske satellitter, Sentinel-2A og Sentinel-2B, der er designet til at operere sammen i samme bane, men 180° udskilt fra hinanden. Dette tillader en høj gensynsfrekvens på 5 dage i den ækvatoriale zone, mens visse områder kan observeres mere ofte på grund af overlappende baneområder. Sentinel-2-missionen blev udviklet af et industrielt ledet konsortium, hvor European Space Agency (ESA) var ansvarlig for den overordnede implementering. Den første satellit, Sentinel-2A, blev udsendt den 23. juni 2015, og den anden satellit, Sentinel-2B, blev udsendt den 7. marts 2017. En tredje satellit, Sentinel-2C, er opsendt i 2024, og skal på sigt overtage fra Sentinel-2A (ESA - *Sentinel-2*, 2024). Sentinel-2 har multispektrale sensorer med 13 spektrale bånd i det synlige-, det nærinfrarøde og infrarøde spektrum, der giver detaljeret information om landoverflade, vegetation og vandkvalitet med en opløsning på 10m, 20m og 60m – afhængigt af lysbåndet.

3.3.1 Grunddata - Sentinel 2

Sentinel 2 optager data i de følgende 13 spektrale bånd svarende til synligt lys (B02-B04), nær-infrarød (B8, B8A), kortbølget infrarød "Short Wave Infrared" (B11, B12), "Red edge" (B05-B07) og atmosfæriske bånd (B01, B09, B10):

En oversigt over de spektrale bånd, deres bølgelængde og at-ground rummelige opløsning ses her under.

Bånd	Forklaring	Bølgelængde	Opløsning (at-ground)
B01	aerosol	443 nm	60m
B02	blue	490 nm	10m
B03	green	560 nm	10m
B04	red	665 nm	10m
B05	vegetation red edge 1	705 nm	20m
B06	vegetation red edge 2	740 nm	20m
B07	vegetation red edge 3	783 nm	20m
B08	nær-infrarød	842 nm (NIR)	10m
B8A	nær-infrarød	865 nm (NIR)	20m
B09	vanddamp	940 nm	60m
B10	skydække detektion	1.375 nm	60m
B11	short wave infrared	1.610 nm (SWIR1)	20m
B12	short wave infrared	2.190 nm (SWIR2)	20m

3.3.2 Afledte data-features – Sentinel 2

De rå satellitdata indeholder meget information, men er ikke altid direkte anvendelige for maskinlæringsmodeller. Ved at udlede specifikke features, som kan fremhæve forskelle mellem trædække og ikke-trædække, kan modellens evne til korrekt at skelne mellem træbevokset eller ej forbedres.

En vigtig egenskab ved maskinlæringsalgoritmer som LightGBM er deres evne til effektivt at håndtere og prioritere features. LightGBM har en indbygget mekanisme, der automatisk identificerer og aggregerer features der ofte er gensidigt udelukkende (Ke, et al., 2017). Denne proces kaldes Exclusive Feature Bundling og nedsætter størrelsen på feature rummet og dermed mængden af forhold modellen skal lære. Denne egenskab gør LightGBM særligt velegnet til at håndtere datasæt med mange features.

Der bruges flere spektrale indeks i modellen, og en kort forklaring af hvert indeks er angivet herunder. Ikke alle features vil her have direkte semantisk sammenhæng med skov segmentering. På grund af LightGBM's indbyggede ekskluderende mekanisme på features og at andre projekter hos SGAV også benytter featuresættet, har vi valgt at beholde dette for at strømline processen omkring fælles projekter.

Normalized Difference Vegetation Index (NDVI) eller "grønhedsindeks"

NDVI bruges til at vurdere vegetationsdækning og -tilstand. NDVI udnytter det faktum, at sunde planter reflekterer meget nær-infrarødt lys (NIR, B08) og absorberer meget rødt lys (RED, B04). Ved at måle forskellene mellem disse to typer refleksioner kan man kvantificere vegetationsmængden i et givet område. (<https://custom-scripts.sentinel-hub.com/sentinel-2/ndvi/>)

Anthocyanin Reflectance Index 1 (ARI1) (Gitelson et al., 2009)

Anthocyaniner er pigmenter, der er almindelige i karplanter og giver dem deres røde, blå og lilla farver. De giver værdifuld information om planters fysiologiske tilstand, da de betragtes som indikatorer for forskellige typer af plantestress (Gitelson, Chivkunova, & Merzlyak, 2009).

Anthocyanin Reflectance Index 2 (ARI2)

For at korrigere for bladets tæthed og tykkelse tilføjes det nær-infrarøde spektralband (i de anbefalede bølgelængder 760-800nm), som er relateret til bladspredning, til det grundlæggende ARI-indeks (Gitelson, Chivkunova, & Merzlyak, 2009).

Chlorophyll Red Edge (CRE)

Indholdet af klorofyl i bladene, er en indikator for helbredet af planten. (Gitelson, Gritz, & Merzlyak, 2003)

Enhanced Vegetation Index (EVI)

I områder med tæt løvtag, hvor bladarealindekset er højt, kan de blå bølgelængder bruges til at forbedre nøjagtigheden af NDVI. Dette sker ved at korrigere for jordbaggrundssignaler og atmosfæriske påvirkninger (<https://custom-scripts.sentinel-hub.com/sentinel-2/evi/>).

Plant Senescence Reflectance Index (PSRI)

Senescence refererer til den proces, hvor planter ældes og går igennem fysiologiske forandringer, der fører til nedbrydning af klorofyl og ophobning af andre pigmenter som carotenoider, der er de af gule og brune pigmenter. (<https://custom-scripts.sentinel-hub.com/sentinel-2/psri/>)

Clay Minerals Ratio (CMR)

CMR er en afledt datafeature, som anvendes til at identificere og kvantificere tilstedeværelsen af lerholdige mineraler i jord og sedimenter. (https://knowledge.dea.ga.gov.au/_modules/dea_tools/bandindices/)

Ferrous Minerals Ratio (FMR)

Ferrojernholdige mineraler som hematit og goethit har specifikke absorptionsegenskaber, der kan detekteres ved hjælp af spektrometri og fjernmålingsdata. Disse mineraler absorberer lys ved bestemte bølgelængder, hvilket resulterer i karakteristiske reflektansmønstre, der kan bruges til at beregne FMR. (Segal & Sellers, 1982)

Iron Oxide Ratio (IOR)

Jernoxidforholdet er forholdet mellem de røde og blå bølgelængder, som måler ændringer i jernoxidforholdet. Karakteren af jernoxidforholdet gør det muligt for dette indeks at afbøde belysningsforskelle forårsaget af terrænskygge. (Segal & Sellers, 1982)

Burned Area Index (BAI)

Måler områder i terrænet som er påvirket af brand (Chuvieco, 2002)

Modified Normalized Difference Water Index (MNDWI)

Udnytter det grønne bånd og kortbølget infrarøde bånd til at beskrive åbent vand. (Xu, 2006)

Normalized Difference Built-up Index (NDBI)

Udnytter NIR- og SWIR båndene til at måle på bebyggede områder. Ved at udnytte forholdet mellem disse bånd, kan man afbøde effekterne af belysningsforskelle i terrænet såvel som atmosfæriske effekter (J., 2018)

3.3.3 Filtrering af data – Sentinel 2

I beregningen af features, inddeles året først i 4 biologisk-betinget sæsoner.

Vinter: 01.12.2022 – 28.02.2023 (ultimo februar afhængig af skudår),

Forår: 01.03.2023 – 31.05.2023

Sommer: 01.06.2023 – 30.08.2023

Efterår: 01.09.2023 – 30.11.2023

Inden for hver sæson beregnes alle grunddata og afledte features baseret på Sentinel 2 spektrale bånd som beskrevet ovenfor. For at beregne features, skal der bruges én observation pr pixel pr sæson.

Tidsserie data fra Sentinel 2 stammer fra flere overflyvninger over Danmark i løbet af året. For hver overflyvning optages der data for en delmængde af det danske landfaste areal.

Det betyder, at den samme pixel har observationer fra flere forskellige datoer inden for hver sæson. Samtidigt kan Sentinel 2 ikke "se" gennem skyer. Så hvis der er skydække over det område som Sentinel 2 satellitten overflyver vil der opstå brud i de observerede tidsserier.

En enkel pixel kan derved have f.eks. 20 observationen i sommersæsonen mellem 1. juni og 30 august, mens der kun er en enkelt observation i vinterperioden.

3.3.4 Sentinel-2 Kompositstrategi

For at udvælge en observation per pixel af de fx 20 der kan eksistere i en sæson, benyttes et filter der kan udvælge den bedst egnede pixel. Dermed laves et *kompositbillede* af Danmark for en given sæson, kun baseret på de bedste pixels. Der findes mange forskellige kompositstrategier der bygger på fx fænologi (Frantz, Stellmes, Hill, & Röder, 2017), (Morresi, Marzano, Lingua, Motta, & Garbarino, 2022)) eller på beregninger på de spektrale bånd (Doninck & Tuomisto, 2017), (Francini, et al., 2023) (Griffiths, Nendel, & Hostert, 2019).

Nyere algoritmer såsom Best Available Pixel og Mediod algoritmerne viser meget gode resultater på temporal konsistente kompositter. Disse kan dog også være komplicerede at implementere på skalerbare måder. Der er derfor anvendt en af de mest traditionelt anvendte metoder, nemlig "den grønneste pixel". Her beregnes NDVI-værdien for hver enkelt observation, og den observation med den højeste NDVI bliver repræsentant for den pixel. Dette filter bruges for at

udvælge de observationer, hvor skydækket har været lavest. Dermed produceres et kort for hver sæson hvor hver pixel er den pixel med størst umiddelbar sandsynlighed for vegetation, og hermed filtreres skyer ofte fra. Det landsdækkende kort for en sæson bliver derved opbygget som en mosaik, hvor der indgår observationer fra flere forskellige datoer.

3.3.5 Tidsseriedata – Sentinel 2

I klassifikation af trædække og ikke-trædække områder, komplementeres med datafeatures, der beskriver forskellen i værdier af grunddata og spektrale indeks mellem vinter- og sommer-sæsonerne.

Delta-beregninger fremhæver de sæsonbetingede forskelle, hvilket hjælper med at identificere og klassificere skovområder i løv eller nåleskov. Derudover kan sne om vinteren dække jorden og skjule vegetationen, hvilket ændrer spektralresponsen. Delta-beregninger mellem vinter og sommer kan hjælpe med at korrigere for sneens indflydelse og give en klarere indikation af det underliggende jord- eller vegetationsdække.

3.3.6 Fænologi-data

Fænologi er studiet af periodiske biologiske hændelser i vegetation i relation til klimatiske og sæsonmæssige ændringer.

Baseret på Sentinel 2 optiske data, er der via WEKEO apiet hentet en række features ud fra High Resolution Vegetation Phenology and Productivity indekset (HR-VPP) (WEKEO, 2024) & (European Environment Agency, 2024). Selvom disse features er bygget på det samme data-grundlag som de spektrale indekser der forklares i 3.3.2, så inkluderer de en tidlig forståelse der ikke er inkluderet i de resterende features.

