České vysoké učení technické v Praze

Fakulta stavební

Katedra geomatiky

Diplomová práce

Klasifikace vybraných tříd pokrytí území z CORINE systému s využitím družicových dat Sentinel-2

Vedoucí práce: prof. Ing. Lena Halounová, CSc.

Studijní program Geodézie a kartografie

Studijní obor Geomatika

Bc. Lucie Stará

Praha, 2021

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta stavební
Thákurova 7, 166 29 Praha 6

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Bc. Stará Jméno: Lucie Osobní číslo: 458932

Zadávající katedra: geomatiky

Studijní program: Geodézie a kartografie

Studijní obor: Geomatika

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce: Klasifikace vybraných tříd pokrytí území z CORINE systému s využitím družicových dat Sentinel-2

Název diplomové práce anglicky: Classification of selected land cover CORINE system classes using Sentinel-2 remote sensing data

Pokyny pro vypracování:
1) Vybrat vhodná území z družicových dat Sentinel 2A, 2B, kde se obtížně klasifikovatelné třídy, jako jsou pastviny zavlažovaná místa apod., vyskytují v Evropě.
2) Určit vhodné spektrální charakteristiky vybraných tříd pro klasifikace.
3) Navrhnout výběr trénovacích a testovacích ploch.
4) Provést klasifikace včetně vyhodnocení jejich přesnosti.
5) Posoudit podmínky klasifikovatelnosti těchto vybraných tříd pro budoucí klasifikace tříd pokrytí území metodou CORINE.

Seznam doporučené literatury:
CORINE Land Cover — Copernicus Land Monitoring Service. https://land.copernicus.eu/pan-european/corine-land-cover
Image Classification Techniques in Remote Sensing. https://gisgeography.com/image-classification-techniques-remote-sensing/
HASHIM, H., Z. ABD LATIF a N. A. ADNAN. Urban vegetation classification with NDVI threshold value method with very high resolution (vhr) pleiades imagery. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences [online]. 2019, roč. XLII-4/W16, s. 237-240.
BOUČEK, T.: Testování způsobů klasifikace pokrytí území vybraných evropských oblastí. Diplomová práce. 2020

Jméno vedoucího diplomové práce: prof. Ing. Lena Halounová, CSc.

Datum zadání diplomové práce: 15. 2. 2021 Termín odevzdání diplomové práce: 17. 5. 2021
Údaj uveďte v souladu s datem v časovém plánu příslušného ak. roku

Podpis vedoucího práce Podpis vedoucího katedry

III. PŘEVZETÍ ZADÁNÍ

Beru na vědomí, že jsem povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je nutné uvést v diplomové práci a při citování postupovat v souladu s metodickou příručkou ČVUT „Jak psát vysokoškolské závěrečné práce“ a metodickým pokynem ČVUT „O dodržování etických principů při přípravě vysokoškolských závěrečných prací“.

Datum převzetí zadání Podpis studenta(ky)

Čestné prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a s použitím uvedených zdrojů a literatury.

V Praze dne 4.5.2021 .....

Bc. Lucie Stará

Poděkování

Ráda bych poděkovala své vedoucí, paní prof. Ing. Leně Halounové, CSc., za vstřícnost a cenné připomínky, které mi v během zpracování poskytovala. Za odborné konzultace a obohacující diskuze nad řešeným problémem děkuji panu Ing. Lukáši Brodskému, PhDr. V neposlední řadě děkuji rodině a přátelům za podporu a trpělivost, které mi projevovali po celou dobu studia.

Anotace

Diplomová práce se věnuje klasifikaci problematických tříd zavlažovaná orná půda, pastviny a přírodní travní porosty. Klasifikace byla provedena ve třech evropských lokalitách (Španělsko, Makedonie, Turecko). Kromě optických dat Sentinel-2 ke klasifikaci přispěly kanály NDVI a topografická data. Trénovací data byla vytvořena na podkladu databáze CORINE. Pro klasifikaci byla využita metoda Random Forest a určeny nejdůležitější příznaky. Nejlépe se podařilo klasifikovat třídu zavlažovaná orná půda (uživatelská přesnost 98,25 %), dále přírodní traviny (89,30 %) a nakonec pastviny (81,17 %).

Klíčová slova

CORINE, Sentinel-2, klasifikace, Random Forest, land cover, zavlažovaná orná půda, pastviny, přírodní travní porost

Abstract

The diploma thesis is focused on classification of problematic classes irrigated arable land, pastures and natural grassland. The classification was carried out for three European locations (Spain, Macedonia, Turkey). Apart from Sentinel-2 optical data, the NDVI channels and the topographic data contributed in the classification. The training data were created from CORINE. The Ranfom Forest classifier was used and the most important features were determined. The best classification results were obtained for irrigated arable land (producer's accuracy 98,25 %), followed by natural grassland (89,30 %) and pastures (81,17 %).

Keywords

CORINE, Sentinel-2, classification, Random Forest, land cover, irrigated arable land, pastures, natural grassland

Obsah

Seznam zkratek

ANN umělé neuronové sítě (Artificial Neural Network)

ASM druhý úhlový moment (Angular Second Moment)

CLC CORINE Land Cover

CORINE Coordination of Information on the Environment

DEM digitální model terénu (Digital Elevation Model)

DMP digitální model povrchu

DMR digitální model reliéfu

DMT digitální model terénu

DPZ dálkový průzkum Země

DSM digitální model povrchu (Digital Surface Model)

DT rozhodovací strom (Decision Tree)

DTM digitální model terénu (Digital Terrain Model)

EEA Evropská agentura pro životní prostředí (European Environmental Agency)

ESA Evropská kosmická agentura (European Space Agency)

FI důležitost příznaků (Feature Importance)

GIS geografický informační systém

GLCM koincidenční matice stupňů šedi (Grey Level Co-occurrence Matrix)

IDE integrované vývojářské prostředí (Integrated Development Environment)

LC pokryv území (land cover)

LU/LC land use land cover

ML metoda maximální pravděpodobnosti (Maximum Likelihood)

MMU nejmenší mapovací jednotka (Minimum Mapping Unit)

OSGeo Open Source Geospatial Foundation

MSI multispektrální senzor (MultiSpectral Instrument)

NDVI normovaný rozdílový vegetační index (Normalized Difference Vegetation Index)

NIR blízké infračervené pásmo (Near InfraRed)

PCA metoda hlavních komponent (Principal Component Analysis)

RF náhodný strom (Random Forest)

SW software

SWIR krátké infračervené vlny (Short Wave InfraRed)

TOA horní část atmosféry (Top Of Atmosphere)

Kapitola 1

Úvod

Obsah práce vychází z mezinárodního projektu Geo-harmonizer, na kterém se Fakulta stavební podílí. Geo-harmonizer se zaměřuje na propojení a harmonizaci dostupných geografických dat a jejich poskytování skrz webově orientovaný systém. Data s celoevropským rozsahem v něm budou poskytována se zaměřením na různé tematické okruhy jako např. kvalita životního prostředí. Rámec zpracování stojí na využití otevřených dat a metodách strojového učení dostupných v open source software. [1], [2]

Dalším z připravovaných okruhů je i land cover (jinak také pokrytí území, dále LC). Jedná se o charakteristiku, která popisuje fyzický pokryv povrchu Země (např. tráva, voda nebo les). Metody dálkového průzkumu Země jsou pro určení LC často využívané.

Postup klasifikace LC ve zmíněném projektu byl z obecného hlediska již zpracován [3]. Klasifikace byla provedena pro různou tematickou podrobnost (až 28 tříd) ve třech odlišných evropských lokalitách. Využita byla data družice Sentinel-2 a produkt CORINE Land Cover (CLC) jako referenční data. Práce představila postupy pro klasifikaci se zaměřením na metodu maximum likelihood, zhodnotila její úspěšnost a identifikovala některá omezení.

V návaznosti i tato práce používá družicová data Sentinel-2 a trénovací plochy z produktu CLC. Pozornost se zde soustředí na vybrané třídy, které se v provedených klasifikacích ukázaly jako problematické. Dle nomenklatury CLC se jedná o třídy trvale zavlažovaná orná půda, pastviny a přírodní traviny. Pro práci byly zvoleny tři oblasti, kde se všechny vybrané třídy nachází. Na základě odlišnosti těchto lokalit bude možné posoudit univerzální využití zvoleného postupu. Jde o oblasti v Makedonii, Španělsku a Turecku.

Vzhledem ke stanoveným datovým zdrojům bude provedena řízená klasifikace metodou Random Forest z oblasti strojového učení.

Kapitola 2

Cíle

Tato práce se zabývá klasifikací problematických tříd LC. Její cíl lze rozdělit do několika kroků.

Problém klasifikace těchto tříd spočívá především v jejich spektrální podobnosti. Jedním z cílů je proto identifikovat vhodné charakteristiky při tvorbě příznakového prostoru.

Práce má dále za cíl určit vhodný výběr trénovacích a testovacích ploch. Tento krok obnáší navržení úprav podkladových dat CORINE a jejich rozdělení do trénovacího a testovacího setu.

Na základě připravených dat provést klasifikaci a vyhodnocení přesnosti.

Posledním dílčím cílem je zhodnocení dosažených výsledků v závislosti na příznacích, metodě a lokalitách. Posoudit možnosti a případné limity klasifikace těchto tříd s využitím metody CORINE.

Kapitola 3

Řešená problematika

Následující kapitola shrnuje dosavadní nálezy a příklady řešení obdobných úloh.

Obdobná úloha byla řešena s využitím dat Landsat a WorldView [4]. Celková přesnost klasifikace se napříč třídami pohybuje v rozmezí 54 % a 87 %. Nejlepších výsledků (přesnost kolem 87 %) dosahuje orná půda, jelikož se od ostatních liší jak spektrálně, tak texturou. Kultivované plochy jsou také zpravidla homogennější. Mimo to jsou zemědělsky využívané plochy zřetelně ohraničeny, což mělo pozitivní vliv v této objektově orientované klasifikaci. Naproti tomu běžné travní porosty vychází s nejnižší přesností (54 %). Do této třídy jsou řazeny různé druhy trav, čímž je také zdůvodněna nízká přesnost klasifikace. Pro řešení je použit algoritmus Random Forest. Mezi nejvýznamnější příznaky klasifikace patří výšková data (nadmořská výška, orientace svahu) nebo textury. Ze spektrálních dat je to blízké infračervené pásmo. Pro travní porosty je důležité, v jakém období vegetačního vývoje se nachází. Studie uvažuje klasifikaci na základě jednoho období. Jako referenční podklad byla použita data z produktu Tasmanian Land Conservancy (TLC).

Použití časových řad Landsat v podobné úloze uvažuje britská studie [5]. Klasifikace se týká komplexního území, které zahrnuje jak zemědělské oblasti, tak travní porosty. Klasifikace LC byla provedena na základě rozdílů ve fenologii tříd, s využitím vegetačních indexů, a texturálních měr. U homogenních a zřetelně definovaných oblastí (lesy, voda) přesnost klasifikace přesáhla 80 %, zatímco pro více heterogenní a více komplexní oblasti byly výsledky horší (částečně kultivované travní porosty). Za účelem vylepšení výsledků těchto tříd navrhuje studie použití digitálního modelu terénu, a to především z toho důvodu, že nejhůře klasifikované třídy se vykytují v oblastech se specifickou topografií.

Výsledky s opačným poměrem prokazuje studie z Německa [6]. V tomto případě byly

travní porosty klasifikovány s přesností přes 80 %, čímž se řadí mezi nejúspěšnější třídy této klasifikace. I v této úloze byla použita data Landsat a dále vegetační indexy (NDVI, EVI, SAVI a NDMI). Jako referenční data byla použita databáze IACS.

Samostatným problémem je pak klasifikace jednotlivých druhů travin. Problémem odlišitelnosti travin spočívá v podobnosti spektrálních příznaků [7]. Studie používá časové řady NDVI odvozené z produktu MODIS. Data pro trénink a validaci byla shromážděna během terénního průzkumu. Na základě analýzy jsou pro klasifikaci určena období, kdy se od sebe třídy v průběhu roku liší. Toto řešení dosahuje celkové přesnosti 73 %. Ovšem při srovnání jednotlivých tříd jsou velké rozdíly mezi uživatelskou a produkční přesností v řádu desítek procent. Studie dále podotýká, že fenologie závisí na podnebí, liší se tudíž nejen rok od roku, ale i dle lokace. Pro klasifikaci jsou vstupní data nesmírně důležitá a jejich kvalita se promítá do výsledku. Nelze předem stanovit vhodné období, pro spolehlivě příznivé výsledky.

Klasifikace přírodních a zemědělsky využívaných travin byla řešena i s použitím dat z LiDAR [8]. Tento přístup je založen na sledování lidského zásahu – pokud je v oblasti viditelná lidská činnost, což je v tomto případě přítomnost traktorových stop, oblast je jednoznačně kultivovaná. Výzkum má nicméně problémy s odlišením oblastí, kde se tyto třídy míchají, kde probíhá přeměna a objevují se opět původní druhy. Klasifikace pomocí LiDAR dosahuje přesnosti 57 %.

Kapitola 4

Použitá data

Tato kapitola podrobně rozebere použitá data a jejich původ. Pro vytvoření trénovacích dat byla použita data ze systému CORINE. Jako zdroj spektrálních dat byly použity scény z družic Sentinel-2. Přesnost klasifikace byla zvýšena zahrnutím dalších dat jako např. digitální model terénu (DEM).

4.1 CORINE Land Cover

CORINE [1] Land Cover (CLC) představuje seznam 44 tříd, do kterých je LC v Evropě klasifikován. CLC je produkt vytvořený v rámci evropského programu Copernicus, který se zabývá monitorováním Země. Vzniká jako jeden z produktů monitorování LU/LC a zapojeny jsou všechny státy EEA39 [9].

Iniciativa k vytvoření soupisu přišla v roce 1985 a první produkt pochází z roku 1990. Aktualizované verze následovaly v letech 2000, 2006, 2012 a 2018. Vedle vektorové vrstvy LC vznikají i změnové vrstvy (CLC-Changes) [9]. Data CLC jsou volně přístupná [10].

44 tříd je uspořádáno v nomenklatuře (obr. 4.1), která se dle podrobnosti dělí na tři úrovně. První úroveň je obecná a skládá se z 5 tříd - zástavba, zemědělské oblasti, lesy a přírodní oblasti, mokřady a konečně vodní plochy. Druhá úroveň je rozšířená do 15 tříd a třetí úroveň tvoří již zmíněných 44 tříd [9].

[1] (CO-oRdination of INformation on the Environment)

Obrázek 4.1: Nomenklatura CORINE [11]

Obsah databáze CLC je vytvářen na úrovni jednotlivých zemí. Většina z nich produkuje data pomocí vizuální interpretace družicových scén s velmi vysokým rozlišením. Nejmenší mapovací jednotka (MMU) je 25 hektarů u plošných objektů a 100 m šířka u liniových objektů. V důsledku toho mohou být objekty, jejichž rozloha je menší než stanovený limit, generalizované a zahrnuté do jiné třídy. Je proto nutné mít na paměti, že přesnost tohoto modelu je ≥ 85 % [9].

4.1.1 Sledované třídy

Třídy, které jsou sledovány v této práci, pochází z nejpodrobnější - třetí - úrovně systému CLC. Dle CLC směrnice [12] jsou definovány takto:

- 2.1.2 Trvale zavlažovaná orná půda: *„Obdělávané a zemědělsky využívané parcely pro ornou půdu, které jsou trvale nebo periodicky zavlažované s použitím trvalé infrastruktury (zavlažovací kanály, drenážní síť a přídavné zavlažovací zařízení). Většinu těchto plodin nelze pěstovat bez umělé dodávky vody. Nezahrnuje sporadicky zavlažovanou půdu.“*

- 2.3.1 Pastviny, louky a ostatní trvalé travní porosty se zemědělským využitím: *„Oblasti jsou soustavně využívány (minimálně po dobu 5 let) pro produkci krmiva. To zahrnuje přírodní nebo oseté byliny, nekultivované nebo lehce kultivované louky a spásané nebo mechanicky sklízené louky. (...) Pastviny lze popsat jako extenzivně nebo intenzivně obdělávané trvalé traviny, kde se nachází prvky zemědělské infrastruktury, jako například ploty, přístřešky, napajedla, a probíhají zde tyto procesy: spásání, zavlažování, osev a hnojení. Typickým znakem je pravidelný tvar pozemků a/nebo vyšlapané cestičky od zvířat.“*

- 3.2.1 Přírodní travní porosty: *„Travní porosty s žádným nebo mírným zásahem člověka. Traviny s nízkou produktivitou. Často se nachází v drsném, nerovném terénu, ve strmých svazích; nezřídka zahrnují i skalnaté oblasti nebo shluky jiné vegetace. (...) Typickou charakteristikou této třídy je velká rozloha, nepravidelný tvar, zpravidla se nacházejí ve větší vzdálenosti od lidských obydlí.“*

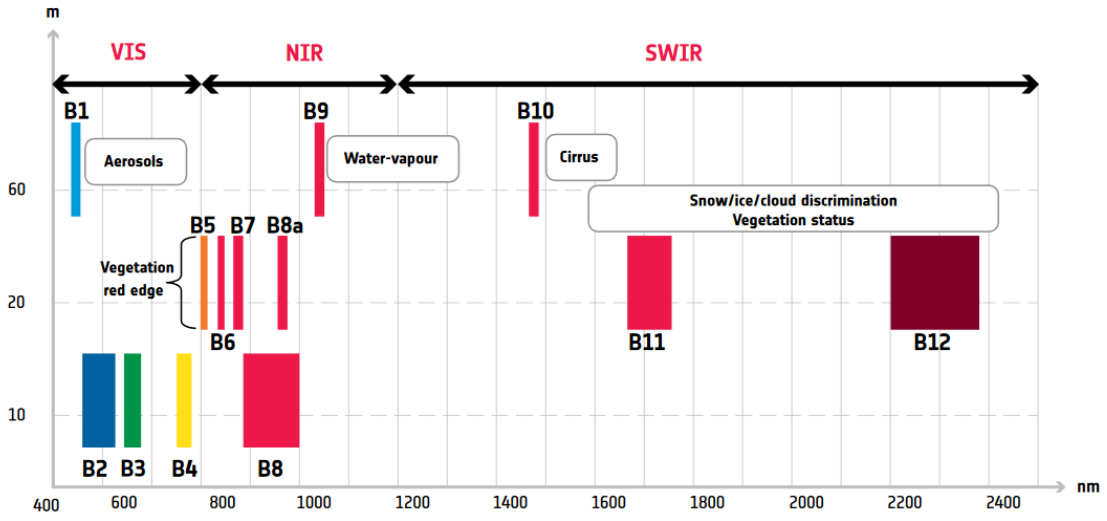
4.2 Sentinel-2

Použití družicových dat pro mapování má oproti pozemním metodám mnoho výhod. Jsou to například častá aktualizace, v mnoha případech téměř okamžitá dostupnost dat a především schopnost zachytit velké území ve velmi krátkém čase. Výběr družice pro konkrétní úlohu se může odvíjet od dostupnosti dat pro řešené období nebo požadovaného rozlišení (prostorové, spektrální, časové). Pro účely této práce byla vybrána data Sentinel-2.

Data Sentinel (včetně mise 2) jsou bezplatně poskytována skrz Copernicus Open Access Hub [13]. Vyhledávání lze limitovat např. dle mise, oblačnosti nebo i zakreslením polygonu. Od roku 2019 Copernicus ukládá některá data do archivu (Long Term Archives) a taková se v rozhraní zobrazí s označením offline. Uživateli je k dispozici náhled a metadata, následně je možné o ně zažádat a jsou přibližně do hodiny zpřístupněna ke stažení [14].

Sentinel-2 je mise programu Copernicus, která je technicky zajišťována Evropskou kosmickou agenturou (ESA). Tvoří ji 2 družice (Sentinel-2A a Sentinel-2B) vypuštěné v roce 2015, resp. 2017. Obíhají po stejné slunečně synchronní dráze a pohybují se v průměrné výšce 786 km. Časové rozlišení jedné družice je 10 dní, při zvážení vzájemného odfázování družic o 180 stupňů je to pouze 5 dní [15], [16].

Spektrální data jsou na palubě každé družice snímána pomocí senzoru, tzv. MSI (MultiSpectral Instrument). Ten snímá informace ve 12 pásmech elektromagnetického záření od viditelného až po střední infračervené (SWIR). Pásma mají různou plošnou rozlišovací schopnost v rozsahu 10 až 60 m (obr. 4.2). Šířka záběru senzoru je 290 km. MSI používá tzv. push-broom skener, což mu umožňuje zaznamenat informace v jeden moment po celé šířce záběru (na rozdíl od whisk-broom skeneru) [15].



