



VILNIAUS UNIVERSITETAS
CHEMIJOS IR GEOMOKSLŲ FAKULTETAS
GEOMOSKLŲ INSTITUTAS
KARTOGRAFIJOS IR GEOINFORMATIKOS KATEDRA

Vaiva Venckauskaitė

Kartografijos magistro baigiamasis darbas

**MEDŽIŲ INVENTORIZAVIMO URBANIZUOTOSE
TERITORIJOSE GALIMYBĖS TAIKANT NUOTOLINIUS
METODUS**

Darbo vadovas
Doc., Dr. Linas Bevainis

Leidžiama ginti _____
(parašas)

Darbo įteikimo data 2024-05-26

Registracijos Nr. _____

Vilnius, 2024



VILNIUS UNIVERSITY
FACULTY OF CHEMISTRY AND GEOSCIENCES
GEOSCIENCES INSTITUTE
DEPARTMENT OF CARTOGRAPHY AND GEOINFORMATICS

Vaiva Venckauskaitė

Master's thesis of Cartography study program

**ASSESSMENT OF TREE INVENTORY POSSIBILITIES IN
URBAN AREAS USING REMOTE SENSING**

Work supervisor
Doc., Dr. Linas Bevainis

Allowed to defend _____
(signature)

Date of work submission 2024-05-26

Registration number _____

Vilnius, 2024

TURINYS

Anotacija	4
ĮVADAS.....	6
1. LITERATŪROS APŽVALGA.....	8
1.1 Medžių segmentavimo tyrimai.....	8
1.2 Medžių charakteristikų ir struktūros įtaka segmentavimui	10
1.3 Spektrinių duomenų naudojimas medžių rūšies nustatymui.....	10
2. METODIKA	14
2.1 Tiriama teritorija	15
2.2 Naudojami duomenys ir R moduliai	16
2.3 Medžių aukščio modelio (CHM) generavimas	17
2.3 Medžių segmentavimas R aplinkoje	19
2.4 Spektrinių ir vektorinių duomenų apdorojimas bei paruošimas mašiniam mokymui	20
3. REZULTATAI.....	24
3.1 Sudarytų CHM ir R aplinkoje nustatytų medžių viršūnių rezultatai.....	24
3.2 CHM ir taškų debesies metodų medžių segmentavimui palyginimas	28
3.3 Medžių rūšių identifikavimo rezultatai	32
DISKUSIJA	35
IŠVADOS	36
NAUDOTA LITERATŪRA.....	37
SANTRAUKA	41
SUMMARY	42

Venckauskaitė V. Medžių inventorizavimo urbanizuotose teritorijose galimybės taikant nuotolinius metodus. Magistro darbas. Vilnius: VU. 2024.

Anotacija. Želdynų ir želdinių inventorizavimo reikšmingumas ne visiems suprantamas, tačiau tai yra įstatymu įtvirtinta veikla, kurią įgyvendinti tampa sudėtinga. Trūkstant kvalifikuotos darbo jėgos, augant želdinių apimtims bei siekiant kuo didesnio efektyvumo, nagrinėjami automatinio inventorizavimo nuotoliniais būdais klausimai, kurie pasitelkiami šiame tyrime. Naudojantis skaitmeninio lazerinio skenavimo duomenimis bei hiperspektriniais vaizdais siekiama nustatyti medžių fizinius parametrus, susidedančius iš medžio viršūnės lokacijos, jo vainiko apimamo ploto, medžio aukščio ir rūšies. Pirmiems trimis parametrams naudojami skirtingo detalumo LiDAR duomenys. Atliekama medžių aukščio modelio ir tiesioginio taškų debesies segmentavimo palyginamoji analizė medžių parametrams išskirti, naudojantis R programavimo kalbos bibliotekomis. Taip pat palyginamos skirtingos medžių aukščio modelių sudarymo technologijos, LiDAR duomenų tankumo bei sudaromo rastro raiškos įtaka medžių viršūnėms ir vainikams išgauti. Rūšies nustatymui pasitelkiamas 64-ių kanalų hiperspektrinis vaizdas. Klasifikacija atliekama atraminių vektorių ir atsitiktinių miškų modeliais objektų-pagrindu penkioms dažniausiai pasitaikančioms medžių rūšims.

Tekstas: 40 psl., 20 pav., 5 lentelės. Santrauka lietuvių ir anglų kalbomis.

Raktiniai žodžiai: LiDAR, hiperspektras, SVM, medžiai, ForestTools, lidR, rūšis, vainikai.

TERMINŲ ŽODYNAS

C – scikit-learn Python bibliotekos SVC modelio reguliarizacijos parametras, valdantis kompromisą tarp ribos maksimizavimo ir mokymo klaidų minimizavimo (Scikit learn, 2024a).

Coef0 – scikit-learn Python bibliotekos SVC modelio parametras, kontroliuojantis kokią įtaką modeliui daro aukšto laipsnio polinomialai ir žemo laipsnio polinomialai. Aktualus tik polinomialiam ir sigmoidiniam modelio tipui (Scikit learn, 2024a).

DSM (angl. *Digital Surface Model*) – iš LiDAR duomenų sudarytas skaitmeninio paviršiaus modelio produktas, rodantis aukščiausių paviršių aukštį duotame taške (Wasser, 2024).

DTM (angl. *Digital Terrain Model*) – iš LiDAR duomenų sudarytas skaitmeninio reljefo modelio produktas, rodantis žemės paviršiaus aukštį (Wasser, 2024).

F1-score – metrika, mašiniame mokyme kombinuojanti tikslumo ir atkūrimo (angl. *recall*) rodiklius, kaip harmoninį vidurkį. Dažnai naudojama kaip vertinimo metrika dvinariniame ir daugiaklasiame klasifikavime (Wood, 2024).

LiDAR (angl. *Light Detection and Ranging*) – nuotolinio stebėjimo metodas, kurį taikant lazerio pulsais matuojami atstumai iki Žemės. Šie šviesos impulsai kartu su kitais duomenimis, kuriuos užregistruoja sistema iš oro, leidžia gauti tikslią trimatę informaciją apie Žemės formą ir jos paviršiaus savybes (National Oceanic and Atmospheric Administration, 2023).

Maksimalus gylis (angl. *max. depth*) – scikit-learn Python bibliotekos Random Forest (liet. atsitiktinių miškų) modelio parametras, nurodantis kiekvienam sprendimų medžiui leidžiamų padalijimų skaičių (Scikit learn, 2024b).

Multispektriniai ir hiperspektriniai vaizdai – daugiakanaliai vaizdai, kurių kiekvienas kanalas fiksuoja skirtingą elektromagnetinės spinduliuotės dalį. Griežto nustatyto kanalų kiekio tarp multispektrinių ir hiperspektrinių vaizdų nėra (eoPortal, 2023).

PCA – pagrindinių komponentų analizė (angl. *principal component analysis*), skirta didelių duomenų apimties (dimensiškumo) supaprastinimui, išlaikant svarbius dėsningumus ir tendencijas (Jaadi, 2024).

R aplinka – R Studio darbalaukio programa, kurioje valdomos R programavimo kalbos bibliotekos, kuriamos vizualizacijos bei rašomas ir vykdomas kodas.

Recall – metrika, kuri parodo, kaip dažnai mašininio mokymosi modelis teisingai nustato teigiamus atvejus (tikruosius teigiamus atvejus) iš visų faktinių teigiamų pavyzdžių duomenų rinkinyje (Evidently AI, 2024).

RGB – skaitmeninė trijų kanalų spalvų sistema, sudaryta iš: raudonos (angl. *red*), žalios (angl. *green*) ir mėlynos (angl. *blue*), jas maišant išgaunama bet kokia kita spalva (Spaudos departamentas, 2024).

IVADAS

Lietuvos Respublikos želdynų įstatymas galioja nuo 2008 m., įpareigojantis savivaldybių vykdomąsias ar atstovaujamasias institucijas juos saugoti, valdyti tvarkymą ir kūrimą. Jame yra įtrauktas želdynų ir želdinių inventorizavimas, kuris yra privalomas ir skirtas rinkti erdvinis duomenis apie įvairius želdynų parametrus, kad būtų galima surinktus duomenis teikti Topografijos ir inžinerinės infrastruktūros informacinėje sistemoje (TIIS), stebėti jų būklę ar remdamasi šia informacija vykdyti kitus darbus. Įstatymas nurodo, jog inventorizacijos duomenys atitinkamai leistų efektyviau valdyti, teikti informaciją ir kontroliuoti želdynus ar želdinius, aplinkos būklę, vertinti galiojančių normų atitikimą. Šiame tyrime dėmesys sutelkiamas į medžių, priskiriamų želdiniams, inventorizavimą. Medžiai teikia naudą žmonių sveikatai valydamis orą nuo kenksmingų medžiagų ar sulaikant dulkes, kaitriomis dienomis mažina saulės radiacijos poveikį žmogui ir prisideda prie karščio salos mažinimo mieste, praturtina architektūrinę kompoziciją, įtvirtina ekologinį stabilumą, tad jų privalumai pasiskirsto plačiame spektre: estetinės, apsauginės, triukšmo ir aibe kitų funkcijų (Lietuvos Respublikos Seimas, 2007; Lietuvos arboristikos centras, 2024).

Problema. Su želdynų įstatymu susijusios želdynų ir želdinių inventorizavimo ir apskaitos taisyklės (Lietuvos Respublikos Seimas, 2008), kuriose nurodyta, jog inventorizavimą gali atlikti tam tinkamą aukštąjį išsilavinimą ar sertifikatą ir pakankamą patirtį (bent vieneri metai želdynų priežiūros ir tvarkymo srityje) turintys asmenys. Inventorizaciją galima atlikti ir lauko darbais ir nuotoliniais metodais nustatytu tikslumu. Naudoti žmogiškuosius resursus yra brangu, taip pat darosi sudėtinga pritraukti kvalifikuotos darbo jėgos. Vilniuje inventorizacijos duomenys yra atnaujinami kas penkis metus, nes viešųjų želdinių ir želdynų duomenys negali būti senesni nei penkeri metai (praktiškai tiek ir užsitęsia lauko darbai sostinėje), tačiau 2023 metais likus maždaug 100 tūkst. želdinių, Vilniaus miesto savivaldybei pritrūko kvalifikuotų specialistų darbui atlikti.

Problemos aktualumas. Norint nuotoliniais metodais inventorizuoti medžius, dideliu iššūkiu tampa jų rūšinė, amžiaus, pasiskirstymo įvairovė, kuri egzistuoja mieste, tačiau urbanizuotai teritorijai ji yra naudinga. Gyventojai linkę susijaudinti dėl medžių kirtimo, nors tai daroma ne tik dėl projektų įgyvendinimo, bet ir dėl jų pačių saugumo. Melburno mieste medžių amžius gali tapti problema, nes daugelis jų buvo sodinti vienu metu (įskaitant garsius prospektus ir parkus) bei yra perlipę 100 metų ribą, o tai reiškia, jog sveikatingumas mažės ir teks juos pašalinti vienu metu (City of Melbourne, 2020). Pagal Europos aplinkos agentūros 2021 metų duomenis, Vilnius yra ketvirtas miestas pagal žaliosios infrastruktūros, žaliųjų erdvių ir medžių dangos dalį (Europos aplinkos agentūra, 2022). Taip pat Vilnius buvo tituluotas Europos žaliaja sostine 2025-iesiems metams. Europos komisija pasiūlė gamtos atkūrimo įstatymą, pagal kurį valstybės narės iki 2030 metų užtikrintų, jog grynasis miesto žaliosios erdvės plotas ir miesto medžių lajų danga nesumažėtų lyginant su 2021 m. Taip pat bendras miesto žaliosios erdvės plotas padidėtų 3 ir 5% atitinkamai iki 2040 ir 2050 m. (Eurocities, 2023). Lietuvos hidrometeorologijos tarnyba, sudariusi įvairius klimato kaitos scenarijus, teigia, jog dažniau pasitaikys ekstremalių oro sąlygų, su kuriomis reikės kovoti. Želdynai padeda ne tik kovoti su karščiu, bet ir reguliuoti vandens kiekius, o medžių

lajų danga vertinama kaip svarbus rodiklis. Jeigu būtų priimtas komisijos įstatymas, medžių mieste tik padaugėtų ir inventorizuoti įprastu būdu taptų dar sudėtingiau.

Tyrimo naujumas. Lietuvoje nėra daug nuotolinių tyrimų želdinių inventorizacijai urbanizuotoms teritorijoms, tai dažniau miškų apskaitos tyrimai, kuriems skiriama daug dėmesio. Labai daug tyrimų fokusuojasi ir į žemės ūkio sferą. Tokiai didelei teritorijai kaip Vilnius verta pagalvoti apie nuotolinių metodų taikymą, ypač, kai miesto savivaldybei priklauso tarnybos ar institucijos, turinčios aerofotografines nuotraukas ar LiDAR skenavimo duomenis bei galinčios savarankiškai vykdyti analizę ir kurti produktus. Šiuo metu yra ieškoma geriausių ir tiksliausių sprendimų, kaip būtų galima pakeisti ar bent palengvinti rankinį lauko patikrų darbą. Šiame tyrime bus integruojami hiperspektriniai aerovaizdai ir LiDAR skenavimo duomenys, taikant GIS metodus, programavimo sprendinius.

Rezultatų praktinis pritaikomumas. Tikimasi, jog su dabartinėmis technologijomis būtų galima jei ne pilnai, tai bent iš dalies automatizuoti medžių inventorizavimą. Tam, kad lauko darbams tektų skirti kuo mažiau resursų – kuriama metodika medžių ir jų fizinių parametrų nustatymui. Darbams palengvinti gali būti taikomi atskiri metodikos produktai kitiems želdiniams, pavyzdžiui, krūmų aukščio nustatymui. Nuotoliniai metodai leistų sukaupti duomenų apie daugiau želdinių ar sudaryti kompromisą, nes dalis jų būna nuosavuose sklypuose, tad ne visi medžiai yra inventorizuojami lauko patikromis, kaip ir nuotoliniais metodais gali būti neužfiksuojami jaunesni ir žemesni medžiai takaus apsodinimo atveju.

Padėka. Už įkvėpimą ir paskatą norėčiau padėkoti savo darbovietai – „ID Vilnius“. Atrastos atviro kodo alternatyvos sulaukė susidomėjimo ir buvo įtrauktos į palyginimą su komerciniu sprendimu. Taip pat buvau įtraukta į medžių inventorizacijos veiklas, kurios susidarė iš metodų testavimo, dalyvavimo aktualiose konferencijose ir susitikimuose su mokslo atstovais ar įmonėmis, vykdančiomis panašius projektus.

Tyrimo hipotezė: distancinių tyrimų pagalba galima identifikuoti dominuojančias medžių rūšis bei fizinius parametrus.

Tyrimo objektas: medžių fizinių parametrų nustatymo metodika.

Tikslas: hiperspektrinių vaizdų ir skaitmeninio lazerinio skenavimo duomenų pagrindu sukurti medžių rūšių ir fizinių parametrų vertinimo metodiką.

Uždaviniai:

- Taikant LiDAR duomenis ir R programavimo bibliotekas nustatyti medžių fizinius parametrus;
- Palyginti skirtingais metodais gautus medžių viršūnių ir vainikų rezultatus;
- Paruošti medžių duomenis mašiniam mokymuisi, susiejant hiperspektrinius vaizdus ir išskirtas lajas su inventorizacijos duomenimis;
- Python programavimo kalba sukurti skirtingus modelius medžių rūšių identifikacijai ir juos validuoti.

1. LITERATŪROS APŽVALGA

1.1 Medžių segmentavimo tyrimai

Yra labai daug tyrimų, besiremiančių tik RGB vaizdais, LiDAR ar multispektriniais vaizdais, tačiau naujausios tendencijos miškininkystėje rodo, jog siekiant geresnių rezultatų tobulinami ne tik vieno, bet ir kelių sensorių naudojimas, pavyzdžiui, RGB vaizdų derinimas su LiDAR, hiperspektriniais ar terminiais duomenimis (Chehreh ir kt., 2023). Žinoma, optiniai vaizdai buvo pradėti naudoti anksčiausiai. Atsiradus LiDAR technologijoms padaugėjo tyrimų pavienių medžių išskyrimui ir apskritai naudojant vien šiuos duomenis (Jakubowski, 2013) Pagrindiniai du stambūs medžių segmentavimo metodai naudojant LiDAR duomenis yra sugeneruoto CHM medžių aukščio modelio (angl. *Canopy Height Model*) panaudojimas arba tiesioginis paties taškų debesies segmentavimas.

Jakubowski ir kt. (2013) lygino medžių vainikų išskyrimą naudojant 3D LiDAR taškų debesies (9 pulsai/m²) segmentavimo algoritmą su vektoriniais duomenimis ir naudojant rastrinius CHM duomenis vaizdo analizei objektų pagrindu. CHM yra sudaromas atimant DTM (skaitmeninis reljefo modelis) iš DSM (skaitmeninis paviršiaus modelis) (Kwak, 2007). Abu metodai pasirodė gerai lyginant tarpusavyje ir su in-situ medžių aukščių duomenimis, tačiau rezultatai suprastėjo tankiai apaugusiose teritorijose. Ekosistemose, kur medžių vainikai tankūs ir persidengiantys, atsiranda nemažai klaidų naudojant CHM. Gautos medžių vainikų ribos skyrėsi. Naudojant taškų debesį gauti sudėtingesni, didesni, tačiau artimesni realioms poligonai, bet jų buvo išskirta mažiau (išskirdavo per mažai, o objektų pagrindu išskirdavo per daug). Ma ir kt. (2021) bandė išvengti klaidų (susidarančių dėl CHM naudojimo) sukuriant augmenijos taškų debesies modelį. Jis sukurtas ištiriant erdvinį tankio pasiskirstymą horizontalioje plokštumoje bei panaudojant lokalaus maksimumo (angl. *local maximum*) ir patobulintą vandenskyros transformacijos (angl. *watershed*) algoritmus-filtrus atitinkamai medžių aukščiams ir vainikams gauti. Tyrimu parodyta, jog metodas turi potencialo. CHM panaudojimas, vis dėlto buvo šiek tiek pranašesnis, kalbant apie nustatymo tikslumą, tuo tarpu pasiūlytas modelis turėjo geresnį aptikimo lygį. Optimalaus filtro lango dydis lokalaus maksimumo algoritmui pateikė tiksliausią rezultatą medžių skaičiui ir vietos nustatymui.

