

Česká zemědělská univerzita v Praze

Fakulta životního prostředí

Katedra prostorových věd



Diplomová Práce

Využití metod hlubokého učení pro klasifikaci
krajinného pokryvu z dat Sentinel-2

Vedoucí práce: Ing. Tomáš Klouček Ph.D.

Diplomant: Bc. Jakub Fér

2024

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Jakub Fér

Prostorové vědy v životním prostředí

Název práce

Využití metod hlubokého učení pro klasifikaci krajinného pokryvu z dat Sentinel-2

Název anglicky

Using deep learning methods and Sentinel-2 for land cover classification

Cíle práce

Cílem diplomové práce je klasifikovat krajinný pokryv pomocí vybraných algoritmů (komerčních – například v softwaru ArcGIS Pro, i volně dostupných) založených na metodách hlubokého učení (deep learning) a volně dostupných multispektrálních družicových dat s vysokým prostorovým rozlišením Sentinel-2.

Díličí cíle práce souvisí s těmito výzkumnými otázkami:

- (a) Lze využít pro klasifikaci krajinného pokryvu předtrénovaných modelů?
- (b) Je možné v běžných komerčních softwarech jednoduše vytvořit vlastní model hlubokého učení pro klasifikaci krajinného pokryvu?
- (c) Dosahují větší úspěšnosti klasifikace krajinného pokryvu komerční nebo volně dostupná řešení?

Metodika

Metodiku lze rámcově rozdělit na:

- (1) Předzpracování vstupních dat;
- (2) Použití volně dostupných předtrénovaných modelů hlubokého učení;
- (3) Tvorbu referenčních dat (trénovacích vs. validačních) pro hluboké učení;
- (4) Tvorbu vlastního klasifikačního modelu využívajícího hluboké učení;
- (5) Vyhodnocení přesnosti klasifikace;
- (6) Zodpovězení výše uvedených výzkumných otázek.

Doporučený rozsah práce

cca 40 normovaných stran

Klíčová slova

Předtrénované modely, ArcGIS Pro, Volně dostupná data, Družicová data, Multispektrální snímky

Doporučené zdroje informací

- ALEM, Abebaw; KUMAR, Shailender. Deep learning methods for land cover and land use classification in remote sensing: A review. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2020. p. 903-908.
- JO, Wonho; LIM, Yongho; PARK, Key-Ho. Deep learning based land cover classification using convolutional neural network-a case study of korea. Journal of the Korean geographical society, 2019, 54.1: 1-16.
- KARRA, Krishna, et al. Global land use/land cover with Sentinel 2 and deep learning. In: 2021 IEEE international geoscience and remote sensing symposium IGARSS. IEEE, 2021. p. 4704-4707.
- SAHA, Sudipan, et al. Unsupervised deep learning based change detection in Sentinel-2 images. In: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, 2019. p. 1-4.
- ZHANG, Xin, et al. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery?. Remote Sensing, 2020, 12.3: 417.

Předběžný termín obhajoby

2023/24 LS – FŽP

Vedoucí práce

Ing. Tomáš Klouček, Ph.D.

Garantující pracoviště

Katedra prostorových věd

Elektronicky schváleno dne 6. 3. 2024

prof. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 7. 3. 2024

prof. RNDr. Michael Komárek, Ph.D.

Děkan

V Praze dne 27. 03. 2024

Čestné prohlášení

Prohlašuji, že jsem diplomovou/závěrečnou práci na téma:

Využití metod hlubokého učení pro klasifikaci krajinného pokryvu z dat Sentinel – 2 vypracoval/a samostatně a citoval/a jsem všechny informační zdroje, které jsem v práci použil/a a které jsem rovněž uvedl/a na konci práce v seznamu použitých informačních zdrojů. Jsem si vědom/a, že na moji diplomovou/závěrečnou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů, především ustanovení § 35 odst. 3 tohoto zákona, tj. o užití tohoto díla. Jsem si vědom/a, že odevzdáním diplomové/závěrečné práce souhlasím s jejím zveřejněním podle zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů, ve znění pozdějších předpisů, a to i bez ohledu na výsledek její obhajoby.

Svým podpisem rovněž prohlašuji, že elektronická verze práce je totožná s verzí tištěnou a že s údaji uvedenými v práci bylo nakládáno v souvislosti s GDPR.

V Praze dne 28. 3. 2024

.....

Poděkování

Touto cestou bych chtěl poděkovat vedoucímu diplomové práce Ing. Tomáši Kloučkovi Ph.D. za vedení práce, cenné rady, trpělivost a podporu. Dále patří poděkování rodině, přátelům a kolegům z oboru, za psychickou podporu, bez které by bylo překonání překážek při tvorbě této práce mnohem těžší.

Abstrakt

V posledních letech dochází k výraznému a stále se zrychlujícímu rozvoji technologií umělé inteligence. Tento rozvoj se nevyhnul ani oblastem dálkového průzkumu Země (DPZ). V DPZ se neustále potýkáme s úkoly, které vyžadují klasifikaci krajinného pokryvu.

Tato práce se zaměřuje na využití technologií hlubokého učení, porovnání různých dostupných způsobů klasifikace krajinného pokryvu a vytvoření vlastního modelu klasifikace pomocí hlubokého učení.

V práci jsem pracoval s družicovými daty s vysokým rozlišením z družic Sentinel-2. Na tato data jsem následně aplikoval předučený model pro klasifikaci krajinného pokryvu dostupný v rámci Esri Living Atlas a také jsem na jejich základě vytvořil vlastní model. Následně jsem porovnal jejich výsledky a náročnost použití. Výchozími referenčními daty byl produkt Corine Land Cover (CLC), dostupný v rámci programu Evropské Unie Copernicus.

Předučený model poskytuje klasifikaci vycházející z dat CLC, ovšem pouze na druhé úrovni dělení jednotlivých tříd. Poskytuje tedy v případě mnou zvoleného území 13 tříd. Svůj model jsem trénoval na datech CLC třetí úrovně, a tedy poskytl stejnou skladbu 26 tříd krajinného pokryvu, které jsou v zájmovém území zastoupeny.

