

**UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA
V KOŠICIACH**

PRÍRODOVEDECKÁ FAKULTA

**MAPOVANIE DRUHOVÉHO ZLOŽENIA STROMOV
V MESTE Z DÁT DPZ METÓDAMI STROJOVÉHO
UČENIA**

2024

Peter MICHALÍK

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH

PRÍRODOVEDECKÁ FAKULTA

**MAPOVANIE DRUHOVÉHO ZLOŽENIA STROMOV
V MESTE Z DÁT DPZ METÓDAMI STROJOVÉHO UČENIA**

BAKALÁRSKA PRÁCA

Študijný program: Geografia a geoinformatika

Študijný odbor: Vedy o Zemi

Vedúci práce: doc. Mgr. Michal Gallay, PhD.

Školiace pracovisko: Ústav geografie, Prírodovedecká fakulta UPJŠ v Košiciach

Košice 2024

Peter MICHALÍK

Zadanie záverečnej práce



Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Peter Michalík
Študijný program: geografia a geoinformatika (jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: Vedy o Zemi
Typ záverečnej práce: Bakalárska práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Mapovanie druhového zloženia stromov v meste z dát DPZ metódami strojového učenia

Názov EN: Mapping tree species composition in urban areas using remote sensing data and machine learning methods

Cieľ: Stromy v meste plnia viacero ekologických úloh a za účelom vhodnej starostlivosti o ne je dôležité poznať ich botanický druh. Cieľom bakalárskej práce bude využitie dostupných dát z leteckého laserového skenovania a fotogrametrie na príklade mesta Košice, pričom sa predpokladá použitie metód strojového učenia a porovnanie s reálnym druhovým zložením. Všetky potrebné dáta sú dostupné na Ústave geografie.

Literatúra: Martins, G.B., Cué La Rosa, L.E., Happ, P.N., Filho, L., Santos, C., Feitosa, R., Ferreira, M. 2023. Deep learning-based tree species mapping in a highly diverse tropical urban setting, *Urban Forestry & Urban Greening*, 64, 127241.

Goga, T., Bobál'ová, H., Sačkov, I., Kopecká, M. 2019. Klasifikácia poškodenia lesa vo veľkej mierke na báze leteckých multispektrálnych snímok a lidarových dát – prípadová štúdia CHKO Dunajské luhy. *Geografický časopis*, 71(1), 51-71.

Cherlinka, V., Dmytruk, Y., Cherlinka, L., Gunchak, M., Sobko V. 2022. Methods of modeling and mapping of the soil bulk density: a case study from Chernivtsi region, Ukraine. *Geographia Cassoviensis*, 16(2), 147-163.

Vedúci: doc. Mgr. Michal Gallay, PhD.
Konzultant: doc. Vasyl Cherlinka, DrSc.
Ústav : ÚGE - Ústav geografie
Riaditeľ ústavu: doc. Mgr. Ladislav Novotný, PhD.

Dátum schválenia: 15.05.2024

„Čestne vyhlasujem, že celú bakalársku prácu na tému „Mapovanie druhového zloženia stromov v meste z dát DPZ metódami strojového učenia“ som vypracoval samostatne, a to s použitím literatúry uvedenej v priloženom zozname.

V Košiciach, 22.6.2024

.....

Peter Michalík

Pod'akovanie

Rád by som sa pod'akoval doc. Gallayovi za odborné rady poskytnuté počas písania práce a Mgr. Martinovi za asistenciu pri práci s rastrovými dátami.

Abstrakt

Táto bakalárska práca sa venuje určeniu druhového zloženia stromov pomocou dát z Diaľkového prieskumu Zeme (DPZ) a s použitím metód strojového učenia. Jedná sa o problematiku, ktorej riešenie má za úlohu zefektívniť a zlepšiť kvalitu mapovania druhového zloženia stromov na väčších územiach. V rámci práce sú použité ako vstupné dáta multispektrálne snímky CIR (Color infrared) a trénovacie polygóny stromov zo štyroch lokalít v rámci mesta Košice. Keďže každý druh stromu má iné špecifické vlastnosti odrazu svetla, multispektrálne snímky nám na základe týchto vlastností dokážu pomôcť určiť druhové zloženie stromov. V práci sme sa zaoberali územím 2x2 kilometre v rámci mesta Košice, kde je v mnohých častiach mesta pestrá druhová rozmanitosť stromov. Na základe poskytnutých dát Ústavom geografie sme vykonali potrebné korekcie a prípravu dát na následné spracovanie. Následné dáta sme klasifikovali použitím riadenej klasifikácie pomocou softvéru ArcGIS pro s rôznou výslednou presnosťou.

Kľúčové slová: diaľkový prieskum Zeme, multispektrálne snímky, trénovacie polygóny, ArcGIS pro, klasifikácia, druhové zloženie stromov, strojové učenie

Abstract

This bachelor's thesis focuses on determining tree species composition using remote sensing data (RS) and machine learning methods. The objective of this research is to enhance the efficiency and quality of tree species mapping over larger areas. The input data utilized in this study include multispectral CIR (Color Infrared) images and training polygons of trees from four locations within the city of Košice. Since each tree species has distinct light reflectance characteristics, multispectral images can help us determine tree species composition based on these properties. The study area encompasses a 2x2 kilometer region within the city of Košice, which features a diverse range of tree species in many parts of the city. Using data provided by the Institute of Geography, we performed the necessary corrections and data preparation for subsequent processing. The data were then classified using supervised classification in ArcGIS Pro software, resulting in varying levels of accuracy.

Keywords: remote sensing, multispectral images, training polygons, ArcGIS Pro, classification, tree species composition, machine learning

Obsah

Úvod.....	11
1.1 Diaľkový prieskum zeme	12
1.2 Mapovanie druhového zloženia stromov v meste.....	13
1.3 Letecké laserové skenovanie (LiDAR)	13
1.4 DSM, DTM, DEM A CHM	15
1.4.1 DSM	15
1.4.2 DTM.....	16
1.4.3 DEM.....	17
1.4.4 CHM.....	18
1.5 Fotogrametria	19
1.6 Strojové učenie	20
1.7 Hlboké učenie	22
1.8 Klasifikácia obrazu.....	23
1.9 Riadená klasifikácia.....	23
1.10 Neriadená klasifikácia.....	23
2. Zber a spracovanie dát	24
2.1. Záujmové územie a jeho charakteristika	26
2.2.Úprava dát pre potreby práce.....	26
2.3. Použitie riadenej klasifikácie.....	30
3. Výsledky	
3.1. Kvalita a vlastnosti výsledných dát.....	33
3.2. Identifikácia jednotlivých druhov stromov.....	35
3.3. Identifikácia druhového zloženia stromov v častiach mesta.....	39
3.4. Výpočet presnosti klasifikácie.....	40
Diskusia.....	41
Záver.....	42
Zoznam použitej literatúry.....	44

Zoznam obrázkov

Obr. 1. Základný princíp získavania dát pomocou leteckého laserového skenovania (LiDAR) Zdroj: Gallay, 2013.....	14
Obr. 2. Príklad DSM snímky na skúmanom území, zdroj: Vlastné spracovanie.....	15
Obr. 3. Príklad DTM snímky na skúmanom území, zdroj: Vlastné spracovanie.....	16
Obr. 4. Rozdiel medzi DSM a DTM dátami, zdroj:JOUAV, 2023.....	17
Obr. 5. Znáznornenie modelu DEM. Zdroj: GISGeography, 2024.....	17
Obr. 6. Výpočet a znázornenie modelu CHM. Zdroj: Stackexchange, 2020.....	18
Obr. 7. Fotografia UAV s kamerou určenou na snímkovanie územia. Zdroj:JOUAV, 2023.....	20
Obr. 8. Časť záujmového územia v RGB kompozícii.....	25
Obr. 9. Časť záujmového územia v CIR kompozícii.....	25
Obr.10. Farebné rozdelenie druhov stromov v tréningových polygónoch.....	28
Obr.11. tréningové polygóny v oblasti Hvozdíkov park.....	28
Obr.12. tréningové polygóny v oblasti Moyzesová ulica.....	29
Obr.13. tréningové polygóny v oblasti spievajúcej fontány na Hlavnej ulici.....	29
Obr.14. tréningové polygóny v oblasti Mestského parku pri vlakovej stanici.....	30
Obr. 15. Pripravené polygóny druhov stromov na klasifikáciu.....	31
Obr. 16. nastavenie hodnôt klasifikátora SVM.....	32
Obr. 17. výsledný polygón riadenej klasifikácie.....	33
Obr.18. nastavenia klipovania klasifikovaného rastra.....	34
Obr.19. rozdelené druhy stromov na základe farebnej škály v záujmovom území.....	35
Obr.20. Územie parku pri spievajúcej fontáne po klasifikácii.....	36
Obr.21. Územie mestského parku pri železničnej stanici po klasifikácii.....	37
Obr.22. Územie Moyzesovej ulici po klasifikácii.....	38

Obr.23. Územie Hvozdkoveho parku po klasifikácii.....	38
Obr.24. Územie pri starom evanjelickom cintoríne po klasifikácii.....	39
Obr.25. Územie parku pri poliklinike Sever po klasifikácii.....	40
Obr.26. Výpočet hodnôt presnosti klasifikácie.....	41

Úvod

Zeleň je základná zložka krajiny a plní mnoho dôležitých úloh v prostredí, v ktorom žijeme. Vegetácia sa vyskytuje takmer na celej Zemi, či už v mestskom prostredí ale takisto aj vo voľnej prírode. Medzi najvýznamnejšie úlohy, ktoré plní patrí produkcia kyslíka, regulovanie klímy, fixáciu CO₂ ale takisto aj poskytuje rôznorodé prostredie pre mnoho živočíchov, čím podporuje biodiverzitu.

Mestská zeleň sú v mestskom prostredí kľúčovým prvkom, ktorý ovplyvňuje kvalitu života obyvateľov. Okrem estetickej funkcie prispievajú stromy v mestách výrazne k zlepšeniu ovzdušia a regulácii teploty. Stromy taktiež regulujú škodlivé emisie z dopravy a priemyslu a zachytávajú prach a iné nečistoty. Mestské parky a stromy poskytujú útočisko pre mnohé druhy živočíchov ako hmyz alebo vtáctvo, čo prispieva k ekologickej rovnováhe v mestskom prostredí. V dnešnej dobe je však urbanizácia na veľmi vysokej úrovni, a s tým sú späté mnohé problémy ako nedostatok zelene, degradácia ekosystémov a strata biodiverzity. To zvýrazňuje potrebu lepšej ochrany a obnovy zelene v mestských oblastiach. (Reháčková, Pauditšová, 2006)

Klasifikácia druhového zloženia stromov na základe dát z DPZ môže výrazne pomôcť pri skvalitňovaní mapovania zdravotného stavu stromov ale aj ich druhového zloženia, čo môže pomôcť k vytvoreniu pestrejšej druhovej rozmanitosti v mestských častiach Košíc. Táto metóda umožňuje vykonávať zber dát a ich následne spracovanie oveľa rýchlejšim a nahrádza komplikovanú a zdĺhavú prácu v teréne.

Cieľom tejto bakalárskej práce je s čo najväčšou úspešnosťou zaklasifikovať zloženie stromov na základe ich druhu vo vybranom území mesta Košice pomocou metód strojového učenia a to konkrétne pomocou riadenej klasifikácie. Riadená klasifikácia sa môže vykonávať rozličnými programami ako je QGIS, MultiSpec, ENVI alebo Random Forest. My v našej práci sme túto klasifikáciu vykonávali pomocou programu ArcGIS pro. Pre správne zhodnotenie výsledkov bude potrebné vytvoriť vhodné kartografické zobrazenie, ktoré umožní interpretovať výsledky čo najpresnejšie. Aj napriek tomu, že skúmané územie je konkrétna oblasť vrámci mesta Košice, predpokladáme, že nadobudnuté výsledky a metódy budú aplikovateľné aj v inej ľubovoľnej lokalite. Aj napriek vysokému nárastu a modernizácii metód DPZ je táto konkrétna problematika málo skúmaná na Slovensku, preto má táto bakalárska

práca potenciál byť vhodným podkladom pre budúce riešenia riadenej klasifikácie na druhovom zložení stromov na Slovensku.

Prehľad problematiky

1.1 Diaľkový prieskum zeme

Diaľkový prieskum zeme (DPZ) je metóda, pomocou ktorej dokážeme získavať spoľahlivé a presné informácie o fyzikálnych objektoch na zemi pomocou meraní a skenovaní a následnej interpretácie snímok a digitálnych záznamov. Tieto dáta sa zaznamenávajú mnohými metódami. Medzi najpoužívanéjšie patria multispektrálne zobrazovanie, hyperspektrálne zobrazovanie, radarové zobrazovanie alebo zobrazovanie pomocou LiDARu. Tieto metódy sa využívajú rôzne na základe toho, čo a v akej kvalite potrebujeme zaznamenať. Letecké skenovanie je technológia, pri ktorej sú nosiče, ktoré zaznamenávajú objekty na zemi pripevnené na lietadlá alebo drony. Môže sa jednať o pilotné lietadlá alebo vrtuľníky ale takisto aj bezpilotné. Pre menšie spoločnosti je táto metóda najvýhodnejšia, pretože dokážeme zaznamenať pomerne detailne veľkú časť územia za prijateľný čas. Jednou z najväčších výhod tejto metódy je zaznamenávanie vysoko kvalitných snímok, čo nám umožňuje veľmi detailnú analýzu povrchu Zeme. Avšak na druhej strane, pri zlých atmosférických podmienkach ako oblačnosť alebo hmla, snímky strácajú svoju kvalitu a v niektorých prípadoch sú kvôli zlej viditeľnosti mapovaného územia nepoužiteľné na následné spracovanie a analyzovanie. Keď potrebujeme zmapovať väčšie časti územia, tak siahneme po dátach, ktoré boli získané satelitnými a radarovými nosičmi. Tieto dáta pokrývajú oveľa väčšiu plochu ako letecké snímky a sú taktiež aj voľne dostupné. (Lillesand, Kiefer, Chipman, 2008)

Celkovo môžeme povedať, že všetky metódy DPZ majú mnoho pozitív, a to hlavne získavanie kvalitných informácií a dát bez potreby fyzického kontaktu s terénom, možnosť efektívne a rýchlo monitorovať veľké oblasti na Zemi a takisto možnosť získavania týchto dát pre bežného používateľa zadarmo. Najväčšou nevýhodou je prítomnosť nepriaznivých atmosférických podmienok alebo potrebný výkon počítačov na následné spracovanie získaných dát. Takisto niektoré metódy DPZ sú nákladné na uskutočnenie a následné spracovanie daných dát.