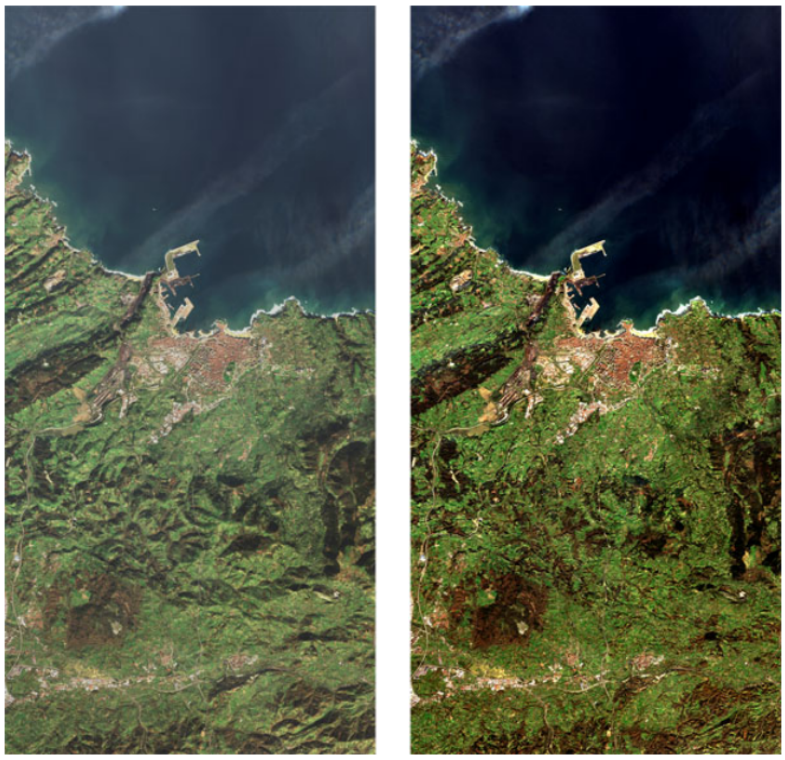
Obrázek 4.2: Pásma Sentinel-2: Vlnová délka vs. prostorová přesnost [17]

4.2.1 Restaurace a rektifikace dat

Originální družicová data zpravidla nejsou ihned vhodná pro využití, jelikož obsahují různá zkreslení, projevuje se vliv atmosféry nebo šum. Tyto nedostatky vyžadují opravu v procesu, kterému se říká předzpracování. Dělí se na dvě části: restauraci a rektifikaci. Restaurace řeší radiometrické chyby tak, že opravuje naměřené hodnoty. Rektifikace řeší geometrickou úpravu a georeferencování (přiřazení polohy pomocí souřadnic) a polohové chyby [18].

Podle úrovně zpracování mají data Sentinel různá označení. Level-0 jsou komprimovaná neupravená data, která stojí na začátku procesu. Postupně jsou vytvořeny produkty Level-1A a Level-1B, které jsou nekomprimované a obsahují radiometrické korekce. Pro uživatele jsou dostupné produkty Level-1C a Level-2A. L1C jsou ortorektifikovaná data z horní části atmosféry (top of atmosphere, TOA), L2A k tomu obsahuje i atmosférické korekce (obr. 4.3). Oba produkty jsou poskytovány po dlaždicích o rozměrech 100 x 100 km v souřadnicovém systému UTM [15],[19].

Veškeré zpracování dat zajišťuje pozemní segment mise, ačkoli pro finální zpracování z Level-1C do Level-2A je uživateli k dispozici tzv. Sentinel Toolbox. [19].



Obrázek 4.3: Atmosférické korekce - srovnání produktů L1C (vlevo) a L2A (vpravo) [20]

4.2.2 Zvýraznění obrazu

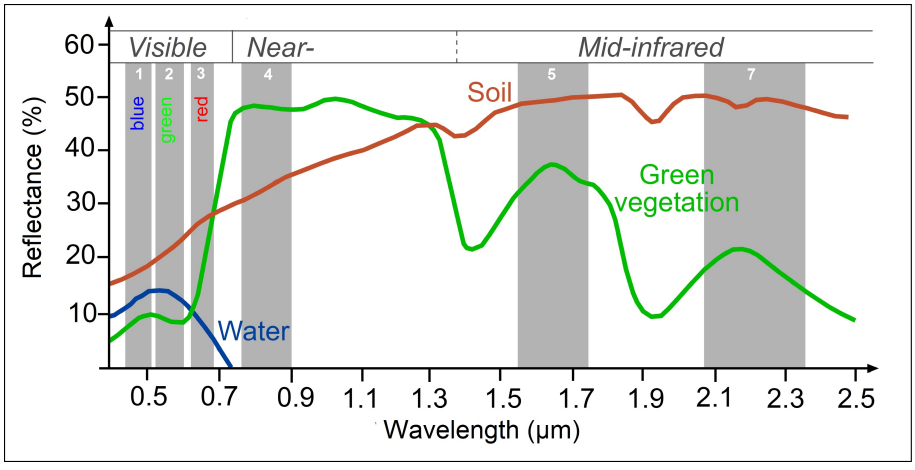
Dalším krokem pro zlepšení vizuální interpretace dat je zvýraznění obrazu. Zavedení těchto úprav zvýrazní odlišnosti, které z původních optických dat nejsou patrné, čímž potenciálně zvýší i úspěšnost klasifikace. Mezi základní se řadí například roztažení kontrastu, aplikace různých filtrů, dále použití více pásem pro výpočet indexů nebo metoda hlavních komponent [18].

Vegetační index NDVI

Normovaný rozdílový vegetační index (normalized difference vegetation index, NDVI) je někdy zvaný také „index zelenosti“. Jedná se o nejrozšířenější typ ze skupiny vegetačních indexů. Vegetační indexy se vypočítávají z kombinací různých pásem, která zvýrazní některé vlastnosti sledovaného jevu. NDVI využívá rozdílné odrazivosti vegetace v červeném (R) a blízkém infračerveném (NIR) pásmu [21].

$$NDVI = \frac{NIR - R}{NIR + R} \tag{4.1}$$

V červeném pásmu vykazuje vegetace významnou absorpci, zatímco v infračerveném pásmu sílí její odrazivost (obr. 4.4), což umožňuje snadné odlišení od ostatních povrchů. Při uplatnění vztahu 4.1 získáme hodnotu, která indikuje množství vegetace v rámci jednoho pixelu. Výsledek nabývá hodnot od -1 do 1 (tab. 4.1). Jednoduchý způsob výpočtu a snadná interpretace z něj právem činí nejpoužívanější z těchto indexů [21].



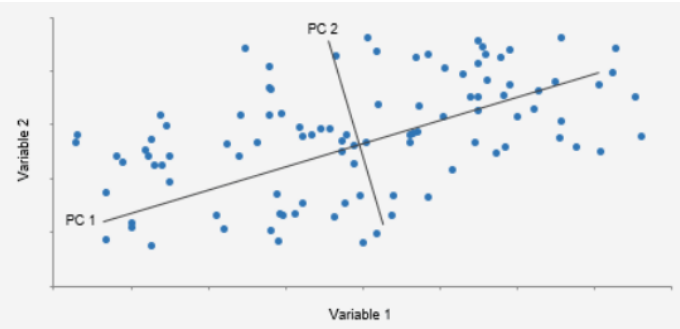
Obrázek 4.4: Spektrální křivky půdy, vody a vegetace (zelená), rozdíl odrazivosti v pásmu R (3) a NIR (4) [22]

hodnota NDVI	typ povrchu
< 0	voda
0 - 0,2	holá půda, zástavba
0,2 - 0,5	řídká až středně hustá vegetace
$> 0,5$	velmi hustá vegetace

Tabulka 4.1: Orientační rozmezí hodnot NDVI pro různé typy povrchů [23], [24]

Metoda hlavních komponent

Použití multispektrálních dat s sebou nese riziko korelace mezi jednotlivými pásmy. Ta může být eliminována použitím metody hlavních komponent (Principal Component Analysis, PCA). PCA stejně jako výpočet indexů využívá více pásem najednou. Jedná se o transformaci, která zavádí nově orientované osy a nový počátek. Nová hlavní osa vede ve směru, kde je rozptyl hodnot ze všech použitých pásem největší. Nový počátek je určen průměrem a druhá osa jím vede kolmo na osu první (obr. 4.5). Data podél ní (a každé další osy) mají mnohem menší rozsah než podél předchozí osy. Vzájemná kolmost os zajišťuje, že korelace mezi daty je odstraněna [18].



Obrázek 4.5: PCA - konstrukce os [25]

Texturální míry

Texturální míry patří k metodám lokálního zvýraznění obrazu [18]. Častou metodou určení textury jsou tzv. Haralickovy funkce [26]. Matice (gray level co-occurence matrix, GLCM) určuje, kolikrát se ve sledované oblasti vyskytuje dvojice pixelů o dané hodnotě a v dané vzdálenosti. Na základě této matice jsou vypočteny konkrétní metriky. Mezi nejčastější patří kontrast, homogenita, korelace, entropie a druhý úhlový moment (angular second

moment, ASM). Textury mohou být využity v případech, kdy jsou spektrální rozdíly sledovaných tříd malé [27].

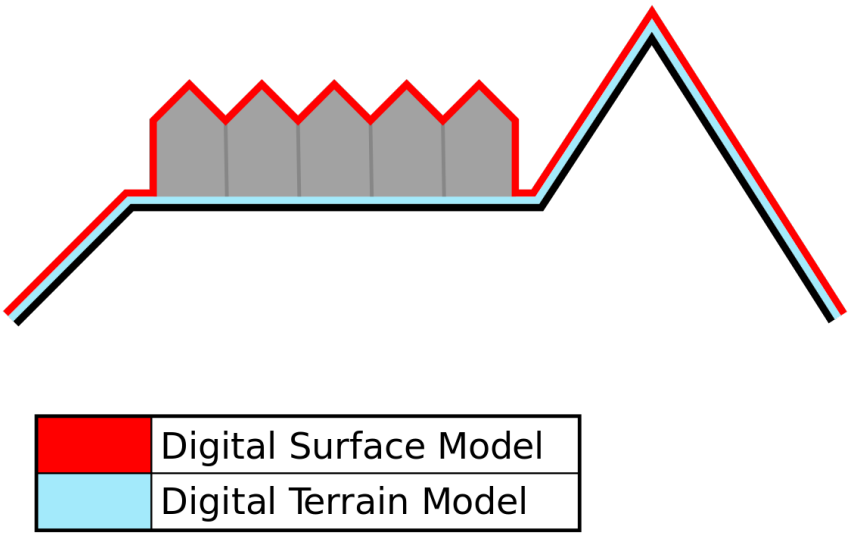
4.3 Digitální model terénu

Výsledky a přesnost klasifikace lze podpořit přidáním dalších informací, které dodají optickým datům širší kontext. Příkladem toho můžou být topografické informace v podobě digitálního modelu terénu.

Digitální model terénu (DMT) je souborné označení produktů reprezentujících výšková data. Zahrnuje digitální model reliéfu (DMR), což je model holé Země a digitální model povrchu (DMP), který zahrnuje i objekty, které se na Zemi nacházejí, vč. vegetace, zástavby ad. [28].

Označení těchto produktů v anglické literatuře se liší. Obecné označení je digital elevation model (DEM). Zobrazení holého zemského povrchu bez vegetace a antropogenních objektů se značí digital terrain model (DTM), zatímco digital surface model (DSM) je reprezentace zemského povrchu se zahrnutím objektů [29].

Rozdíl mezi DMR a DMP, resp. DTM a DSM znázorňuje obr. 4.6.



Obrázek 4.6: Rozdíl mezi DTM a DSM [30]

4.3.1 EU-DEM

Produkt EU-DEM je digitální model povrchu vydaný roku 2016 a je dostupný opět pod hlavičkou programu Copernicus [31]. Jedná se o hybridní model vytvořený na základě vážených průměrů dřívějších výškových modelů misí SRTM a ASTER GDEM.

Použita byla verze EU-DEM v1.1, ve které jsou provedeny opravy jak v polohové přesnosti (s využitím SPOT 2011), tak opravy vertikální přesnosti (pomocí dat ICESat [2]). Model je poskytovaný ve formátu GeoTIFF po dlaždicích 1000 x 1000 km s rozlišením 25 m a směrodatnou odchylkou výškových dat 7 m. Referenční epochou je rok 2011 [33].

[2]Mise pod hlavičkou NASA, monitorující především tloušťku a změny ledu, ale mj. i nadmořskou výšku povrchu [32].

Kapitola 5

Software

V souladu s ideou projektu Geo-Harmonizer byla data zpracována pomocí open source SW.

5.1 QGIS

Pro úvodní zpracování dat byl použit SW QGIS verze 3.10. QGIS je geografický informační systém (GIS), který se řadí do skupiny open source. Jedná se o projekt organizace Open Source Geospatial Foundation (OSGeo), která podporuje a rozvíjí využívání otevřených technologií. V souladu s tím je i QGIS komunitně řízený projekt založený na činnosti dobrovolníků. Mimo zabudované funkce QGIS poskytuje další funkcionalitu pomocí zásuvných modulů, tzv. pluginů, které jsou vytvářeny kýmkoli z komunity [34]. Možnosti použití rozšiřuje knihovna GDAL a nástroje dalších GIS SW jako jsou GRASS a SAGA.

5.2 Spyder

Klasifikace byla zpracována s využitím programovacího jazyka Python v prostředí Spyder. Spyder (Scientific PYthon Development EnviRonment) je integrované vývojářské prostředí (IDE) určené pro práci v jazyce Python [35]. Tento open source SW byl zvolen pro svou celkovou uživatelskou přívětivost (vhodná kombinace možností editace skriptů, zobrazení výsledků ad.). Při klasifikaci byla hojně využívána knihovna scikit-learn, která implementuje metody strojového učení včetně metody RandomForestClassifier [36].

Kapitola 6

Zájmové území

6.1 Zájmové území

Pro srovnání klasifikace v různých částech Evropy byly vybrány tři odlišné lokality. Výběr měl několik kritérií:

- všechny třídy uvnitř jedné dlaždice,
- rozloha třídy,
- geografická poloha dlaždice.

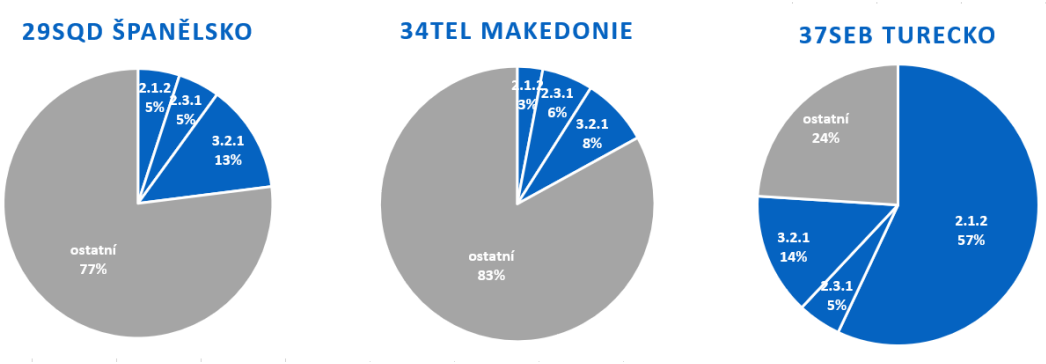
Pro detekci vhodných oblastí byl použit prostorový dotaz a vizuální interpretace. Použita byla síť dlaždic Sentinel-2 [37] a z CLC dat byly vybrány pouze sledované třídy. Výběr byl proveden na základě výskytu všech 3 tříd v rámci jedné dlaždice. Zatímco třídy 2.3.1 a 3.2.1 se vyskytují téměř po celé Evropě, třída 2.1.2 je dominantou Středomoří. Výběr proto směřoval do této oblasti.

Dle vizuálního posouzení jsou na kombinaci těchto tříd jednoznačně nejbohatší Turecko a Španělsko. V rámci zachování rozmanitosti bylo nutné najít třetí lokalitu. V tomto případě byly rozhodující jak geografická poloha, tak rozloha tříd v dané oblasti. Itálie v tomto ohledu byla velmi chudá, jelikož třída 2.1.2 na většině dlaždic měla rozlohu menší než 1 %[1]. V oblasti balkánského poloostrova je situace obdobná v případě třídy 2.3.1. Přijatelné zastoupení všech tříd bylo nalezeno v oblasti v Severní Makedonii. Polohu vybraných lokalit ilustruje obr. 6.1.

[1] Rozloha byla určena pomocí pluginu *Dissolve with stats*.

Obrázek 6.1: Zájmová území: poloha scén Sentinel-2 a jejich označení

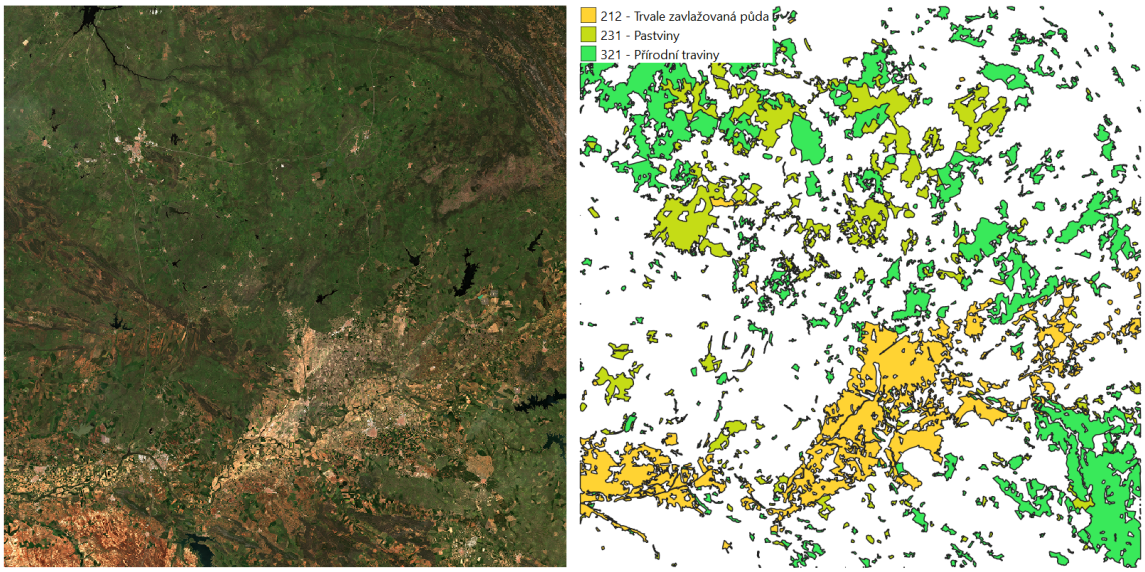
Obr. 6.2 znázorňuje procentuální zastoupení sledovaných tříd ve vybraných lokalitách. Je zřejmé, že rozloha tříd v rámci jednotlivých lokalit se liší. Sledované třídy zabírají nejvíc rozlohy v lokalitě v Turecku, zatímco nejmenší území zabírají v Makedonii. Různé je i zastoupení konkrétní třídy v různých lokalitách. Pro třídu 2.1.2 se liší od 3 do 57 %. Oproti tomu nejstabilnější zastoupení má třída 2.3.1, jejíž rozloha se pohybuje mezi 5 a 6 %.



Obrázek 6.2: Procentuální zastoupení tříd ve vybraných lokalitách

6.1.1 Španělsko

Scéna zachycuje oblast v okolí města Merida v provincii Badajoz na západě Španělska. Jedná se především o zemědělsky využívanou oblast s minimem zástavby a vodních ploch. Nadmořská výška se zde pohybuje od 200 do 1 500 m.n.m. Sledované třídy jsou rozptýleny po celé oblasti, s výjimkou třídy 2.1.2, která se rozkládá spíše v jižní nížinaté části. Jedná se o suchou oblast se středomořským podnebím, léta jsou velmi suchá, zimy spíše mírné a vlhké [38].