Lokalaus maksimumo ar vandenskyros transformacija gana dažnai pasitaikantys metodai. Kwak ir kt. (2007) individualių medžių ribų išskyrimui pritaikė vandenskyros transformaciją euklidiniais atstumais transformuotui vaizdui, kuriame buvo nustatytos medžių viršūnės. Įprastai, kaip ir šiame darbe, medžių viršūnėmis laikoma maksimali vertė segmentuotame medžio vainiko plote – tam skirtas lokalaus maksimumo algoritmas. Yra nustatomos medžių viršūnės, po to vainikų ribos, arba atvirkščiai. Tyrime išryškinta prie kokių skirtingų išplėstinio maksimumo (angl. *extended maxima*) transformacijos formulės parametro reikšmių rezultatai skirtingoms medžių rūšims buvo geresni. Transformacija medžių aukščio modeliui buvo panaudota tam, jog būtų išvengta netikrų medžių viršūnių. Ieškant lokalaus maksimumo Zhang ir kt. (2015) naudojo *tree climbing* algoritmą, tačiau geresnius rezultatus pasiekė su apribotu *tree climbing* algoritmu, kuriame įvedė horizontalią ribinę vertę, kurią buvo galima nustatyti pagal lauko patikrų duomenis.

Qin ir kt. (2014) derino optinius ir LiDAR duomenis didelėms teritorijoms. Miško medynų ribas išskyrė iš labai aukštos raiškos optinių vaizdų ir kiekvieną medį segmentavo iš LiDAR sugeneruoto CHM. Buvo pastebėta, jog daugelis medžių vainikų buvo per daug sudalinti. Jeigu galvojama apie keletą skirtingų duomenų naudojimą, procesas sudėtingėja suderinant skirtingas erdves, spektrines ar laiko raiškas. Taip pat metodai, kurie naudojami taškų debesų formatui ir optinių vaizdų duomenims suderinti, praranda nemažai informacijos (Wan ir kt., 2021). Chen, X. ir kt. (2023) pagerino taškų debesies segmentavimą 8% panaudojant ir aukštos kokybės optinius vaizdus, tačiau teigia, jog metodas netinka didelėms teritorijoms dėl jo sudėtingumo ir užimamo laiko, be to reikalauja labai aukštos kokybės nuotolinių duomenų. Individualių medžių išskyrimo strategijos gali būti padalintos į keturis metodus: pikselių-pagrindo, regionų-pagrindo, kontūrų-pagrindo ir empirinio metodo. Empirinis metodas naudoja daug lauko patikros duomenų, kad išvestų tokius miško parametrus kaip vainikų plotis (medžio aukštis ir vainiko plotis turi tiesinį ryšį regresijos analizėje), medžių amžius bei kitus medžių parametrus. Vandenskyros transformacijos segmentavimas priskiriamas pikselių-pagrindo metodui, kuris tiria medžių geometrinę struktūrą vaizde (Chen ir kt., 2005). Vandenskyros transformacijos segmentavimas susiduria su medžių skaičiaus perversinimo ar nepakankamo įvertinimo dėl didelių aukščio skirtumų jų topografijoje ar mažesnių medžių po aukštesnių medžių vainikais problema, tad labai svarbu tinkamai parinkti funkciją ir jos parametrus (Dougherty ir Lotufo, 2003). Li ir kt. (2012) apibendrina, jog aptiktų medžių kiekis kai kuriuose tyrimuose gali varijuoti nuo vos 40 iki 80 procentų naudojant LiDAR duomenis, tačiau tai buvo gana seni tyrimai. Savajame nusprendė naudoti ne lokalaus maksimumo filtrą, kadangi sunku nustatyti paieškos lango dydį, o globalų maksimumą, jį segmentuojant pagal sudarytas klasifikacijos taisykles. Rastas medis pašalinamas ir taip ieškomas sekantis globalus maksimumas. Segmentavimo rezultatai tiesiogiai daro įtaką rūšių identifikavimo tikslumui vėlesniuose procesuose. Tinkama segmentacija gali sumažinti spektrinę rūšių variaciją, o prasta – daryti įtaką su lajos dydžiu susijusiems kintamiesiems (Wang ir kt., 2018; Dian ir kt., 2017).

Taškų debesies segmentavimo metodai plečiasi nuo paprasto klasterizavimo iki sudėtingų mašininio mokymosi metodų, naudojančių ir prižiūrimą ir neprižiūrimą būdus. Tai gali būti daug laiko atimantis procesas turint didelį taškų debesies tūrį. Kadangi prižiūrimam metodui reikalingi sužymėti duomenys, o tam reikalingi dideli laiko ir rankinio darbo resursai, yra tyrimų, kurie sudaro sintetinius duomenis juos simuliuojant – taip greičiau gaunant reikalingą duomenų kiekį algoritmų apmokymui (Bryson ir kt., 2023). *Watershed* kaip ir *local maxima* algoritmai priskiriami neprižiūrimam mašininiam mokymuisi. Keletas kitų būtų augančių regionų ir kontūrų aptikimas. Pastarasis dažnai naudojamas pirminiam apdorojimui prieš sukuriant regionus ar objektus. Vaizduose identifikuoja objektų ar elementų ribas. Augantys regionai naudoja pradinius taškus, kurie paskleisti po tiriamą teritoriją, ir pagal juos plečia regioną, jei šalimais esantys pikseliai atitinka kriterijus kaip, pavyzdžiui, spalva, tekstūra ar jų kombinacijos. Vandenskyros transformacija, vis dėlto viena dažniausiai pasitaikančių metodų miškininkystėje (Chehreh ir kt., 2023).

1.2 Medžių charakteristikų ir struktūros įtaka segmentavimui

Individualių medžių išskyrimo tikslumas naudojant LiDAR duomenis gana žemas heterogeniniuose miškuose. Qiao ir kt. (2023) savo tyrime nustatė, jog vidutinė kvadratinė paklaida (RMSE) didėja tankėjant miškui bei buvo mažiausia spygliuočių miške. Taip pat spygliuočių miškas turėjo gerai atskiriamus vainikus, palyginus su plačialapiais ir maišytais miškais, kurie, žinoma, nemažai persidengia dėl didesnių ir įvairesnių formų vainikų. Atitinkamai medžių lokacijos RMSE buvo didesnė. Apibendrinant medžių aukščius geriausią rezultatą pasiekė spygliuočių rūšis. Medžių vainikų forma turi nemažą įtaką viršūnių nustatymui, nuo ko gali priklausyti ir medžių ribų išskyrimas. Kwak ir kt. (2007) taip pat pastebėjo, jog geriau segmentuoti spygliuočiai. Jų forma gali būti apibendrinta kūgiu, turinčiu aukštą nuolydį, todėl aukščio skirtumai labai aiškūs keliaujant nuo viršūnės iki kraštų. LiDAR apskritai gali pateikti informaciją apie skirtingų medžių rūšių geometrinę struktūrą. Wan ir kt. (2021) sudarė kiekvieno apibendrinto medyno profilio kreivę, sudarytą iš vertikalios LiDAR duomenų dažnio pasiskirstymo, kuri leido išskirti struktūrinės medynų savybes. Šios kreivės labai priklauso nuo taškų debesies tankio. Mišriuose miškuose reikalingas tankus taškų debesis, kad būtų galima išskirti skirtingas rūšių struktūras kreive. LiDAR galima užfiksuoti skirtingų medžių lapijos pasiskirstymą ir lajos išsišakojimo struktūrą, tad skenavimo duomenys ne tik naudojami medžių vainikų riboms išskirti, bet ir gali pasitarnauti tolimesnėje rūšies identifikavimo problematikoje papildomais duomenimis (Michałowska ir Rapiński, 2021; Wang ir kt., 2018) Iš lazerinio skenavimo duomenų galima išskirti tokias ypatybes kaip vainiko poringumą, kuris galėtų pagerinti klasifikacijos kokybę medžiams su didesniais, tačiau retesniais vainikais. Be poringumo apskaičiuotinas ir lajos tankio profilis, kuris gali būti pateiktas pikselių vertėse (Dian ir kt., 2016), tačiau yra tyrimų, kurie išgauna ir vertikalų medžio profilį, pasinaudodami tankiu.

Medžių vainikai gali būti labai sudėtingi. Nustatant vainikus dažnai daroma prielaida, jog jie yra artimi apskritimui, bet miesto medžiai gali turėti gana sudėtingas geometrines formas. Jų vainikų dydis ne visada vertinamas miesto želdiniuose, tačiau gali būti naudingas parametras modeliuojant ekosistemas. Vainikai įprastai būna arba pamatuojami, arba išgaunami iš kitų parametrų. Anizotropijos įtraukimas į algoritmus, tikėtina, pagerintų rezultatus (Zhang ir kt., 2015).

Tyrimuose analizuojant skirtumus tarp aptiktų medžių viršūnių ir inventorizuotų medžių taškų aiškinama, jog nuokrypių gali atsirasti dėl keleto faktorių. Vienas jų – tai netinkamas prietaiso naudojimas lauko darbuose, pavyzdžiui vietose, kur medžiai vieni kitus perdengia tankiai apaugusiose teritorijose. Naudojant lazerinį atstumo matuoklį darbuotojas gali praleisti viršūnę ir pamatuoti šalimais esančių medžių aukštį. Skenuojant paviršių nuotoliniais metodais, LiDAR pulsas gali nepataikyti į medžio viršūnę, jeigu tarpai taškų debesyje yra dideli (Zhang ir kt., 2015).

1.3 Spektrinių duomenų naudojimas medžių rūšies nustatymui

Dirbant su spektrine informacija egzistuoja ne tik erdvinė ar laiko raiška, bet ir spektro detalumas. Turint mažai kanalų ar ieškant naudingų savybių kuriami ir dirbtiniai kanalai, atsižvelgiant į objektų sąveiką su spinduliuote. Vykiant klasifikaciją prie originalių kanalų gali

būti pridėti įvairūs išskaičiuoti indeksai, dažnai susidedantys iš kanalų santykio, taip pat naudojamos paprastesnių matematinių išraiškų reikšmės kaip kanalų vidurkiai. Hiperspektrinio duomenys turi didelį dimensijų kiekį, tokiais atvejais gali būti naudojamas nestabilumo indeksas ar dažnai pasitaikanti pagrindinių komponenčių analizė (PCA). Jensen ir kt., (2011) klasifikuodami medžių rūšis naudojant hiperspektrinius vaizdus pasitelkė šešias pagrindines komponentes. Komponentių kiekį galima pasirinkti pamatuotai arba visiškai atsitiktinai. Rezultatai pasiekė 82%, bet pagerėjo iki 91,4% prie naudotų komponenčių pridėjus augalijos indeksus, kanalų vidurkius ir kitas kanalų santykinės reikšmės. Visi kanalai nebūtinai neša vienodą vertę, o apjungiant skirtingus duomenis dimensijų kiekis gali dar labiau padidėti, jei nėra vykdoma reikšmingų savybių atranka. Mozgeris ir kt. (2018) lapuočių rūšims išskirti naudojo visus hiperspektrinio vaizdo kanalus (16 kanalų), kad paskaičiuotų zonavą statistiką, t. y. įvairius minimumo, maksimumo, vidurkio ir panašius parametrus kiekvienam vektorizuotui medžio vainikui. Vėliau vykdė savybių atranką taip pat pasinaudojant PCA ir papildomais algoritmais. Ghosh ir kt. (2014) atkreipia dėmesį, jog kombinuojant spektrinę ir struktūrinę informaciją (išgaunamą iš LiDAR duomenų) pateikia geresnius rezultatus, nors ir skirtingų duomenų naudojimas reikalauja tam tikro agregavimo ar prisideda prie erdvinio nesutapimo.

Nors rūšinė įvairovė skirtingose klimato juostose kinta, svarbiausios spektro dalys išlieka. Ferreira ir kt. (2013) tyrė kintamumą tarp rūšių bei jų viduje tropinėse rūšyse ir nustatė, jog skirtumai tarp rūšių yra penkis kartus didesni nei jų viduje. Modeliuodami lapų optines savybes pastebėjo, jog pigmento koncentracija ir vidinė lapo struktūra yra pagrindiniai parametrai, nulemiantys spektrines rūšių savybes, dėl kurių taip pat gali skirtis staigi raudonojo spektro reikšmių pokyčio vieta (angl. *red edge*). Nustatytos ir reikšmingiausios spektro dalys: matomoji, *red edge* ir trumpoji infraraudonoji spinduliuotės iš hiperspektrinių duomenų. Tuo tarpu Immitzer ir kt. (2019) tyrinėdami Centrinės Europos medžių rūšis multispektrinio Sentinel-2 instrumento pagalba, taip pat išskyrė trumpąją raudonąją spinduliuotę šalia raudonosios matomosios. Atitinkamai pabrėžė pastarosios svarbą spygliuočiams, o trumposios – plačialapiams. Eilė tyrimų naudoja ir išskirtas radiometrines savybes (vidutinis vainiko intensyvumas, procentilių išskyrimas), kurios padeda atskirti rūšis ir net dažnai turi didesnę pranašumą prieš spektrinę informaciją (Hartling ir kt., 2021; Liu ir kt., 2017).

Renkant žemės paviršiaus duomenis nuotoliniu būdu natūralu, jog ne visada būna tinkamos oro sąlygos, per ilgą laiką keičiasi saulės kampas, dėl kurio struktūros mieste gali mesti šešėlius ant stebimo objekto. Hiperspektriniai duomenys kiekviename savo kanale fiksuoja dar mažesnę spinduliuotės kiekį, negu multispektriniai jutikliai, be abejo, tiesioginė spinduliuotė yra labai svarbi, antraip būtų labai mažas signalo ir triukšmo santykis, pikselių atspindėta šviesa gali būti maišyta (Trier ir kt., 2018). Į šiuos nepatogumus tyrimai stengiasi atsižvelgti kurdami vaizdams kaukes. Jos sudaromos naudojantis paskaičiuotu NDVI (normalizuoto skirtumo augmenijos indeksas), spektro reikšmėmis šešėliui/apšviestumui ar sudaromos jų kombinacijos. Trier ir kt. (2018) pasiūlė žaliosios ir mėlynosios spinduliuotės santykį kaip šešėlių kriterijų (mėlynoji spinduliuotė intensyvesnė už žaliąją), tad buvo atfiltruoti pikseliai, kurių mėlyno spektro reikšmės buvo didesnės už žaliojo bei panaudotas NDVI atskirti antropogenines struktūras. Kadangi naudota NDVI reikšmė gana didelė (> 0.55), ši augalija buvo laikoma gyva, o žemai augalijai sukurta

atskira kaukė. Ankstesni tyrimai nurodo, jog klasifikacijos rezultatai buvo geriausi, kai naudojami tik gerai apšviesti pikseliai (Dian ir kt., 2016).

Norint identifikuoti kuo daugiau medžių rūšių gali būti reikšmingi skirtingi mėnesiai, priklausomai nuo jų fenologinių ypatumų. VDU dėstytojas dr. G. Mozgeris akcentuoja, jog rugpjūčio pirma pusė tinka būklei stebėti, o išskirti rūšis tinkamesni gegužės-birželio mėnesiai. Nagrinėjant segmentavimo praktikas daug kalbama apie medžių struktūrą, kuri lazerinio skenavimo duomenyse gali pasirodyti aiškiau sezonais, kai medžiai lapų dar neturi. Nors svarbu turėti vienu metu rinktus duomenis, kad būtų kuo mažiau nesuderinamumų, vykdomi ir daugialaikiai tyrimai, kurie gali pagerinti ne tik bendrą klasifikacijos tikslumą, bet ir atskirai spygliuočių ir lapuočių kategorijoms. Konkrečiai daugialaikię spektrinę analizę pasinaudojo Immitzer ir kt. (2019), panaudojant 6-7 Sentinel-2 vaizdus. Pagal juos nustatyta, jog svarbiausi mėnesiai pasiekti gerą klasifikacijos rezultatą tęsiasi nuo balandžio iki rugpjūčio. Yra tyrimų, naudojusį ir kelių skirtingų metų Sentinel-2 palydovo vaizdus apjungiant su kitais duomenimis miškų vienetams, tačiau naudojami buvo tik keturi vaizdai dėl debesų taršos. Vis dėlto dar kartą pasitvirtina pavasario-vasaros mėnesių naudojimas (Wan ir kt., 2021). Jeigu planuojama naudoti ir lazerinio skenavimo duomenis, Michałowska ir Rapiński (2021) dar kartą patvirtina, jog norint išgauti tai, kas geriausia iš radiometrinių ypatybių, gali prireikti panaudoti laikotarpį be lapijos, nes gaunama daugiau grįžtančio signalo iš žemesnių medžio dalių, taip surenkama daugiau informacijos apie jų vertikalią struktūrą. Kiekvienai rūšiai klasifikacijos tikslumas gali būti gana skirtingas, dėl to bendras klasifikacijos tikslumas gali būti žemesnis, tačiau dažniausiai pasitaikančių ir vietovei charakteringiausių rūšių atpažinimo tikslumas būna didesnis (Liu ir kt., 2017).

