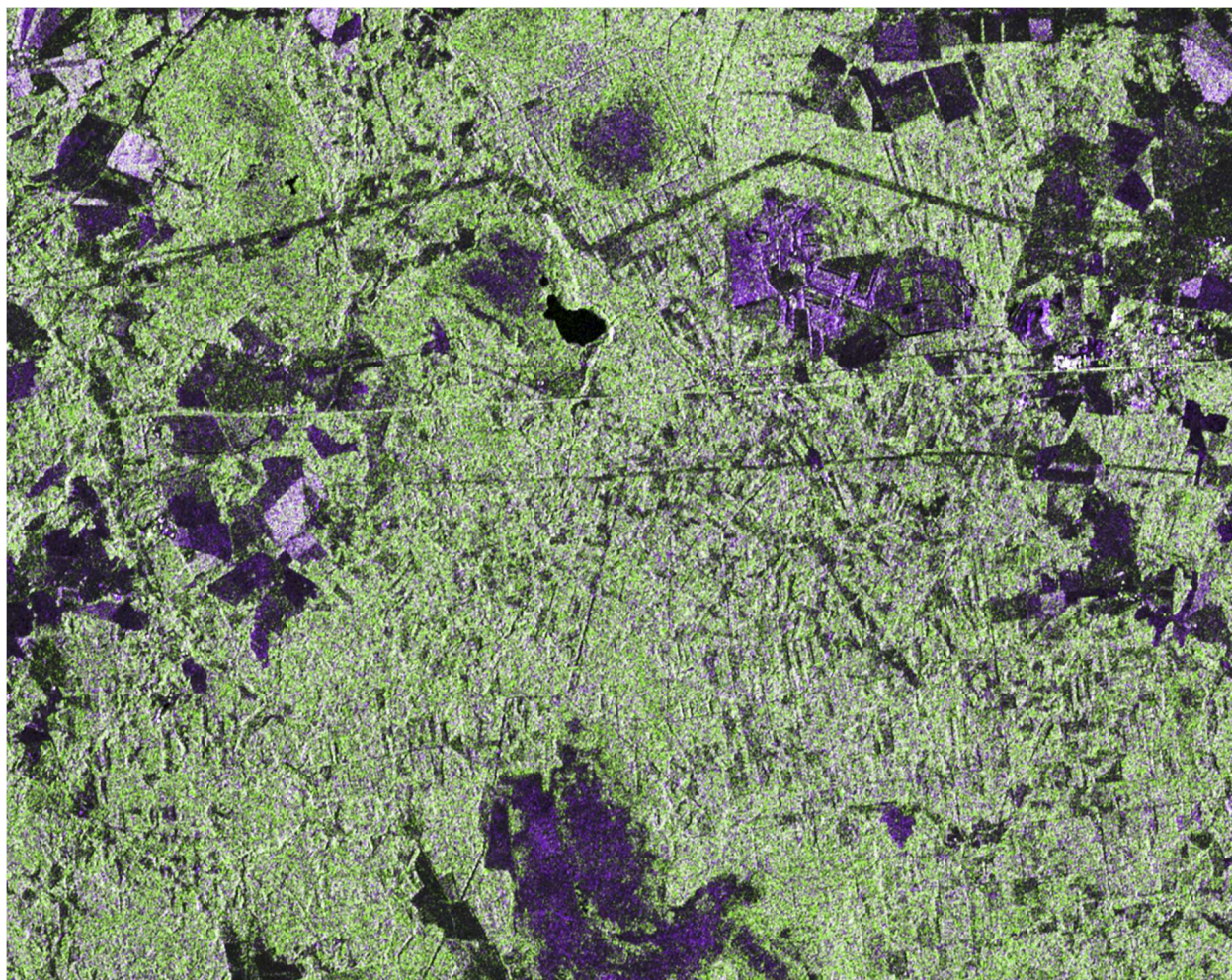


Metoodika kirjeldus



Koostas: Tauri Tampuu

Tadus- ja arendusjuht, SAR ekspert

KappaZeta OÜ

KAPPAZETA

Sisukord

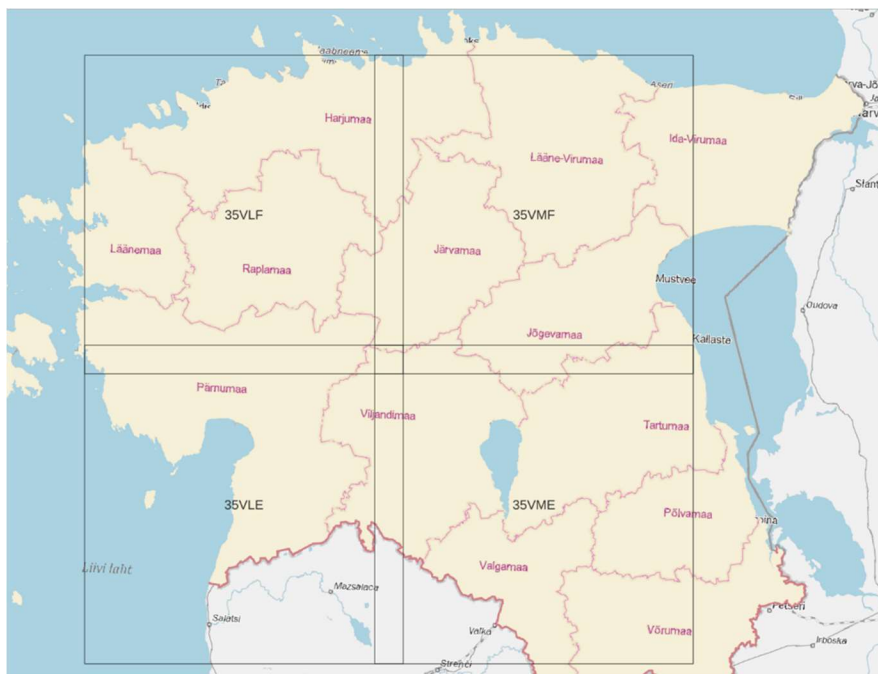
1.	Kokkuvõtte tehtud tööst.....	1
2.	Tarne kirjeldus	1
3.	Kasutatud satelliidid, sisendkanalid, satelliidiandmete aegread ja võrdlusandmed	2
4.	Mudeli kirjeldus.....	3
5.	Täpsuse hindamise meetoodika kirjeldus	5
5.1.	Täpsuse hindamine mudeli arendamise käigus	5
5.2.	Tarnitud tulemuste täpsuse hindamine	5
6.	Täpsuse hinnang.....	6
6.1.	Aasta 2022	6
6.2.	Aasta 2023	9
7.	Vea-allikad ja mudeli edasiarendamine	11

1. Kokkuvõtte tehtud tööst

KappaZeta OÜ on arendanud välja närvivõrkudel põhineva mudeli, mis piloteerib lageraite operatiivse tuvastamise võimalikkust tuleviku lahenduste tarbeks. Loodud mudelil on suur üldistusvõime, sest mudel suudab sisendina võtta iga metsaeraldise (sõltumata sellest, kas eraldisel kasvab noor või vana mets; mudeli sisendiks sobivad ka metsaeraldised, kus mets on hiljuti või aastaid tagasi maha võetud) ja väljastada hinnangu, kas eraldisel toimus vaadeldava perioodi jooksul lageraie. Mudel klassifitseerib metsaeraldised sündmuste klassi (koos toimumise aja hinnanguga) ja mittesündmuste klassi. Mudel töötab metsaeraldistel, mis on vähemalt 0.5 ha suured. Mudel saavutas lageraite tuvastamisel täpsuse (precision) kuni 0.95; F-skoor (F-score) üle sadade tuhandete metsaeraldise oli kuni 0.89, mis annaks metsateatiste puhul F-skoori 0.92. Mudeli täpsust kontrolliti RMK 2022 ja RMK 2023 andmete vastu ja eksperthinnanguga Sentinel-2 RGB piltide ja mudeli sisendina kaustatud aegridade põhjal.