S istotou môžeme povedať, že DPZ je veľmi dôležitý odbor v geoinformatickom smere, ktorý nám pomáha monitorovať celú Zem aj z pohodlia domova a získavať tým rozličné informácie o oblastiach na jej povrchu a následne tieto informácie zužitkovať.

1.2. Mapovanie druhového zloženia stromov v meste

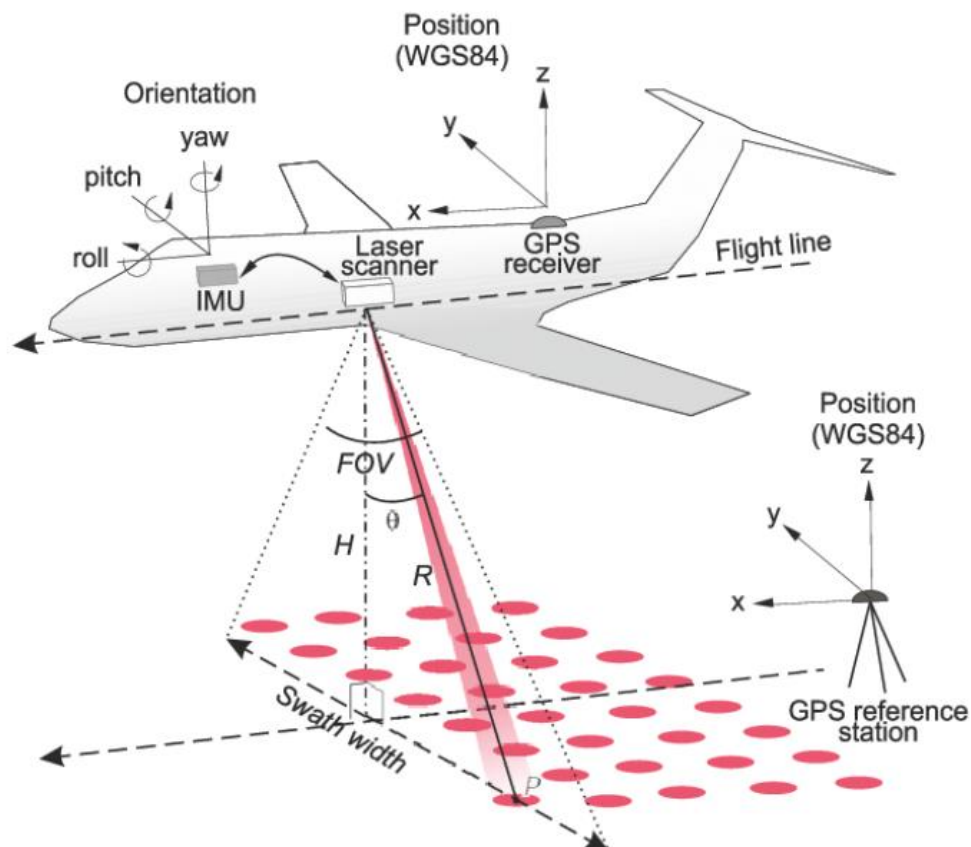
Mapovanie vegetácie pomocou metód DPZ je bežnou praktikou pre získavanie informácií o zeleni. Je to využívaná metóda či už v lesníctve, poľnohospodárstve ale takisto aj v mestských oblastiach. Metódy DPZ sa v mestských oblastiach využívajú na určenie druhového zloženia stromov, vekového zloženia ale aj na získanie informácií o ich zdravotnom stave. To nám umožňuje dozvedieť sa viac informácií o mestskej zeleni, čo môže pomôcť či už pri analýze zlepšenia ovzdušia v oblastiach, o druhovom rozložení stromov a ich následnú diverzifikáciu na základe vhodných podmienok na rast alebo o vysadení a nahradení poškodených nezdravých stromov. Okrem metód diaľkového prieskumu, ako je satelitné, radarové alebo letecké skenovanie sa takisto využíva aj terénne mapovanie, kde sa človek, ktorý dané územie preskúmava nachádza priamo v teréne. To je však oveľa náročnejšie najmä časovo, a preto sme aj v tejto práci siahli po dátach získaných leteckým snímkaním.

1.3. Letecké laserové skenovanie (LiDAR)

Stavebníctvo, kartografia, geodézia, lesníctvo, či ochrana životného prostredia, sú len niektoré z mnohých oblastí, v ktorých sa využívajú technológie skenovania terénu a objektov pomocou leteckých nosičov. Technológia LiDAR (Light Detection and Ranging) je inovatívna a moderná metóda čoraz častejšie používaná v oblastiach geoinformatiky. Svoj rozmach zažíva v posledných dvadsiatich rokoch a stále sa vylepšuje. LiDAR je založený na princípe merania času, ktorý prejde medzi vyslaním laserového lúča zo zdroja smerom na zemský povrch a časť lúča sa odrazí naspäť. Letecké laserové skenovanie je teda aktívnou metódou DPZ a jej princípy sú analogické radarovému skenovaniu. (Wasser, 2024)

Je to vysoko efektívna metóda s vysokou presnosťou, rýchlosťou a hustotou zberu údajov. Na lietadle alebo drone je pripevnený senzor, ktorý obsahuje technológiu LiDAR. Tento letecký nosič postupne prelietava nad záujmovým územím a vysiela krátke impulzy laserového svetla na zem. Množstvo impulzov vyslaných za sekundu môže byť vyše 300 000. Lúč vyslaný senzorom preniká až na pevnú zem, no počas svojho pohybu prekonáva rozličné prekážky ako sú koruny stromov, rohy budov alebo

mnoho iných a tým pádom sa na zem nedostane rovnaké množstvo impulzov, ako bolo vyslané. Pri technológii LiDAR sa najčastejšie využívajú vlny s vlnovou dĺžkou od 1064 do 1540 nm. Tieto sa nachádzajú v blízkosti infračerveného svetla (angl. NIR – Near – Infrared). Časti lúča, ktoré sa dostali až na zemský povrch sa následne odrazia naspäť k senzoru, kde sú získavané mračná bodov, ktoré senzor dokázal zachytiť. Je dôležité podotknúť, že pri tomto postupe sa meria čas, odkedy bol lúč vyslaný na zemský povrch a následne sa vrátil. Tento proces sa opakuje mnohokrát za sekundu, pričom senzor získava milióny bodových dát z rôznych častí povrchu. K určenie presnej polohy daného bodu je využívaný globálny navigačný systém (GNSS) a takisto prístroj, ktorý určuje presnú polohu a rýchlosť lietadla (IMU) . Bez týchto dvoch informácií by sme neboli schopní určiť presnú polohu lietadla v danom momente vyslania a prijatia laserového lúča a následne vypočítať polohu zameraného bodu. To nám dokonca umožňuje lokalizovať merané body s presnosťou až do 10 – 15cm v horizontálnom aj vertikálnom smere. (Straka, 2023)



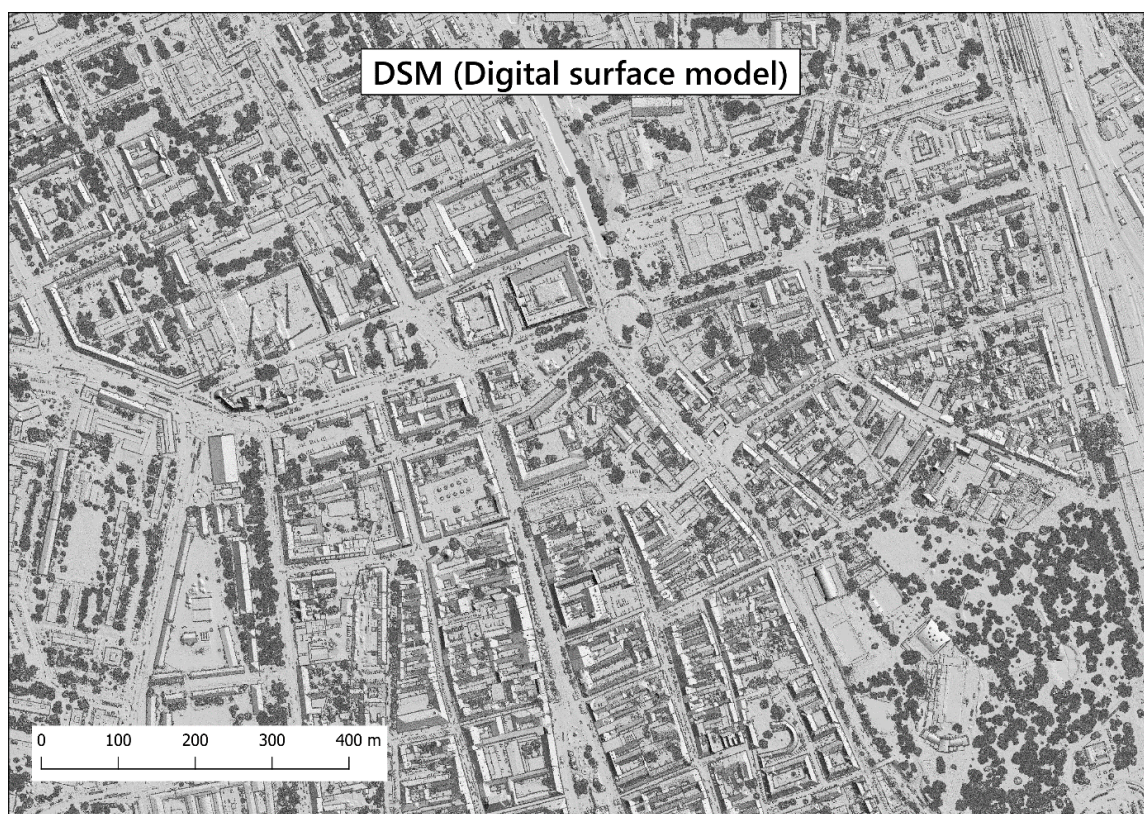
Obr. 1 Základný princíp získavania dát pomocou leteckého laserového skenovania (LiDAR) Zdroj: Galloway, 2013

1.4.DSM, DTM, DEM A CHM

Výstupy LiDARového skenovania sú DigitalElevation Model (DEM), Digital Surface Model (DSM) a Digital Terrain Model (DTM) snímky, ktoré sa na prvý pohľad zdajú podobné no je medzi nimi rozdiel. V lesnom hospodárstve sa taktiež využíva Canopy height model(CHM) a je to samostatný model, ktorý je odvodený z údajov o nadmorskej výške v mračne bodov. V našej práci sme využívali DSM a DTM snímky, ktoré majú svoje opodstatnené využitie v geoinformatike.

1.4.1.DSM

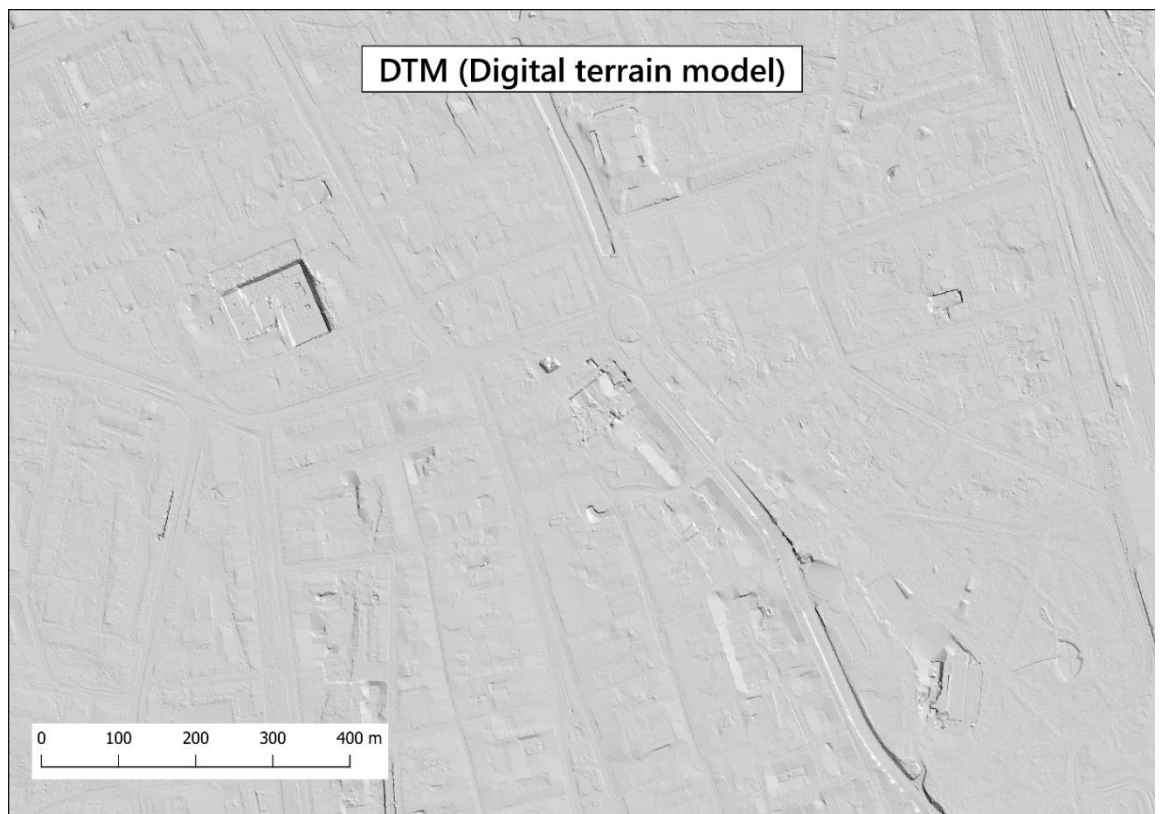
DSM je digitálna trojrozmerná reprezentácia zemského povrchu a prvkov, ktoré sa na ňom nachádzajú. To zahŕňa budovy, stromy a ostatné viditeľné prvky. Využitie takejto vrstvy má svoje opodstatnenie v mnohých oblastiach priemyslu. Napríklad pri výstavbe budov alebo telekomunikačných veží je potrebné vedieť výšku okolitých prekážok a predísť tak neočakávaným nepriaznivým podmienkam. Vrstva DSM sa využíva na tvorbu vizuálnej reprezentácie terénu s výškovými informáciami o všetkých povrchových prvkoch. Tento model je využívaný v oblasti architektúry, predpovede prírodných katastrof, výstavby telekomunikácií ale môže byť využívaný aj na tvorbu virtuálnych prehliadok budov. (Wasser, 2024)