Ud fra HR-VPP indekset er der fra Copernicus beregnet en række parametre, og i datafeatures. Udvalgt til dette projekt indgår følgende data features:

SOSD	Dag for sæsonstart
EOSD	Dag for sæsons slut
MAXD	Dag for sæson maxproduktivitet
SOSV	VPP værdi for sæsonstart
EOSV	VPP værdi for sæsons slut
LSLOPE	Hældning på kurven for VPP ved sæsonstart
RSLOPE	Hældning på kurven for VPP ved sæsons slut

Hver parameter udregnes af Copernicus baseret på et helt års Sentinel 2 observationsdata, og interpolering for perioder med manglende observationer grundet skydække.

I modellen er anvendt afledte fænologidata, henholdsvis middelværdien og variansen, baseret på to års data.

3.3.7 Opsummering

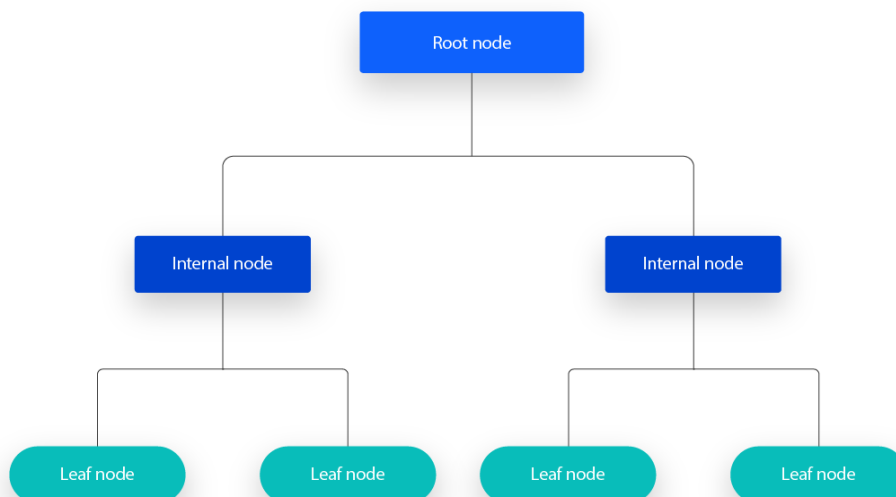
Det endelige produkt består af 116 data-features og et geografisk indeks med en opløsning på 10x10 m landsdækkende for Danmark – i alt ca. 0,5 mia. pixel.

3.4 Model-metode

Til klassificering af trædække/ikke-trædække for de enkelte pixels, er anvendt LightGBM-frameworket. LightGBM er en gradient boosting maskinlæringsalgoritme udviklet af Microsoft. Det er konstrueret med fokus på hurtighed og effektivitet og er anerkendt for sin evne til at håndtere store datamængder. LightGBM står for Light Gradient Boosting Machine og implementerer beslutningstræ-baserede algoritmer (Ke, et al., 2017).

En sammenlignende analyse af gradient boosting-algoritmer evaluerede ydeevnen af forskellige varianter af gradient boosting, herunder LightGBM, i forhold til Random Forest. Resultaterne antydede, at selvom Random Forest er en robust og pålidelig metode, opnår gradient boosting-algoritmer som LightGBM ofte bedre generaliseringsydelse og hurtigere træningstid. LightGBM er klart den hurtigste model til træning og inferens. Dette skyldes i høj grad modellens evne til parallel behandling. Alle gradient boosting modellerne der er evalueret i studiet performer dog bedre end Random Forest (Bentéjac et al., 2021).

LightGBM-modellen består af flere "generationer" af delmodeller, som hver især er simple beslutningstræer som kun tager hensyn til en lille mængde af variansen af data. Et beslutningstræ er en algoritme, som bruges i flere forskellige machine learning frameworks, både til klassificering og regression (Figur 3.1).



Figur 3.1 Illustration af et beslutningstræ. Øverst er der en root node, der har grene ned til 2 Internal nodes. Disse to Internal nodes har derefter grene ned til 4 Leaf nodes. (What is a decision tree?, 2024)

Et beslutningstræ består af flere niveauer af noder, og ved hver node foretages der split i data baseret på en enkel af de features, som beslutningstræet har fået stillet til rådighed. Et beslutningstræ minder således om klassiske flowdiagrammer, hvor "vejen" ned gennem træet afgøres af værdien for features for den enkelte observation.

Et split i træet vil typisk kunne formuleres som "Er værdien af feature X større end Y?", hvor størrelsen kaldes tærskelværdien. Netop tærskelværdien for de enkelte features er dem som skal læres i træningen af modellen.

Gevinsten ved at tilføje en ekstra node til træet beregner reduktionen i fejlen mellem forudsigelser og faktiske værdier i træningssættet.

Forbindelserne mellem noderne repræsenterer beslutningsvejene, hvor hver forbindelse viser retningen af opdelingen baseret på tærskelværdien. Terminal- eller bladnoderne indeholder de endelige forudsigelsesværdier og indikerer udfaldet af beslutningsvejen, uden yderligere opdeling, og vil i dette tilfælde enten have værdien 1 (træ) eller 0 (ikke-træ).

Delmodellerne bygges sekventielt, hvor hver model tager forbehold for fejlene fra tidligere delmodeller. Hver delmodel indgår i slutmodellen som en del af et ensemble.

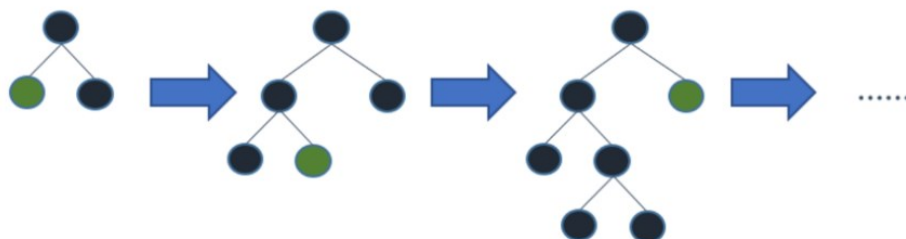
De generelle trin i metoden er herunder:

1. Start: Begynder med et simpelt beslutningstræ

2. Beregning af Fejl: For hver iteration, beregnes fejlene (forskellen mellem forudsagte og faktiske værdier).
3. Træning: Et nyt beslutningstræ bygges ud fra information om de beregnede fejl.
4. Opdatering: Den nye model kombineres til de eksisterende beslutningstræer.
5. Gentagelse: Processen fortsætter, indtil den samlede model ikke længere forbedres.

Det er særegent for LightGBM, at den er hurtigere og mere effektiv til at opnå den optimale ensemble af beslutningstræer via "Leaf-wise tree growth".

I stedet for at opbygge træer niveau for niveau, vokser LightGBM træer ved at splitte de "blade", der giver den største reduktion i fejl. Dette fører til dybere og mere komplekse træer (se nedenstående *Figur 3.2*).



Figur 3.2: LightGBM – "Leaf-wise tree growth". Illustrerer hvordan LightGBM algoritmen vokser. Beskrevet ovenfor.

Klassificering er en af de maskinlærings-tekniker, der bruges til at identificere potentielle klasser. I dette tilfælde bruger vi metoden til at klassificere, hvorvidt der er træ-dække eller ej.

3.5 Forklarbarhed og Gennemsigtighed

Forklarbarhed er typisk en af de store bekymringer ved en induktiv model, såsom en maskinlæringsløsning, sammenlignet med en deduktiv, teori-drevet model. Induktive modeller er blevet meget populære grundet deres evne til præcist at fange meget komplicerede sammenhænge i data. Dette kan dog også give udfordringer da disse komplicerede sammenhænge kan blive så uigennemsigtige at eksperter kan have svært ved at forstå modellens beslutningsgrundlag.

Denne udfordring kan effektivt mitigeres gennem både valg af model, samt statistiske efteranalyser.

3.5.1 Feature Importance i LightGBM

LightGBM er, som beskrevet i 3.4, en forklaringstræs-baseret model. Disse er bredt anerkendt for deres evne til at eksportere den globale vigtighed af features i adskillelsen af træningsdata. I LightGBM er feature importance opdelt i 2 kategorier der er logges i forbindelse med træning af SGAVs skovmodel:

Split Importance: Måler hvor ofte en feature benyttes i et *split* i beslutningstræet. Hver gang modellen foretager et split, skal det forstås som "*hvis feature_x > Y → A ellers → B*", og dermed at featuren har en eller anden funktion i at opdele dataen mest effektivt.

Gain Importance: Denne metrik måler hvor meget modellen gennemsnitligt forbedre sig ved splits der involverer en given feature.

Hvis en feature benyttes i meget få splits, eller at de splits featuren indgår i har meget lille effekt i loss-funktionen er featuren altså ikke udslagsgivende for modellens overordnede performance. Altså kan denne metrik generelt fortælle hvilke features der er vigtigst for korrekt at klassificere størstedelen af dataen.

3.5.2 SHAP Værdier

Shapely Additive Explanations (SHAP) (Lundberg & Lee, 2017) værdier er blevet meget populære grundet deres evne til at modelagnostisk estimere bidraget fra givne features. Værdierne beregnes ud fra spilteori, hvor hver features bidrag til det endelige output kvantificeres – positive tal trækker i en positiv retning, og negative tal trækker i en negativ retning.

Modellen kan, meget lig feature importance, beregne globale indflydelser for hver feature på forudsigelserne.

Derudover kan disse beregnes, ikke kun globalt for modellen, men også effektivt beregne hvordan værdien af alle features påvirker en enkelt forudsigelse positivt eller negativt. Dermed kan man foretage lokale analyser på enkelte områder og estimere hvorfor modellen har klassificeret et givent område på en bestemt måde. Dette værktøj er derfor enormt præcist, men også meget beregningstungt at benytte på alle pixels. Derfor laves lokale analyser kun på udvalgte områder.

3.6 Træning af maskinlæringsmodel

Træning af modellen til binær klassifikation foregår ved at splitte datasættet i 3 dele: trænings-, validerings- og testdata. LightGBM algoritmen justerer sine parametre for at minimere fejl i klassifikationen af træningsdataene – men med udgangspunkt i afprøvning af valideringsdata. Når modellen ikke opnår bedre nøjagtighed på valideringsdata efter et forudbestemt antal iterationer fastlåses maskinlæringsmodellens parametre.