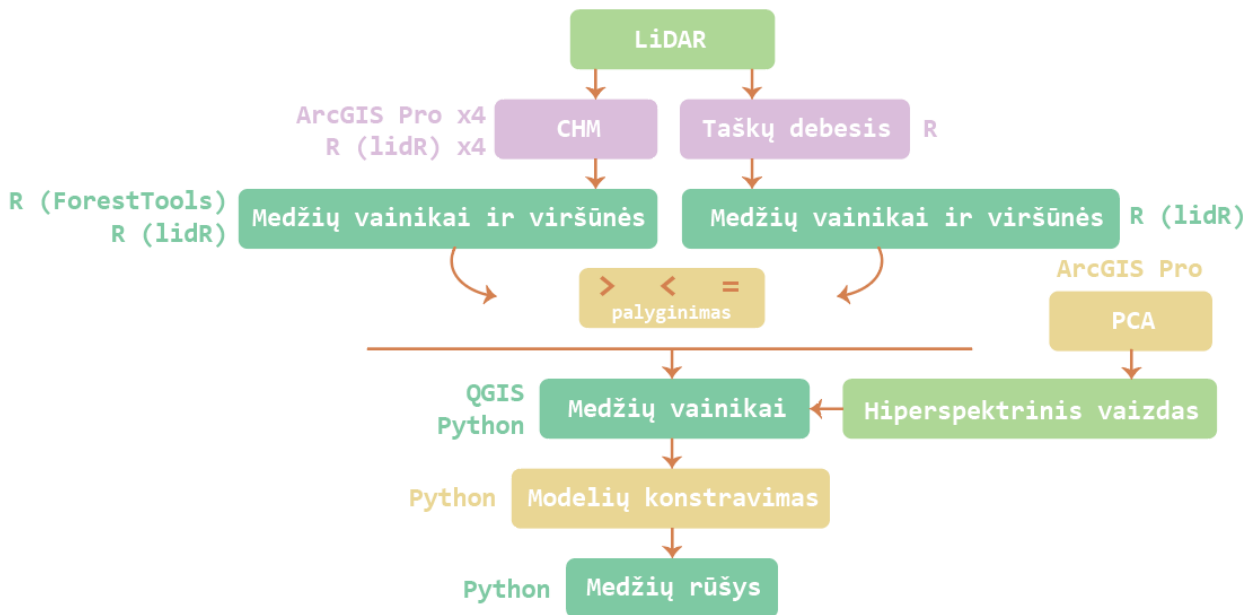
Tyrimai, naudojantys Sentinel-2, turi nemenką erdvinės gardelės dydį (maksimali raiška 10 m) ir klasifikacija iš esmės nėra labai detali – daugiau apibendrinanti plačius miškais apaugusius plotus. Vis dėlto, mažesnės gardelės atveju turima mažiau variacijos klasių viduje. Kuo gardelė smulkesnė, tuo daugiau reikšmių galima pastebėti vienos rūšies viduje, kas apsunkina klasifikaciją. Variacija gali būti mažinama filtrų pagalba. Carleer ir Wolff (2004) panaudojo vidurkio filtrą visiems kanalams bei papildomoms NDVI bei PCA reikšmėms. Nors jų naudojamas metodas buvo pakankamai paprastas, t. y. naudotas maximum-likelihood algoritmas prižiūrimai klasifikacijai pikselių-pagrindu, pasiekė 86 % bendrą tikslumą septynioms skirtingoms rūšims su pora amžiaus grupių. Filtrai sėkmingai panaudojami ir naujesniuose tyrimuose (Chen ir kt., 2023). Wan ir kiti (2021) apjungė daugelį elementų: aukštos kokybės ortofotografijas, Sentinel-2, LiDAR ir daugialaikię analizę, tačiau tai darė miškų struktūriniais vienetams, kuriuos susikūrė segmentuojant. Tai nebuvo individualių medžių klasifikacija, bet labiau dominuojančių jų grupių miško vienetu ir tikslumas pasiekė virš 90 % 11-ai rūšių. Nagrinėjant stambiais vienetais galima pasiekti puikių rezultatų, tačiau tai tampa sudėtingiau tyrinėjant vis detalesnius duomenis dėl didėjančios variacijos.

Norint pasirinkti tinkamą modelį problemai spręsti, įprastai tenka išbandyti kelis. Iš patirties ar algoritmų savybių, turimų duomenų įvertinimo galima nebent numatyti, nuo kurių vertėtų pradėti. Ghosh ir kt. (2014) apibendrinę eilę tyrimų ir pastebėjo, jog daugelis, naudojančių hiperpsektrinių ir LiDAR duomenų sintezę savo tyrimuose, pabrėžia neparimetrinių algoritmų naudojimą. Parametriniai klasifikavimo algoritmai supaprastina funkciją iki žinomos funkcijos. Tai

paspartina mokymosi procesą ir algoritmas suranda funkcijos koeficientus iš turimų duomenų. Tai tinkamesnis variantas paprastoms problemoms spręsti. Tuo tarpu neparimetriniai algoritmai neaproksimuoja funkcijų, todėl yra lankstesni (Almaliki, 2018; Brownlee, 2020). Kaip pavyzdys tyrime paminėti atraminiai vektoriai (angl. *support vector machines*, toliau – SVM). Jie sudaryti taip, jog galėtų atlaikyti didelį dimensijų kiekį, todėl yra gana standartinis pasirinkimas darbams, susijusiems su hiperspektriniais vaizdais. Sumažinus dimensijų kiekį kiti algoritmai gali šį algortimą ir aplenksti, kaip, pavyzdžiui, atsitiktiniai miškai (Hartling ir kt., 2021). Vis dėlto, jie gali būti ir parametriniai ir neparimetriniai (kai naudoja Gauso filtrus). SVM ir atsitiktiniai miškai (angl. *random forest*) plačiai naudojami sudėtingų, daugelį dimensijų turinčių duomenų klasifikavimui (Marrs ir Ni-Meister, 2019; Michałowska ir Rapiński, 2021). Mozgeris ir kt. (2018) išbandė rečiau sutinkamų algoritmų, kuriuos atsitiktiniai miškai taip pat lenkė, bet pasiekė mažesnius rezultatus prieš daugiasluoksnį perceptroną (angl. *multilayer perception*). Turint aukšto dimensiškumo duomenis yra atsisakoma kai kurių, mažai reikšmės turinčių savybių. Dian ir kt. (2016) pastebėjo, jog šių savybių išėmimas SVM nedarė didelės įtakos, ir kad algoritmais pagrįstas ypatybių atsikratymas yra tinkamesnis, negu manualinis.

2. METODIKA

Tyrimas atliekamas distanciniais metodais, pasitelkiant lauko patikrų duomenis validavimui. Medžių viršūnių, vainikų ir aukščių išskyrimui buvo naudojami žemės lazerinio skenavimo duomenys, apdorojami Esri ArcGIS Pro programine įranga bei R programavimo kalbos bibliotekų pagalba. Pirmojoje palyginamojo tyrimo dalyje nagrinėjami du pagrindiniai metodai. Pirmasis, kuriame naudojamas medžių aukščių modelis (CHM), išsiplečia į papildomą GIS įrangos ir R programavimo kalbos bibliotekos (lidR) galimybių, priklausančių nuo skirtingo detalumo LiDAR duomenų bei naudojamų įrankių funkcionalumo, analizę. Sudaroma po keturis medžių aukščio modelius naudojantis ArcGIS Pro ir lidR. Toliau aukščių modeliai yra naudojami medžių viršūnių lokacijos nustatymui ir vainikų išskyrimui iš kiekvieno CHM, pasitelkiant ForestTools ir lidR R programavimo kalbos bibliotekas. Antrame iš pagrindinių metodų LiDAR taškų debesis yra tiesioginis duomenų šaltinis medžių vainikams ir viršūnių lokacijoms nustatyti – naudojamas lidR bibliotekos funkcionalumas.

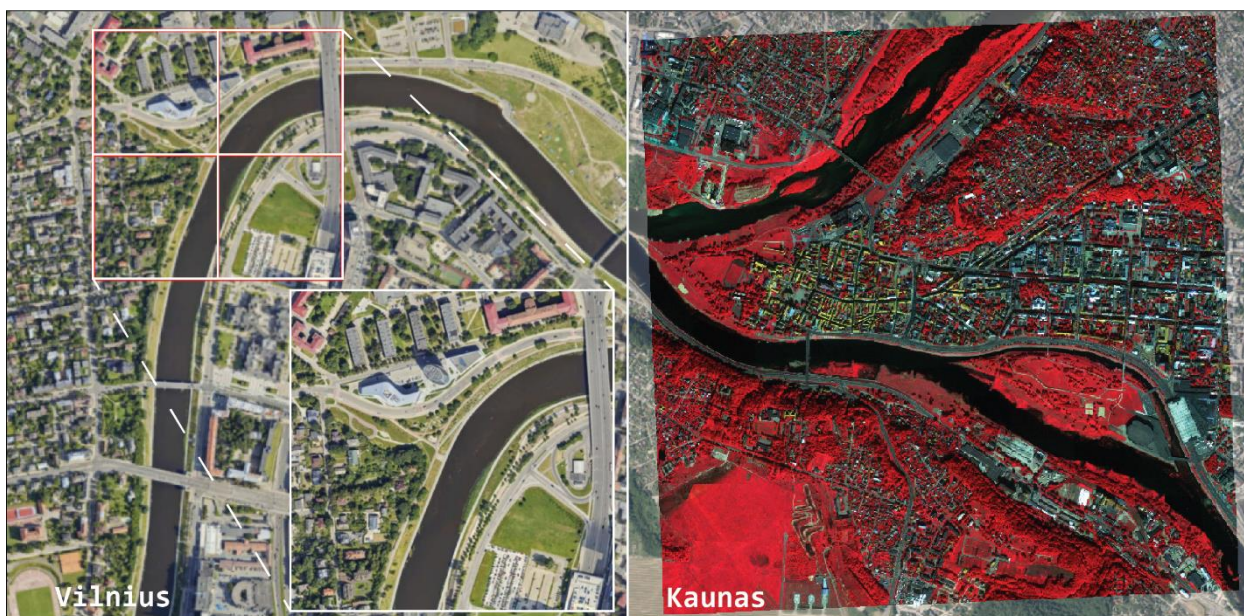


1 pav. Supaprastinta metodikos schema

Gauti medžių fiziniai parametrai yra palyginami ir pasirenkamas optimaliausias variantas. Pereinama prie hiperspektrinio vaizdo, kuriam pritaikyta pagrindinių komponentų analizė (PCA), leidžianti akivaizdžiai pastebėti vaizde esančias klaidas ir jas pašalinti iš tyrimo. Python programavimo kalbos pagalba, gauti medžių vainikai sulyginami su lauko inventorizacijos duomenimis, kad medžiai įgautų rūšies atributą. Hiperspektrinė informacija taip pat yra apibendrintai perkeliama į medžių vainikų duomenis naudojantis atviro kodo QGIS programinės įrangos Orfeo Toolbox įskiepiu. Naudojantis Python išanalizuojami ir pašalinami duomenys su dideliais pasiskirstymo nuokrypiais. Galutinai sukuriama mašininio mokymo modeliai palyginamajai medžių rūšių identifikavimo analizei, jog būtų galima atrinkti tinkamiausią modelį.

2.1 Tiriama teritorija

Tyrime naudojamos dvi skirtingos teritorijos: Kauno ir Vilniaus centrų dalys. Tai yra du didžiausi Lietuvos miestai, kuriems kaupti įvairią erdvinę informaciją yra paprasčiau, nei daugeliui savivaldybių centrų, nekalbant apie dar mažesnius miestelius. Ši dviejų miestų erdvinė atskirtis atsiranda dėl to, jog darbas yra tęstinis ir hiperspektriniai duomenys tyrimo vykdymo metu prieinami tik Kaune, antraip tyrimas būtų tęsiamas sostinėje. Vilniuje tyrimas apima tik 25 ha plotą ir dėmesys skiriamas medžių lajų išskyrimo testavimui ir panaudotų algoritmų lyginimui Žvėryno ir Naujamiesčio seniūnijų dalyse (2 pav.), kur galima pastebėti skirtingą medžių išdėstymo tankį, nes susiliejančioms juos atskirti yra sudėtingiau, negu tikslingais tarpais pasodintus medžius. Įprastai medžiai, kurie yra tvarkingai pasodinti, arčiau gatvių – yra miesto atsakomybės želdiniai. Ne miesto atsakomybės medžių galima rasti ne tik privačiuose sklypuose, bet ir kiemuose tarp daugiabučių. Visi želdiniai, prie kurių įmanoma prieiti, yra inventorizuojami. Kauno aprėptis didesnė – 400 ha, nes tokį plotą užima iš Vytauto Didžiojo universiteto (toliau – VDU) gautas hiperspektrinis vaizdas, kuris naudojamas rūšių identifikavimui jau pasirinkus lajų nustatymui tinkamesnį metodą.



2 pav. Tiriama teritorija ir naudojamų LiDAR lapų apimtis ant „ID Vilnius“ ortofotografinio 2022 m. foninio žemėlapiu (kairėje). Kauno hiperspektrinis vaizdas, vizualizuotas 44-u, 22-u ir 7-u kanalais (dešinėje).

Kadangi ieškoma tinkamo lajų išskyrimo inventorizacijos tikslais metodo, procesai vykdomi tik ant miesto dalies. Taip pat testuojamas ir komercinės programinės įrangos metodas „ID Vilnius“ GIS eksperto, tad komercinio ir atviro sprendimų rezultatų lyginimui įmonėje buvo sutarta naudoti teritoriją. Spektrinės informacijos nagrinėjimas vykdomas tik šio tyrimo autorės, todėl hiperspektrinių vaizdų galimybėms tirti panaudotas atviro kodo lajų išskyrimo sprendimas, kadangi tenka naudoti kito miesto duomenis. Tai tuo pačiu galimybė toliau stebėti metodų panaudojamumą ir suderinamumą, kai pradiniai duomenys gali kisti.

2.2 Naudojami duomenys ir R moduliai

Vilniaus medžių inventorizacijos ir LiDAR duomenys gauti iš UAB „ID Vilnius“, o Kauno inventorizacijos duomenys pateikti VDU, tuo tarpu lazerinio skenavimo duomenys buvo imti iš Lietuvos erdvinės informacijos portalo (geoportal.lt). Vilniaus skenavimas atliktas 2022 m., o taškų debesies tankis ~ 45 taškai/ m^2 . Tokio tankaus skenavimo duomenys suskaidyti į 6 2500 m^2 dydžio nomenklatūrinį tinklėlį visam miestui. Kauno centrui galima panaudoti Lietuvos Respublikos apskričių centrų skaitmeninius erdvinius lazerinio skenavimo taškų duomenis, rinktus 2017 m. Jie taip pat būtų tankūs (vidutiniškai 45 taškai/ m^2). Vis dėlto, atsižvelgiant, jog ir kiti miestai bei miesteliai turi vykdyti inventorizaciją, nuspręsta pasitelkti ir mažiau detalų visos Lietuvos skaitmeninio erdvinio lazerinio skenavimo taškų duomenų rinkinį, sudarytą per 2019–2022 metus. Jų tankis ne mažesnis nei 6,5 taškai/ m^2 , o horizontalus ir vertikalus tikslumas identiškas apskričių centrų skenavimui.

Medžių inventorizacijos duomenys per visą teritoriją yra įvairaus atnaujinimo laipsnio. Stebint naujausią 2022 m. „ID Vilnius“ ortofoto vaizdą galima matyti, kad ne visi medžiai yra suinventorizuoti, nors ir neauga privačiuose gyvenamuosiuose sklypuose arba jau yra atsiradę naujų ar smulkių sodinukų, kurių ortofotografijoje nesimato (3 pav.). Pagal inventorizacijos duomenis vidutinis medžių aukštis 10,9 m. 2514 medžių neturi užfiksuoto aukščio ir 4171 medžių aukštis lygus 0. Kauno inventorizacijos duomenys gauti iš VDU dėstytojo G. Mozgerio. Jie dengia



3 pav. Inventorizacijos duomenys teritorijoje ir trūkumų vietų pavyzdžiai.

tik aktualią hiperspektriniui vaizdui miesto dalį. Dėl naudojamų skirtingų tankių LiDAR duomenų, kurie gali būti surinkti ne tais pačiais metais kaip ir kiti lyginamieji ar naudojami vaizdai, gali atsirasti nesutapimų. Apdorotu ir kalibruotu specialiomis programomis hiperspektriniu Kauno centro vaizdu testavimo ir mokslo tikslais pasidalino G. Mozgeris. Vaizdo raiška yra 0,5 metro, o hiperspektrinį rastrą sudaro 64 skirtingi kanalai nuo mėlynojo iki infraraudonojo spektro dalių.

Medžių viršūnių detekcijai ir vainikų riboms išskirti iš CHM naudojama ForestTools (Plowright ir Roussel, 2023) R programavimo kalbos biblioteka. Ji pateikia įrankius analizuoti

nuotoliniais metodais gautus miško duomenis. Bibliotekos pagalba galima gauti medžių viršūnes ir juos atitinkančius vainikus. Medžių segmentavimui tiesiogiai iš taškų debesies naudojama lidR (Roussel ir kt., 2020; Roussel ir Auty, 2022) biblioteka, kuri taip pat medžių paieškai gali naudoti ir sugeneruotą CHM vietoje taškų debesies. Tai ne originali, bet optimizuota greičiui biblioteka, kuri buvo pagrindas Python programavimo kalbos bibliotekai PyCrown. lidR paketas platesnio

funktionalumo, negu ForestTools. Jis taip pat skirtas LiDAR duomenų manipuliacijai ir vizualizavimui.

