



**Hodnocení krajinného pokryvu Vojenského újezdu Libavá s použitím
klasifikátoru Random Forest**

**Evaluation of the land cover in the military training area Libavá
using Random Forest classifier**

Bakalářská práce

Vít Žďánský

Vedoucí práce: doc. RNDr. Přemysl Štych, Ph.D.

Bakov nad Jizerou, 2019

Vysoká škola: Univerzita Karlova
Katedra: Aplikované geoinformatiky a kartografie

Fakulta: Přírodovědecká
Školní rok: 2018/2019

Zadání bakalářské práce

pro Víta Žďánského

obor Geografie a kartografie

Název tématu: Hodnocení krajinného pokryvu Vojenského újezdu Libavá s použitím klasifikátoru Random Forest

Zásady pro vypracování

Bakalářská práce se zabývá hodnocením přesnosti klasifikátorů Random Forest (RF) a Maximum Likelihood (ML) pomocí družicových dat Sentinel-2 v oblasti Vojenského újezdu Libavá. Tento vojenský újezd prošel velmi specifickým vývojem se specifickým dopadem na změny a stav místní krajiny.

Hlavním cílem této bakalářské práce je klasifikace land cover (krajinného pokrytí) Vojenského újezdu Libavá pomocí volně dostupných družicových dat Sentinel-2 z roku 2019. Za tímto účelem bude použit klasifikátor Random Forest (RF) dostupný v open source programu SNAP. Použité metody a dosažené výsledky budou porovnány s tradičně používaným klasifikátorem Maximum Likelihood. Budou vyhodnoceny výsledky a přesnosti klasifikací obou klasifikátorů a vhodnost a perspektivy použitých družicových dat Sentinel-2. Součástí práce bude též návrh a použití klasifikačního systému vytvořeného pro řešení vojenský újezd.

V úvodu práce bude sepsána podrobná rešerše dostupné literatury a použitých metod. Bakalářská práce vyhodnotí a interpretuje dosažené výsledky. Na závěr budou shrnuty dosažená zjištění a poznatky a diskutovány použité metody.

Rozsah grafických prací: cca 10 stran

Rozsah průvodní zprávy: cca 40 – 50 stran

Seznam odborné literatury:

BELGIU, M., DRAGUT, L. (2016): Random Forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 8 s.

BREIMAN, L. (2001): Random Forest. Statistics Department, University of California, Berkley, CA, 33 s.

NOI, P. T., KAPPAS, M. (2017): Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18, 20 s.

SKOKANOVÁ, H., HAVLÍČEK, M., KLUSÁČEK, P., MARTINÁT, S. (2017): Five military training areas – five different trajectories of land cover development? Case studies from the Czech Republic. Geographie Cassoviensis XI, 13 s.

Vedoucí bakalářské práce: doc. RNDr. Přemysl Štych, Ph.D.

Datum zadání bakalářské práce: 8. 12. 2016

Termín odevzdání bakalářské práce: 31. 7. 2019

.....
Vedoucí bakalářské práce

.....
Vedoucí katedry

V Praze dne 30/10/2018

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Bakove nad Jizerou dne 8. 2. 2019

Poděkování

Rád bych poděkoval vedoucímu bakalářské práce doc. RNDr. Přemyslu Štychovi, Ph.D., za cenné rady, ochotu a věnovaný čas při zpracování této bakalářské práce. Dále bych rád poděkoval rodině za to trápení, kterým si se mnou procházeli.

Abstrakt

Data land cover nám pomáhají lépe pochopit krajinu, jak se vyvíjí, její využití a vliv lidské činnosti na ni. Díky novým metodám v oblasti DPZ jsme schopni tyto procesy zaznamenávat rychleji a v širším měřítku, než tomu bylo dříve. Tato práce se zabývá hodnocením přesnosti klasifikátorů Random Forest (RF) a Maximum Likelihood (ML) pomocí družicových dat Sentinel-2 v oblasti Vojenského újezdu Libavá. Právě vojenské újezdy prošly velmi specifickým vývojem a údaje o krajinném pokrytí chybí. Klasifikační legenda obsahuje 8 tříd. Výsledky klasifikace obou algoritmů jsou vyšší než 80 %. Lepších výsledků bylo dle očekávání dosaženo použitím klasifikátoru Random Forest. Nejpresněji byly klasifikovány třídy vodních ploch a listnatých a jehličnatých lesů, nejhůře orná půda a řídká vegetace. Ostatní třídy se svou přesností odlišovaly. Výsledky práce jsou hodnoceny pomocí chybových matic, celkové přesnosti a kappa koeficientu.

Klíčová slova: klasifikace, Random Forest, Maximum Likelihood, vojenský újezd, dálkový průzkum země, Sentinel 2, krajinné pokrytí, Libavá

Abstract

Data land cover help us understand nature, how it develops, its uses and the influence that human actions have on it. Thanks to new methods in the remote sensing area, we can record these processes faster and at a larger scale than before. This thesis evaluates accuracy of the Random Forest (RF) and Maximum Likelihood (ML) classifiers using satellite data Sentinel-2 from the military training area Libavá. The military area went through a very specific development and the information regarding natural coverage in the region is missing. The classifier documentation contains 8 classes. The classification results from both algorithms are higher than 80 %. As expected, more accurate results were achieved using the Random Forest classifier. The most accurate classifications were of water surfaces and forests. The least accurate classifications were of agricultural land and sparse vegetation. Other classes varied in accuracy levels. This thesis' results are evaluated using error matrices, overall accuracy and the kappa coefficient.

Keywords: classification, Random Forest, Maximum Likelihood, military training area, remote sensing, Sentinel 2, land cover, Libavá

Obsah

Seznam použitých obrázků.....	6
Seznam použitých tabulek.....	6
Seznam použitých grafů	6
1 Úvod.....	9
2 Literární rešerše a úvod do problematiky.....	10
2.1 Dálkový průzkum Země	10
2.2 Elektromagnetické záření	12
2.2.1 Spektrální informace	13
2.2.2 Spektrální odrazivost.....	13
2.3 Klasifikace	15
2.3.1 Neřízená klasifikace	16
2.3.2 Řízená klasifikace	16
2.3.3 Objektová klasifikace	19
2.3.4 Random Forest	21
3 Zájmové území.....	26
3.1 Vojenský újezd	26
3.1.1 Vojenský újezd Libavá.....	27
4 Data a metodika.....	28
4.1 Data.....	28
4.1.1 Sentinel 2.....	28
4.1.2 ORTOFOTO ČR	30
4.1.3 ZABAGED.....	30
4.1.4 LPIS.....	30
4.1.5 ArcČR 500.....	30
4.2 Software.....	31
4.3 Metodika	31
4.3.1 Předzpracování	31
4.3.2 Land cover a klasifikační legenda.....	32
4.3.3 Trénovací plochy.....	36
4.3.4 Klasifikace.....	38
4.3.5 Post-klasifikační úpravy a hodnocení přesnosti klasifikace.....	40

5	Výsledky	41
5.1	Hodnocení přesnosti klasifikací.....	41
5.1.1	Podíly a rozložení jednotlivých tříd land cover na území VÚ Libavá.....	44
6	Diskuze a závěr	47
	Použité zdroje a literatura.....	50
	Seznam příloh.....	55

Seznam použitých obrázků

Obrázek 1: Proces sběru dat dálkového průzkumu Země.....	11
Obrázek 2: Elektromagnetická vlna v prostoru	12
Obrázek 3: Elektromagnetické spektrum.....	13
Obrázek 4: Spektrální křivky odrazivosti (voda, vegetace, půda).....	14
Obrázek 5: Vybrané klasifikátory řízené klasifikace.....	17
Obrázek 6: Neuronová síť – Uspořádání neuronů do vrstev	19
Obrázek 7: Využití klasifikátorů v letech 2007 – 2017	21
Obrázek 8: Základní schéma principu fungování algoritmu Random Forest.....	22
Obrázek 9: Poloha Vojenského újezdu Libavá.....	27
Obrázek 10: Vojenský újezd Libavá (2019).....	30
Obrázek 11: Vodní plochy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	32
Obrázek 12: Listnaté lesy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	33
Obrázek 13: Jehličnaté lesy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	33
Obrázek 14: Travnatý porost Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	34
Obrázek 15: Řídká vegetace Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	34
Obrázek 16: Orná půda Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	35
Obrázek 17: Bez vegetace Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	35
Obrázek 18: Zástavba Sentinel 2 (vlevo) a Ortofoto ČR (vpravo).....	36
Obrázek 19: Výsledek klasifikace Random Forest a Maximum Likelihood.....	45
Obrázek 20: Detailní výřez výsledků klasifikací RF a ML	46

Seznam použitých tabulek

Tabulka 1: Specifikace Sentinel 2A.....	29
Tabulka 2: Trénovací plochy	37
Tabulka 3: Oddělitelnost jednotlivých tříd	37
Tabulka 4: Testované parametry klasifikátoru RF	38
Tabulka 5: Testované parametry klasifikátoru ML	39
Tabulka 6: Chybová matice klasifikace Random Forest	42
Tabulka 7: Chybová matice klasifikace Maximum Likelihood.....	43

Seznam použitých grafů

Graf 1: Podíl jednotlivých tříd krajinné pokrývky ve VÚ Libavá.....	44
---	----

Seznam použitých zkratek

ANN	Artificial Neural Networks
BOA	Bottom of Atmosphere
CART	Classification and Regression Tree
CLC	CORINE Land Cover
ČR	Česká republika
ČUZK	Český úřad zeměměřičský a katastrální
DPZ	Dálkový průzkum Země
ERTS-1	Earth Resources Technology Satellite 1
ESA	Evropská vesmírná agentura
ESRI	Environmental Systems Research Institute
GMES	Global Monitoring for Environment and Security
GPS	Global Positioning System
IRECI	Inverted Red-Edge Chlorophyll Index
LC	Land Cover
LPIS	Land Parcel Identification System
ML	Maximum Likelihood
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
NDVI	Normalised Difference Vegetation Index
NDWI	Normalized Difference Water Index
NDBI	Normal Difference Built-up Index
NN	Neural Networks
OBIA	Object-based Image Analysis
RF	Random Forest
ROI	Region of interests
S2rep	Sentinel 2 Red-Edge Position Index

SAR	Synthetic Aperture Radar
SHP	Shapefile
SNAP	Sentinel Application Platform
SVM	Support Vector Machine
UMT	Universal Transverse Mercator
USGS	United States Geological Survey
VÚ	Vojenský újezd
WMS	Web Map Service
WGS	World Geodetic System
ZABAGED	Základní báze geografických dat

1 Úvod

Odlišné způsoby hospodaření s krajinou Česka v minulosti ovlivnily nynější vzhled a využití území. Oblastmi, které se svým vývojem liší od okolního prostředí jsou vojenské újezdy. Ty byly od svého vzniku (1927) uzavřeny pro veřejnost a sloužily především pro vojenské účely, které specifickým způsobem určovalo vývoj ekosystémů v území. Na rozdíl od území vně újezdů zde neprobíhaly žádné ekonomické nebo sociální nátlaky, které by krajinu na území znehodnocovaly, a i díky tomu je společným znakem vojenských újezdů v Česku velká biodiverzita (Skokanová a kol 2017). Území vojenských újezdů bývala kvůli nepřístupnosti pro veřejnost a zkreslováním map zahalena tajemstvím. Tento stav se změnil po konci studené války, kdy docházelo ke snižování stavu armády a tím i rušení nebo zmenšování rozlohy. Prvním zrušeným vojenským újezdem byl v roce 1993 VÚ Ralsko. V současné době tento trend zpřístupňování újezdů pokračuje. To otevírá nové možnosti ve studiu změn krajiny.

Jednou z metod, jak krajinné pokrytí zkoumat je pomocí dálkového průzkumu země. Metoda dálkového průzkumu se za posledních několik desítek let stala velmi populární při získávání informací o zemském povrchu a jeho okolí. Dříve byl sběr informací omezen pouze na terénní průzkum, který byl časově i finančně náročný. S rozvojem techniky dálkového průzkumu se začaly rozšiřovat i nové techniky a metody.

Díky dálkovým datům o krajinném pokryvu jsme schopni do určité míry předpovídat a monitorovat přírodní katastrofy (Doris a kol. 2010), sledovat globální změny a v případě potřeby reagovat (Quintano a kol. 2017)

Hlavním cílem této bakalářské práce je klasifikace land cover (krajinného pokrytí) Vojenského újezdu Libavá pomocí volně dostupných družicových dat Sentinel 2 z roku 2019. Za tímto účelem bude použit klasifikátor Random Forest (RF) v procesu klasifikace využitím programu SNAP. Výsledky krajinného pokryvu budou porovnány s klasifikátorem Maximum Likelihood. Nedílnou součástí této klasifikace bude navržení vhodné legendy land cover. V závěru bude vyhodnocení přesnosti obou klasifikátorů.

Práce je členěna do dvou hlavních částí. První, teoretická část, se zabývá úvodem a seznámením s problematikou. Jsou zde vysvětleny základní principy dálkového průzkumu země, představen proces klasifikace a klasifikátory použité v této práci, spektrální projevy základních krajinných prvků a stručně charakterizovaná zkoumaná oblast. Druhá polovina práce je zaměřena na praktickou část, ve které budou přiblížena data, využitá při zpracování, použitý software a metody postupu při zpracování jednotlivých cílů. V samotném závěru práce jsou prezentovány výsledky práce, závěr a diskuse. Každá část je pro větší přehlednost členěna do kapitol.

2 Literární rešerše a úvod do problematiky

2.1 Dálkový průzkum Země

Dálkový průzkum je metoda, při které získáváme kvalitativní a kvantitativní informace o objektech a jevech bezkontaktními metodami (Halounová, Pavelka 2008). Z geografického pohledu používáme pojem „Dálkový průzkum Země“ (dále jen DPZ) zabývající se sbíráním informací o zemském povrchu nebo jeho blízkém okolí. Vymezení DPZ a se zabývalo mnoho autorů. Definice se liší rozsahem, chápáním pojmu nebo dobou, ve kterém vznikla.

„Dálkový průzkum je způsob získávání informací o zemském povrchu a vodních plochách s využitím snímků pořízených z ptačí perspektivy, využívá elektromagnetického záření v jednom nebo více intervalech spektra, toto záření je odraženo nebo emitováno ze zemského povrchu.“ (Campbell 2011)

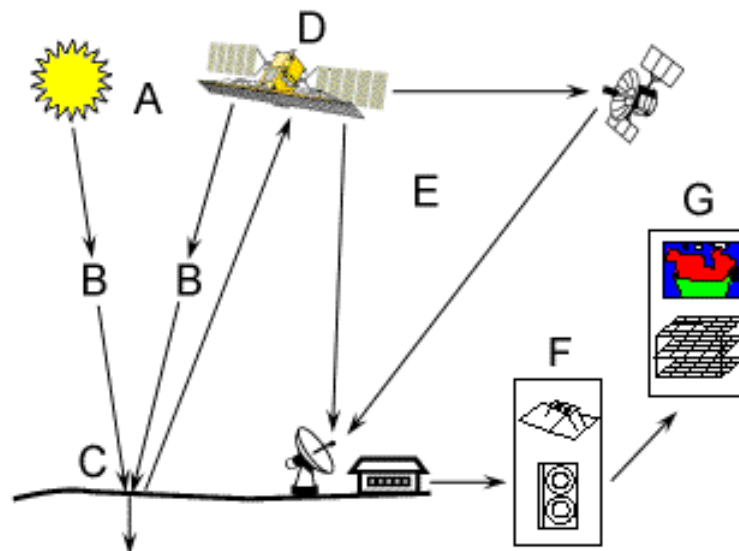
„Získávání informací o zemském povrchu na dálku pomocí snímacích zařízení, umístěných obvykle v letadlech nebo družicích. Zahrnuje jednak samostatné snímání, jednak vyhodnocování pořízených záznamů.“ (Čapek 1988)

„Dálkový průzkum je věda a umění získávat užitečné informace o objektech, plochách či jevech prostřednictvím dat měřených na zařízeních, která s těmito zkoumanými objekty, plochami ani jevy nejsou v přímém kontaktu.“ (Lillesand, Kiefer, 1994, Cit. in: Dobrovolný 1998)

Data DPZ mají velmi široké využití ve spoustě různých odvětvích – tvorba map, posuzování škod na životním prostředí, monitorování využívání půdy, monitorování radiace, územní plánování, hodnocení kvality půdy nebo hodnocení výnosu plodin (Sisodia a kol. 2014), na poli geografie (meteorologii a klimatologii, hydrologie, geomorfologii) a jiných odvětvích – environmentalistice, ekonomii, bezpečnosti, zemědělství nebo astronomii. Jednou z výhod využívání dálkového průzkumu je rychlost zpracování dat. Ve velmi krátkém okamžiku zachytí daleko více detailů než při běžném pozemním průzkumu, který je časově i finančně náročný (Dobrovolný 1998). Získávání dat z DPZ se dnes neomezuje jen na cesty přímo vnímané lidskými smysly, ale i na jevy a údaje, které jsou pro člověka pouhým okem nezjistitelné (Kolář 1990).

Nositelem informace je v dálkovém průzkumu Země elektromagnetické záření (viz kapitola 2.2), pomocí kterého je možné zkoumat i energii nepostřehnutelnou lidským okem. Elektromagnetické záření, které dopadá na zkoumaný objekt je pohlceno nebo odraženo zpět do okolí. Intenzitu vyzářeného nebo odraženého záření objektu detekuje senzor, umístěný na různých typech nosičů, nejčastěji balónech, letadlech a umělých družicích (Dobrovolný 1998). Proces získávání dat se skládá z několika základních částí (Obrázek 1). Základem je elektromagnetické záření vyzářené aktivním povrchem Země nebo odražené záření Slunce nebo umělé družice. Na základě vlastností jednotlivých objektů na zemském povrchu dojde k odražení záření. To je zaznamenáváno speciálními přístroji umístěnými na letadlech nebo družicích. Změřená data jsou přenesena zpět na zemský povrch a zpracována podle potřeb výzkumné otázky.