2. Tarne kirjeldus

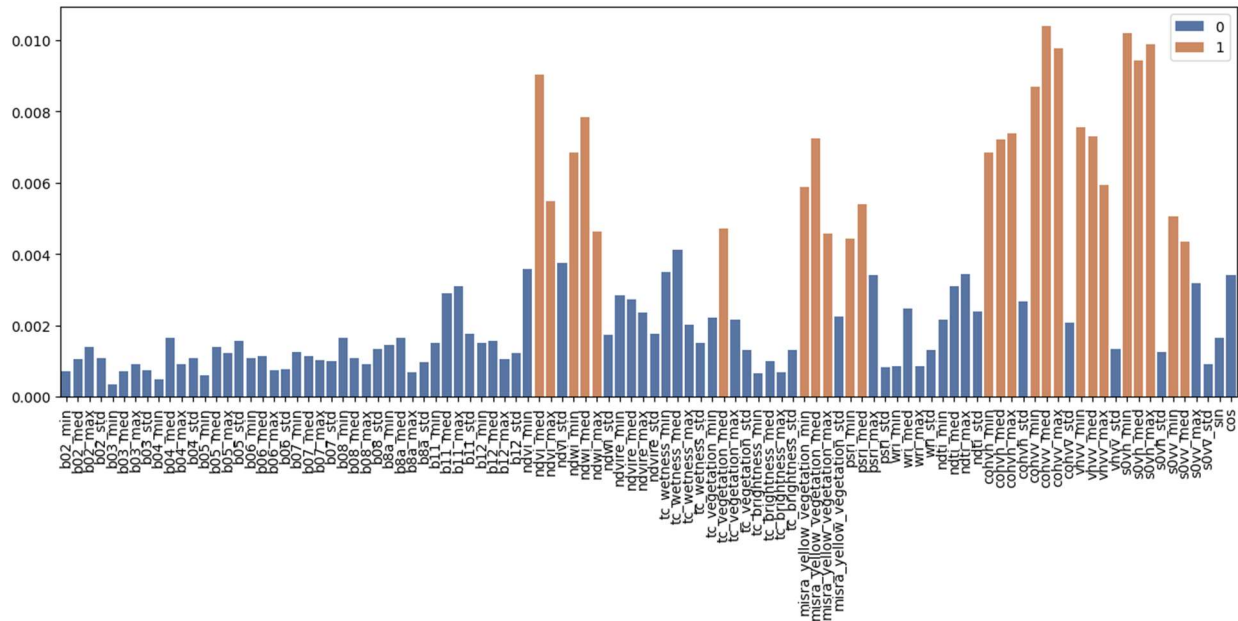
Mudel väljas shp failid, mis sisaldavad mitte-lageraiesündmusi ja lageraiesündmusi koos lageraieaja hinnanguga üle suurema osa Mandi-Eesti (ruudustiku sisse jääv ala Joonisel 1) vaatlusperioodil 2022–2023. Kummagi aasta kohta on eraldi shp fail. Mudel väljastab hinnangud kõigile metsaeraldistele (nii riigi- või eraomanduses olevatele), mis on alla laetud Metsaregistrist (veebruaris 2024), st ruumikujud, millele hinnang antakse pärinevad Metsaregistrist. Mudeli treenimiseks kasutati RMK-lt saadud ajaloolist raieinfot ja RMK-lt saadud ruumikujusid (st mitte Metsaregistrist alla laetavaid ruumikujusid).



Joonis 1. Mudeli väljastab hinnangud neljale Sentinel-2 jalajäljele vastaval alal, näidatud ruuduna Eesti kaardil, mis katab suurema osa Mandri-Eestist.

3. Kasutatud satelliidid, sisendkanalid, satelliidiandmete aegread ja võrdlusandmed

Mudeli sisendiks on satelliitkaugseire (ingl. Earth Observation, EO) andmed, mis on mudelisse sisestatud kui statistikute aegread metsaeraldiste kaupa. Andmesisendina kasutati järgnevaid kaugseiresensoreid: tehisavaradar (ingl. Synthetic Aperture Radar, SAR) Sentinel-1 (S1) ja multispektraalne satelliit Sentinel-2. Mudeli väljatöötamisel kasutatud S1 kanalid olid: VV tagasihajumine, VH tagasihajumine, VV interferomeetriline koherentsus ja VH interferomeetriline koherentsus. Kasutatud S2 kanalid ja neist tuletatud indeksitest olid: kanalid 2–12 (v.a 9 ja 10) ning indeksid NDVI, NDWI, NDVIRE, Tasseled Cap (TC) MYVI (Misra Yellow Vegetation Index), PSRI (Plant Senescence Reflectance Index) ja WRI indeksid. Igas S1 või S2 kanali ja indeksi kohta arvutati iga pildi jaoks kõigi uuritud ruumikujude (mis vastab metsaeraldisele) statistikud: mediaanväärtus (med), keskmine, standardhälve, miinimum (min) ja maksimum (max) väärtus. Statistikud koondati iga ruumikuju jaoks kanalite kaupa aegridadeks. Joonis 2 näitab sisendkanalite ja statistikute olulisust mudeli tulemusel. Lõpuks valminud mudelis kasutati vaid neid kanali-statistiku kombinatsioone, mis omasid mudeli otsusele olulist positiivset mõju: NDVI med, max; NDWI med, min, max; TC med; MYVI med, min, max; PSRI med, min, VH koherentsus med, min, max; VV koherentsus med, min, max; VH/VV tagasihajumise suhe med, min, max; VH tagasihajumine med, min, max; VV tagasihajumine med, min. Mudeli otsusele mitteolulist mõju omavate sisendite kaasamine ajaks mudelit segadusse ja vähendaks väljundi täpsust.



Joonis 2. Mudeli sisendkanalid ja statistikud olulisuse järgi mudeli poolt tehtavasse otsusesse. 1 tähendab oluline, 0 tähendab mitteoluline. Valminud mudeli sisendina kasutati vaid olulisi (1) kanali-statistikute kombinatsioone. Mudeli Y-teljel on sisendi olulisusskoor (Feature attribution score).