2.3 Medžių aukščio modelio (CHM) generavimas

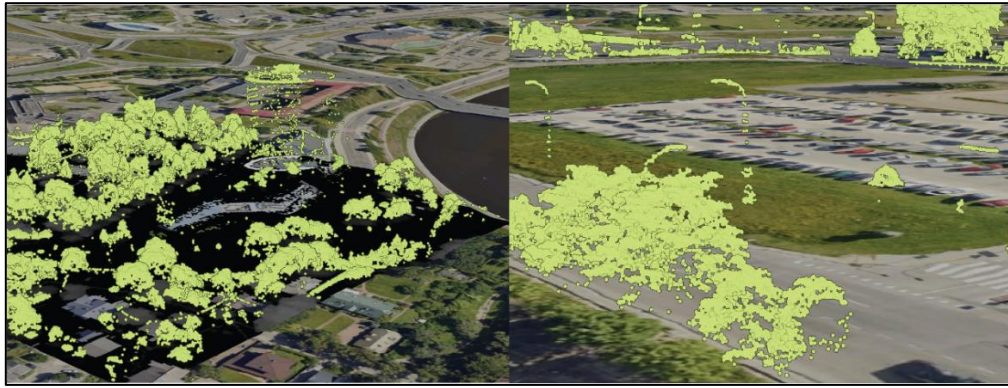
Medžių aukščio modelis įprastai gaunamas iš pirmo atspindžio taškų rastro atimant žemės paviršiaus rastrą. Mielcarek ir kt. (2018) tyrimo rezultatai rodė, jog CHM generavimo metodas daro įtaką medžių aukščių nustatymo tikslumui. Jeigu kuriant rastrą naudojamas metodas, kuris pašalina tarpus paprasta filtracija – tai daro įtaką visiems rastro pikseliams (ne tik tarpams, angl. *pits*), tad nemažai pamenkina medžių aukštį. Tik parinkimo metodai, kurie išima nereikalingus pikselius (pavadinta juodosiomis skylėmis), išlaiko medžių aukščius ir nuokrypiai nuo tikrųjų matavimų yra mažesni.



4 pav. Iš kairės į dešinę: 76_32-0165 lapo pirmo atspindžio rastras, žemės paviršiaus rastras ir sugeneruotas CHM.

Naudojantis ArcGIS Pro, keturi Vilniaus miesto taškų debesų (.las) failai buvo atfiltruoti pagal pirmą atspindį tik aukštos augmenijos taškams, nes mieste yra kitų objektų, kurie gali pasimaišyti tolimesniuose etapuose. Konvertuojant taškų debesį į rastrą *LAS dataset to raster* įrankiu buvo pasirinktas *binning* interpoliacijos metodas, jis leidžia pasirinkti kokia pikselio vertė nustatoma ar paskaičiuojama pagal į jį patenkančius taškus. Celių priskyrimas nustatomas, jog parinktų maksimalią aukščio (Z) reikšmę. Tarpai nebuvo pildomi, kad nesudarytų išstętos „uodegos“ objektų kraštuose dėl interpoliavimo. Galutinai parenkamas 1 m celės dydis. Žemės paviršiui gauti, taškų debesis filtruojamas pagal žemės paviršiaus atspindžius ir atliekamas identiškas vertimo į rastrą procesas, tačiau celių priskyrimas pakeičiamas į minimumą ir tarpai yra užpildomi, nes žemė dengia visą teritoriją. ArcGIS Pro *Minus* įrankiu iš medžių pirmo atspindžių rastro atimamas žemės paviršiaus rastras – gaunamas CHM.

Kadangi nenormalių reikšmių nesimato (ne daugiau 50 m), tad CHM generavimas laikomas baigtu (4 pav.). Vis dėlto pastebėtina, jog LiDAR taškų debesis nėra visiškai švarūs, tad aukštai augmenijai priskiriamos pastatų, turėklių, tiltų dalys ir kitokie artefaktai. Gatvių šviestuvai interpretuoti kaip aukšta augmenija. Dėl tankios lapijos kartais užfiksuotos tik pačios viršutinės medžių vainikų dalys ir dažnai kamienų nesimato (5 pav.).



5 pav. „Nešvarūs“ LiDAR duomenys.

ForestTools pirminė rasta rekomenduojama CHM raiška buvo 1 metras, tačiau Kauno hiperspektrinis vaizdas yra 0,5 m, todėl ArcGIS Pro buvo sudaromas ir šis tankesnis CHM. Gaminant detalesnį rasterizuotą taškų debesies vaizdą naudojant mažesnio tankumo LiDAR duomenis buvo galima pastebėti medžių lajose esančius tarpus (6 pav.), kuriems užpildyti, kaip minėta, tyrimuose ieškoma tinkamiausių būdų. Gardelėms sumažėjus paprasčiausiai nebėra taškų, padedančių jas padengti, tad mažiau tankiems debesų taškams reikalingas *simple* tarpų užpildymo parametras, kuris neįmanomas be *binning* interpoliacijos tipo ArcGIS Pro.



6 pav. Skylės nenaudojant užpildo nustatymuose ArcGIS Pro (kairė) ir užpildytos

Kadangi tyrimas tęstinis, jis buvo praplėstas gilesniu CHM nagrinėjimu. Medžių aukščio modelį galima naudoti ir lidR funkcijose vietoje taškų debesies. Bibliotekos funkcionalumas platus ir suteikia galimybę kurti rastrus iš lazerinio skenavimo duomenų. Su lidR taip pat buvo sukurtos 1 bei 0,5 metro gardelės CHM iš skirtingų skenavimo duomenų (~ 45 taškai/m² ir $\sim 6,5$ taškai/m²) R programavimo kalbos aplinkoje. Pirma yra nuskaitomas taškų debesies failas, normalizuojamas aukštis, kad objektai atsidurtų ant lygaus paviršiaus. Tam metodas gali naudoti pagamintą žemės vietovės modelį arba jį sudaryti tuo momentu vykdant kodo eilutę. Panaudotas *tin* algoritmas *normalize_height* funkcijoje ir atsikratomos tuščios reikšmės papildomu parametru. Vėliau paliekamos tik lygios ir aukštesnės už 0 aukščio (Z) reikšmės filtruojant taškus su *filter_poi*.

Pasirenkami taškai, priklausantys aukštos augalijos klasei (LASHIGHVEGETATION) – lidR klasifikacija paremta ASPRS standartu. Iš atrinktų medžių taškų gaminamas CHM rastras. Funkcija *rasterize_canopy* leidžia nustatyti kokio maksimalaus ilgio trianguliacines trikampių kraštines atmesti, kad būtų užpildyti tarpai medžių lajose, o nuo jų kraštų nenusitęstų suinterpoliuotos „uodegos“. Sėkmingas CHM sudarymas R kalba leistų nenaudoti papildomos atviros ar komercinės GIS įrangos ir procesus paprasčiau integruoti. Iš viso buvo sudaryti 8 CHM: po 4 naudojantis ArcGIS Pro ir R, jog abu generavimo metodai panaudotų abiejų tankių LiDAR skenavimo duomenis dvejomis skirtingomis (1 ir 0,5 m) raiškomis.

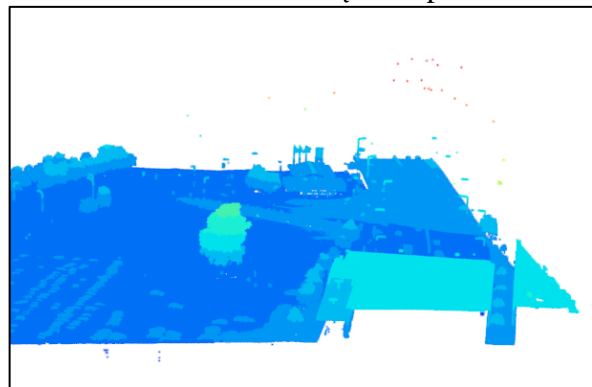
2.3 Medžių segmentavimas R aplinkoje

Persikėlus 1 m raiškos ArcGIS Pro medžių aukščių modelį į R aplinką, ForestTools pagalba galima iš karto pradėti išskirti medžių viršūnes ir vainikus, bet prieš tai reikalinga apibrėžti slenkančio lango funkciją. Kadangi medžių charakteristikos mieste labai įvairios, statistiniai metodai, antraip nei miškininkystėje (kur pasitaiko vienodų charakteristikų medynų), nepadėtų įvertinti, kokios konstantos turėtų būti įrašytos į funkciją:

$$(1) f(x) = x * 0,03 + 2,5$$

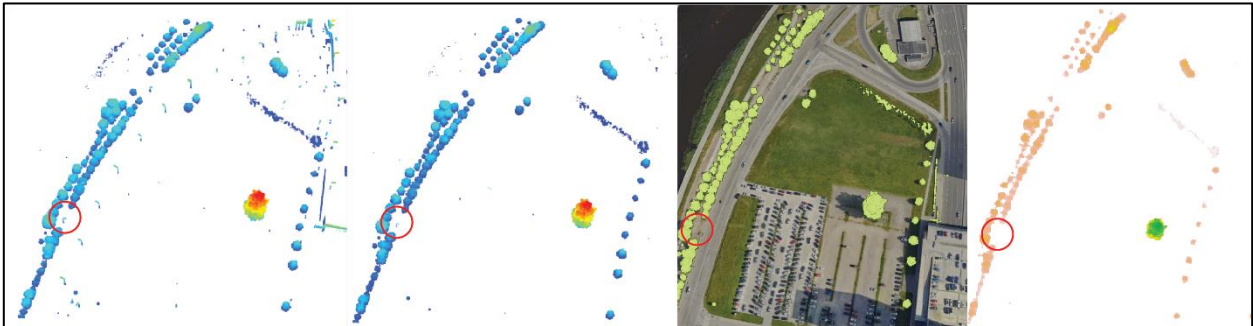
Medžių viršūnės aptinkamos naudojant kintamojo lango filtrą (angl. *variable window filter*). Ši paketo funkcija naudoja rastro raišką (x – rastro pikselio reikšmė), kad nuspręstų kiek celių langui reikia padengti. Nurodomas sugeneruotas CHM, lango funkcija (1) ir minimalus aukštis. Tenka eksperimentuoti su parametrais ir parinkti tinkamiausias konstantas. Formulėje kaip pavyzdys paimtos konstantos, kurios buvo naudotos 76_32-0165 nomenklatūrinio „ID Vilnius“ lapo taškų debesies sugeneruotam 1 m raiškos CHM, naudojantis ArcGIS Pro. Turint viršūnes pereinama prie žymių kontroliuojamos vandenskyros transformacijos (angl. *marker-controlled watershed*) algoritmo medžių vainikų riboms segmentuoti. Medžių viršūnės reikalingos segmentacijai – atlieka žymių funkciją. Šiai funkcijai reikalinga pateikti viršūnių duomenis, CHM, minimalų aukštį, kuris paliktas kaip 4 metrai iš viršūnių detekcijos. Gaunamas rastras paverčiamas į poligonus ir eksportuojamas *.shp* formatu, taip pat išeksportuoti ir viršūnių taškai.

Baigus segmentavimą visiems keturiems lapams Vilniuje naudojant CHM, pereinama prie tiesioginio taškų debesies segmentavimo naudojant lidR. Nuskaitant *.las* failą ir taupant resursus, galima nuskaityti tik dominančius taškus ar klasifikacijos atributus. 76_32-0186 lapas buvo šiek tiek apvalytas perklasifikuojant taškus ArcGIS Pro. Nuskaitomos x , y ir z taškų reikšmės, taip pat atspindžio numeris (reikalingas vertimo į rastrą funkcijai) bei klasifikacijos duomenys. Kaip ir galima tikėtis, buvo pastebėta, jog vaizdas turi artefaktų (7 pav.). Tam, jog segmentuoti medžiai turėtų realistišką aukštį, taškų debesies normalizuojamas naudojant *tin* algoritimą bei



7 pav. Artefaktai 76_32-0186 taškų debesyje.

pašalinamos tuščios reikšmės, taip pat atfiltruojami taškai tarp 0 ir 30 aukščio reikšmių ir parenkami tik aukštos augmenijos taškai. Vizualizuojant apvalytą *las* failą R aplinkoje, kai kurios klaidos vis tiek pasirodo, kurių nesimato ArcGIS Pro arba taškų debesį pavertus rastru (8 pav.).

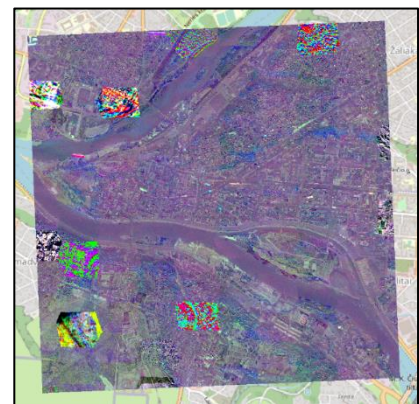


8 pav. LiDAR nesutapimai originaliame, apvalytame debesyje R aplinkoje, apvalytame ArcGIS Pro aplinkoje ir rasterizuotame variante R aplinkoje.

Panašiai į ForestTools eigą, panaudojama (1) funkcija su konstantomis 0,1 ir 5, kuri paskaičiuoja lango dydį kaip taško ar pikselio aukščio funkciją. Nurodant *.las* failą ir *lmf* algoritimą, kuris priklauso nuo funkcijos, iš taškų debesies su *locate_trees* identifikuoja medžių viršūnes gana neblogo rezultatu. Vis dėlto, dirbant su taškų debesimi, medžių viršūnių ir medžių segmentavimas vyksta atskirai. *li2012* algoritmas skirtas medžių segmentavimui. Priklausomai nuo parametrų gali išskirti medžius pagal šakos viršūnę arba medžio viršūnę, todėl naudojami artumo parametrai. Dėl jų tektų derinti parametrus skirtingiems *.las* lapams, atsižvelgiant į medžių išsidėstymo ypatumus. Visi keturi lapai buvo segmentuoti su numatytais algoritmo nustatymais, o 76_32-0186 lapą papildomai bandyta paderinti, nustatant kitas pirmą ir antrą ribas, taško aukštį (nuo kurio atitinkamai naudotų kitą ribą) ir minimalus aptikto medžio aukštis neturi būti mažesnis, nei 4 metrai. Su biblioteka taip pat galima gauti vainikų formas, tuo pačiu apskaičiuojant išgaubtą ar išgaubtą korpusus (*crown_metrics* funkcija). Tai buvo įtraukta į analizės procesą ir išeksportuota *.shp* formatu.

2.4 Spektrinių ir vektorinių duomenų apdorojimas bei paruošimas mašiniam mokymui

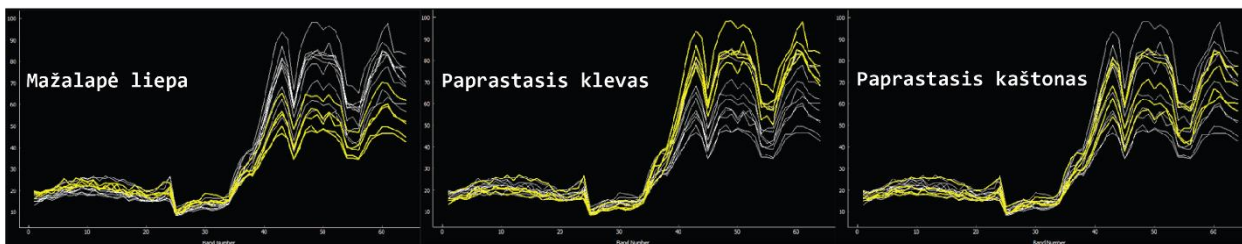
Pirminė idėja buvo panaudoti PCA, jog būtų sumažintas hiperspektrinio vaizdo dimensiskumas paliekant išraiškingiausias savybes. Prieš PCA, stebint vizualizuotą netikrų spalvų (angl. *false colour*) kompoziciją, išryškinančią augaliją, kilo klausimas ar mozaika korektiška. Dėl įvairių vaizdo vietų (ypatingai pastebimų infraraudonuosiuose kanaluose), po PCA buvo akivaizdu, kad kanalai keliuose vietose yra prastai registruoti (9 pav.). Tam, kad neatsirastų klaidingų duomenų dėl kanalų pikselių pasislinkimo, inventorizuoti Kauno medžiai šiose vietose buvo atmesti iš duomenų rinkinio. PCA



9 pav. Prastai registruoti kanalai.

vertino visus vaizdo aprėptyje esančius objektus, tai įskaito ir pastatus, kitą infrastruktūrą. Dėl to galėjo sumažėti PCA efektyvumas medžių rūšims išskirti, kas be ko, ji sudaro sintetinius kanalus.

Su QGIS EnMAP-Box įskiepiu, kuris buvo sudarytas specialiai vokiško hiperspektrinio palydovo vaizdų analizei, panagrinėti įvairių medžių rūšių pikselių spektriniai parašai. Iš pavienių pikselių parašų galima susidaryti įvaizdį, kokia variacija egzistuoja rūšių viduje (10 pav.). Kanalu reikšmėse nuo maždaug 38 iki 64 pastebimi didesni parašų skirtumai.