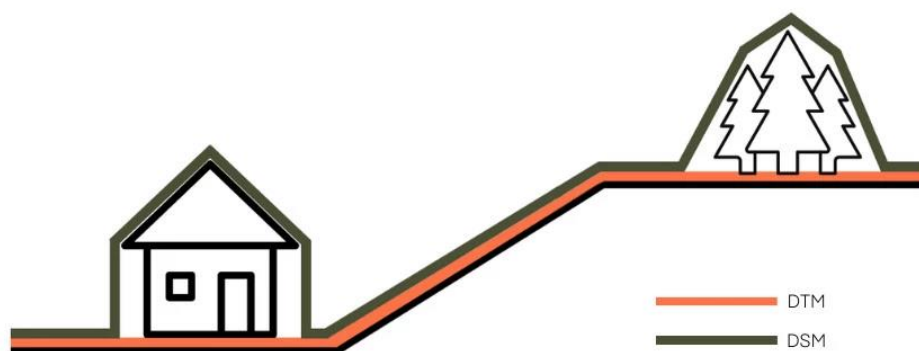
Obr. 2. Príklad DSM snímky na skúmanom území, zdroj: Vlastné spracovanie

1.4.2.DTM

DTM je digitálna reprezentácia základnej topografie Zeme, ktorá odstraňuje prvky ako budovy, stromy a iné aby sa zobrazil iba skutočný povrch terénu. Táto metóda je využívaná pre geodetické účely, plánovanie infraštruktúry, hydrologické analýzy, analýzy terénu a iné aplikácie, pre ktoré je dôležité poznať charakter terénu. (Lux carta)



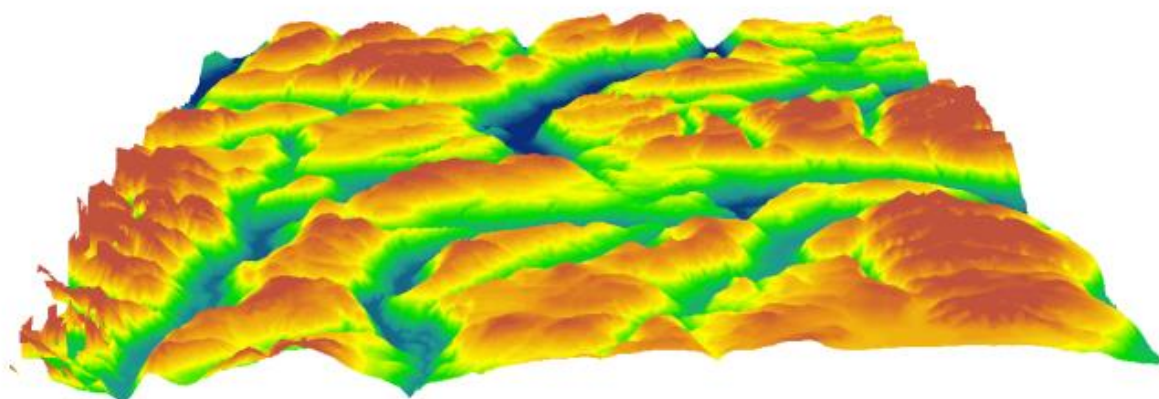
Obr.3. Príklad DTM snímky na skúmanom území, zdroj: Vlastné spracovanie



Obr. 4. Rozdiel medzi DSM a DTM dátami, zdroj: JOUAV, 2023

1.4.3. DEM

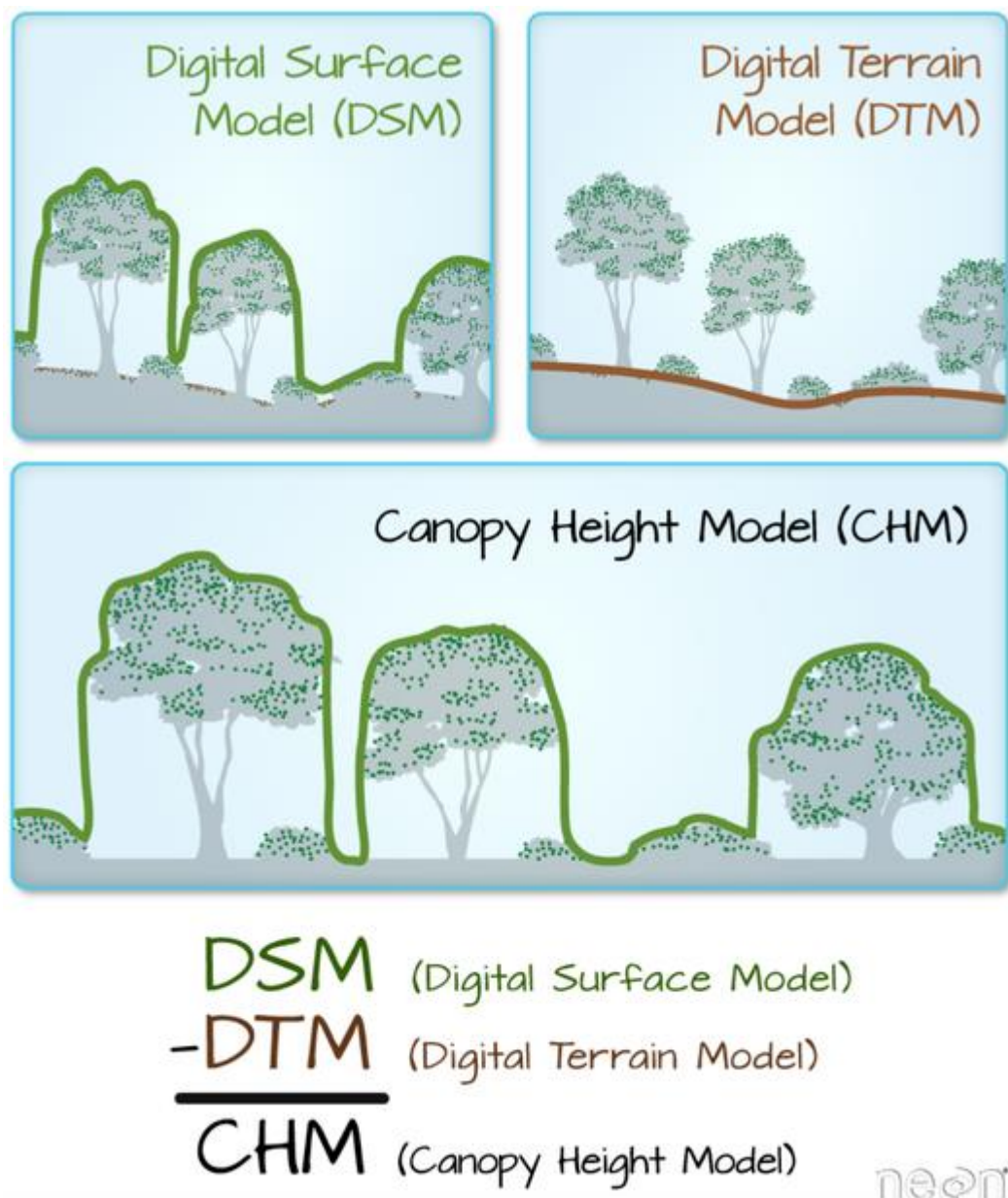
DEM je digitálny model, ktorý nám poskytuje informácie o nadmorskej výške povrchových objektov na Zemi. V tomto modeli sú odstránené všetky povrchové prvky ako stromy, budovy alebo mosty. V tomto modeli je znázornená iba „holá Zem“ a poskytuje nám výškové informácie o topografickom mapovaní ako je nadmorská výška kopcov, údolí, svahov, korýt alebo riek. Nadmorská výška týchto objektov sa znázorňuje farebnou škálou napríklad nadmorská výška od 0 až 10m.n.m. je väčšinou modrej farby, 11 až 20m.n.m. oranžovou alebo 21 až 30 m.n.m. červenou.



Obr. 5. Znázornenie modelu DEM. Zdroj: GISGeography, 2024

1.4.4. CHM

Model CHM reprezentuje výšku vegetácie nad Zemou. Tento model poskytuje cenné údaje o výške stromov, kríkov a inej vegetácie najmä pre ekologické a lesnícke štúdie. Na vytvorenie takého modelu je potrebné mať modely DSM a DTM a následne vypočítať ich rozdiel. Výsledná hodnota nám udáva výšku vegetácie nad zemou, na rozdiel od modelu DEM, ktorý udáva výšku topografických objektov nad úrovňou mora.



Obr.6. Výpočet a znázornenie modelu CHM. Zdroj: Stackexchange, 2020

1.5.Fotogrametria

„Fotogrametria je umenie, veda a technológia získavania spoľahlivých informácií o fyzických objektoch a prostredí prostredníctvom procesov zaznamenávania, merania a interpretácie fotografických snímok a vzorov zaznamenananej žiarivej elektromagnetickej energie a iných javov.“ (Wolf a Dewitt, 2000; McGlone, 2004).

Fotogrametria je metóda, pri ktorej získavame dáta pomocou vyhotovovania snímok prostredníctvom fotoaparátov a kamier pripevnených na lietadlách, vrtuľníkoch alebo bezpilotných lietadlách. Na rozdiel od metódy LiDAR, pri tejto technológii získavame vysoko kvalitné ortofotosnímky, ortofotomapy a dáta odvodené zo snímok pomocou objektívu a nie vysielaním laserového lúča. Z hľadiska zdroja EMG žiarenia ide o pasívnu metódu DPZ. Celý proces získavania trojrozmerného obrazu z dvojrozmerných fotografií prebieha v dvoch fázach a to zber údajov a následné spracovanie. Lokalita, ktorú chceme preskúmať sa najprv nasníma z rôznych uhlov a pohľadov, a získané zábery sa následne spracujú softvérom, ktorý na základe jednotlivých snímok vygeneruje spoločné body vo vzťahu k polohe. Takáto sústava bodov sa nazýva „Pointcloud“ a je základom pre následne generovanie 3D modelov územia. Presnosť výsledku závisí od hustoty vygenerovaných bodov na danej ploche. Je dôležité podotknúť, že každá takáto snímka obsahuje metadáta, napríklad informácie o nadmorskej výške, zemepisnej šírke alebo typu senzora a optike, ktoré boli pri vyhotovovaní snímok použité. Tieto informácie hrajú dôležitú úlohu pri presnosti následného vytvoreného 3D modelu. Populárnou technikou je v dnešnej dobe zbieranie takýchto dát pomocou dronov (ďalej UAV), ktoré priniesli väčšiu efektivitu a nižšiu nákladnosť na rozdiel od pilotovaných lietadiel. Výhodou pri leteckých nosičoch je rýchlosť zberu údajov, oveľa väčšia výdrž, ale naopak s UAV je možné získať detailnejšie zábery a šikmé až horizontálne snímky. Napriek zlepšovaniu kvality UAV kamier je kvalita ortofotosnímok nižšia ako u veľkoformátových kamier, ktoré sú nesené lietadlom či helikoptérou. Snímkovanie s dronmi je častejšie využívané pri mapovaní menších území, tvorbe 3D modelov budov a objektov, analýze škôd po požiaroch a živelných pohromách alebo pri výskume oblastí životného prostredia. (Mubanga, 2024)



Obr. 7. Fotografia UAV s kamerou určenou na snímkovanie územia. Zdroj: JOUAV, 2023

1.6. Strojové učenie (Machine learning)

Strojové učenie (machine learning) je odvetvie umelej inteligencie (AI), ktoré sa zaoberá vývojom algoritmov a modelov. Tieto algoritmy a modely umožňujú počítačom učiť sa z dát a skúseností, a následne tieto znalosti využívať na riešenie konkrétnych úloh bez explicitného programovania. Hlavným cieľom strojového učenia je umožniť počítačom identifikovať vzory a vzťahy v dátach. To je veľmi dôležitým nástrojom pri riešení problémov, ktoré by boli inak ťažko alebo aj nemožno riešiteľné tradičným programovaním. Algoritmy strojového učenia využívajú prvky matematickej štatistiky, metódy štatistickej analýzy a hĺbkovú analýzu. Táto technológia sa v dnešnej dobe využíva v mnohých oblastiach ako napríklad v zdravotníctve pri analyzovaní rádiologických snímok, vo finančnom sektore pri algoritmickej obchodovaní na burzách, v modernej doprave pri výrobe autonómnych vozidlách, pri detekovaní tvári a objektov v informatickom sektore ale taktiež aj v oblastiach geoinformatiky na uľahčovaní a zlepšovaní identifikácie objektov, klasifikácie druhov a typov objektov alebo vegetácie alebo aj na 3D modelovaní a analýze terénu. (IBM)

Existujú štyri typy strojového učenia, ktoré sa využívajú a to:

1. **Učenie s učiteľom (Supervised Learning)** – model AI je tréňovaný na základe vstupných a výstupných dát. Cieľom tohto typu je naučiť model predikovať výstupné hodnoty pre nové a neoznačené dáta.
2. **Učenie bez učiteľa (Unsupervised Learning)** – model AI sa učí z neoznačených dát a jeho cieľom je objaviť štruktúru alebo vzory v dátach
3. **Učenie s čiastočným učiteľom (Semi-Supervised Learning)** - Kombinovanie prvkov učenia s učiteľom a učenia bez učiteľa, kde model je tréňovaný na základe malého množstva označených dát a veľkého množstva neoznačených dát.
4. **Zosilňovanie (Reinforcement Learning)** – model AI sa učí rozhodovať na základe interakcie s prostredím a získava odmenu alebo trest na základe svojich rozhodnutí. Cieľom je naučiť model, ako optimalizovať svoje rozhodnutia, aby dosiahol maximálnu odmenu.