3.7 Træningsdata og metode for udvælgelse af træningsdata

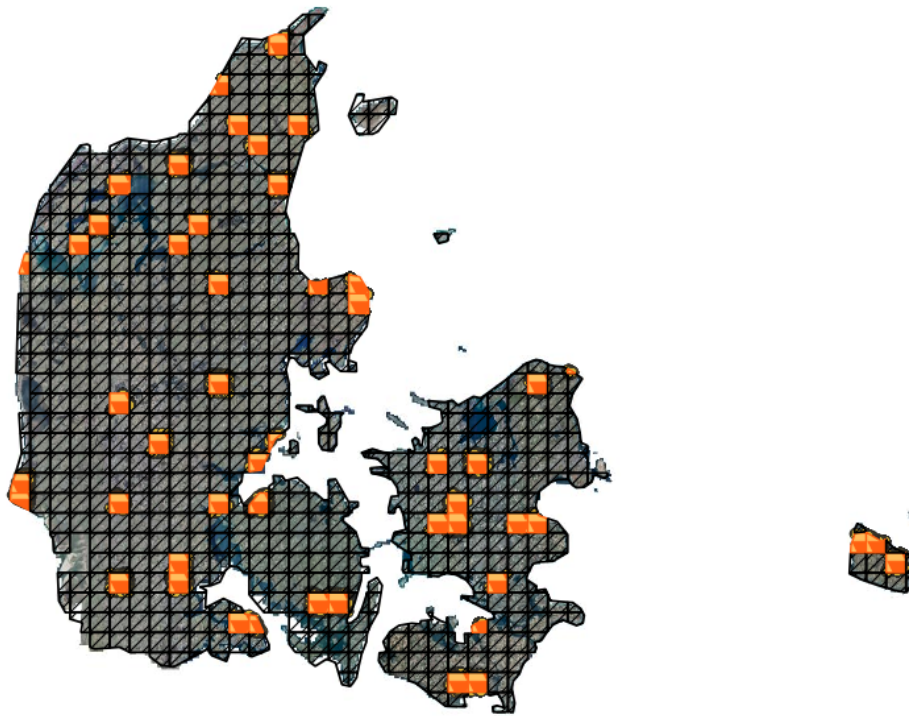
Træningsdata beskriver sammenhængen mellem de indhentede satellitdata og hvorvidt en enkel 10m x10m pixel er trædækket eller ej. Træningsdata refererer til data, der er blevet forsynet med beskrivende etiketter eller tags, der angiver, hvad hvert datapunkt repræsenterer eller hvilken kategori det tilhører – eksempelvis: træ-dække/ikke-træ-dække.

3.7.1 Træpixels

En pixel med trædække udvælges baseret på træningskortlag produceret internt i SGAV. Til at træne modellen for arealkortet, har SGAV skabt datagrundlaget i GIS. Der er blevet foretaget en faglig visuel genkendelse af trædækkede skovområder som de ser ud på ortofotos for 2022¹. Trædækkede skovområder er indtegnet som polygoner og markeret som værende løv, nål eller blandet skov.

Områder til træningsdata er udvalgt ved stratificering. Danmark er blevet opdelt i 600 10x10 km kvadrater. Derefter blev 50 kvadrater valgt ud fra deres geografiske fordeling i Danmark, for at sikre bred geografisk dækning af træningsdata. I hvert udvalgt kvadrat blev der optegnet polygoner med skovbevoksede arealer. Som hjælpværktøj til optegning af skovpolygoner blev fredskovsarealet i hver 10x10 km pixel benyttet som ramme. Der er tegnet polygoner, som er grundlag for pixels. I figuren nedenfor vises den geografiske spredning af træningsgrundlaget. I alt er udlagt 12.082 polygoner fordelt på 50 10x10km kvadrater.

¹ (Ortofoto 12,5 forår 2022, EPSG:25832 – ETRS89 / UTM 32N fra KDS' kortforsyning WMS).



Figur 3.3: Illustration af 50 10x10km kvadrater der er udlagt i Danmark. I de viste kvadrater er der udlagt polygoner for træ-dække og prikker for ikke-trædække

Optegning af polygoner blev foretaget i sommeren 2023.

De indtegnede polygoner med trædække er herefter konverteret fra polygoner til 10x10m pixels via det anvendte geospatiale indeks (Geospatielt indeks.3.2.1). Til dette anvendes Databricks Mosaics (databrickslabs, 2024) funktionen *grid_polyfill*. Denne funktion finder alle pixels hvis centrum ligger inden for et givent polygon. Denne metode vil resultere både i pixels der ligger delvis uden for polygonet og dele af det oprindelige polygonareal som går tabt.

Et eksempel på processen kan ses i Figur 3.4. Her ses hvordan gule pixels svarende et blå underliggende polygon vil se ud. I kanten af det oprindelige polygon ses hvordan nogle pixels går ud over polygonets grænser, samt hvordan dele af polygonet ikke dækkes af pixels.



Figur 3.4: Illustration af mosaics grid_polyfill funktionen. Den viser hvordan træningspixels bliver udvalgt på et indtegnet polygon, og især hvordan pixels tætte på grænsen for polygonet bliver udvalgt; ved at se om centrum for pixelen er omfattet af polygonet.

3.7.2 Ikke-træ pixels

I de samme 10x10 km kvadrater, hvor træpolygoner er blevet tegnet, blev der også markeret ikke-træ pixels på 10x10 meter, som er identificeret som områder uden trædække. Pixels, der blev identificeret ved tolkning af flyfotos, blev der også anvendt feltverificerede ikke-træ punkter fra overvågning af lysåben natur (habitatnaturtyper, som identificeret og kortlagt i den Nationale Overvågning af Vand og Natur NOVANA), og pixels placeret i marker med afgrøder, som ikke giver anledning til trædække (se afsnit 4.1.1). I alt er der blevet udlagt ca. 260.000 pixels uden trædække. Pixels uden trædække fra marker er beskrevet yderligere i listen over afgrødekoder (Bilag 1.1), og pixels med lysåben natur fra NOVANA og Digital Naturovervågning (se Bilag 1.2).

3.7.3 Afbalancering mellem positiv og negative træningspunkter

Der er udtaget ca. 3 mio. træningspunkter for træ-dække og 260.000 pixels for ikke træ-dække. Efterfølgende er antallet af træ-dække pixels ned-samplet til 260.000 for at opnå et balanceret datasæt.

4. For- og efterbehandling

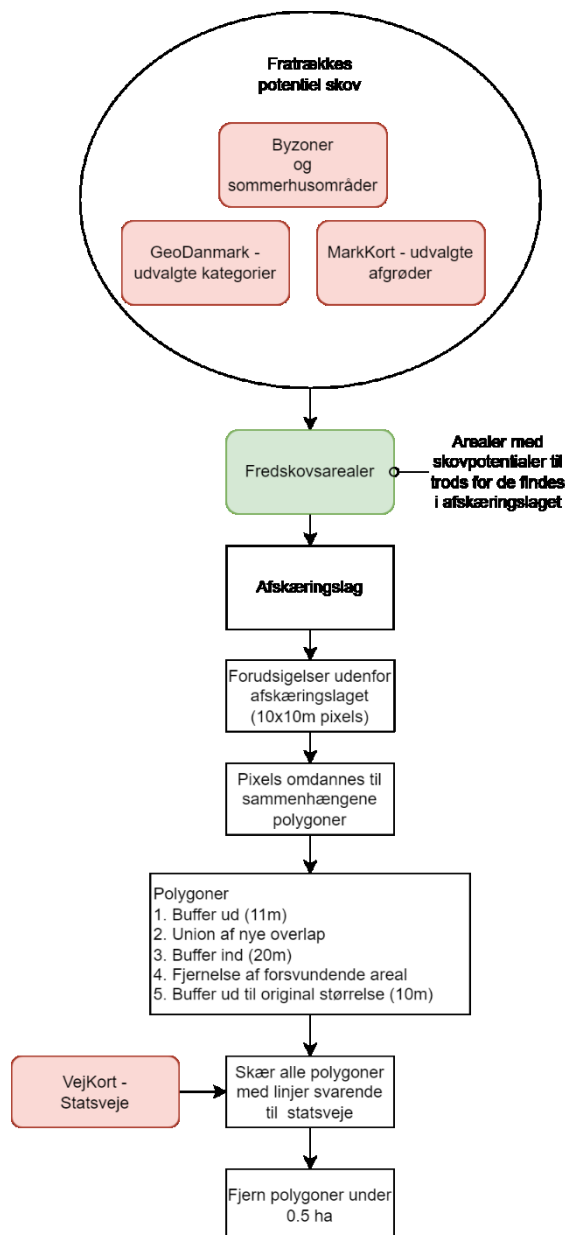
4.1 Afskæringslag før modellen prædikerer træ-dække

Før maskinlæringsmodellen skal finde trædækkede områder reduceres det geografiske område modellen skal arbejde med, for at spare processeringstid og skabe en mere effektiv model. Det gøres ved at skære områder ud som med sikkerhed ikke er dækket af træer. Derudover anvendes et kystlinjelag ("landpolygon" fra GeoDanmark) til at definere udstrækning af kystlinjen, for at fjerne alle havområder i Danmark. Dette samlede "ikke-træ" område defineres som "afskæringslag". Afskæringslagene er beskrevet herunder.

4.1.1 Landbrugsarealer

Arealer, som med sikkerhed er landbrugsarealer uden trædække indgår i afskæringslaget. Landbrugsstyrelsen udarbejder detaljerede markkort, som på polygonniveau indeholder information om hvilke afgrøder, der findes på den enkelte mark. Kortlaget er grundlag for udbetaling af EU-landbrugsstøtte og er underlagt detaljeret kontrol og har derfor en høj grad af nøjagtighed.

Langt de fleste afgrøder indgår i afskæringslaget, dog udelukker nogle afgrøder ikke, at der kan være skov, fx visse flerårige græsarealer og andre arealer som ikke pløjes hvert år, såsom juletræsarealer, skovrejsning på markjord og poppel til energitræ. Ligeledes indgår visse arealer med træer i afskæringslaget. Det drejer sig om skovlandbrug, som defineres som landbrug og frugtplantager som specifikt er nævnt som ikke værende skov i FAO's skovdefinitioner. Læhegn indgår ikke i markblokkortene, da de per definition ikke er marker/dyrket areal. Det betyder også at de ikke indgår i afskæringslaget, men i stedet bliver fjernet ved efterbehandlingen (afsnit 4.2), fordi de er for smalle til at leve op til skov-definitionen (afsnit 2.1).



Figur 4.1: Grafisk oversigt over dannelse af færdigt skovkort.

Afgrænsning af hvilke afgrødekoder, som indgår i grundlaget for skovkortet eller som ikke gør (afskæringslaget) er anbefalet af Københavns Universitet. Den fulde liste over de ca. 600 afgrødekoder og deres anvendelse i det digitale skovkort fremgår af bilag til notatet "skov eller ej" (Nord-Larsen, Skov eller ej, 2024).

Landbrugstyrelsens markkort opdateres løbende. Data opdateres i løbet af foråret for indeværende år og udgives løbende i en ufærdig version. Den endelige version udgives ca. 1. juni, og denne version afspejler dermed årets marker. Se evt. <https://lbst.dk/landbrug/kort-og-mark-blokke/oversigt-over-korttemaer#c30679>

4.1.2 GeoDanmark - befæstede arealer

Klimadatastyrelsen udgiver sammen med kommunerne under www.geodanmark.dk en lang række data af høj præcision om bl.a. befæstede arealer og bygninger mv. Der indgår 10 lag fra GeoDanmark i afskæringslaget, som med sikkerhed ikke kan omfatte skov.