Masinõppemudeli arendamisel kasutati RMK-lt saadud metsaeraldisi koos raieinfoga. Iga metsaeraldise kohta oli kasutada info, kas seal toimus raie (raieliik ja sündmuse toimumise aeg) või ei toimunud raie (nn mittesündmus, mida samuti kasutati mudeli treenimisel). RMK-lt saadud andmeid kasutati mudeli treenimisel kui referentsandmeid (nn tõekriteeriumi). Samuti kasutati RMK andmeid mudeli täpsuse hindamiseks. RMK poolt olid kasutada andme 2018–2023, mudeli treenimiseks kasutati neist aastaid 2018–2021 üle Eesti (Joonisel 1 näidatud neli S2 jalajäljele vastav ala). Mudelit trenniti jagades RMK metsaeraldised treening-, valideerimis- ja testiandmekoguks. Mudel ei ole treenimise käigus näinud aastaid 2022 ja 2023, mida kasutatakse mudeli täpsuse hindamiseks.

RMK andmed puhastati dubleeritud või vigastest kirjetest; raieinfo viidi vastavusse raie toimumise aastale vastanud metsaeraldise ruumikujuga. Need sammud tagasid RMK andmete maksimaalse koostööla maapealsele reaalsusele, filtreerides maksimaalselt välja potentsiaalsed andmete RMK andmebaasi sisestamisel tekkinud vead, RMK andmepaasist pärimisel tekkinud ebatäpsused ja vähendades metsaeraldise ruumikujude läbi aja muutumisest tingitud mõju mudeli täpsusele treenimise käigus.

4. Mudeli kirjeldus

Tulenevalt lageraiete tuvastamise ülesande olemusest formuleeriti probleem kui aegridade binaarse klassifitseerimise ülesanne, mille käigus mudel otsib aegridadel libiseva aknaga LR sündmust ja jaotab ruumikujud kaheks: kus toimus lageraie vs mitte-lageraie (kaasaarvatud toimus mõni muud tüüpi raie, mis ei olnud lageraie). Mudel kasutab 1D-CNN-i arhitektuuri ja on treenitud optimeerima täpsust (ingl. Precision). Enne kandidaatmudelite treenimist tuli andmekogu teha mõningad kohandused, sealhulgas:

- 1) Klasside tasakaalustamatusega arvestamine: mudelisse sisestati sündmused lageraie vs mitte-lageraie suhtega 1:3, alamesindatud raieliikide oversampling, nende parema esindatuse tagamiseks
- 2) S1 ja S2 aegreapunktid pandi vastama ühtsele 6-päevasele ajasammule (resampling), kasutades ajaperioodi keskmisi ja või pikemate andemaukude korral interpoleeritud väärtusi
- 3) RMK poolt raporteeritud raite (sündmuste) ümber konverteerimine 14-päeva pikkuseks, mis lõpevad RMK poolt antud raie lõpukuupäevaga, tagamaks referentsandmete ja satelliidiandmete ülekattuvus. Niimoodi saadud sündmuse alguskuupäevale lisati lubatud puhver 14 päeva.

Mudel väljastab lageraie tõenäosuse (0–100%) kui tõenäosuskõvera läbi aja. Mudeli poolt antava tõenäosushinnangu aluseks on sündmuse (lageraie) signatuur, mis koosneb kolmest komponendist: 1) sündmuse eelne 2) sündmuse aegne ja 3) sündmuse järgne signatuur.

Mudeli arendamisel on kasutatud järgmisi parameetreid ja tehtud järgmisi kohandusi:

- Mudel otsib sündmust libiseva akna sees. Mudel vaatles perioodi, mis ulatub kuni ± 60 kaugusele potentsiaalsest sündmusest. 60 päeva on valitud selliselt, et oleks piisavalt suur tõenäosus, et akna sisse satuks enamasti vähemalt üks optiline pilt sõltumata aastaajast (1 pilt on absoluutne miinimum, millest üldjuhul kvaliteetse otsuse tegemiseks ei pruugi siiski piisata). See tähendab, et pilviste olude puhul ootas mudel pilvevabade S2 piltide ilmumist kuni 60 päeva. Pilvitutel perioodidel suudab mudel kindla otsuse teha vähemate päevad jooksul.
- Mudel on optimeeritud maksimeerima täpsust (precision)

Tõenäosuskõvera kuju ja tõenäosuse väärtuse järgi klassifitseerisime mudeli tulemused kahe kasutajuhi tarbeks. Tulemuste klassideks jaotamine on arbitraarne – st klasside arv ja piirid on defineeritavad vastavalt vajadusele. Tarne käigus simuleerisime kahte stsenaariumit: 1) simuleerib KeA kasutusjuhtu, kus oluline on punase ja rohelise klassi võimalikult kõrge täpsus ja ebakindlamad tuvastused koguti vähem usaldusväärsesse kollasesse klassi 2) simuleerib KAURi kasutusjuhtu, kus oluline on tulemuste terviklikkus ja tasakaalustatus, mis saavutati läbi kollase klassi lävendipõhise jaotamise punase ja rohelise klassi vahel.

Kolme klassiga KeA kasutusjuht:

- Lageraie (punane klass): tõenäosuskõvera keskmine väärtus on adaptiivselt kas ≥ 0.90 või ≥ 0.75 (talvekuudel: jaan, veeb, märts, nov, dets) perioodil >5 päeva
- Potentsiaalne lageraie (kollane klass): tõenäosuskõvera keskmine väärtus ≥ 0.40 perioodil >5 päeva
- Mitte-lageraie (roheline klass): tõenäosus alla < 0.40 perioodil >5 päeva, sh kõik need juhud, kui tõenäosuskõver ei kerkinud üle 0

Kahe klassiga KAURi kasutusjuht:

- Lageraie (punane klass): tõenäosuskõvera keskmine väärtus on adaptiivselt kas ≥ 0.50 või ≥ 0.40 (talvekuudel: jaan, veeb, märts, nov, dets) perioodil >5 päeva
- Mitte-lageraie (roheline klass): tõenäosus adaptiivselt alla < 0.50 või < 0.40 perioodil >5 päeva, sh kõik need juhud, kui tõenäosuskõver ei kerkinud üle 0

Mudeli väljud on arvutatud selliselt, et 2022. aasta aegridu on pikendatud 2021. aasta novembrisse, võimaldamaks mudelile parem juurdepääs sündmuse eelsele ajale pilvisel

talveperioodil. Kuna mudeli sisendiks olnud EO aegread lõppesid 2023 detsembriga, siis 2023. aasta lõpus on vähem tuvastatud sündmusi, sündmuste tõenäosuse hinnang on madalam ja valepositiivsete arv oluliselt suurem. Põhjuseks see, et sündmuse järgset signatuur ei saa enamasti veel piisava kindlusega tuvastada, kuna puuduvad seda võimaldavad pilvevabad optilised pildid, mille võiks saada talve teisest poolest või kevade algusest. Seetõttu pidasime vajalikuks 2023. aasta lõpuosa (kolm viimast kuud – oktoober kuni detsember) mitte tarnida. 2022. aasta on tarnitud 1. jaanuarist kuni 31. detsembrini ja annab ülevaate mudeli võimekusest rasketes (pilvestes oludes). Samuti 2023 aasta suvi (juulis ja augustis) oli väga pilvine, mis annab võimaluse hinnata mudeli täpsust ka pilvise suve tingimustes. Sellest, kuidas pilvisest talvest tingitud raskustega tulevikus paremini toime tulla on täpsemalt peatükis 7 (Vea-allikad ja mudeli edasiarendamine).