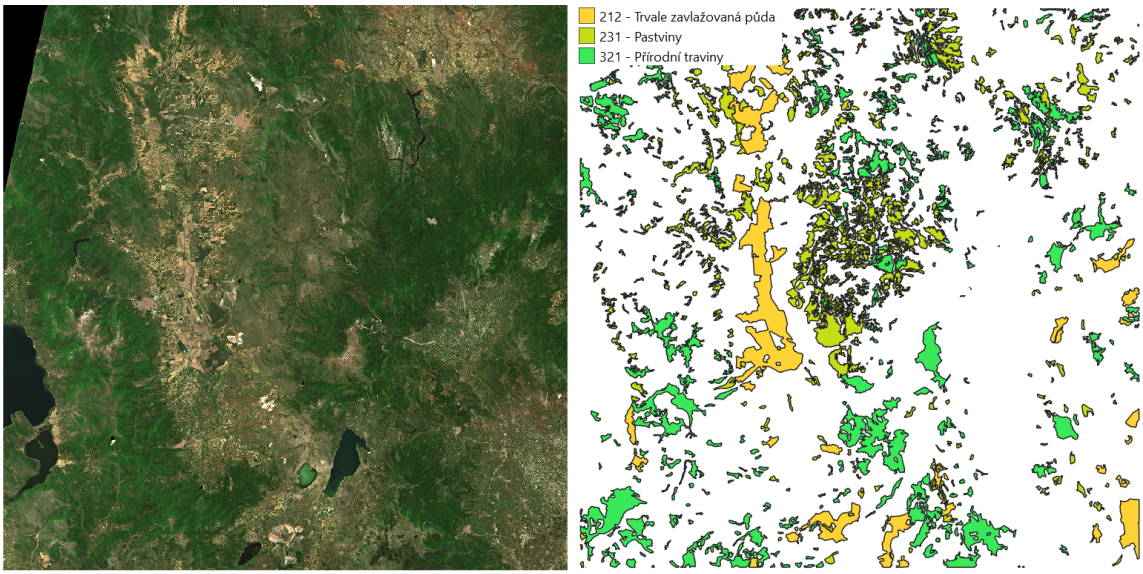


(a) scéna v pravých barvách (b) distribuce LC tříd v oblasti

Obrázek 6.3: Scéna 29SQD (Španělsko) - srovnání družicových a CLC dat

6.1.2 Severní Makedonie

Tato scéna zobrazuje oblast v regionu Pelagonie v jižní části země na hranici s Řeckem. Z vybraných oblastí jsou zde nejvýraznější přechody mezi nadmořskou výškou, která se rozpíná od 0 do 2 500 m.n.m. Významná část oblasti se nachází ve výškách nad 1 500 m.n.m., kde se nachází i nejpočetnější LC této oblasti - lesy a polopřírodní oblasti. Druhou nejpočetnější třídou jsou zemědělské oblasti. Klima je zde kontinentální, s mírnějšími letními teplotami a studenými zimami [38].

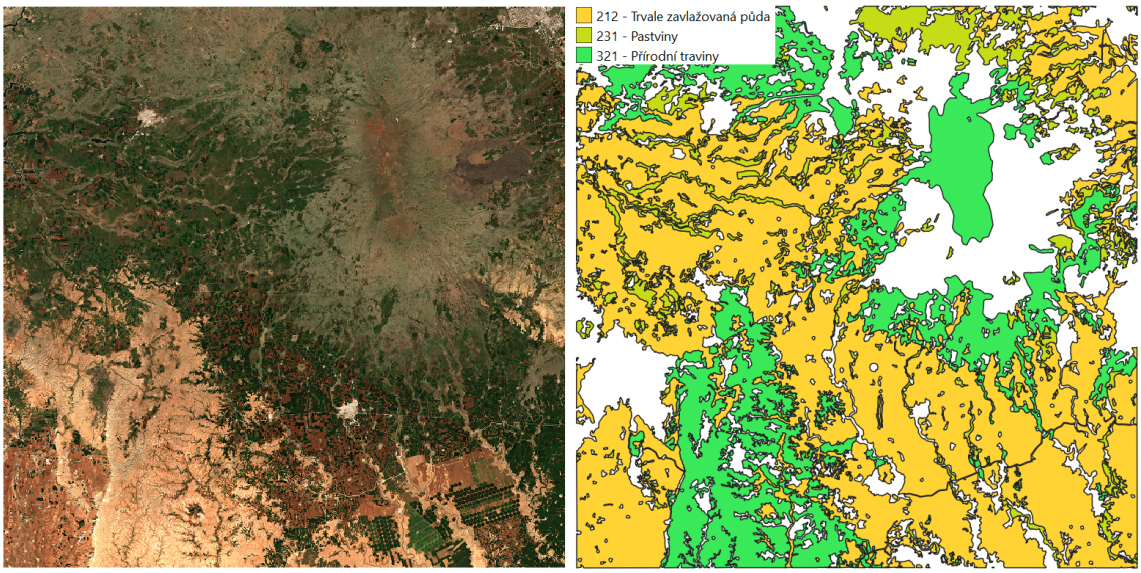


(a) scéna v pravých barvách (b) distribuce LC tříd v oblasti

Obrázek 6.4: Scéna 34TEL (Makedonie) - srovnání družicových a CLC dat

6.1.3 Turecko

Zachycená oblast se nachází v okolí města Viransehir na jihovýchodě Turecka. Nadmořská výška od jihu (400 m.n.m.) stoupá až ke 2 000 m.n.m. Sledované třídy zde zabírají největší část oblasti. Opět dominuje zemědělství, ve vyšších nadmořských výškách se nacházejí přírodní traviny. Pastviny se zde vyskytují především v severní části a v okolí řek. Obdobně jako ve Španělsku i tato oblast je velmi suchá [38].



(a) scéna v pravých barvách (b) distribuce LC tříd v oblasti

Obrázek 6.5: Scéna 37SEB (Turecko) - srovnání družicových a CLC dat

6.2 Sběr družicových dat

Klasifikace na základě jedné scény může být velkou výzvou [7], obzvlášť v případech, kdy je obtížné některé LC třídy odlišit [39]. V takových případech je vhodné použití multitemporálních dat. Jejich výhoda tkví v možnosti klasifikovat dané LC třídy na základě odlišností v různých obdobích [39].

Při výběru družicových dat byla zohledněna i vegetační aktivita v dané oblasti. Ta se v závislosti na klimatických podmínkách a přístupu k zemědělství může lišit napříč lokalitami, a to jak v délkou, tak částí roku, kdy probíhá. V Turecku a Španělsku byla aktivita zjevná již v březnu. Od května přirozená vegetace postupně usychala v závislosti na stoupající teplotě. V makedonské lokalitě svou roli sehrála i nadmořská výška, jelikož zde byla část oblasti do konce dubna pokryta sněhem. Stárnutí vegetace se začala projevovat během srpna.

Výběr dat omezovala i často přítomná oblačnost, a to především v jarních měsících, kdy byla vegetace nejaktivnější. Turecko bylo v tomto ohledu nejméně problematické, s výjimkou velmi oblačného května.

Touto vylučovací metodou byly vybrány scény shrnuté v tab. 6.1. Družicová data pro Španělsko a Turecko byla vybrána z roku 2018, jelikož k tomuto roku byla vydána použitá verze CLC. V Makedonii byly scény z roku 2018 velmi oblačné, proto byla použita data z roku 2019.

Španělsko (2018)	Makedonie (2019)	Turecko (2018)
28.3.	8.6.	19.3.
17.4.	3.7.	23.4.
17.5.	7.8.	23.5.
16.6.	16.9.	7.6.
16.7.	16.10.	12.7.

Tabulka 6.1: Multitemporální data - vybrané scény

Kapitola 7

Metodika

Klasifikace je automatický proces, během kterého jsou vstupní data převedena do tematické mapy [40]. Samotná klasifikace je součástí komplexního procesu. Jeho průběh se liší v závislosti na zvolené metodě klasifikace. Proto budou nejprve představeny různé druhy klasifikace, poté postup a zásady pro zvolený typ klasifikace, v závěru kapitoly bude popsána použitá metoda a její vyhodnocení.

7.1 Metody klasifikace

Typy klasifikačních metod lze rozdělit podle různých aspektů. V následujícím oddílu jsou shrnuty některé z těch hlavních.

Nejčastěji se metody dělí na řízenou a neřízenou. To se odvíjí od toho, zda je předem k dispozici informace třídách, do kterých bude klasifikace provedena. Pokud ano, použije se klasifikace řízená (supervised classification), kdy je na základě trénovacích ploch vytvořeno klasifikační pravidlo a s tím pak probíhá klasifikace. Řadí se sem metody maximální pravděpodobnosti (maximum likelihood, ML), umělé neuronové sítě (artificial neural network, ANN) a rozhodovací stromy (decision tree, DT). Pokud není k dispozici a priori informace, probíhá neřízená klasifikace (unsupervised classification). Data jsou setříděna do shluků (clusters) na základě podobnosti a třída je jim přiřazena až následně. Běžné metody pro tuto klasifikaci jsou např. ISODATA a k-means [41].

Dále se metody dělí na parametrické a neparametrické. Parametry se určují na základě trénovacích dat. Vychází z předpokladu, že data odpovídají normálnímu rozdělení. Nejčastější používaná parametrická klasifikace je ML, jelikož je jednoduchá na implementaci.

Neparametrické klasifikátory nevyužívají normální rozdělení. Jsou vhodné pro klasifikace, které využívají data z více zdrojů. V takovém případě mají lepší výsledky než klasifikátory parametrické. Jsou to již zmíněné ANN, DT a metoda podpůrných vektorů (support vector machine, SVM) [41].

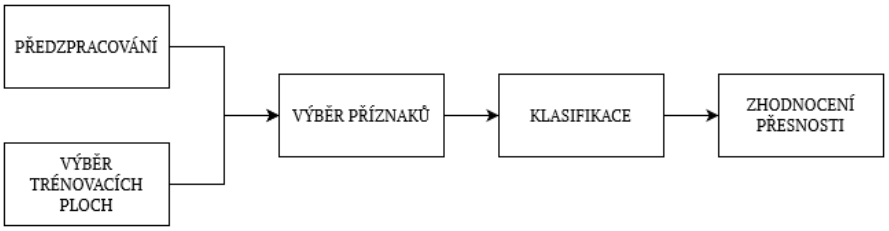
V neposlední řadě se rozdělení odvíjí od přístupu k pixelu. Většina doposud zmíněných metod používá techniku pixel po pixelu (per pixel), která klasifikuje každý pixel na základě jeho hodnot. U družicových dat s menším rozlišením často vznikají smíšené pixely. Pro tento typ problému se používají klasifikace uvnitř pixelu (subpixel), např. fuzzy klasifikace. Posledním typem jsou objektově orientované algoritmy (také per-field). Základní jednotkou není pixel, nýbrž skupiny pixelů, které tvoří jednotlivé objekty. Tato metoda se hodí pro scény s vysokým rozlišením [41].

Jak z výše zmíněných způsobů vyplývá, před výběrem metody pro konkrétní úlohu je nutné vzít v potaz nejen dostupná data, ale i schopnosti jednotlivých metod. Dalším rozhodovacím faktorem může být i dostupný SW a samozřejmě požadovaná přesnost [41]. Je však důležité podotknout, že výsledek klasifikace se neodvíjí pouze od zvolené metody. Je to komplexní proces, ve kterém je neméně důležitý samotný výběr trénovacích ploch, zpracování obrazových dat nebo vhodné klasifikační schéma [40].

Vzhledem k datům určeným pro tuto práci byla zvolena řízená klasifikace. S použitím CLC je k dispozici apriorní informace. Prostorové rozlišení Sentinel-2 indikuje tradiční klasifikaci na základě pixelu. Mimo optická data bude pro klasifikaci použit i DEM, vhodnější se tedy jeví použití neparametrické metody. Touto vylučovací metodou byl pro klasifikaci zvolen klasifikátor Random Forest.

7.1.1 Řízená klasifikace

Postup řízené klasifikace popisuje obr. 7.1. Prvními kroky jsou předzpracování a výběr trénovacích ploch. Následuje výběr vhodných příznaků, klasifikace a vyhodnocení výsledků.



Obrázek 7.1: Schéma řízené klasifikace

Výsledky závisí nejen na použité metodě. Do klasifikace vstupuje celá řada faktorů, které se ve výsledcích projeví. S ohledem na dosažení uspokojivých výsledků lze celý proces opakovat se zaměřením na jednotlivé činitele:

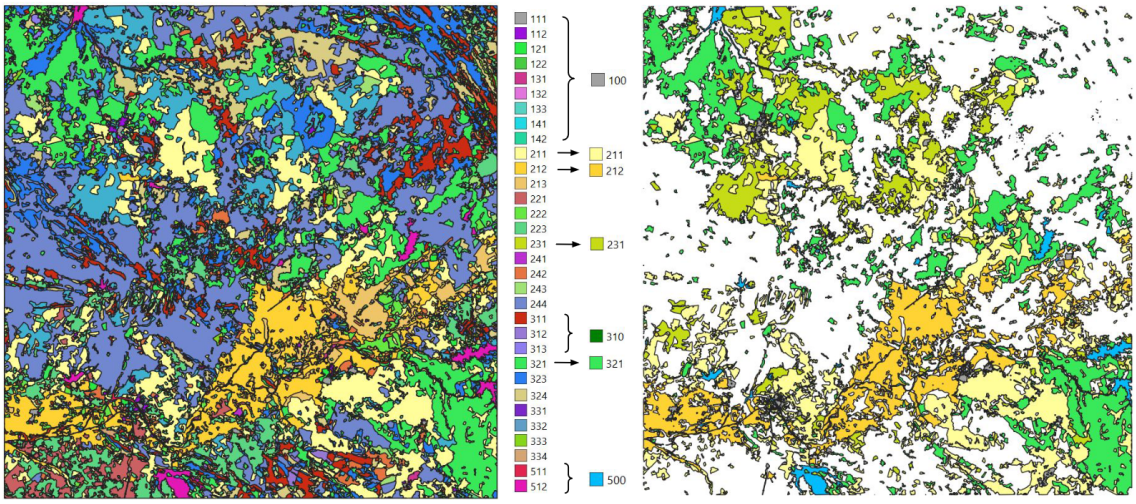
- trénovací plochy,
- vstupní data,
- použitá metoda a parametry.

7.1.2 Klasifikační schéma

Před výběrem trénovacích ploch bylo nutné vybrat třídy, pro které bude klasifikace probíhat. Klasifikační schéma bylo posouzeno pro každou lokalitu zvlášť v závislosti na přítomných třídách.

V každé oblasti se vyskytuje přes 20 tříd ze třetí úrovně CLC nomenklatury. Hlavním cílem úpravy bylo vybrat třídy tak, aby počet ostatních tříd výrazně nepřevažoval nad počtem sledovaných tříd. Zároveň bylo žádoucí zachovat rozmanitost výstižnou pro danou lokalitu. Některé třídy byly na základě vzájemné podobnosti LC agregovány, jiné byly vyřazeny.

Úpravu schématu demonstruje obr. 7.2. Do třídy s označením 100 byly zahrnuty všechny třídy třetí úrovně CLC, které reprezentují zástavbu. Obdobně vznikly i třídy 310 a 500, které reprezentují lesní porost a vodní plochy. Jedná se o označení typů LC, které jsou v oblasti zachovány, ale pouze ve zobecněné podobě. Nové klasifikační schéma pochopitelně obsahuje sledované třídy (212, 231, 321). Pro srovnání byla zahrnuta i třída 211 - nezavlažovaná orná půda.



Obrázek 7.2: Úprava klasifikačního schématu - eliminace a agregace tříd LC

Ostatní třídy byly ponechány stranou. Jedním z důvodů byla zanedbatelná rozloha těchto tříd ve scéně, např. třídy mokřadů, jejichž rozloha byla menší než 1 % a pro tuto práci nepotřebná. Druhým důvodem byla možná záměna se sledovanými třídami na základě vizuální podobnosti. Byly vyřazeny ostatní třídy z kategorie zemědělské oblasti (označení tříd začíná číslem 2), a zbylé třídy kategorie lesy a polopřírodní oblasti (označení začíná 32 nebo 33).

Tímto postupem byla schémata upravena pro všechny oblasti. Výsledek je shodný pro Španělsko a Makedonii, v Turecku chybí zástupci třídy 310 (tab. 7.1).

Španělsko	Makedonie	Turecko	popis
100	100	100	zástavba
211	211	211	nezavlažovaná orná půda
212	212	212	trvale zavlažovaná orná půda
231	231	231	pastviny
310	310	-	lesy
321	321	321	přírodní travní porost
500	500	500	vodní plochy

Tabulka 7.1: Klasifikační schéma - všechny lokality

7.1.3 Trénovací data

Po vytvoření klasifikačního schématu následovala příprava trénovacích dat.

Pro provedení klasifikace je nutné vytvořit tzv. klasifikační pravidlo. V případě řízené klasifikace se pro vytvoření klasifikačního pravidla použijí trénovací plochy. Trénovací plochy (jinak také referenční data) nesou informaci o rozdělení sledované oblasti do jednotlivých tříd. Analýzou vztahu mezi příznaky a jejich zařazením vzniká klasifikační pravidlo.

Příznaky jsou jednotlivé složky obrazu, ve kterých lze jednotlivé třídy odlišit. Jelikož klasifikační pravidlo není většinou předem známé, je nutné ho pro každou úlohu stanovit zvlášť. Na základě klasifikačního pravidla klasifikátor rozpoznává a zařazuje pixely do tříd podle vzoru [18].

Trénovací plochy jsou sestaveny na základě apriorní znalosti daného území. Informace mohou být získány průzkumem v terénu (zpravidla body), sestavením polygonů na základě vizuální interpretace nebo využitím nějakého existujícího produktu [18].

Úspěšná klasifikace vychází z dobře sestaveného klasifikačního pravidla. Proto je nutné i vhodně upravit trénovací plochy, které by měly splňovat tyto zásady:

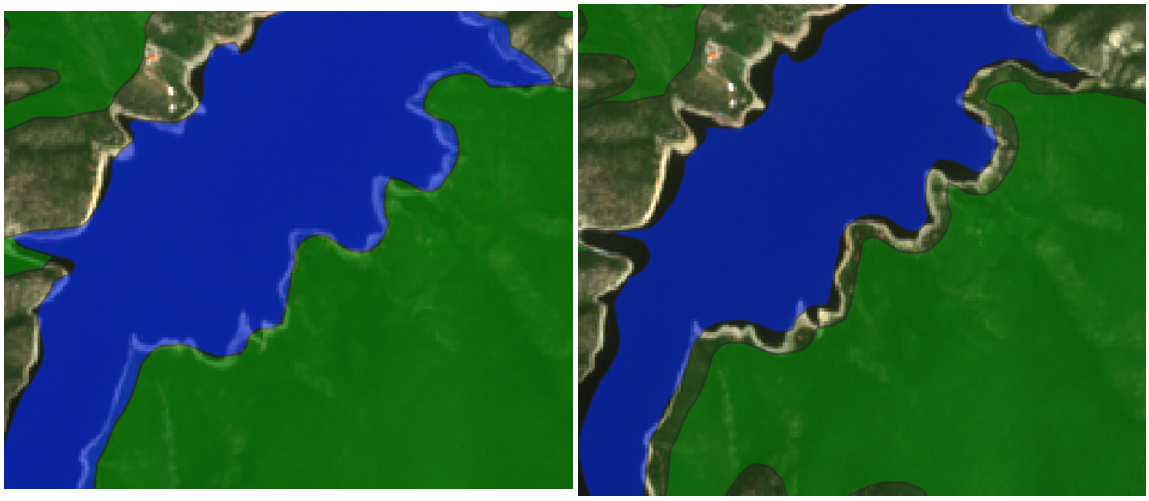
- oddělitelnost,
- reprezentativnost,
- kompletnost [18].

Příprava trénovacích dat

Apriorní informace byla v tomto případě obsažena v datech CORINE. Přípravné práce byly provedeny v SW QGIS. Vrstva CLC byla nejprve oříznuta podle příslušné dlaždice Sentinel-2 (*Extract/clip by extent*), tak aby pokrývala pouze území vybrané scény. Následně byla vektorová vrstva převedena do odpovídajícícho zobrazení (*Reproject*)[1].

Hranice mezi jednotlivými LC jsou generalizované a nemusí věrně vystihovat realitu. V důsledku toho je možné, že mezi pixely na této hranici dojde k záměně. Proto byla použita funkce *Buffer* s hodnotou -40 metrů (2 pixely při rozlišení 20 m), aby mezi jednotlivými LC vznikla mezera a zamezilo se tím případné nesprávné klasifikaci (obr. 7.3).

[1] V závislosti na dané lokalitě šlo o projekce s EPSG kódy 32629, 32634, 32637.