10 pav. Skirtingų medžių rūšių parašų įvairovė pikseliuose

Kauno centro medžių vainikų riboms sudaryti naudojami Lietuvos skaitmeninio erdvinio lazerinio skenavimo taškų duomenys ($\geq 6,5$ taškai/m²). Parašytas R kodo ciklas, kuris, kaip ir anksčiau aprašyta CHM generavimo metodika, R aplinkoje normalizuoja visus nomenklatūrinius LiDAR duomenis ir atfiltruotus medžius konvertuoja į rastrą bei išsaugo kiekvieną medžių aukščio modelį naujame faile. ArcGIS Pro turimi CHM apjungiami su *Mosaic to New Raster* įrankiu ir suvienodinamos aprėptys CHM apkirpus pagal hiperspektrinį vaizdą su *Extract by Mask*. Gautas vientisas CHM perkeliamas atgal į R aplinką medžių viršūnių ir vainikų detekcijai. Aukščiausioms viršūnėms nustatyti naudojamas fiksuoto dydžio ($ws = 9$) apvalios formos langas *locate_trees* funkcijoje. Su *sp* R biblioteka aptiktos *simple features* tipo viršūnės paverčiamos į *spatial* tipo duomenis, jog būtų galima išsieksportuoti į Esri Shapefile formatą GIS programinei įrangai skaityti. Toliau kuriami vainikų poligonai. Sukuriamas algoritmo objektas pagal *dalponte2016* funkciją, kurioje nurodomas CHM bei nustatytos viršūnės. Jis panaudojamas kaip funkcija, pritaikoma naujai kuriamam objektui R aplinkoje, šiuo atveju, tai yra medžių vainikai, pateikiami rastro formatu. Sukuriami vektoriniai duomenys panaudojant *rasterToPolygons* funkciją, kuriuos jau galima eksportuoti shapefile formatu.

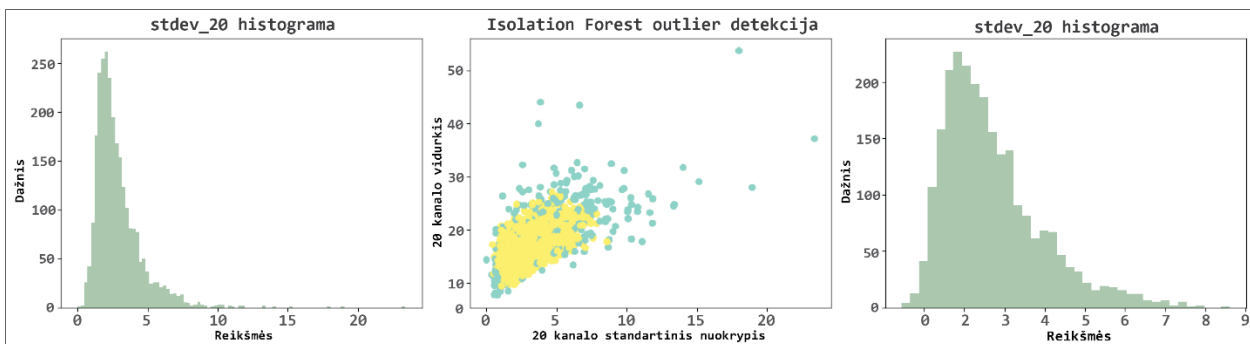
Kiekvienas pikselis buvo paverstas vektoriumi turinčiu atributą, priskiriantį jį prie atitinkančio medžio vainiko. Panaudotas *Dissolve* ArcGIS Pro įrankis, jog vieno vainiko vektoriniai pikseliai būtų sujungti į vieną. GitHub svetainėje rasta *tree-detection-evo* repozitorija, kurioje patalpintas Mäyrä ir kt. (2021) medžių rūšių identifikavimo tyrimo kodas. Pasinaudojant jų repozitorija vykdomas inventorizacijos ir nustatytų duomenų apjungimas automatizuotam paruošimui vykdyti mašininį mokymą, kodo dalis pritaikant šiam tyrimui. Pirmiausia medžių vainikams sukuriama papildomi atributai, kurie saugotų medžių viršūnių koordinatas ir vainikų apimamą aprėptį. Inventorizuotiems medžiams taip pat yra išskiriamos į atributus ilgumos ir platumos koordinatės GIS įrangos pagalba. Su *Summary Statistics* ArcGIS Pro paskaičiuoti kiekvienos rūšies medžių vienetai, tad buvo atrenkamos tik 5-ios plačiausiai paplitusios medžių rūšys.

Turimas pilnos aprėpties vektorinis sluoksnis suskaidomas dalimis, sukuriant vainikų vektorines kaladėles, jog būtų taupomi kompiuteriniai resursai apdorojimo metu. Į atmintį buvo pakraunami visi vainikai ir apjungtas hiperspektrinis vaizdas su CHM. Parinktas 625 pikselių kaladėlių dydis, naudojamas ir rastrui ir vektoriams. Naujame kodo faile užkraunami inventorizacijos duomenys bei kaladėlės. Šioje dalyje surandamas artimiausias inventorizuotas medis ir priskiriama medžių rūšis identifikuotiems vainikams. Tai padeda pamatuoti nustatytą ir inventorizuotą medžių koordinatų informacija, gauta anksčiau. Jeigu vainikas turi tik vieną nustatytą tašką, jis ir yra priskiriamas prie medžio bei įgauna inventorizacijos duomenis apie rūšį. Originaliame Python kodo cikle naudojamas *Pandas* bibliotekos *pd.append* metodas jau panaikintas, tad buvo pakeistas į naująjį *pd.concat*, užkomentuotos vietos, kuriose paliktos neegzistuojančios funkcijos, priderinta *utils.py* kodo failo kopija, aprašanti reikalingas funkcijas, kurias galima importuoti į pagrindinį kodą. *label_contours* funkcijai išimama GPS dalis, kadangi Kauno inventorizacijos duomenys rinkti vienu būdu ir prioretizuoti GPS duomenų nereikia. Taip pat funkcijoje buvo reikalinga pakeisti laukų pavadinimus į atitinkančius turimus duomenis. Nustatyti medžiai be juos atitinkančio inventorizuoto buvo pašalinami. Galutinis išvesties rezultatas yra bendras visos aprėpties vektorinis vainikų sluoksnis, turintis aktualią rūšies informaciją. Medžių skaičius, naudojamas tyrime, susitraukė ne tik nuo pasirinkto nagrinėjamo rūšių kiekio ar prastai registruotų kanalų, bet ir nuo šio automatizuoto savybių paskirstymo proceso, padedančio paspartinti darbų eigą ir greičiau apdoroti didelį kiekį duomenų (1 lentelė).

1 lentelė. Duomenų pokytis po automatinio apdorojimo.

Inventorizuota medžių rūšis	Inventorizuotas kiekis	Modelių apmokymui likęs medžių kiekis
Liepa mažalapė	3740	2034
Klevas paprastasis	1055	573
Kaštonas paprastasis	652	331
Klevas uosialapis	603	208
Beržas karpotasis	420	205

Panaudojamų duomenų kiekis modelio apmokymui objektų pagrindu maždaug du kartus (mažalapė liepa, paprastasis klevas ir kaštonas) ir daugiau (uosialapis klevas) mažesnis, negu inventorizuotų medžių rūšių. Galiausiai QGIS Orfeo ToolBox (toliau – OTB) įskiepio pagalba paskaičiuojama zoninė statistika (*ZonalStatistics*), kuri papildo vainikų vektorinius duomenis kiekvieno kanalo minimalaus, maksimalaus, vidurkio ir standartinio nuokrypio reikšmių atributais, tačiau prieš tai pridėdamas *class_id* atributas rūšies užkodavimui skaitine reikšme. Python aplinkoje importavus *train_test_split* iš *Scikit-learn* bibliotekos. Nustatomas 0,2 (20%) validavimo duomenų kiekis nuo bendro duomenų rinkinio. Atskirai išsaugomi apmokymo ir validavimo duomenys. Tikrosios reikšmės iš abiejų duomenų rinkinių surenkamos į atskirus objektus, išsaugoma geometrija, jog būtų galima ją pridėti atgal peržiūrai GIS įrangoje. Atsikratomi visi nereikalingi atributai, paliekant tik OTB statistikos sugeneruotus laukus. Toliau renkami ir formuojami mašininio mokymo modeliai, ieškant geriausių tikslumo (angl. *accuracy*) rezultatų.



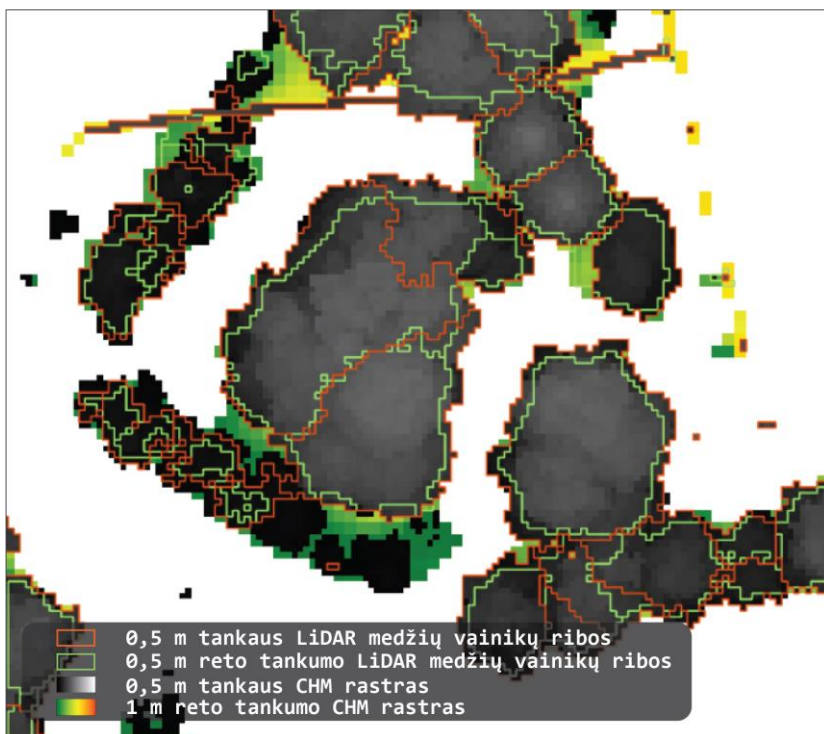
11 pav. Reikšmių pasiskirstymas prieš ir po IsolationForest panaudojimo 20-ame kanale.

Vykdomi bandymai pakartoti modelio sudarymą sumažinus naudojamų atributų skaičių ar pašalinus papildomus, nukrypstančius nuo daugumos, duomenis (angl. *outliers*). Įtartinų duomenų pašalinimui naudojamas *IsolationForest* (taip pat iš *Scikit-learn*) – jam reikia nurodyti tik atmetamą duomenų dalį (užterštumą), ją pritaikant pagal turimus duomenis. Šiuo atveju paliktas 0,1. Duomenų pokytį padeda stebėti konkrečių atributų histogramos (11 pav.). Tikėtina tarp jų turėtų atsirasti vainikai, kurių ribos prastai dera su vaizdais ar turi daugiau ploto šešėlyje ar patenka ant kitų struktūrų, negu kiti rūšies atstovai.

3. REZULTATAI

3.1 Sudarytų CHM ir R aplinkoje nustatytų medžių viršūnių rezultatai

Tankesnio LiDAR skenavimo duomenys sudaro mažiau problemų renkantis CHM detalumą, kas yra labai intuityviai suvokiama. Tuo tarpu smarkiai sumažinus taškų kiekį kvadratiniam metre tenka pagalvoti apie naudojamo įrankio parametro (*void fill method*) pakeitimus tarpams vainikuose užpildyti, naudojant ArcGIS Pro. Sudarant 0,5 m medžių aukščių modelius GIS įranga su parametru pokyčiu bei R biblioteka pavyko gauti panašius rezultatus. ArcGIS Pro būtinai turi būti nustatytas *binning* interpoliacijos metodas, jog būtų galima rinktis tarpų užpildymo metodą, o su lidR bibliotekos siūlomu *tin()* metodu galima konkrečiai nurodyti skaitinę reikšmę, tad rezultatų panašumas priklausytų nuo parenkamo skaitmens ilgiausiai trikampio kraštinei. Su R padidėja gaunamų duomenų valdymas. Mažesnio tankumo lazerinio skenavimo duomenis verčiant į rastrą R aplinkoje svarbu nepamiršti pritaikyti tinkamo dydžio trikampio kraštinės reikšmę. 0,5 m raiškiai



12 pav. CHM skirtumai priklausomai nuo LiDAR tankio ir gardelės dydžio.

tinkantis skaitmuo gali sudaryti tarpus 1 m medžių aukščio modeliui.

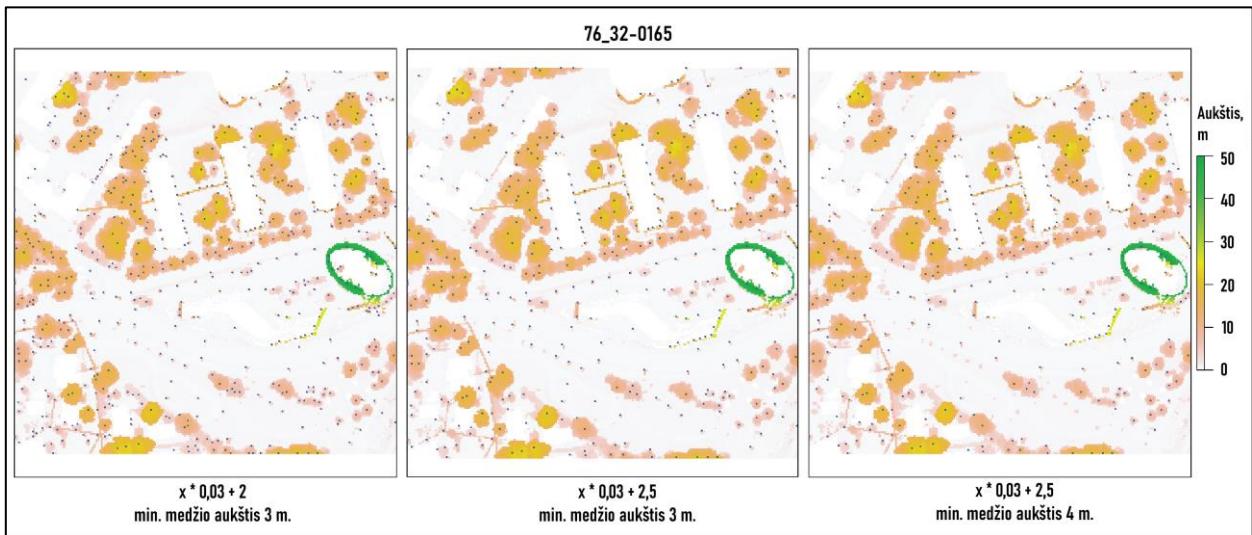
CHM matomi medžių vainikai didesni naudojant tankesnę lazerinį skenavimą. Žinoma, tai greičiausiai daugiausia lemia taškų tankis, taip pat yra tikimybė, jog nesutampa skenavimo sezonas, todėl lapijos yra daugiau, ar medžiai paprasčiausiai paaugo, jeigu būtų praėję keletas metų tarp lyginamų duomenų. Taip pat gardelės sustambėjimas gali nežymiai padidinti vainiką bei bendrą CHM užpildomą plotą, tačiau bendras padidėjęs CHM plotas nereiškia, jog bus

išskiriama daugiau medžių, tai paprasčiausiai interpoliavimo išdava. Padidėjęs plotas gali būti žemos aukščio reikšmės, kurios įtakos segmentavimo algoritmui neturi, jei yra žemiau nustatyto tinkamo aukščio arba papildomas plotas yra tolygus (12 pav.).

Detalesnis taškų debesis užfiksuoja daugiau detalių, todėl neturint idealiai klasifikuotų lazerinio skenavimo duomenų, CHM galima išvysti elektros stulpų apjungimų. Nors ir smulkesne gardele (0,5 m), tankaus taškų debesis atveju šios detalės išlieka. Tikimybė juos išvysti mažesnio tankumo taškų debesis CHM sumažėja – galima pastebėti liekanas, bet ne pilnus objektus. Verta pastebėti, jog tokios detalės yra didesnė problema generuojant CHM su ArcGIS Pro. Sudarant CHM lidR bibliotekos pagalba, minėtų apjungimų praktiškai nesimato tankaus LiDAR rezultate,

kad ir naudojamas tas pats neidealus taškų debesis. R metodas leidžia išvengti papildomų klaidų CHM generavimo metu.

Medžių viršūnės yra aukščiausios identifikuotos medžio vietos, kurios nebūtinai sutampa su kamieno padėtimi, tačiau reprezentuoja medžio egzistavimą bei jo aukštį. Sugeneruoti ArcGIS Pro 1 metro raiškos CHM pagal LiDAR duomenis, turėjo artefaktų, kuriuos algoritmas taip pat įvertino. Naudojant ForestTools ir derinant algoritmą sudėtinga jį pritaikyti egzistuojančiai medžių įvairovei. Tiksliau išgaunant stambesnius medžius gaunamas nepakankamas medžių viršūnių skaičius, o derinant modelį prie smulkesnių medžių, jų kiekis yra pervertinamas, rezultatai taip pat nemankai priklauso nuo parinkto minimalaus medžio aukščio (13 pav.).



13 pav. Nustatytos medžių viršūnės pagal skirtingas reikšmes ir parametrus.