5. Täpsuse hindamise metoodika kirjeldus

5.1. Täpsuse hindamine mudeli arendamise käigus

Treenitud kandidaatmudelite võrdlemiseks pandi kõrvale RKM-lt pärit andmetest valideerimis- ja testiandmekogu, mille peal võrreldi mudelite F-skoori (F-score), AUC-d ning lageraiaks (LR) klassifitseeritud klassi täpsushinnanguid (täpsus, ingl. Precision, ja saagis, ingl. Recall). Valideerimise eesmärk oli suurendada positiivse klassi üldist F-skoori ja vähendada valepositiivseid tulemusi, valides kandidaatmudelite hulgast parima ja kasutades sobivaid lävendeid. Peale mudeli väljavalimist, testiti mudelit testandmekogul, hindamaks valitud mudeli täpsust. Niimoodi hangitud teadmisi kasutati lõpliku (piloodi käigus valminud) mudeli loomiseks.

Piloodi käigus valminud mudeli tulemused jaotati tõenäosusklassideks (punane, kollane, roheline). Klassideks jaotamise täpsust hinnati korduvalt võrdluses RMK-lt pärit andmetega ja läbi visuaalse valideerimise, kuni lõplike tulemuste fikseerimiseni. Visuaalne valideerimine kujutas endast Sentinel-2 RGB piltide ja mudeli otsuse aluseks olevate S1 ja S2 aegridade visuaalselt inspekteerimist ning selle põhjal eksperthinnangu langetamist raie toimumise/mittetoimumise kohta. Eramaal toimunud lageraiete tuvastamise täpsust sai hinnata vaid läbi visuaalse valideerimise, kuna andmed raiete kohta eramaalt puudusid.

5.2. Tarnitud tulemuste täpsuse hindamine

Kõik tarnitud metsaeraldised jagunevad omandivormi järgi. Eeldasime kvaliteedikontrollis, et „erametsaks“ saab lugeda eraldised mille omandivorm on E, F, J, S, Y ja „riigimetsaks“ ülejäänud (A, M, R, -). Kvaliteedikontroll oli kaheastmeline:

1) Andmebaaside võrdlusel põhinev automaatne kontroll RMK metsaeraldistel aastate 2022 ja 2023 kohta. Automaatne kvaliteedikontroll sisaldas vaid neid riigi metsaeraldise, millel oli vaste RMK 2023. aasta shp failis eraldis_2023_region.shp. NB! Mitte kõik RMK sündmuste failis RMK_raied_2015_2022_v2.xlsx sisalduvad eraldised ei leidnud 2023 RMK shp failis vastet; RMK 2022. aasta shp failiga oli erinevus oluliselt suurem. Seetõttu kasutasime mõlema aasta puhul 2023. aasta shp faili. Eraldised failist eraldis_2023_region.shp jagasime binaarselt „lageraieteks“ ja „mitte-lageraieteks“ vastavalt sellele, kas vaadeldava aasta jooksul toimus RMK andmebaaside RMK_raied_2015_2022_v2.xlsx ja RMK_raied_2023_19022024.xlsx järgi metsaeraldisel lageraie või mitte. Tabelis 1 on näidatud RMK metsaeraldiste arv, mis läbisid automaatse kvaliteedikontrolli, klasside ja aastate lõikes.

Tabel 1: RMK metsaeraldiste arv, mis läbisid automaatse kvaliteedikontrolli, klassides 3- ja 2-klassilise mudeli korral (vastavalt KeA ja KAUR kasutusjuhud).

Mudel	3-klassiline mudel 2022	2-klassiline mudel 2022	3-klassiline mudel 2023	2-klassiline mudel 2023
Klass	Eraldiste arv klassis	Eraldiste arv klassis	Eraldiste arv klassis	Eraldiste arv klassis
Punane	3839	4741	2245	3344
Kollane	1141		1651	
Roheline	358091	358330	355610	356162
Kokku	363071	363071	359506	359506

2) Visuaalne kontroll, mis tugines Sentinel-2 RGB piltide ja mudeli otsuse aluseks olevate S1 ja S2 aegridade visuaalse inspekteerimise põhjal eksperthinnangu langetamisele kokku 1200-l juhuslikult valitud riigi ja eraomandis metsaeraldisel aastatest 2022 ja 2023:

100 (juhuslikku eraldist) x 3 (klassi) x 2 (omandivormi) x 2 (aastat) = 1200 (eraldist)

Hinnang mudeli väljundile jaotati mõlemas kvaliteedikontrolli astmes neljaks:

- Tõene positiivne (true positive, TP): nii mudel kui RMK andmed või visuaalne inspeksioon, vastavalt, kas tegemist oli automaatse või visuaalse kvaliteedikontrolliga, kinnitasid raiet
- Valepositiivne (false positive, FP): mudel ennustas raie, kuid RMK andmed / visuaalne inspeksioon raiet ei kinnitanud
- Tõene negatiivne (true negative, TN): nii mudeli kui RMK andmete / visuaalse inspeksiooni järgi raiet ei toimunud
- Valenegatiivne (false negative, FN): mudeli järgi raiet ei toimunud, kuid RMK andmetel / visuaalse inspeksiooni järgi toimus raie

6. Täpsuse hinnang

6.1. Aasta 2022

Tabelites 2 ja 3 on kokku võetud mudeli väljundi automaatse kvaliteedikontrolli tulemused 2022. aastal KeA ja KAUR stsenaariumi jaoks. KeA stsenaariumi puhul, milles on vajalik eelkõige maksimaalselt kõrge tuvastamise täpsus (ja kollane klass on, kui ebakindel, kõrvale heidetud) on **punase klassi täpsus 0.95** (ingl. precision) ja saagis 0.84 (ingl. recall), st 95% lageraie klassifitseeritust ka tööpoolest oli lageraie ja kõrge täpsuse hinnaks oli, et 16% tegelikult toimunud lageraieid jäi tuvastamata, kuuludes roheline klassi. Mudel leiab sadadest tuhandetest metsaeraldistest hästi üles need, kus toimus lageraie, mida kinnitab kaalumata keskmine **F-skoor 0.94** (ingl. macro average F-score) (Tabel 2). KAURi stsenaariumi puhul, kus oluline on mudeli tasakaalustatus kahe klassi vahel, on **punase klassi täpsus** mõnevõrra madalam (**0.88**), aga klasside ülene kaalumata keskmine **F-skoor** kõrge **0.93** (Tabel 3).