Předučený model nabízí uživatelsky přívětivé a jednoduché použití s konzistentními výsledky. Jeho nevýhodou je závislost na konkrétních datech a jejich struktuře, což může omezovat jeho využití ve specifických analýzách zaměřených na analýzu krajinného pokryvu.

Tvorba vlastního modelu je oproti tomu náročná jak samotnými kroky tvorby trénovacích dat a volby správné architektury modelu, tak i časově. Výhodou však je, že uživatel není omezen vstupními daty a model si tvoří “na míru” své práci.

Výsledkem závěrečné práce byly klasifikované rastry, u kterých jsem porovnával jejich přesnost. Pro toto porovnání, jsem provedl i klasifikaci algoritmem Support Vector Machine (SVM), který je výchozím algoritmem pixelové klasifikace v softwaru ArcGIS Pro. Celková přesnost předtrénovaného modelu byla 74 %. Přesnost mnou vytvořeného modelu byla 34 % pro 26 tříd a 46 % pro 12 tříd. Přesnost algoritmu SVM byla 21 % pro 26 tříd a 29 % pro 12 tříd.

Klíčová slova: Předtrénované modely, ArcGIS Pro, Volně dostupná data, Družicová data, Multispektrální snímky

Abstract

Recent years have seen significant and accelerating advances in machine learning, deep learning and artificial intelligence technologies. This development has not avoided the field of remote sensing (RS). In remote sensing, we are constantly faced with tasks that require the classification of landscape cover.

This focuses on the use of deep learning technologies, comparing different available methods for landscape cover classification and developing a custom classification model using deep learning.

In this thesis I have worked with high resolution satellite data from Sentinel-2 satellites. I then applied a pre-trained model for landscape cover classification available within Esri Living Atlas to this data and also created my own model based on the data. I then compared their results and the difficulty of application. The initial reference data was the Corine Land Cover (CLC) product, available under the European Union's Copernicus programme.

The pre-trained model provides a classification based on CLC data, but only at the second level of class subdivision. Thus, it provides 13 classes in the case of my chosen area. I have trained my model on the third level CLC data, and thus it provides the same composition of the 26 landscape cover classes that are represented in the area of interest.

The pre-trained model offers a user-friendly and simple to use model with consistent results. Its drawback is its dependence on specific data and their structure, which may limit its use in specific analyses focused on landscape cover analysis.

On the other hand, creating a custom model is both time consuming and demanding in terms of the actual steps of creating training data and choosing the right model architecture. The advantage, however, is that the user is not limited by the input data and can create a model "tailored" to their work.

The final work resulted in classified rasters, for which I compared their accuracy. For this comparison, I also performed classification with the Support Vector Machine (SVM) algorithm, which is the default pixel classification algorithm in ArcGIS Pro software. The overall accuracy of the pre-trained model was 74 %. The accuracy of the model I created was 34 % for 26 classes and 46 % for 12 classes. The accuracy of the SVM algorithm was 21 % for 26 classes and 29 % for 12 classes.

Keywords: Pre-trained models, ArcGIS Pro, Open data, Satellite data, Multispectral imagery

Obsah

1	Úvod.....	1
2	Cíle práce.....	3
3	Literární rešerše	4
3.1	Co je krajinný pokryv a problematika jeho zaznamenávání.....	4
3.2	Projekt CORINE Land Cover	5
3.3	Využití CORINE Land Cover	7
3.4	Družicové systémy používané pro klasifikaci krajinného pokryvu	8
3.4.1	Družice Landsat	8
3.4.2	Sentinel-2	10
3.5	Současný stav řešené problematiky	12
4	Charakteristika studijního území	16
5	Metodika	17
5.1	Vstupní data.....	17
5.2	Předzpracování dat.....	18
5.3	Předtrénovaný model.....	19
5.4	Tvorba vlastního modelu.....	20
5.5	Validace výstupů klasifikace	23
6	Výsledky	25
6.1	Předtrénovaný model.....	25
6.2	Vlastní model.....	26
7	Diskuse	31
7.1	Jednoduchost a uživatelská přívětivost.....	32
7.2	Porovnání výsledků.....	32
8	Závěr a přínos práce.....	34
9	Přehled literatury a použitých zdrojů	35
10	Seznam obrázků a tabulek	41
11	Přílohy	42

1 Úvod

Klasifikace krajinného pokryvu patří od počátků dálkového průzkumu Země mezi jedny z nejdůležitějších tematických oblastí. Původně se tato disciplína zaměřovala na vojenské potřeby, kdy bylo nutné zkoumat krajinu z perspektivy strategického vojenského plánování. Nicméně v dnešní době, kdy čelíme výzvám jako je změna klimatu, ochrana biodiverzity a udržitelný rozvoj, nabývá sledování a klasifikace krajinného pokryvu ještě většího významu. Informace o krajinném pokryvu a využití krajiny jsou základními prvky pro plánování budoucího rozvoje jak na lokální, tak na globální úrovni.

S rozvojem technologií dálkového průzkumu jsme přešli od leteckého snímkování, které bylo omezené na relativně malé oblasti a časově náročné, k technologiím družicového průzkumu. Družice nám již více než 50 let poskytují nepřetržité sledování zemského povrchu v globálním měřítku a mnohá z těchto dat jsou k dispozici zcela bezplatně (Zhu et al., 2019). Paralelně s tím se vyvíjely i metody analýzy těchto dat. S nástupem počítačových technologií se stále více setkáváme s umělou inteligencí, která nabízí velký potenciál pro analýzu dat v mnoha oblastech. V současnosti jsou nejčastěji využívané metody klasifikace krajinného pokryvu metody strojového učení, které poskytují vysokou přesnost klasifikace (Noi & Kappas, n.d.) a jsou široce dostupné, jejich efektivní využití však vyžaduje nemalé úsilí a znalosti na straně uživatele.