Obrázek 1: Proces sběru dat dálkového průzkumu Země



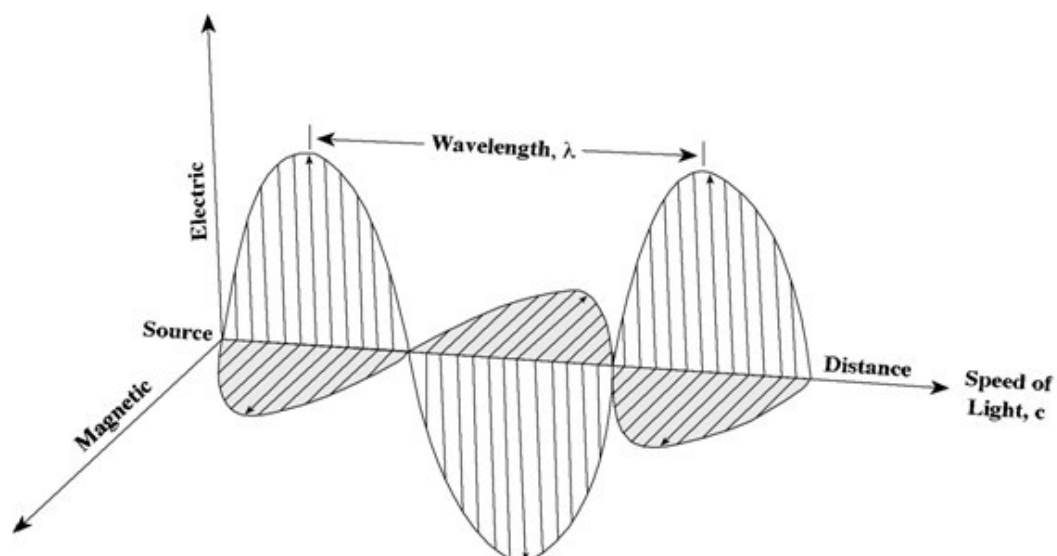
Zdroj: CCRS (2018). A: zdroj záření, B: záření v atmosféře, C: interakce s objektem, D: změření záření přístrojem, E: přenos, příjem dat, F: zpracování dat, G: využití informace v aplikaci

2.2 Elektromagnetické záření

Jak bylo zmíněno výše, dálkový průzkum Země je založený na sbírání údajů o objektech a jevech, které nepřichází do přímého kontaktu a jsou měřeny na větší vzdálenost. Typické je, že při získávání informací o objektech se neomezuje jen na informace vnímané lidskými smysly, ale i ty, které jsou sbírané pomocí přístrojů. Tímto způsobem je možné získávat informace o silovém poli zkoumaných objektů v číselné podobě a převádět do podoby přijatelné pro člověka (Kolář 1990). Informace o silových polích objektů získáváme měřením elektromagnetického záření, jež je považován za nositele informace v dálkovém průzkumu.

Základním principem je, že každý objekt na zemském povrchu vyzařuje elektromagnetické záření a zároveň odráží záření vyzařované jinými objekty. Sledováním tohoto záření je možné získávat informace o všech objektech na zemském povrchu (Campbell 2011). Záření se šíří prostorem prostřednictvím elektromagnetické vlny (Obrázek 2), složené ze dvou částí – sinusoidy elektrické vlny a magnetické vlny, které svírají pravý úhel. Základními charakteristikami elektromagnetického záření je vlnová délka a frekvence. (Dobrovolný, 1998).

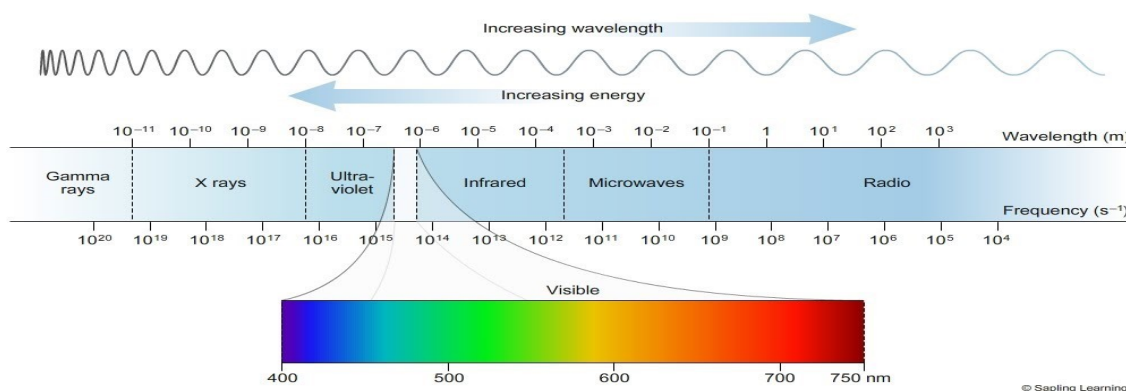
Obrázek 2: Elektromagnetická vlna v prostoru



Zdroj: Jensen (2017)

Rozsah vlnových délek je označován jako elektromagnetické spektrum, jež je rozděleno do několika částí (Obrázek 3) (Halounová a Pavelka 2008). Nejznámějším typem je viditelná část spektra (0,38-0,72 nm), na které je citlivé i lidské oko a jeho odrazy vnímáme jako barvy. Mezi další oblasti vlnových délek, které se začalo rozšiřovat až s pokrokem techniky, patří gama záření (<0,03 nm), x-záření (0,03-300 nm), ultrafialové záření (0,3-0,38 μm), infračervené záření (0,72-1000 μm), mikrovlnné (1 mm-30 cm) a radiové (≥30 cm) (Dobrovolný 1998).

Obrázek 3: Elektromagnetické spektrum



Zdroj: Sapling Learning (2018) dostupné na: <https://sites.google.com/site/chempendix/em-spectrum>

2.2.1 Spektrální informace

Každý satelitní snímek pořízený jinou družicí je specifický svými parametry. Těmi jsou prostorové rozlišení, spektrální rozlišení a časové rozlišení. *Prostorové rozlišení* je rozsah nejmenšího úhlového nebo lineárního odlišení mezi dvěma objekty, které mohou být rozlišeny (Jensen 2007). Prostorové rozlišení definuje, jak velká plocha na zemském povrchu odpovídá jednomu pixelu na snímku (GEO COPERNICUS). Družice jsou schopny pořizovat snímky s prostorovým rozlišením několika desítek cm až kilometrů. *Spektrální rozlišení* je počet a šířka specifických intervalů vlnových délek v elektromagnetickém spektru, na který je senzor citlivý (Jensen 2007). Lze rozlišovat multispektrální data (několik jednotek pásem) a hyperspektrální data (několik desítek až stovek pásem). *Časové rozlišení* lze definovat jako čas, za který je schopen senzor pořádit snímky stejného území.

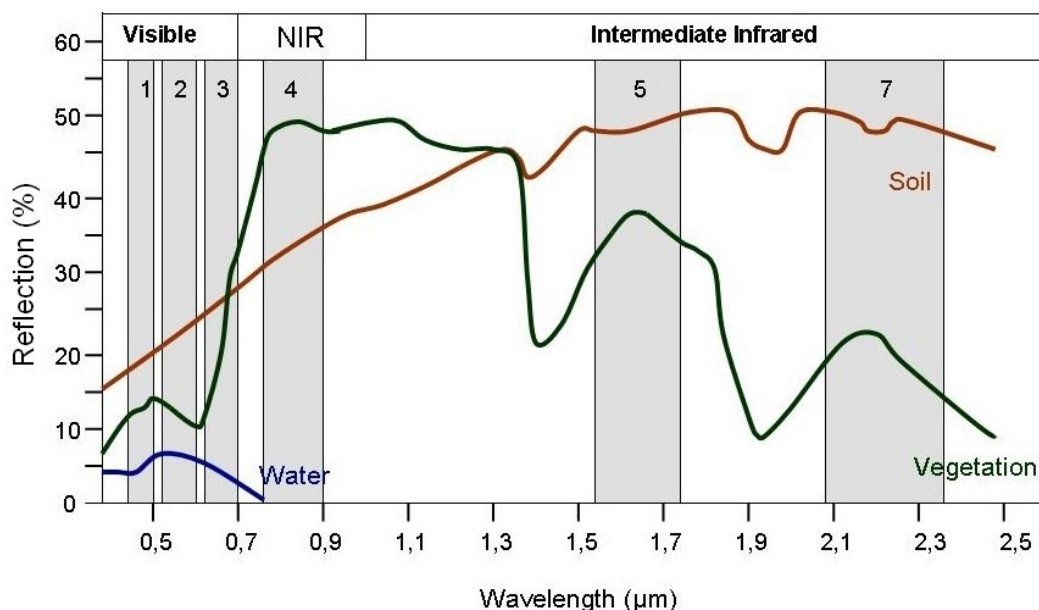
2.2.2 Spektrální odrazivost

Spektrální odrazivost je poměr intenzity odraženého záření a intenzity záření dopadajícího na určité vlnové délce (Dobrovolný 1998).

Sluneční záření, které projde atmosférou je při kontaktu s objekty na zemském povrchu buď absorbováno nebo odraženo zpět do prostředí. Způsob, jakým reaguje povrch na záření ovlivňují jeho odrazové vlastnosti. Ty závisí na fyzickém stavu, chemickém stavu a na okolí objektu (GSP 216 2017). Pod fyzickými vlastnostmi si můžeme představit například hrubost povrchu nebo obsah vody. Chemické mohou být rozlišeny podle obsahu minerálů. Každá látka na zemském povrchu má svou charakteristickou odrazivost. Odrazivost se mění s vlnovou délkou záření. Proto i objekty, které jsou si svou odrazivostí podobné v jednom intervalu vlnových délek, mohou mít různou odrazivost v jiném intervalu (Dobrovolný 1998). Závislost odrazivosti na vlnové délce ukazuje spektrální křivka odrazivosti (Obrázek 4). Křivka by v ideálním případě měla přesně identifikovat a rozlišit jednotlivé typy povrchů. V reálném prostředí tento

stav nenastává, neboť je povrch ovlivňován okolními – například teplotou, vlhkostí, oblačností a typicky u vegetace také ročním obdobím (Campbell 2011).

Obrázek 4: Spektrální křivky odrazivosti (voda, vegetace, půda)



Zdroj: SEOS (2018) dostupný na: <http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html>

V následující části budou definovány 3 základní typy povrchů, které mají své specifické vlastnosti a navzájem se liší. Patří mezi ně půda, voda a vegetace.

Voda, jako příměs ostatních složek krajiny (vegetace, půdy apod.) značně ovlivňuje ostatní objekty krajiny, avšak sama je, co se odrazivosti týče, jednoduchým, homogenním objektem (Kolář 1990). V kapalném skupenství se vyznačuje velmi nízkou odrazivostí na všech vlnových délkách. Záření dopadající na vodní plochu proniká hluboko pod hladinu a je z větší části absorbováno. Maximální odrazivost vody nastává ve viditelné části spektra.

Vegetační povrch je téměř vždy zastoupen na snímcích dálkového průzkumu Země. Přesto, že na snímcích vypadají vegetační plochy svou barvou homogenně, jedná se o velmi různorodý povrch (lesy, zemědělské plodiny, louky, lesy atd.). I přes různorodost vykazuje vegetace podobný průběh spektrální odrazivosti. Obecná vegetační křivku odrazivosti (Obrázek 4) se podle jejího typického průběhu dělí do tří částí: oblast pigmentační absorpce (0,4-0,7 μm), oblast buněčné struktury (0,7-1,3 μm) a oblast vodní absorpce (1,3-3,0 μm). (Kolář 1990). Největší procento odrazivosti a zároveň největší rozdíly odrazivosti jednotlivých druhů vegetace nastávají v oblasti buněčné struktury. Klasickým příkladem je rozlišení jehličnatých a listnatých stromů. Listnaté stromy mají vyšší spektrální odrazivost než stromy jehličnaté (Dobrovolný 1998). Hlavními faktory, které formují a odlišují úroveň odrazivosti jednotlivých druhů rostlin jsou především jejich chemické a fyzikální vlastnosti (Fassnacht a kol. 2016): vnější uspořádání

vegetačního krytu, chemické vlastnosti listové tkáně, morfologie (tloušťka buněčných stěn, vzduchové mezibuněčné prostory a kutikula), struktura korun stromů (hustota listů a větví, úhel sklonu) (Leckie a kol. 2005), výška stromu, stáří stromu nebo zdravotní stav. Kromě vlivů, které napomáhají rozlišit jednotlivé druhy je množství odraženého záření znehodnocován okolními vlivy (půda, mechy, lišejníky).

Půdní typ povrchu je směs organických a anorganických látek, které se liší chemickými, biologickými nebo fyzikálními vlastnostmi. Kombinací těchto vlastností vytváří celkový spektrální projev. Obecné schéma spektrální křivky půdního povrchu naznačuje postupné zvyšování odrazivosti s rostoucí vlnovou délkou (Kolář 1990).

Místem, kde jsou uchovávány informace o spektrálním chování povrchů jsou spektrální knihovny. Uchovávají údaje o absolutních hodnotách odrazivosti stovek nejběžnějších druhů povrchů (minerálů, hornin, půd, fyzikálně konstruovaných i matematicky vypočítaných směsí, rostlin, vegetačních společenství, mikroorganismů a umělých materiálů) (USGS 2019). Tyto informace byly shromážděny za účelem využití spektrálních prvků pro dálkovou detekci. Největšími knihovnami jsou: spektrální knihovna ASTER od NASA (<https://speclib.jpl.nasa.gov/>) a USGS (<https://www.usgs.gov/labs/spec-lab>)

2.3 Klasifikace

Klasifikace je proces, při kterém je jednotlivým obrazovým prvkům (pixelům) podle určitého příznaku¹ a na základě rozhodovacího pravidla přiřazován určitý informační význam (Dobrovolný 1998), kde typ informace záleží na požadovaném výsledku. Pomocí porovnávání a přiřazování známých a neznámých obrazových prvků je možné pixely zařadit do shluků, které jsou spojeny informačními kategoriemi a podle potřeby řazeny do jednotlivých tříd. Pixely patřící do jedné třídy jsou teoreticky homogenní (Campbell 2011). Klasifikace je založena na použití klasifikačních pravidel (tzv. klasifikátorů), které mohou být obecně založeny na spektrálním chování, ale i na vlastnostech objektů a jevů v obraze. Podmínkou je, aby byly objekty v porovnávaných vlastnostech dostatečně odlišitelné (Kolář 1990). Ačkoliv existuje celá řada klasifikátorů, nelze říci, který je nejlepší. Výběr vhodného klasifikačního pravidla je základní předpoklad pro dosažení kvalitních výsledků. Závisí na výzkumné otázce, tréninkových datech, uživatelském využití, prostorovém a spektrálním rozlišení dat (Asamoahl a kol. 2017), klimatických podmínkách a dalších faktorech, které mohou výsledek ovlivnit. Častým řešením bývá kombinace více klasifikačních metod.

Klasifikátory dělíme na bodové (pixelové) a objektově orientované (Campbell 2011). Bodové (per-pixelové) klasifikátory jsou nejjednoduššími a také nejčastěji používanými

¹ Veličina, kterou můžeme pro jednotlivé pixely měřit a třídit (Kupková 2010).

klasifikátory založené na spektrálních vlastnostech jevů a objektů samotného pixelu. Bodové klasifikátory dělíme na řízené a neřízené, podle toho, jak zpracovatel zasahuje do procesu klasifikace (Dobrovolný 1998). Objektově orientované klasifikátory jsou složitější. Umožňují složitější klasifikační procesy, které mohou zahrnovat kontextuální a další informace o skupinách obrazových prvků (Campbell 2011).

2.3.1 Neřízená klasifikace

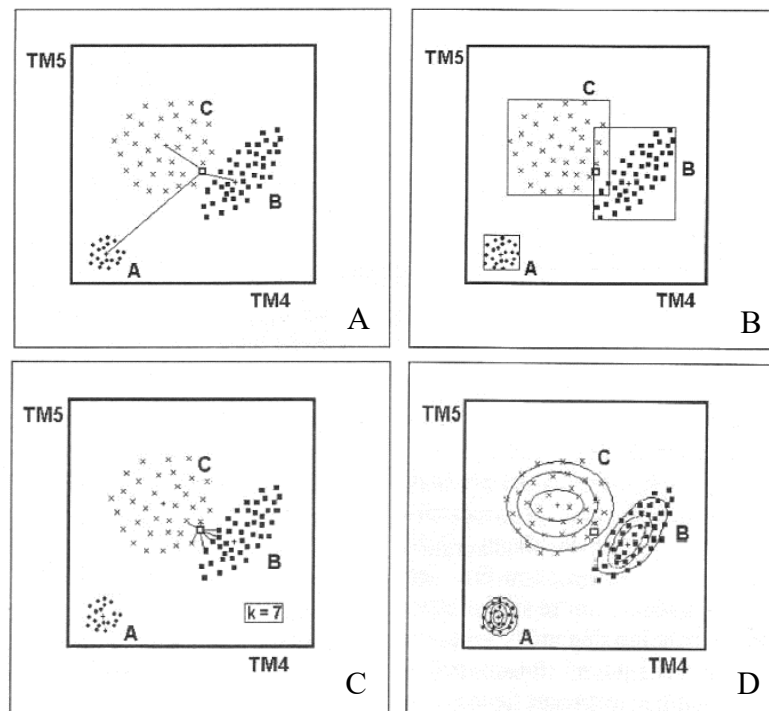
Při neřízené klasifikaci jde o dvoustupňový proces, při kterém nejsou využívána trénovací data. V prvním kroku jsou všechny obrazové prvky zařazeny automaticky podle spektrálních vlastností do tříd. V druhém kroku rozhoduje zpracovatel o informačním obsahu jednotlivých obrazových prvků. K odlišení shluků se používají shlukové analýzy (Campbell 2011). Výhodou oproti řízené klasifikaci je, že neřízená klasifikace toto spektrální chování objektů dokáže při vhodně zvolených analýzách roztrždit sama a nevyžaduje před samotnou klasifikací zásah zpracovatele (Sisodia a kol. 2014). Nejčastěji využívanými klasifikátory neřízené klasifikace jsou algoritmus K-průměrů (K-means) a ISODATA.