Vis dėlto, visiems nomenklatūriniais lapams vienas iš optimalių variantų pasirodė $f(x) = x * 0,03 + 2,5$ (visada yra galimybė atrasti dar labiau tinkančias konstantas), nustatant 4 metrų minimalų medžio aukštį, jeigu norima per daug nenukrypti nuo inventorizuotų medžių skaičiaus. Inventorizuotų medžių ir nustatytų medžių viršūnių 1 m raiškos ArcGIS įranga generuoto CHM rezultatai (14 pav.) rodo, kokie skirtumai susidaro tarp rezultatų skirtinguose lapuose. Medžių viršūnės, susidariusios ant pastatų kraštų dėl klaidų/artefaktų, buvo pašalintos parenkant 1 metro atstumą nuo pastatų, į kuriuos galėjo patekti vienas kitas tikras medis, tačiau bendrai skirtumai sumažėja (2 lentelė).

2 lentelė. Inventorizuotų medžių ir nustatytų medžių viršūnių skirtumai LiDAR lapuose.

LiDAR lapai	76_32_0165	76_32_0166	76_32_0185	76_32_0186
	Medžių kiekis			
Prieš apvalymą	391	205	321	252
Po apvalymo	284	199	286	183
Inventorizuota	324	168	263	138
Skirtumas prieš	67	37	58	114
Skirtumas po	-40	31	23	45

76_32_0165 lape (šiaurės vakarai) vis vien lieka elektros stulpų, šviestuvų, kuriems buvo nustatytos medžių viršūnės. Medžių sankaupose su stambiais vainikais rezultatas dažnai prastas. Yra vietų, kur smulkesni medžiai labai tiksliai identifikuoti, tačiau yra ir vietų, kur, žiūrint į visumą, buvo galima tikėtis geresnio rezultato. Sunku prognozuoti, kuriose vietose panašus erdvinis medžių išdėstymas pateiks geresnius rezultatus. Teigiama tai, jog ten, kur medžiai nebuvo inventorizuoti, vietomis neblogai buvo atpažinti.



14 pav. ForestTools nustatytos medžių viršūnės ir inventorizuoti medžiai.

76_32_0166 rezultatai, vertinant medžių skaičių ir tikslumą, buvo geriausi, iš CHM vizualiai gana gerai matosi stambesnių medžių viršūnės. Smulkūs medžiai po stambiais vainikais neįvertinti. Apvalius sluoksnį pagal atstumą nuo pastatų, vis dar lieka medžių ant tilto struktūrų, tad skirtumas dar labiau sumažėtų. Taip pat medžių skaičius didesnis dėl krantinės augmenijos. Pastebima, jog

vieno stambaus medžio vainikai gali turėti išskirtas dvi viršūnes, o arti vienas kito augantys didelių vainikų medžiai sulieti į vieną viršūnę.

76_32_0185 vyrauja chaotiškas medžių išsidėstymas dėl didelio kiekio gyvenamųjų sklypų. Daug vietų tankiai apaugusios įvairaus dydžio medžiais. Labai sunkiai interpretuojama medžio viršūnės lokacija, nors nustatytų ir inventorizuotų medžių skirtumas mažiausias ir labiausiai išsibalansuoja. Šiame lape, antraip nei prieš tai nagrinėtame, yra inventorizuota nemažai krantinės smulkių želdinių. Šiaurės rytiniame lape matėsi išskirta krantinės augmenija, tačiau šiame lape jos nustatyta mažiau. Šioje vietoje yra inventorizuota nemažai šviežių sodinukų.

76_32_0186 lape matosi medžių eilių tarpai, taip pat problemos dėl panašaus aukščio susiliejusį lajų. Išskirti šviestuvai, kiti infrastruktūros elementai, vėlgi labai smulkūs nauji želdiniai nepateko į išskirtas viršūnes. Norint atsikratyti šviestuvų būtų galima išbandyti juos atfiltruoti pagal lajos ploto-aukščio reikšmes, nes tokių aukštų ir siaurų medžių neturėtų pasitaikyti.

ForestTools labai jautrus aukščio pokyčiams ir vietomis išskiria daugiau viršūnių, kai vainikai stambūs ir sudėtingos struktūros. Galima gauti geresnius rezultatus tokio pobūdžio medžiams, tačiau smulkesnių medžių kaina.

lidR bibliotekoje medžių viršūnės yra nustatomos, tačiau vėliau nenaudojamos medžių segmentavimui tiesiogiai iš taškų debesies (15 pav. 76_32-0186 įkarpa). Dėl šios priežasties lyginti nustatytas medžių pozicijas šiais skirtingais būdais turi menką reikšmę. Nustatytų medžių ir viršūnių skaičius skiriasi, norint šia biblioteka išskirti medžius tiesiogiai iš LiDAR taškų debesies.

Lyginant 1 m ArcGIS Pro tankaus CHM medžių viršūnes su 0,5 tankaus CHM nustatytų viršūnių pozicijomis pastebimas menkas nuokrypis (dažniausiai 35 cm) tarp tų pačių viršūnių taškų, kuris atsiranda dėl gardelių dydžio skirtumo. Taip pat dėl detalesnės gardelės galima pastebėti daugiau išskirtų viršūnių ant miesto struktūrų. Mažesnio tankumo 1 m CHM vienodų medžių viršūnės skiriasi didesniu atstumu (1 – 1,41 m) nuo tankesnio 1 m CHM rezultato arba visiškai sutampa. Tarp visų skirtingų nustatytų medžių viršūnių variantų galima atrasti absoliučiai sutampančių. Kaip ir su tankiu CHM, mažiau tankaus CHM 1 ir 0,5 m raiškos rastrų medžių viršūnių padėtis skiriasi 35 cm atstumu.

R aplinkoje generuotų CHM rezultatai atkartoja ArcGIS Pro, kada tarp aukštesnio ir žemesnio rastrų raiškos detalumų dominuoja 35 cm medžių viršūnių skirtumai, o tarp skirtingų LiDAR tankių, bet vienodos raiškos CHM, pozicijų skirtumai siekia dažniausiai 1-1,41 m atstumą. Visais atvejais pastebima sutampančių taškų. Yra medžių, kurie daugeliu skirtingų rezultatų metu absoliučiai sutampa, jeigu turi tą pačią raišką ir nesvarbu, koks buvo LiDAR tankis ar naudojama technologija CHM sukurti, kaip ir skirtingos konstantos kintančio lango funkcijoje. Atliktas viršūnių kiekio apvalymas ir likusiems 7 CHM viršūnių detekcijos rezultatams (3 lentelė). Konstantos kintančio lango funkcijoje visiems pritaikytos skirtingos, nes palikus tas pačias medžių nustatoma per daug arba per mažai.

Pasirinkus fiksuoto lango metodą, kada nėra vertinama esamo pikselio vertė pagal formulę – užtenka parinkti lango dydį. Eksperimentai su keliais skirtingais CHM parodė, jog fiksuoti langai yra stabilesni ir reikalauja mažesnių pastangų derinant funkciją. Generuotiems CHM R aplinkoje, medžių kiekis išlieka panašus, lyginant pagal skirtingą LiDAR skenavimo detalumą, tuo tarpu ArcGIS Pro skirtumai didesni kai yra naudojamas vienodas lango dydis. Tankesnio LiDAR CHM, sudarytame ArcGIS Pro, skirtumas sumažėja iki 66-ių pakoregavus lango dydį į lygiai 6.

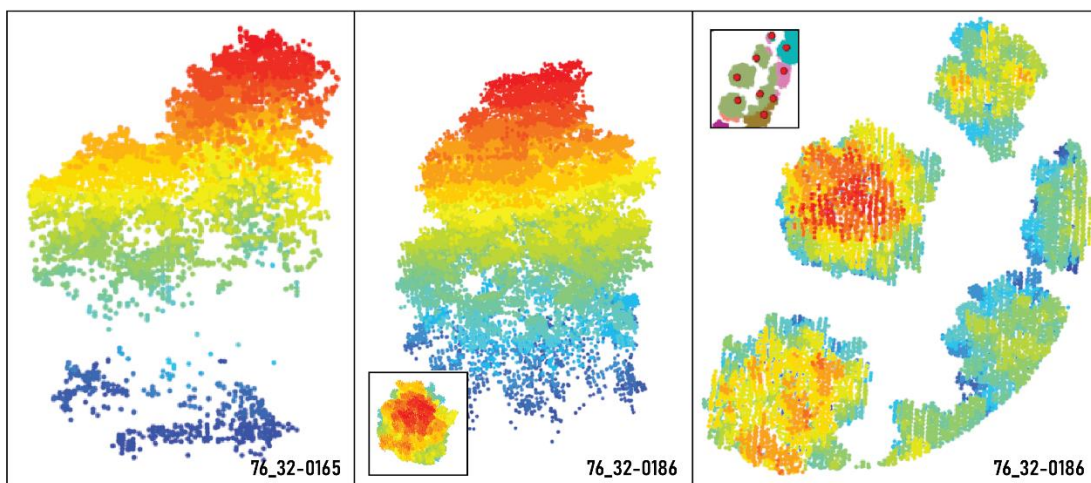
3 lentelė. Medžių viršūnių skirtumai priklausomai nuo rastro raiškos, LiDAR detalumo ir CHM generavimo technologijos.

	ArcGIS Pro				R			
	ForestTools							
CHM raiška	1 m		0,5 m		1 m		0,5 m	
Taškų tankis, tšk./m ²	≥ 45	≥ 6,5	≥ 45	≥ 6,5	≥ 45	≥ 6,5	≥ 45	≥ 6,5
Konstantos	0,3 + 2.5	0.1 + 2	0,08 + 2	0,02 + 1	0,06 + 2	0,05 + 2	0,04 + 2	0,15 + 1
Medžių kiekis								
Prieš apvalymą	391	333	401	477	306	355	374	315
Po apvalymo	284	261	314	353	282	322	339	282
Inventorizuota	324	324	324	324	324	324	324	324
Skirtumas	40	63	10	29	42	2	15	42
lidR								
CHM raiška	0,5 m							
Taškų tankis, tšk./m ²	≥ 45		≥ 6,5		≥ 45		≥ 6,5	
Konstanta	5,5 (fiksotas langas)							
Medžių kiekis								
Prieš apvalymą	529		478		393		354	
Po apvalymo	421		378		346		299	
Inventorizuota	324		324		324		324	
Skirtumas	97		54		22		25	

Tačiau užfiksuotų medžių skaičius yra preliminari metrika. Tenka atminti, jog detalus ir mažiau detalus LiDAR skenavimai atlikti skirtingais metais, tad „ID Vilnius“ duomenyse didesnė tikimybė užfiksuoti mažesnius jaunesnius medžius. Vienas svarbus fiksuoto lango privalumas yra tai, jog nenurodomas minimalus medžio aukštis ir algoritmas gali juos užfiksuoti, tuo tarpu kintančio lango funkcija, naudota ForestTools, nustačius minimalaus aukščio, gali nustatyti daugiau viršūnių, negu tikimasi. Medžių kiekis taip pat gali padidėti dėl užfiksuotų viršūnių ant pastatų artefaktų. Didesnė tikimybė tokių taškų turėti daugiau yra naudojant tankų LiDAR. Kadangi inventorizacijos duomenyse nebūna fiksuoti visi matomi nuotoliniuose vaizduose medžiai, juos gali užfiksuoti algoritmai, tad taikant metodus reikalinga vizuali patikra, jog būtų galima įvertinti nustatytų medžių viršūnių tinkamumą.

3.2 CHM ir taškų debesies metodų medžių segmentavimui palyginimas

lidR nederinant algoritmo labai smulkiai suskaldo vainikus pagal didesnius išsišakojimus. Kaip ir literatūroje rašoma apie šį metodą – gaunama šakos arba medžio viršūnė. Norint gauti didelių medžių vainikus, langas pasidaro labai didelis ir keli medžiai, esantys vienas šalia kito, segmentuojami kaip vienas. Gali pasitaikyti ir atvejų, kada į vieną unikalų medį patenka ir kiti, akivaizdų tarpą tarpusavyje turintys medžiai (15 pav.)



15 pav. lidR segmentuotų medžių įvairovė LiDAR lapuose.

Lyginant ForestTools ir lidR bibliotekų rezultatus, akivaizdžiai geriau atrodo ForestTools medžių vainikai, ypač lyginant su nederintu lidR algoritmu, nors ir poligonai, išlaikę ribas po rastro konversijos, kartais ne visiškai tikslingi (16 pav.). Yra vainikų dalių, patenkančių ant pastatų. Vizualios inspekcijos metu lyginant sudarytų skirtingų CHM išgautus medžių vainikus su ForestTools, nepastebima labai išskirtinų variantų. Tankesnio LiDAR CHM vainikai atitinkamai pagal CHM rastrą yra šiek tiek didesni, kaip ir ArcGIS Pro generuotų CHM plotas ir gauti vainikai didesni, negu R generuotų. Teisingai detaliau ar stambiau išskirti medžiai pasitaiko ir 1 m ir 0,5 m CHM raiškose. Rezultatai nestebina, nes priklauso nuo nustatytų medžių viršūnių. Išorinės ribos iš esmės vienodos, o jeigu yra išskirtas papildomas medis, jis dažniau atsidurs vieno vainiko ribose ir jį padalins. Skirtumai įdomesni lyginant ForestTools kintančio ir lidR fiksuoto lango vainikus. Yra ribų, kurios tarp dviejų skirtingų metodų visiškai sutampa, tačiau bendrai tikslesnės bendros medžių lajos pasiekiamos su kintančiu langu. Fiksuoto lango vainikuose atsiranda kampų ir lygių kraštinių. Šios savybės teigiamu aspektu tampa tai, jog gaminant duomenis mašininio mokymo modeliams rūšių nustatymui, gali patekti mažiau neaktualių pikselių vainikų kraštuose, kurių spektrinė reikšmė gali būti maišyta su žemės paviršiumi ar kita danga. Taip pat fiksuoto lango rezultatai rečiau įtraukia kitų medžių vainikus į vieną. Dėl vainikų užimamo ploto, naudojant fiksuotą langą rezultatų esmė atsikartoja kaip ir naudojant kintantį langą, tad priklauso nuo naudojamo LiDAR ir CHM generavimo. Skirtumai tarp ArcGIS Pro ir R generuotų vainikų mažiau tikėtini turint kuo tankesnę LiDAR, tuo tarpu su mažesnio detalumo LiDAR duomenimis reikalingas didesnis dėmesys kiekvienam procesui.

Segmentuotus medžius tiesiogiai iš taškų debesies pagal lidR rašoma, jog galima juos gauti įvairiomis formomis, tačiau šiuo atveju labiausiai domina vainiko ribos, nurodančios ir vainiko apimamą tūrį (*crown_metrics*). Jos dažnai labai netaisyklingų formų, primenančios klaidas, kurios atsirastų dėl interpoliavimo.



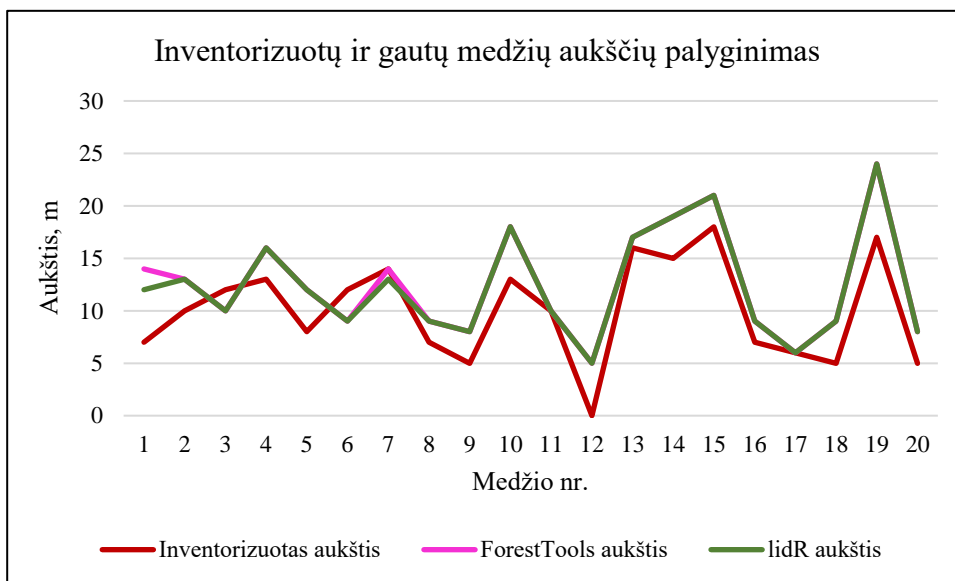
16 pav. Išskirtos medžių vainikų ribos pagal skirtingas R bibliotekas.

lidR vainikai, antraip nei ForestTools, persidengia vieni po kitais dėl didesnių atsišakojimų interpretavimo atskirais medžiais, bet tai nebūtinai reiškia, jog gali būti tinkamai išskiriami smulkūs medžiai po dideliais vainikais. Vietos patikroje buvo įvertintas stambus medis A. Tumėno stovėjimo aukštelėje prie seimo. lidR išskyrė penkis poligonus jo vainiko ribose. Panašų skaičių stambių atsišakojimų galima interpretuoti (~6), nors smulki penkta atšaka tarp poligonų nėra labai stambi, tačiau labai atsikišusi, tad algoritmo buvo interpretuotas dar vienas šakos aukščio maksimumas (17 pav.). Nors lidR pateikia vainikų tūrius, jie, kaip ir vainikų ribos, būtų labai iškreipti daugelyje vietų.

Vis dėlto, lidR *crown_metrics* gali pateikti gražius vainikus ir jų tūrį pavieniams mažesniems medžiams, kurių vainikai dar nėra suformavę sudėtingos struktūros. Galbūt būtų įmanoma vertinti kai kurių medžių išsišakojimą. ForestTools pateikia natūraliai priimtinesnį rezultatų variantą, nors ir vainiko tūrio apskaičiuoti negali, bet iš karto pateikia vainiko plotą, o medžių išskyrimo problematika apibendrinta tyrimo dalyje apie medžių viršūnes.