S rozvojem hlubokého učení (deep learning) se otevírají nové možnosti využití těchto technologií i v oblasti klasifikace krajinného pokryvu. I když v současnosti není jejich použití ještě běžné, mají potenciál v budoucnosti automatizovat analýzy krajinného pokryvu (Karra et al., 2021). Vzhledem k rostoucímu objemu dat o zemském povrchu je takový pokrok velmi žádoucí. Manuální analýzy takového množství dat jsou velmi časově náročné a vyžadují velké množství pracovníků. Umělá inteligence právě z tohoto množství dat dokáže velmi dobře těžit. Čím více dat bude o povrchu Země dostupných, tím více materiálu pro učení bude moci umělá inteligence využít, a její potenciál se tak bude stále zvětšovat (Vali et al., 2020). Může se tím uvolnit značné množství lidské síly pro další zpracovávání těchto dat, či jinou vědeckou činnost. Přestože dnes je vstup člověka stále nenahraditelný, snahy o automatizaci procesů, které jsou dnes rozšířené ve všech vědeckých disciplínách se propisují i do odvětví DPZ. Vzhledem k tomu, že výpočetní výkon běžnému uživateli dostupných počítačů stále roste, stávají se i technologie hlubokého učení stále dostupnějšími.

Tato diplomová práce je věnována zkoumání dostupných možností klasifikace krajinného pokryvu pomocí metod hlubokého učení, jejich uživatelské přívětivosti a porovnání přesnosti s doposud používanými volně dostupnými metodami. Tyto poznatky by mohli využít budoucí uživatelé těchto technologií při volbě svého pracovního postupu, a pro dosažení optimálních výsledků svých klasifikací.

2 Cíle práce

Hlavním cílem diplomové práce je klasifikovat krajinný pokryv pomocí vybraných algoritmů v softwaru ArcGIS Pro založených na metodách hlubokého učení (deep learning) a volně dostupných multispektrálních družicových dat s vysokým prostorovým rozlišením Sentinel-2.

Dílčí cíle práce souvisí s těmito výzkumnými otázkami:

- (a) Lze využít pro klasifikaci krajinného pokryvu předtrénovaných modelů?;
- (b) Je možné v běžných komerčních softwarech jednoduše vytvořit vlastní model hlubokého učení pro klasifikaci krajinného pokryvu?;
- (c) Dosahují větší úspěšnosti klasifikace krajinného pokryvu komerční nebo volně dostupná řešení?

Dílčími cíli práce bylo vybrat vhodný předučení model pro klasifikaci krajinného pokryvu na základě dat Sentinel-2, natrénovat vlastní model pro stejný účel a ověřit přesnost výsledných klasifikací. Pro tento účel porovnávám klasifikace provedené modely deep learning s výsledkem klasifikace provedené běžně užívaným klasifikátorem Support Vector Machine.

3 Literární rešerše

3.1 Co je krajinný pokryv a problematika jeho zaznamenávání

V první řadě si musíme stanovit, co je vlastně krajinný pokryv (angl. Land cover; LC). Krajinný pokryv je klíčovým popisem povrchu Země a prostorové informace o něm. Na krajinný pokryv navazující statistiky jsou nezbytné pro řadu rozhodnutí v oblasti řízení přírodních zdrojů na lokální, národní a mezinárodní úrovni. Krajinný pokryv poskytuje informace o funkčním vztahu mezi terénem, klimatem a půdou, a může tak být indikátorem řady faktorů ovlivňujících jejich různé změny (Townshend et al., 2012). Typicky je popisován jako soubor tříd charakterizujících většinové zastoupení prvků na zemském povrchu (Grekousis et al., 2015). Naopak Evropskou environmentální agenturou je krajinný pokryv definován jako biofyzikální pokryv povrchu Země. Tato definice umožňuje dělení do různých kategorií, například vegetace (les, travní porost), budovy, vodní plochy (*Land Cover — European Environment Agency*, n.d.).

Rovněž je třeba rozlišovat termíny krajinný pokryv (Land Cover; LC) a využití krajiny (Land use; LU). Tyto pojmy jsou často zaměňovány, avšak je v nich zásadní rozdíl. Land cover představuje samotný popis povrchu Země na sledované ploše. Land use, naproti tomu, popisuje využití daného povrchu. Například zástavba by byla kategorie krajinného pokryvu, pokud bychom se zabývali i jejím využitím, pak bychom mohli rozlišovat zástavbu obytnou, komerční, rekreační apod. Stejně tak trvalý travní porost bychom mohli rozlišovat na louky a pastviny podle způsobu jejich využití člověkem (*The Difference between “Land Use” and “Land Cover” - MSU Extension*, n.d.). Většina systémů pro klasifikaci krajinného pokryvu v sobě zahrnuje zároveň i klasifikace využití těchto pokryvů (Grekousis et al., 2015). Tyto dvě informace nám společně umožňují udělat si obrázek o sledovaném území a dále například plánovat jeho rozvoj, či sledovat jeho vývoj v čase a chránit tak například zranitelné ekosystémy apod. (*The Difference between “Land Use” and “Land Cover” - MSU Extension*, n.d.)

Dále si je potřeba uvědomit rozdíl mezi geografickými daty a geografickými informacemi, neboť to nám umožní lépe chápat problematiku získávání a kategorizování údajů o krajinném pokryvu. O geografických datech a geografických informacích přemýšlíme jako o dvou odlišných a samostatných konceptech. Data chápeme jako výsledek měření nějakého předem definovaného jevu, zatímco informace je výsledkem interpretace, kategorizace, klasifikace nebo jiného typu zpracování (Comber et al., 2005). Geografická data mohou být například představována měřením přítomnosti vegetace, budov, vody atd., ovšem sama o sobě uživateli neposkytnou příliš informací. Naproti tomu geografické informace vyžadují vnímání dat v kontextu situace a jsou ovlivněny měřítkem snímaných dat (Woodcock

& Strahler, 1987). V zásadě je klasifikace krajinného pokryvu do jisté míry subjektivní, různé organizace, státy, úřady, či lidé mohou stejnou plochu klasifikovat jako jinou třídu. Například pouhá přítomnost stromů nemusí nutně pro každého znamenat, že se jedná o les (Comber et al., 2005).