2.3.2 Řízená klasifikace

Při řízené klasifikaci zpracovatel využívá tzv. trénovacích ploch. To jsou části obrazu, u kterých může zpracovatel na základě dostupných dat definovat známý povrch. Pomocí trénovacích ploch je možné vygenerovat statistické charakteristiky spektrálních příznaků, se kterými je postupně porovnán každý obrazový prvek a podle vhodného klasifikačního pravidla zařazen do určité třídy. Tato pravidla jsou založena na předpokladu, že obrazové prvky představující určitou třídu se budou shlukovat v určité oblasti vícerozměrného příznakového prostoru. Pixely, které nenáleží do žádného ze shluků jsou následně přiřazeny podle klasifikačního pravidla (Dobrovolný 1998). Výběr klasifikátoru závisí na povaze a kvalitě vstupních dat a na požadovaném výstupu. Typickými klasifikátory používanými v DPZ jsou klasifikátor nejvyšší pravděpodobnosti, minimální vzdálenosti, nejbližšího souseda nebo klasifikátor pravoúhelníků (Campbell 2011). V následující kapitole budou některé z nich přiblíženy a pro lepší pochopení doplněny o obrázky.

Na obrázku č. 5. jsou středy shluků (centroidy) označeny znakem + a zařazované pixely znakem □. Jednotlivé shluky jsou označeny písmeny A, B a C.

Obrázek 5: Vybrané klasifikátory řízené klasifikace



Zdroj: Dobrovolný 1998

A. Klasifikátor minimální vzdálenosti (Minimum Distance)

Pixely, které svou polohou nenáležejí do žádného shluku jsou analyzovány a zjištěny jejich vzdálenosti od centroidů okolních shluků. Pixel je zařazen do toho shluku, kde je vzdálenost mezi ním a centroidem nejbližší. Nevýhodou tohoto pravidla je necitlivost algoritmu na různý stupeň rozptylu ve spektrálním chování, a proto se v takovýchto případech nepoužívá (Dobrovolný 1998).

B. Klasifikátor pravoúhelníků (Parallelepiped Classification)

Každému ze shluků jsou definovány pixely s minimální a maximální hodnotou, kterými jsou vedené rovnoběžné linie, ohraničující shluk. Pixely, které nepatří do žádného ze shluků jsou přiřazeny podle pravoúhelníku, ve kterém se po klasifikaci nachází (obrázek 9–B). Pixely, které se nacházejí ve více než jedné skupině mohou být klasifikovány podle jiného pravidla nebo je možné je zařadit podle priority jednotlivých tříd (tzn. podle předpokládané četnosti výskytu pro některou z tříd). Jinými možnostmi je definování nové smíšené třídy nebo definování pravoúhelníku na základě jiných hodnot než maximální a minimální (Dobrovolný 1998). Nevýhodou této metody je možnost velkého počtu vzdálených pixelů, které nejsou přiřazeny do žádné kategorie (Campbell 2011).

C. Klasifikátor nejbližšího souseda (k-Nearest Neighbors)

Klasifikátor nejbližšího souseda se podobá metodě minimální vzdálenosti s tím rozdílem, že u tohoto klasifikátoru je brána v úvahu nejen vzdálenost od centroidů, ale i početní zastoupení pixelů jednotlivé třídy v okolí nezařazeného prvku. Ten je poté zařazen do třídy, která mezi sousedy převažuje (Dobrovolný 1998).

D. Klasifikátor maximální pravděpodobnosti (Maximim Likelihoody)

Přiřazuje každý pixel v obraze k třídě s nejvyšší pravděpodobností nebo je označen jako nezařazený, pokud jsou hodnoty pravděpodobnosti pod prahovou hodnotou nastavenou uživatelem (Ahmad a Qeugan 2012).

Klasifikátor maximální pravděpodobnosti je nejvíce využívaný řízený klasifikátor (Sisodia a kol. 2014). Je založený na předpokladu, že shluky pixelů mají normální rozdělení. Při zařazování kvantitativně hodnotí rozptýl hodnot, korelaci a kovariaci každé třídy (Dobrovolný 1998). Berou v úvahu odchylky, které mohou být přítomny v jednotlivých třídách a bere v úvahu problémy, které vznikají při překrytí. Výhodou této metody je větší přesnost, ale je velmi citlivá na kvalitu trénovacích dat (Campbell 2011).

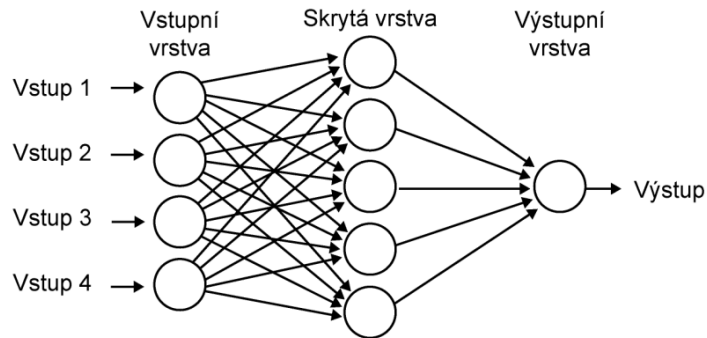
Podle studie Hagner a Reese (2007) je nevýhodou této metody to, že algoritmus při nedostatku dobrých referenčních dat předpokládá stejnou pravděpodobnost výskytu u všech tříd. Pokud se v tréninkových datech nepoužijí žádné váhy a četnost výskytu tříd se nebere v úvahu, může dojít k tomu, že jsou méně často se vyskytující třídy nadhodnoceny. Ve své studii upravuje tuto metodu a vytváří upravenou metodu maximální pravděpodobnosti „kalibrovaný klasifikátor maximální pravděpodobnosti“. Tato metoda je založená na opakované klasifikaci, kdy je z výsledků první klasifikace zjištěna četnost výskytu jednotlivých tříd a váha jednotlivých tříd je pak použita v opakované klasifikaci.

Neural Networks (NN)

NN je metoda strojového učení, navržena tak, aby napodobovala učící se procesy probíhající v nervové soustavě člověka prostřednictvím vytváření vazeb mezi vstupními a výstupními daty (Campbell 2011). Neuronová síť (Obrázek 6) se skládá z množství umělých neuronů, které jsou vzájemně propojeny, předávají si signál a mění ho na informaci. Neuron má neomezený počet vstupů, ale pouze jeden výstup. Neuronová síť se skládá ze *vstupní vrstvy*, která zajišťuje příjem vstupů zvenčí a rozděljuje a přenáší signálu do dalších vrstev. Následujících vrstvách (*skrytá vrstva*) dochází ke zpřesňování výpočtu, kde prvek sčítá všechny příchozí hodnoty vážených spojení a vyhodnotí výslednou hodnotu. Počet skrytých vrstev ovlivňuje celkovou přesnost výstupu (vyšší počet skrytých vrstev je schopný řešit komplexnější problémy), neplatí však pravidlo, že

čím více vrstev, tím lepší výsledek. Konec celé sítě představuje *výstupní vrstva* s jedním výstupem. Klasifikace využívá k trénovacím procesům zpětného šíření (Back propagation). Díky tomu je zmenšována chyba mezi vstupními a výstupními daty. Na základě trénovacích dat jsou pak upravovány váhy spojení mezi neurony (Atkinson a Tatnall 1997)

Obrázek 6: Neuronová síť – Uspořádání neuronů do vrstev



Zdroj: Matematickabiologie (2019); Dostupné z: <http://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologickyh-dat--umela-inteligence--neuronove-site-jednotlivy-neuron--uvod-do-neuronovych-siti--koncept-umele-neuronove-site>

Support Vector Machine (SVM)

Metoda SVM je neparametrická lineární metoda strojového učení. Klasifikační proces je založen na hledání ideální nadroviny (hyperplane, lineární oddělovací rovina), která by byla schopná rozdělit příznakový prostor tak, aby podle ní byla data rozdělena do klasifikačních tříd. Ideálně umístěná nadrovina je taková, kde nejmenší vzdálenost nejbližších bodů od roviny je co největší. Tyto body se nazývají support vectors (Mahesh 2011).

2.3.3 Objektová klasifikace

Zatímco bodové klasifikátory jsou založeny výhradně na spektrálních informacích o každém pixelu, klasifikace založená na objektech (OBIA) zkoumá i kontextové informace o krajině na základě spektrálních vlastností, barvy, velikosti, tvaru, textury, okolních objektech atd. a jejich vzájemných vztazích. Tento typ klasifikace se pokouší napodobit typ analýzy prováděné lidmi během vizuální interpretace (GSP 2015). Základní fází objektové klasifikace je segmentace. Při procesu segmentace jsou pixely v obraze seskupeny do segmentů (shluků pixelů) nebo objektů, které sdílejí homogenní spektrální nebo kontextovou podobnost. Způsobů, jakým je přistupováno k segmentaci je několik, například top-down (od největších objektů k menším), bottom-up (od nejmenších objektů k větším), boundary seeking, object seeking atd.

Důležitou částí před samotnou klasifikací je nastavení parametrů – definování jednotlivých vlastností a pravidel segmentů, podle kterých je klasifikátor schopný jednotlivé segmenty zařadit do příslušné třídy (Veljanovsky a kol. 2010). Klasifikační model používaný v objektových klasifikacích může být založen na tzv. tvrdých (např metoda rozhodovacího stromu) nebo měkkých klasifikátorech (Fuzzy model využívající příkazových operátorů a matematických operací) (GISAT 20??). Mezi další užívané klasifikátory patří SVM, neuronové sítě, rozhodovací stromy, Maximum Likelihood a další.

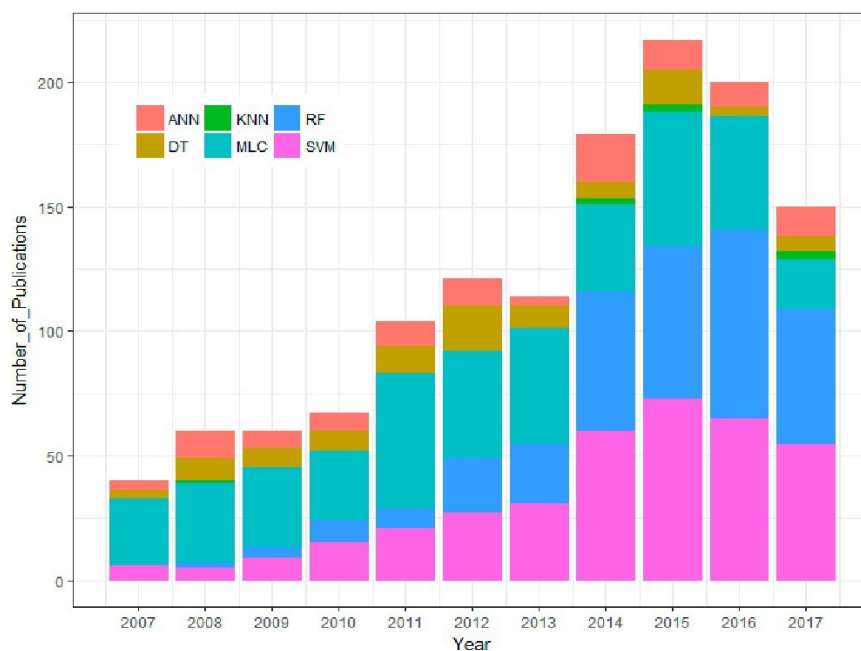
Veljanovsky a kol. (2010) se ve své studii zabývali teoretickým představením a metodologií objektové analýzy obrazu dat dálkového snímání. Poskytuje přehled o vývoji možnostech jeho využití objektové klasifikace i v jiných oborech. Mimo jiné se snaží poukázat na určitá omezení a výhody, které z použití objektové analýzy plynou.

Mezi nejvíce používané klasifikátory do nedávna patřily řízený parametrický klasifikátor maximální pravděpodobnosti a výše zmíněné neřízené klasifikátory K-means a ISODATA (Moore a Bauer 1990). Se zvyšujícím se prostorovým rozlišením se metody klasifikace založené na pixelech staly méně efektivní, protože vztah mezi velikostí pixelu a rozměrem pozorovaných objektů na zemském povrchu se významně změnil. Proto se stále více využívanější metodou stává objektově orientovaná klasifikace (Veljanovsky a kol. 2010). Také díky rozvoji výpočetní techniky (volně dostupné algoritmy, open source, programovací jazyka atd.) pro aplikace dálkového průzkumu země přijato několik algoritmů vyvinutých pro strojové učení (Fassnacht a kol. 2016). Mezi ně bychom mohli zařadit neparаметrické SVM, umělé neuronové sítě (ANN) nebo Random Forest (Niculescu a kol. 2018).

Tento trend ve své studii ukazují i Noi a Kappas (2017), kteří se mimo jiné zabývali četností využití jednotlivých klasifikátorů. V letech 2007-2017 (obrázek 7) patřili mezi nejvyužívanější klasifikátory právě MLC, SVM a RF. Z obrázku je patrný zvýšený trend používání SVM a RF. Oproti tomu je vysoká četnost využití MLC v průběhu let stabilní.

Podle autorů je dán vysoký počet příspěvků MLC tím, že je tento klasifikátor využíván jako jedna z kritérií pro porovnání algoritmů strojového učení.

Obrázek 7: Využití klasifikátorů v letech 2007 – 2017



Zdroj: Noi a Kappas (2017). ANN = Artificial Neural Network, KNN = *K-Nearest Neighbor*, RF = Random Forest, DT = Decision Tree, MLC = Maximum Likelihood Classifier, SVM = Support Vector Machine

Jednou z největších výhod, oproti klasickým neparametrickým metodám je, že u strojového učení není pro přesnou klasifikaci vyžadováno normálové rozdělení a jsou schopny se vypořádat i s nevyváženými daty. I proto se stávají stále populárnějšími pro klasifikaci dálkově snímaných dat (Belgiu a Dragut 2016).

V následující části bude stručně představena jedna z výše zmíněných metod – Random Forest klasifikátor, který byl i zvolen pro účely této bakalářské práce.

2.3.4 Random Forest

Random Forest (česky náhodný les, dále jen Random Forest nebo RF) je kombinovaná metoda (ensemble learning)² strojového učení určená pro klasifikaci a regresi. RF ve svém rozhodovacím klasifikačním procesu využívá velké množství rozhodovacích stromů s použitím náhodně vybrané podmnožiny tréninkových souborů a proměnných (Fawagheh a kol. 2014).

Za obecný základ pro metodu náhodného rozhodovacího lesa byl rok 1995, kdy Ho navrhl metodu, díky které bylo možné vytvoření složitější strukturu stromů bez negativního ovlivnění přesnosti klasifikace (Fawagheh a kol. 2014). Rozšířením

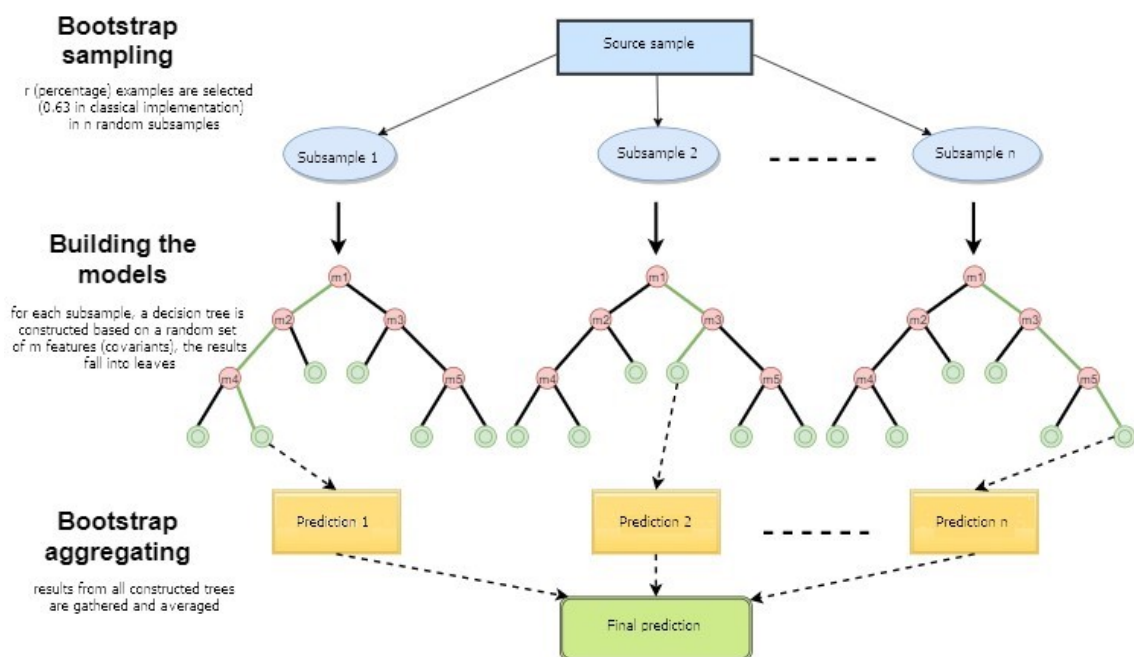
² Kombinovaný/souborový klasifikátor je ten, který kombinuje různé algoritmy strojového učení a následné agregování jejich výsledků za účelem přesnější klasifikace. Základní myšlenkou je spojení slabých klasifikátorů (stromů) ve skupinu silných (les).

algoritmu vyvinul v roce 2001 Leo Breiman metodu s názvem Random Forest. Metoda kombinuje Breimanovu metodu bagging vyvinutou v roce 1996 a metodu náhodného výběr vzorků představenou nezávisle Ho (1995, 1998) a Amitem a Gemanem (1997) (Fawaghech a kol. 2014). Amit a Geman navrhli přístup založený na rozpoznávání pomocí velkého množství geometrických tvarů (Breiman 2001).

Základním procesem (Obrázek 8) klasifikátoru je bagging (neboli Bootstrap aggregating). Bagging je meta-algoritmus strojového učení, jehož principem je vytváření velkého množství učících se podmnožin z tréninkových dat. Při baggingu na sobě nejsou jednotlivé podmnožiny závislé – každá je nezávisle konstruována pomocí bootstrappingu (viz níže) (Liaw a Wiener 2001). Každá podmnožina ve stromové struktuře navrhne klasifikování vektoru do určité třídy. Poté je z výsledků/predikce všech podmnožin, na základě „hlasování“ o klasifikaci, vektor zařazen do třídy s největším počtem hlasů (Kulkarni a Lowe 2016).