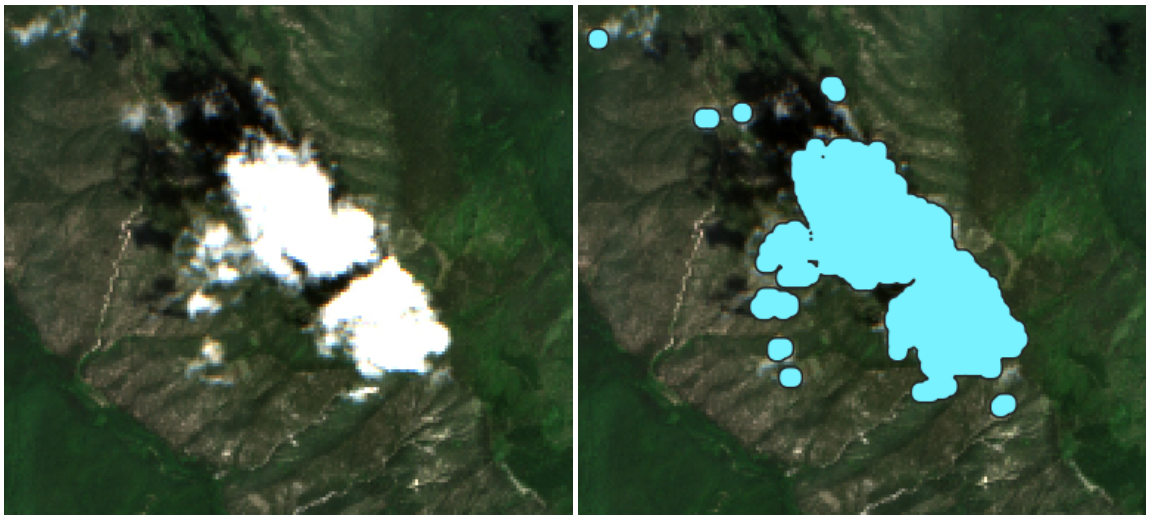


(a) CLC bez úprav

(b) vnitřní buffer

Obrázek 7.3: Úprava trénovacích ploch - hranice LC tříd

V případě oblačnosti bylo nutné tyto oblasti z trénovacích dat odstranit. Byla k tomu využita vrstva SCL (Scene classification), která je součástí dat Sentinel-2. Z této rastrové vrstvy byly pomocí nástroje *Raster Calculator* vybrány pixely, které byly klasifikovány jako mraky. Vytvořená vrstva byla dále vektorizována s využitím nástroje *Polygonize (raster to vector)* z knihovny GDAL. Aby vytvořená maska zakrývala mraky až do okrajů, byl navíc aplikován buffer 60 metrů, aby se daná oblast rozšířila (obr. 7.4). Oblasti byly z CLC vrstvy odstraněny funkcí *Difference*.



(a) oblačnost

(b) maska

Obrázek 7.4: Úprava trénovacích ploch - oblačnost

Posledním krokem bylo vytvoření generalizované bodové vrstvy, která byla použita pro další zpracování a klasifikaci. Nejprve byla vytvořena pomocná pravidelná čtvercová

mřížka s rozestupem 500 m (nástroj *Create grid*). Vrstva CLC byla touto mřížkou rozdělena použitím nástroje *Intersection* a na základě jeho výstupu byly vytvořeny centroidy (funkce *Centroids*).

Každému centroidu byly pomocí nástroje *Field Calculator* vypočteny souřadnice. Pomocí funkce *Join Attributes by Location* byla bodům přiřazena příslušná LC třída z vrstvy CLC na základě geometrického vztahu *within*. Pokud byl centroid vytvořen v místě, kde se žádný CLC polygon nenacházel, byl vyřazen na základě zvolené možnosti *Discard records which could not be joined*. Informace obsažené v rastrových vrstvách byly centroidům zapsány pomocí funkce *Add Raster Values to Points*. Bodová vrstva byla poté exportována ve formátu CSV a připravena pro další zpracování.

Úprava trénovacích dat

V průběhu klasifikace byly v trénovacích datech zjištěny některé nedostatky, které zde budou popsány. Pro vyhodnocení potřebných úprav byly výsledky zobrazeny v QGIS. Chybně klasifikované body byly porovnány s družicovými daty. Na základě vizuálního posouzení byly zřetelné tyto typy záměn:

- vegetace v zástavbě,
- vodní plochy v zástavbě,
- vegetace ve vodní plochách,
- zástavba ve vegetaci.

V jednotlivých třídách CLC se mohou objevit jiné typy LC, které do nich správně nepatří, ale jejichž velikost je menší než stanovená MMU při produkci CLC (např. silnice užší než 100 metrů nebo plochy menší než 25 ha). Právě tyto případy byly nejčastější příčinou výše zmíněných záměn. Zřetelný důvod záměny mezi jednotlivými LC vegetace pozorován nebyl.

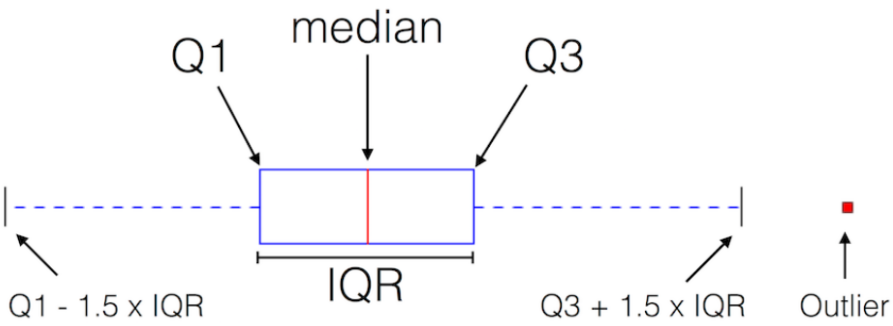
Zjištěné nedostatky byly v referenčních datech opraveny na základě hodnoty NDVI. Z daného typu LC byly vyřazeny všechny body, jejichž hodnoty NDVI do daného intervalu nespadaly. Mezní hodnoty NDVI (tab. 4.1) byly zvoleny kompromisem mezi studovanou literaturou a pozorovanými hodnotami v dané oblasti. Rastr NDVI byl vytvořen na základě vztahu 4.1 s použitím pásem 4 (R) a 8a (NIR) a nástroje *Raster Calculator*.

Další provedenou úpravou trénovacích dat bylo vyřazení odlehlých měření. Hodnoty

těchto bodů nebyly pro sledovaný jev reprezentativní, a tak byly odstraněny. Pro stanovení odlehlých měření byly použity následující vztahy:

$$\text{odlehlé měření} < 1QR - 1,5 \times IQR,$$
$$\text{odlehlé měření} > 3QR + 1,5 \times IQR,$$

kde 1QR je první kvartil, 3QR je třetí kvartil a jejich rozdíl tvoří mezikvartilové rozpětí (IQR) (viz obr. 7.5) [42].



Obrázek 7.5: Odlehlá měření [43]

Tento proces byl aplikován pro každý příznak, ve kterém byly postupně upraveny body jednotlivých tříd. Celkový počet odebraných bodů závisel na použitých příznacích. V důsledku toho se mohl počet bodů lišit při klasifikacích, ve kterých byly použity odlišné příznaky.

Klasifikace byla nejprve prováděna pro určené klasifikační schéma (tab. 7.1). V druhé fázi klasifikace byly ostatní třídy ponechány stranou a do klasifikace byly vybrány pouze tři sledované třídy (212, 231, 321). Cílem výběru bylo zjistit, jak klasifikace proběhne, pokud budou odstraněny zbylé potenciálně rušivé třídy, a získat představu o odlišitelnosti pouze mezi vybranými typy LC.

V posledním kroku se úprava referenčních dat týkala počtu vstupních bodů. Ten dosud nebyl nijak regulován a každá třída byla v klasifikaci zastoupena jiným počtem vzorků. V tomto kroku byl z každé třídy vybrán stejný počet bodů, tak aby byl jejich počet vyrovnán. K tomu byla použita metoda *RandomUnderSampler*.

7.1.4 Příznaky

Za účelem zlepšení výsledků klasifikace, byly zařazeny různé druhy příznaků, které by pomohly v odlišitelnosti jednotlivých tříd dle různých aspektů. Do klasifikace byly kromě optických dat zařazeny tyto příznaky: NDVI, PCA, texturální míry a topografická data. V této části bude popsáno zpracování těchto příznaků před klasifikací.

Družicová data

V práci bylo použito 9 optických pásem Sentinel-2 s rozlišením 20 m (B2, B3, B4, B5, B6, B7, B8a, B11 a B12). Hlavní komponenty (PCA) byly vytvořeny pomocí pluginu *PCA4CD* na základě těchto 9 optických pásem. Do klasifikace byly použity první 3 komponenty (PCA1, PCA2, PCA3), jelikož na rozdíl od zbylých pásem v nich byly patrny spektrální rozdílnosti. Pásma použitá pro výpočet PCA musela být umístěna v jednom rastru, k čemuž byl využit nástroj *Build Virtual Raster*.

Texturální příznaky byly vytvořeny s použitím GRASS funkce *r.texture*. Velikost pohyblivého okna byla nastavena na 3 pixely a vzdálenost mezi dvěma vzorky na 1 pixel. Na základě studované literatury [44], [45] byly vytvořeny míry, jejichž použití se opakuje: korelace, homogenita, entropie a druhý úhlový moment. Literatura doporučuje vytvoření měr na základě prvního pásma PCA, to se ale v tomto případě neosvědčilo a lepších výsledků bylo dosaženo pomocí pásma B2.

Tvorba rastru NDVI byla popsána v předchozím oddílu 7.1.3.

Topografická data

Produkt EU-DEM bylo nejprve nutné limitovat pouze na zájmové území. Bylo tak učiněno ve dvou krocích. První byla změna zobrazení z původní ETRS89-LAEA (EPSG 3035) do WGS84 (EPSG 32634) pomocí GDAL nástroje *Warp (Reproject)*. Následně byl rastr oříznut podle příslušné dlaždice (funkce *Clip Raster by Extent*).

Na základě tohoto rastru byl následně vytvořen rastr sklonitosti (nástroj *Slope*). Sklonitost neboli míra svahu udává sklon terénu v daném místě (pixelu) v hodnotách ve stupních.

7.2 Strojové učení

Strojové učení (machine learning) je založeno na principu, kdy počítač řeší daný problém na základě nabyté zkušenosti [46]. Na základě dat, která jsou stroji poskytnuta, se samostatně učí vztahy mezi nimi, aniž by uživatel definoval jednotlivé kroky postupu. Na základě naučených vztahů pak předpovídá výsledek [47].

Obvyklým postupem při řešení úlohy pomocí strojového učení je rozdělení dat na trénovací a testovací. Pomocí trénovacího datasetu probíhá fáze učení (tvoří se klasifikační pravidlo). V trénovací fázi má klasifikátor k dispozici jak příznaky, tak informaci, do které třídy data zařadit (labels), a učí se vztahy mezi nimi. Při testování má klasifikátor k dispozici pouze příznaky, které klasifikuje na základě natrénovaných vztahů. Labels jsou v trénovací fázi ponechány stranou a využity až ve fázi zhodnocení [42].

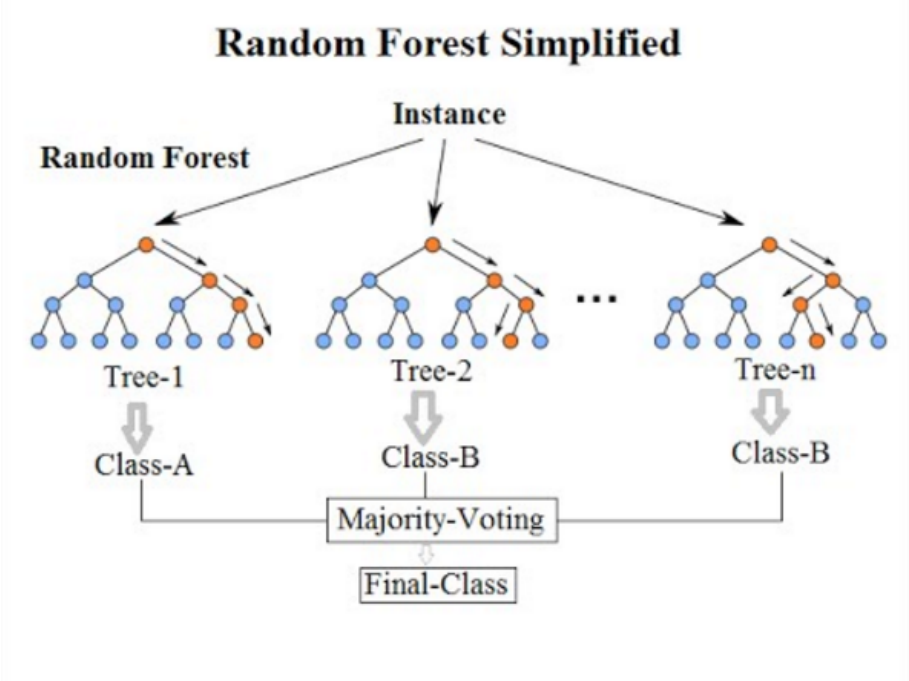
Zpravidla se větší část dat používá pro trénování, obvyklé poměry rozdělení na trénovací a testovací jsou např. 60:40 nebo 80:20. [42] K tomuto rozdělení byl použit nástroj *train_ test_ split*, pro testovací set bylo vyčleněno 20 % bodů a poměr zastoupení tříd v trénovacím a testovacím setu byl zachován podle původních dat.

7.2.1 Random Forest

Náhodný les (dále Random Forest, RF) je metoda strojového učení používaná pro vysokou přesnost, rychlost a nízké riziko přetrénování. RF se skládá z určitého počtu rozhodovacích stromů (obr. 7.6). Ze všech příznaků, které jsou k dispozici na vstupu, používá každý strom pouze náhodný a nezávislý výběr (subset) těchto veličin. Klasifikace následně probíhá nezávisle v každém stromu. Výsledná hodnota je modus, tedy nejčastěji zvolená hodnota napříč všemi stromy. Každý strom rozhoduje nezávisle, čímž se snižuje korelace mezi výsledky [48].

V porovnání s podobnými metodami (bagging, boosting) je výpočetně méně náročný, a tedy rychlejší [48], minimalizuje přetrénování [48], [50] a oproti samostatnému rozhodovacímu stromu je metoda RF robustnější [44].

V DPZ byl použit v mnoha úlohách a byl úspěšně aplikován i na klasifikace LC [50], [51]. RF poskytuje velmi dobré výsledky pro klasifikace, které používají data z různých zdrojů, tedy nejen optická, ale i jejich deriváty, výšková data, příp. texturální informace [44], [50]. Umí si poradit i s problémy, které obsahují velké množství (tisíce) příznaků [48].



Obrázek 7.6: Zjednodušený RF a klasifikace dle hlasu většiny [49]

Výběr příznaků

Kvůli časové náročnosti výpočtu může být v některých úlohách přínosnější určit příznaky ještě před klasifikací. RF nejenže dokáže pracovat i s větším množstvím příznaků a v porovnání s podobnými metodami je srovnatelně rychlejší [48], ale významnost jednotlivých příznaků dokáže samostatně určit. Na základě významnosti (feature importance, FI) lze eliminovat méně přínosné příznaky nebo jim přiřadit odpovídající váhu a výsledek klasifikace dále zlepšit [50]. V procesu klasifikace k tomu byl použit atribut metody *RandomForestClassifier* s příznačným názvem *feature_importances_*.

Hyperparametry

Ladění hyperparametrů bývá jednou z finálních úprav klasifikace metodou RF. Jedná se o proces, při kterém se určují takové hodnoty parametrů, které zvýší přesnost. Běžně jsou laděny tyto parametry:

- *n_estimators* - počet rozhodovacích stromů (vyšší počet může zvýšit časovou náročnost), def = 100,
- *max_depth* - maximální výška stromu; s větší výškou se zvyšuje přesnost, ale od určité hodnoty může nastat přetrénování, def = None,

- *min_ samples_ split* - minimální počet bodů v uzlu (node) než se rozroste o další uzly, def = 2,
- *min_ samples_ leaf* - minimální počet bodů, které musí být v uzlu po rozdělení (split), def = 1,
- *max_ features* - počet příznaků, které model uvažuje při hledání nejlepšího splitu, def = auto.

Určení hyperparametrů lze rozdělit na 2 kroky. Prvním je náhodné hledání (ve scikit learn funkce *RandomizedSearchCV*), kdy jsou v rámci širšího rozmezí určeny přibližné ideální hodnoty. V druhém kroku probíhá přesnější hledání (funkce *GridSearchCV*) na základě hodnot určených v předchozím kroku [52].

Proces klasifikace proběhl opakovaně za účelem optimalizace výsledků. Pro srovnání byla klasifikace provedena nejprve bez úprav podkladových dat a s užitím monotemporálních dat (data ze začátku vegetační aktivity). Následně byla upravena trénovací data a do klasifikace byla zařazena data ze všech 5 scén. Byly zahrnuty různé skupiny příznaků shrnuté v tab. 7.2.

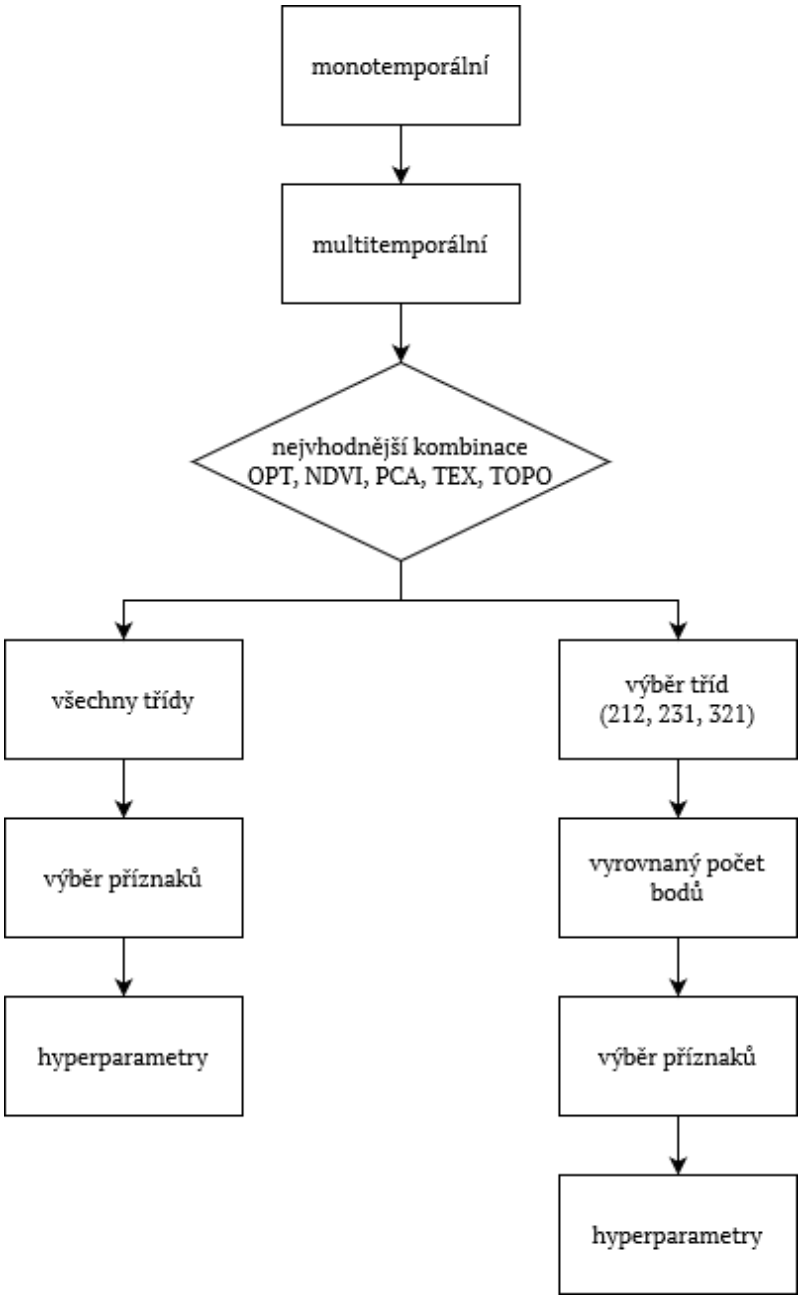
Jedná se o 9 optických pásem (OPT) a vegetační indexy jednotlivých scén (NDVI). Dále byla použita 3 pásma hlavních komponent (PCA), která v některých kombinacích nahradila optická data. Zařazena byla i texturální (TEX) a topografická data (TOPO). Do klasifikace byly tyto skupiny zaváděny postupně, tak aby bylo možné rozlišit vliv daných příznaků.

kategorie	jednotlivé příznaky	celkový počet příznaků (5 scén)
optická (OPT)	B2, B3, B4, B5, B6, B7, B8a, B11, B12	45
NDVI	NDVI	5
PCA	PCA1, PCA2, PCA3	15
texturální (TEX)	ASM, korelace, homogenita, entropie	20
topografická (TOPO)	DEM, sklon	2

Tabulka 7.2: Použité příznaky

Klasifikace byla nejprve provedena pro všechny třídy ve zvoleném klasifikačním schématu. Následně byly do klasifikace zařazeny pouze vybrané třídy, aby byla posouzena jejich klasifikovatelnost nezávisle na ostatních třídách (označení varianty - výběr). V tomto pří-

padě byl pro další klasifikaci počet bodů na vstupu upraven na stejný počet pro všechny tři třídy (označení - vyrov.). Pro obě varianty (všechny třídy i výběr tříd) byly určeny důležité příznaky z dané kombinace a hyperparametry. Popsaný postup shrnuje obr. 7.7.