Postup fungovania strojového učenia zahŕňa niekoľko krokov, pomocou ktorých celý princíp funguje. Je to dynamická oblasť, s neustálym vývojom a novými technológiami, preto je dôležité vedieť, že sa mnoho vecí a postupov neustále obmieňa. Konkrétne kroky, ako strojové učenie pracuje pozostávajú zo:

1. Zberu dát – získanie alebo generovanie dát, pomocou ktorých je daný model tréňovaný. Dáta môžu byť vo forme obrázkov, meraní alebo textových dokumentov
2. Spracovanie dát – surové dáta potrebujú istú formu úpravy, aby boli vhodné na tréňovanie modelu
3. Výber a tréňovanie modelu – model je zvolený na základe typu úlohy a dát. Následne je model tréňovaný na dátach, čo znamená, že sa AI snaží naučiť vzťahy a vzory medzi dátami
4. Evaluácia modelu – po skončení tréňovacej fázy je úspešnosť a presnosť modelu vyhodnocovaná a posudzovaná
5. Optimalizácia a ladenie modelu – pokiaľ model nebol dostatočne presný a úspešný pri testovaní, nastáva optimalizácia a úprava parametrov, alebo zlepšenie predprípravy dát

6. Predikcia – po skončení tréningu AI modelu sa môže začať používať na predikciu výstupov pre nové vstupné dáta. (FBE, 2023)

1.7.Hlboké učenie (Deep learning)

Metóda hlbokého učenia je podmnožinou strojového učenia, ktorý navrhuje sofistikovaný model a je založený na neurónových sieťach inšpirovaných ľudskou myslou. Modely strojového učenia si vyžadujú zásah človeka na zlepšenie presnosti, naopak modely hlbokého učenia sa po každom dosiahnutom výsledku zdokonaľujú aj bez zásahu človeka. Tieto modely pozostávajú z viacerých vrstiev algoritmov nazývaných neuróny. Zlepšujú sa rovnako ako kognitívna myseľ, ktorá sa zdokonaľuje praxou a časom. Neurónové siete tvoria proces strojového učenia a umožňujú počítačom riešiť problémy v oblasti AI, strojového a hlbokého učenia. Neurónové siete majú vrstvy uzlov, ktoré obsahujú vstupnú vrstvu, jednu alebo viacero skrytých vrstiev a výstupnú vrstvu. Tieto siete sa spoliehajú na tréningové údaje, aby sa naučili a zlepšovali svoje predpovede v priebehu času, čo umožňuje ich lepšiu využiteľnosť. Hlboké učenie hrá dôležitú rolu v mnohých súčasných technológiách, ako napríklad Alexa alebo Siri, ktoré bežne používame. Na rozdiel od klasického strojového učenia, ktoré zvyčajne využíva údaje, ktoré sú štruktúrované a riadne označené, hlboké učenie pracuje aj s neštruktúrovanými údajmi. To znamená, že príprava tréningových dát nie je potrebná až tak podrobne ako pri strojovom učení. Algoritmy hlbokého učenia spracúvajú údaje, ktoré sú tvorené textom, obrázkami a automatizujú extrakciu rozličných funkcií. Hlboké učenie využíva rozličné typy metód zlepšovania sa. Môžu sa zlepšovať taktiež bez učiteľa, to znamená, že nevyžaduje riadne označené súbory údajov, ale odhaľuje vzory v údajoch na základe určitých charakteristík bez pomoci ľudského faktoru.

Hlboké učenie je dôležitým a komplexným nástrojom, ktorý zahŕňa štatistiku či prediktívne modelovanie. Je to mimoriadne prospešná metóda pre vedcov, ktorí zhromažďujú a analyzujú veľké množstvo údajov. Vďaka hlbokému učeniu je tento proces oveľa rýchlejší a menej náročný bez nutnosti hľadania zložitých a často neexistujúcich riešení. Hlboké učenie sa dá v podstate považovať za spôsob automatizácie prediktívnej analýzy. (IBM)

1.8.Klasifikácia obrazu

Klasifikácia obrazu je jednou z výsledných častí spracovania obrazu. V širšom slova zmysle ide o proces, kde obrazovým prvkom priradíme informačný význam. Jeho hlavným cieľom je nahradiť hodnoty rádiometrických charakteristík pôvodného obrazu, ktoré vyjadrujú spektrálne vlastnosti objektov a javov na ňom zobrazených, hodnotami vyjadrujúcimi informačné triedy. Klasifikáciu môžeme obecné rozdeliť do dvoch skupín a to na riadenú a neriadenú klasifikáciu. Výber typu klasifikácie závisí na tom, akým spôsobom a v ktorom okamžiku spracovateľ zasiahne do procesu klasifikácie. (Esri, 2021)

1.9.Riadená klasifikácia

Riadená klasifikácia je technika používaná pri spracovaní dát získaných diaľkovým prieskumom zeme, kedy sa tréningové sady dát používajú na klasifikovanie pixelov v obraze. Tréningové dáta pozostávajú z rozličných vzoriek typov krajiny pokrývky, ktoré používateľ chce zmapovať. Výber tréningových dát si vyberá používateľ ručne, a takisto ich aj ručne priradzuje k príslušným triedam. Klasifikátor následne použije tieto tréningové údaje na klasifikáciu zostávajúcich pixelov v obrázku. Riadená klasifikácia sa považuje za oveľa presnejšiu ako neriadená klasifikácia, pretože používateľ má kontrolu nad procesom klasifikácie. Avšak výber vhodných vzorových dát si vyžaduje väčšie úsilie a odbornú znalosť. Hlavnou výhodou riadenej klasifikácie je umožnenie presnej a efektívnej klasifikácie veľkých oblastí. Je to takisto flexibilná technika, ktorá má využitie v rôznych aplikáciách. Výber tréningových dát je kľúčovým prvkom pri riadenej klasifikácii. Pre každú triedu je nevyhnutné správne vybrať reprezentatívne pixely, ktoré zachytávajú spektrálnu variabilitu danej triedy. Počet tréningových pixelov závisí od zložitosti skúmanej oblasti a počtu tried. Presnosť klasifikácie závisí od kvality tréningových dát a výberu algoritmu. Na určenie presnosti klasifikácie sa používa súbor validačných údajov, čo je tvorené pixelmi, ktoré sa vo tréningovej fáze nevyužili. Používateľ následne porovnáva klasifikované pixely s validačnými dátami a vypočítava celkovú presnosť a presnosť pre špecifickú triedu. (Spatialpost, 2023)

1.10. Neriadená klasifikácia

Neriadená klasifikácia je technika DPZ, kde klasifikačný algoritmus automaticky zoskupuje pixely s podobnými vlastnosťami do klasterov. Používateľ tým pádom nemusí poskytovať žiadne tréningové datasety pre algoritmus. Naopak, algoritmus analyzuje spektrálne vlastnosti obrazu a identifikuje zhľuky pixelov s podobnými vlastnosťami. Neriadená klasifikácia je tým pádom rýchlejšia metóda klasifikácie ako riadená. Avšak táto metóda je menej presná, pretože algoritmus, ktorý sa pri neriadenej klasifikácii využíva nedisponuje žiadnymi informáciami o type krajinej pokrývky. Neriadená klasifikácia používa techniku klastrovania pre pixely s podobnými spektrálnymi vlastnosťami. Pri klastrovaní algoritmus priradzuje každý pixel do klastra na základe spektrálnej podobnosti s ostatnými pixelmi. Každému zoskupeniu potom používateľ priradí triedu krajinej pokrývky na základe svojich znalostí o skúmanej oblasti. Po dokončení klasifikácie algoritmom, je nutné určiť počet tried ručne. Optimálne určenie tried však niekedy môže byť náročné, pretože správne určenie závisí na vedomostiach používateľa. Čo sa presnosti neriadenej klasifikácie týka, podobne ako pri riadenej klasifikácii sa musia triedy krajinej pokrývky priradiť manuálne ku každému zhľuku klastrov. To je však častokrát časovo veľmi náročné a priradzovanie tried môže byť veľmi subjektívne. (Donald, Weih, 2005)

2. Zber a spracovanie dát

Pri postupnom spracovaní dát sme využili rozličné metódy DPZ a umelej inteligencie, ktoré nám pomohli lepšie pochopiť problematiku záujmového územia. Ako prvé sme poskytnuté dáta museli odfiltrovať, a rozhodnúť sa, s ktorými budeme počas našej práce pracovať. Rozhodovali sme sa medzi RGB a CIR snímkami, no vzhľadom na cieľ práce a to druhovo rozdeliť zloženie stromov sme sa rozhodli pre CIR snímky s rozlíšením 10cm. Dáta CIR zahŕňajú tri spektrálne pásma a to blízke infračervené (NIR/Near infrared), červené (Red) a zelené (Green). Infračervené pásmo je veľmi citlivé na vegetáciu a umožňuje lepšiu druhovú identifikáciu stromov. Výrazne červené farby znázorňujú zdravé stromy a vegetáciu a slabšie odtiene červenej naopak.



Obr. 8. Časť záujmového územia v RGB kompozícii.



Obr.9. Časť záujmového územia v CIR kompozícii.

2.1. Záujmové územie a jeho charakteristika

Územie spracované v tejto bakalárskej práci sa zameriava na mesto Košice, druhé najväčšie mesto Slovenskej Republiky. Vegetáciu mesta Košice tvoria prevažne lesy košického lesoparku, pasienky a zeleň v blízkosti obytných sídlach – záhrady, sady, parky a mestská zeleň. Na území mesta Košice prevláda pestrá druhová rozmanitosť a je možné tu nájsť takmer všetky druhy pôvodných európskych drevín. Najväčšie zastúpenie majú listnaté dreviny (83%) z toho buk (54%), dub (16%) ale aj topoľ, breza, javor či lipa. Zvyšná časť zastúpenia drevín je tvorená ihličnatými drevinami (17%) ako jedľa (8%), smrek (6%), borovica (2%) smrekovec (1%) (zdroj).

Pre analýzu územia sme v rámci mesta vymedzili štvorcové územie s rozmermi 2 x 2 kilometre. V tejto oblasti sa nachádzajú rôzne typy zelene, od malých kríkov až po vysoké stromy v parkoch. Najväčšiu koncentráciu a druhovú rozmanitosť stromov môžeme nájsť v mestských parkoch Košíc. Pre podrobnejšie skúmanie zelene sme si vytýčili štyri lokality, kde budeme algoritmus riadenej klasifikácie trénovať a to: Hvozdíkov park, Mestský park pri autobusovej stanici, park pri spievajúcej fontáne na Hlavnej ulici a Moyzesovú ulicu. Po natrénovaní algoritmu ho následne dokážeme použiť na zaklasifikovanie stromov v celom záujmovom území.

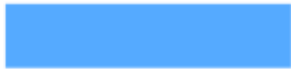
2.2. Úprava dát pre potreby práce

Surové poskytnuté dáta boli prvotne upravované v softvéri QGIS vo verzii 3.36.3. CIR snímky, s ktorými sme sa rozhodli pracovať boli rozdelené do menších častí, ktoré bolo pre lepšiu prehľadnosť a presnosť pri práci potrebné zjednotiť do jedného veľkého rastra. Táto úprava sa vykonávala v prostredí QGIS pomocou funkcie merge raster. Tento krok nám následne vygeneroval jeden veľký polygón, s ktorým bola následná práca omnoho jednoduchšia.