Klasifikátor Random Forest je s rozvojem výpočetní techniky stále více využívanou metodou, a to nejen v dálkovém průzkum země, ale i jiných disciplínách. Ekologii (duhová klasifikace) medicíně (klasifikace nádorů), astronomii (klasifikace objektů na obloze), plánování dopravy, zemědělství (klasifikace zemědělských plodin), bioinformatice (Fawagreh a kol. 2014), bankovníctví (vyhodnocování důvěryhodnosti klienta), akciový trh (predikce vývoje trhu), e-marketing (Polamury 2017) a ve spoustě dalších odvětvích.

Obrázek 8: Základní schéma principu fungování algoritmu Random Forest



Zdroj: Maxim Dmitrijevsky; dostupný na: https://www.mq15.com/en/articles/3856?source=post_page

Trénovací vzorky jsou do podmnožin přidávány náhodně a s opakováním metodou bootstrapping. Každý náhodný výběr vzorku probíhá pokaždé z celého testovacího výběru. To znamená, že stejný vzorek může být zvolen několikrát, zatímco jiné nemusí být vůbec vybrány. Při vytváření trénovacích podmnožin je statisticky použito asi 64 % testovaných dat (in-bag vzorky). Ty jsou použity k trénování stromů. Zbývajících 36 % (out-of-bag) je použito pro křížovou validaci, technice pro odhad, jak dobře výsledný RF model provádí (Breiman 2001).

Parametry RF

Jedním z důležitých kroků vedoucích k úspěšné klasifikaci a minimalizaci obecné chyby (generalized error) je nastavení parametrů klasifikátoru. U Random Forestu jsou vyžadovány 2 parametry:

- 1) nTree – počet vytvořených stromů v lese.
- 2) Mtry – počet náhodně vybraných proměnných pro rozdělení uzlu stromu.

Citlivostí RF klasifikátoru na zvolené parametry se ve své studii mimo jiné zabývali Belgiu a Dragut (2016), kteří na základě zkoumání zjistili, že nejběžnějším doporučením pro parametry počtu stromů (nTree) je 500 a Mtry na druhou odmocninu počtu vstupních proměnných. Podle studie Noi a Kappas (2017) mnoho autorů dosahuje uspokojivých výsledků s defaultním nastavením. Breiman (Liaw a Wiener 2001) mimo defaultního nastavení doporučuje pro mtry $\frac{1}{2}$ defaultního nastavení, dvojnásobek defaultního nastavení nebo mtry = 1. Podle autorů nejsou rozdíly při různém nastavení nijak dramatické. Navzdory výše zmíněným doporučením, velké množství autorů doporučuje zpětný postup pro eliminování chyb, který je založený na predikci přesnosti s použitím OOB (out-of-bag). Použitím tohoto postupu se podle autorů výrazně zvyšuje přesnost.

Rešerše

Studie Pal (2005) se zabývá prezentováním svých výsledků dosažených s klasifikátorem Random Forest a následným porovnáním výsledků s klasifikátorem SVM z hlediska klasifikační přesnosti, rychlosti klasifikace a parametrů potřebných ke klasifikaci. Ve své práci autor s využitím dat Landsat 7 klasifikuje stejně jako v této bakalářské práci land cover nedaleko města Littleport ve Velké Británii. S cílem dosáhnout nejlepší přesnosti, otestoval autor přesnost klasifikace s různým nastavením parametrů. Konkrétně počet stromů. Testován byl počet stromů mezi 100 a 12000 a nejlepší uživatelské přesnosti bylo dosaženo při počtu stromů 100. Jak ukázaly výsledky RF (88,37 %) je, co se přesnosti týče, porovnatelný s SVM (87,9 %) klasifikátorem. Výhodou RF je počet parametrů, které jsou potřeba pouze 2. Podle autora je schopný RF zvládnout i nevyvážená data a data s chybějícími hodnotami.

Dalšími autory, kteří se zabývali hodnocením land cover jsou Gislason a kol. (2005). Ve studii byly porovnávány přesnosti nejpoužívanějších kombinovaných metod jako jsou boosting a bagging s metodou RF. Jsou použita data Landsat a geografická data vrstevnic (po 10 m), sklonitosti a orientace. Zkoumanou oblastí je pohoří v Coloradu a klasifikováno je 10 tříd. Různá měření ukázala, že klasifikátor RF dosahuje velmi dobrých výsledků, ale ne ve všech měřeních dosahuje vyšší přesnosti jako boosting nebo bagging. Výhodou klasifikátoru je jeho rychlost. Nejvyšší přesnosti u RF (82,8 %) klasifikátoru bylo dosaženo při 500 stromech a 3 uzlech.

V jiné studii (Nitze a kol. 2012) bylo hodnocení land cover provedeno za pomoci multispektrálních dat RapidEye. Autoři použili data ze 4 různých období (2. června, 10 srpna, 25 srpna a 5 září), která doplnila o 5 vegetačních indexů jako doplňující zdroje pro zvýšení přesnosti klasifikace. Porovnávány byly nejčastěji používané metody strojového učení: SVM, ANN, RF a tradiční klasifikátor Maximum likelihood. Zkoumaná oblast byla Indian Head v Kanadě, kde bylo do klasifikace zahrnuto 512 zemědělských polí v roce 2009. Klasifikováno bylo 10 různých plodin. Počet stromů v klasifikaci byl 100, 400, 500 a 600, každý použit u snímku z jiného období. Nejnižší přesnosti klasifikace dosáhl ve všech 4 obdobích ML klasifikátor, nejlepších většinou SVM. RF a ANN dosáhli nepatrně horších výsledků než SVM. Špatných výsledků dosáhl RF na snímcích z 2. června, kde měl klasifikátor při nastavení nTree 500 znatelně nejhorší přesnost ze všech metod strojového učení.

Jinou, zajímavou studií se zabýval Noi a Kappas (2017), kteří porovnávali přesnost metody k-nejbližší soused, SVM a RF v deltě řeky Red River ve Vietnamu na snímcích Sentinel 2. Na celkové ploše 30 x 30 km² bylo klasifikováno 6 různých typů povrchu. Zajímavou částí na této studii bylo 14 tréninkových dat, některá vyvážená a některá nevyvážená o celkové velikosti 50–1250 pixelů/třída. Stanoveny byly 2 základní cíle. 1) porovnat klasifikační přesnost všech tří klasifikátorů na snímcích Sentinel 2 a 2) vyhodnotit vliv velikosti tréninkových dat na přesnost klasifikace. Pro každý klasifikátor bylo vyzkoušeno několik parametrů. U RF bylo testování nTree = 100, 200, 500 a 1000 a mtry = 1:10. Nejlepších výsledků bylo dosaženo při mtry = 2, 3 a 4 a ntree 200, 500 a 1000 ukazují velmi podobné výsledky, ale nejlépe dopadl nTree = 500. V této studii byla tréninková data rozdělena do podmnožin tréninkových dat, kde sada 1 obsahovala 5 % z původních tréninkových dat, sada 2 10 %, další 20 %, 40 %, 60 %, 80 % a nakonec 100 %. Nejvyšší přesnosti u všech klasifikátorů bylo dosaženo při velikosti trénovacích ploch 60 % (=760 pixel/třída). Autoři zjistili, že pokud jsou tréninková data příliš velká, může dojít ke změnám mezi třídami a tím znehodnocovat přesnost. Pro všechny třídy v této studii bylo dosaženo nejvyšší přesnosti, když tréninková data reprezentují asi 0,25 % celkové studované plochy.

Cílem studie Niculescu a kol. (2018) byla klasifikace vegetačního pokryvu v oblasti Pays de Brest ve Francii při použití série snímků radarových dat Sentinel 1 a optických dat Sentinel 2 a SPOT-6. Do procesu klasifikace byly zahrnuty i 4 indexy (NDVI, NDWI, S2rep, IRECI), které byly vytvořeny nad daty Sentinel 2. Jako klasifikátor byl využit Random Forest. Klasifikovány byly třídy: letní plodiny, bez vegetace, les, voda, travnatý porost, močály, zimní plodiny a písečné duny. Zajímavou myšlenkou bylo spojení radarových dat, optických dat a indexů. Byly porovnávány 3 klasifikační metody: 1) SAR data + NDVI a NDWI, 2) SAR data + S2rep a IRECI, 3) SAR data + optická data. Z výsledků vyplývá, že díky vegetačním indexům bylo možné dobře rozlišit lesy, ornou půdu a travnatý porost. Dále pásma red-edge u snímků Sentinel 2 pozitivně obohatilo klasifikaci při mapování vegetace. Série snímků měla také pozitivní vliv zejména na mapování vegetačního krytu.

Na a kol. (2010) porovnávali výsledky klasifikačních algoritmů Random Forest, CART (Classification and Regression Trees) a Maximum Likelihood. Mapovány byly wetlands v oblasti Sanjang Plain v Číně pomocí dat Landsat 5, geografických dat (sklon, nadmořská výška, topografická data, geologické podloží a klimatická data) a dvou vegetačních indexů (NDVI, EVI). Klasifikovány byly třídy: suchá půda, lesy, louky, bažiny, zástavba, vodní plochy a rýžová pole. Hodnocena byla celková přesnost klasifikátoru a citlivost na velikost trénovacích dat. Bylo zjištěno, že RF a CART mohou dosahovat podstatně vyšší přesnosti než běžně používaný klasifikátor ML. Nejvyšší celková přesnost byla získána klasifikací RF (91,3 %). Za druhé, RF byl oproti CART a ML nejméně citlivý na velikost trénovacích dat a jeho výsledky byly nejstabilnější. Naopak nejhorších výsledky z všech 3 testovaných klasifikací dosáhl ML.

Cílem studie Jhonnerie a kol. (2015) bylo za použití objektově orientované klasifikace určit a porovnat přesnost klasifikátoru RF a klasifikátoru ML. Zkoumanou oblastí byl mangrovový les na řece Kembung v Indonésii. K procesu 7 klasifikací byla použita kombinace dat snímku Landsat 5TM, ALOS PALSAR FBD (Advanced Land Observation Satellite Phase Array L-band Synthetic Aperture Radar Fine Beam Double Polarization) a spektrálních transformací (NDVI, NDWI, NDBI). K vyhodnocení přesnosti byl použit celková přesnost a kappa koeficient. Pro všechny kombinace dat dosáhl RF vyšší přesnosti než ML. Nejvyšší přesnost RF klasifikátoru (81,1 %, 0,76 kappa) byla při použití všech zmíněných dat, zatímco u ML bylo nejvyšší přesnosti (77,7 %, 0,71 kappa) dosaženo při kombinaci dat Landsat 5TM a ALOS PALSAR.

3 Zájmové území

3.1 Vojenský újezd

„Vojenský újezd je vymezená část území státu určená k zajišťování obrany státu a k výcviku ozbrojených sil. Újezd tvoří územně správní jednotku“ (ACR 2018).

Nejstarší vojenský újezd Brdy a Dědice vznikli v roce 1927. Nejvíce VÚ vzniklo v období komunistického režimu, kdy vznikaly nové vojenské újezdy a staré byly rozšiřovány. Od konce studené války jsou VÚ rušeny nebo zmenšovány a zpřístupňovány veřejnosti. V současné době se na území České republiky nachází 4 VÚ (Boletice, Březina, Hradiště, Libavá) (MVCR 2010). Od roku 1994 ve VÚ probíhá výcvik armády české armády a jejich spojenců v rámci programu Pro mír a nyní i spojenci NATO (ACR 2018).

Ačkoliv jsou hodnocení land use/land cover a jejich změny jednou z nejčastěji dálkově získávaných informací, jsou vojenské újezdy zkoumány jen sporadicky i přesto, že vývoj krajiny uvnitř újezdů odlišuje od okolní krajiny. Autoři, kteří se zabývali změnami land coveru ve vojenských újezdech jsou Skokanová a kol. (2017). Ve své studii zkoumali dlouhodobý vývoj krajinného pokrytí na území 5 vojenských újezdů a zdali se jejich vývoj odlišuje. Součástí výzkumu bylo zjistit jaká je závislost mezi krajinným pokrytím a populací na území. Použito bylo 6 sad topografických map a leteckých snímků (z let 1836–2014). Klasifikováno bylo 7 tříd (orná půda, trvalé travní porosty, sady, dřevinná vegetace, vodní plochy, zástavba a ostatní – zahrnující těžbu, ruiny a cvičicí plochy). Kategorie lesa byla nahrazena dřevinnou vegetací, která zahrnuje i keře a stromy stojící osamoceně mimo plochu lesa). Výsledky se liší v závislosti na historických podmínkách a změnách. Dále byly zjištěny podobnosti mezi újezdy Březina a Brdy, a Hradiště a Libavá. První, menší, skupina byla charakteristická nižší populací s méně vojenskými tábory. Převažoval spíše dřevinný pokryv na rozdíl od druhé skupiny, kde převládala otevřená prostranství. Po 2. světové válce se snížila populace a zvýšila se dřevinná vegetace. V posledních 20. letech vede upadající vojenská činnost ke zvyšování dřevinné vegetace a ochraně krajiny vojenských újezdů.

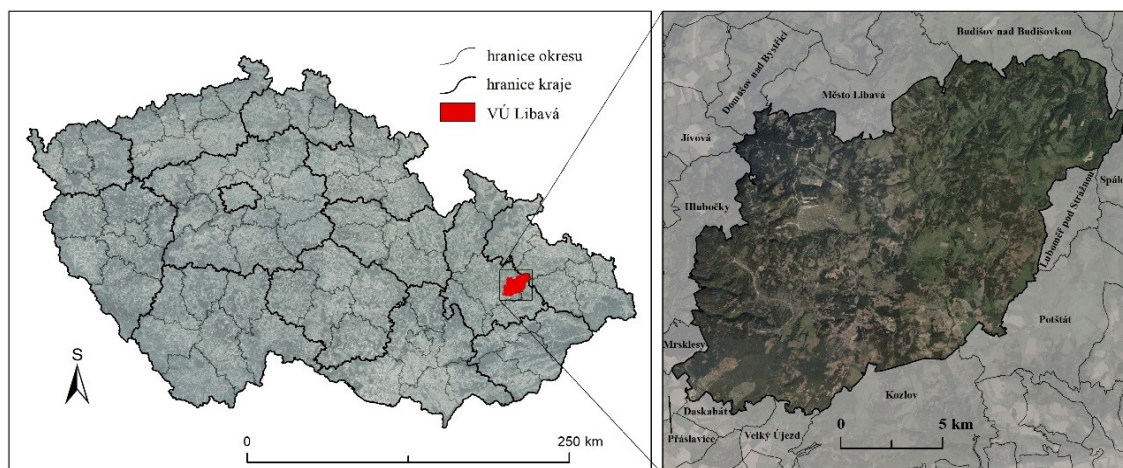
Podobným výzkumem se zabýval i Havlíček a kol. (2018), kteří zkoumali land use a silniční síť v újezdu Libavá a porovnávali výsledky s vývojem jeho okolí. Jejich cílem bylo potvrdit hypotézu, že založení vojenského újezdu významně ovlivní land use a silniční síť v tom směru, že dojde ke zvyšování lesních ploch a/nebo travních porostů. Tato hypotéza byla potvrzena. Využity byly topografické mapy a letecké snímky z let 1837–2014). Rozlišováno bylo 11 land use kategorií (orná půda, trvalé travní porosty, zahrady a sady, chmelnice, les, vodní plochy, zástavba, rekreační plochy, vojenské kempy, a ostatní (těžba a dopadové plochy) a opuštěné ruiny). Podobně jako ve studii

výše, využití land use velmi záviselo na historickém období. Od 2. poloviny 20. byl využíván VÚ k trénování tanků a pěchoty, který ke cvičení využívali opuštěná obydlí. To mělo za následek ubývání zastavené plochy a změnu orné půdy v travnaté povrchy. Dnes se již nevyužívané vojenské plochy pokrývají vegetací a zalesňují se.

3.1.1 Vojenský újezd Libavá

Vojenský újezd Libavá byl založen v roce 1950 a nachází se v Olomouckém kraji (Obrázek 9), severovýchodní částí svého území zasahuje do Moravskoslezského kraje. Leží severovýchodně od krajského města Olomouc v oblasti Oderských vrchu. Hlavní sídlem je Město Libavá. 1.1.2016 byly podle zákona podle zákona č. 15/2015 Sb. změněny hranice Vojenského újezdu Libavá. Újezd se tak zmenšil z 327 km² na 245 km² (ztráta území Města Libavá, Budišov nad Budišovkou, Vítkov, Luboměř p. Strážnou, Kozlov a Slavkov, Mrsklesy, Hlubočky, Jívová a Domašov n. B.) (ACR 2018). Vojenský újezd Libavá je druhým největším újezdem v Česku. Pouze 1/3 území je určena pro tréninkové účely, převážně pro tanky, dělostřelectvo a pěchotu.

Obrázek 9: Poloha Vojenského újezdu Libavá



Zdroj: Vlastní tvorba

Krajinu VÚ tvoří zvlněný reliéf Oderských vrchů a Vítkovské a Domašovské vrchoviny. Průměrná nadmořská výška je kolem 500 m n. m. Nejvyšší vrchol Chlum dosahuje výšky 705,7 m, Nejnižší místo, 320 m, se nachází u příslavnických kasáren. Nejnižší a nejvyšší naměřené teploty dosahují hodnot +/- 35 °C. Průměrná denní teplota vegetačního období (duben-září) je 12 °C. Roční úhrn srážek činí v průměru 863 mm. (VOJUJEZD-LIBAVA.CZ 2015). Většina území leží v povodí Odry. Území představuje významnou akumulační oblast povrchových vod. Největší je vodní nádrž Barov (10 ha) v severovýchodní části vojenského újezdu (Smolová a kol 2005).