- a. Bykerne (Tæt bebygget område med gader, bolig og andre bygninger)
- b. Lav bebyggelse
- c. Høj bebyggelse
- d. Erhverv
- e. Parkeringsområder
- f. Bassin
- g. Tekniske anlæg
- h. Startbane (flyvepladser)
- i. Bygninger
- j. Vindmøller
- k. Kirkegårde

Tekniske anlæg (g) er en stedfast klart afgrænset konstruktion, der er opført til et bestemt formål, og som ikke karakteriseres som en bygning. Fx en olietank, gyllebeholder, halmfyr, bio-brændselsanlæg mm. Bygninger (i) er data fra BBR og består af en eller flere konstruktioner, som udgør en helhed der som minimum har et tag.

Data fra GeoDanmark er hentet for år 2022 fra <https://datafordeler.dk/dataoversigt/geodanmark-vektor/geodanmark-vektor-praedefineret-shape/>, som er samstemmende med udlagte træningspunkter og datafeatures fra Sentinel 2.

4.1.3 Byzoner og sommerhusområder

I byer og sommerhusområder findes træer i haver, parker og andre arealer med træer (*other land with treecover*). Disse er ikke en del af skovarealet jf. skovdefinitionen. Byzoner og sommerhusområder er udlagt i planlægningszoner, som fremgår af plandata om kommuneplanrammer. Disse zoner indgår i afskæringslaget. Der forekommer dog skove i byer eller som mindre områder inden for sommerhusområder. Disse vil i hovedsagen være pålagt fredskovspligt. Alle fredskovspligtige arealer større end 0,5 ha indgår som grundlag for det digitale skovkort uanset om det ligger inden for disse zoner eller i andre afskæringslag. Se figur i afsnit 3.1. Zonekortet hentes hos Plandata.dk via deres WFS service.

Plandata anvendes ligeledes fra 2022, som øvrige datafeatures.

4.2 Efterbehandling

Efter maskinlæringsmodellen har fundet alle pixels med trædække, inden for det afgrænsede område, skal disse pixels omdannes til polygoner som lever op til skovdefinitionens afgrænsninger.

4.2.1 Fra pixel til polygon

Alle pixels med forudsigelser af trædække omdannes til sammenhængende polygoner (skovarealer).

En af forudsætningerne i skovdefinitionen er, at en skov er mere end 20 m bred. SGAV har på baggrund af anbefalinger fra pilotprojekt udarbejdet af Københavns Universitet (Nord-Larsen, et al., 2023) fortolket det som, at polygoner som ligger tættere på hinanden end 20 meter defineres som værende en del af den samme polygon og dermed det samme skovområde. Som et led i dannelsen af polygoner udvides hvert polygon derfor med 11 meter, og polygoner der overlapper efter udvidelsen, sammenføjes til en samlet polygon. På tilsvarende vis fortolkes små områder inde i skoven, som er mindre end 500 m², som en del af skovområdet. De nye, større polygoner formindskes efterfølgende med 22 meter, hvilket medfører, at små områder uden trædække fjernes fra skovkortet. For at gendanne til størrelsen på den originale polygon udvides polygonerne igen med 10 meter.

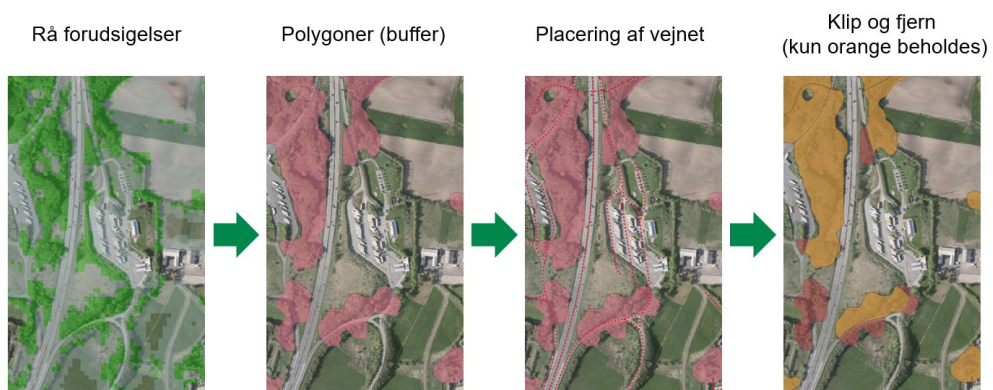
4.2.2 Veje

Det endelige sæt af polygoner skæres visse steder igennem af statsveje, som vejdirektoratet er vejmyndighed for. Disse veje er typisk større hovedveje mellem større byer samt motor og motortrafikveje. Disse veje er karakteriseret som brede veje, som tydeligt deler skovene op og er medtaget i afskæringslaget. Mindre landeveje, biveje og andre kommunale veje er ikke medtaget i afskæringslaget, da skovens kronedække kan nå over vejene og ikke danner barrierer i forhold til skovens funktioner og drift.

Data om statsveje er hentet fra vejdirektoratets dataudveksling (<https://du-portal-ui.dataudveksler.app.vd.dk/>) – hentet april 2024.

Afslutningsvis fjernes alle polygoner, der herefter er under 0,5 hektar, idet disse arealer ikke lever op til FAO's skovdefinition om minimumsstørrelsen for en skov.

Resultatet af disse trin er et skovkort, der er baseret på pixelbaserede forudsigelser om trædække, men som består af sammenhængende polygoner (skove), der er justeret ved at fjerne små trædækkede områder, som er for små til at leve op til FAO skovdefinitionen (under 0,5 ha) og medtager små lysåbne områder uden træer i større skovpolygoner. Endelig er der taget højde for de væsentligste veje (statsveje).



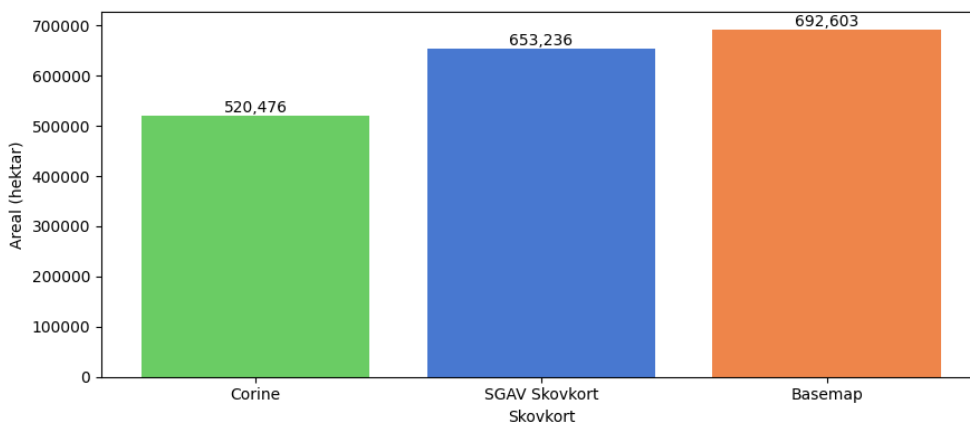
Figur 4.2: Illustration af metoder til efterbehandling af prædiktionskortlag. Trædækkeforudsigelserne bliver lavet om til polygoner, hvor der bliver lavet bufferøvelser på. Herefter bliver vejene illustreret og hvordan de skærer gennem skovpolygonerne og fjerner alt skov under 0,5 hektar. Dette er et eksempel, og vejene stemmer ikke overens, med hvilke veje der bliver brugt i metoden.

5. Resultater og kvalitetssikring

5.1 Resultater fra maskinevaluering

De data der er udlagt til træning af LightGBM til genkendelse af træ/ikke-træ er opdelt i tre grupper

1. Træningsdata – består af 64% af ikke-træ pixel, og et tilsvarende antal af træ-pixel. Disse data bruges i træningen af modellen til at indstille algoritmen til genkendes af træ/ikke-træ på pixel-niveau
2. Valideringsdata – består af 16% af ikke-træ pixel, og et tilsvarende antal af træ-pixel. Disse data bruges i træning af modellen til Early.stopping, hvor fejlen i forudsigelserne løbende evalueres via valideringssættet, og træning stopper, når der ikke er opnået forbedringer i resultaterne. Dermed undgår man overfitting af modellen
3. Testdata – består af 20% af ikke-træ pixel, og et tilsvarende antal af træ-pixel. Disse data bruges først når træningen er afsluttet til at måle modellens nøjagtighed i klassificeringen af træ/ikke-træ på pixels som hverken har indgået i træning eller validering.



Baseret på testdata er der lavet to evalueringer af modellens nøjagtigt, en "Confusion Matrix" og et "Accuracy" mål.

5.1.1 Confusion Matrix

En confusion matrix er et værktøj til at evaluere ydeevnen af en klassifikationsmodel ved at sammenligne de faktiske og prædikterede værdier inden for testdata. Forudsigelserne opdeles i 4 forskellige kategorier

1. True Positive (TP) – Forudsigelse af "træ" i en pixel, som i testdata også er angivet som "træ"
2. False Positive (FP) – Forudsigelse af "træ" i en pixel, som i testdata er angivet som "ikke-træ"
3. False Negative (FN) – Forudsigelse af "ikke-træ" i en pixel, som i testdata er angivet som "træ"
4. True Negative (TN) – Forudsigelse af "ikke-træ" i en pixel, som i testdata også er angivet som "ikke-træ"

Ud fra disse fire mål, opstilles en matrice, hvor hvert mål er angivet som total antal eller procent af forudsigelser inden for den respektive kategori.

Confusion Matrix for binær klassifikation kan således opstilles som herunder

	Kortlagt Træ (1)	Kortlagt ikke-træ (0)
Forudsagt Træ (1)	True Positive (TP)	False Positive (FP)
Forudsagt Ikke-træ (0)	False Negative (FN)	True Negative (TN)

For modellen anvendt til at skabe trædække/ikke-trædække forudsigelserne på pixelniveau opnås der følgende resultater:

	Kortlagt Træ (1)	Kortlagt ikke-træ (0)
Forudsagt Træ (1)	98.7%	1.3%
Forudsagt Ikke-træ (0)	1.5%	98.5%

Det betyder at over 98% af de pixels der er kortlagt indenfor hver af de to kategorier klassificeres korrekt af den trænedede LightGBM model, og at modellen evner at genkende ikke-trædække på samme niveau som træ-dække pixels.

5.2 Sammenligning med Referencekort

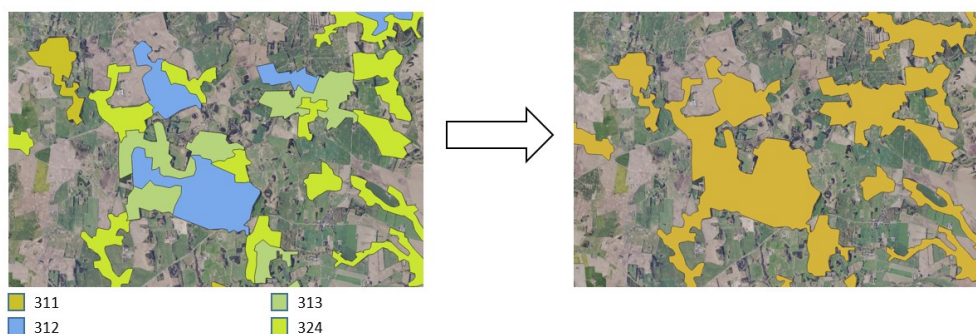
5.2.1 Corine

CORINE Land Cover (Europa-Kommissionen, 2024) er en serie af kort produceret af ESA Land Monitoring Service. CORINE er et projekt, hvor land use på tværs af Europa kortlægges, og der er derfor medtaget flere forskellige kategorier. For alle kategorier gælder det at det skal være et minimums areal på 25 hektar, som dog kan være på tværs af landgrænser, og en minimumsbredde på 100 meter.