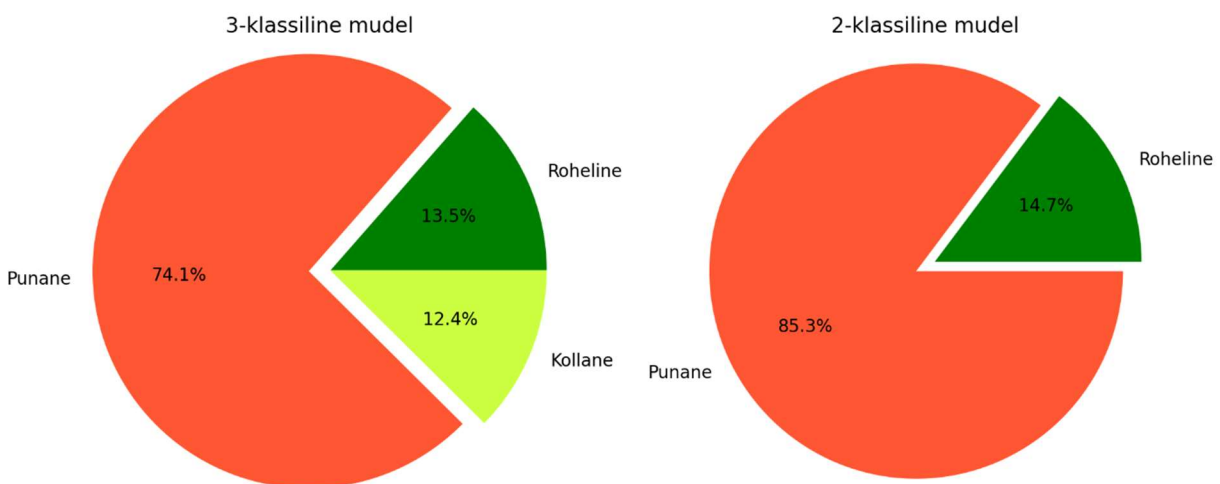
Tabel 2: 3-klassilise mudeli (KeA kasutajajuht) automaatse kvaliteedikontrolli tulemus RMK metsaeraldistel aastal 2022, kui kollast, st ebakindlat klassi mitte arvesse võtta. Võrdluse alus (referents andmed RMK-lt) näitab eraldiste hulka, mis sellesse klassi tegelikult oleksid pidanud kuuluma (mis on mudeli väljundi seisukohast TP + FN).

	Täpsus	Saagis	F-skoor	Võrdluse alus
Roheline klass	1.00	1.00	1.00	357602
Punane klass	0.95	0.84	0.89	4328
Õigsus (ingl. accuracy)			1.00	361930
Kaalumata keskmine	0.97	0.92	0.94	361930
Kaalatud keskmine	1.00	1.00	1.00	361930

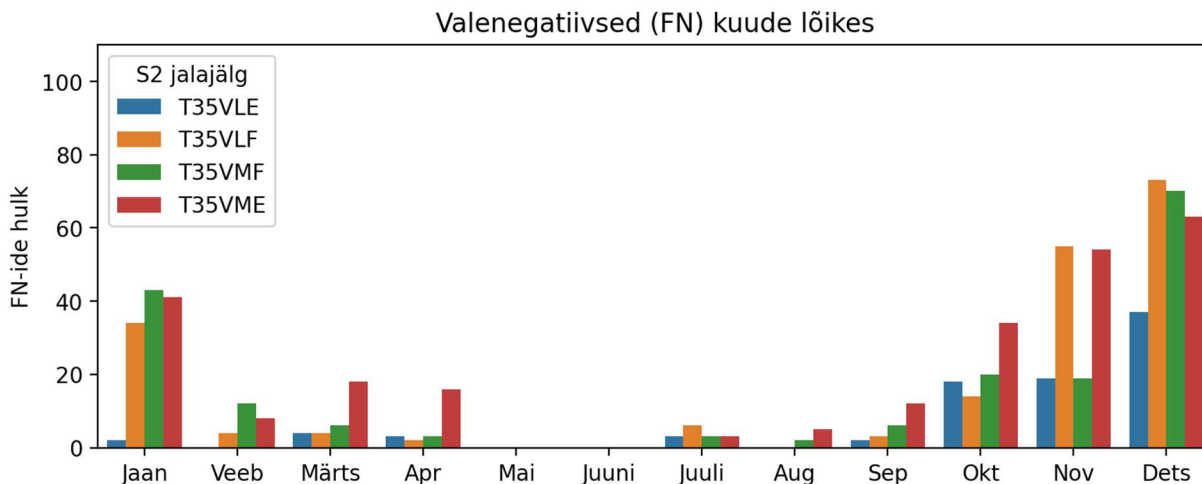
Tabel 3: 2-klassilise mudeli (KAUR kasutusjuht) automaatse kvaliteedikontrolli tulemus RMK metsaeraldistel aastal 2022. Võrdluse alus (referentsandmed RMK-lt) näitab eraldiste hulka, mis sellesse klassi tegelikult oleksid pidanud kuuluma (mis on mudeli väljundi seisukohast TP + FN).

	Täpsus	Saagis	F-skoor	Võrdluse alus
Roheline klass	1.00	1.00	1.00	358133
Punane klass	0.88	0.85	0.86	4938
Õigsus (ingl. accuracy)			1.00	363071
Kaalumata keskmine	0.94	0.92	0.93	363071
Kaalatud keskmine	1.00	1.00	1.00	363071

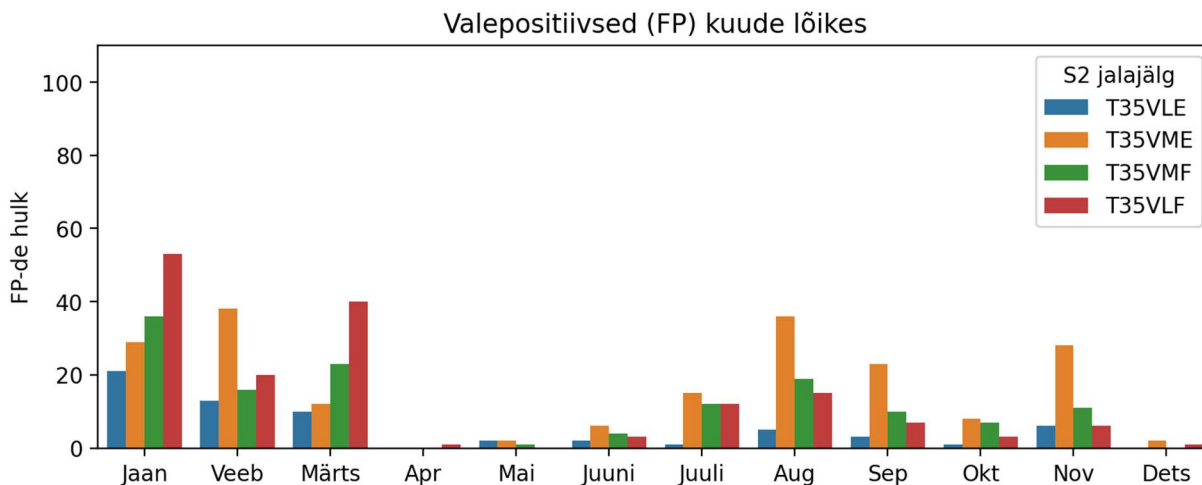
Joonis 2 näitab RMK lageraiete (tegelikud sündmused) jaotumist mudeli väljundklasside vahel aastal 2022. Joonistel 3 ja 4 on näidatud ajas ja ruumis – millistel kuudel ja millistes S2 jalajälgedes – mudeli vead (FP ja FN) tekivad. FN-ide ja FP-de suurem arv jaanuaris ja detsembris on osalt seletatav sellega, et pilvistel perioodidel võib mudel lageraie tuvastada kuni kuu aega selle tegelikust toimumisest hiljem. Seda tulenevalt tõenäosusekõvera väärtuse ja kuju sõltumisest S2 kättesaadavusest koosmõjus sündmuste klassidesse jaotamise algoritmi omapäraga.



Joonis 2. Tegelikke lageraiete (RMK järgi) jagunemine mudeli väljastatud klasside vahel protsendina 3- ja 2-klassilise mudeli puhul aastal 2022.