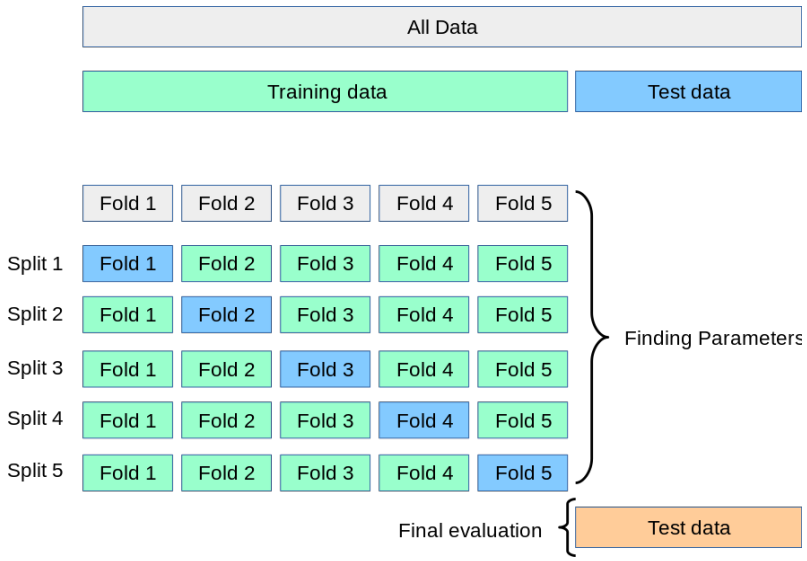
Obrázek 7.7: Pracovní postup klasifikace

7.3 Zhodnocení výsledků

Při hodnocení klasifikátoru se používají tzv. testovací data, která se liší od trénovacích. Jedná se o porovnání výsledku klasifikace se skutečností. Při tomto posouzení je nutné mít na paměti, že *„kvalita jakéhokoliv odhadu přesnosti je pouze tak dobrá, jak je dobrá informace o skutečném stavu“* [18].

7.3.1 Křížová validace

K vyhodnocení samotného modelu lze použít metodu křížové validace. Tato metoda umožňuje zhodnotit, jak dobře bude model pracovat na nezávislém datasetu (testovacím vzorku), a odhalit případné přetrénování [36]. Obr. 7.8 zobrazuje tzv. k-fold validaci, která byla použita (funkce *cross_ validate*). Při k-fold validaci jsou trénovací data rozdělena na k počet podmnožin (folds), jedna z nich je ponechána stranou a na zbylých podmnožinách proběhne trénink klasifikátoru. Výsledek je zhodnocen pomocí nepoužité části dat. Tento proces se opakuje podle stanoveného počtu. [36].



Obrázek 7.8: K-fold validace [36]

7.3.2 Chybová matice

Jeden ze způsobů vyhodnocení výsledků je tzv. chybová matice (obr. 7.9). S její pomocí lze zhodnotit výsledky jak jednotlivých tříd, tak celkové klasifikace. Na jedné straně jsou umístěna referenční data, tedy skutečné hodnoty pixelů, proti nim stojí výsledky klasifikace, tedy pixely zařazené pomocí klasifikátoru [40]. V příslušném směru lze zhodnotit

záměnu u referenčních a klasifikovaných bodů. Při porovnání skutečných a modelovaných hodnot lze provést podrobnější zhodnocení.

	D	C	BA	SB	row total
D	65	4	22	24	115
C	6	81	5	8	100
BA	0	11	85	19	115
SB	4	7	3	90	104
column total	75	103	115	141	434

Land Cover Categories
D = deciduous
C = conifer
BA = barren
SB = shrub

Obrázek 7.9: Příklad chybové matice: skutečnost (sloupce) vs. výsledky klasifikace (řádky) [53]

7.3.3 Precision, recall, F1

Základní charakteristikou výsledku klasifikace je celková přesnost - poměr mezi správně klasifikovanými pixely (na diagonále chybové matice) a celkovým počtem pixelů. Použití jediné hodnoty pro zhodnocení více tříd nemusí být v některých případech vhodné, zvlášť pokud jsou jednotlivé třídy zastoupeny jiným počtem pixelů [54]. Informaci o výsledku klasifikace jednotlivých tříd podávají metriky precision a recall (obr. 7.10).

Precision, jinak také spolehlivost, je poměr správně klasifikovaných pixelů dané třídy ku všem pixelům, které byly do této třídy klasifikovány. Značí, nakolik klasifikované pixely odpovídají skutečnému stavu. Lze ji určit na základě vztahu

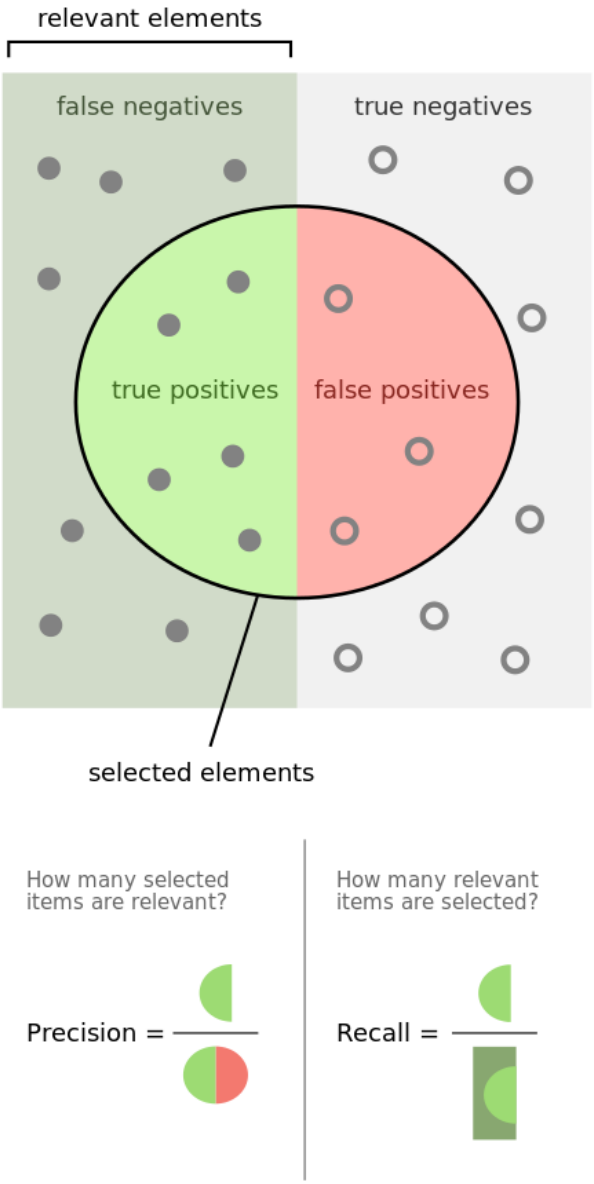
$$precision = \frac{TP}{TP + FP}, \tag{7.1}$$

kde TP je true positive (skutečně pozitivní) a FP je false positive (falešně pozitivní).

Recall, jinak úplnost nebo senzitivita, je poměr správně klasifikovaných pixelů dané třídy a celkového počtu pixelů, které do této třídy patří ve skutečnosti. Říká, nakolik jsou pixely z dané třídy rozpoznané při klasifikaci [40],[53] a popisuje jej vztah

$$recall = \frac{TP}{TP + FN}, \tag{7.2}$$

kde FN je false negative (falešně negativní).



Obrázek 7.10: Precision, recall [55]

Precision a recall jsou v literatuře označovány také jako uživatelská (producer's) a tvůrčí přesnost (user's accuracy) [40]. Tyto metriky se chovají jako spojené nádoby a v úlohách DPZ je žádoucí mezi nimi dosáhnout vyrovnaných hodnot. Soulad těchto metrik shrnuje ukazatel F1 (rov. 7.3). Jedná se o vážený harmonický průměr precision a recall. Říká, nakolik je model přesný, stejně jako nakolik je robustní. Nabývá hodnot 0 až 1, čím je hodnota vyšší, tím je model lepší [54].

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{7.3}$$

Kapitola 8

Výsledky

Postup klasifikace byl testován na lokalitě v Turecku, pro kterou budou výsledky rozebrány podrobněji. Následně budou popsány výsledky aplikace stejného postupu ve zbylých lokalitách.

8.1 Turecko

Výsledky klasifikace na základě monotemporální (březnové) scény zobrazuje tab. 8.1. Klasifikace proběhla bez úprav trénovacích dat a s výchozím nastavením klasifikátoru. Levá část tabulky je chybová matice, která zobrazuje počet bodů a jak byly body dané třídy klasifikovány. Pravá část výsledky shrnuje pomocí ukazatelů recall, precision a F1.

		klasifikace									
		100	211	212	231	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	166	0	72	44	96	1	379	43,80	61,25	51,08
	211	10	15	324	33	55	0	437	3,43	45,45	6,38
	212	51	14	4936	109	200	2	5312	92,92	83,89	88,17
	231	21	2	255	402	662	2	1344	29,91	46,64	36,45
	321	22	2	288	272	2254	0	2838	79,42	68,85	73,76
	500	1	0	9	2	7	50	69	72,46	90,91	80,65

Tabulka 8.1: Turecko: výsledky klasifikace - výchozí pozice (březen)

V tomto stadiu se záměna objevila v jisté míře mezi všemi třídami. Víc než polovina bodů zástavby (100) byla klasifikována jako jiné typy LC. Zdaleka nejlepší výsledky lze pozorovat u zavlažované orné půdy (212), u které bylo správně rozpoznáno téměř 93 %

bodů a z bodů klasifikovaných do této třídy bylo přes 83 % relevantních. Naproti tomu měly pastviny (231) pouze 30 % úspěšnost rozpoznaných bodů, z bodů klasifikovaných v této třídě bylo necelých 47 % správně. Při pohledu do levé části je možné posoudit, kde nastala záměna. V případě této třídy byla valná část bodů klasifikována jako přírodní travní porost (321), nicméně významná záměna se objevila i se zavlažovanou půdou. Nakonec, z přírodních travin bylo správně rozpoznáno přes 79 %, body zařazené do této třídy byly správně klasifikovány z téměř 69 %. Nejnižší hodnoty lze pozorovat u nezavlažované orné půdy (211). Vývoj klasifikace této třídy bude podrobněji rozebrán později.

Se zařazením multitemporálních dat (březen - červenec) byly použity různé kombinace příznaků (tab. 7.2). V této fázi bylo potřeba určit, která kombinace příznaků bude pro výsledky nejvhodnější. Vybrané kombinace a jejich výsledky zobrazuje tab. 8.2. Průběžné výsledky klasifikace byly hodnoceny na základě vztahu hodnot precision a recall, který jednočíselně charakterizuje hodnota F1. Tato hodnota byla použita i při určení vhodné kombinace příznaků pro dané třídy. Tabulka obsahuje jak průměry hodnot pro všechny uvažované třídy v dané lokalitě, tak průměry pouze pro 3 sledované třídy, aby v další části klasifikace byly použity relevantní kombinace.

kombinace	F1 ∅ (vše) [%]	F1 ∅ (212, 231, 321) [%]
OPT	74,68	79,18
OPT + NDVI	75,57	79,50
PCA + NDVI	78,15	79,61
PCA + NDVI + TEX	78,62	78,34
PCA + NDVI + TEX + TOPO	**80,20**	81,27
OPT + NDVI + TOPO	76,94	**82,15**

Tabulka 8.2: Turecko: srovnání hodnot F1 pro vybrané kombinace příznaků

Určení nejvhodnější kombinace nebylo v tomto případě jednoznačné. V kontextu všech klasifikovaných tříd byly nejlepší výsledky při použití kombinace PCA + NDVI + TEX + TOPO (tab. 8.3). Využití hlavních komponent zde převážilo nad původními optickými příznaky. Pro sledované třídy se nejlépe osvědčila kombinace OPT + NDVI + TOPO (tab. 8.4). Texturální míry pro dané třídy v této lokalitě nepatří mezi vhodné příznaky.

		klasifikace									
		100	211	212	231	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	50	0	0	0	0	0	50	100,00	100,00	100,00
	211	0	40	85	11	29	0	165	24,24	81,63	37,38
	212	0	1	1408	15	23	0	1447	97,30	91,43	94,28
	231	0	1	21	278	213	0	513	54,19	73,54	62,40
	321	0	7	26	74	1261	0	1368	92,18	82,63	87,15
	500	0	0	0	0	0	25	25	100,00	100,00	100,00

Tabulka 8.3: Turecko: výsledky klasifikace - kombinace PCA + NDVI + TEX + TOPO

		klasifikace									
		100	211	212	231	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	78	0	0	2	1	0	81	96,30	100,00	98,11
	211	0	21	148	23	24	0	216	9,72	70,00	17,07
	212	0	2	2309	33	28	0	2372	97,34	90,66	93,88
	231	0	0	35	439	273	0	747	58,77	72,92	65,09
	321	0	7	55	105	1722	0	1889	91,16	84,08	87,48
	500	0	0	0	0	0	25	25	100,00	100,00	100,00

Tabulka 8.4: Turecko: výsledky klasifikace - kombinace OPT + NDVI + TOPO

V obou případech přetrvala záměna mezi pastvinami a přírodními travinami, nicméně v porovnání s předchozími výsledky se ji na základě použitých příznaků a úpravě trénovacích dat podařilo zmenšit. Zástavba a vodní plochy byly klasifikovány téměř bezchybně, což se může odvíjet od jejich nepočetného zastoupení v podkladových datech.

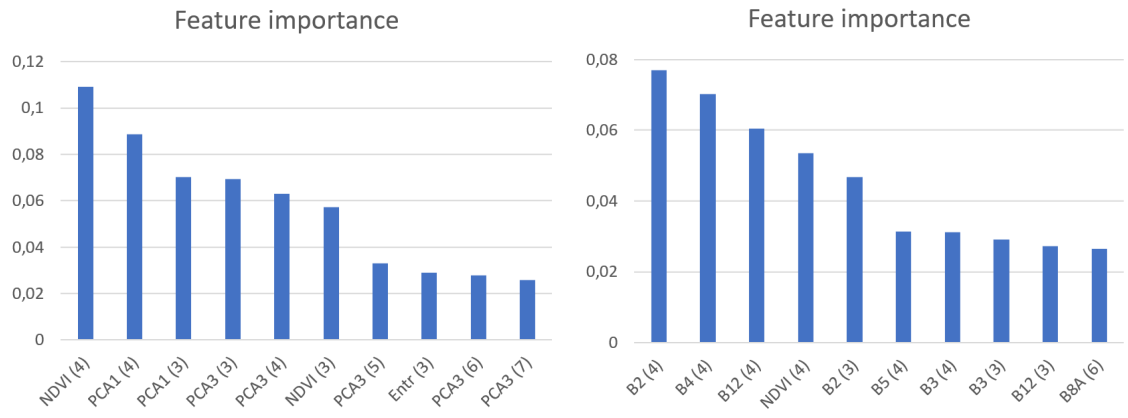
Následně byly v klasifikačním schématu ponechány pouze 3 sledované třídy. Na tuto strategii navázal další krok - byl upraven počet vstupních bodů tak, aby každá třída byla zastoupena stejným počtem bodů a podmínky klasifikace byly pro každou třídu vyrovnané. Na základě dřívějších poznatků byla pro klasifikaci použita kombinace příznaků OPT + NDVI + TOPO, která pro sledované třídy v dané lokalitě vykázala nejlepší výsledky. Výsledky tohoto kroku zobrazuje tab. 8.5.

		klasifikace						
		212	231	321	celkem	recall [%]	precision [%]	F1 [%]
reference	212	701	21	7	729	96,16	96,03	96,09
	231	12	644	74	730	88,22	80,50	84,18
	321	17	135	578	730	79,18	87,71	83,23

Tabulka 8.5: Turecko: výsledky klasifikace - vyrovnaný počet bodů (OPT + NDVI + TOPO)

V závěru byly určeny významné příznaky dané klasifikace a namísto výchozích parametrů klasifikátoru byly nalezeny hyperparametry.

Srovnání deseti nejdůležitějších příznaků ilustruje obr. 8.1. Příznaky, které začínají písmenem B, značí jednotlivá spektrální pásma Sentinel-2 (např. B2 - pásmo 2), čísla v závorce značí měsíc, ke kterému se daný příznak vztahuje (3 - březen, 4 - duben atd.).



(a) všechny třídy (PCA + NDVI + TEX + TOPO)

(b) vybrané třídy (OPT + NDVI + TOPO)

Obrázek 8.1: Turecko: srovnání důležitých příznaků

Co se týče všech tříd, nejvýznamnější se jeví příznak NDVI z dubnové scény, druhý nejdůležitější příznak je 1. pásmo PCA ze stejného období. Příznaky PCA jednoznačně převažují. Mezi 10 nejvýzamnějších se řadí i entropie pro období březen. Co se týče období, vyrovnané jsou příznaky z března a dubna, ale svůj význam mají i ostatní použité scény.

Na vybrané třídy mají největší vliv spektrální příznaky, jmenovitě B2 (490 nm), B4 (665 nm) a B12 (2190 nm). Významně přispívají příznaky z dubna a března.

S využitím významných příznaků a určenými hyperparametry došlo ještě k mírnému zlepšení (tab. 8.6). Z chybové matice je patrné, že úplnou záměnu mezi pastvinami a přírodními travinami se eliminovat nepodařilo. Nicméně, v porovnání s průběžnými výsledky lze vyrovnání vstupního počtu bodů hodnotit jako velmi úspěšný krok.

		klasifikace						
		212	231	321	celkem	recall [%]	precision [%]	F1 [%]
reference	212	702	20	7	729	96,30	95,77	96,03
	231	16	651	63	730	89,18	81,17	84,99
	321	15	131	584	730	80,00	89,30	84,39

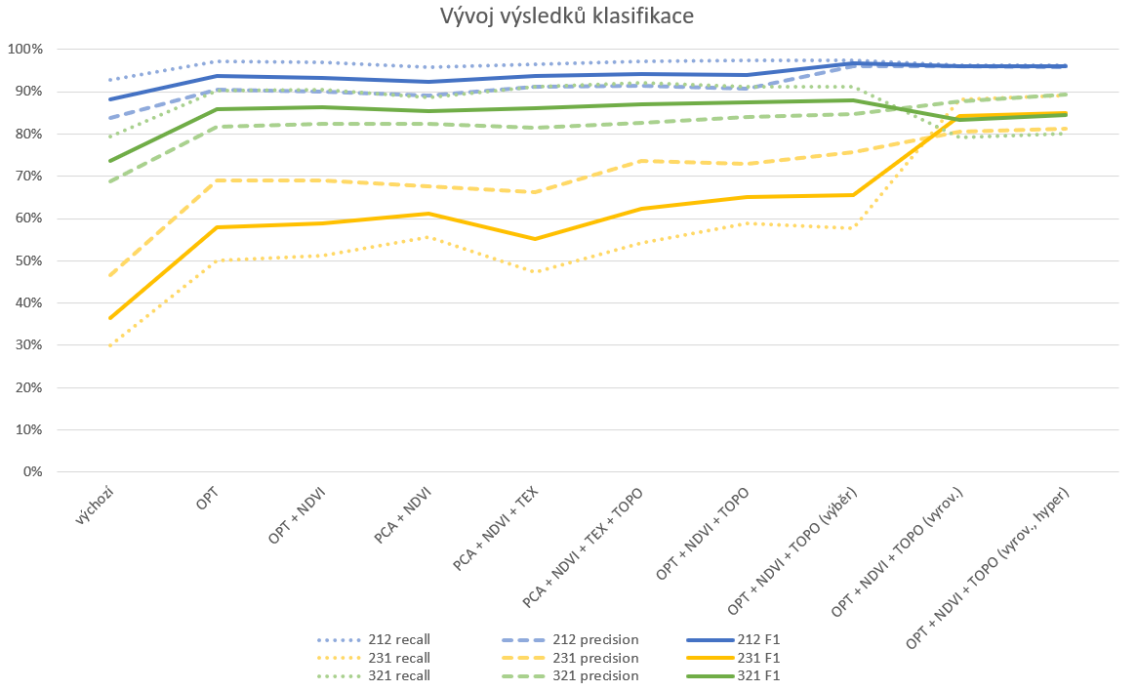
Tabulka 8.6: Turecko: výsledky klasifikace - vybrané třídy (OPT + NDVI + TOPO)

Záměnu mezi třídami této klasifikace zobrazuje tab. 8.7. Pravý sloupec v procentech vyjadřuje záměnu bodů, které byly z referenční třídy (sloupec *reference*) klasifikovány do třídy jiné (sloupec *klasifikace*). Je zjevné, že míra záměny mezi zavlažovanou ornou půdou a ostatními třídami byla maximálně necelá 3 %. Naproti tomu záměna mezi pastvinami a přírodními travinami se v konečné fázi pohybovala mezi 8 a téměř 18 %.