Uzavření vojenského újezdu pro civilní obyvatelstvo pozitivně ovlivnilo přírodní charakter území. Dnes je považován za významnou krajinu s cennými biotopy s výskytem zvláště chráněných rostlin a živočichů (VOJUJEZD-LIBAVA.CZ 2015). Celé území je díky vzácnému ptactvu označeno jako speciálně chráněné území (SPA). 1/3 je na seznamu významných lokalit Natura 2000 (Havlíček a kol. 2018).

4 Data a metodika

V následující části budou přiblíženy jednotlivé typy dat, které byly použity pro zpracování bakalářské práce. Následovat bude metodická část ve které bude popsán proces příprav, zpracování dat, klasifikace a v poslední části vyhodnocení dat.

4.1 Data

4.1.1 Sentinel 2

Pro hodnocení land cover jsem využil data Sentinel 2, která byla vyvinuta v rámci evropského vesmírného programu Copernicus. Program vznikl na základě společné iniciativy evropské komise a evropského vesmírného programu (ESA 2012). Data jsou volně dostupná na internetovém portále Copernicus Open Access Hub (<https://scihub.copernicus.eu/dhus/#/home>).

Systém Sentinel 2 se skládá ze dvou družic: Sentinel 2A uvedenou do provozu 23.6.2015 a Sentinel 2B vypuštěnou 7.3.2017. Družice poskytuje multispektrální snímky s vysokým prostorovým rozlišením a šířkou záběru až 290 km. Součástí družice je optický senzor MSI snímající ve 13 pásmech, z nichž mají prostorové rozlišením 10 m, 6 pásem s rozlišením 20 m a 3 pásma s rozlišením 60 m (Tabulka 2). Časové rozlišení jedné družice je 10 dní, při použití obou družic je možné dosáhnout časového rozlišení 5 dní. Data jsou systematicky kaskádově zpracovávána a řazena do několika úrovní Level 0, Level 1 (1A, 1B, 1C) a Level 2A (ESA 2018).

Tabulka 1: Specifikace Sentinel 2A

spektrální pásma		střední vlnová délka (nm)	šířka pásma (nm)	prostorové rozlišení (m)
1	coastal aerosol	443,9	27	60
2	blue	496,6	98	10
3	green	560	45	10
4	red	664,5	38	10
5	red-edge 1	703,9	19	20
6	red-edge 2	740,2	18	20
7	red-edge 3	782,5	28	20
8	NIR	835,1	145	10
8a	NIRn	864,8	33	20
9	water vapor	945	26	60
10	shortwave infrared cirrus	1372,5	75	60
11	SWIR 1	1613,7	143	20
12	SWIR 2	2202,4	242	20

Zdroj: Sentinel 2 Product Specification Document (2018)

Data Sentinel 2 jsou velmi často využívána především k monitorování krajinného pokryvu a jeho změn, mapování zdravotního stavu vegetace, mapování rizik. K účelům monitorování vegetace jsou nejčastěji využívána blízká infračervená pásma, které Sentinel 2 poskytuje v prostorovém rozlišení 20 m (Addabbo 2016). Při výběru dat pro tuto bakalářskou práci byly hlavními důvody využitelnost při monitorování vegetace, finanční dostupnost a dobré prostorové rozlišení.

Pro tuto bakalářskou práci byla použita data produktu Level 2A, (S2A_MSIL2A_20190630T100031_N0212_R122_T33UXR_20190630T125130) v projekci UTM/WGS 84. Satelitní snímek Sentinel 2 byl je volně dostupný na stránkách společnosti ESA (<https://scihub.copernicus.eu/>). Datum pořízení snímku je 30.6.2019. Hlavními kritérii pro výběr snímku byla nízká oblačnost a vegetační období kvůli spektrálním odezvám vegetace. Na obrázku 10 je náhled upraveného a oříznutého snímku.

Obrázek 10: Vojenský újezd Libavá (2019)



Zdroj: ESA (2019): Dostupné na:
<https://scihub.copernicus.eu/dhus/#/home>

4.1.2 ORTOFOTO ČR

Pro hodnocení přesnosti a ke zvýšení přesnosti a při vytváření trénovacích ploch byla použita WMS služba Ortofoto ČR, kterou poskytuje Český úřad zeměměřičský a katastrální (dále jen ČÚZK). Snímky území Vojenského újezdu Libavá byly pořízeny v roce 2018 v prostorovém rozlišení 20 cm (ČÚZK 2019).

4.1.3 ZABAGED

Tato vrstva byla použita jako referenční primárně pro odlišování plochy orné půdy a travnatého porostu. Data ZABAGED byly použity ve formě WMS služby v měřítku 1:10000, kterou poskytuje ČÚZK. V současné době je tvořena 128 typy geografických objektů zařazených do polohopisné nebo výškopisné části. (ČÚZK 2019)

4.1.4 LPIS

Stejně tak, jako výše uvedená vrstva ZABAGED, byla i vektorová vrstva LPIS využita k odlišení třídy orné půdy a travnatého porostu. LPIS je veřejný registr půdy, který je tvořen primárně evidencí zemědělské půdy. Hlavním cílem je ověřování údajů v žádostech o dotace poskytovaných ve vazbě na zemědělskou půdu (LPIS 2019)

4.1.5 ArcČR 500

Data ArcČR obsahují digitální vektorová data České republiky a byla využita pro ořez zájmového území. Vznikla ve spolupráci ARCDATA PRAHA, s.r.o., Zeměměřického úřadu a Českého statistického úřadu.

4.2 Software

Při zpracování této bakalářské práce byly využity 3 softwary. SNAP 6.0, který je volně ke stažení na internetových stránkách společnosti ESA (<http://step.esa.int/main/download/>). Program byl využit k předzpracování surových satelitních snímků. Dalším využitým programem byl ArcMap 10.2 od společnosti ESRI. Proběhlo zde přetransformování dat do stejné projekce, tvorba trénovacích ploch a část hodnocení přesnosti. Programem, ve kterém proběhla samotná klasifikace a hodnocení přesnosti je ENVI 5.1 od Harris Geospatial.

4.3 Metodika

4.3.1 Předzpracování

Před samotným hodnocením land cover vojenského újezdu bylo nejprve potřeba satelitní snímek Sentinel 2 upravit. Veškeré předzpracování snímku proběhlo v programu SNAP 6.0. V rámci příprav na klasifikace byly vytvořeny trénovací plochy v programu ArcMap (viz níže).

Součástí úrovně 2A satelitního snímku jsou atmosférické korekce. Jeho hodnoty odrazivosti odpovídají úrovni odrazivosti na povrchu BOA (Bottom of Atmosphere) a nejsou ovlivněny atmosférickými vlivy. Proto není potřeba žádné další atmosférické korekce provádět.

Snímky Sentinel 2 obsahují 13 spektrálních pásem, které se liší prostorovým rozlišením (60 m, 20 m, 10 m). Aby bylo možné provést další úpravy vedoucí ke klasifikaci snímku, bylo potřeba snímek transformovat tak, aby prostorové rozlišení bylo jednotné u všech pásem. Toho bylo docíleno pomocí funkce *resampling* (=převzorkování). Převzorkování je proces, při kterém je původní DN hodnota převedena do jiné souřadné soustavy (Dobrovolný 1998). Transformována byla pásma s nižším prostorovým rozlišením, a to 1, 9, 10 (60 m) a pásem 5, 6, 7, 8a, 11 a 12 (20 m). Zvolena byla metoda *nearest neighbor*. Díky tomuto procesu byly všechny pixely snímku převedeny do stejného prostorového rozlišení u všech pásem, a to na 10 m.

V následujícím kroku byla za pomoci funkce *subset* vybrána pásma, která byla následně použita v klasifikačním procesu. Vybráno bylo 9 pásem. Ta, jejichž původní prostorové rozlišení bylo 10 m (pásma 2, 3, 4, 8) a 20 m (5, 6, 7, 8a, 11, 12). Odstraněna byla pásma s původním prostorovým rozlišením 60 m (1, 9, 10) a pásmo 8a (20 m), protože se svým rozsahem vlnové délky překrývá s pásmem 8, které má ale lepší prostorové rozlišení. Po provedených korekcích, převzorkování a vybrání pásem byl snímek oříznut na velikost zkoumaného území. Ořez satelitního snímku byl proveden za pomoci shapefilové vrstvy Vojenského újezdu Libavá dostupného z datasetu ArcČR funkcí *land/sea mask*.

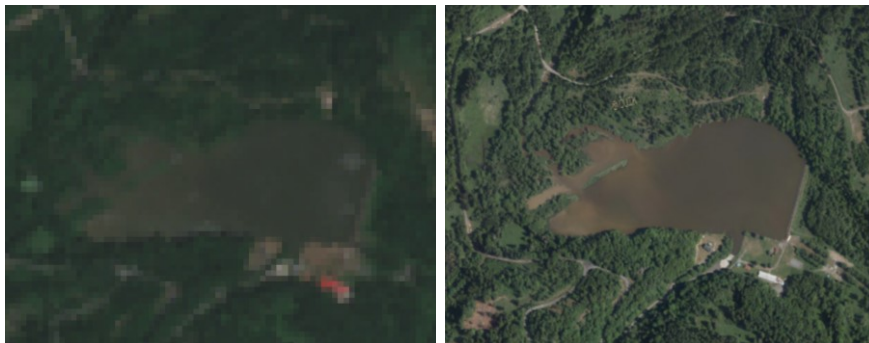
4.3.2 Land cover a klasifikační legenda

Tvorba klasifikační legendy je důležitou součástí každého klasifikačního procesu a také klíčovým faktorem ovlivňující celkový výsledek. Legenda je vždy volena na základě charakteru území a měla by být zahrnovat všechny třídy, které se na daném území nachází. Klasifikační třídy byly vybrány a definovány s pomocí nomenklatury CORINE Land Cover (Copernicus 2019) (viz níže). Hodnocením land cover vojenských újezdů Brdy a Ralsko se ve své studii zabývali Štych a kol. (2018). S ohledem na podobnosti krajiny vojenských újezdů byly s menšími modifikacemi odpovídajícím charakteru VÚ Libavá vybrány stejné klasifikační třídy: vodní plochy, listnaté lesy, jehličnaté lesy, travnatý porost, řídká vegetace, orná půda, bez vegetace a zástavba. Jako referenční data při tvorbě trénovacích ploch byla využita převážně data Ortofoto ČR. Dále volně dostupná WMS služba ZABAGED a vektorová vrstva LPIS (29.5.2019 dostupné z: <http://eagri.cz/public/app/eagriapp/lpisdata/>), která byla využita pro rozlišení mezi ornou půdou a travním porostem. Pro rozlišení jehličnatých a listnatých lesů byla využita barevná syntéza 843 originálního snímku Sentinel. Drobné rozdíly mezi půdou bez vegetace a řídkou vegetací detekoval NDVI index.

Vodní plochy

Na území se vyskytují i drobnější vodní toky (Obázek 11), které s ohledem na prostorové rozlišení, nebyly do procesu klasifikace zahrnuty. Tato třída obsahuje nomenklaturu 5.1.2. Vodní plochy dle CORINE Land Cover.

Obrázek 11: Vodní plochy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)

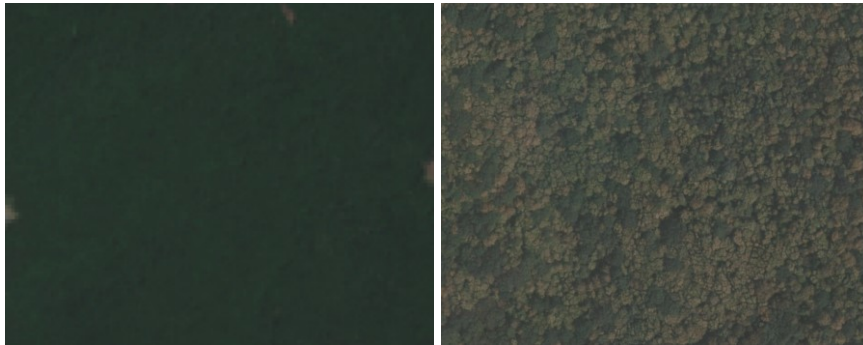


Zdroj: Sentinel 2, Ortofoto ČR

Listnaté lesy

Území, kde je hustota korunové pokrývky listnatých stromů (Obrázek 12) více než 30 %. Odpovídá nomenklatuře 3.1.1 Listnaté lesy. Ve VÚ Libavá převládají listnaté lesy nad jehličnatými. Problémem u klasifikování listnatých a jehličnatých stromů jsou lesní školky – stromy do výšky 5 m mají podobné spektrální projevy a byly zahrnuty do třídy řídké vegetace.

Obrázek 12: Listnaté lesy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Jehličnaté lesy

Území, kde je hustota korunové pokrývky jehličnatých stromů (Obrázek 13) více než 30 %. Odpovídá nomenklatuře 3.1.2 Jehličnaté lesy.

Obrázek 13: Jehličnaté lesy Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Třída smíšených lesů nebyla vytvořena, a proto v případě výskytu smíšeného lesa je o klasifikaci rozhodnuto na základě vyššího zastoupení jehličnatého nebo listnatého lesa.

Travnatý porost

Tato třída (Obrázek 14) zahrnuje 2.3.1 Louky a 3.2.1 Přírodní pastviny. Do třídy louky patří trvalé travní porosty, které jsou nebo byly v minulosti využívány jako zemědělské plochy nebo jsou ovlivněny lidskou činností. Typicky se využívají jako pastviny. Louky se svými spektrálními vlastnostmi podobají orné půdě, a proto je problematické jejich rozlišení. Pro takové případy byly při rozlišení použity i vrstvy WMS ZABAGED a LPIS. Přírodními pastvinami jsou plochy, které jsou minimálně nebo nejsou vůbec ovlivněny lidskou činností.

Obrázek 14: Travnatý porost Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Řídká vegetace

Do kategorie řídké vegetace (Obrázek 15) byly zahrnuty nomenklatury 3.2.2 Slatiny a vřesoviště, křovinaté formace, 3.2.4 Přejídná stadia lesa a křoviny a 3.3.3 Oblasti s řídkou vegetací. Řídká vegetace je typicky reprezentována lesními školkami, novou vegetací rostoucí v místech po těžbě stromů nebo dopadových ploch, nízkou dřevinnou vegetací (keře) a řídkou vegetací pokrývající 10–50 % povrchu.

Obrázek 15: Řídká vegetace Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Orná půda

Tato třída podle definice CORINE odpovídá nomenklatuře 2.1.1 Orná půda (Obrázek 16) mimo zavlažovaných ploch. Jako referenční data byla využito Ortofoto ČR, ZABAGED a LPIS. Orná půda se vyskytuje většinou na okraji území.

Obrázek 16: Orná půda Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Bez vegetace

Tato kategorie reprezentuje plochy bez vegetace (Obrázek 17). Mezi takové patří například oblasti po těžbě dřeva.

Obrázek 17: Bez vegetace Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

Zástavba

V roce 2016 došlo ke zmenšení VÚ Libavá. Nyní se na území nachází jen minimum zastavěné plochy (Obrázek 18). Podle definic CORINE do této třídy řadíme třídy: 1.1.1 Městská souvislá zástavba, 1.1.2 Městská nesouvislá zástavba, 1.2.2 Silniční a železniční síť a přilehlé prostory a 1.3.1 Těžba hornin. Do třídy zástavby byly zahrnuty plochy reprezentující nezpevněné cesty (ZABAGED) a dopadové plochy.

Obrázek 18: Zástavba Sentinel 2 (vlevo) a Ortofoto ČR (vpravo)



Zdroj: Sentinel 2, Ortofoto ČR

4.3.3 Trénovací plochy

Trénovací plochy byly vytvořeny v programu ArcMap 10.5.1. Nejprve bylo vytvořeno 8 shapefilů reprezentující jednotlivé třídy. U každé vrstvy byly vytvořeny trénovací polygony (Tabulka 3) s důrazem na dodržení několika základních pravidel (Dobrovolný 1998):

- Trénovací plochy by měly obsahovat pouze pixely dané kategorie a nezasahovat k okrajovým plochám krajinného pokryvu
- Trénovací plochy musí obsahovat dostatečný počet pixelů
- V případě klasifikace Maximum Likelihood je důležité, aby trénovací plochy měly normální rozdělení (pro účely RF klasifikátoru být nemusí)
- Trénovací plochy by v ideálním případě měly být rovnoměrně prostorově rozmístěny.

Minimální mapovací jednotka pro klasifikaci byla stanovena 0,25 ha (=25 pixelů pro snímek Sentinel 2) (Copernicus 2016). Pro větší úspěšnost klasifikace bylo (tam kde to bylo možné) u problematických tříd (jehličnaté x listnaté, bez vegetace x zástavba, orná půda x travnatý porost, orná půda x bez vegetace) vytvořeno více polygonů (Tabulka 2). Třída vodních ploch, zástavby a orné půdy obsahovala trénovacích ploch, protože jejich zastoupení v území je nižší. Každý vytvořený polygon byl zkontrolován mapou Ortofoto ČR, nahanou pomocí WMS služby do ArcMapu. U orné půdy a travnatého porostu, jak

bylo zmíněno výše, byly polygony kontrolovány na referenčními daty ZABAGED a LPIS.

Tabulka 2: Trénovací plochy

	počet polygonů	počet pixelů
vodní plochy	8	1057
listnaté lesy	93	11945
jehličnaté lesy	91	5377
travnatý porost	48	13282
řidká vegetace	24	3112
orná půda	18	8012
bez vegetace	101	9915
zástavba	30	1368

Zdroj: Vlastní tvorba

Po vytvoření trénovacích ploch byla data naimportována spolu s upraveným snímkem Sentinel 2 do programu SNAP 6.0, kde byly funkcí *Statistics* vygenerovány statistiky všech tříd. Ke statistickému testování byl vytvořen graf spektrální odrazivosti jednotlivých tříd ve všech použitých pásmech a Jeffries-Matusita test oddělitelnosti mezi jednotlivými třídami.

Pro zhodnocení oddělitelnosti byla využita funkce *ROI Separability* (Tabulka 3) v programu ENVI 5.1. ROI Separability porovnává vždy oddělitelnost dvou tříd a hodnoty se pohybují v intervalu od 0 do 2. Čím více se hodnota blíží 2, tím lépe oddělitelné jednotlivé třídy jsou. V případě nižších hodnot je vhodné zvážit sloučení tříd nebo upravení trénovacích ploch. Výsledek oddělitelnosti je zobrazení v tabulce č. 3.