Pro účely této práce bude dále rozebíráno určování krajinného pokryvu z optických družicových senzorů. Prostorové rozlišení takového senzoru je definováno velikostí pixelu. Plocha, kterou na povrchu Země tento pixel představuje většinou obsahuje různé druhy krajinného pokryvu (vegetaci, budovy, holou půdu atp.) a data získaná z tohoto pixelu jsou jejich průměrem (Fisher, 1997). Informace, které lze o krajinném pokryvu ze snímku získat a interpretovat jsou tak ovlivněny minimální mapovou jednotkou (Woodcock & Strahler, 1987)prostorovým rozlišením snímku (Fisher, 1997), zorným polem senzoru, tedy plochou zemského povrchu, kterou jeho snímek zobrazuje (Forshaw et al., 1983), a částmi elektromagnetického spektra, které je schopen měřit (Comber et al., 2005).

3.2 Projekt CORINE Land Cover

Projekt CORINE (Coordination of Information on the Environment) Land Cover vznikl na počátku 80. let 20. století jako iniciativa Evropské komise, která měla za účel vytvořit jednotnou, standardizovanou metodologii pro vytváření map krajinného pokryvu, biotopů a kvality ovzduší v celoevropském měřítku (Moss & Wyatt, 1994).

Projekt rozlišuje celkem 44 tříd krajinného pokryvu a využití krajiny, které jsou děleny do 3 úrovní podrobnosti. První datová sada CORINE Land Cover byla vydána v roce 1998, jako CLC1990, následovala datová sada pro rok 2000 a další jsou již vydávány v šestiletých intervalech. Minimální mapovaná jednotka (MMU) představuje 25 ha, to znamená, že objekty s plochou menší než 25 ha nemohou být zobrazeny. Minimální mapovaná šířka (MMW) pak představuje 100 m (Ořahel, 2000). Přestože se tyto základní údaje v průběhu historie nezměnily, proces tvorby samotných dat, se změnil výrazně. Například produkt CLC1990 vznikl v letech 1986–1998 a vycházel ze snímků družice Landsat 5 se senzory MultiSpectral Scanner a Thematic Mapper – prostorové rozlišení těchto snímků bylo 57x79 m, z čehož následně vycházelo určení MMU a MMW. Následující produkty již vycházely postupně ze snímků družic Landsat 7, SPOT, IRS LISS III a RapidEye. Rovněž se zvyšovala geometrická přesnost těchto produktů a zkracoval se čas potřebný k jejich produkci (Bütner, 2014).

Názvosloví CORINE Land Cover je děleno do pěti hlavních kategorií. Tyto základní kategorie představují:

1. Umělé povrchy (Artificial surfaces)

2. Zemědělské plochy (Agricultural areas)
3. Lesy a polo-přírodní plochy (Forests and semi-natural areas)
4. Mokřady (Wetlands)
5. Vodní tělesa (Water bodies)

Tyto základní kategorie jsou následně děleny do dalších dvou úrovní podrobnosti (Tab. 1) na základě jejich využití a dalších charakteristik (O’ahel’, 2000) Kategorizace probíhá na základě automatické analýzy družicových snímků, manuální interpretace leteckých snímků a sběru dat přímo v terénu. Společně s produkty mapování krajinného pokryvu, jsou vytvářeny i produkty popisující jejich změnu v čase (Bütner, 2014).

Level 1	Level 2	Level 3	
1. Artificial surfaces	1.1. Urban fabric	1.1.1. Continuous urban fabric	
		1.1.2. Discontinuous urban fabric	
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units	
		1.2.2. Road and rail networks and associated land	
		1.2.3. Port areas	
		1.2.4. Airports	
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites	
		1.3.2. Dump sites	
		1.3.3. Construction sites	
	1.4. Artificial, non-agricultural vegetated areas	1.4.1. Green urban areas	
		1.4.2. Sport and leisure facilities	
	2. Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land
			2.1.2. Permanently irrigated land
			2.1.3. Rice fields
		2.2. Permanent crops	2.2.1. Vineyards
			2.2.2. Fruit trees and berry plantations
2.2.3. Olive groves			
2.3. Pastures		2.3.1. Pastures	
2.4. Heterogeneous agricultural areas		2.4.1. Annual crops associated with permanent crops	
		2.4.2. Complex cultivation patterns	
		2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation	
		2.4.4. Agro-forestry areas	
		3.1. Forests	3.1.1. Broad-leaved forest
	3.1.2. Coniferous forest		
	3.1.3. Mixed forest		
3.2. Scrub and/or herbaceous associations	3.2.1. Natural grassland		
	3.2.2. Moors and heathland		
	3.2.3. Sclerophyllous vegetation		
	3.2.4. Transitional woodland-scrub		
3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes, sands		
	3.3.2. Bare rocks		
	3.3.3. Sparsely vegetated areas		
	3.3.4. Burnt areas		
	3.3.5. Glaciers and perpetual snow		
4. Wetlands	4.1. Inland wetlands	4.1.1. Inland marshes	
		4.1.2. Peat bogs	
	4.2. Marine wetlands	4.2.1. Salt marshes	
		4.2.2. Salines	
		4.2.3. Intertidal flats	
		4.2.4. Coastal lagoons	
5. Water bodies	5.1. Inland waters	5.1.1. Water courses	
		5.1.2. Water bodies	
	5.2. Marine waters	5.2.1. Coastal lagoons	
		5.2.2. Estuaries	
		5.2.3. Sea and ocean	

Tabulka 1 - Dělení tříd Land Cover/Land Use dle CORINE Land Cover (Bütner, 2014).