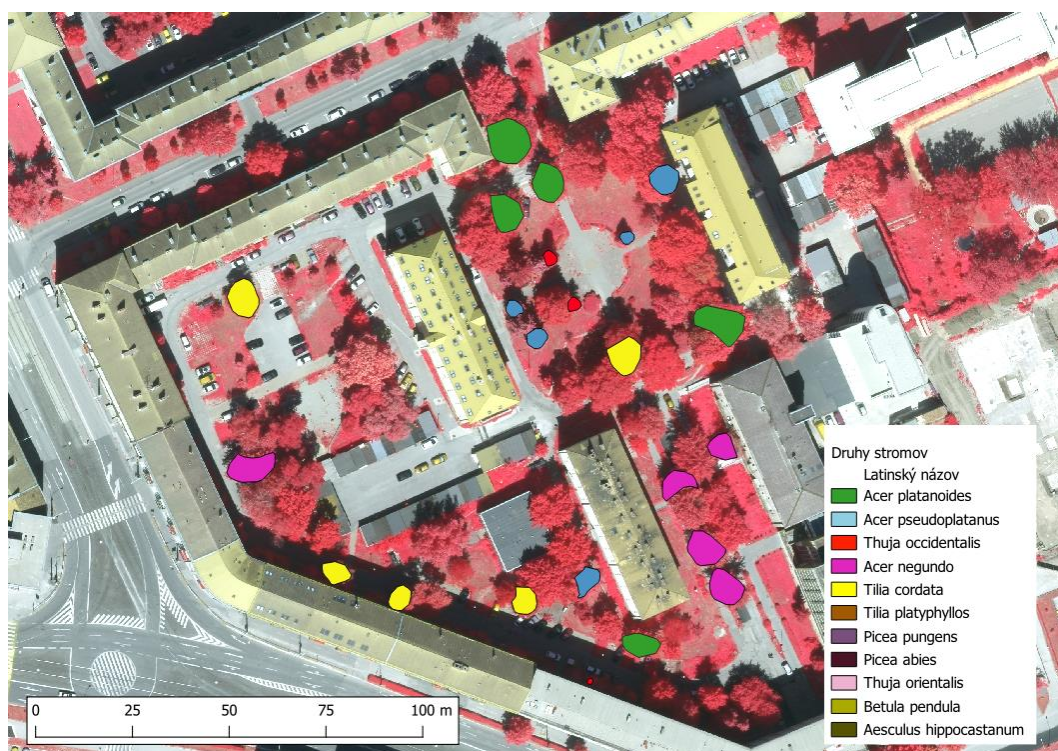
Ako ďalšie bolo potrebné vyčleniť druhy stromov, ktoré chceme pri riadenej klasifikácii používať ako trénovacie dáta. Po podrobnom preskúmaní všetkých štyroch lokalít, z ktorých budeme vyberať tréningové polygóny sme sa rozhodli pre 11 druhov stromov a to:

- 1 - Acer Platanoides
- 2 - Acer Pseudoplatanus
- 3 - Thuja occidentalis
- 4 - Acer negundo
- 5 - Tilia Cordata
- 6 - Tilia platyphyllos
- 7 - Picea pungens
- 8 - Picea abies
- 9 - Thuja orientalis
- 10 - Betula pendula
- 11 - Aesculus hippocastanum

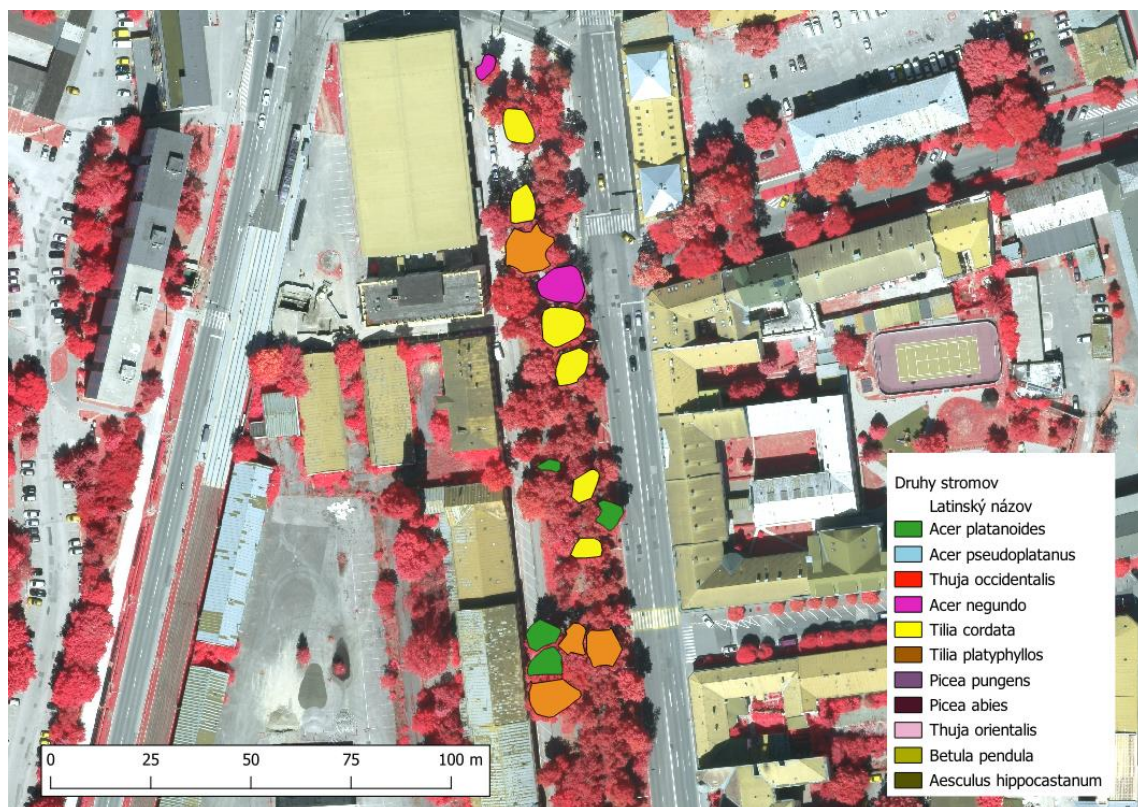
Celkovo sme vytvorili 107 trénovacích polygónov druhového zloženia stromov. V priemere sme vytvorili z každého druhu od 9-10 polygónov, pre čo najpresnejšiu klasifikáciu.

Acer platanoides	
Acer pseudoplatanus	
Thuja occidentalis	
Acer negundo	
Tilia cordata	
Tilia platyphyllos	
Picea pungens	
Picea abies	
Thuja orientalis	
Betula pendula	
Aesculus hippocastanum	

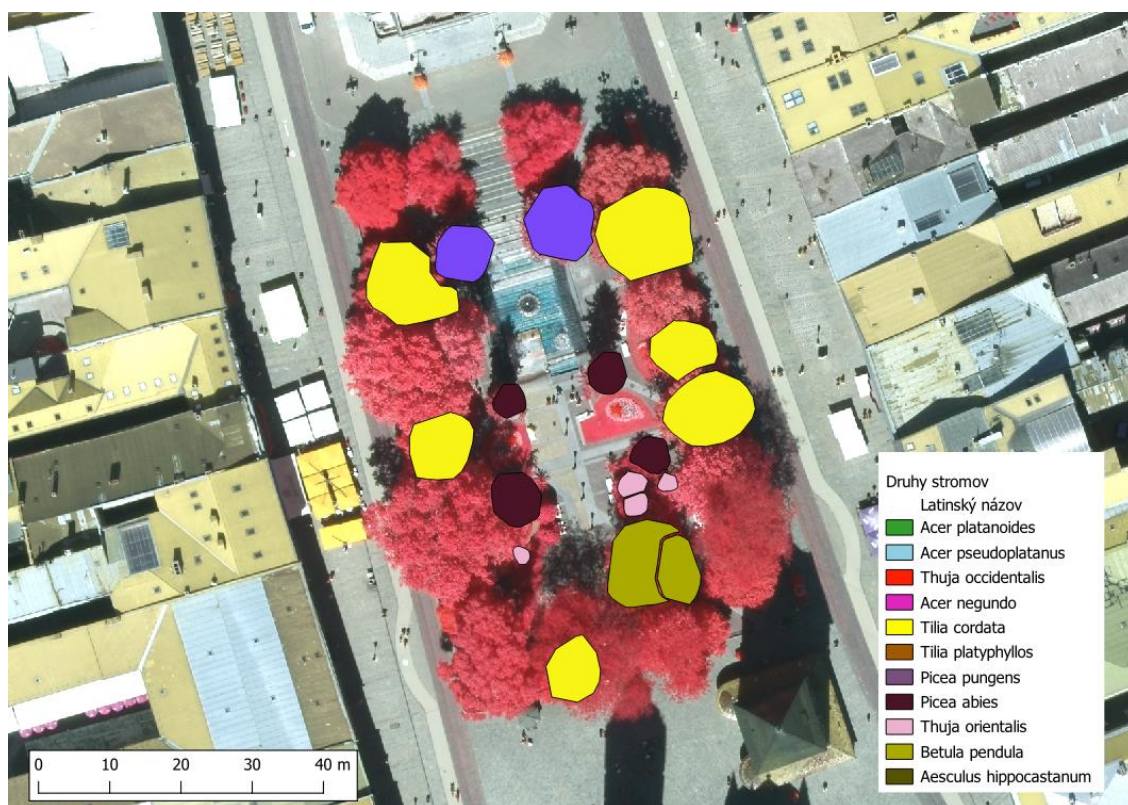
Obr.10. Farebné rozdelenie druhov stromov v tréningových polygónoch.



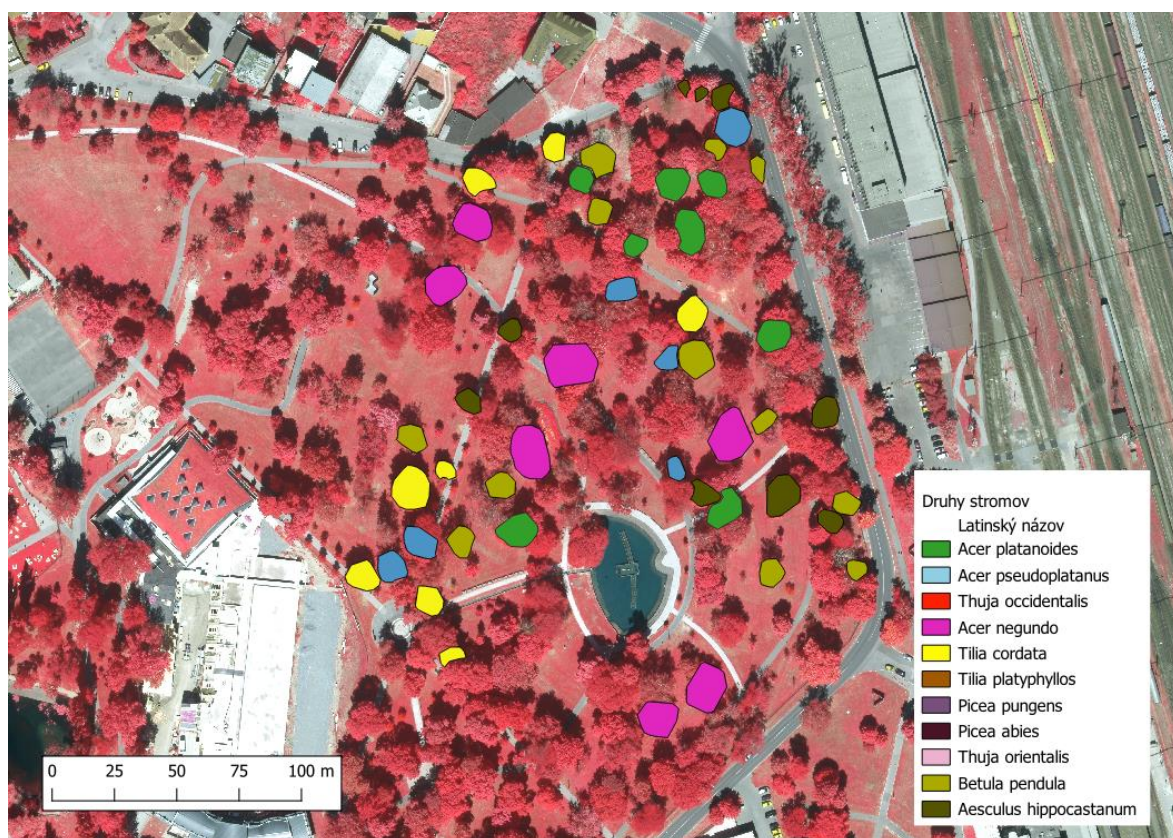
Obr.11. tréningové polygóny v oblasti Hvozdkov park.



Obr.12. tréningové polygóny v oblasti Moyzesová ulica.



Obr.13. tréningové polygóny v oblasti spievajúcej fontány na Hlavnej ulici.



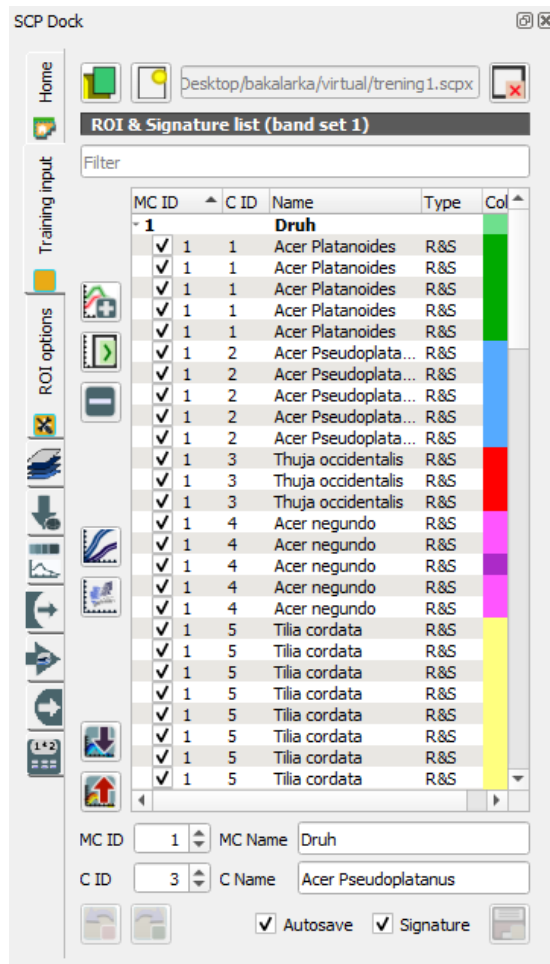
Obr.14. tréningové polygóny v oblasti Mestského parku pri vlakovej stanici.