Joonis 3. 2-klassilise mudeli valenegatiivsete (False Negative, FN) otsuste jaotus ajas (kuude lõikes) ja ruumis (S2 jalajälgede lõikes, vaata Joonis 1).



Joonis 4. 2-klassilise mudeli valepositiivsete (False Positive, FP) otsuste jaotus ajas (kuude lõikes) ja ruumis (S2 jalajälgede lõikes, vaata Joonis 1).

Visuaalne kvaliteedikontroll andis automaatkontrollile sarnased kvaliteedinäitajad. Riigi metsaeraldistel (300 juhuslikult valitud eraldist) oli KeA stsenaariumi puhul kaalumata keskmine **punase klassi täpsus 0.95** ja **F-skoor 0.98**. KAURi stsenaariumi puhul on **F-skoor 0.89**. Kui arvestada, et 1/3 FP-dest pärines eraldistelt, mis olid lagedaks raiutud juba varem, ja selliseid eraldisi mudeli sisendina metsateatiste kasutamisel ei esineks, siis võiks mudel saavutada simuleerides metsateatiste kontrollimisele vastavat juhtu **F-skoori 0.92**.

Eravaldustes metsaeraldistel (300 juhuslikult valitud eraldist) oli KeA stsenaariumi **punase klassi täpsus 0.88** ja **F-skoor 0.93**; KAURi stsenaariumi puhul oli **F-skoor 0.83**. Metsateatiste kasutamisel võiks mudel saavutada **punase klassi täpsuse 0.96** ja **F-skoori 0.98** KeA stsenaariumi puhul ja KUARI stsenaariumi puhul **F-skoori 0.88**. Eramaal on võrreldes riigimaaga suurem FP-de arv. Põhjus on meile teadmata; võibolla on erametsad kuidagi erinevad riigimetsast. Visuaalse kontrolli tolmused on leitavad tabelitest QA_state_22.xlsx ja QA_private_22.xlsx.

6.2. Aasta 2023

Tabelites 4 ja 5 on kokku võetud mudeli väljundi automaatse kvaliteedikontrolli tulemused 2023. aastal (välja arvatud okt–dets) KeA ja KAUR stsenaariumi jaoks. Tulemused tarnisime ilma aasta lõputa, sest aegread lõppesid 2023. aasta 31 detsembriga, aasta lõpu pilvistest oludes ei jõudnud sündmusejärgne signatuur välja kujuneda ja tekkis palju FP-sid. KeA stsenaariumi puhul, milles on vajalik eelkõige maksimaalselt kõrge tuvastamise täpsus (ja kollane klass on, kui ebakindel, kõrvale heidetud) on **punase klassi täpsus 0.96** ja saagis 0.87, st 96% lageraiana klassifitseeritust ka tõepoolest oli lageraie ja 13% tegelikult toimunud lageraieid jäi tuvastamata, kuuludes rohelisse klassi. Mudel leiab sadadest tuhandetest metsaeraldistest hästi üles need, kus toimus lageraie, mida kinnitab kaalumata keskmine **F-skoor 0.98** (Tabel 4). KAURi stsenaariumi puhul, kus oluline on mudeli tasakaalustatus kahe klassi vahel, on **punase klassi täpsus** mõnevõrra madalam (**0.88**), aga klasside ülene kaalumata keskmine **F-skoor** kõrge **0.93** (Tabel 5).

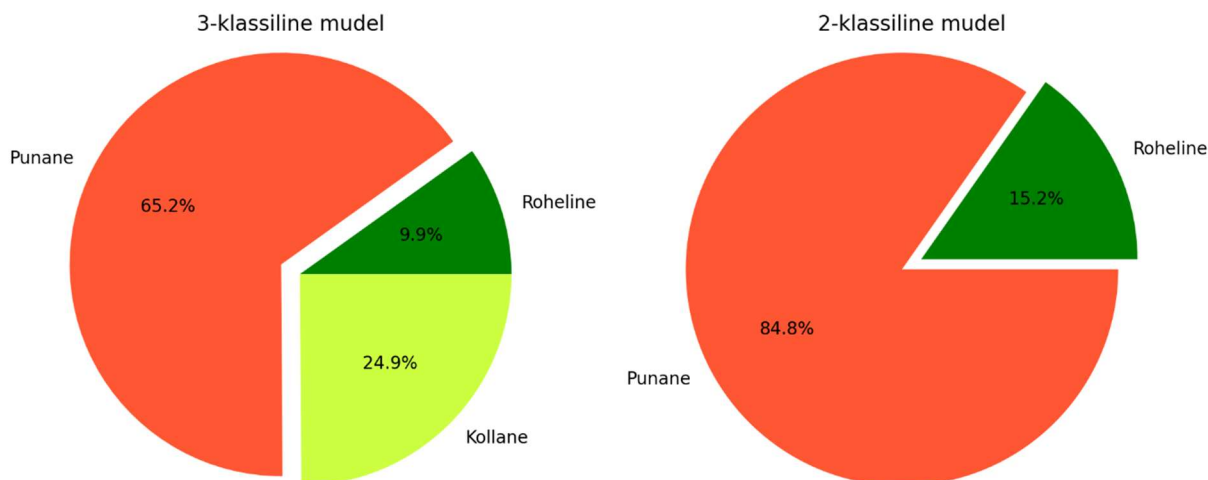
Tabel 4: 3-klassilise mudeli (KeA kasutajajuht) automaatse kvaliteedikontrolli tulemus RMK metsaeraldistel aastal 2023 (okt–dets välja arvatud), kui kollast, st ebakindlat klassi mitte arvesse võtta. Võrdluse alus (referents andmed RMK-lt) näitab eraldiste hulka, mis sellesse klassi tegelikult oleksid pidanud kuuluma (mis on mudeli väljundi seisukohast TP + FN).

	Täpsus	Saagis	F-skoor	Võrdluse alus
Roheline klass	1.00	1.00	1.00	355370
Punane klass	0.96	0.87	0.92	2485
Õigsus (ingl. accuracy)			1.00	357855
Kaalumata keskmine	0.98	0.94	0.96	357855
Kaalatud keskmine	1.00	1.00	1.00	357855