I dette projekt er der lavet en sammenligning mellem CORINE kortet fra 2018 og SGAVs arealkort inden for de kategorier fra CORINE som er relevant for Arealkortet. De udvalgte kategorier er:

1. 311 (Broadleaved Forest)
2. 312 (Coniferous Forest)
3. 313 (Mixed)
4. 324 (Transitional Woodland Shrub)

CORINE 2018 er baseret på Sentinel-2 observationer for årene 2017-2018, med Landsat-8 til at udfylde huller i Sentinel-2 tidsserierne. Denne metode leder til et totalt skovareal på 520 476 hektar.



Figur 5.1 Illustration af hvordan individuelle polygoner fra CORINE datasættet indenfor de 4 udvalgte kategorier samles til sammenhængende skovpolygoner.

5.2.2 Basemap

Der er taget udgangspunkt i Basemap04 der udkom i 2022, baseret på data fra 2021. Basemap er produceret gennem en kompleks

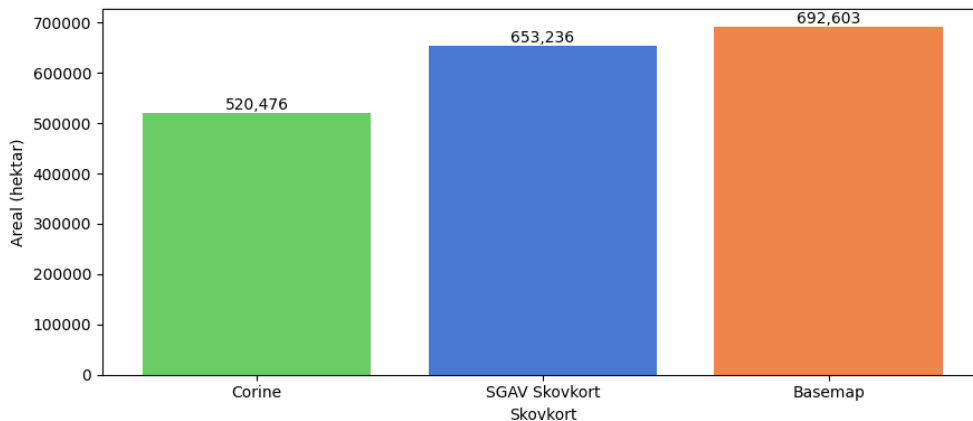
På samme måde som sammenligningen med CORINE, er SGAV Arealkortets skovpolygoner også sammenlignet med de arealer, som er kortlagt i BASEMAP. Fra BASEMAP er brugt laget "sub_tree_cover", hvor kategorierne

1. "Tree cover on other land use/land cover",
2. "Forest/afforestation",
3. "Christmas trees/cut greenery",

Er medtaget, og merget sammen til sammenhængende skovpolygoner. Dette resulterer i et samlet skovareal på 692.603 hektar.

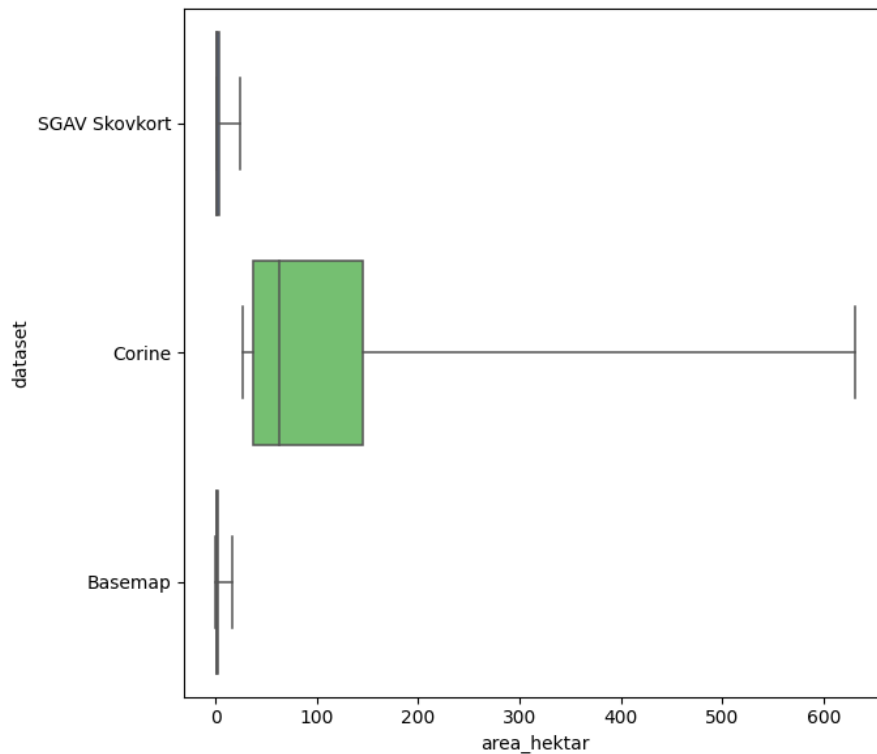
5.2.3 Generel Statistik

De tre metoder er ret forskellige, og metoderne bag Basemap og SGAVs arealkort finder skovarealer der afviger med ca 40 000 hektar. Corine derimod, finder ca 150 000 ha mindre end de to andre metoder.



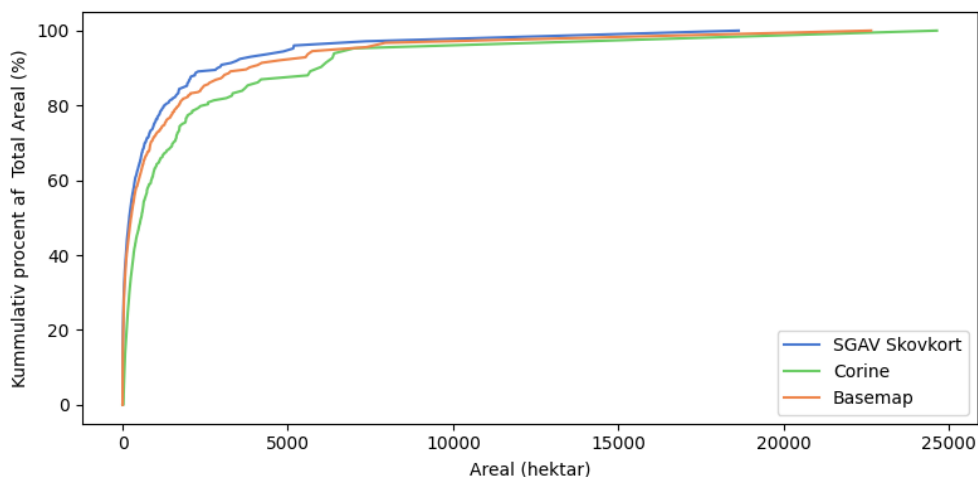
Figur 5.2 Illustration af totalt skovareal i hektar af hhv. CORINE på 520.476, SGAV arealkort på 653.236 og Basemap på 692.603 hektar.

Hvis man sammenligner størrelserne på de polygoner de tre metoder producerer, kan vi se at CORINE producerer nogle meget større sammenhængende skovarealer. Dette skyldes formentlig den meget lavere opløsning på kortet, der er 100m x 100m sammenlignet med Basemaps og SGAVs 10m x 10m.



Figur 5.3: Illustration af fordelingen af størrelserne på de enkelte skov polygoner for hhv. SGAV arealkort, Corine og Basemap.

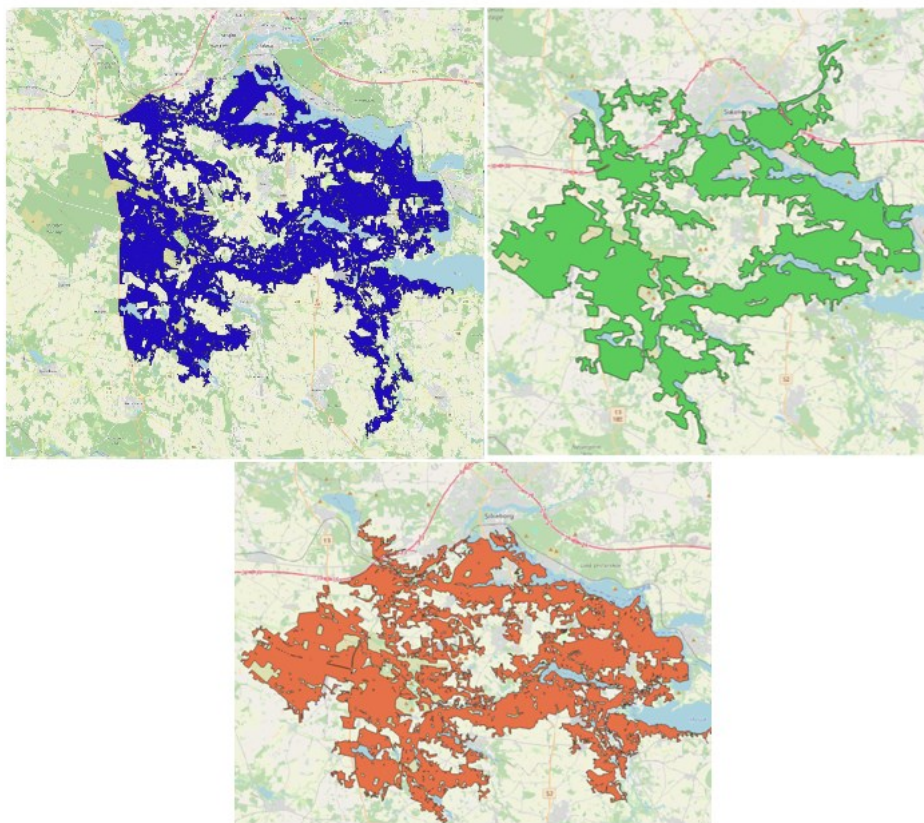
Dette ses også i det kumulative plot, hvor SGAVs areal kort i højere grad består af mange små polygoner, hvorimod CORINE-kortet består af få store polygoner. Basemap er meget tæt på SGAVs kort, men har tendens til at producere marginalt større polygoner.



Figur 5.4: Illustration af den kumulative fordeling af de enkelte polygon arealer for de tre kort over skov.