reference	klasifikace	záměna [%]
212	231	2,74
212	321	0,96
231	212	2,19
231	321	8,63
321	212	2,05
321	231	17,95

Tabulka 8.7: Turecko: procentuální záměna mezi vybranými třídami (OPT + NDVI + TOPO)

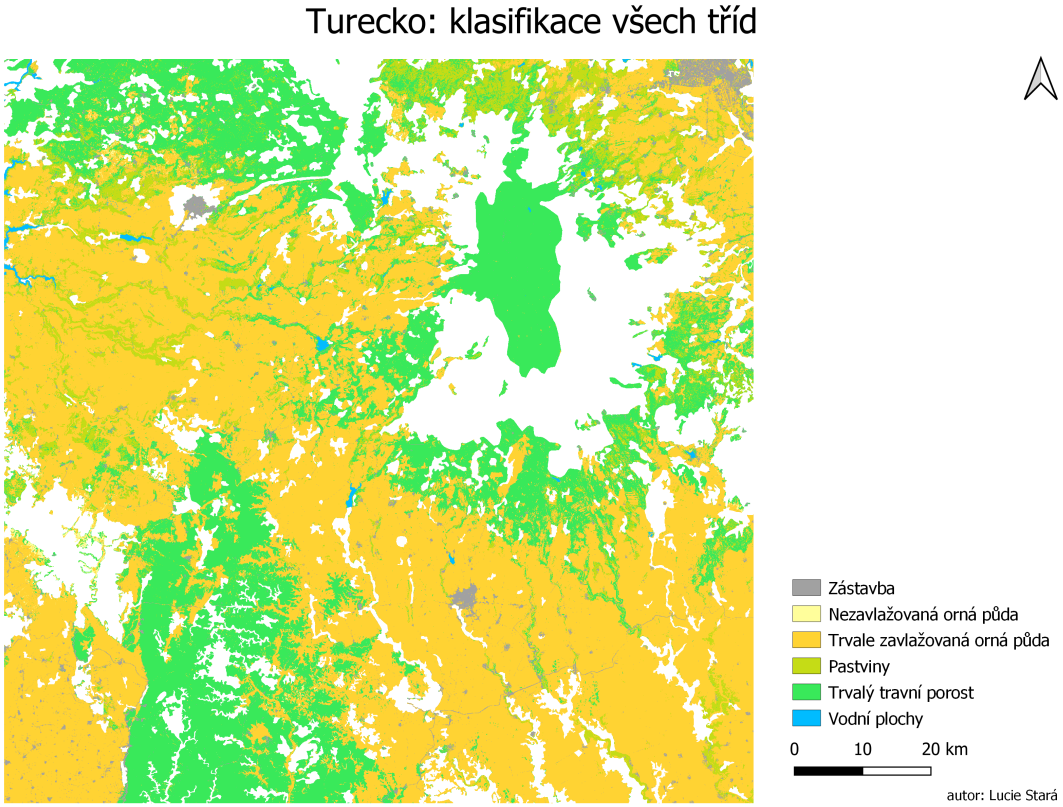
Průběžný vývoj výsledků sledovaných tříd se zahrnutím vybraných kroků zachycuje obr. 8.2. Na ose x jsou vyneseny použité kombinace počínaje první klasifikací s monotemporálními daty (výchozí), přes použití různých kombinací multitemporálních dat (OPT, OPT + NDVI atd.), konče výsledky tří klasifikací, které byly v závěru provedeny pouze pro vybrané třídy. Na ose y jsou vyneseny hodnoty v procentech. Ústřední zobrazenou veličinou je F1, kolem ní vytyčují pás recall a precision.



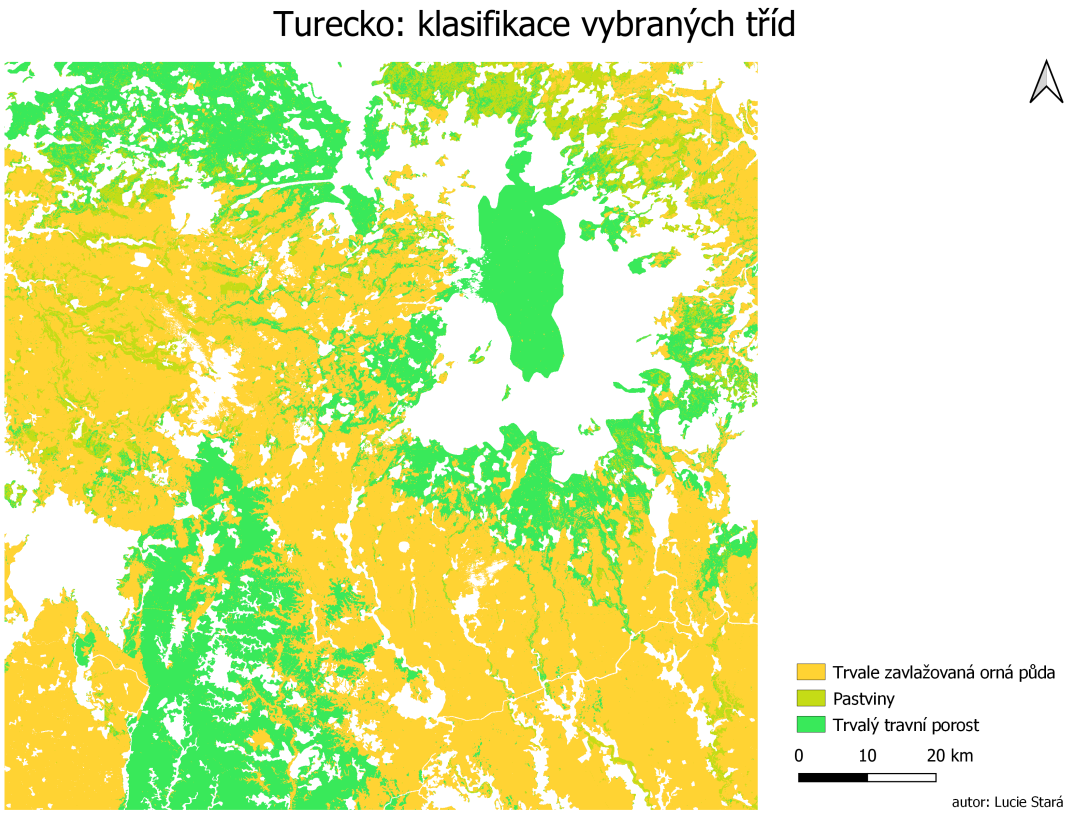
Obrázek 8.2: Turecko: vývoj výsledků klasifikace - vybrané třídy

Je zřejmé, že výsledky pro zavlažovanou ornou půdu i přírodní traviny v celém průběhu vykazovaly velmi dobré výsledky přes 80 %. Výsledky pro pastviny výrazně znázorňují význam jednotlivých kroků pro klasifikaci této třídy. Nejvýraznější nárůst lze pozorovat u přidání multitemporálních dat (OPT) a u vyrovnání počtu bodů (vyrov.). V momentě, kdy se počet bodů vyrovnal, výsledky této třídy se zlepšily o 20 %. V porovnání s výchozí pozicí se výsledky této třídy se zlepšily téměř o 50 %.

Výstupy klasifikace jsou zobrazeny na obr. 8.3.



(a) klasifikace všech tříd (PCA + NDVI + TEX + TOPO)



(b) klasifikace vybraných tříd (OPT + NDVI + TOPO)

Obrázek 8.3: Turecko: výstup klasifikace

8.2 Španělsko

Výsledky výchozí pozice zobrazuje tab. 8.8. S použitím monotemporálních dat a bez úpravy trénovacích ploch se hodnota F1 pohybuje mezi 30 a 84 %.

		klasifikace										
		100	211	212	231	310	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	222	129	65	18	12	30	7	483	45,96	73,75	56,63
	211	27	1595	243	218	85	488	5	2661	59,94	52,31	55,87
	212	27	434	701	21	24	75	12	1294	54,17	64,79	59,01
	231	7	353	25	377	65	622	6	1455	25,91	36,57	30,33
	310	3	98	6	29	903	135	5	1179	76,59	70,16	73,24
	321	13	419	29	364	188	1515	5	2533	59,81	52,70	56,03
	500	2	21	13	4	10	10	268	328	81,71	87,01	84,28

Tabulka 8.8: Španělsko: výsledky klasifikace - výchozí pozice

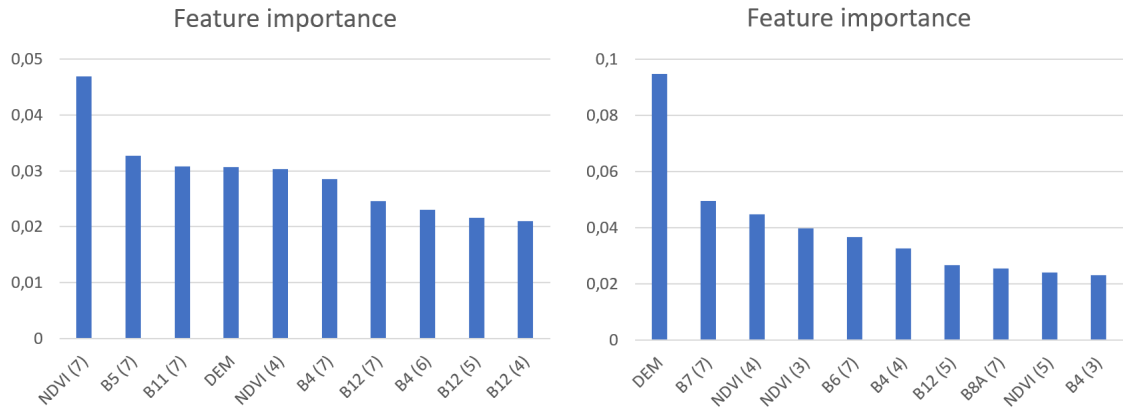
Do multitemporální klasifikace španělského území vstupovala taktéž data z období březen - červenec. Tabulka 8.9 zobrazuje srovnání vybraných kombinací příznaků na základě provedených klasifikací. Použití kanálů PCA se zde projevilo nižšími hodnotami (79,53 %) oproti kombinaci s optickými daty (81,83 %). Pro sledované třídy hodnoty při klasifikaci PCA + NDVI klesly vůbec nejníže (63,44 %). Naopak texturální míry přinesly mírné zlepšení. Další pozitivní změna se projevila přidáním topografických příznaků. Nejvyšších hodnot bylo dosaženo s kombinací OPT + NDVI + TEX + TOPO, která byla použita pro další zpracování.

kombinace	F1 ∅ (vše) [%]	F1 ∅ (212, 231, 321) [%]
OPT	79,31	64,43
OPT + NDVI	81,83	67,06
PCA + NDVI	79,53	63,44
OPT + NDVI + TEX	82,26	68,18
OPT + NDVI + TEX + TOPO	**83,70**	**70,61**
OPT + NDVI + TOPO	83,25	69,76

Tabulka 8.9: Španělsko: srovnání hodnot F1 pro vybrané kombinace příznaků

Pro vybranou kombinaci byly určeny důležité příznaky (obr. 8.4). Ke klasifikaci všech tříd nejvíce přispěly příznaky z července, konkrétně NDVI, B5 (705 nm) a B11 (1610 nm), následované DEM. Mezi další významné příznaky se zařadily B4 (665 nm) a B12 (2190 nm) z různých období. V porovnání s Tureckem mezi 10 nejvýznamnějšími nebyl ani jeden

příznak z březnové scény. Při klasifikaci vybraných tříd jednoznačně vynikl vliv DEM. Následující příznaky byly různorodější, opakovaly se pouze NDVI (březen, duben, květen) a B4 (březen, duben). Ani v jedné variantě se mezi 10 nejvýznamnějšími neobjevil žádný texturální příznak, jejich význam byl v této lokalitě menší.



(a) všechny třídy (OPT + NDVI + TEX + TOPO)

(b) vybrané třídy (OPT + NDVI + TEX + TOPO)

Obrázek 8.4: Španělsko: srovnání důležitých příznaků

Výsledky klasifikace všech tříd s hyperparametry zobrazuje tab. 8.10. S výjimkou zástavby a vodních ploch, jejichž výsledky byly 100 %, měly ostatní třídy své rezervy. Vynikla vzájemná záměna mezi nezavlažovanou ornou půdou, pastvinami a přírodními travinami, která se v Turecku neprojevila. Zavlažovaná orná půda byla zaměněna především s nezavlažovanou ornou půdou. Problematicky se projevily opět pastviny, ze kterých bylo rozpoznáno pouze 47 %. U této třídy se největší záměna objevila s přírodními travními porosty, následně nezavlažovanou ornou půdou. Lesy (310) byly klasifikovány s naprosto vyrovnanými hodnotami precision a recall (96,51 %), záměna s jinými třídami byla mizivá.

		klasifikace										
		100	211	212	231	310	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	90	0	0	0	0	0	0	90	100,00	100,00	100,00
	211	0	895	11	84	3	125	0	1118	80,05	76,89	78,44
	212	0	57	324	0	4	2	0	387	83,72	96,14	89,50
	231	0	111	0	276	3	196	0	586	47,10	63,01	53,91
	310	0	1	2	1	442	12	0	458	96,51	96,51	96,51
	321	0	100	0	77	6	778	0	961	80,96	69,90	75,02
	500	0	0	0	0	0	0	226	226	100,00	100,00	100,00

Tabulka 8.10: Španělsko: výsledky klasifikace - všechny třídy (OPT + NDVI + TEX + TOPO)

Výsledky klasifikace vybraných tříd s vyrovnaným počtem bodů a hyperparametry zobrazuje tab. 8.11. Nejlepší výsledky stejně jako v předchozí lokalitě vykázala zavlažovaná orná půda (precision 99 %, recall 98,25 %). Výrazné zlepšení výsledků bylo pozorováno u pastvin, kde hodnota F1 narostla o téměř 30 %. Ve výsledcích travních porostů došlo především k nárůstu hodnoty precision o více než 20 %.

		klasifikace						
		212	231	321	celkem	recall [%]	precision [%]	F1 [%]
reference	212	392	1	6	399	98,25	99,24	98,74
	231	0	333	66	399	83,46	82,02	82,73
	321	3	72	324	399	81,20	81,82	81,51

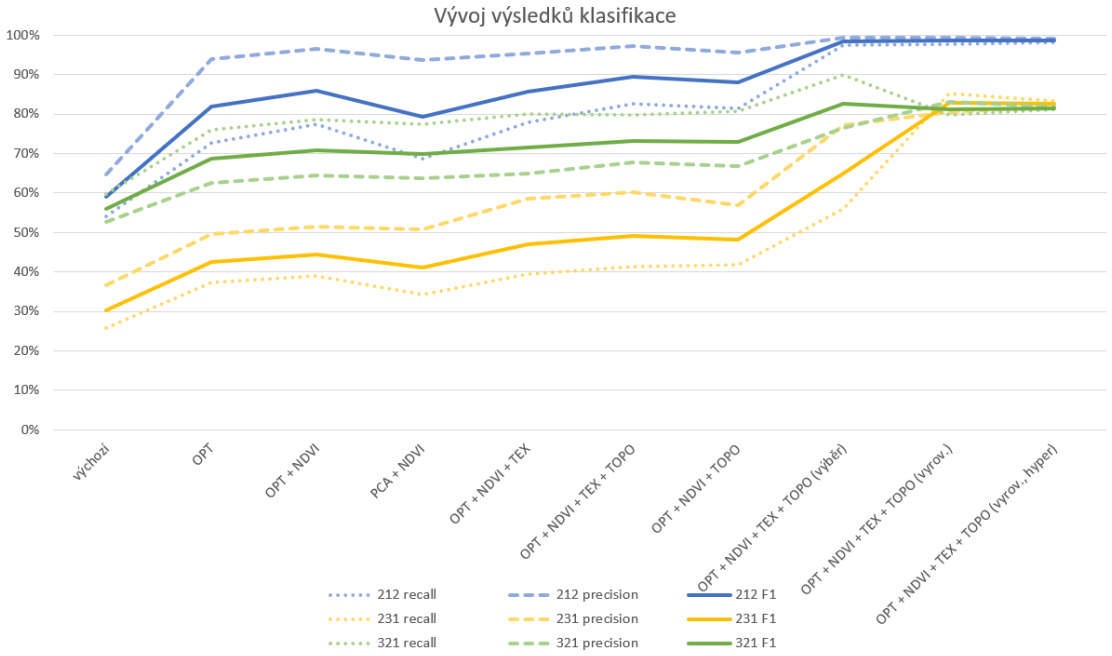
Tabulka 8.11: Španělsko: výsledky klasifikace - vybrané třídy (OPT + NDVI + TEX + TOPO)

Pro všechny sledované třídy se hodnoty recall a precision výrazně zvýšily a vyrovnaly, což je známkou dobře vytrénovaného modelu. Co se týče záměny mezi pastvinami a přírodními travinami, pohybovala se v tomto případě mezi 16 a 18 % (tab. 8.12). Záměnu se podařilo snížit především u pastvin. To může být jak v důsledku vyrovnání počtu bodů, tak odebráním třídy nezavlažovaná orná půda, za kterou byly body obou tříd často zaměňovány.

reference	klasifikace	záměna [%]
212	231	0,25
212	321	1,50
231	212	0,00
231	321	16,54
321	212	0,75
321	231	18,05

Tabulka 8.12: Španělsko: procentuální záměna mezi vybranými třídami (OPT + NDVI + TEX + TOPO)

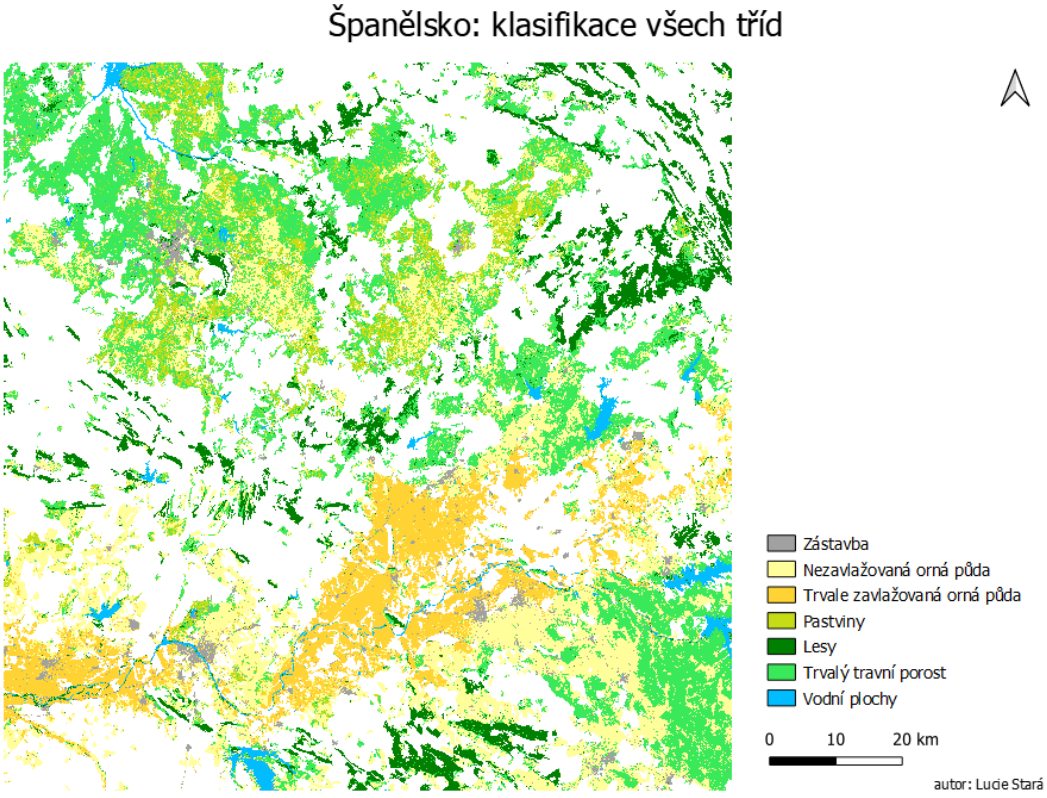
Vývoj výsledků zachycuje obr. 8.5. Výrazný nárůst je patrný především při zařazení multitemporálních dat, dále při klasifikaci pouze vybraných tříd (výběr) a při vyrovnání vstupního počtu bodů. Pás kolem F1 měl u všech tříd rozpětí větší než 10 %, až v posledních klasifikacích se zúžil na hodnotu kolem 1 %. Po určení hyperparametrů se rozdíl mezi hodnotami recall a precision zmenšil a vyrovnal.