3.3 Využití CORINE Land Cover

Změny ve využití krajiny jsou významnou součástí výzkumu našeho prostředí a jejich optimalizace je klíčová pro ochranu životního prostředí, udržitelný rozvoj a řízení přírodních zdrojů. Sledování těchto změn poskytuje důležité informace pro rozhodování na politické, ekonomické a environmentální úrovni a umožňuje tak efektivní rozvoj (Lambin et al., 2001) Produkty CORINE Land Cover Change (CLCC) poskytují informace o těchto změnách v šestiletých obdobích mezi publikacemi jednotlivých produktů CLC (*CORINE Land Cover — Copernicus Land Monitoring Service*, n.d.-a). Tyto produkty nacházejí využití v různých odvětvích, například v urbanismu, či při studiích fragmentace krajiny.

Fragmentace krajiny je jev, který, když je analyzován v různých časových rámcích, může ukázat dynamiku změn ve využití půdy. Z hlediska krajiny, což je nejčastější kontext analýzy, je zvyšování fragmentace obvykle žádoucí, protože naznačuje rozmanitost a zvýšení biodiverzity. Nicméně, pokud se týká pouze oblastí s velkým lidským osídlením, pak z hlediska přírody obvykle znamená snížení odolnosti přírodního prostředí vůči lidské činnosti. Zároveň může být vypovídající o ekonomické situaci a rozvoji ve sledované oblasti. Nízký index fragmentace naznačuje nízký rozvoj, a tedy i nízký ekonomický růst. Tento jev je možné vhodně analyzovat právě z produktů CLCC (Cieślak et al., 2017).

Sledování změn krajinného pokryvu nachází také využití v urbanismu. Sledování rychlosti růstu měst je důležitým ekonomickým ukazatelem. Rovněž je přínosné sledovat, které druhy pokryvu ustupují umělým povrchům, neboť tím získáváme informace o dopadech činnosti člověka na životní prostředí (Yilmaz & Yilmaz, 2010).

3.4 Družicové systémy používané pro klasifikaci krajinného pokryvu

3.4.1 Družice Landsat

Inspirací programu Landsat byly mise Apollo. Během těchto misí byly prováděny experimenty s fotografováním zemského povrchu. První návrh na vypuštění civilní družice snímající zemský povrch se objevil v roce 1965. Tehdy se setkal s nevolí úřadů a ministerstva obrany USA, neboť úřady bylo považováno za ekonomicky výhodnější použít jako nosič senzorů pro sledování povrchu letadla. Ministerstvu obrany se pak nelíbila možnost prozrazení špionážních programů (*History | Landsat Science*, n.d.). Program Landsat představuje nejdéle běžící kontinuální zaznamenávání zemského povrchu. První družice byla vypuštěna v roce 1972, poslední v řadě pak v roce 2021 (Obr. 1). Družice Landsat dodnes poskytují záznamy o zemském povrchu a stále tak rozšiřují archiv snímků sahající 54 let do historie. Každá z družic má časové rozlišení 16 dní (*About | Landsat Science*, n.d.).

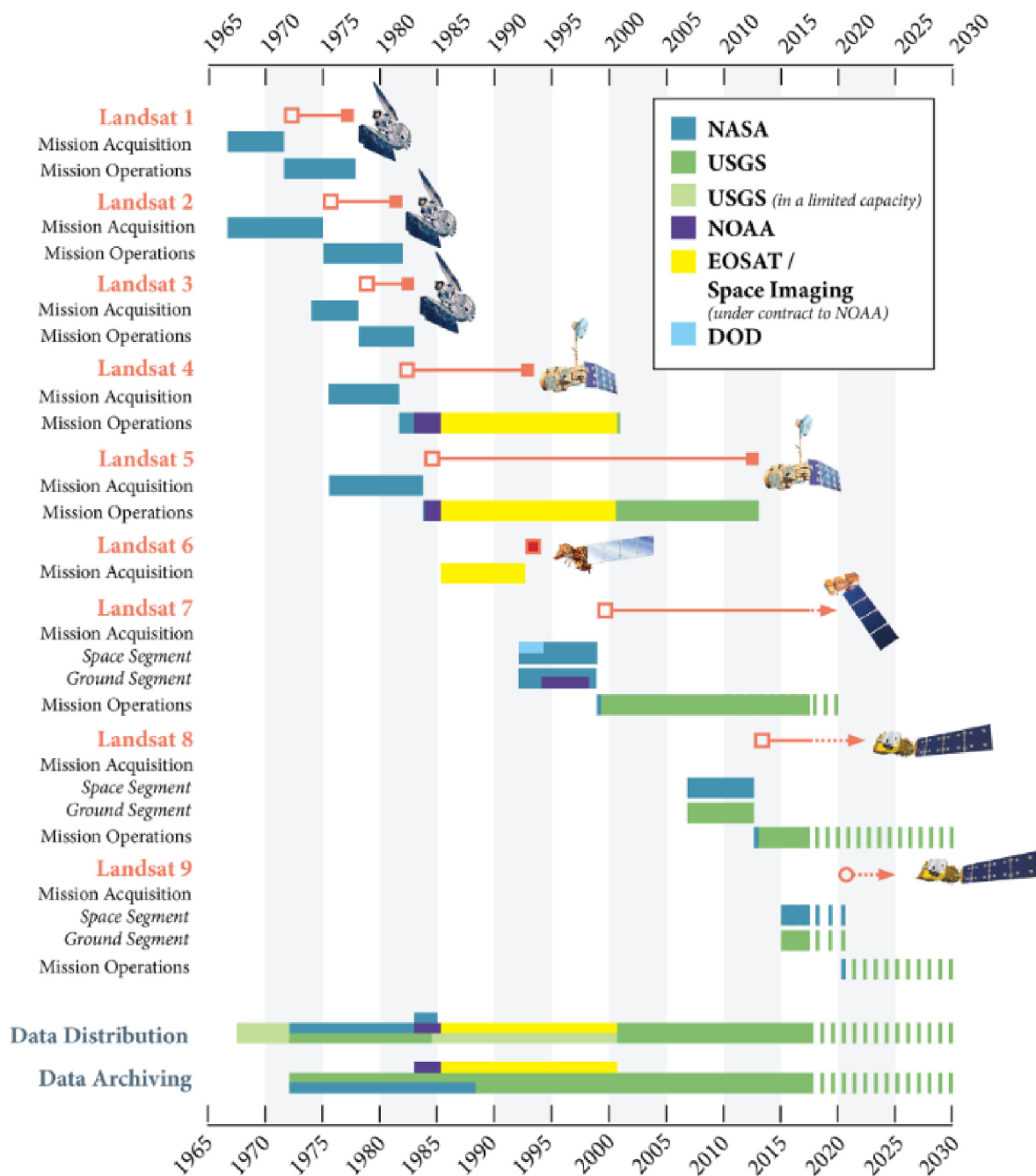
První družice byly vybaveny senzorem MultiSpectral Scanner (MSS). Tyto senzory snímaly rovnoběžně s dráhou družice a skenování v kolmém směru zajišťovalo oscilující zrcadlo. První dvě družice byly vybaveny senzorem MSS snímajícím ve čtyřech pásmech viditelného spektra. Družice Landsat 3 pak byla vybavena senzorem schopným zaznamenávat data i v pásmu tepelného infračerveného záření. Prostorové rozlišení senzorů bylo 68x83 m, většinou bylo převzorkováno na 57 či 60 m pro další práci se snímky. Jeden snímek pokrýval plochu 185x185 km (*Multispectral Scanner (MSS) | Landsat Science*, n.d.).