Et eksempel på denne forskel kan ses ved at sammenligne det største sammenhængende skov-polygon for alle tre kort, nemlig skovområdet syd for Silkeborg, der dækker over bl.a. Støre Højlund Plantage og Skærbæk Plantage. De tre forskellige metoder finder et skov areal på dette område på hhv. 18638 ha (SGAV), 22465 (Basemap) og 24636 (CORINE). Den lave opløsning på CORINE-kortet kan ses da der ikke er nogle "huller" i skovdækket der reflekterer store lysninger, veje eller andet. Derudover så er Gludsted plantage fraregnet netop dette polygon i

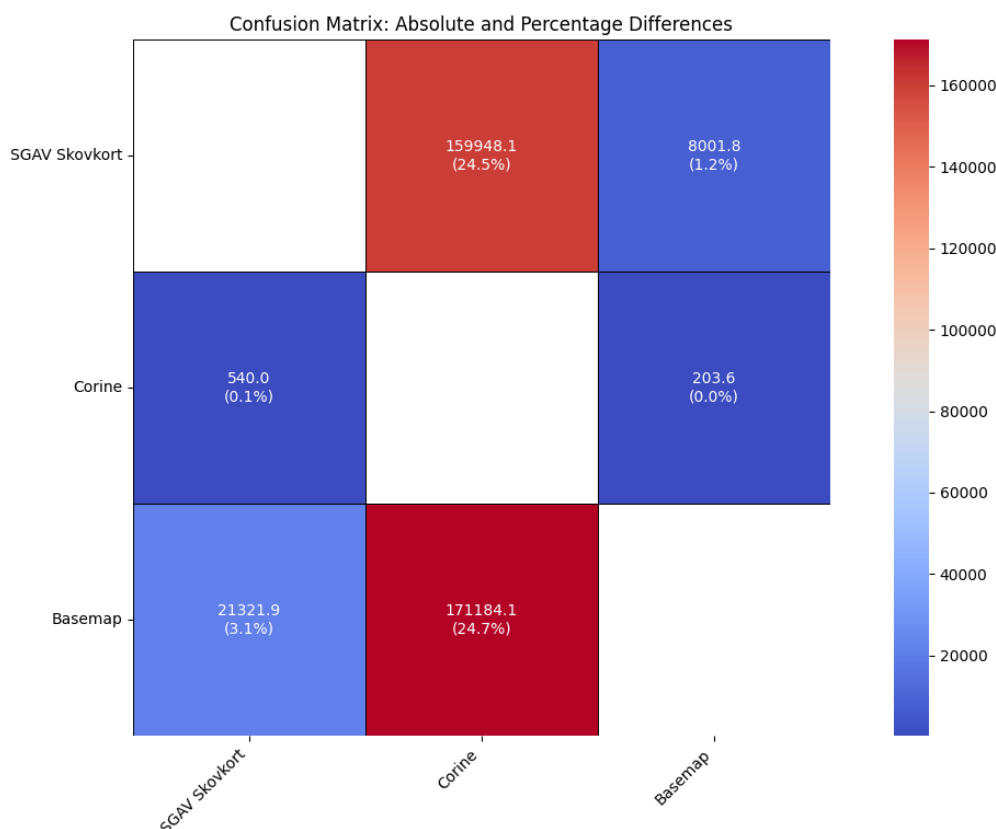
SGAVs skovkort, da skovpolygonet skæres over af en statsvej. Skovarealet i Gludsted plantage er altså ikke fraværende i skovkortet som helhed, men blot ikke en del af det største polygon – hvilket også forklarer hvorfor SGAVs rapporterede areal er væsentligt mindre end Basemaps.



Figur 5.5: Illustration af det største skov polygon for hhv. SGAVs arealkort (Blå), CORINE (Grøn) og Basemap (Orange).

5.2.4 Unikke Polygoner

Der er to forskellige slags uenigheder mellem metoderne; uenigheder med overlap (1) og uenigheder uden overlap (2). Forskel 1 beskriver et scenarie hvor modellerne er enige om at der er en skov i et valgt område, men uenighed om præcis hvor grænserne til denne går, ift udstrækning, skovveje, lysninger mm. Forskel 2 beskriver en uenighed hvor der er intet overlap mellem modellernes skovareal, og at de dermed er fundamentalt uenige. Der er lagt mest vægt på analyser af forskel (2), da disse vidner om mere metodiske forskelle.



Figur 5.6: Illustration af forvirringsmatrice for de tre kort. Antal hektar ét kort (venstre) kan bidrage med til et andet kort (bunden), samt hvor stor en procentdel af referencekortet (bunden) dette areal udgør.

Hvis vi sammenligner de tre kort kan vi se hvor mange hektar ét skovkort (venstre) har kortlagt som ikke har nogen snitflade til skov polygoner i et andet kort (bunden). F.eks. ses det, at det størst mulige bidrag er SGAVs skovkort der har kortlagt 159 948 hektar der ikke har nogen snitflade til skovpolygoner i CORINE, svarende til 24.5% af CORINE-kortets fulde størrelse. Derimod har CORINE kortet kun kortlagt 540 hektar der ikke eksisterer i SGAVs skovkort, svarende til under 0.1% af SGAVs rapporterede skovareal.

Mere interessant er sammenligningen mellem Basemap og SGAVs skovkort. SGAV kortlægger 8000 hektar der ikke findes i Basemap, svarende til 1.2% af Basemaps areal. Derimod kortlægger Basemap 21321 hektar (3.1%) der ikke har nogen snitflade til skovpolygoner i SGAVs arealkort, på trods af at SGAV rapporterer et større nationalt skovareal.

Det vil sige at Basemap generelt finder mere skov end SGAV der hvor der er enighed om at der er skov, og at Basemap anser flere steder som skov som SGAV ikke forudsiger noget skov i. Dette kan skyldes en række faktorer, heriblandt afskæringslaget. Det største polygon hvor der er grundlæggende uenighed om skov, er i Aalborg Lufthavn. Dette område er ekskluderet for SGAVs skovmodel grundet afskæringslaget, men er medtaget i Basemap. Dermed forudsiges der ingen skov i dette område. Det kan tydeligt ses i figur *Figur 3.1*, hvor hele Aalborg Lufthavn befinder sig i afskæringslaget, hvor de største af de grundlæggende uenigheder der er mellem SGAVs kort og Basemap befinder sig.

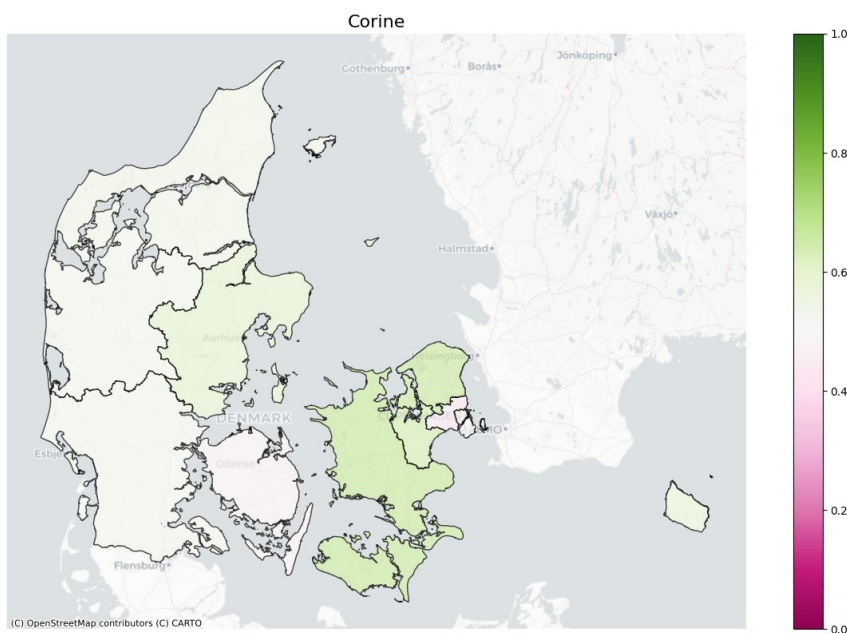
Fælles for alle kort er at det er de mindste polygoner i datasættet der er unikke. Der er dermed ingen store områder modellerne generelt er uenige om.



Figur 5.7: Illustration af skovpolygoner fra Basemap der ikke har noget overlap med SGAVs skovkort (venstre) og SGAVs afskæringslag (højre). Områder der forudsiges i er grønne, og ellers røde.

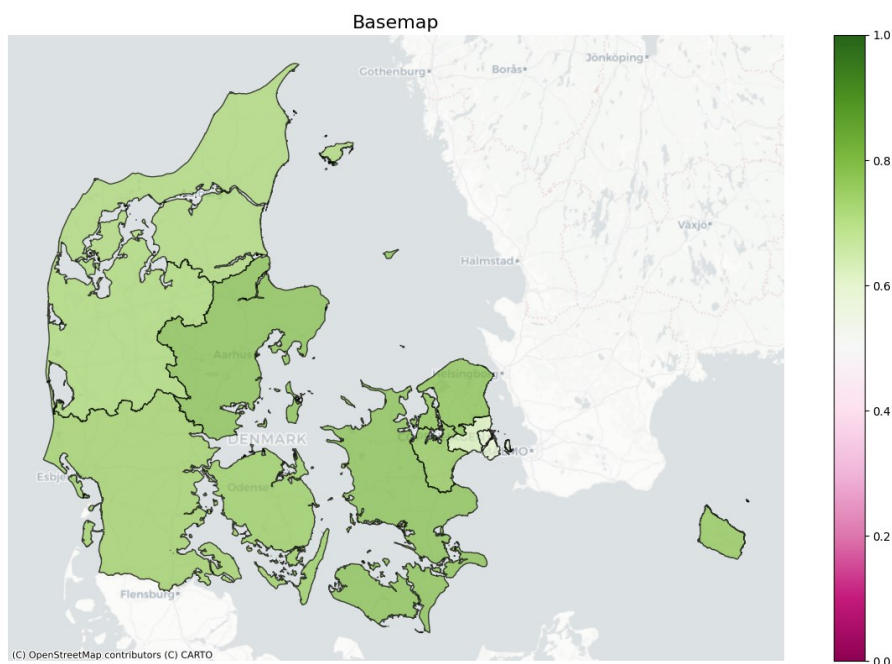
5.2.5 Regionale Forskelle

Ved at benytte sig af Jaccard Indekset $\frac{|A \cap B|}{|A \cup B|}$ kan man beregne en enkelt *enighedsmetrik* der måler hvor enige de to kort er om arealet af et specifikt område, hvor 1 er fuldstændig enig, og 0 er fuldstændig uenighed. Målt på regioner kan man se at SGAV er relativt uenig med CO-RINE kortet.



Figur 5.8: Illustration af regionalt landkort farvelagt efter Jaccard indeks mellem SGAV og CO-RINE

Vi kan til gengæld se at der er meget stor enighed mellem Basemap og SGAVs arealkort. Den region hvor der er størst generel uenighed er i region Hovedstaden, samt Storkøbenhavn. Ved visuel inspektion ses det at det primært drejer sig om at store dele af København og Storkøbenhavn er i SGAVs afskæringslag og bliver dermed aldrig behandlet.