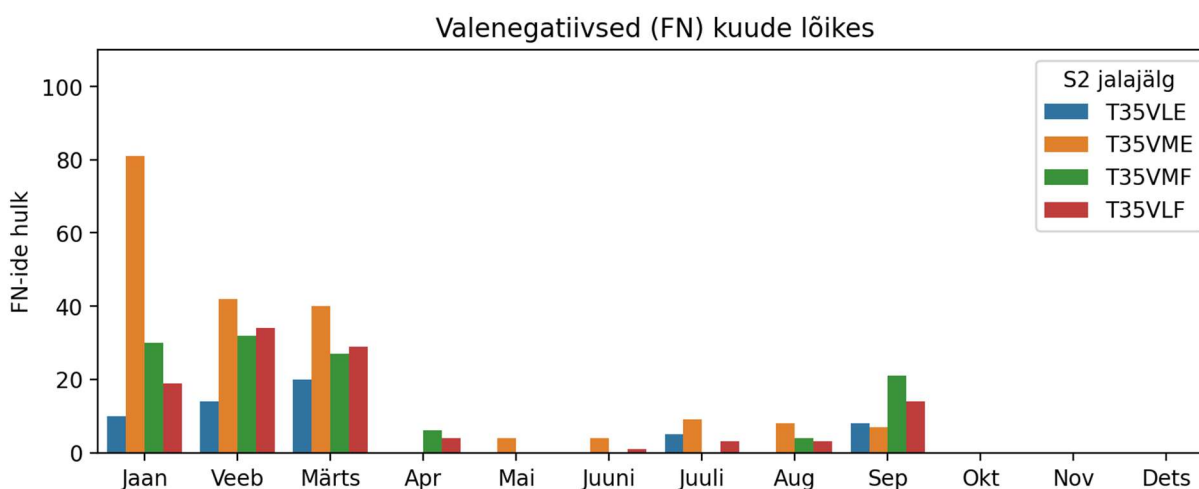
Tabel 5: 2-klassilise mudeli (KAUR kasutusjuht) automaatse kvaliteedikontrolli tulemus RMK metsaeraldistel aastal 2023 (okt–dets välja arvatud). Võrdluse alus (referentsandmed RMK-lt) näitab eraldiste hulka, mis sellesse klassi tegelikult oleksid pidanud kuuluma (mis on mudeli väljundi seisukohast TP + FN).

	Täpsus	Saagis	F-skoor	Võrdluse alus
Roheline klass	1.00	1.00	1.00	356066
Punane klass	0.88	0.86	0.87	3440
Õigsus (ingl. accuracy)			1.00	359506
Kaalumata keskmine	0.94	0.93	0.93	359506
Kaalatud keskmine	1.00	1.00	1.00	359506

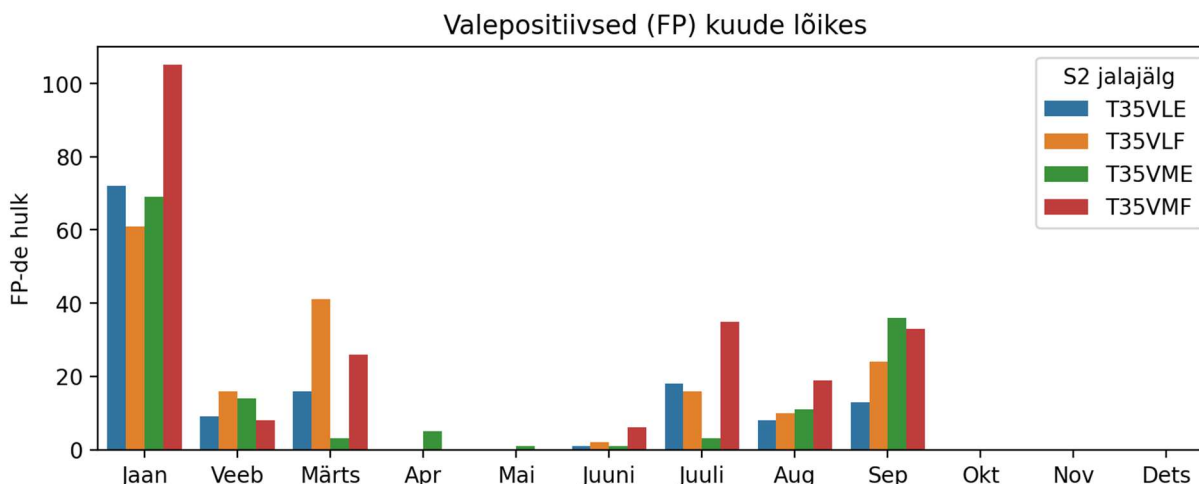
Joonis 5 näitab RMK lageraie (tegelikud sündmused) jaotumist mudeli väljundklasside vahel aastal 2023. Joonistel 6 ja 7 on näidatud ajas ja ruumis – millistel kuudel ja millistes S2 jalajälgedes – mudeli vead (FP ja FN) tekivad. Aasta 2023 automaatse kvaliteedi kontrolli tulemused on sarnased 2022. aasta tulemustele, siinjuured peab aga meeles pidama, et 2023 tulemusid on tarnitud ilma aasta lõpuosata (oktoober kuni detsember). Visuaalne kontroll kinnitas automaatse kontrolli tulemusi. Eraomanduses metsaeraldistel oli täpsus suurem kui 2022. aastal: KeA stsenaariumi **punase klassi täpsus 0.96** ja **F-skoor 0.98**, KAURi stsenaariumi **F-skoor 0.88**, metsaeraldiste kasutusjuhu puhul oleks **F-skoor 0.91**). Kõrgemat täpsust saab põhjendada sellega, et puudu on aasta 3 viimast kuud, kust oleks tulnud juured valepositiivseid ja ka valenegatiivseid. Visuaalse kontrolli tulemused on leitavad tabelitest QA_state_23.xlsx ja QA_private_23.xlsx.



Joonis 5. Tõeste positiivsete (TP) lageraiesündmuste leidumine protsendina kõigist RMK poolt raporteeritud lageraiesündmustest aastal 2022 ja 2023 klassi kaupa 3- ja 2-klassilise mudeli puhul.



Joonis 6. 2-klassilise mudeli puhul valenegatiivsete (False Negative, FN) jaotus ajas (kuude lõikes) ja ruumis (S2 jalajälgede lõikes, vaata Joonis 1).



Joonis 7. 2-klassilise mudeli puhul valeneposiitivsete (False Positive, FP) jaotus ajas (kuude lõikes) ja ruumis (S2 jalajälgede lõikes, vaata Joonis 1).

7. Vea-allikad ja mudeli edasiarendamine

Loodud mudel suudab katta kõik aastaajad, st mudel on universaalne. Mudel käitub era- ja riigimetsas suhteliselt sarnaselt. Mudeli arendamise käigus ja tuvastasime, et parimaid tulemusi andis optilise S2 ja SAR S1 kombineerimine. S2 oli siinjuures olulisem sisend.

S2 osas põhjustasid mudelile raskusi pikad pilvised perioodid talvel (ja ka erakordselt pilvine suvi 2023), varjud ja pilvemaski vead (pilved ja pilvevarjud, mis said valesti maskitud). Pilvemaski vigadega seonduvaid isoleeritud tõenäosuskõvera hüppeid sai järeltötluse käigus kompenseeritud sellega, et tõenäosuskõvera väärtus pidi olema üle klassidesse jaotamise lävendi vähemalt 5 päeva.

Enamasti ei piisa usaldusväärsete tulemsute saamiseks vaid ühe S2 pildi olemasolust. Selles saab veenduda vaadates meie visuaalse kontrolli materjalidega kaasa pandud S2 RGB pilte.

Visuaalsete kontrollide järgi umbes 1/3 valepositiivseid, pärines metsaeraldistelt, kus raieküpset metsa ei esinenud (st eraldised olid juba varem lagedaks raiutud). Operatiivse rakenduse puhul, kus raiumisi otsitakse vaid metsateatiste hulgast selliseid valepositiivseid ei esineks ja mudeli täpsus paraneks kuni 10%.