Další družice byly vybaveny senzorem Thematic Mapper (TM), který byl navržen, aby poskytoval vyšší prostorové rozlišení než MSS. Tento senzor snímá v

sedmi pásmech, s prostorovým rozlišením 30 m pro pásma viditelného a infračerveného světla a 120 m pro tepelné infračervené pásmo. Zlepšila se i geometrická přesnost senzoru. Snímky pokrývají plochu 185x172 km (*Thematic Mapper (TM)* | *Landsat Science*, n.d.).

Družice Landsat 7 byla jako první vybavena senzorem Enhanced Thematic Mapper Plus (ETM+), který nově snímá v panchromatickém pásmu (pásmu pokrývajícím celé viditelné spektrum) s prostorovým rozlišením 15 m. Ostatní pásma, tedy R, G, B, NIR a SWIR mají stále rozlišení 30 m. U pásma termálního pak došlo k významnému zvýšení rozlišení oproti TM, a to na 60 m. Plocha pokrytá snímkem je pak 183x170 km (*Enhanced Thematic Mapper Plus (ETM+)* | *Landsat Science*, n.d.).

Pokračování programu Landsat je zaručeno vyhláškou US Land Remote Sensing Act z roku 1992, která stanovuje, že vláda Spojených Států zaručí stálou globální archivaci snímků Landsat pro umožnění dlouhodobého pozorování zemského povrchu a jeho změn. Tato kontinuálnost je definována globálním pokrytím s časovým rozlišením 16 dní a bezplatným přístupem k celému archivu snímků (Zhu et al., 2019) Pro usnadnění práce s daty byl celý archiv snímků Landsat 1-8 přepracován do podoby tzv. Collection 1, které zajišťují konzistentní radiometrickou kvalitu těchto dat. Tímto krokem potenciálním uživatelům odpadá čas nutný k přípravě dat při používání snímků Landsat napříč různými obdobími a pořízených různými družicemi (Dwyer et al., n.d.).



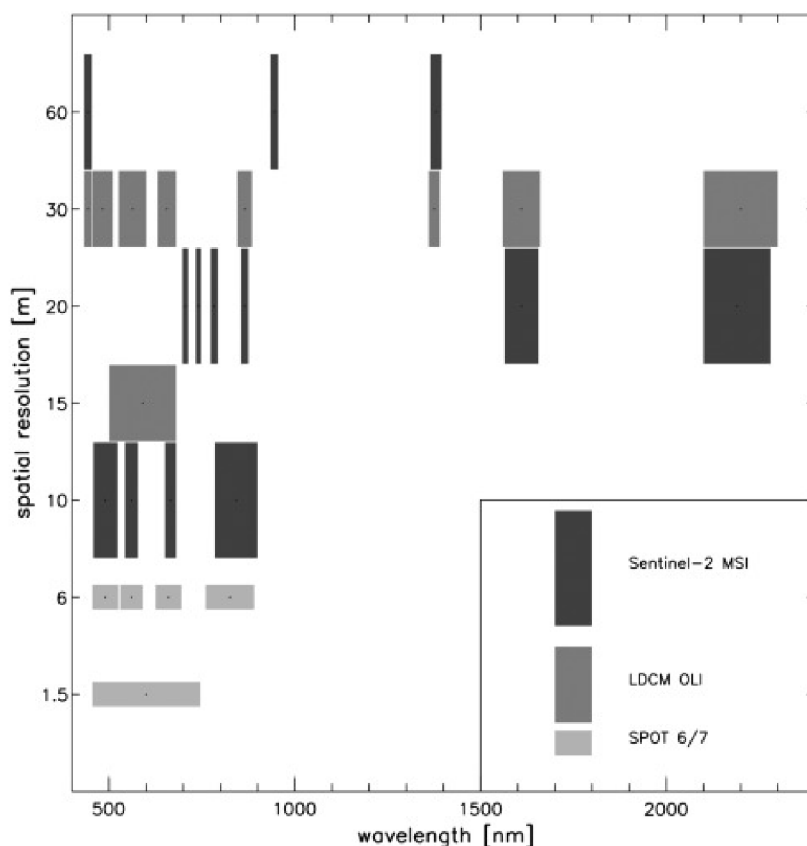
Obrázek 1 - Časová osa programu Landsat (Wulder et al., 2019)

3.4.2 Sentinel-2

Program Sentinel zahrnuje řadu družic určených pro snímání povrchu Země, a to s použitím rozličných senzorů. Mise Sentinel-2 představuje dvě identické družice nazvané Sentinel-2A a Sentinel-2B. Tyto družice obíhají Zemi na heliosynchronních drahách s časovým rozlišením 10 dní na rovníku. Družice mají oběžné dráhy posunuté proti sobě o 180°, dohromady tedy zvládnou pořídít snímek stejné oblasti na rovníku každých 5 dní. Tato vlastnost zajišťuje pokrytí s podrobným časovým rozlišením (Sentinel-2 - Missions - Sentinel Online - Sentinel Online, n.d.).