Figur 5.9: Illustration af regionalt landkort farvelagt efter Jaccard index mellem SGAV og CO-RINE

Afsluttende, så er konklusionen at SGAV og Basemap meget ens på nær enkelte uenigheder ift. Afskæringslag. Begge kort er relativt uenige med CORINE kortet på små og mellemstore polygoner da disse ikke klassificeres af CORINE grundet lav opløsning.

5.3 Faglig evaluering

Til yderligere evaluering af arealkortet har SGAV's Skov-enhed kvalitativt evalueret overordnet på Danmarks fem regioner. Hvor maskinlæringsevalueringen udelukkende er blevet brugt til at finde ud af hvilken fejlprocent trædækkegenkendelsen har, kan en menneskelig kvalitativ evaluering finde fejl der er systematiske for metoden og synlige outliers. Skov-enheden har gennemgået arealkortet med ortofotos som reference og indtegnet de områder i arealkortet hvor der er opmærksomhedspunkter. Nedenfor er der beskrevet de opmærksomhedspunkter der er blevet fundet ved den faglige evaluering.

5.3.1 Trædækkegenkendelse

Denne type opmærksomhedspunkter har sit udgangspunkt i trædækkegenkendelsen. Det vil sige at det er når trædækkemodellen genkender noget som værende skov, som evt. ikke er skov eller omvendt.

Afdrifter

Trædækkemodellen er ikke trænet til at genkende områder, hvor der ikke er trædække. Derfor vil afdriftsområder der er større end bufferzonen på 10x10m, ikke tælle som skov. Det betyder at der kan findes større områder i skovene, hvor der er renafdrift, der ikke er med i arealkortet.

Krat

Trælige arter såsom krat, kan også blive genkendt som trædække. Det betyder at der er flere områder, hvori der er pilekrat og bjergfyr som bliver klassificeret som skov.

Ung skov og juletræer

Da trædækkegenkendelsesmotoren ikke specifikt er trænet til at genkende unge træer, vil ung skov og unge juletræsplantager kun blive delvist genkendt, afhængigt hvor mange 10x10 m pixels modellen klassificerer som trædække.

Midlertidigt oversvømmede landbrugsarealer

Der er blevet identificeret et område i landbruget, hvor der er oversvømmelse. Da trædækkegenkendelsesmodellen ikke er trænet op i områder med vand, vil det være tilfældigt hvorvidt den klassificerer disse områder som trædække eller ej.

5.3.2 For- og efterbehandling

Denne type opmærksomhedspunkter har sit udgangspunkt i enten for- eller efterbehandlingsmetoderne. Det vil sige at det kan handle om trædække områder der enten bliver eller ikke bliver taget med pga. afskæringslaget. Det kan også dreje sig om at efterbehandlingen enten laver områder der ikke skulle have været skov om til skov eller omvendt.

Fredskov & Sommerhusområder

Sommerhusområder er med i afskæringslaget, hvorimod fredskov bliver medtaget i grundlaget for arealkortet. Dette betyder at fredskov i sommerhusområder kan blive klassificeret som skov, mens resten af sommerhusområderne ikke vil.

Gældende fredskov

Det anvendte fredskovslag er oprettet pr. 31. december 2022, for at synkronisere fredskovslaget til det de øvrige arealdata som er grundlag for arealkortet. Det er imidlertid blevet identificeret, at fredskovsoprettelsesdatoen også gælder for fredskov der er blevet rettet til. Dette vil sige at oprettelsesdato ikke kan bruges effektivt til arealkortet og fejlen vil blive rettet til i løbet af 2025.

Campingpladser og kolonihaver

Campingpladser og kolonihaver er ikke kortlagt i Danmark, derfor er der ikke et tilgængeligt afskæringslag. Hvis disse arealer ligger udenfor byzone eller på arealer som er kortlagt som friluftsurealer/rekreative arealer, vil de kunne blive kortlagt som skov. Det er tilfældet, hvis trædækket lever op til skovdefinitionens mindstekrav til arealstørrelse. Sådanne arealer er ikke skov men defineres som andet areal med trædække. Eftersom der kan være højt trædække i begge områder er der en vis sandsynlighed for at dele af disse områder bliver kortlagt som skov.

Baghaver med træarealer

Der kan forekomme haver, med trædække som lever op til skovdefinitionens mindstekrav til arealstørrelser, som derfor bliver genkendt som skov. Jf. FAOs skovdefinition er sådanne arealer ikke skov, men det der kaldes andet areal med trædække. Der findes ikke et særskilt datasæt over haver, som ville kunne anvendes som afskæringslag.

Små veje og jernbaner med træer på hver side

Træer langs jernbaner og på begge sider af mindre veje kan teknisk set blive skov, selv om det reelt er to rækker træ på hver side af vejen. Det sker hvis begge træækker og vejen imellem dem samlet er mere end 20 meter bredt og større en 0,5 ha.

Fejlregistreringer – Markbloklag/sommerhusområder/byzoner/industri

Fejlregistreringer af markbloklaget/sommerhusområder/byzoner/industri kan betyde, at der kan være skov i afskæringslaget og skoven derfor ikke er med i arealkortet.

5.4 Fremtidig udvikling

Denne første iteration af Digitalt Skovkort – arealkort er hverken endeligt udtalende eller fuldt dækkende. Der vil være fejl, såsom fredskov oprettelsesdato, der løbende rettes og mangler. Derudover vil der også være opmærksomhedspunkter såsom ung skov og afdrifter, der kan blive tilgodeset i senere udvikling af Digitalt Skovkort. Hvis du som læser finder andre opmærksomhedspunkter i arealkortet eller har idéer til fremtidig udvikling er du velkommen til at sende en mail til skov@sgav.dk med emnet "Digitalt Skovkort – Feedback".

Bilag 1. Supplerende data

Bilag 1.1 Afgrødekoder til udlægning af ikke-træ pixels.

Afgrødekode	Afgrøde
1	Vårbyg
2	Vårhvede
3	Vårhavre
4	Blanding af vårsåede arter
5	Majs til modenhed
6	Vårhvede, brødhvede
7	Korn + bælgssæd under 50% bælgssæd
8	Vårspelt
9	Vinterspelt
10	Vinterbyg
11	Vinterhvede
13	Vinterhvede, brødhvede
14	Vinterrug
15	Vinterhybridrug
16	Vintertriticale
17	Blanding af efterårssåede arter
18	Korn og bælgssæd (over 50 % bælgssæd)
21	Vårraps
22	Vinterraps
23	Rybs
24	Solsikke
25	Sojabønner
30	Ærter
31	Hestebønner
32	Sødlupin
35	Bælgssæd, flerårig blanding
36	Bælgssæd, andre typer til modenhed blanding
40	Oliehør
41	Spindhør
42	Hamp
51	Blanding bredbladet afgrøde, frø/kerne
52	Quinoa
53	Boghvede
54	Bælgssæd blanding
55	Vårrug
56	Vårtriticale
57	Vinterhavre

58	Sorghum
101	Rajgræsfrø, alm.
102	Rajgræsfrø, alm. 1. år, efterårsudlagt
103	Rajgræsfrø, ital.
104	Rajgræsfrø, ital. 1. år efterårsudlagt
105	Timothefrø
106	Hundegræsfrø
107	Engsvingelfrø
108	Rødsvingelfrø
109	Rajsvingelfrø
110	Svingelfrø, bakke- (tidl. Stivbladet)
111	Svingelfrø, strand-
112	Engrapgræsfrø (marktype)
113	Engrapsgræsfrø (plænetype)
114	Rapgræsfrø, alm.
115	Hvenefrø, alm. og krybende
116	Rajgræs, hybrid
117	Rajgræs, efterårsudl. hybrid
118	Rajsvingelfrø, efterårsudlagt
120	Kløverfrø
121	Bælgplanter, frø
123	Valmuefrø
124	Spinatfrø
125	Bederoefrø
126	Blanding af markfrø til udsæd
149	Kartofler, lægge- (certificerede)
150	Kartofler, lægge- (egen opformering)
151	Kartofler, stivelses-
152	Kartofler, spise- (pakkeri, vejsalg)
154	Kartofler, spise- (proces, skrællet kogte)
155	Kartofler, pulver/granules-
156	Kartofler, friteret/chips/pommes frites
160	Sukkerroer til fabrik
161	Cikorierødder
162	Blanding, andre industriafgr.
170	Græs til fabrik (omdrift)
171	Lucerne, slæt
172	Lucernegræs, over 25% græs til slæt inkl. eget foder
173	Kløver til slæt
174	Kløvergræs til fabrik
180	Gul sennep
182	Blanding af oliearter
210	Vårbyg, helsæd
211	Vårhvede, helsæd
212	Vårhavre, helsæd
213	Blandkorn, vårsået, helsæd
214	Korn og bælgssæd, helsæd, under 50% bælgssæd
215	Ærtehelssæd
216	Silomajs
217	Korn og bælgssæd, helsæd (over 50 % bælgssæd)
220	Vinterbyg, helsæd
221	Vinterhvede, helsæd
222	Vinterrug, helsæd
223	Vintertriticale, helsæd

224	Blandkorn, efterårssået helsæd
230	Blanding af vårkorn, grønkorn
234	Korn og bælgssæd, grønkorn, under 50% bælgssæd
235	Blanding af vinterkorn, grønkorn
236	Græs med kløver/lucerne, under 50 % bælgpl. (omdrift) efterårsudlagt i vinterkorn til grønkorn
237	Græs med kløver/lucerne, over 50 % bælgpl. (omdrift) efterårsudlagt i vinterkorn til grønkorn
250	Permanent græs, meget lavt udbytte
251	Permanent græs, lavt udbytte
252	Permanent græs, normalt udbytte
255	Permanent græs, under 50% kløver/lucerne
256	Permanent kløvergræs, over 50% kløver/lucerne
257	Permanent græs, uden kløver
259	Permanent græs, fabrik, over 6 tons
260	Græs med kløver/lucerne, under 50 % bælgpl. (omdrift)
261	Kløvergræs, over 50% kløver (omdrift)
262	Lucernegræs, over 50% lucerne (omdrift)
263	Græs uden kløvergræs (omdrift)
264	Græs og kløvergræs uden norm, under 50 % kløver (omdrift)
266	Græs under 50% kløver/lucerne, ekstremt lavt udbytte (omdrift)
267	Græs under 50% kløver/lucerne, meget lavt udbytte (omdrift)
268	Græs under 50% kløver/lucerne, lavt udbytte (omdrift)
269	Græs, rullegræs
270	Græs til udegrise, omdrift
272	Permanent græs til fabrik
274	Permanent lucernegræs over 25% græs, til fabrik
276	Permanent græs og kløvergræs uden norm, under 50 % kløver
278	Permanent lucerne og lucernegræs over 50% lucerne
279	Permanent kløvergræs til fabrik
280	Fodersukkerroer
281	Kålroer
282	Fodermarvkål
284	Græs med vikke og andre bælgplanter, under 50 % bælgpl.
285	Græs og kløvergræs uden norm, over 50 % kløver (omdrift)
286	Permanent græs og kløvergræs uden norm, over 50 % kløver
287	Græs til udegrise, permanent
305	Permanent græs, uden udbetaling af økologi-tilskud
306	Græs i omdrift, uden udbetaling af økologi-tilskud
308	MFO-brak, sommerslåning
310	Brak, sommerslåning
324	Blomsterbrak
325	MFO-Blomsterbrak
327	MFO-bræmme, sommerslåning
328	MFO-bræmme med blomsterblanding
334	MFO-bræmme, forårsslåning
335	MFO-bræmme, permanent græs, forårsslåning
336	MFO-bræmme, permanent græs, sommerslåning
338	Brak, forårsslåning
339	MFO-brak, forårsslåning
342	Bestøverbrak
343	MFO-bestøverbrak
344	Brak langs vandløb og søer, forårsslåning (alternativ til efterafgrøder)
345	Brak langs vandløb og søer, sommerslåning (alternativ til efterafgrøder)