Tabulka 3: Oddělitelnost jednotlivých tříd

	vodní plochy						
listnaté lesy	2,000	listnaté lesy					
jehličnaté lesy	1,999	1,986	jehličnaté lesy				
travnatý porost	2,000	1,999	1,999	travnatý porost			
řidká vegetace	2,000	1,914	1,967	1,848	řidká vegetace		
orná půda	2,000	2,000	2,000	1,942	1,999	orná půda	
bez vegetace	2,000	2,000	1,999	1,998	1,999	1,944	bez vegetace
zástavba	1,999	1,999	1,999	1,999	1,999	1,999	1,999

Zdroj: Vlastní tvorba

Jeffries-Matusita test oddělitelnosti ukazuje, že nejnižší hodnoty podle předpokladů dosahují třídy řídké vegetace a travnatého porostu (1,848), déle řídké vegetace a listnaté (1,914) a jehličnaté lesy (1,967). Řídká vegetace se vyskytuje na zarůstajících plochách, vzniklých po těžbě lesa nebo právě na rozhraní travnatých porostů a lesů. Příkladem mohou být lesní školky, kde jsou vysazovány nové stromy. Relativně nižších hodnot dosahují i třídy travnatého porostu a orné půdy (1,942). Mezi nižšími hodnotami, v porovnání s ostatními třídami, se pohybuje orná půda a plochy bez vegetace (1,944). Všechny ostatní oddělitelnosti tříd dosahují hodnot 1,98 a více. Obecně jsou hodnoty 1,9 a více považovány za dobře oddělitelné, čehož je dosaženo téměř u všech tříd.

4.3.4 Klasifikace

Random Forest

Klasifikace Random Forest byla provedena nad satelitním snímkem v programu SNAP 6.0. Po vytvoření trénovacích ploch v programu ArcMap byly polygony nahrány do programu SNAP, kde byl následně spuštěn proces klasifikace. K nastavení procesu klasifikace slouží 2 volitelné parametry: Number of training samples. Zde byla ponechána přednastavená hodnota 5000 a Number of Trees hodnota 500. Tyto hodnoty byly vybrány na základě testování volitelných parametrů a jejich celkové přesnosti (Tabulka 4). Podle studií (kapitola 2.3.4) bylo dosahováno nejlepších výsledků při použití number of trees v rozmezí 100–1000, proto byly k testování využity hodnoty 100, 500, 1000. Posledním krokem před klasifikací bylo zvolení spektrálních pásem a tréninkových ploch, které vstupovaly do procesu. Jak bylo zmíněno výše, použita byla pásma s původním prostorovým rozlišením 10 a 20 m. Vynecháno bylo také pásmo 8a, které se svou vlnovou délkou překrývá s pásmem 8 a má menší prostorové rozlišení.

Tabulka 4: Testované parametry klasifikátoru RF

Random Forest			
Number of training samples	Number of trees	celková přesnost	kappa koeficient
5000	100	70,7	0,665
5000	500	85,5	0,834
5000	1000	74,3	0,706

Zdroj: Vlastní tvorba

- Number of training samples udává počet pixelů v tréninkových polygonech, které by měly být náhodně vybrány do procesu klasifikace. V případě že třída obsahuje pixelů méně, jsou některé pixely v rozhodovacím procesu využity vícekrát. V této bakalářské práci byl počet pixelů nastaven na hodnotu 5000 i přesto, že některé třídy (vodní plochy, zástavba atd.) takové velikosti nedosahují. Nejmenší možná hodnota u klasifikátoru RF byla 10 a nejvyšší 5000.
- *Number of Trees* představuje počet vytvořených trénovacích stromů. Vyšší počet stromů bývá využíván při testování velkého množství vstupních dat.

Maximum Likelihood

Klasifikace satelitního snímku algoritmem Maximum Likelihood byla provedena v programu SNAP. Ačkoliv program SNAP obsahuje mnoho užitečných funkcí, tak v možnostech nastavení parametrů klasifikátoru ML je dost omezený. Nabízí pouze nastavení jediného parametru Number of training samples, tedy pouze počet pixelů zahrnutých v procesu klasifikace. Vzhledem k počtu pixelů trénovacích dat nebyly testovány hodnoty vyšší než 10 000. Nejlepších výsledků v procesu testování (Tabulka 5) bylo dosaženo při hodnotě 10000, a proto byla použita jako výsledná hodnota pro klasifikaci porovnání s RF klasifikací. Při testování nižších hodnot bylo dosaženo horších výsledků. V procesu klasifikace byly použity pásma 2, 3, 4, 5, 6, 7, 8, 11 a 12, tedy pásma s původním prostorovým rozlišením 10 m a 20 m jako u klasifikátoru Random Forest.

Tabulka 5: Testované parametry klasifikátoru ML

Maximum Likelihood		
Number of training samples	celková přesnost	kappa koeficient
1000	80,3	0,774
5000	83,0	0,806
10000	83,8	0,817

Zdroj: Vlastní tvorba

4.3.5 Post-klasifikační úpravy a hodnocení přesnosti klasifikace

Po klasifikaci a vizuální kontrole byl klasifikovaný snímek exportován do programu ENVI. Pixelové klasifikátory často způsobují efekt „salt and pepper“, kde jsou některé samostatné pixely chybně klasifikovány. Proto bylo nutné snímek filtrovat. K tomuto procesu byl využit filtr *Majority/Minority Analysis*. V nastavení parametrů byly zvoleny všechny třídy a použit filtr 3x3. Díky této úpravě byla klasifikace zbavena rušivých pixelů.

Ačkoliv filtr pomohl vyladit plochy a zbavit je nejmenších samostatných pixelů, stále klasifikovaný obraz obsahoval i plochy, které byly větší než nejmenší MMJ. Za účelem odstranění těchto malých ploch byl snímek převeden do programu ArcMap. Zde byl klasifikovaný snímek funkcí *Raster to polygon* převeden do shapefilu. Pomocí *Select by attributes* byly vybrány plochy o velikosti menší než 0,25 ha a funkcí *Eliminate* odstraněny. Tento proces byl opakován, než bylo dosaženo podmínky MMJ. Posledním krokem před hodnocením klasifikace bylo funkcí *Create random point* vytvořit náhodné body pro každou z tříd. K hodnocení klasifikace bylo pro každou třídu vytvořeno 75 náhodných bodů (Lillesand a kol. 2008, Cit. in: Červená 2018). Kontrola všech bodů proběhla v programu ArcMap, kde byly v atributové tabulce k náhodným bodům klasifikace pomocí Ortofota ČR přiřazeny kontrolní body.

Bodová vrstva byla následně exportována zpět do ENVI, převedena do formátu ROI a funkcí *Confusion Matrix using Ground Truth ROIs* byla vyhodnocena celková přesnost klasifikace, uživatelská přesnost, zpracovatelská přesnost a kappa koeficient.

Všechny post-klasifikační úpravy byly aplikovány stejným způsobem na klasifikovaný snímek Random Forest i Maximum Likelihood.

5 Výsledky

V této kapitole jsou prezentovány výsledky obou klasifikací. Jak bylo zmíněno výše, pro zhodnocení klasifikace byla zvolena chybová matice s 600 body, rovnoměrně rozdělených do 8 klasifikovaných tříd.

5.1 Hodnocení přesnosti klasifikací

Celková přesnost obou klasifikací dosáhla 85,5 % (RF) a 83,8 % (ML), což bychom mohli považovat za uspokojivý výsledek. Podle očekávání dosáhl klasifikátor Random Forest lepších celkových výsledků než Maximum Likelihood. Obecně lze říci, že celkové výsledky za jednotlivé třídy nejsou ani u jedné z klasifikací vyvážené. Třídy, které dosahují nižší přesnosti jsou řídká vegetace, která byla nejčastěji zaměňována s lesy a travnatým porostem. Dále orná půda byla nejčastěji chybně klasifikována jako travnatý porost nebo jako třída bez vegetace. Právě tyto třídy vykazovaly nižší oddělitelnost, a proto se vzájemné chyby v klasifikaci předpokládaly. S tím je spojený i stejný trend výsledků travnatého porostu, kde uživatelská přesnost dosahovala relativně vysokých hodnot (RF 84,0 % a ML 82,7 %), ale zpracovatelská přesnost byla u obou klasifikátorů jednou z nejnižších (RF 67,7 % a ML 74,7 %). Tento rozdíl byl způsoben chybně klasifikovanou ornou půdou a řídkou vegetací jako travnatý porost. Zajímavé je, že orná půda byla zaměňována nejčastěji s travnatým porostem, ale travnatý porost byl chybně klasifikován jako orná půda pouze v jednom případě. Další problematickou třídou, snižující celkovou přesnost klasifikace byla orná půda, která se se svým projevem nacházela na hranici mezi třídou bez vegetace a travnatým porostem. Zatímco u RF dosáhl výsledek orné půdy uživatelské přesnosti pouze 60,0 %, ML dosáhl o 16 procentních bodů vyšší přesnosti. Společným znakem obou klasifikací je vyšší uživatelská přesnost vodních ploch, listnatých lesů, jehličnatých lesů a ploch bez vegetace. Podobné přesnosti bylo dosaženo u řídké vegetace, kde se uživatelská přesnost obou klasifikátorů pohybovala lehce nad 70 %, u ML byla zpracovatelská přesnost o 10,6 procentních bodů nižší. Podobné uživatelské přesnosti bylo dosaženo i u travnatého porostu, a to 84,0 (RF) a 82,7 (ML). Největší chybou v uživatelské přesnosti jsou plochy zástavby, u kterých byl rozdíl mezi klasifikátorem RF a ML 25,3 procentních bodů. To je dáno pravděpodobně rozdílnými metodami zpracování obou klasifikátorů. Z mapy 2 je na první pohled patrný rozdíl u tříd zástavby a řídké vegetace. Jde o třídy s nejmenším počtem pixelů trénovacích ploch (mimo vodu), což mohl být pravděpodobně hlavní faktor chybné klasifikace u ML. Jako třída zástavby byly u ML chybně klasifikovány plochy bez vegetace a v menší míře plochy jehličnatých lesů.

Random Forest

Celkový výsledek klasifikace Random Forest je 85,5 % a kappa koeficient³ 0,834. V tabulce 6 je vidět, že nejvyšší přesnosti podle očekávání dosáhly vodní plochy (100 %). Vysokých hodnoty uživatelské přesnosti bylo dosaženo u ploch bez vegetace (96,0 %), zástavby (92,0 %), listnatého lesa (89,3 %) a jehličnatého lesa (88,0 %). Naopak nejnižší přesnost je dle očekávání u orné půdy (60,0 %) a řídké vegetace (74,7 %). Chybně klasifikované třídy relativně dobře odpovídají statistickým ukazatelům oddělitelnosti (ROI separability), kde byla horší oddělitelnost naměřena právě u řídké vegetace s lesy a travnatým porostem. U orné půdy je nízký výsledek dán obdobím, kdy a jakým způsobem je půda obhospodařována. Tím, že byl satelitní snímek pořízen 30.6., mohou být některé plodiny v různé fázi růstu a některé zcela bez vegetace. Orná půda byla nejčastěji chybně klasifikována jako bez vegetace nebo travnatý porost. Nepřesnosti u řídké vegetace mohou být částečně způsobeny chybným vytvořených tréninkových ploch dané rozdílem mezi pořízením satelitního snímku a poslední aktualizací Ortofota ČR. VÚ Libavá prochází intenzivní změnou krajinného pokryvu a ploch, které touto změnou prochází je velké množství. To může mít negativní vliv na celkový výsledek klasifikace. Problematické oddělení řídké vegetace, listnatých lesů a řídké vegetace a travnatého porostu naznačují i výsledky klasifikace. Řídká vegetace byla chybně klasifikována jako travnatý porost a listnatý a jehličnatý les.

Tabulka 6: Chybová matice klasifikace Random Forest

RF	vodní plochy	listnaté lesy	jehličnaté lesy	travnatý porost	řídká vegetace	orná půda	bez vegetace	zástaba	celkem	uživatelská přesnost (%)	
vodní plochy	75	0	0	0	0	0	0	0	75	100,0	
listnaté lesy	0	67	2	2	4	0	0	0	75	89,3	
jehličnaté lesy	0	8	66	0	1	0	0	0	75	88,0	
travnatý porost	0	1	1	63	6	0	4	0	75	84,0	
řídká vegetace	0	7	4	8	56	0	0	0	75	74,7	
orná půda	0	0	0	19	0	45	11	0	75	60,0	
bez vegetace	0	0	0	1	0	2	72	0	75	96,0	
zástaba	0	0	0	0	0	6	0	69	75	92,0	
celkem	75	83	73	93	67	53	87	69	600		
zpracovatelská přesnost (%)	100,0	80,7	90,4	67,7	83,6	84,9	82,8	100,0			
	Celková přesnost: 0,855				Kappa Index: 0,834						

Zdroj: Vlastní tvorba

³ **Kappa koeficient** je udává přesnost provedené klasifikace s přesností dosažitelnou čistě náhodným zařazením pixelů do jednotlivých tříd. Hodnota koeficientu se může pohybovat v rozmezí 0-1. Vyšší hodnoty znamenají lepší výsledek. U hodnot nad 0,75 se jedná o dobrý výsledek klasifikace (Červená 2018)

Maximum Likelihood

Dle očekávání je výsledek klasifikátoru ML nižší (Tabulka 7) než u klasifikace RF. Celková přesnost klasifikace je 83,8 % a kappa koeficient 0,817. Nejvyšší uživatelská přesnost byla opět dosažena u vodních ploch, a to 100 %, zpracovatelská přesnost byla také 100 %. Dobrých přesností bylo dosaženo u ploch bez vegetace (96,0 %), jehličnatého lesa (90,7 %) a listnatého lesa (88,0 %). Oproti klasifikaci Random Forest byla dosažena vyšší uživatelská (76,0 %) i zpracovatelská přesnost (95 %) u orné půdy. Nejčastěji byla opět chybně označena jako travnatý porost a bez vegetace. Třídou s nejnižší uživatelskou přesností byla zástavba, která byla nejčastěji zaměňována s třídou bez vegetace a s jehličnatými lesy. Uživatelská přesnost byla pouze 66,7 %. Takto nízká přesnost je dána pravděpodobně nevyváženými trénovacími daty a jejich rozložením v území, na což je klasifikátor Maximum Likelihood citlivý. Oproti nízké uživatelské přesnosti u třídy zástavby byla zpracovatelská přesnost 100 %.

Tabulka 7: Chybová matice klasifikace Maximum Likelihood

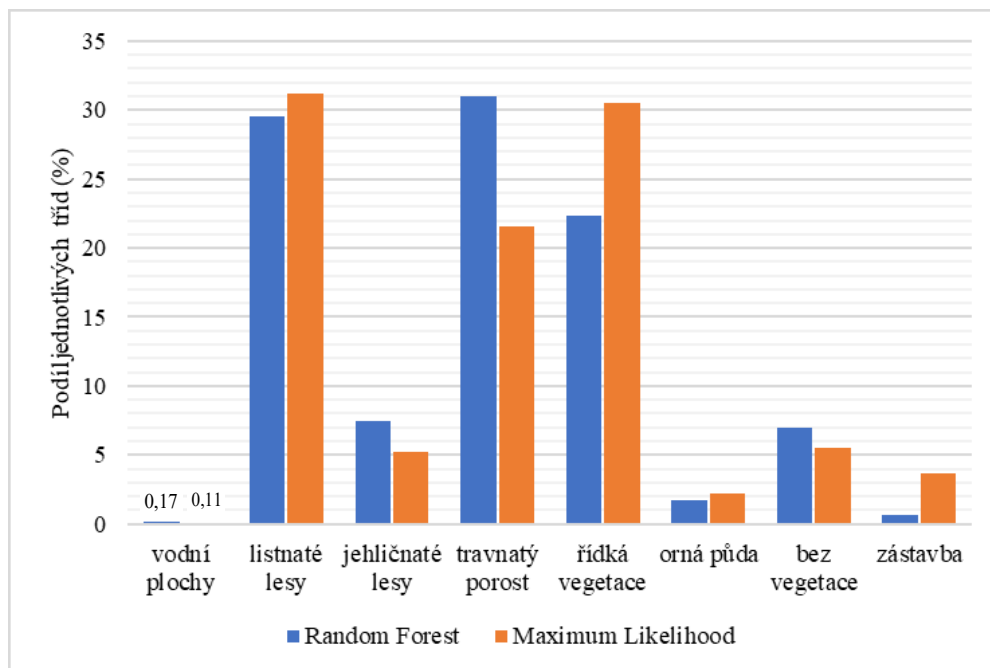
ML	vodní plochy	listnaté lesy	jehličnaté lesy	travnatý porost	řídka vegetace	orná půda	bez vegetace	zástava	celkem	uživatelská přesnost (%)
vodní plochy	75	0	0	0	0	0	0	0	75	100,0
listnaté lesy	0	66	2	1	6	0	0	0	75	88,0
jehličnaté lesy	0	4	68	0	3	0	0	0	75	90,7
travnatý porost	0	0	2	62	8	1	2	0	75	82,7
řídka vegetace	0	9	4	8	54	0	0	0	75	72,0
orná půda	0	0	0	10	0	57	8	0	75	76,0
bez vegetace	0	0	0	1	1	1	72	0	75	96,0
zástava	0	1	6	1	2	1	14	50	75	66,7
celkem	75	80	82	83	74	60	96	50	600	
zpracovatelská přesnost (%)	100,0	82,5	82,9	74,7	73,0	95,0	75,0	100,0		
	Celková přesnost: 0,838				Kappa Index: 0,817					

Zdroj: Vlastní tvorba

5.1.1 Podíly a rozložení jednotlivých tříd land cover na území VÚ Libavá

Podíl jednotlivých tříd na rozloze klasifikovaného vojenského újezdu je u obou klasifikací rozdílný (Graf 1). Zatímco listnaté lesy, jehličnaté lesy, vodní plochy a bez vegetace se liší minimálně, u řídké vegetace a travnatého porostu jsou rozdíly velké. Největším rozdílem v podílu tříd na rozloze je travnatý porost, který je u klasifikátoru ML o 9,5 procentních bodů nižší než u klasifikace RF. Podobným rozdílem (8,5 procentních bodů) v podílu zastoupení je řídké vegetace, která má zastoupení v RF klasifikátoru nižší než u ML. Právě řídká vegetace a travnatý porost jsou mezi sebou nejčastěji chybně klasifikovány. Větší rozdíl je patrný také u zástavby, která byla u klasifikátoru ML nejčastěji chybně klasifikována jako třída bez vegetace. To se projevilo také na podílu na rozloze, který je u ML 2,5x vyšší než u RF. Třída zástavba byla u klasifikátoru Maximum Likelihood nejčastěji chybně klasifikována a její podíl v rozloze je vyšší na úkor ostatních tříd.