Hlavním cílem mise družic Sentinel-2 bylo navázat na mise dálkového průzkumu Země evropských družic SPOT a amerických družic Landsat. Dalším cílem bylo poskytnout novou generaci produktů jako jsou například právě mapy krajinného pokryvu, či jeho změn. 13 pásem (viz Obr. 2) snímače Multi Spectral Instrument (MSI) v tomto ohledu zajišťuje široké spektrum aplikací dat Sentinel-2 (*Sentinel-2 - Mission Objectives - Sentinel Online - Sentinel Online, n.d.*) Životnost družice byla počítána na 7,25 let, družice však mají zásobu paliva pro manévrovací motory na 12 let + rezervu pro de-orbiting na konci života družice (*Sentinel-2 - Satellite Description - Sentinel Online - Sentinel Online, n.d.*).

MSI je tzv. “push broom” skener, snímá tedy řádky dat napříč dráze letu a využívá dopředného pohybu družice k snímání nových řádků. Průměrná doba snímání je 17 minut. Radiometrické kalibrace jsou prováděny pomocí difuzéru, který je součástí závěrky senzoru. Pro zajištění šířky záběru 290 km je snímač vybaven dvanácti detektory, umístěnými ve dvou řadách s vzájemně se překrývajícím zorným polem. Snímač je rozdělen na 3 části, jednu pro pásma viditelného světla a blízkého infračerveného světla a dvě pro snímání v pásmech krátkovlnného infračerveného záření (SWIR) (*Sentinel-2 - Missions - Instrument Payload - Sentinel Handbook - Sentinel Online, n.d.*).



Obrázek 2- Skladba pásem družic Sentinel-2 v porovnání s ostatními družicemi používanými pro sledování zemského povrchu (Drusch et al., 2012).

Jak je patrné z obrázku 3, družice Sentinel-2 poskytují informace ve třech prostorových rozlišeních. Pásma viditelného světla a blízkého infračerveného záření jsou snímána v rozlišení 10 m. Pásma krátkovlnného infračerveného záření (SWIR) jsou pak snímána v rozlišení 20 m. Družice poskytují také 3 pásma snímána v rozlišení 60 m, ta jsou určena především pro sledování atmosféry, korekce a detekci oblačnosti (Drusch et al., 2012). Družice Sentinel zajišťují nepřetržité snímání zemského povrchu ve vysokém prostorovém rozlišení (Martimort et al., 2007).

Heliosynchronní oběžné dráhy družic Sentinel-2 způsobují, že každé snímané území, je snímáno v dalším cyklu přibližně ve stejnou denní dobu. Tato vlastnost je zvláště důležitá pro efektivní využití senzorů snímajících v optické části elektromagnetického spektra (Paek & Kim, 2018). Čas průletu nad rovníkovými oblastmi byl zvolen na 10:30. Tento čas byl vybrán z důvodu minimalizace oblačnosti v době přeletu a pro zaručení optimálních světelných podmínek (Phiri et al., n.d.).

3.5 Současný stav řešené problematiky

V současné době dochází k výraznému nárůstu počtu studií zaměřených na klasifikaci krajinného pokryvu pomocí dat z družic Sentinel-2. Zatímco v minulosti byl hlavním zdrojem dat pro tyto účely program Landsat, družice Sentinel-2 přinášejí nové možnosti v oblasti analýz krajinného pokryvu. Důvodem tohoto posunu je zejména vysoké prostorové rozlišení snímků, které umožňuje detailnější a přesnější zobrazování krajiny (Topaloğlu & Sertel, n.d.), a také vysoké časové rozlišení, díky němuž je možné sledovat změny v krajinném pokryvu v průběhu času s větší časovou citlivostí. Díky bezplatné dostupnosti snímků z družic Sentinel-2 jsou tyto data stále atraktivnější pro vědecký výzkum a aplikace v oblasti environmentálního monitoringu, územního plánování, a dalších oblastech. Vzhledem k těmto vlastnostem se družice Sentinel-2 stávají ideálními kandidáty pro provádění klasifikací a analýz krajinného pokryvu na celosvětové úrovni (Phiri et al., n.d.).

Pro tyto klasifikace, bylo vyvinuto mnoho metod již v době předcházející spuštění programu Sentinel-2. V dnešní době jsou potom nejrozšířenější klasifikátory založené na metodách strojového učení, jako jsou Random Forest (RF), K-Nearest Neighbour (KNN), či Support Vector Machine (SVM) (Talukdar et al., n.d.). Z těchto klasifikátorů je pak uživateli nejčastěji využíván Random Forest (Inglada et al., 2015).

Klasifikace můžeme dělit dle vstupu ze strany uživatele na řízené (supervised) a neřízené (unsupervised). Z těchto dvou možností jsou v současnosti častěji využívané klasifikace řízené, a to z toho důvodu, že většina aktuálních pokročilých klasifikátorů byla navržena právě pro tento druh klasifikací (Miranda et al., 2018). Řízené klasifikace spočívají v definování klasifikačních tříd uživatelem, vyžadují tedy jistou znalost území, či alespoň předpoklad výskytu těchto tříd ve zkoumaném území. Dále je vyžadováno definování trénovacích dat uživatelem, tato data následně slouží pro