Želdinių aukštis fiksuojamas sveikais skaičiais. Parinkus atsitiktinių medžių, sudaryta diagrama (18 pav.). ForestTools ir lidR medžių aukščiai pateikiami su dešimtainėmis dalimis. Suapvalinus jų aukščius iš esmės gaunami vienodi rezultatai. Kadangi R bibliotekų sutapimas toks artimas, nesutampantys medžių aukščiai gali būti dėl lidR vainikų fragmentiškumo – parinktas netinkamas vainikas medžiui reprezentuoti.

Nors medžių aukštis turėtų būti tikslesnis juos segmentuojant tiesiai iš taškų debesies, medžių aukščio modelis nenusileidžia savo tikslumu. Daugelio šių medžių atributai (neaišku kurie) buvo redaguoti 2022 ar 2021 metais, tad skirtumai nuo realybės gali priklausyti ir nuo redagavimo datos, ypač jaunesniems medžiams. Į grafiką įtrauktas ir 0-inį aukštį turintis medis, kuris greičiausiai buvo per žemas pirminės inventurizacijos vertinimui arba įsivėlė klaida. Šių duomenų naudojimas



18 pav. Inventorizuotų ir R bibliotekomis gautų medžių aukščių palyginimas



17 pav. Medis aikštelėje ir išskirti šakų aukščio maksimumai kaip atskirų vainikų ribos.

padėtų išvengti klaidų dabartinės inventurizacijos duomenyse, nes aukščiausias užfiksuotas medis mieste 1021 metrų (spėjama, jog tai klaidingo dešimtainių skirtuko naudojimo išdava), o visa aibė medžių turi fiksuotą aukštį virš 100. Pagal tokį didelį kiekį aukštų

medžių daroma prielaida, kad nelogiškas aukštis gali būti fiksuojamas kitais matavimo vienetais, t. y. centimetrais, bet pažvelgus į inventorizuotus taškus akivaizdu, jog ten pasitaiko ir aukštų medžių, tad duomenų įvedimas galimai vyko su trūkumais.

ForestTools (CHM) ir lidR (taškų debesys) turi panašių bei skirtingų problemų segmentuojant medžius, taip pat pateikia skirtingus atributus. Su lidR galima gauti tūrį, o su ForestTools tik lajų plotą, kurį iš esmės būtų galima pasiskaičiuoti ir lidR rezultatams. Fiksuoto lango lidR sprendimas gali pateikti ir tikslių vainikų ir išlygintomis kraštinėmis dėl pačio metodo principo, bet tai apriboja įsikišimą į gretimo medžio lajas, o tai prisideda prie tikslesnės spektrinės informacijos. Taip pat galima užfiksuoti smulkesnius medžius, nenurodant minimalaus aukščio reikalavimo. ForestTools rezultatas puikiai tinka bendrų medžių lajų išskyrimui vektorine forma. Iš taškų debesies rezultatų bei CHM medžių segmentavimo galima gauti patenkinamą aukščio informaciją, tačiau taškų debesies segmentavimu negaunama medžius atitinkančių jų viršūnių. Abi bibliotekos gali būti naudingos pavieniams medžiams, tačiau tankiau apaugusioms teritorijoms reikėtų tobulinti procesą ar ieškoti kitokių metodų.

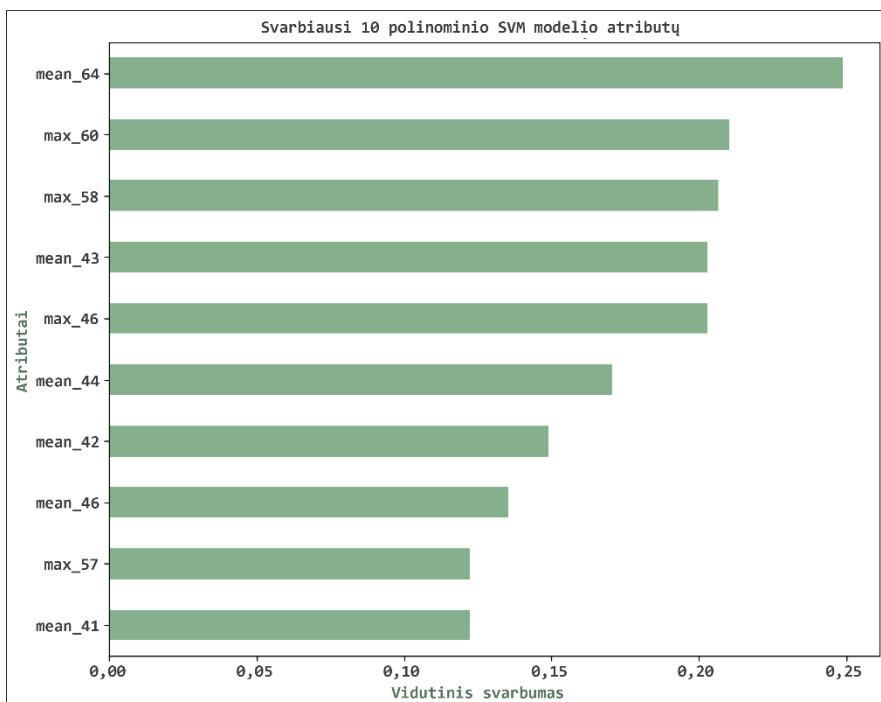
3.3 Medžių rūšių identifikavimo rezultatai

Medžių rūšių nustatymui panaudojus visus zoninės statistikos sugeneruotus duomenis, geriausiu tapo polinominio tipo SVM modelis, pasiekęs 75,26 % tikslumą (4 lentelė). Atmetus didelį nuokrypį turinčius duomenis ir apmokius naują modelį vienodais parametrais su grynesnių savybių objektais, tikslumas nukrito 0,89 % tam pačiam validavimo duomenų rinkiniui.

4 lentelė. Derinti modelių parametrai ir jų tikslumo rezultatai.

Duomenys	SVM modelio parametrai				Tikslumas, %
	Tipas	C	Laipsnis	Coef0	
Pilni	Polinominis	1	3	9	75,26
Apvalyti	Polinominis	1	3	9	74,37
Apmažinti	Polinominis	1	4	5	70,34
Pilni	Tiesinis	1	3	-	69,45
Pilni	Tiesinis	0,01	4	-	73,77
RF modelio parametrai					
	Tipas	Medžių sk. miške	Maksimalus gylis		
Pilni	Gini	110	22		71,38
Pilni	Entropy	110	22		70,23
Pilni	Log loss	100	20		71,23

Validavimo duomenys nebuvo apvalyti, kadangi jie reprezentuoja realius naujus nematytus duomenis. Po duomenų apvalymo su *IsolationForest*, eilučių kiekis buvo sumažėjęs tik 268



19 pav. 10 svarbiausių polinominio SVM modelio atributų.

į vienetą tikslumas tapo 69,45 %. Polinominis SVM modelis taip pat apmokomas greičiau, negu tiesinis. RBF SVM modelis gana dažnas pasirinkimas šiai užduočiai, bet jo efektyvumas šiuo atveju taip pat nesiekė 70 % tikslumo.

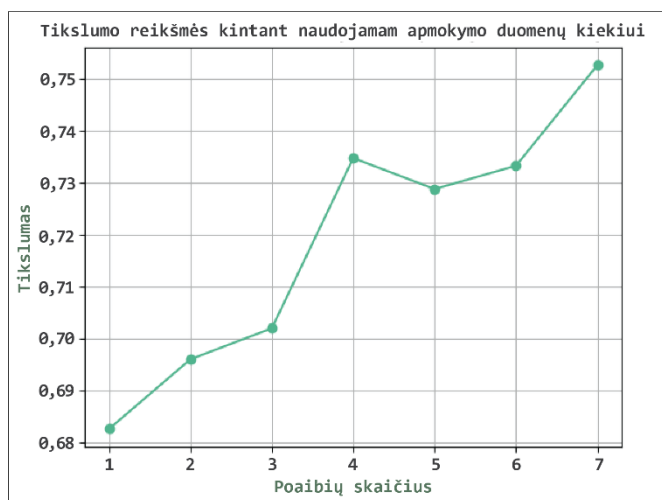
Pagal nagrinėtus spektrinius profilius, kanaluose nuo 1 iki maždaug 37 skirtumai menkesni. Nors SVM gerai dorojasi su aukštu dimensijų kiekiu, išbandyta minėtus kanalus atmesti iš analizės. Polinominio SVM tikslumas nukrito iki 70,34, tad trumpesnės spektro bangos, panašu, neša vertę. Atributų svarba paprastai pamatuojama, jeigu naudojamas tiesinis SVM modelis. Jo svarbiausiomis savybėmis išskirtinai tapo 41 ir 39 kanalų vidurkiai, o trečioje vietoje maksimali pirmo kanalo, reprezentuojančio medžių aukščio modelį (bei atitinkamai medžių aukštį), reikšmė. Po jos seka 2-o kanalo vidurkis bei 15 kanalo minimali reikšmė. Tačiau polinominio SVM atveju tenka naudoti permutacijos svarbą (angl. *permutation importance*), kada sumaišomos laukų reikšmės ir stebima, kaip keičiasi modelio tikslumas. Maksimalios ir vidutinės infraraudonojo spektro dalies kanalų reikšmės turi didelę svarbą pagal permutacijos analizę (19 pav.). Teigiamos reikšmės indikuoja, jog maišant lauko reikšmes tikslumas mažėja, o neigiamos – jog lauko reikšmių sumaišymas prisideda prie modelio tikslumo. Šiuo atveju neigiamos reikšmės pasiskirsto daugiau per kanalų minimalius ir standartinio nuokrypio laukus. Neigiamų permutacijos svarbumo vidurkio reikšmių yra žymiai mažiau, negu teigiamų – atitinkamai 75 ir 185. Neigiamos reikšmės labai mažos (didžiausia neigiama reikšmė siekia tūkstantines dalis, o didžiausia teigiama siekia

vienetų. Palyginus su kito tipo modeliais, potencialą parodė ir tiesinis SVM, pasiekęs 73,77 % tikslumą, kurį buvo galima gerinti su C parametru, tačiau kadangi jis nustato baudą už mokymo metu neteisingai suklasifikuotus duomenis ir mažina modelio perspaudimą, šio parametro sumažinimas gali sumažinti modelio efektyvumą naujų duomenų atveju. Minėtas tikslumas nebūtų pasiektas, jeigu C nebūtų lygi 0,01, tačiau sugrąžinus

0,25), todėl turi gana mažą įtaką, kaip ir reikšmės arti nulio. Neigiamos reikšmės atsiranda dėl originalių klaidinančių objektų savybių ar modelio perspaudimo dėl triukšmo duomenyse.

Išbandytas ir *Random Forest* (liet. atsitiktinių miškų, toliau – RF) klasifikatorius, naudojant visus atributus. Yra keletas pasirenkamų kriterijų, aprašančių kaip skaidomi sprendimų medžiai. Iš *scikit-learn* pateiktų trijų variantų – du pateikia panašias bendro tikslumo reikšmes (4 lentelė). Vis dėlto, daugelis rezultatų ir su RF ir SVM siekia maždaug 71 % tikslumą, kurį yra galimybė pagerinti toliau derinant modelį.

Panagrinėjus geriausio modelio klasifikacijos ataskaitą akivaizdu, jog geriausius rezultatus pateikia medžių rūšys, kurių buvo daugiausia (mažalapė liepa ir paprastasis klevas), o panašaus kiekio rūšių rezultatai artimesni vieni kitiems (5 lentelė). Liepos sudaro labai didelę dalį visų medžių, įskaitant ir šešėlių paveiktose vietose. Modeliui gali turėti įtakos ir neišbalansuotos klasės. Jeigu vienos klasės yra labai daug, modelis gali linkti dažniau priskirsti medį būtent tai rūšiai. Kaštonai, išsidėstę intensyvaus šešėlio teritorijoje, dažniau priskiriami liepoms, negu tikrajai klasei.



20 pav. Modelio tikslumo pokytis didėjant imties kiekiui.

5 lentelė. Polinominio SVM medžių rūšių klasifikacijos ataskaita

Rūšis	Tikslumas	Recall	F1-score	Klasės kiekis rinkinyje
Karpotasis beržas	0,56	0,41	0,47	34
Paprastasis kaštonas	0,58	0,32	0,41	56
Paprastasis klevas	0,75	0,65	0,69	127
Uosialapis klevas	0,63	0,33	0,43	52
Mažalapė liepa	0,78	0,93	0,85	402

Jeigu būtų išskiriamos gentys (klasifikuojama lengivau) ir nedetalizuojama rūšis, iš 179 klevų validavimo imtyje, 117 būtų klasifikuoti teisingai. Tai šiek tiek pagerintų bendrą tikslumą, tačiau kadangi uosialapių klevų labai mažai, pokytis būtų menkas ir esmę sudarytų paprastieji klevai. Iš imties dydžio įtakos modelio tikslumui analizės (20 pav.) akivaizdu, kad modeliui praverstų didesnis kiekis medžių apmokymui, tačiau būtų galima ir papildyti savybių kiekį bei jomis pakeisti mažiau naudingus atributus.

DISKUSIJA

Didelis dėmesys tyrime skirtas didiesiems miestams, tačiau išnagrinėjus CHM generavimo technologijas bei tam naudojamus skirtingus LiDAR tankius, galima teigti, jog Lietuvos lazerinio skenavimo duomenis būtų galima naudoti ir mažiau urbanizuotose teritorijose želdynų inventorizacijos darbams pagal šią metodiką. Medžių išskyrimas labai priklauso nuo pradinių duomenų kokybės. Reikėtų skirti daugiau laiko LiDAR perklasifikavimui, jog rezultatai turėtų mažiau išlikusių artefaktų. Jų šiek tiek mažiau pasitaikytų naudojant mažesnio tankumo lazerinio skenavimo duomenis. Rezultatą taip pat galima koreguoti ir po medžių išskyrimo, tačiau tam reikėtų vizualaus patikrinimo per visą teritoriją atsižvelgiant į papildomas duomenis arba atlikti papildomas analizes. Kaip minėta rezultatų skyriuje, tokius objektus, kaip šviestuvai, būtų galima bandyti šalinti pagal vainiko ploto ir jam priskirto aukščio santykį.

Rastro raiška (1 m arba 0,5 m) kiekybiškai turi mažą įtaką išskiriamų medžių rezultatui, palyginus su naudojamais parametrais – ypatinai lango funkcija. Medžių rūšims nustatyti dažnai kuriamos šešėlių kaukės ar naudojami kiti metodai, jog būtų galima gauti grynesnes spektrines savybes ir atsikratyti prastų pikselių vainikų objektuose. Tačiau jeigu medis pilnai patenka į šešėlį, jis yra nebeįvertinamas, o tai realūs duomenys, kurie ne visada ar ne visoje erdvėje būna idealūs. Tokių idealių duomenų surinkimas apsunkintų ir prailgintų procesą. Fiksuoto lango sprendimas padeda atpažinti ir smulkius medelius ir gauti šiek tiek grynesnius spektro duomenis klasifikavimui objektų pagrindu, tačiau smulkių vainikų medžiai taip pat gali likti neįvertinti, jeigu sukurtos šešėlių kaukės pašalina visus jo pikselius.

Medžių aukščio modelio maksimalių reikšmių svarba tiesiniam SVM modeliui netikėta. Medžiai gali būti susodinami skirtingu metu ir ta pati rūšis gali pasiekti didelę brandos variaciją. Dirvožemis ir kitos lokalsios aplinkos sąlygos kaip mikroklimatas ar želdinio priežiūra taip pat gali turėti įtakos medžio augimo spartai ar sveikatingumui. Panašu, jog rūšies viduje variacija nebuvo didelė ar bent jau nepakankama. Verta išbandyti modeliuose panaudoti apskaičiuotus augalijos indeksus, o kadangi hiperspektrinis kanalas fiksuoja mažesnę kiekį šviesos, būtų galime stebėti ir keletos apjungtų kanalų augalijos indeksų reikšmingumą rūšių identifikavimo modeliui, nes hiperspektru galima susikurti dirbtinius multispektrinius duomenis bei taip padidinti panaudojamą savybių kiekį. Yra daug galimybių tyrinėti indeksus bei iš LiDAR duomenų gaunamas savybes kaip intensyvumas.

IŠVADOS

1. R bibliotekų pagalba iš skirtingo tankio LiDAR duomenų galima nustatyti medžių lokaciją, aukštį, ir vainiko apimamą plotą, tačiau nuokrypia neišvengiami ir reikalingas vizualinis vertinimas. Priklausomai nuo parametrų tikslumas kinta stambesnių ar smulkesnių medžių sąskaita, tačiau galimi naudoti ir mažiau tankūs lazerinio skenavimo duomenys.
2. Fizinių parametrų nustatymas R bibliotekomis pasinaudojant medžių aukščio modeliu yra pranašesnis, nei tiesioginio taškų debesies naudojimas, kuriam reikalingi detalūs duomenys ir sudėtingesni algoritmai, negu lidR siūlomas sprendimas.
3. Automatizuojant duomenų paruošimą mašiniam mokymui prarandama apie pusę duomenų, tačiau tai pagreitina analizę.
4. Su polinominiu SVM algoritmu pasiektas didžiausias penkių rūšių identifikavimo tikslumas – 75,26%, kuris turi potencialo pagerėti padidėjus apmokomųjų duomenų kiekiui.