Tréningové polygóny boli tvorené pomocou klasifikačného pluginu Semi-Automatic Classification Plugin (SCP), avšak so samotnou klasifikáciou sa nám v tomto plugine vyskytoval problém, preto musela byť uskutočnená v softvéri ArcGIS pro. Tréningové polygóny sme vyberali na základe čo najpresnejšej viditeľnosti a bez prekrytov s okolitými druhmi stromov. Najpestrejšia druhová rozmanitosť bola v mestskom parku pri železničnej stanici, čo je pochopiteľné vzhľadom na jeho veľkosť. Naopak najmenšiu plochu osadenú stromami zaberal park pri spievajúcej fontáne na Hlavnej ulici. Najrozšírenejší druh vo všetkých štyroch lokalitách bola *Tilia cordata* (Lipa malolistá) a najmenšie zastúpenie mal *Piceus pungens* (smrek pichľavý).

2.3. Použitie riadenej klasifikácie

Pri spracovaní dát sme sa v bakalárskej práci použiť čo najlepšiu a najpresnejšiu metódu klasifikácie. Zo začiatku sme sa snažili ju vytvoriť už spomínaným SCP pluginom v prostredí QGIS, avšak pri spustení samotnej klasifikácie sme sa stretli s mnohými chybami. Riadená klasifikácia využíva pripravené dáta (v našom prípade

polygóny) a my sme si ich pripravili v prostredí QGIS, pretože nám to prišlo najprehľadnejšie.



Obr. 15. Pripravené polygóny druhov stromov na klasifikáciu.

Vytvorený súbor s trénovacími polygónmi sme si exportovali vo formáte geopackage, a následne sme boli pripravení vykonať klasifikáciu.

Klasifikácia prebiehala v softvéri ArcGIS pro a celkový čas od spustenia po dokončenie klasifikácie zabral viac ako 24 hodín. Pre úspešnú a čo najpresnejšiu klasifikáciu sme sa rozhodli pre nástroj (classifier) Support Vector Machine (SVM). Tento klasifikátor je vhodný pre klasifikovanie objektov na veľkom území a je vhodný

pre

segmentovaný

rastrový

výstup.

Image Classification

Classify : virtualka.tif

Classifier

Support Vector Machine

Training Samples

E:\BA\bakalarka (2)\prjbc\training1.shp

Maximum Number of Samples per Class

500

Segmented Image (optional)

Output Classified Dataset

E:\BA\bakalarka (2)\prjbc\prjbc.gdb\Classified_202406270015327315803

Output Classifier Definition File (.ecd)

E:\BA\bakalarka (2)\prjbc\ClassDefinitions_202406270015327315543.ecd

Run

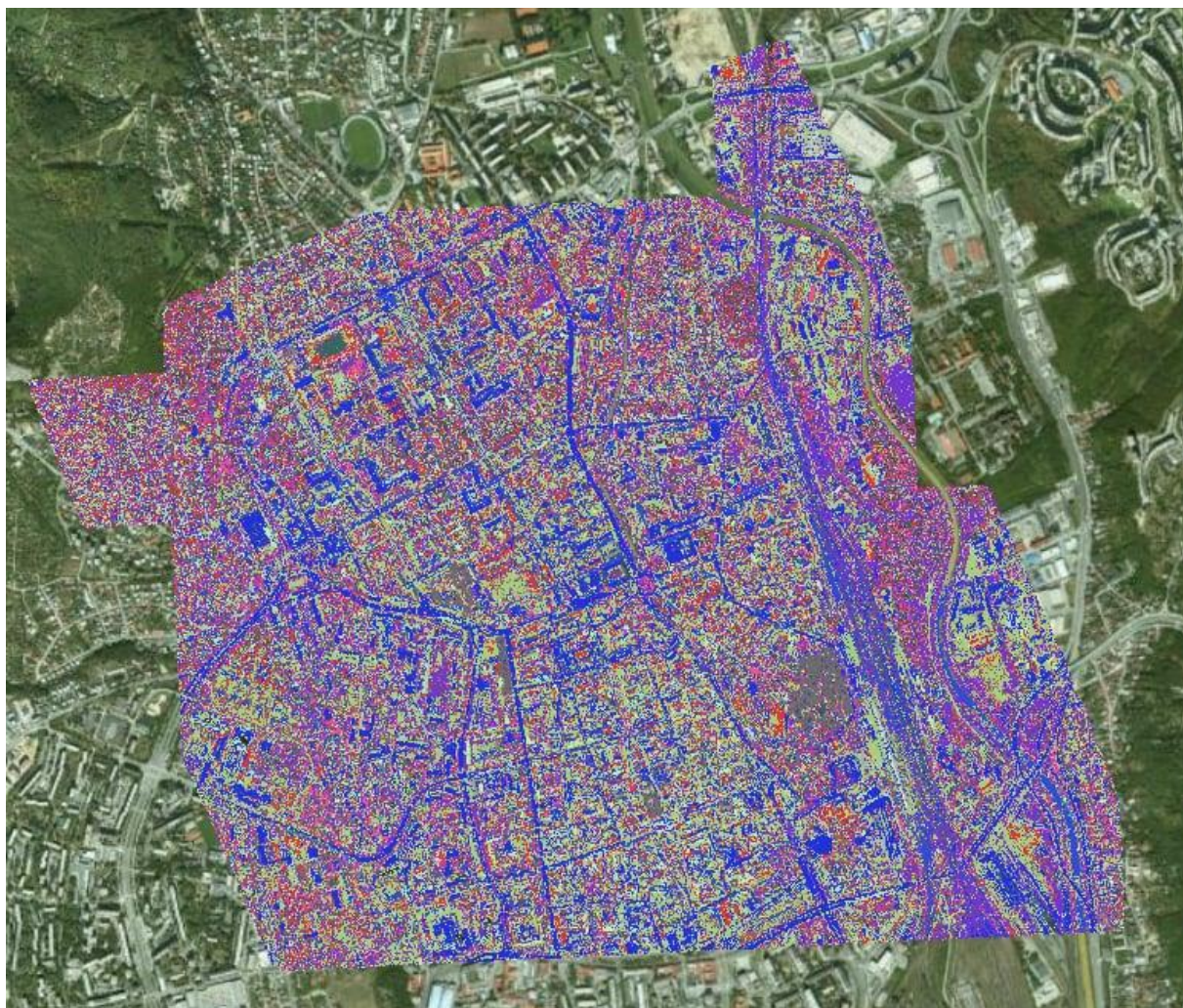
Obr. 16. nastavenie hodnôt klasifikátora SVM.

V samotných nastaveniach klasifikátora sme si nastavili, aby maximálny počet prvkov danej triedy bol maximálne na hodnote 500, pre prehľadnejší a výstup. Následne sme spustili klasifikáciu a čakali viac ako 24 hodín na jej dokončenie.

3. Výsledky

3.1. Kvalita a vlastnosti výsledných dát

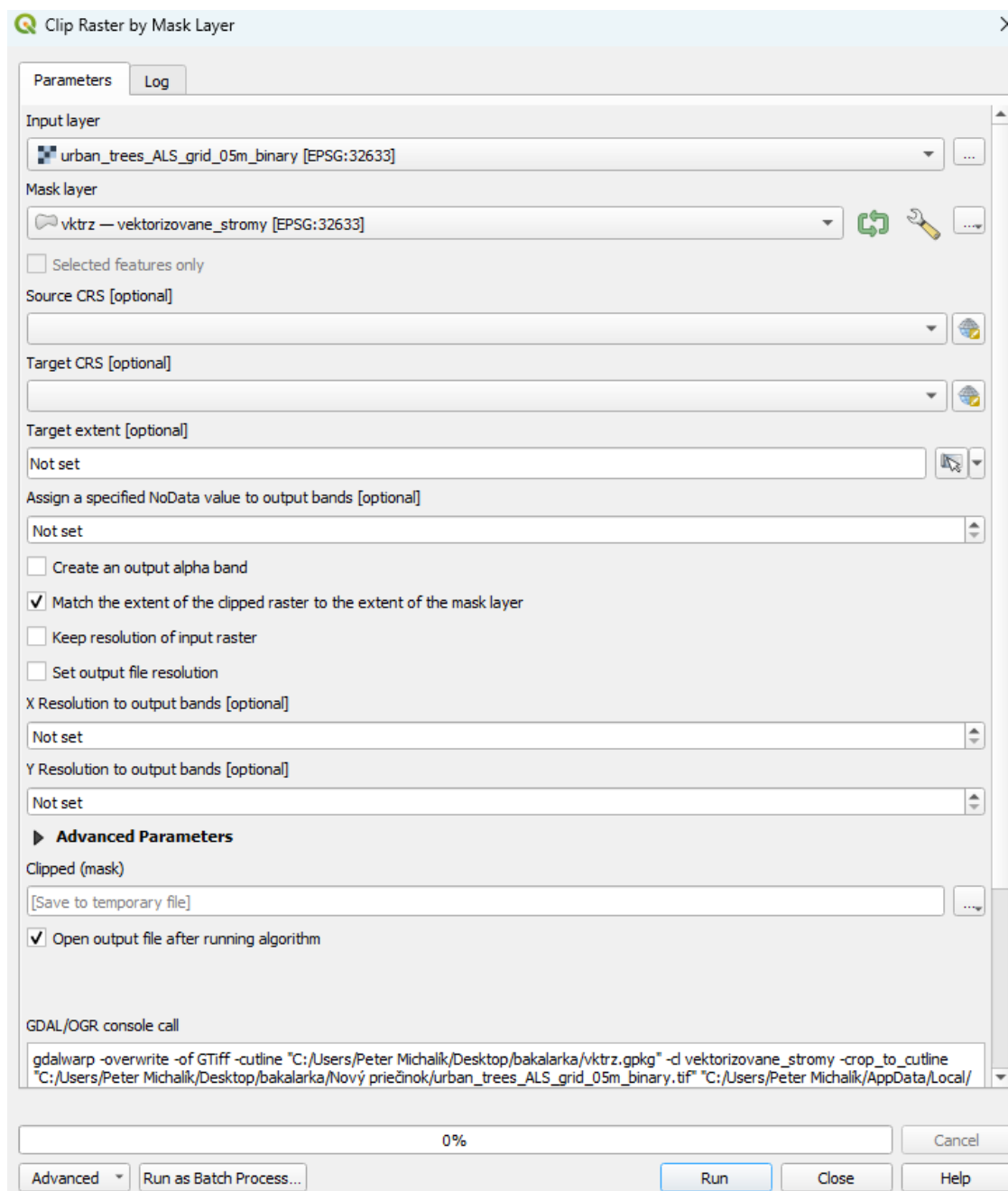
Výsledkom klasifikácie je jeden polygón, ktorý okrem tried stromov nám klasifikoval takisto nové kategórie, ktoré zobrazovali budovy, cesty, rieky a iné prvky na danom území.



Obr. 17. výsledný polygón riadenej klasifikácie

Vytvorený polygón bol príliš neprehľadný, preto bolo potrebné odfiltrovať nepotrebné triedy a zachovať iba druhy stromov, čo bolo aj cieľom našej práce. Pre správne filtrovanie nepotrebných dát sme využili nástroj v programe QGIS clip raster by mask extent. Ako masku, podľa ktorej sme oklipovanie klasifikovaného rastra vykonali sme si vybrali vrstvu všetkých stromov na skúmanom území. Pre správnosť a presnosť tohto procesu, bolo potrebné klasifikovaný raster spolu s vrstvou stromov