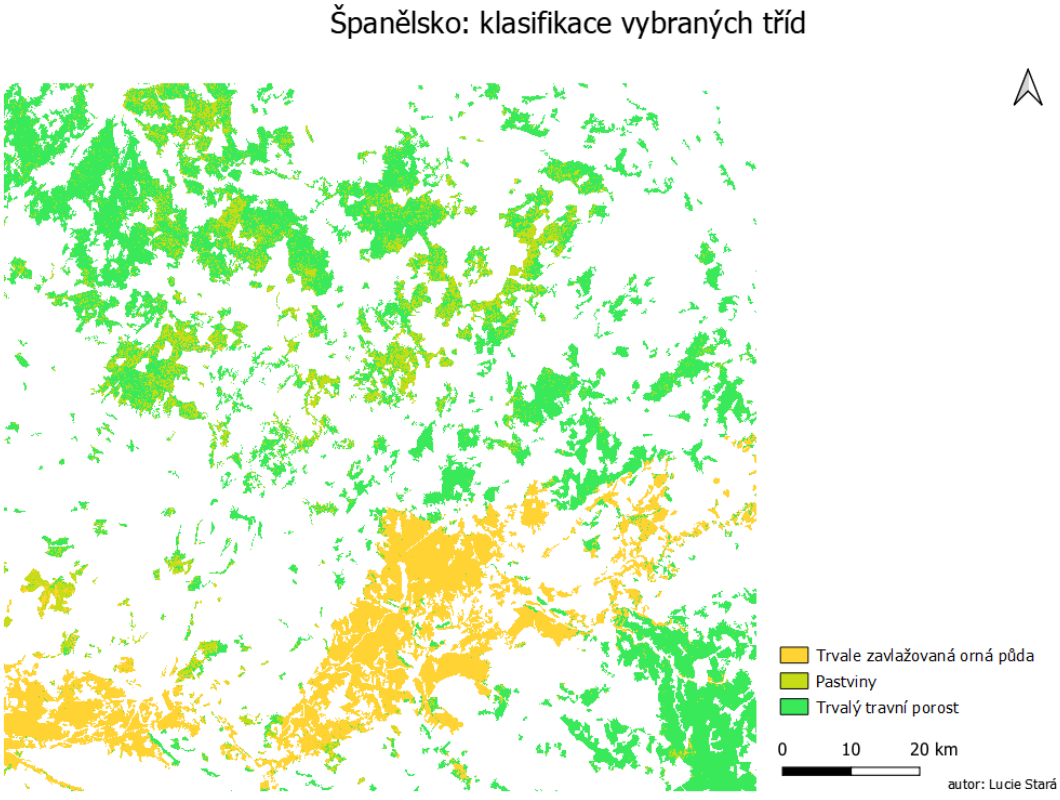
Obrázek 8.5: Španělsko: vývoj výsledků klasifikace - vybrané třídy

Pro zavlažovanou půdu a pastviny se objevil znatelný pokles při zařazení příznaků PCA. Jejich použití se pro tuto lokalitu neosvědčilo. Co se týče zavlažované orné půdy, po zařazení multitemporálních dat se hodnoty pohybovaly nad 80 %, což je dobrý výsledek, nicméně v závěru hodnoty vyrostly až k 99 %, což lze považovat za výborné. Při porovnání počátečních a koncových hodnot třídy pastviny, zlepšily se výsledky o více než 50 %, což je zdaleka největší nárůst. Oproti ostatním třídám se u této třídy projevilo výrazné zlepšení i po vyrovnání vstupního počtu bodů. To může být v důsledku toho, že tato třída byla v klasifikaci zastoupena nejmenším počtem bodů. Vývoj výsledků třídy přírodní traviny byl v porovnání s ostatními nejpozvolnější. Nicméně, i výsledky této třídy v závěru překonaly hranici 80 %.

Výstupy klasifikace zobrazuje obr. 8.6.



(a) klasifikace všech tříd (OPT + NDVI + TEX + TOPO)



(b) klasifikace vybraných tříd (OPT + NDVI + TEX + TOPO)

Obrázek 8.6: Španělsko: výstup klasifikace

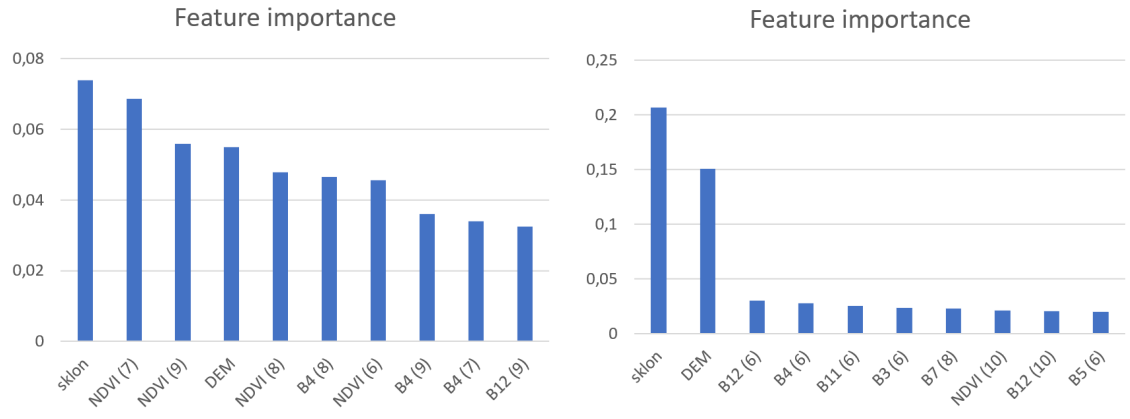
8.3 Makedonie

Vzhledem k odlišenému klimatu byla klasifikace této oblasti provedena na základě scén z období červen - říjen. Na základě srovnání (tab. 8.13) byly pro klasifikaci vybrány kombinace OPT + NDVI + TEX + TOPO (všechny třídy) a OPT + NDVI + TOPO (výběr). I v této lokalitě použití kanálů PCA znamenalo pokles výsledků oproti kombinaci s optickými daty (více než 1 % pro všechny třídy, necelá 3 % pro sledované třídy). Texturální příznaky měly pozitivní vliv na klasifikaci v kontextu všech tříd, pro vybrané třídy však byly výsledky s použitím textur nižší.

kombinace	F1 ∅ (vše) [%]	F1 ∅ (212, 231, 321) [%]
OPT	79,95	62,60
OPT + NDVI	80,93	64,37
PCA + NDVI	79,33	61,69
OPT + NDVI + TEX	79,77	61,50
OPT + NDVI + TEX + TOPO	**85,71**	71,19
OPT + NDVI + TOPO	85,45	**71,40**

Tabulka 8.13: Makedonie: přehled hodnot F1 pro vybrané kombinace příznaků

Co se týče významných příznaků (obr. 8.7), v obou klasifikacích měl největší roli sklon. V klasifikaci všech tříd patřily mezi významné příznaky NDVI (červen, červenec, srpen, září) a pásmo 4 (červenec, srpen, září). Pro klasifikaci vybraných tříd významně převyšují sklon a DEM, topografický aspekt měl v této oblasti největší vliv. Na dalších pozicích se nejčastěji objevily různé příznaky z června.



(a) všechny třídy (OPT + NDVI + TEX + TOPO)

(b) vybrané třídy (OPT + NDVI + TOPO)

Obrázek 8.7: Makedonie: srovnání důležitých příznaků

Opět byly provedeny klasifikace pro vybrané kombinace s hyperparametry. Oproti Španělsku byla chybová matice (tab. 8.14) čistější. Body ze zástavby a vodních ploch měly nulovou záměnu, následovány lesy, jejichž body byly klasifikovány téměř 100 %. I v tomto případě se projevila záměna mezi třídami nezavlažovaná orná půda, pastviny a přírodní traviny, nicméně v menší míře. Převažující byla opět záměna pouze mezi třídami pastviny a přírodní traviny. Nízkou úspěšnost lze pozorovat u zavlažované orné půdy, u které bylo 70 bodů ze 166 klasifikováno jako nezavlažovaná orná půda. V kontextu všech tříd zde patřily sledované třídy k těm hůře klasifikovaným.

		klasifikace										
		100	211	212	231	310	321	500	celkem	recall [%]	precision [%]	F1 [%]
reference	100	11	0	0	0	0	0	0	11	100,00	100,00	100,00
	211	0	548	8	38	0	7	0	601	91,18	82,53	86,64
	212	0	70	94	1	0	1	0	166	56,63	87,04	68,61
	231	0	38	6	602	0	191	0	837	71,92	67,26	69,52
	310	0	0	0	0	1785	0	0	1785	100,00	99,89	99,94
	321	0	8	0	254	2	728	0	992	73,39	78,53	75,87
	500	0	0	0	0	0	0	22	22	100,00	100,00	100,00

Tabulka 8.14: Makedonie: výsledky klasifikace - všechny třídy (OPT + NDVI + TEX + TOPO)

Odebráním tříd 100, 211, 310 a 500 se výsledné hodnoty změnily. Chybová matice v tab. 8.15 ukazuje výsledek klasifikace vybraných tříd pro vyrovnaný počet vstupních bodů a s použitím hyperparametrů. Hodnoty precision a recall se vyrovnaly pouze u zavlažované orné půdy. Hodnota F1 narostla téměř o 30 % a tradičně se výrazně přiblížila 100 %. Hodnoty zbylých tříd zůstaly pod hranicí 80 %. Ani v tomto případě záměna mezi třídami pastviny a přírodní traviny nebyla zcela odstraněna, v konečném stadiu se pohybovala mezi 17 a téměř 27 % (tab. 8.16).

		klasifikace						
		212	231	321	celkem	recall [%]	precision [%]	F1 [%]
reference	212	264	2	0	266	99,25	97,78	98,51
	231	5	215	46	266	80,83	74,65	77,62
	321	1	71	194	266	72,93	80,83	76,68

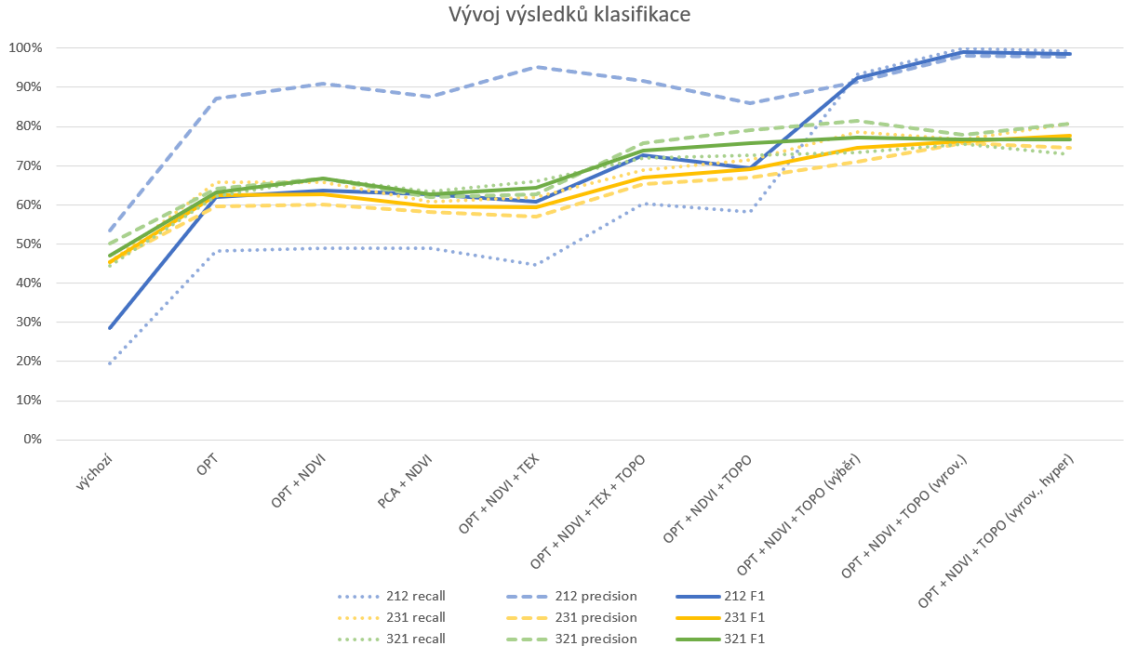
Tabulka 8.15: Makedonie: výsledky klasifikace - vybrané třídy (OPT + NDVI + TOPO)

Vývoj výsledků (obr. 8.8) byl v této oblastí pozvolnější, a to především pro pastviny

reference	klasifikace	záměna [%]
212	231	0,75
212	321	0,00
231	212	1,88
231	321	17,29
321	212	0,38
321	231	26,69

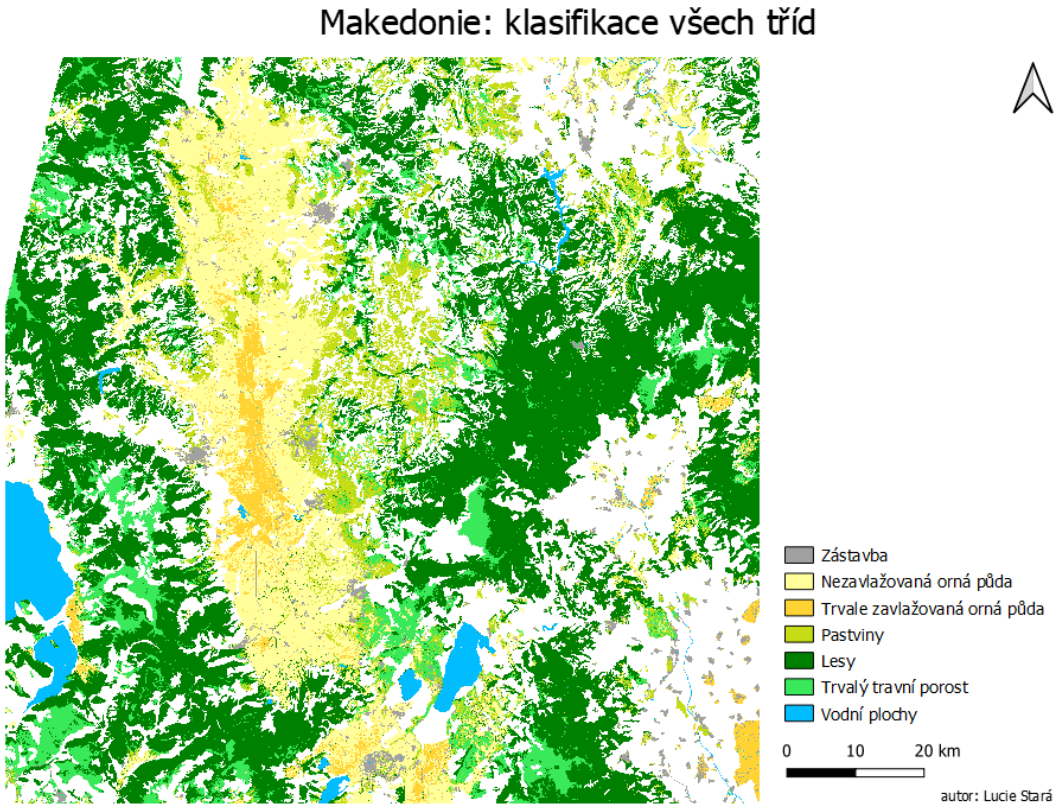
Tabulka 8.16: Španělsko: procentuální záměna mezi vybranými třídami (OPT + NDVI + TOPO)

a přírodní travní porosty. Výsledky klasifikace těchto tříd byly velmi podobné. To mohlo souviset i s faktem, že počet bodů těchto tříd byl oproti jiným lokalitám přibližně stejný. Shodně se pro obě třídy projevil i rozdíl mezi hodnotami precision a recall v průběhu celého klasifikačního procesu (do 10 %). Odfiltrování ostatních tříd se významně projevilo u třídy zavlažovaná orná půda, která v přechozích výsledcích vykazovala významnou záměnu s nezavlažovanou ornou půdou. Tento jev se výrazně projevil i na hodnotách precision a recall.

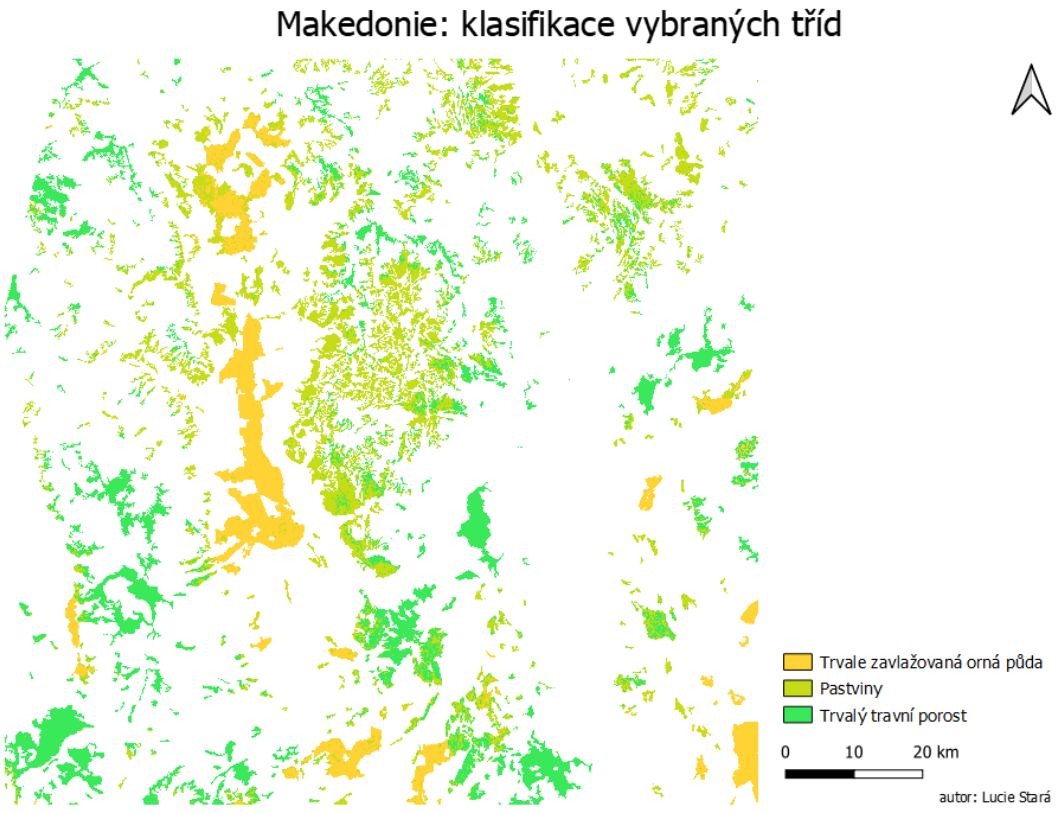


Obrázek 8.8: Makedonie: vývoj výsledků klasifikace - vybrané třídy

Výstupy klasifikace znázorňuje obr. 8.9.



(a) klasifikace všech tříd (OPT + NDVI + TEX + TOPO)



(b) klasifikace vybraných tříd (OPT + NDVI + TOPO)

Obrázek 8.9: Makedonie: výstup klasifikace

Kapitola 9

Diskuse

Obdobný postup klasifikace byl aplikován ve třech lokalitách s rozdílným podnebím i topografií. Odlišnost se projevila i ve výsledcích, ačkoli některé jevy lze pozorovat shodně.

Dosažené výsledky jednotlivých tříd budou zhodnoceny ve dvou rovinách. První jsou výsledky sledovaných tříd v kontextu ostatních tříd z klasifikačního schématu. V tomto případě se počet vstupních bodů odvíjí od velikosti původních trénovacích ploch CORINE a je pro každou třídu jiný. Druhá rovina je klasifikace pouze třech sledovaných tříd, ve které byl počet vzorků pro každou třídu vyrovnaný.

9.1 Referenční data

Co se týče referenčních dat, množství podkladových dat je jeden z aspektů, které se při tomto druhu analýzy ovlivnit nedá, tedy alespoň v rovině dostatečnosti. Na základě dosažených výsledků však lze soudit, že výsledky jsou závislé i na počtu nebo spíše na poměru mezi počtem bodů jednotlivých tříd. V první části analýzy počet vstupních bodů do klasifikace nebyl nijak regulován. Odvíjel se od početnosti zastoupení jednotlivých tříd v trénovacích datech, a byl pro každou třídu různý. Počet bodů sledovaných tříd byl v závěru klasifikace vyrovnán. Tento krok nejen že přinesl významné zlepšení výsledků jednotlivých tříd, navíc se v závěru ve všech lokalitách shodně projevila rozdílná klasifikovatelnost zavlažované orné půdy od pastvin a přírodních travin.

Na základě poznatků lze soudit, že mezi poměrem bodů jednotlivých tříd a výsledky klasifikace nelze hledat přímou úměru. Jak se projevilo například v Makedonii, zavlažovaná orná půda byla zastoupena přibližně pětinou bodů oproti pastvinám a přírodním travinám,

avšak hodnota F1 byla ve výsledku přibližně stejná (obr. 8.8). Nicméně, počet bodů může sehrát svou roli.