400	Asieagurker
401	Asparges
402	Bladselleri
403	Blomkål
404	Broccoli
405	Courgette, squash
406	Grønkål
407	Gulerod
408	Hvidkål
409	Kinakål
410	Knoldselleri
411	Løg
412	Pastinak
413	Rodpersille
415	Porre
416	Rosenkål
417	Rødbede
418	Rødkål
420	Salat (friland)
421	Savoykål, spidskål
422	Spinat
423	Sukkermajs
424	Ærter, konsum
429	Jordskokker, konsum
430	Bladpersille
431	Purløg
432	Krydderurter (undtagen persille og purløg)
434	Grøntsager, andre (friland)
448	Medicinpl., en- og toårige
449	Medicinpl., stauder
450	Grøntsager, blandinger
486	Hønsegård uden plantedække
488	Hønsegård, permanent græs
499	Lukket system
501	Stauder
502	Blomsterløg
503	En- og to-årige planter
504	Solbær, stiklingeopformering
506	Stikkelsbær, stiklingeopformering
512	Rabarber
513	Jordbær
540	Tomater
541	Agurker
542	Salat (drivhus)
543	Grøntsager, andre (drivhus)
544	Snitblomster og snitgrønt
545	Potteplanter
547	Planteskolekulturer, stauder
548	Småplanter, en-årige
551	Moskusgræskar
552	Mandelgræskar
553	Centnergræskar
563	Svampe, champignon
564	Containerplads

570	Humle
579	Tagetes, sygdomssanerende plante
596	Elefantgræs
597	Rørgræs
650	Chrysanthemum Garland, frø
651	Dildfrø
652	Kinesisk kålfrø
653	Karsefrø
654	Rucolafrø
655	Radisefrø (inklusive olieræddikefrø)
656	Bladbedefrø, rødbedefrø
657	Grønkålfrø
658	Gulerodsfrø
659	Kålfrø (hvid- og rødkål)
660	Persillefrø
661	Kørvelfrø
662	Majroefrø
663	Pastinakfrø
664	Skorzonerrod/skorzonerrodfrø
666	Purløgsfrø
667	Timianfrø
668	Blomsterfrø
701	Grønkorn af vårbyg
702	Grønkorn af vårhvede
703	Grønkorn af vårhavre
704	Grønkorn af vårrug
705	Grønkorn af vårtriticale
706	Grønkorn af vinterbyg
707	Grønkorn af vinterhvede
709	Grønkorn af vinterrug
711	Grønkorn af vintertriticale
921	Intern kode: Bar jord

Bilag 1.2 Habitatnaturtyper til ikke-træer pixels.

'Hav',
'Marint fladvand',
'Strand',
'Strandvold med enårige',
'Strandvold med flerårige',
'Kystklint/klippe',
'Enårig strandengsvegetation',
'Vadegræssamfund',
'Strandeng',
'Indlandssalteng',
'Forklit',
'Hvid klit',
'Grå/grøn klit',
'Klithede',
'Havtornklit',
'Grårisklit',
'Klitlavning',
'Enebærklit',
'Revling-indlandsklit',

'Græs-indlandsklit',
'Sø',
'Våd hede',
'Lagt krat',
'Højt krat',
'Tør hede',
'Rydning',
'Enekrat',
'Græsmark',
'Tør overdrev på kalkholdigt sand',
'græsland',
'Kalkoverdrev',
'Surt overdrev',
'Tidvis våd eng',
'Urtebræmme',
'Eng',
'Mose',
'Højmose',
'Nedbrudt højmose',
'Hængesæk',
'Tørvelavning',
'Avneknippemose',
'Kildevæld',
'Rigkær',
'Indlandsklippe'

6. Bibliography

- Chuvieco, E. M. (2002). Assessment of Different Spectral Indices in the Red-Near-Infrared Spectral Domain for Burned Land Discrimination. *Remote Sensing of Environment*, 112, 2381-2396.
- Copernicus. (2024). *Copernicus Land Monitoring Service*. Hentet fra land.copernicus.eu: <https://land.copernicus.eu/en/technical-library/product-user-manual-of-seasonal-trajectories/@@download/file>
- databrickslabs. (2024). *Mosaic*. Hentet fra Mosaic: <https://databrickslabs.github.io/mosaic/index.html>
- Doninck, J. V., & Tuomisto, H. (2017). Influence of Compositing Criterion and Data Availability on Pixel-Based Landsat TM/ETM+ Image Compositing Over Amazonian Forests. *IEEE*.
- Francini, S., Hermosilla, T., Coops, N., Wulder, M., White, J., & Chirici, G. (2023). An assessment approach for pixel-based image composites. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Frantz, D., Stellmes, M., Hill, J., & Röder, A. (2017). Phenology-adaptive pixel-based compositing using optical earth observation imagery. *Remote Sensing of Environment*.
- GeoDanmark. (2025). *Baggrundskort*. Hentet fra GeoDanmark: <https://www.geodanmark.dk/home/vejledninger/baggrundskort/>
- Gillis, M., Omule, A., & Brierly, T. (2005). Monitoring Canada's forests: The National Forest Inventory. *The Forestry Chronicle*.
- Gitelson, A., Chivkunova, O., & Merzlyak, M. (2009). Non-Destructive Estimation of Anthocyanins and Chlorophylls in. *American Journal of Botany*.
- Gitelson, A., Gritz, Y., & Merzlyak, M. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*.

- Griffiths, P., Nendel, C., & Hostert, P. (2019). Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*.
- <https://land.copernicus.eu/>. (9 2024). *CORINE Land Cover*. Hentet fra <https://land.copernicus.eu/en/products/corine-land-cover>
- J., X. R. (2018). Extraction of High-Precision Urban Impervious Surfaces from Sentinel-2 Multispectral Imagery via Modified Linear Spectral Mixture Analysis. *Sensors* , 2873.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting. *Proceedings : Annual Conference on Neural Information Processing Systems*.
- Levin, G. (14. 02 2024). *Basemap*. Hentet fra AU - Department of Environmental Science: <https://envs.au.dk/en/research-areas/society-environment-and-resources/land-use-and-gis/basemap>
- Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems*.
- Matasci, G., Hermosilla, T., Wulder, M., White, J., Coops, N., Hobart, G., & Zald, H. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sensing of Environment*.
- Morresi, D., Marzano, R., Lingua, E., Motta, R., & Garbarino, M. (2022). Mapping burn severity in the western Italian Alps through phenologically coherent reflectance composites derived from Sentinel-2 imagery. *Remote Sensing of Environment*.
- Nord-Larsen, T. (2024). *Skov eller ej*. København: Institut for Geovidenskab og Naturforvaltning, Københavns Universitet.
- Nord-Larsen, T., Tsatsakis, M., Li, S., Avila, L., Brandt, M., Liu, S., . . . Fensholt, R. (2023). *Forest Plan - pilot project on mapping of forest resources*. Copenhagen: Department of Geosciences and Natural Resource Management, University of Copenhagen.

- Segal, M., & Sellers, R. (1982). Kinetics of metal oxide dissolution. Reductive dissolution of nickel ferrite by tris(picolinato)vanadium(II). *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*.
- SGAV. (2025). *Fredskovpligten og tilsyn*. Hentet fra Styrelsen for Grøn Arealomlægning og Vandmiljø: <https://sgavmst.dk/skovbrug-og-landbrug/skovbrug/fredskovpligten-og-tilsyn>
- SGAV. (2025). *Skovovervågning og skovstatistik*. Hentet fra Styrelsen for Grøn Arealomlægning og Vandmiljø: <https://sgavmst.dk/skovbrug-og-landbrug/skovbrug/skovovervaagning-og-skovstatistik>
- Vicharnakorn, P., Shrestha, R., Nagai, M., & Salam, P. (2014). Carbon Stock Assessment Using Remote Sensing and Forest Inventory Data in Savannakhet, Lao PDR. *MDPI Remote Sensing*.
- wekeo. (2024). *High Resolution Vegetation Phenology and Productivity*. Hentet fra wekeo.eu: <https://www.wekeo.eu/use-cases/hr-vegetation-phenology-and-productivity-service>
- What is a decision tree?* (11. September 2024). Hentet fra IBM: <https://www.ibm.com/topics/decision-trees>
- Xu, H. (2006). Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *International Journal of Remote Sensing*, 14 , 3025-3033.

Bilag 1.3 Litteraturliste.

- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Chergui, N., Saker, Y., & Belayat, Y. (2023). Forest Tree Species Classification Using Machine Learning and Satellite Images. *2023 International Conference on Earth Observation and Geo-Spatial Information (ICEOGI)*, 1–5. <https://doi.org/10.1109/ICEOGI57454.2023.10292964>

- ESA - Sentinel-2. (2024). https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2
- European Environment Agency. (2024, juni 6). *Copernicus Land Monitoring Service* (L. Forslund, Red.).
- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87.
<https://doi.org/10.1016/j.rse.2016.08.013>
- Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany*, 96(10), 1861–1868. <https://doi.org/10.3732/ajb.0800395>
- Klimadatastyrelsen. (2024, august 21). *Dataforsyningen*. <https://dataforsyningen.dk/>
- LightGBM - Microsoft Research. (2016). **microso**<https://www.microsoft.com/en-us/research/project/lightgbm/>
- Mellor, A., Haywood, A., Jones, S., & Wilkes, P. (2012). *FOREST CLASSIFICATION USING RANDOM FORESTS WITH MULTISOURCE REMOTE SENSING AND ANCILLARY GIS DATA*.
- Miljøstyrelsen. (2022). *MST AI model til prædiktion af §25 skov*. <https://mst.dk/media/pirgw35c/presentation-af-ai-til-25-skov-290322.pdf>
- Nord-Larsen, T., Østergaard, M. J., Riis-Nielsen, T., Thomsen, I. M., Bentsen, N. S., & Jørgensen, B. B. (2022). *Skovstatistik 2022. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet*. https://static-curis.ku.dk/portal/files/376995779/Rapport_Skovstatistik_2022_web.pdf
- WEkEO. (2024, august 21). *New HR Vegetation Phenology and Productivity service*. <https://www.wekeo.eu/use-cases/hr-vegetation-phenology-and-productivity-service>



Styrelsen for Grøn
Arealomlægning og Vandmiljø
Tolderlundsvej 5
5000 Odense C