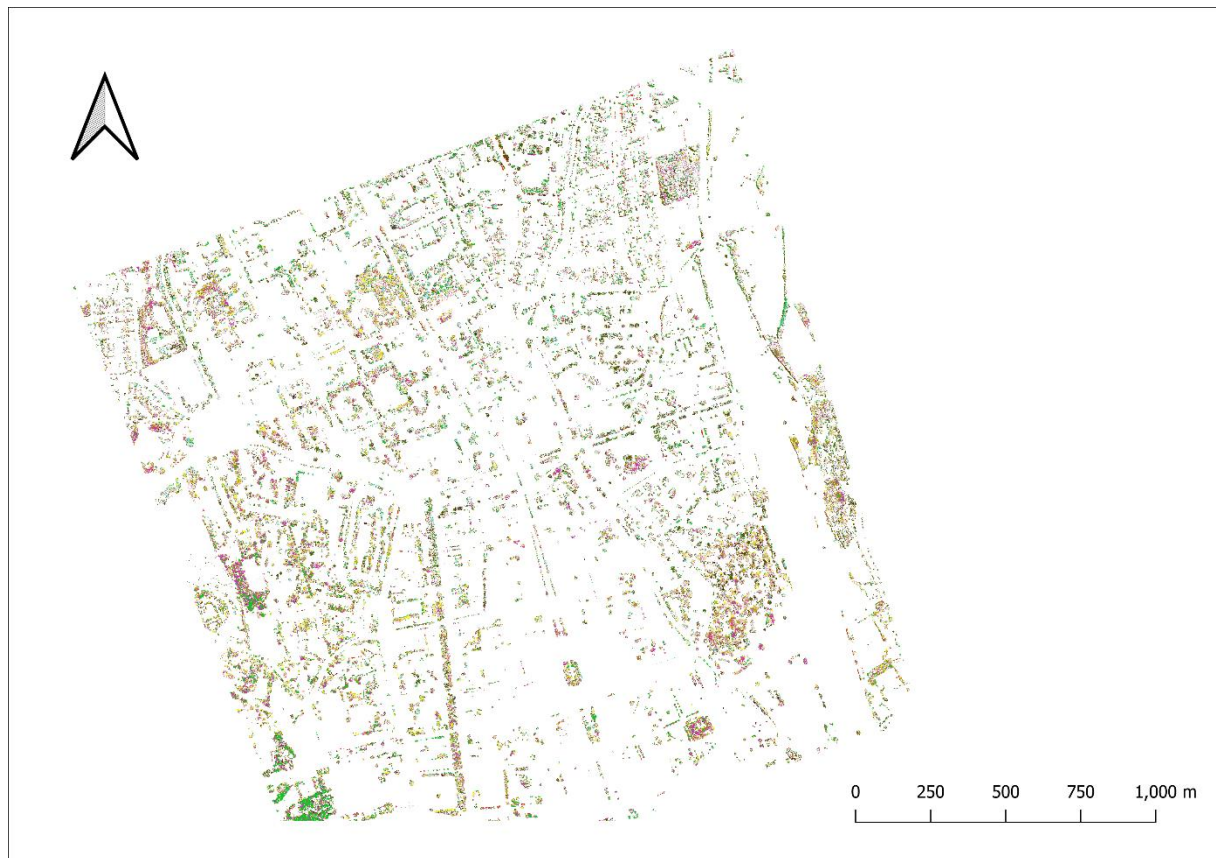
mať nastavenú v rovnakom súradnicovom systéme, čo bol v našom prípade EPSG:32633 – WGS 84 / UTM zone 33N.



Obr.18. nastavenia klipovania klasifikovaného rastra

Po dokončení klipovania nám vznikla rastrová vrstva, ktorá vyobrazovala všetky druhy stromov, ktoré sme si pri klasifikácii vybrali ako záujmové. Na prvý pohľad sa dáta zdajú chaotické a neprehľadné, čo je zapríčinené obrovským rozdielom vo veľkosti jedného stromu v porovnaní s veľkosťou celého územia a počtom všetkých

stromov v danom území. Pri podrobnejšej analýze sme sa už zamerali na konkrétne druhy stromov, ktoré nás zaujímali, a ktoré boli hlavným cieľom riadenej klasifikácie.



Obr.19. rozdelené druhy stromov na základe farebnej škály v záujmovom území.

3.2. Identifikácia jednotlivých druhov stromov

Predpokladali sme, že výsledné dáta nebudú tak prehľadné ale naopak, budú skôr chaotické a nepresné a to sa aj potvrdilo. Pri bližšom skúmaní bolo už na prvý pohľad jasné, že vo výslednej klasifikácii sa vyskytujú veľké odchýlky a nepresnosti. Na obrázkoch nižšie si bližšie zanalyzujeme konkrétne druhy stromov v jednotlivých častiach územia. V lokalite na Hlavnej ulici pri spievajúcej fontáne sme ako tréningové polygóny použili celkovo 5 druhov stromov a to: žltá – *Tilia cordata*, fialová – *Picea pungens*, bordová – *Picea abies*, ružová – *Thuja orientalis* a žltozelená – *Betula pendula*. Celkovo sme v tejto lokalite využili 18 tréningových polygónov. Prekvapivo najpresnejšie nám algoritmus zaklasifikoval druh *Picea pungens* aj napriek tomu, že sme použili iba dva tréningové polygóny. Takisto pomerne presná klasifikácia prebehla aj na druhu *Picea abies*, ktorý bol farebne označený bordovou farbou. Dôvodom toho je veľkosť koruny stromu ale takisto odlišnosť druhov *Picea* od zvyšných listnatých

stromov, pretože ihličnaté stromy majú inú nižšiu farebnú odrazivosť v NIR pásme na rozdiel od listnatých. Čo sa zvyšných stromov týka, je už na prvý pohľad jasné, že klasifikácia v nich nebola veľmi úspešná. Príčinou toho je fakt, že v danej lokalite je vysoká rozmanitosť druhového zloženia stromov, ktoré sú veľmi husto osadené a preto nie je možné určiť ich druhové zloženie s dokonalou presnosťou. Je zrejmé, že farebné anomálie klasifikácie sa budú opakovať nie len v tejto lokalite, ale v celom záujmovom území.



Obr.20. Územie parku pri spievajúcej fontáne po klasifikácii.

Ako ďalšia lokalita, ktorou sme sa zaoberali je mestský park pri stanici. Svojou veľkosťou to je najväčšia plocha s koncentráciou stromov na jednom mieste v celom záujmovom území. Použili sme tu celkovo 50 tréovacích polygónov z toho 5 druhov stromov – *Acer platanoides*, *Acer pseudoplatanus*, *Tilia cordata*, *Acer negundo* a *Aesculus hippocastanum*. Podobne ako v prvej lokalite tu presnosť klasifikácie v oblastiach s husto zarastenými stromami nebola presná. Najpresnejšie zaklasifikovaný bol druh *Aesculus hippocastanum*, ktorý je najvýraznejšie vidieť v oblasti celého parku. Zvyšné druhy stromov nie sú tak jednoznačne rozlíšiteľné, to ma za príčinu husté osadenie a podobnosť farebnej odrazivosti daných druhov.

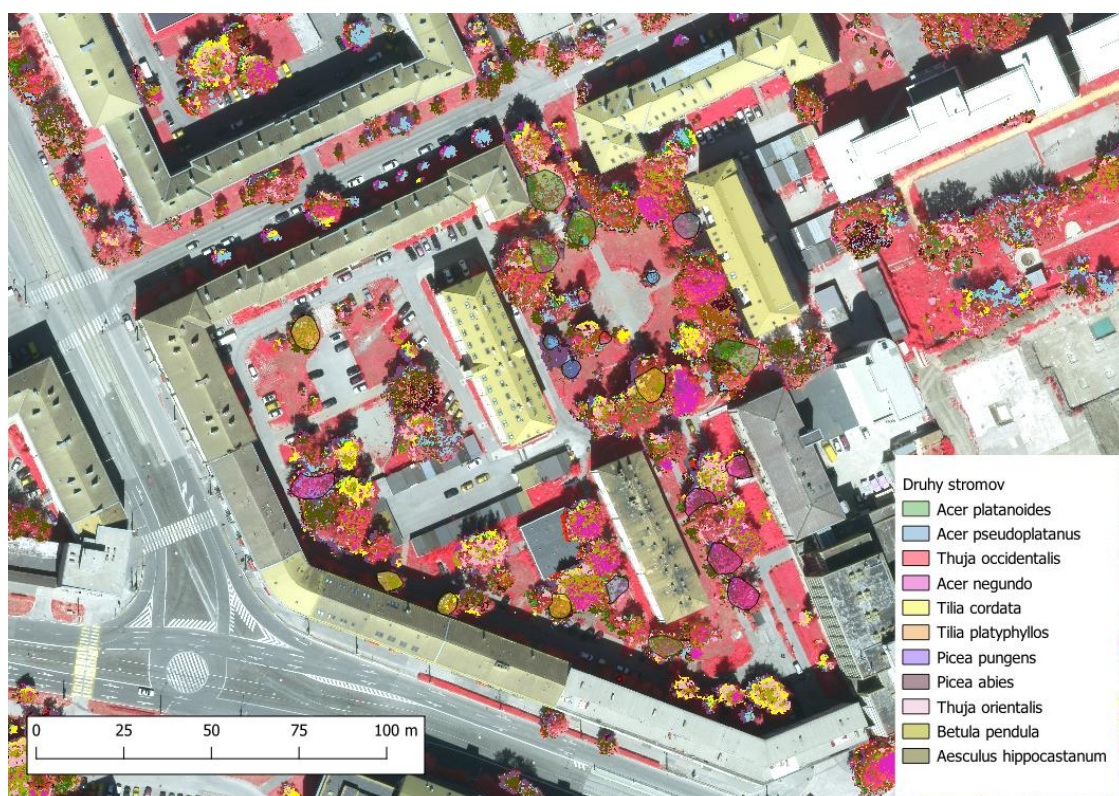


Obr.21. Územie mestského parku pri železničnej stanici po klasifikácii.

Čo sa týka stromov na Moyzesovej ulici a v Hvozdíkovom parku, klasifikácia tu bola najmenej úspešná. V oboch lokalitách je nejasne rozoznateľné druhové zloženie stromov, nachádza sa tam mnoho prekrytov a algoritmus mal problém správne a jasne identifikovať druhy stromov. Jedine druh *Acer negundo* bol spomedzi ostatných druhov pomerne presne identifikovateľný.



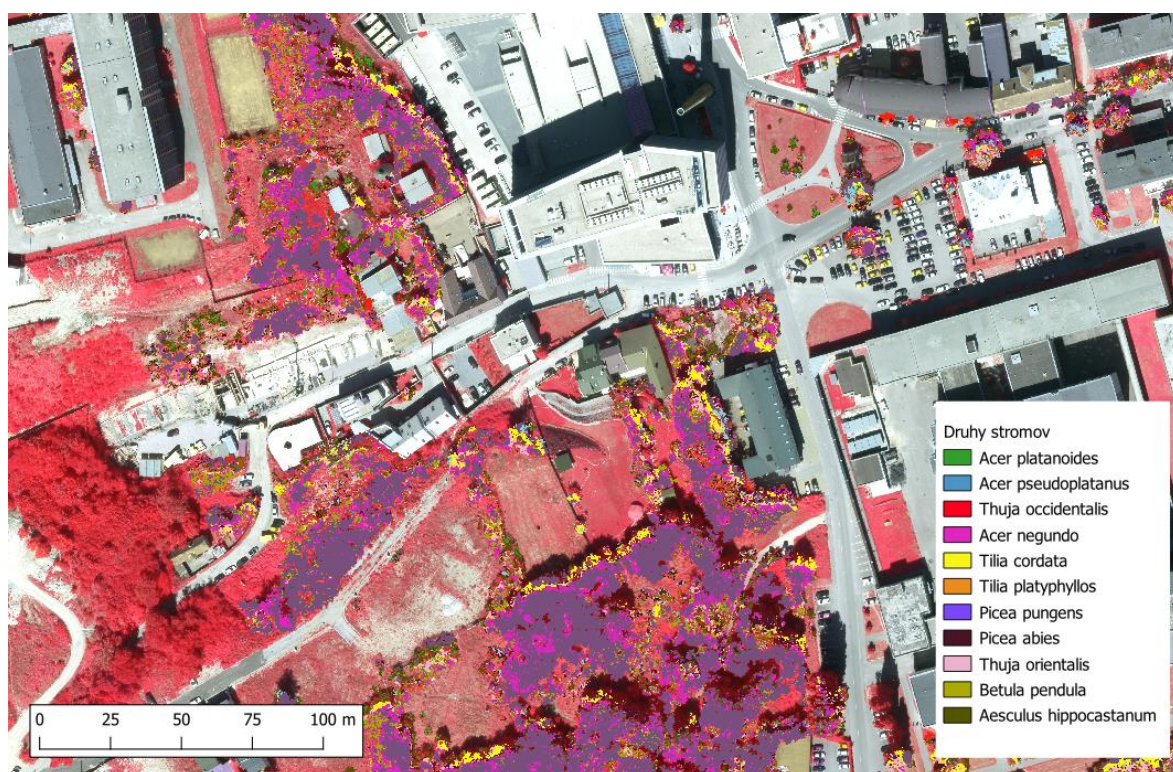
Obr.22. Územie Moyzesovej ulici po klasifikácii.



Obr.23. Územie Hvozdkoveho parku po klasifikácii.

3.3. Identifikácia druhového zloženia stromov v častiach mesta

V záujmovom území sa po dokončení klasifikácie objavilo niekoľko zaujímavých oblastí, kde klasifikácia bola na rozdiel od zvyšných oblastí veľmi úspešná. Oblasť v okolí starého evanjelického cintorína nad Cassovarom obsahuje výrazné množstvo ihličnanov a to konkrétne druh *Picea pungens*. Podobne ako v ostatných lokalitách, ihličnatý strom bol klasifikovaný pomerne presne a to vďaka svojim odrazovým vlastnostiam.



Obr.24. Územie pri starom evanjelickom cintoríne po klasifikácii

Ďalšou zaujímavou oblasťou je park pri poliklinike Sever. Najrozšírenejším druhom v tejto oblasti je vyobrazený svetlo-modrou farbou a to *Acer pseudoplatanus* ale takisto tu má pomerne vysoké zastúpenie aj *Picea pungens*, ktorý je vyobrazený fialovo. Okrem týchto dvoch druhov tu môžeme vidieť takisto zastúpenie žltej farby – *Tilia cordata* a svetloružovou *Thuja orientalis*. Zvyšné druhy stromov sú menej jasne viditeľné a to je spôsobené tým, že sa jedná o listnaté stromy, s ktorými mal algoritmus v celom území najväčší problém ich druhovo rozoznať.



Obr.25. Územie parku pri poliklinike Sever po klasifikácii.