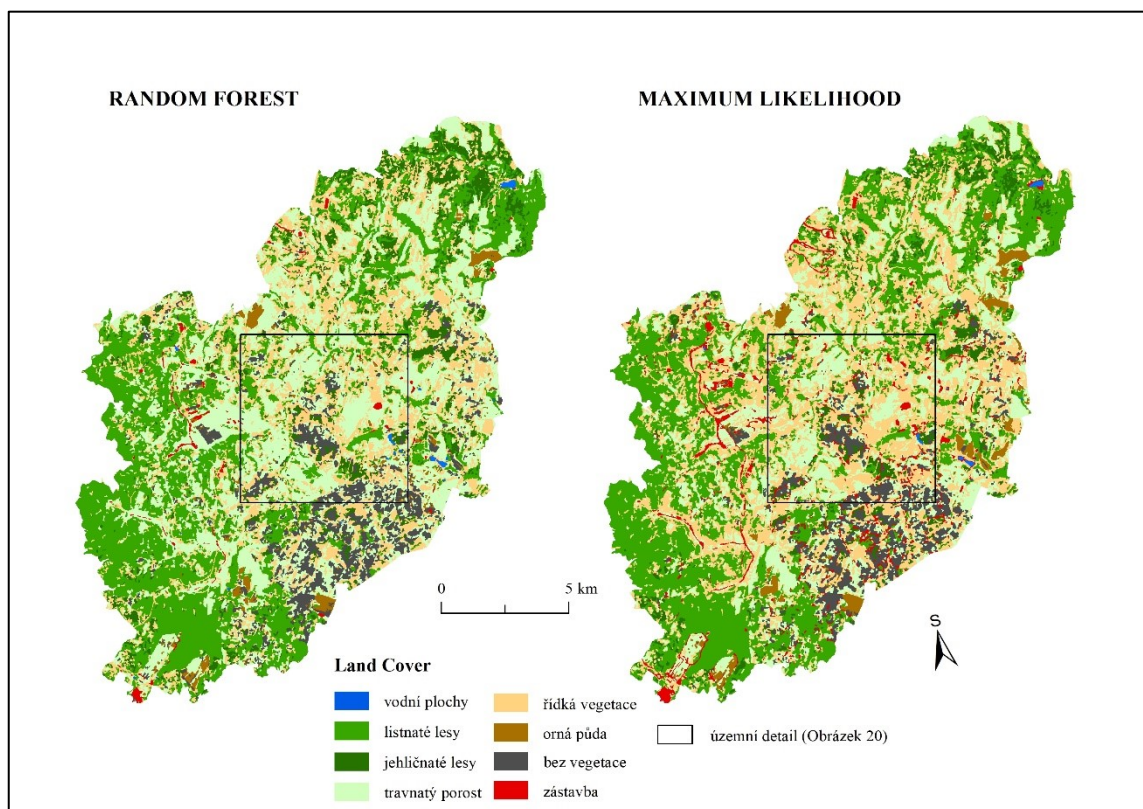
Graf 1: Podíl jednotlivých tříd krajinné pokrývky ve VÚ Libavá



Zdroj: Vlastní tvorba

Rozložení klasifikovaných tříd v krajině vojenského újezdu není rovnoměrné. Z mapy (Obrázek 19) je na první pohled patrné rozložení lesních porostů. Největší procento lesů se nachází v jižní a západní části území. Dále pak v části severní. Centrální a východní část Vojenského újezdu je lesy pokryto řídce. Další oblastí, která je na mapě patrná je jihovýchodní část, kterou dříve pokrývaly jehličnaté lesy, ale dnes, v důsledku vykácení, se zde nachází holá půda. Plochy zástavby se na území vyskytují jen roztroušeně a představují většinou silnice, nezpevněné cesty nebo dopadové plochy. Jediná souvislá plocha zástavby je v nejižnější části. Vodní plochy se na území vyskytují jen v podobě nádrží, a ačkoliv se na území nachází několik drobných toků, nebyly kvůli prostorovému rozlišení satelitního snímku klasifikovány. Největší vodní nádrž se nachází na severu. Vojenský újezd byl v roce 2016 zmenšen a přišel tak o plochy, která byla obhospodařována obyvateli okolních obcí. Pozůstatky těchto ploch jsou orné půdy, které se vyskytují jen na okrajích v okolí blízkých měst. Centrální část území je víceméně rovnoměrně pokryto travnatým porostem a řídkou vegetací.

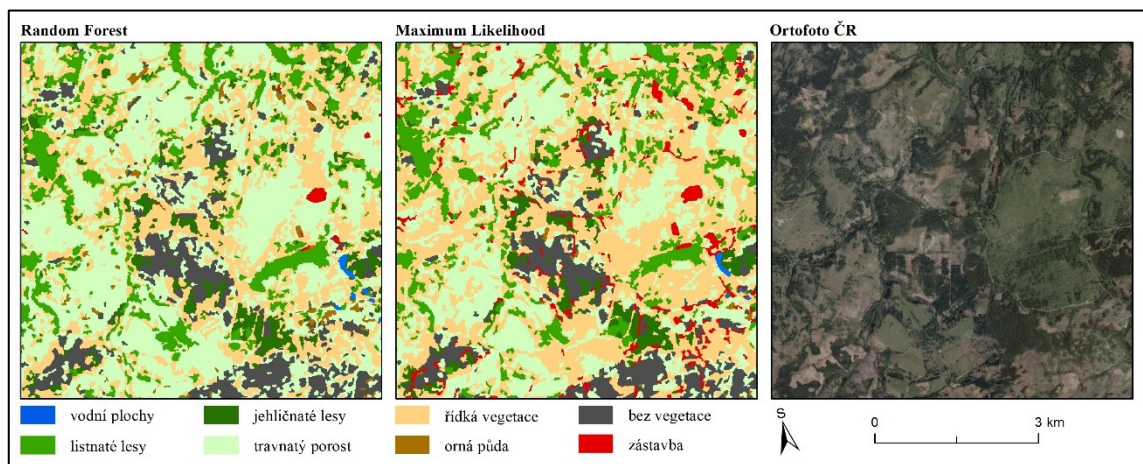
Obrázek 19: Výsledek klasifikace Random Forest a Maximum Likelihood



Zdroj: Vlastní tvorba

Pro větší představu o výsledcích klasifikace byl vytvořena detailní mapa – výřez centrální části území VÚ (Obrázek 20). Z detailní mapy jde vidět rozdíl obou klasifikátorů při klasifikování řídké vegetace a zástavby. Plochy zástavby jsou často chybně klasifikovány na rozhraní jehličnatého lesa a územím bez vegetace. Na některých místech je vidět snaha algoritmu ML klasifikovat i nejmenší silnice. Oproti tomu RF tyto silnice neklasifikuje. Za povšimnutí stojí i plochy lesa a travnatého porostu, které jsou, zejména u ML, chybně klasifikovány jako řídká vegetace. Rozdílovou třídou byly také orné půdy. Ty se ve výřezu klasifikace RF nachází v podobě menších ploch na travnatém porostu.

Obrázek 20: Detailní výřez výsledků klasifikací RF a ML



Zdroj: Vlastní tvorba

6 Diskuze a závěr

V této kapitole je pozornost zaměřena na zhodnocení metodiky a výsledků, které práce přinesla.

Cílem bakalářské práce byla klasifikace krajinného pokryvu na území Vojenského újezdu Libavá se zaměřením na klasifikátor Random Forest a následným porovnáním výsledků s nejčastěji využívaným klasifikátorem Maximum Likelihood. Pro zpracování všech dílčích úkolů byla využita data Sentinel 2 level 2A. K vytvoření trénovacích ploch byla použita volně dostupná referenční data Ortofoto ČR, ZABAGED, LPIS a CORINE Land Cover. Dílčí části práce byly zpracovávány v programech ENVI 5.1, ArcMap 10.2 a SNAP 6.0, který byl využit k procesu obou klasifikací.

Celkový výsledek klasifikátoru Random Forest dosáhl vyšších hodnot než ML. Důvodem je nižší citlivost na nevyváženost trénovacích dat (Na a kol. 2010). Kvalita trénovacích dat a celé klasifikace byla ovlivněna výběrem počtu klasifikovaných tříd, rozlišitelností jednotlivých tříd a jejich rozdílnými spektrálními vlastnostmi. Byly zahrnuty třídy odvozené definicí nomenklatur CORINE, které jsou rozpoznatelné ze satelitního snímku, Ortofota ČR nebo jiných referenčních dat. Třídy, které vstoupily do procesu klasifikace jsou: voda, listnaté lesy, jehličnaté lesy, řídká vegetace, travnatý porost, orná půda, bez vegetace a zástavba. Největší problém při výběru trénovacích ploch bylo určení počtu tříd a oddělitelností, protože některé krajinné prvky se svými spektrálními vlastnostmi podobaly a bylo obtížnější je klasifikovat. Takovou třídou byla orná půda, která se svými spektrálními vlastnostmi podobala třídě travnatý porost a třídě bez vegetace. Trénovací plochy orné půdy byly vytvořeny pomocí Ortofota ČR, ZABAGED a LPIS. Spektrálnímu odlišení orné půdy v klasifikaci by se dalo při použití snímku z jiného ročního období. Jiným řešením by mohlo být použití multitemporálního přístupu (Nitze a kol. 2012), kdy do procesu klasifikace vstupují data v několika ročních obdobích. Jiným problémem je zastoupení některých tříd a četnost jejich výskytu – takovými příklady typickými pro VÚ Libavá je nesouvislá zástavba nebo řídká vegetace. Vzhledem k účelům využití vojenského újezdu se plochy zástavby vyskytují minimálně nebo jsou v malých zastoupeních roztroušeny po celém území. Při sběru trénovacích ploch bylo problematické podle dat Ortofoto ČR vytvořit trénovací plochy některých tříd. Ačkoliv bylo ortofoto ČR nad zkoumaným územím aktualizováno v roce 2018, nastalo v území mezi pořízením satelitního snímku (30.6.2019) a Ortofotem ČR velké množství změn, spojené zejména s vykáčením velké části jehličnatých lesů poškozených kůrovcovou kalamitou. Tudiž některé plochy na satelitním snímku nekorespondovaly s plochami Ortofota ČR. Optimálním řešením pro získání aktuálních referenčních dat by mohlo být provedení InSitu měření, tedy měření v terénu, při kterém by byly přesně nasnímány problematické oblasti, zjištěny vlastnosti a výskyt jednotlivých typů povrchů

a následně využity při procesu klasifikace. Vojenský újezd je ale pro veřejnost uzavřený. Pro účel sběru trénovacích dat v této práci byla, zejména u ploch bez vegetace, využita barevná syntéza satelitního snímku a index NDVI, díky kterému bylo možné dobře odlišit místa bez vegetace a míst, která jsou po vykácení a jsou v procesu zarůstání vegetací. Stejný problém nastal u řídké vegetace, u které se spektrální vlastnosti podobají vlastnostem travnatého porostu a lesům.

V první části samotné klasifikace je hodnocen krajinný pokryv celého území obou klasifikátorů. V části druhé jsou oba výsledky porovnány. Pro proces klasifikací byl proveden v programu SNAP 6.0, který nabízí přívětivé uživatelské rozhraní s mnoha funkcemi. Na druhou stranu je jednodušší a zejména u ML nabízí jen malou možnost nastavení klasifikačních parametrů. Lepší se mi pro tyto účely jeví program ENVI, který ale neobsahuje klasifikátor RF. V klasifikaci byly testovány různé hodnoty nastavení parametrů klasifikátorů a nejlepších výsledků (Tabulka 4 a 5) bylo dosaženo při nastavení RF na *number of training samples* na 5000 a *number of trees* 500. U ML *nuber of training samples* 10000. Hodnocení přesnosti bylo provedeno pomocí chybových matic v programu ENVI. Celková přesnost klasifikátoru Random Forest je 85,5 % a přesnost klasifikátoru Maximum Likelihood je 83,8 %. Ačkoliv se zdá, že jsou celkové výsledky klasifikací podobné, uživatelské přesnosti jednotlivých tříd se liší. Stejně tak procentuální zastoupení je rozdílné. Podobných výsledků u obou klasifikací dosáhly jehličnaté lesy, listnaté lesy, travnatý porost, řídká půdy a plochy bez vegetace. Naopak největším rozdílem v přesnosti a zastoupení tříd je zástavba, která je chybně klasifikována u algoritmu Maximum Likelihood a orná půda, která naopak dosahovala nejnižších výsledků u klasifikátoru Random Forest. V této práci bylo prokázáno, že i přes stejná vstupní data mohou být výsledky klasifikátorů rozdílné. Metodu, která se osvědčila využil mimo jiné Gislason a kol (2005). Pro zlepšení výsledků klasifikace použil informace o nadmořské výšce, sklonitosti a dalších geografických datech. Niculescu a kol. (2018) využil kombinace optických dat Landsat 5 a radarových dat Sentinel 1. Za účelem zvýšení přesnosti klasifikace mohou být využity různé indexy (Niculescu a kol. 2018), zejména vegetační index NDVI se podle jiných studií zdá jako velmi platný. V praktické části byla, zejména kvůli rozložení a četnosti výskytu zástavby vyzkoušena funkce *buffer*. Cílem tohoto pokusu bylo vytvoření většího množství a lepší prostorové rozložení trénovacích dat. Ani jeden z výše zmíněných metod se pro území Vojenského újezdu Libavá neprokázala jako příliš přínosná.

Vojenské újezdy patří mezi oblasti, které se svým vývojem liší od okolního prostředí. Především uzavřenost pro veřejnost a využívání ploch k vojenským účelům určuje vývoj ekosystémů v území. V dnešní době se tento stav mění a vojenské újezdy jsou postupně zpřístupňovány veřejnosti. Větší pozornost se VÚ dostává i v odborné sféře. Změnami land coveru VÚ se zabývali Skokanová a kol. (2017), kteří zkoumali změny 5 vojenských újezdů za posledních 180 let. Podobnou studii vypracovali Štych a kol. (2018), kteří

hodnotili změny krajiny VÚ Brdy a Ralsko. Také Vojenský újezd Libavá je svou krajinou zajímavý nejen z pohledu změn, které probíhají v důsledku zpřístupňování některých částí VÚ pro veřejnost, ale i přírodních změn, které už probíhají (například redukce jehličnatých lesů způsobených kůrovcem). Další zkoumání může být přínosné pro vývoj české krajiny nebo pro boj s kůrovcovou kalamitou.

Stanovené cíle uvedené v úvodu této práce byly splněny. Tato práce potvrdila, že ačkoliv je metoda dálkového průzkumu využívána desítky let, je zde stále místo pro zlepšení. Dnešními metodami a možnostmi nedokážeme získat pomocí dálkového průzkumu bezchybné výsledky. Zjišťování informací o zemském povrchu pomocí dálkového průzkumu země je relativně mladá disciplína, která nabízí neuvěřitelné možnosti, a to nejen při klasifikace krajinného pokryvu. Proto je potřeba ve výzkumu pokračovat a metody dálkového průzkumu země zlepšovat.

Použité zdroje a literatura

Literární zdroje:

- ACR (2018): Informace o vojenských újezdech. Dostupné z (30.6.2019):
<http://www.acr.army.cz/scripts/detail.php?id=215>
- ADDABBO, P., FOCARETA, M., MARCUCCIO, S., VOTTO, C., ULLO, S. L. (2016): Contribution of Sentinel-2 data for applications in vegetation monitoring. ACTA IMEKO. 5, 2, 44-54, 11 s.
- AHMAD, A., QUEGAN, S. (2012): Analysis of maximum likelihood classification on multispectral data. Applied Mathematical Science, 6, 129, 13 s.
- ASAMOAHL, J. N., OSEI, E. M. Jnr.2., ACQUAHL, P. C., AMO AHL, A. S. (2017): Comparison of Decision Tree and Maximum Likelihood Using a Landsat Image of Ejisu-Juaben Municipality. International Conference on Applied Science and Technology (ICAST). 11 s.
- ATKINSON, P. M., TATNALL, A. R. L. (1997): Introduction Neural Networks in remote sensing. International Journal of Remote Sensing, 18:4, 12 s.
- BELGIU, M., DRAGUT, L. (2016): Random Forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 8 s.
- BREIMAN, L. (2001): Random Forest. Statistics Department, University of California, Berkley, CA, 33 s.
- CAMPBELL, B., J., WYNNE, H., R. (2011): Introduction to Remote Sensing: Fifth Edition. New York, Guilford Press, 662 s.
- CCRS: Natural Resources of Canada (2017). Dostupné z (30.6.2019):
<http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/14621>
- COPERNICUS (2016): Mapping Guide for a European Urban. European Commission. Dostupné z (30.6.2019): <https://land.copernicus.eu/user-corner/technical-library/urban-atlas-mapping-guide>
- COPERNICUS (2019): Updated CLC illustrated nomenclature guidelines. Dostupné z (30.6.2019): https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf
- ČAPEK, R. (1988): Dálkový průzkum Země. Praha, Ministerstvo školství ČSR. 244 s.