NAUDOTA LITERATŪRA

1. Almaliki, Z. A. (2018). „Parametric vs Nonparametric models“, *Medium* [internetė]. Prieiga internetu: <https://dataakkadian.medium.com/what-are-parametric-vs-nonparametric-models-8bfa20726f4d> (Žiūrėta: 09-03-2024)
2. Bryson, M., Wang, F., ir Allworth, J. (2023). Using Synthetic Tree Data in Deep Learning-Based Tree Segmentation Using LiDAR Point Clouds. *Remote Sensing*, 15(9), 2380. <https://doi.org/10.3390/rs15092380>
3. Brownlee, J. (2020). „Parametric and Nonparametric Machine Learning Algorithms“, *Machine Learning Mastery*. Prieiga internetu: <https://machinelearningmastery.com/parametric-and-nonparametric-machine-learning-algorithms/> (Žiūrėta: 09-03-2024)
4. Carleer, A., & Wolff, E. (2004). Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing*, 70(1), 135-140. <https://doi.org/10.14358/PERS.70.1.135>
5. Chehreh, B., Moutinho, A., ir Viegas, C. (2023). Latest Trends on Tree Classification and Segmentation Using UAV Data—A Review of Agroforestry Applications. *Remote Sensing*, 15(9), 2263. <https://doi.org/10.3390/rs15092263>
6. Chen, L., Chiang, T., ir Teo, T. (2005, November). Fusion of LIDAR data and high resolution images for forest canopy modelling. In *Proc. 26th Asian Conference on Remote Sensing*.
7. Chen, X., Wang, R., Shi, W., Li, X., Zhu, X., & Wang, X. (2023). An Individual Tree Segmentation Method That Combines LiDAR Data and Spectral Imagery. *Forests*, 14(5), 1009. <https://doi.org/10.3390/f14051009>
8. City of Melbourne. (2020). *Explore Melbourne's Urban Forest*. Prieiga internetu: [Melbourne Urban Forest - Visual \(melbourneurbanforestvisual.com.au\)](https://melbourneurbanforestvisual.com.au) (Žiūrėta: 08-03-2023)
9. Dian, Y., Pang, Y., Dong, Y., & Li, Z. (2016). Urban tree species mapping using airborne LiDAR and hyperspectral data. *Journal of the Indian Society of Remote Sensing*, 44, 595-603. <https://doi.org/10.1007/s12524-015-0543-4>
10. Dougherty, E. R., ir Lotufo, R. A. (2003). *Hands-on morphological image processing* (Vol. 59). SPIE press, <https://doi.org/10.1117/3.501104>
11. eoPortal. (2023). Hyperspectral Imaging. Prieiga internetu: <https://www.eoportal.org/other-space-activities/hyperspectral-imaging#references> (Žiūrėta: 11-05-2024)
12. Eurocities. (2023). EU Nature Restoration Law. Eurocities reaction on the proposal to restore nature. Prieiga internetu: [Title \(eurocities.eu\)](https://eurocities.eu) (Žiūrėta: 12-06-2023)
13. Europos aplinkos agentūra. (2022). Percentage of total green infrastructure, urban green space, and urban tree cover in the area of EEA-38 capital cities (excluding Liechtenstein). Prieiga internetu: [Percentage of total green infrastructure, urban green space, and urban tree cover in the area of EEA-38 capital cities \(excluding Liechtenstein\) — European Environment Agency \(europa.eu\)](https://environment.ec.europa.eu/eu-environmental-statistics/percentage-of-total-green-infrastructure-urban-green-space-and-urban-tree-cover-in-the-area-of-eea-38-capital-cities-excluding-liechtenstein) (Žiūrėta: 12-06-2023)

14. Evidently AI. (2024). Accuracy vs. precision vs. recall in machine learning: what's the difference? Prieiga internetu: <https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall> (Žiūrėta: 11-06-2024)
15. Ferreira, M. P., Grondona, A. E. B., Rolim, S. B. A., & Shimabukuro, Y. E. (2013). Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling. *Journal of Applied Remote Sensing*, 7(1), 073502-073502. <https://doi.org/10.1117/1.JRS.7.073502>
16. Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26, 49-63. <https://doi.org/10.1016/j.jag.2013.05.017>
17. Hartling, S., Sagan, V., & Maimaitijiang, M. (2021). Urban tree species classification using UAV-based multi-sensor data fusion and machine learning. *GIScience & Remote Sensing*, 58(8), 1250-1275. <https://doi.org/10.1080/15481603.2021.1974275>
18. Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., & Atzberger, C. (2019). Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. *Remote Sensing*, 11(22), 2599. <https://doi.org/10.3390/rs11222599>
19. Jaadi, Z. (2024). A Step-by-Step Explanation of Principal Component Analysis (PCA). Prieiga internetu: <https://builtin.com/data-science/step-step-explanation-principal-component-analysis> (Žiūrėta: 11-05-2024)
20. Jakubowski, M., Li, W., Guo, Q., ir Kelly, M. (2013). Delineating Individual Trees from Lidar Data: A Comparison of Vector- and Raster-based Segmentation Approaches. *Remote Sensing*, 5(9), 4163–4186. <https://doi.org/10.3390/rs5094163>
21. Jensen, R. R., Hardin, P. J., & Hardin, A. J. (2012). Classification of urban tree species using hyperspectral imagery. *Geocarto International*, 27(5), 443-458.
22. Kwak, D. A., Lee, W. K., Lee, J. H., Biging, G. S., ir Gong, P. (2007). Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, 12(6), 425-434, <https://doi.org/10.1007/s10310-007-0041-9>
23. Li, W., Guo, Q., Jakubowski, M. K., ir Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering and Remote Sensing*, 78(1), 75-84. <https://doi.org/10.14358/PERS.78.1.75>
24. Lidar_DR_LT - skaitmeniniai erdviniai lazerinio skenavimo taškų duomenys (2019-2022 m.) © Nacionalinė žemės tarnyba prie Aplinkos ministerijos, 2024
25. Lietuvos arboristikos centras. (2024). *Želdiniai ir jų įtaka aplinkai*. Prieiga internetu: <https://www.arboristai.lt/articles/poveikis-zmogui/205-straipsniai22/1160-zeldiniai-ir-j-itaka-aplinkai> (Žiūrėta: 24-02-2024)
26. Lietuvos hidrometeorologijos tarnyba. (datos nėra). Klimato kaita Vilniuje XXI a. Prieiga internetu: [Klimato kaita Vilniuje XXI a. - Meteo.lt](https://www.meteo.lt/klimato-kaita-vilniuje-xxi-a) (Žiūrėta: 12-06-2023)
27. Lietuvos Respublikos Seimas. (2007). Lietuvos Respublikos želdynų įstatymas (2007, birželio 28 d., Nr. X-1241). Prieiga internetu: [X-1241 Lietuvos Respublikos želdynų įstatymas \(lrs.lt\)](https://www.lrs.lt/litdoc/lrsdoc.nsf/(open)/X-1241) (Žiūrėta: 23-02-2024)

28. Lietuvos Respublikos Seimas. (2008). Dėl Želdynų ir želdinių inventorizavimo ir apskaitos taisyklių patvirtinimo (2008, sausio 8 d., Nr. Nr. D1-5). Prieiga internetu: [D1-5 Dėl Želdynų ir želdinių inventorizavimo ir apskaitos taisyklių patvirtinimo \(lrs.lt\)](https://www.lrs.lt/doc/lrs/11-05-2023) (Žiūrėta: 11-05-2023)
29. Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, 200, 170-182. <https://doi.org/10.1016/j.rse.2017.08.010>
30. Ma, K., Xiong, Y., Jiang, F., Chen, S., ir Sun, H. (2021). A Novel Vegetation Point Cloud Density Tree-Segmentation Model for Overlapping Crowns Using UAV LiDAR. *Remote Sensing*, 13(8), 1442. <https://doi.org/10.3390/rs13081442>
31. Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., ... & Vihervaara, P. (2021). Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, 256, 112322. <https://doi.org/10.1016/j.rse.2021.112322>
32. Marrs, J., & Ni-Meister, W. (2019). Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sensing*, 11(7), 819. <https://doi.org/10.3390/rs11070819>
33. Michałowska, M., & Rapiński, J. (2021). A review of tree species classification based on airborne LiDAR data and applied classifiers. *Remote Sensing*, 13(3), 353. <https://doi.org/10.3390/rs13030353>
34. Mielcarek, M., Stereńczak, K., ir Khosravipour, A. (2018). Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation. *International journal of applied earth observation and geoinformation*, 71, 132-143. <https://doi.org/10.1016/j.jag.2018.05.002>
35. Mozgeris, G., Juodkienė, V., Jonikavičius, D., Straigyte, L., Gadala, S., & Ouerghemmi, W. (2018). Ultra-light aircraft-based hyperspectral and colour-infrared imaging to identify deciduous tree species in an urban environment. *Remote sensing*, 10(10), 1668. <https://doi.org/10.3390/rs10101668>
36. National Oceanic and Atmospheric Administration. (2023). What is lidar? Prieiga internetu: <https://oceanservice.noaa.gov/facts/lidar.html> (Žiūrėta: 11-05-2024)
37. Plowright, A., Roussel, J. (2021). ForestTools: Tools for Analyzing Remote Sensing Forest Data. R package version 0.2.5, <https://CRAN.R-project.org/package=ForestTools> (Žiūrėta: 13-01-2023)
38. Qiao, Y., Zheng, G., Du, Z., Ma, X., Li, J., ir Moskal, L. M. (2023). Tree-Species Classification and Individual-Tree-Biomass Model Construction Based on Hyperspectral and LiDAR Data. *Remote Sensing*, 15(5), 1341. <https://doi.org/10.3390/rs15051341>
39. Qin, Y., Ferraz, A., Mallet, C., ir Iovan, C. (2014, July). Individual tree segmentation over large areas using airborne LiDAR point cloud and very high resolution optical imagery. In 2014 IEEE Geoscience and Remote Sensing Symposium (pp. 800-803). IEEE. [10.1109/IGARSS.2014.6946545](https://doi.org/10.1109/IGARSS.2014.6946545)

40. Roussel J. R., Auty, D. (2022). Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 4.0.2. <https://cran.r-project.org/package=lidR> (Žiūrēta: 13-01-2023)
41. Roussel, J.R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Sánchez Meador, A., Bourdon, J.F., De Boissieu, F., Achim, A. (2020). lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, 251 (August), 112061. doi:10.1016/j.rse.2020.112061
42. Scikit learn. (2024a). SVC. Prieiga internetu: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html> (Žiūrēta: 11-05-2024)
43. Scikit learn. (2024b). RandomForestClassifier. Prieiga internetu: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (Žiūrēta: 11-05-2024)
44. Spaudos departamentas. (2024). DUK – pamokēlēs pradedantiesiems. Prieiga internetu: <https://www.spaudosdepartamentas.lt/pamokeles-pradedantiesiems/> (Žiūrēta: 11-05-2024)
45. Trier, Ø. D., Salberg, A. B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., & Aarsten, D. (2018). Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. *European Journal of Remote Sensing*, 51(1), 336-351. <https://doi.org/10.1080/22797254.2018.1434424>
46. Wan, H., Tang, Y., Jing, L., Li, H., Qiu, F., ir Wu, W. (2021). Tree Species Classification of Forest Stands Using Multisource Remote Sensing Data. *Remote Sensing*, 13(1), 144. <https://doi.org/10.3390/rs13010144>
47. Wang, K., Wang, T., & Liu, X. (2018). A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment. *Forests*, 10(1), 1. <https://doi.org/10.3390/f10010001>
48. Wasser, L. A. (2024). What is a CHM, DSM and DTM? About Gridded, Raster LiDAR Data. Prieiga internetu: <https://www.neonscience.org/resources/learning-hub/tutorials/chm-dsm-dtm> (Žiūrēta: 11-05-2024)
49. Wood., T. (2024). What is the F-score? Prieiga internetu: <https://deepai.org/machine-learning-glossary-and-terms/f-score> (Žiūrēta: 11-05-2024)
50. Zhang, C., Zhou, Y., ir Qiu, F. (2015). Individual Tree Segmentation from LiDAR Point Clouds for Urban Forest Inventory. *Remote Sensing*, 7(6), 7892–7913. <https://doi.org/10.3390/rs70607892>

SANTRAUKA

VILNIAUS UNIVERSITETAS CHEMIJOS IR GEOMOKSLŲ FAKULTETAS

VAIVA VENCKAUSKAITĖ

Medžių inventorizavimo urbanizuotose teritorijose galimybės taikant nuotolinius metodus

Europos komisija priėmė gamtos atkūrimo įstatymą, pagal kurį žaliosios erdvės turėtų plėstis iki patenkinamo lygmens, ir šių erdvių bei medžių lajų dengiamas plotas neturėtų sumažėti iki 2030 metų, lyginant nuo reglamento įsigaliojimo. Lietuvos Respublikoje želdinių ir želdynų inventorizavimas yra reglamentuojamas, įstatymo pokytis leidžia darbus atlikti ir nuotoliniu būdu. Pasitelkiant žmogiškuosius išteklius padengti didelį miestą inventorizacijos duomenimis brangu, užtrunka ilgą laiką, o darbuotojų trūksta, todėl šiame darbe siekiama įvertinti kaip galima panaudoti nuotolinius duomenis miestų želdinių inventorizavimui.

Kadangi tokie miestai kaip Vilnius gali sau leisti atnaujinti nuotoliu renkamus duomenis kaip LiDAR ar ortofoto ir juos pernaudoti, nagrinėjama fizinių medžių parametrų išskyrimo galimybė skirtingais metodais, naudojantis LiDAR ir R programavimo kalbos bibliotekas, bei medžių rūšies identifikavimas hiperspektrinių vaizdų pagalba. Tyrime lyginami du medžių fizinių parametrų išskyrimo metodai, reikalingas duomenų detalumas bei naudojama programinė įranga. Darbų automatizavimui pasitelkiami R ir Python programavimo kalbų sprendimai. Išrinkus patikimesnį metodą medžių parametrų išskyrimui, parenkami kuo optimaliausi parametrai didesnei imčiai apdoroti. Gauti duomenys buvo tinkami klasifikacijai objektų pagrindu. Jos metu pasirinkta nustatyti penkias dažniausiai pasitaikančias Kauno miesto medžių rūšis, turimo hiperspektrinio vaizdo aprėptyje.

Medžių viršūnės ir vainikai išskiriami patikimiau naudojant medžių aukščio modelį, negu tiesiogiai segmentuojant taškų debesį. Ruošiant duomenis mašininiam mokymui, galimų panaudoti medžių sumažėjo apie du kartus, tačiau tai paspartino procesą. Geriausias pasiektas tikslumas identifikuojant rūšis polinominiu SVM algoritmu – 75,26 %. Duomenų analizė parodė, jog rezultatai turėtų tapti geresni padidinus medžių kiekį modelio apmokymui, tačiau dar yra daug vietos nagrinėti papildomų savybių, išgaunamų iš spektrinių vaizdų, reikšmingumą modelio tobulinimui.

SUMMARY

VILNIUS UNIVERSITY FACULTY OF CHEMISTRY AND GEOSCIENCES

VAIVA VENCKAUSKAITĖ

Assessment of tree inventory possibilities in urban areas using remote sensing

The European Commission has adopted a nature restoration law that requires green spaces to expand to a satisfactory level, and the area covered by these spaces as well as tree canopy should not decrease by 2030, compared to the year the regulation came into force. In the Republic of Lithuania, the inventory of trees and green areas is regulated and the change in the law allows the work to be done with remote methods and the municipalities can choose their preferred method. It is expensive and time-consuming to cover a large city with inventory data using human resources, hence this paper attempts to assess how remote sensing data can be used for inventorying urban trees.

Since cities like Vilnius can afford to update and reuse remotely collected data like LiDAR or orthophotos, the possibility of extracting physical tree parameters by different methods using LiDAR and R programming language libraries, and tree species identification using hyperspectral imagery is explored. The study compares the two methods for extracting physical parameters of trees, the required data granularity and the software used. R and Python programming language solutions are used to automate parts the work. Once the more robust method for tree parameter extraction has been chosen, the most optimal parameters are selected to process a larger sample. The data obtained were suitable for object-based classification. It was chosen to identify the five most frequent tree species in Kaunas city within the available hyperspectral image coverage.

Tree tops and crowns are extracted more reliably using the tree height model than by directly segmenting the point cloud. Preparing the data for machine learning reduces the number of trees that can be used by about a factor of two, but it speeds up the process. The best accuracy achieved was 75,26% for species identification using the polynomial SVM algorithm. The data analysis showed that the results should improve with an increase in the number of trees to train the model, but there is still room to investigate the relevance of the additional features extracted from the spectral images to improve the model.

BAIGIAMOJO MAGISTRO DARBO VERTINIMO LAPAS

Darbo autorius: Vaiva Venckauskaitė
(vardas, pavardė) (parašas)

Mokslinis darbo
vadovas: Doc., Dr. Linas Bevainis
(mokslinis laipsnis, mokslinis vardas, vardas, pavardė) (parašas)

Recenzentas:
(mokslinis laipsnis, mokslinis vardas, vardas, pavardė) (parašas)

Kartografijos ir
geoinformatikos
katedros vadovas:
(mokslinis laipsnis, mokslinis vardas, vardas, pavardė) (parašas)

Darbo gynimo data:

Darbo įvertinimas:
(balas skaičiumi, balas raštu)

Baigiamųjų darbų gynimo
komisijos pirmininkas:
(mokslinis laipsnis, mokslinis vardas, vardas, pavardė) (parašas)

Baigiamųjų darbų gynimo
komisijos sekretorius:
(vardas, pavardė) (parašas)