3.4. Výpočet presnosti klasifikácie

Po dokončení samotnej klasifikácie bolo potrebné vypočítať presnosť a úspešnosť daných postupov. Ako už bolo spomenuté, bolo už na prvý pohľad jasné, že o vysokej úspešnosti a presnosti klasifikácie nemožno hovoriť. Pri porovnaní klasifikovaného rastra s testovacími dátami boli opticky viditeľné mnohé nepresnosti a anomálie. Najväčší vplyv na úspešnosť mala práve veľkosť celého záujmového územia. Klasifikované územie malo cez 4 km² a v porovnaní s veľkosťou polygónu jedného stromu nie je možné dokonale klasifikovať a odlíšiť druhové zloženie stromov. Pre výpočet presnosti klasifikácie sme využili plugin SCP, a v ňom konkrétne postprocessingovú funkciu Accuracy. Ako vstupný raster sme využili novo vzniknutý zaklasifikovaný raster a ako validačné dáta sme použili nové polygóny daných druhov. Po spustení funkcie na výpočet presnosti sme získali konkrétne hodnoty, ktoré su vyobrazené na obrázku nižšie. Číslo, ktoré nás najviac zaujímalo bola hodnota celkovej presnosti klasifikácie. Táto hodnota dosiahla číslo 8.1018%, čo značí, že predpoklad o nízkej presnosti klasifikácie sa naplnil.

211	10	15	416	4.52			
212	11	15	73111	794.57			
213	12	15	5958	64.75			
214	13	15	1341	14.57			
215	14	15	2119	23.03			
ERROR MATRIX [pixel count]							
		> Reference					
V_Classified	0	1	2	3	4	5	6
0	0	778	31646	1054	1541	17479	16338
1	0	471	24383	11666	2721	5107	13382
2	0	297	23429	2524	2119	6356	9351
3	0	0	382	153	233	144	785
4	0	20	3502	382	246	1364	31364
5	0	3030	10797	511	3802	5339	15102
6	0	32	30674	74	398	6464	23347
7	0	390	16719	283	1859	11583	17747
8	0	37	26356	118	270	8159	9512
9	0	499	23912	314	2122	5907	45027
10	0	181	6975	199	502	3173	40040
11	0	253	71345	1112	1331	104369	28065
12	0	131	18685	16	645	7281	88617
13	0	153	38482	329	856	20255	19557
14	0	233	8801	745	1683	1855	3922
15	0	640	80328	727	1219	35678	30595
Total	0	7145	416416	20207	21547	240513	392751
AREA BASED ERROR MATRIX							
		> Reference					
V_Classified	0	1	2	3	4	5	6
0	0	0.0005	0.0192	0.0006	0.0009	0.0106	0.0099
1	0	0.0003	0.0148	0.0071	0.0016	0.0031	0.0081
2	0	0.0002	0.0142	0.0015	0.0013	0.0039	0.0057
3	0	0	0.0002	0.0001	0.0001	0.0001	0.0005
4	0	0	0.0021	0.0002	0.0001	0.0008	0.0190
5	0	0.0018	0.0065	0.0003	0.0023	0.0032	0.0092
6	0	0	0.0186	0	0.0002	0.0039	0.0142
7	0	0.0002	0.0101	0.0002	0.0011	0.0070	0.0108
8	0	0	0.0160	0.0001	0.0002	0.0049	0.0058
9	0	0.0003	0.0145	0.0002	0.0013	0.0036	0.0273
10	0	0.0001	0.0042	0.0001	0.0003	0.0019	0.0243
11	0	0.0002	0.0433	0.0007	0.0008	0.0633	0.0170
12	0	0.0001	0.0113	0	0.0004	0.0044	0.0537
13	0	0.0001	0.0233	0.0002	0.0005	0.0123	0.0119
14	0	0.0001	0.0053	0.0005	0.0010	0.0011	0.0024
15	0	0.0004	0.0487	0.0004	0.0007	0.0216	0.0186
Total	0	0.0043	0.2525	0.0123	0.0131	0.1458	0.2381
Estimated area	0	77.65	4525.58	219.61	234.17	2613.88	4268.39
SE	0	0.0001	0.0003	0.0001	0.0001	0.0003	0.0003
SE area	0	0.91	5.89	1.46	1.57	4.62	5.54
95% CI area	0	1.78	11.55	2.87	3.08	9.06	10.86
PA [%]	nan	6.5920	5.6263	0.7572	1.1417	2.2198	5.9445
UA [%]	0.0000	0.6689	35.5071	8.1862	0.5202	7.3377	27.5292
Overall accuracy [%] = 8.1018							
Area unit = metre^2							
SE = standard error							
CI = confidence interval							
PA = producer's accuracy							
UA = user's accuracy							

Obr.26. Výpočet hodnôt presnosti klasifikácie.

Diskusia

V tejto bakalárskej práci sme sa použitím riadenej klasifikácie snažili čo najpresnejšie a s čo najväčšou úspešnosťou určiť druhové zloženie stromov v danom záujmovom území. Pri práci sme používali softvéry ako ArcGIS pro a QGIS, v programe QGIS prebehala príprava dát a ich úprava na následne použitie a spracovanie. Takisto sme v tomto softvéri vytvorili tréningové polygóny použité pri riadenej klasifikácii ale taktiež aj validačné polygóny, pomocou ktorých sme vypočítali presnosť samotnej klasifikácie. Okrem týchto dvoch programov, ktoré sme využili existujú aj iné softvéry, pomocou ktorých by sa dala riadená klasifikácia vykonať, avšak rozhodli sme sa využívať práve tieto dve, kvôli prehľadnosti a nadobudnutými skúsenosťami s prácou v týchto prostrediach.

Cieľom práce bolo na základe vytvorených trénovacích polygónov čo najpresnejšie určiť druhové zloženie stromov v území 2x2 kilometre. To sa nám podarilo, avšak s veľmi nízkou presnosťou. Pri našej práci kvalitu výsledkov ovplyvňovalo mnoho faktorov ako je veľkosť územia, prekryty korún stromov, hustá druhová rozmanitosť stromov na malom území alebo aj počet trénovacích polygónov.

Na dosiahnutie presnejších výsledkov by bolo ideálne záujmové územie rozdeliť na niekoľko menších častí, a postupnými krokmi ich klasifikovať, a následne ich spojiť do jedného veľkého rastrového vyobrazenia. Taktiež použitie viacerých tréningových polygónov, resp. použitie každého druhu stromu v danom území by pomohlo výrazne zvýšiť presnosť klasifikácie.

Záver

Touto prácou sme chceli poukázať na možnosti použitia metód strojového učenia, konkrétne riadenej klasifikácie pri určovaní druhového zloženia stromov na väčších územiach bez nutnosti vykonávať klasifikáciu manuálne v teréne. Existuje množstvo programov, ktoré dokážu riadenú klasifikáciu vykonať a to napríklad ENVI, MultiSpec, QGIS, ArcGIS pro a iné, my sme sa však rozhodli pracovať so softvéromi QGIS a ArcGIS pro. Tieto programy poskytujú množstvo voľne dostupných pluginov, ktoré sú nápomocné či už pri samotnej klasifikácii, ale aj pri predpríprave a následnej analýze dát.

Ako prvé sme museli vytýčiť záujmové územie, zamerali sme sa na územie mesta Košice. Pri výbere dát, na ktorých riadenú klasifikáciu chceme vykonať sme sa rozhodovali medzi RGB a CIR snímkami. Rozhodli sme sa pracovať s CIR snímkami, pretože je to veľmi efektívna metóda pri určovaní druhovej rozmanitosti vegetácie. Pracovali sme s rasterovými CIR snímkami v rozlíšení 10x10cm. Následne sme museli vytvoriť polygóny korún stromov, na základe ktorých bude klasifikácia prebiehať. Celkovo sme vytvorili 107 trénovacích polygónov pre 11 druhov stromov. Po exportácii dát vo vhodných formátoch a po dokončení príprav sme mohli začať vykonávať klasifikáciu.

Pre vykonanie riadenej klasifikácie sme využili program ArcGIS pro, a metódu klasifikácie Support Vector Machine. Celkový čas, ktorý klasifikácia zabrala bol viac ako 24 hodín a to kvôli obrovskému objemu dát. Po dokončení klasifikácie sme dáta

vizuálne upravili pre väčšiu prehľadnosť a ľahšiu analýzu. Následne sme bližšie zanalyzovali oblasti, v ktorých sa tréningové polygóny vytvárali. Na prvý pohľad už bolo zrejmé, že klasifikácia veľmi presná nebola. Po zanalyzovaní celého územia sme boli pripravení vypočítať presnosť klasifikácie.

Na výpočet presnosti sme využili plugin SCP a konkrétnu funkciu Accuracy. Ako vstupné dáta sme využili klasifikované územie, a ako validačné dáta sme využili časť polygónov stromov, ktoré sme použili pri klasifikácii, ale takisto sme vytvorili aj nové. Po výpočte presnosti sme dospeli k hodnote 8,1018%, čo značí veľmi vysokú nepresnosť.

Napriek neuspokojivému výsledku klasifikácie druhového zloženia stromov, možno povedať, že pre budúcich autorov, ktorí sa budú venovať podobnej problematike sú tieto výsledky použiteľné. Pre dosiahnutie lepšieho a presnejšieho výsledku budú mať poznatky o tom, čomu sa vyvarovať a aké metódy použiť. Skúmané územie je obrovských rozmerov a nachádzajú sa tam stovky stromov a takisto množstvo druhov stromov, preto dosiahnuté výsledky a zistenia je potrebné brať do úvahy pre budúce práce.

Zoznam použitej literatúry

REHÁČKOVÁ, T., PAUDITŠOVÁ, E., 2006: *Vegetácia v urbánnom prostredí*. Dostupné

online: https://www.researchgate.net/profile/EvaPauditsova/publication/289406406_Vegetacia_v_urbannom_prostredi/links/568c94d108ae153299b66ccd/Vegetacia-v-urbannom-prostredi.pdf

LILLESAND, KIEFER, CHIPMAN, 2008: *Remote Sensing and Image Interpretation*. NY: Wiley, 756 s.

DOBROVOLNÝ, P., 1998: *Dálkový průzkum Země, Digitální zpracování obrazu*. Brno (Masarykova Univerzita).

STRAKA. R., 2023: *Čo je LiDAR?*. Dostupné online: <https://skymove.sk/skenovanie-technologiu-lidar/>

WASSER. A., 2024: *What is a CHM, DSM and DTM? About Gridded, Raster LiDAR Data*. Dostupné online: <https://www.neonscience.org/resources/learning-hub/tutorials/chm-dsm-dtm>

AEVEX AEROSPACE, 2024: *DEM, DSM & DTM: DIGITAL ELEVATION MODEL – WHY IT'S IMPORTANT*. Dostupné online: <https://geodetics.com/dem-dsm-dtm-digital-elevation-models/>

COURSERA., 2024: *What Is Machine Learning? Definition, Types and Examples*.

Coursera.org, Dostupné online: <https://www.coursera.org/articles/what-is-machinelearning>

IBM: *What is machine learning (ML)?* Dostupné online:

<https://www.ibm.com/topics/machine-learning>

IBM: *What is deep learning?* Dostupné online: <https://www.ibm.com/topics/deep-learning>

ESRI, 2021: *What is image classification?* Dostupné online: <https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/what-is-image-classification-.htm>

LUXCARTA TEAM, 2024: *DEM vs DTM vs DSM: know the difference between key mapping concepts*. Dostupné online: <https://www.luxcarta.com/blog/dem-dtm-dsm>

ABER. S., 2010: *Photogrammetry*. Dostupné online: <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/photogrammetry>

MUBANGA. K., 2024: *What is photogrammetry*. Dostupné online: <https://www.artec3d.com/learning-center/what-is-photogrammetry>

FBE, 2023: *Machine learning (strojové učenie)*. Dostupné online: <https://fbe.sk/machine-learning-strojove-ucenie/>

FBE, 2023: *Deep learning (hlboké učenie)*. Dostupné online: <https://fbe.sk/deep-learning-hlboke-ucenie/>

EŠTOK, P. 2018: *Dolovanie znalostí a strojové učenie*. Časopis Katedry počítačov a informatiky FEI TUKE. [cit. 2024-04-22]. dostupné online: <https://magazin.kpi.fei.tuke.sk/2018/05/dolovanie-znalosti-a-strojove-ucenie/>

SAJID, H. 2023 : *Strojové učenie vs. hlboké učenie – kľúčové rozdiely*. Dostupné online: <https://www.unite.ai/sk/machine-learning-vs-deep-learning-key-differences/>

SPANO, M. 2020: *UMELÁ INTELIGENCIA (NIELEN) PRE PODNIKANIE, ČASŤ 5: HLBOKÉ UČENIE A UMELE NEURÓNOVÉ SIETE*. Dostupné online: <https://www.kros.sk/blog/umela-inteligencia-nielen-pre-podnikanie-cast-5-hlboke-ucenie-a-umele-neuronove-siete/>

SPATIALPOST,2023: *Difference Between Supervised and Unsupervised Classification In Remote Sensing*. Dostupné online: <https://www.spatialpost.com/supervised-vs-unsupervised-remote-sensing/>

CHAOVALIT, P., and ZHOU, L.,2005:*Comparison between Supervised and Unsupervised Classification Approaches*,

DONALD,E.,WEIH, R.,2005:Integrating Supervised and Unsupervised Classification Methods to Develop a More Accurate Land Cover Classification

F.E. FASSNACHT, H. LATIFI, K. STEREŃCZAK, A. MODZELEWSKA, M.

LEFSKY, L.T. WASER, C. STRAUB, A. GHOSH 2016: *Review of studies on tree*

species classification from remotely sensed data. Remote Sens. Environ

CONGEDO, L., 2014. *Semi-Automatic Classification Plugin Documentation.*