Meie hinnangul saab käesoleva piloodi arendada kord kuus toimivaks teenuseks, kusjuures peab arvestama, et pikkade pilviste perioodide jooksul tuleb oodata enne tulemuste saamist perioodi lõppu. Tagamaks mudeli poolt antud hinnangu kvaliteet, peaks laskma aegrea signatuuril pärast sündmust välja kujuneda (sündmuse signatuur koosneb sündmuse eelsest, sündmusele vastavast ja sündmusejärgsest mustrist). Seetõttu hindame, et operatiivse teenuse puhul ei peaks pilvistel talvekuudel tarnima värskemaid kui 1–3 kuu vanuseid lageraieid, sest muidu kasvab valepositiivsete hulk.

Mudel tuvastas enamasti ka poolikud lageraied. Mudel tuvastas ka sanitaarreied, mis sarnanesid lageraieetele. Siit tulenevalt on olemas potentsiaal tuvastada ka intensiivsemaid muud tüüpi raieid.

Järgnevalt on loetletud võimalused, kuidas mudelit saaks parendada:

- RMK andmete mudeli treenimiseks ja automaatseks tulemuste kvaliteedikontrolliks kasutamine nõudis teadmisi, kuidas võrdlusandmeid kõige õigem oleks kasutada. Tihe suhtlus RMK-ga oli peamiseks sellekohaseks abiks. Vaatamata selle selgus töö käigus, et mõningad asjad oleks pidanud tegema algusest peale teisiti. Näiteks oleks RKM 2022. aasta sündmused pidanud kokku viima RMK 2023 aasta shp failiga, mitte 2022 shp failiga, nagu meie tegime. 811 RMK poolt aastal 2015–2022 raporteeritud raietest puudus 2022 aasta shp failis ruumikujuline vaste, samas kui 2023 shp failis puudus ruumikujuline vaste 14 raietele. Mudeli kohaldamisel operatiivseks rakenduseks peab maapealsete võrdlusandmete kasutamise käesoleva piloodi õppetundide valgelt täpsemalt läbi mõtlema. Mida täpsem on treeningandmestik, seda täpsem tuleb mudel.
- Võrdlus pilvevabadest päevades pärit S2 optiliste satelliidiandmetega näitas, et maapealsed kasutada olnud võrdlusandmed (RMK-lt saadud raieandmed ja mitte raiutud alad) ei ole alati ajaliselt ja ruumiliselt täpsed. Samuti muutuvad metsaeraldiste kujud ajas, mis eksitab mudelit ja raskendab kvaliteedikontrolli. Püüdsime sellest tulenevaid vigu mudelis minimeerida andmete masinõppe jaoks ettevalmistamise etapis, puhastades andmed dublikaatidest ja vigastest kirjetest. Sellest hoolimata, meie juhuvalimiga visuaalsused kontrollid mudeli arendamise käigus tuvastasid mitmeid lageraieid, mida ei esinenud RMK poolt edastatud raie-andmetes (vaata ka tarnega kaasa pandud kvaliteedikontrolli xlsx faile). Sellised ebatäpsused (tekkinud selle käigus, kui RMK võrdlusandmeid puhastas ja üle andis või mujal) jäid mudeli arendamise käigus treeningandmetesse sisse ja mõjutasid mudelit. **Mudeli operatiivseks rakenduseks arendamise etapis peab rõhku panema treeningandmekogu puhastamisele läbi visuaalse kontrolli.**
- Visuaalne kvaliteedikontroll näitas, et mudeli täpsus 2022. aastal erametsas oli riigimetsaga võrreldes mõnevõrra madalam. Mudeli täpsust aitaks tõsta, kui mudeli treenimiseks oleksid kasutada ka andmed erametsadest, millele Eesti Erametsaliidu kaudu võiks olla juurdepääs.
- S1C orbiidile saatmine on plaanitud 2024. aasta 15 novembrisse. Seejärel muutub S1 aegrida tihedamaks ja on võimalik kasutada 6-päevast koherentsust (praeguse 12-päeva asemel), millel on oluline positiivne mõju mudeli täpsusele. Koherentsus ei ole küll enamasti lume ja jääga (st talvel) kasutatav. Integreerimistööd saaks teha piloodi operatiivseks rakenduseks arendamise käigus.
- NISAR (õnnelikult) orbiidile jõudmine (edasi lükkunud 2024. aasta teise poolde) võimaldab kasutusele võtta S-riba (12 cm lainepikkus) ja L-riba (24 cm) Sentinel-1 C-riba (3 cm) kõrval. See võimaldab metsa ja lageraielangi oluliselt paremat eristamist võrreldes S1-ga, kasutades vaid tagasihajumise intensiivsust. Nii väheneks sõltumine S2-st ja seega pilvkattest ning võimaldaks potentsiaalselt eristada ka alasid alla 0.5 ha, sest koherentsuse olulisus mudeli sisendina väheneks tagasihajumise kasuks. NISAR andmeid võiks saada kasutada umbes aasta pärast, kui sensor on saanud koguda mudeli treenimiseks vajaliku koguse andmeid. **NISARi lisandumisel on potentsiaal tarneaegu tihendada ja tagada sõltumatus pilvkattest.**
- Maa-amet on läbirääkimistes Planet Labs-iga ostmaks Eesti ala kohta Planeti 1-päevase sammuga 3.5 m resolutsiooniga optilisi satelliidipilte. Planet optiliste piltide kasutusele võtmine võimaldaks potentsiaalselt rohkem pilvevabu ülesvõtteid. Samuti võiks olla kasu Planeti teravamast resolutsioonist (sh lageraie ala piiritlemisel, kui sellise rakenduse poole

on soov liikuda). Planet andmete miinuseks on vaid 4 spektraalkanalit: sinine, roheline, punane ja lähis-infrapunane (NIR).

- Talvekuude operatiivsemaks katmiseks või lihtsalt mudeli täpsuse tõstmiseks on mõeldav ka mudeli tulemustest tuntav osa visuaalsesse kontrolli suunata – KappaZeta võimekus on 1000 eraldist päevas.

Väiksemate kui 0.5 ha alade jälgimine oleks võimalik, kui:

- Koguda piirnevad ja ühel ajal võetud metsateatiseid kokku lankideks, eeldades, et lageraie toimub tervel sellisel langil üheaegselt, ja jooksutada käesoleva piloodi käigus loodud mudelit neil lankidel.
- Arendada paralleelset ainult optilistel satelliitidel (S2 ja ka Landsat) põhinev mudel 0.2–0.5 ha alade jaoks. Ajaline lahutus ja tuvastamise täpsus oleks kehvem, kui käesoleva piloodi näitajad. Multispektraalse Landsati andmete lisamine aitaks optilisi aegridu tihendada, aga ei aita oluliselt pikkade pilviste perioodide puhul (nagu 2023 aasta suvi või pilvised talvekuud).
- Maa-ameti Planet andmete ostu järel (kui see saab teoks), kasutada optilises mudelis S2 kõrval või asemel Planet andmeid, mis 1-päevase sammu ja 3.5 m ruumilise lahutusega võiksid lühendada ka tarneaegu ja suurendada tuvastamise täpsust. Planet andmetel on tänud paremale ruumilisele lahutusele potentsiaal tuvastada ka <0.2 ha alasid. Planetil on vaid 4 spektraalkanalit, mis võib takistada sensori ajalis-ruumilise potentsiaali rakendamist.
- NISARi orbiidile jõudmise järel treenida, mudel, mis ei kasuta koherentsuse väärtusi, vaid ainult tagasihajumist ja optilisi kanaleid.