- ČERVENÁ, L. (2018): Hodnocení přesnosti klasifikace. Katedra aplikované geoinformatiky a kartografie, Přírodovědecká fakulta, Univerzita Karlova, Praha, 13 s.
- ČUZK (2019): ZABAGED. Dostupné z (30.6.2019):
[https://geoportal.cuzk.cz/\(S\(aworqik024421urylxmogg43q\)\)/Default.aspx?menu=3113&mode=TextMeta&side=wms.verejne&metadataID=CZ-CUZK-WMS-ZABAGED-P&metadataXSL=metadata.sluzba](https://geoportal.cuzk.cz/(S(aworqik024421urylxmogg43q))/Default.aspx?menu=3113&mode=TextMeta&side=wms.verejne&metadataID=CZ-CUZK-WMS-ZABAGED-P&metadataXSL=metadata.sluzba)
- DOBROVOLNÝ, P. (1998): Dálkový průzkum Země: Digitální zpracování obrazu. Brno. Masarykova Univerzita. 210 s.
- DORIS, K. L., REMMEL, T. K., DREZNER, T. D. (2010): Tracking Desertification in California Using Remote Sensing: A Sand Dune Encroachment Approach. Remote Sensing, 3, 13 s.
- EAGRI (2019): O aplikaci Registr půdy (2019). Dostupné z (30.6.2019):
<http://eagri.cz/public/web/mze/farmar/LPIS/>
- ESA (2012): Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. ESA Communications. 80 s. Dostupné z (30.6.2019):
https://sentinel.esa.int/documents/247904/349490/S2_SP-1322_2.pdf
- ESA: Copernicus (2018). Dostupné z (30.6.2019):
https://m.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview4
- ESA: Copernicus Open Access Hub (2018) Dostupné z (30.6.2019):
<https://scihub.copernicus.eu/dhus/#/home>
- ESA (2018): Sentinel-2 Product Specification Document. Dostupné z (30.6.2019):
<https://sentinel.esa.int/documents/247904/685211/Sentinel-2-Products-Specification-Document>
- FASSNACHT, F. E., LATIFY, H., STERENCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, CH., GHOSH, A. (2016): Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment, 186, 64-87, 25 s.
- FAWAGREH, R., GABER, M., M., ELYAN, E. (2014): Random forests: from early developments to recent advancements. Systems Science & Control Engineering: An Open Access Journal, 2, 8 s.
- GEO/COPERNICUS: Základní principy DPZ (2016): Dostupné z (30.6.2019):
<http://copernicus.gov.cz/zakladni-informace-a-princip-dpz>
- GISAT (20??): Klasifikace. Dostupné z (30.6.2019):
<http://www.gisat.cz/content/cz/dpz/zpracovani-dat/klasifikace>
- GISLASON, P., O., BENEDIKTSSON, J., A., SVEINSSON, J., R. (2005): Random Forest for Land Cover Classification. Pattern Recognition Letters, 27, 7 s.

- GSP 216: Object-based Classification (2015): Dostupné z (30.6.2019):
http://gsp.humboldt.edu/olm_2015/Courses/GSP_216_Online/lesson6-1/object.html).
- GSP 216: Spectral Reflectance (2017). Dostupné z (30.6.2019):
http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson2-1/reflectance.html
- HAGNER, O., REESE, H. (2007): A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment*. 110, 438–444, 7 s.
- HALOUNOVÁ, L., PAVELKA, K. (2008): *Dálkový průzkum Země*. Praha, České vysoké učení technické, 192 s.
- HAVLÍČEK, M., SKOKANOVÁ, H., DOSTÁL, I., VYMAZALOVÁ, M., PAVELKOVÁ, R., PETROVIČ, F. (2018): The consequences of establishing military training areas for land use development—A case study of Libavá, Czech Republic. *Land Use Policy*, 73, 10 s.
- JENSEN, R., J. (2007): *Remote Sensing of the Environment: An Earth Resource Perspective (Second Edition)*. Pearson Prentice Hall, 592 s.
- JHONNERIE, R., SIREGAR, V., P., NABABAN, B., PRASETYO, L., B., WOUTHUYZEN, S. (2015): Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries. *Procedia Environmental Sciences*, 24, 7 s.
- KOLÁŘ, J. (1990): *Dálkový průzkum Země*. Praha, SNTL – Nakladatelství technické literatury, 170 s.
- KULKARNI, A., D., LOWE, B. (2016): Random Forest Algorithm for Land Cover Classification, *International Journal on Recent and Innovation Trends in Computing and Communication*, 4, 3, 6 s.
- LIAW, A., WIENER, M. (2001): Classification and Regression by RandomForest. *Forest*, 23, 5 s.
- LILLESAND, T., M., KIEFER, R., W. (1994): *Remote Sensing and Image Interpretation*. John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 750 s.
- LILLESAND, T. M.; KIEFER R. W.; CHIPMAN, J. W. 2008. *Remote Sensing and Image Interpretation*. 6th ed. United States of America: Wiley, 2008. 756 s.
- LECKIE, D. G., TINIS, S., NELSON, T., BURNETT, CH., GOUGEON, F. A., CLONEY, E., PARADINE, D. (2005): Issue in species classification of trees in old growth conifer stands. *Canadian Journal of Remote Sensing*. 31, 2, 16 s.
- MAHESH, P. (2011): Support vector machines/relevance vector machine for remote sensing classification: A review. *Computing Research Repository*. 18 s.

- MATEMATICKÁ BIOLOGIE (2019): Koncept umělé neuronové sítě. Dostupné z (30.6.2019) <http://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologickych-dat--umela-intelligence--neuronove-site-jednotlivy-neuron--uvod-do-neuronovych-siti--koncept-umele-neuronove-site>
- MOORE, M. M., BAUER, M. E. (1990): Classification of Forest Vegetation in North-Central Minnesota Using Landsat Multispectral and Thematic Mapper Data. *Forest Science*, 36, 2, 330 – 342, 14 s.
- Maxim Dmitrievsky (2018): Random Decision Forest in Reinforcement Learning. Dostupné na (30.6.2019): https://www.mql5.com/en/articles/3856?source=post_page-----
- MVCR (2010): Neznámý svět vojenských újezdů. Dostupné z (30.6.2019): <https://www.mvcr.cz/clanek/neznamy-svet-vojenskych-ujezdu-974223.aspx>
- NA, X., ZHANG, S., LI, X., YU, H., LIU, CH. (2010): Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic Data. *Photogrammetric Engineering & Remote Sensing*, 76, 7, 7 s.
- NASA (2019): Spectral Library. Dostupné z (30.6.2019): <https://speclib.jpl.nasa.gov>
- NICULESCU, S., BILLEY, A., ALI, H. T. O. (2018): Random forest classification using Sentinel-1 and Sentinel-2 series for vegetation monitoring in the Pays de Brest (France). *Remote Sensing for Agriculture, Ecosystems, and Hydrology XX*, 17 s.
- NITZE, I., SCHULTHESS, U., ASCHE, H. (2012): Comparison of Machine Learning Algorithms Random Forest, Artificial Neural Network and Support Vector Machine to Maximum Likelihood for Supervised crop type classification. Institut für Geographie, Universität Potsdam, 6 s.
- NOI, P., T., KAPPAS, M. (2017): Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18, 20 s.
- PAL, M. (2005): Random Forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 1, 6 s.
- POLAMURY, S. (2017): How the Random Forest Algorithm Works in Machine Learning. *Dataaspirant*, Dostupné z (30.6.2019): <https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning/>
- QUINTANO, C., FERNANDÉZ-MANSO, A., FERNANDÉZ-MANSO, O. (2017): Combination of Landsat and Sentinel-2 MSI data for initial assessing of burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 5, 5 s.

- SAPLING LEARNING: Electromagnetic Spectrum. Dostupné z (30.6.2019):
<https://sites.google.com/site/chempendix/em-spectrum>
- SEOS: Introduction to Remote Sensing. Dostupné z (30.6.2019): <http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html>
- SISODIA, P. S., TIWARI, V., KUMAR, A. (2014): Analysis of Supervised Maximum Likelihood Classification for Remote Sensing Image. IEEE International Conference on Recent Advances and Innovations in Engineering, Jaipur, India, 4 s.
- SKOKANOVÁ, H., HAVLÍČEK, M., KLUSÁČEK, P., MARTINÁT, S. (2017): Five military training areas – five different trajectories of land cover development? Case studies from the Czech Republic. *Geographie Cassoviensis XI*, 13 s.
- SMOLOVÁ, I., FŇUKAL, M., SZCZYRBA, Z. (2005): Krajina a lidé vojenského újezdu Libavá. *Geografické rozhledy*, 1, 2 s.
- ŠTYCH, P., PALUBA, D., LAŠTOVIČKA, J., OTRATA, D., HLADKÝ, R. (2018): Hodnocení změn krajinného pokryvu bývalého vojenského újezdu Brdy pomocí dálkového průzkumu Země. *Bohemia centralis*, Praha, 34, 23 s.
- USGS (2019): Spectral library. Dostupné z (30.6.2019):
<https://www.usgs.gov/labs/spec-lab>
- VELJANOVSKI, T., KANJIR, U., OŠTIR, K. (2011): Object-based Image Analysis of Remote Sensing Data. *Geodetski vestnik*, 55, 24 s.
- VOJENSKÝ ÚJEZD LIBAVÁ (2015): Dostupné z (25.4.2018): http://www.vojujezd-libava.cz/vismo/dokumenty2.asp?u=9342&id_org=9342&id=3381

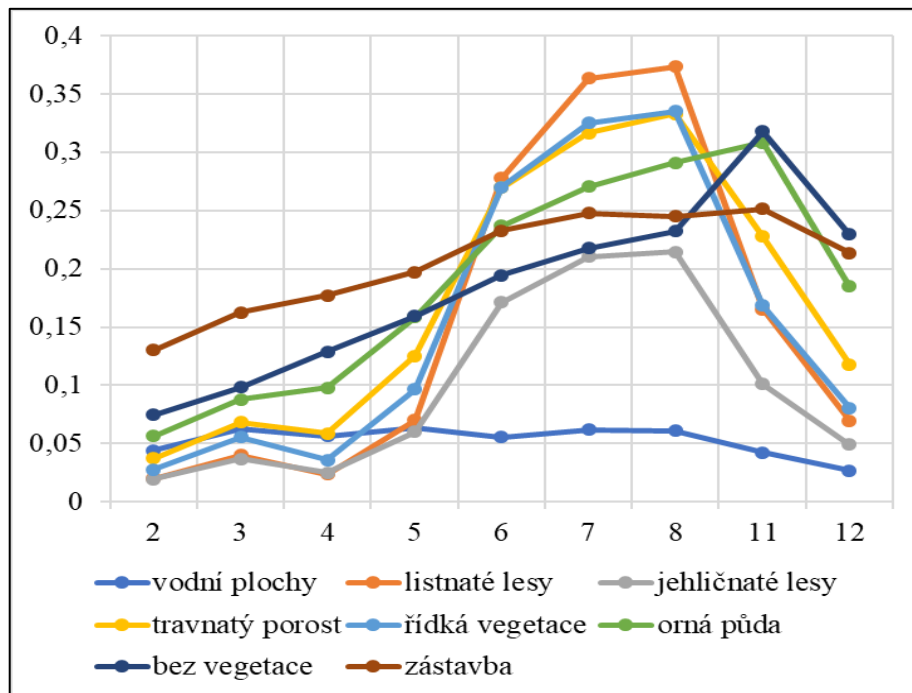
Datové zdroje

- ArcČR 500: Digitální geografická databáze, verze 3.3. Praha: ARCDATA Praha s.r.o.
- ČÚZK (2019): Služba WMS – Ortofoto ČR (2018). Praha. Dostupné z (30.6.2019):
geoportal.cuzk.cz/WMS_ORTOFOTO_PUB/WMSservice.aspx.
- ČÚZK (2019): Služba WMS – ZABAGED (2019). Praha. Dostupné z (30.6.2019):
<http://ags.cuzk.cz/arcgis/services/ZABAGED/MapServer/WmsServer?>
- ESA: Sentinel 2 – level 2A (2019). Copernicus Open Access Hub. Dostupné z (6.7.2019): <https://scihub.copernicus.eu/dhus/#/home>
- EAGRI (2019): LPIS (2019). Dostupný z (6.7.2019):
<http://eagri.cz/public/app/eagriapp/lpisdata/>

Seznam příloh

Příloha 1: Průměrné hodnoty tříd spektrální odrazivost v pásmech Sentinel 2	55
Příloha 2: Zastoupení jednotlivých tříd krajinné pokrývky VÚ Libavá	55
Příloha 3: Mapa klasifikace Random Forest	56
Příloha 4: Mapa klasifikace Maximum Likelihood	56

Příloha 1: Průměrné hodnoty tříd spektrální odrazivost v pásmech Sentinel 2



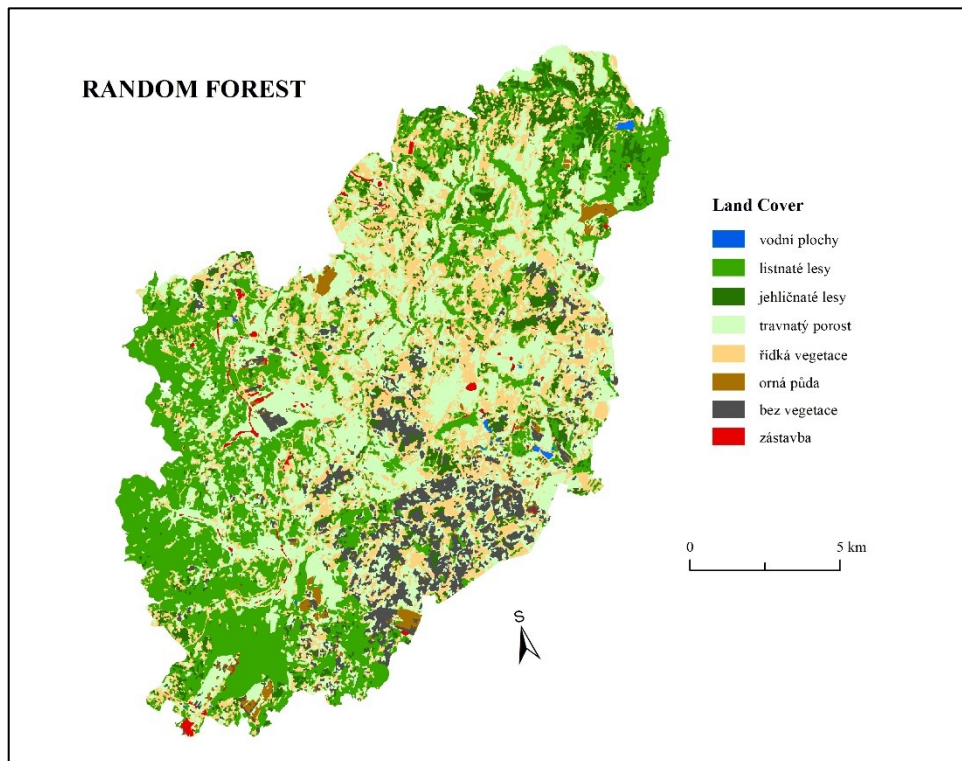
Zdroj: Vlastní tvorba

Příloha 2: Zastoupení jednotlivých tříd krajinné pokrývky VÚ Libavá

	RF	ML	RF	ML
	km ²		%	
vodní plochy	0,4	0,3	0,2	0,1
listnaté lesy	72,3	76,3	29,6	31,2
jehličnaté lesy	18,3	12,8	7,5	5,2
travnatý porost	75,8	52,6	31,0	21,5
řídká vegetace	54,7	74,6	22,4	30,5
orná půda	4,3	5,4	1,7	2,2
bez vegetace	17,1	13,6	7,0	5,6
zástavba	1,6	9,0	0,7	3,7
celkem	244,58		100,0	

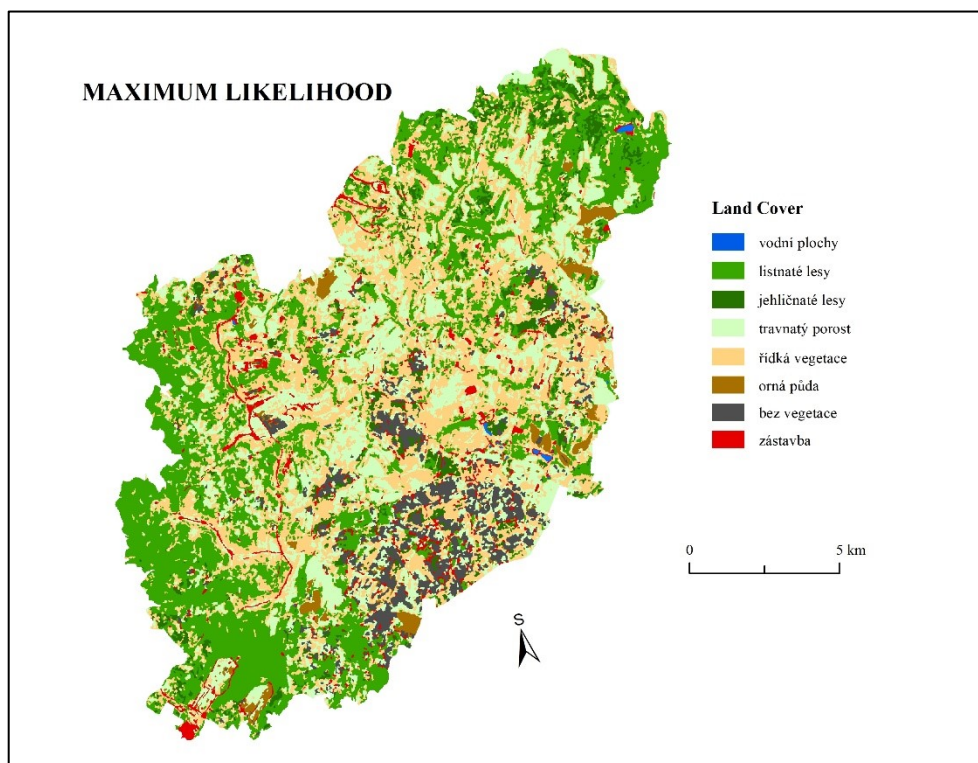
Zdroj: Vlastní tvorba

Příloha 3: Mapa klasifikace Random Forest



Zdroj: Vlastní tvorba

Příloha 4: Mapa klasifikace Maximum Likelihood



Zdroj: Vlastní tvorba