9.2 Metoda klasifikace

S použitím metody Random Forest bylo dosaženo velmi dobrých výsledků. Výsledky mezi trénovacím a testovacím setem se pohybovaly na podobných hodnotách, což indikuje, že referenční data byla vybrána vhodně a neobjevilo se přetrénování. Klasifikátor navíc pomohl v určení důležitých příznaků jednotlivých oblastí. Určení hyperparametrů mělo rozdílný efekt v závislosti na lokalitě. Ve Španělsku byly hodnoty precision a recall mezi jednotlivými třídami vyrovnány. V Makedonii i Turecku použití hyperparametrů vedlo naopak k většímu rozdílu mezi precision a recall, nicméně v Turecku se hodnoty celkově mírně zlepšily (přibližně o 1 %).

9.3 Význam příznaků

Ze spektrálních příznaků se mezi nejdůležitějšími nejčastěji objevila pásma B4 (665 nm) a B12 (2190 nm). Významná byla i další pásma z viditelné části spektra, menší význam měla pásma z oblasti blízkého infračerveného spektra. Za významné příznaky byly opakovaně určeny kanály NDVI.

Jelikož sledované třídy vykazovaly spektrální podobnost, byly do klasifikace zařazeny kanály PCA, které by zvýraznily odlišnosti. Jejich přínos se ale osvědčil pouze v turecké lokalitě. V ostatních lokalitách byly výsledky s použitím PCA nižší než v kombinaci s optickými daty.

Ani texturální příznaky jednoznačně nepřispěly k lepší klasifikovatelnosti sledovaných tříd. U těchto příznaků by bylo možné se dále zaměřit na jejich vliv v závislosti na jejich prostorovém rozlišení (ačkoli to nemusí být určující [4], [44]) nebo zvážit vytvoření textur na základě jiných pásem, např. těch, která jsou pro klasifikace nejvýznamnější [4].

Jednoznačně přínosná byla pro klasifikaci sledovaných tříd ve všech lokalitách topografická data (obr. 8.1, 8.4, 8.7). Ve Španělsku byl DEM určen jako téměř dvakrát důležitější než příznak na druhém místě. V Makedonii byly jako několikanásobně významnější příznaky oproti ostatním označeny oba topografické příznaky - DEM a sklon. Pro sledované třídy může být význam v tom, že přírodní travní porosty se nacházejí na obtížně pří-

stupných místech, což je i důvod, proč nejsou nijak kultivovány. Obtížně přístupná místa mohou být například i velmi svažité oblasti, které jsou zachyceny právě pomocí topografických dat. Naproti tomu pastviny a orná půda se nacházejí ve spíše rovinatých oblastech, které jsou zemědělské činnosti přístupnější. Komplikovaná členitost terénu ve vybraných lokalitách může být důvodem, proč se zde tyto příznaky osvědčily.

9.4 Sledované třídy

V kontextu dalších tříd patří sledované třídy k těm hůře klasifikovaným, což potvrzuje předpoklad práce. Nižších hodnot klasifikace ve všech oblastech bylo dosaženo u pastvin. Nejlepší výsledek byl dosažen v Makedonii (precision 67,26 %, recall 71,92 %). U této třídy se objevila záměna s přírodními travními porosty, dále pak s nezavlažovanou ornou půdou. Všechny tyto třídy lze hodnotit jako spektrálně podobné, obzvlášť v letních měsících a období sucha. Pastviny a louky jakožto kultivované plochy sice můžou být sečené, hnojené a jinak obhospodařované, nicméně na základě pozorování použitých scén tyto rozdílnosti oproti přírodním travinám pozorovány nebyly.

Lepší přesnosti v kontextu dalších tříd bylo docíleno u přírodních travních porostů. Nejvyšších hodnot bylo dosaženo v Turecku (precision 84,23 %, recall 91,74 %). Vývoj výsledků této třídy v závislosti na různých kombinacích příznaků byl velmi pozvolný. Nejvýraznější vliv na klasifikaci této třídy měly topografické příznaky v Makedonii. Nejlépe klasifikovanou třídou byla trvale zavlažovaná orná půda (Turecko: precision 91,16 %, recall 96,96 %), pouze s výjimkou Makedonie, kde došlo k významné záměně s nezavlažovanou ornou půdou. Opačný případ se vyskytl v Turecku, kde většina nezavlažované orné půdy byla klasifikována jako zavlažovaná orná půda. V obou případech mohl být příčinou přístup k zemědělství v souvislosti s klimatickými podmínkami.

Po vyrovnání počtu bodů se ve výsledcích objevily podobnosti napříč lokalitami. Ve všech třech lokalitách se vzájemná klasifikovatelnost tříd projevila shodně. Orná půda byla klasifikována nejhůře v Turecku (precision 95,77 %, recall 96,30 %) a nejlépe ve Španělsku (precision 99,24 %, recall 98,25 %). Výrazně se tak vzdálila zbylým třídám. Za daných podmínek je její klasifikovatelnost velmi dobrá.

Záměna mezi pastvinami a přírodními travinami se projevila ve všech lokalitách. V porovnání s výchozími hodnotami se výsledky těchto tříd použitými metodami podařilo výrazně zlepšit. Nejlépe byly pastviny klasifikovány v Turecku (precision 81,17 %, recall

89,18 %). Ve stejné lokalitě bylo nejvyšších hodnot dosaženo i pro přírodní travní porosty (precision 89,30 %, recall 80,00 %).

Použité klasifikační schéma poukázalo také na záměnu, která vznikla mezi sledovanými třídami a nezavlažovanou ornou půdou (třída 2.1.1). Způsob záměny se ovšem liší v závislosti na lokalitě. Ve Španělsku se projevila výrazná záměna s pastvinami a přírodními travinami. Ve zbylých oblastech byla chybná klasifikace zjevná především se zavlažovanou ornou půdou. Příčinou by ve všech případech mohlo být klima nebo podmínky a přístup k zemědělství. Klíčem k jejich rozlišení by také mohlo být použití takových scén, kde je tato zemědělská činnost patrná, a kdy je možné tyto třídy rozlišit. Zkoumání vztahů těchto tříd může být námětem k další práci.

Kapitola 10

Závěr

Cílem práce bylo provedení klasifikace problematických tříd zavlažovaná orná půda, pastviny a přírodní travní porosty a zhodnotit možnosti jejich klasifikace na podkladu dat CORINE. Součástí práce byly dále výběr vhodných příznaků, návrh zpracování trénovacích ploch a jejich rozdělení pro klasifikaci včetně hodnocení výsledků.

Práce byla zpracována v souladu s myšlenkou projektu Geo-Harmonizer s použitím volně přístupných dat (Sentinel-2, CORINE ad.) a pomocí open source SW (QGIS).

Pro klasifikaci byly vybrány tři evropské lokality, kde se tyto třídy nacházejí. Největší zastoupení měly třídy v Turecku, dále ve Španělsku a také v Severní Makedonii.

Pro řešení se osvědčilo použití multitemporálních dat. Uplatněna byla kombinace optických dat, kanálů NDVI a topografických informací. V případě Španělska přispěla ke klasifikaci navíc i texturální data, ačkoli v ostatních lokalitách se jejich využití neprojevilo zlepšením výsledků.

Referenční data byla použita z celoevropské databáze CORINE. V klasifikaci se projevil generalizační limit spojený s MMU, kterou CORINE používá. Na tento problém byly aplikovány odpovídající úpravy.

Klasifikace byla provedena s použitím metody Random Forest. Na základě této metody byly zjištěny i významné příznaky. Význam jednotlivých příznaků byl v závislosti na dané lokalitě odlišný. Mezi nejvýznamnější byl ve Španělsku a Makedonii zařazen DEM, v Makedonii k tomu ještě sklon terénu. Ze spektrálních příznaků byly ve všech oblastech nejčastější optická pásma 4 (665 nm) a 12 (2190 nm). Mezi deseti nejvýznamnějšími příznaky byly zaznamenáni také zástupci kanálů NDVI. Spíše neaplikovatelné byly příznaky texturální a kanály PCA.

Navržený postup vykázal nejlepší výsledky klasifikace vybraných tříd při vyrovnaném počtu vstupních bodů. Provedení klasifikace v různých lokalitách umožnilo odhalit společné rysy sledovaných tříd. Nejlépe klasifikovatelná se za daných podmínek ukázala zavlažovaná orná půda, která dosáhla nejlepších hodnot ve Španělsku (precision 99,24 %, recall 98,25 %). Pastviny a trvalé travní porosty dosáhly sice nižších výsledků, ale hodnoty F1 těchto tříd se v závěru pohybovaly na přibližně stejných hodnotách. Nejlepších výsledků bylo dosaženo v Turecku - pastviny (precision 81,17 %, recall 89,18 %), přírodní travní porosty (precision 89,30 %, recall 80,00 %).

Ve všech lokalitách se projevila záměna mezi pastvinami a přírodními travinami a přes všechny aplikované kroky tato záměna přetrvala. V závislosti na lokalitě se pohybovala v rozpětí 8 až 27 %. Ačkoli se sledovanou záměnu mezi třídami nepodařilo zcela eliminovat, bylo pro sledované třídy dosaženo uspokojivých výsledků.

Na základě dosažených výsledků lze říci, že použití CORINE jako podkladových dat s adekvátními úpravami se osvědčilo. Pro tyto polygony byl navržen způsob výběru trénovacích a testovacích dat (bodů) s využitím přístupů strojového učení. Pomocí metody Random Forest byla provedena klasifikace i výběr vhodných příznaků, které přispěly ke klasifikovatelnosti sledovaných tříd. Na základě provedené práce lze vytčené cíle považovat za splněné.

Seznam obrázků

Seznam tabulek

Bibliografie

[1] Open Data Science. *Geo-harmonizer: EU-wide automated mapping system for harmonization of Open Data based on FOSS4G and Machine Learning.* [Online]. URL: https://opendatascience.eu/geoharmonizer-project/. [cit. 3.3.2021].

[2] FSv ČVUT - Aktuality. *Iniciační fond fakulty stavební.* [Online]. URL: https://web.fsv.cvut.cz/aktuality/490/. [cit. 3.3.2021].

[3] T. Bouček. *Testování způsobů klasifikace pokrytí území vybraných evropských oblastí.* Diplomová práce. 2020. [cit. 3.3.2021].

[4] Bethany Melville, Arko Lucieer a Jagannath Aryal. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. In: *International Journal of Applied Earth Observation and Geoinformation* 66 (2018). [Online], s. 46–55. ISSN: 0303-2434. DOI: https://doi.org/10.1016/j.jag.2017.11.006. [cit. 3.3.2021].

[5] Richard Lucas et al. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 62.3 (2007). [Online], s. 165–185. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2007.03.003. [cit. 3.3.2021].

[6] Isaac Kyere et al. Multi-Temporal Agricultural Land-Cover Mapping Using Single-Year and Multi-Year Models Based on Landsat Imagery and IACS Data. In: *Agronomy* 9.6 (2019). [Online]. ISSN: 2073-4395. DOI: 10.3390/agronomy9060309. [cit. 3.3.2021].

[7] W. S. McInnes, B. Smith a G. J. McDermid. Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.4 (2015). [Online], s. 1395–1403. DOI: 10.1109/JSTARS.2015.2416713. [cit. 3.3.2021].

[8] Ryan J. Fisher, Ben Sawa a Beatriz Prieto. A novel technique using LiDAR to identify native-dominated and tame-dominated grasslands in Canada. In: *Remote Sensing of Environment* 218 (2018). [Online], s. 201–206. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2018.10.003. [cit. 3.3.2021].

[9] European Enviromental Agency. *CORINE Land Cover — Copernicus Land Monitoring Service*. [Online]. URL: https://land.copernicus.eu/pan-european/corine-land-cover. [cit: 24.9.2020].

[10] *CLC 2018 Download*. [Online]. Copernicus. URL: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download. [cit. 24.9.2020].

[11] Copernicus. *Corine land cover classes*. [Online]. URL: https://land.copernicus.eu/Corinelandcoverclasses.eps.75dpi.png/image_view_fullscreen. [cit. 24.9.2020].

[12] European Enviromental Agency. *Updated CLC illustrated nomenclature guidelines*. [Online]. 2019, s. 42, 52, 75. URL: https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf. [cit. 3.3.2021].

[13] *Copernicus Open Access Hub*. [Online]. Copernicus. URL: https://scihub.copernicus.eu/. [cit. 24.9.2020].

[14] Copernicus. *Long Term Archive Access*. [Online]. URL: https://scihub.copernicus.eu/userguide/LongTermArchive. [cit. 5.3.2021].

[15] ESA. *Sentinel-2 User Handbook*. [Online]. 2015. URL: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook. [cit. 3.3.2021].

[16] Copernicus. *Launch of Sentinel-2B satellite*. [Online]. URL: https://land.copernicus.eu/user-corner/events/launch-of-sentinel-2b-satellite-for-copernicus. [cit. 3.3.2021].

[17] European Space Agency. *Sentinel-2 ESA Bulletin 161*. [Online]. 2015. URL: http://esamultimedia.esa.int/docs/EarthObservation/Sentinel-2_ESA_Bulletin161.pdf. [cit. 3.3.2021].

[18] Lena Halounová, Karel Pavelka a ČVUT v Praze. *Dálkový průzkum Země*. Praha: Vydavatelství ČVUT, 2008.

[19] ESA. *Sentinel-2 Technical Guide*. [Online]. URL: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/products-algorithms. [cit. 3.3.2021].

[20] European Space Agency. *Sentinel User Guide*. [Online]. URL: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a. [cit. 3.3.2021].

[21] Pettorelli Nathalie. *The Normalized Difference Vegetation Index*. [Online]. OUP Oxford, 2013. ISBN: 9780199693160. DOI: 10.1093/acprof:osobl/9780199693160.001.0001. [cit. 5.4.2021].

[22] SEOS. *Classification Algorithms and Methods*. [Online]. URL: https://seos-project.eu/classification/classification-c01-p05.html. [cit. 5.4.2021].

[23] Rutkay Atun, Kaan Kalkan a Önder Gürsoy. Determining The Forest Fire Risk with Sentinel 2 Images. In: 1 (2020). [Online], s. 22–26. URL: https://dergipark.org.tr/en/pub/turkgeo. [cit. 15.4.2021].

[24] H. Hashim, Z. Abd Latif a N. A. Adnan. URBAN VEGETATION CLASSIFICATION WITH NDVI THRESHOLD VALUE METHOD WITH VERY HIGH RESOLUTION (VHR) PLEIADES IMAGERY. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W16 (2019). [Online], s. 237–240. DOI: 10.5194/isprs-archives-XLII-4-W16-237-2019. [cit. 15.4.2021].

[25] statistiXL. *Principal Component Analysis*. [Online]. URL: https://www.statistixl.com/features/principal-components/. [cit. 15.4.2021].

[26] Robert M. Haralick, K. Shanmugam a Its'Hak Dinstein. Textural Features for Image Classification. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (1973). [Online], s. 610–621. DOI: 10.1109/TSMC.1973.4309314. [cit. 1.5.2021].

[27] *GRASS GIS 7.9.dev Reference Manual*. [Online]. GRASS Development Team. URL: https://grass.osgeo.org/grass79/manuals/index.html. [cit. 1.5.2021].

[28] Terminologická komise ČÚZK. *Terminologický slovník zeměměřictví a katastru nemovitostí*. [Online]. URL: https://www.vugtk.cz/slovnik/index.php. [cit. 4.3.2021].

[29] GIS Geography. *DEM, DSM DTM Differences*. [Online]. URL: https://gisgeography.com/dem-dsm-dtm-differences/. [cit. 4.3.2021].

[30] Yodin. *Surfaces represented by a Digital Surface Model and Digital Terrain Model*. CC BY-SA 4.0, via Wikimedia Commons. [Online]. [cit. 15.4.2021].

[31] *EU-DEM v1.1 Download*. [Online]. Copernicus. URL: https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=download. [cit. 31.12.2020].

[32] NASA. *ICESat Science Rational*. [Online]. URL: https://icesat.gsfc.nasa.gov/icesat/sciencemis.php. [cit. 4.3.2021].

[33] Copernicus Programme. *EU-DEM*. [Online]. URL: https://land.copernicus.eu/imagery-in-situ/eu-dem. [cit. 4.3.2021].

[34] *The Open Source Geospatial Foundation*. [Online]. OSGeo. URL: https://www.osgeo.org/. [cit. 5.4.2021].

[35] *The Scientific PYthon Development EnviRoment*. [Online]. SPYDER. URL: https://www.spyder-ide.org/. [cit. 13.4.2021].

[36] *User Guide*. [Online]. Scikit learn. URL: https://scikit-learn.org/stable/user_guide.html. [cit. 5.4.2021].

[37] ESA. *Sentinel-2 tiling grid*. [Online]. URL: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/data-products. [cit. 24.9.2020].

[38] A. John Arnfield. *Köppen climate classification*. [Online]. Encyclopedia Britannica. URL: https://www.britannica.com/science/Koppen-climate-classification. [cit. 24.9.2020].

[39] Susan Kathleen Langley, Heather M. Cheshire a Karen S. Humes. A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. In: *Journal of Arid Environments* 49.2 (2001). [Online], s. 401–411. ISSN: 0140-1963. DOI: https://doi.org/10.1006/jare.2000.0771. [cit. 26.4.2021].

[40] Siamak Khorram et al. *Remote Sensing*. 2012. vyd. [Online]. Boston, MA: Springer US. ISBN: 2191-8171. DOI: 10.1007/978-1-4614-3103-9. [cit. 13.4.2021].

[41] D. Lu a Q. Weng. A survey of image classification methods and techniques for improving classification performance. In: *International Journal of Remote Sensing* 28.5 (2007). [Online], s. 823–870. URL: https://doi.org/10.1080/01431160600746456. [cit. 26.4.2021].

[42] *A Complete Machine Learning Project Walk-Through in Python: Part One*. [Online]. Towards Data Science. URL: https://towardsdatascience.com/a-complete-machine-learning-walk-through-in-python-part-one-c62152f39420. [cit. 13.4.2021].

[43] Sebastian Raschka. *Boxplot*. CC BY-SA 4.0, via Wikimedia Commons. [Online]. 2016. [cit. 26.4.2021].

[44] Yuhao Jin et al. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. In: *International Journal of Remote Sensing* 39.23 (2018). [Online], s. 8703–8723. URL: https://doi.org/10.1080/01431161.2018.1490976. [cit. 26.4.2021].

[45] Przemysław Kupidura. The Comparison of Different Methods of Texture Analysis for Their Efficacy for Land Use Classification in Satellite Imagery. In: *Remote Sensing* 11.10 (2019). [Online]. ISSN: 2072-4292. DOI: 10.3390/rs11101233. [cit. 26.4.2021].

[46] Ethem Alpaydin. *Introduction to Machine Learning, Fourth Edition*. 2020. vyd. MIT Press. ISBN: 978-0262043793.

[47] *Machine Learning*. [Online]. W3Schools. URL: https://www.w3schools.com/python/python_ml_getting_started.asp. [cit. 26.4.2021].

[48] Leo Breiman. Random Forests. In: *Machine Learning* 45.1 (2001). [Online], s. 5–32. URL: https://doi.org/10.1023/A:1010933404324. [cit. 26.4.2021].

[49] Venkata Jagannath. *Random Forest*. CC BY-SA 4.0, via Wikimedia Commons. [Online]. [cit. 26.4.2021].

[50] Pall Oskar Gislason, Jon Atli Benediktsson a Johannes R. Sveinsson. Random Forests for land cover classification. In: *Pattern Recognition Letters* 27.4 (2006). [Online], s. 294–300. DOI: https://doi.org/10.1016/j.patrec.2005.08.011. [cit. 26.4.2021].

[51] Serhii Havryliuk et al. Using the Random Forest Classification for Land Cover Interpretation of Landsat Images in the Prykarpattya Region of Ukraine. In: [Online]. 2018. DOI: 10.1109/STC-CSIT.2018.8526646. [cit. 26.4.2021].

[52] *Hyperparameters of Random Forest Classifier*. [Online]. Geeks for Geeks. URL: https://www.geeksforgeeks.org/hyperparameters-of-random-forest-classifier/. [cit. 26.4.2021].

[53] A review of assessing the accuracy of classifications of remotely sensed data. In: *Remote Sensing of Environment* 37.1 (1991). [Online], s. 35–46. ISSN: 0034-4257. DOI: https://doi.org/10.1016/0034-4257(91)90048-B. [cit. 26.4.2021].

[54] *Metrics to Evaluate your Machine Learning Algorithm*. [Online]. Towards Data Science. URL: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234. [cit. 26.4.2021].

[55] Walber. *Precision and recall*. CC BY-SA 4.0, via Wikimedia Commons. [Online]. [cit. 26.4.2021].