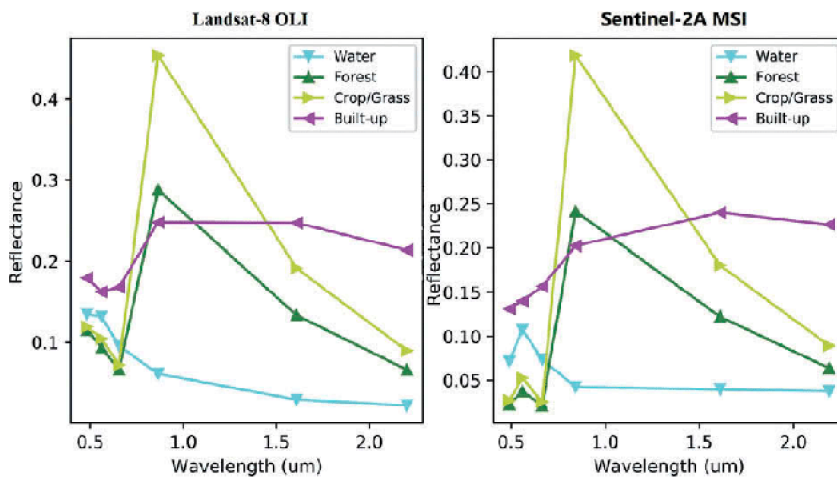
učení klasifikátoru. Tyto klasifikace vykazují často vyšší přesnost než klasifikace neřízené, právě z důvodu definování tříd a jejich vlastností pomocí trénovacích vzorků uživatelem (Ahmad & Quegan, 2013). U neřízených klasifikací pak bývá obvykle jediným vstupem uživatele definování počtu tříd, případně počtu iterací, které klasifikátor při analýze provede. Prvotní výsledek neřízené klasifikace pak většinou vyžaduje revize spočívající ve sloučení některých tříd uživatelem apod. Tento přístup ale může být výhodný, pokud není k dispozici dostatečné množství trénovacích dat, či pokud je cílem minimalizovat vstup uživatele při jejich tvorbě (Thomson et al., 1998).

Dalším způsobem dělení klasifikace je na objektové a pixelové. Toto rozdělení rozlišuje přístup klasifikátoru k samotné klasifikaci. Pixelové klasifikace analyzují spektrální vlastnosti každého pixelu v zájmovém území, ovšem nevnímají tyto údaje v kontextu k danému pixelu. S rozvojem technologií poskytující vyšší prostorové rozlišení tyto klasifikační přístupy vedou ke stále lepším výsledkům (Dwivedi et al., 2004). Při vysokém prostorovém rozlišení snímků však může při použití tohoto pixelového přístupu docházet k tzv. "salt & pepper efektu", který naopak přesnost klasifikace může snižovat, protože jednotlivé pixely jsou klasifikovány bez ohledu na jejich kontext. Může tak vznikat například situace kdy jednotlivé stromy jsou klasifikovány jako pixely lesa uprostřed louky (De Jong et al., 2001). V návaznosti na tento problém byly vyvinuty metody klasifikace objektové. Objektové klasifikátory zahrnují nejen spektrální informace uložené v jednotlivých pixelech, ale také zvažují informace o prostorovém uspořádání těchto dat. To znamená, že jsou schopny brát v úvahu sousedící pixely a jejich vzájemné vztahy v rámci snímku. Jedním z klíčových kroků při použití objektových klasifikátorů je segmentace obrazu, což znamená rozdělení obrazu do různých oblastí na základě jejich podobnosti ve spektrálních vlastnostech a prostorovém uspořádání. Tento postup umožňuje dosažení vyšší přesnosti klasifikace tím, že zachycuje větší kontext a strukturální informace v datech, což vede k robustnějším výsledkům a lepšímu porozumění prostředí. (Blundell & Opitz, n.d.).

V souladu s tématem a cíli diplomové práce se dále zaměřím na pixelové řízené klasifikace. Tyto klasifikace odlišují definované třídy na základě rozdílů ve spektrálních příznacích jednotlivých tříd krajinného pokryvu. Při definování těchto tříd je tak potřeba brát v potaz jejich podobnost, neboť vysoká podobnost spektrálních profilů jednotlivých tříd bude mít za následek nízkou přesnost klasifikace. Tato problematika je zkoumána mnoha studiemi a existují řady databází spektrálních příznaků nejrůznějších druhů povrchů (Bojinski et al., 2003).

Pokud budeme uvažovat spektrální profily základních tříd klasifikace dle CORINE Land Cover, tak zjistíme, že zastavěné oblasti a umělé povrchy mají spektrální profily ploché, často s o něco menší odrazivostí v pásmech viditelného světla než v pásmech SWIR. Naproti tomu trávy a zemědělské plodiny vykazují

vysokou odrazivost v pásmech zeleného světla a blízkého infračerveného záření. Lesní oblasti mají spektrální profil podobný, ovšem odrazivost v zeleném a infračerveném pásmu bývá oproti jiné vegetaci nižší. Voda pak vykazuje relativně nízkou odrazivost ve všech pásmech používaných senzory DPZ, s možnými vyššími hodnotami v pásmech modrého či zeleného světla v závislosti na jejím složení či růstu vodní vegetace, viz Obr. 3 (XI et al., 2019).



Obrázek 3 - Spektrální profily základních druhů krajinného pokryvu dle dat Landsat 8 a Sentinel-2 (XI et al., 2019).

Co se týká různých klasifikátorů pro řízené pixelové klasifikace, nejčastěji používanými jsou klasifikátory využívající strojové učení jako jsou Random Forest (RF), K-Nearest Neighbour (KNN) a Support Vector Machine (SVM). Všechny tyto klasifikátory obvykle vykazují vysokou přesnost klasifikace. Samozřejmě v závislosti na spektrální podobnosti klasifikovaných tříd, počtu trénovacích vzorků definovaných uživatelem pro jednotlivé třídy a prostorovém rozlišení používaných snímků. Rozdíly v přesnosti klasifikace se obvykle pohybují v jednotkách procent (Noi & Kappas, n.d.).

V poslední době se začínají objevovat již i studie využívající pokročilé klasifikační metody na bázi hlubokého učení. O tyto technologie je v oblasti dálkového průzkumu Země stále větší zájem, což je i motivací pro potenciální investory. Vzhledem k tomu, že množství dat ke zpracování a analýzám se neustále zvětšuje, jsou stále hledány nové metody, jak tyto analýzy zjednodušit, či automatizovat. Právě jedním ze způsobů, které se nabízejí, je využívání pokročilých klasifikátorů hlubokého učení, které svůj potenciál zvětšují úměrně množství dat použitelných pro jejich učení (Vali et al., 2020).

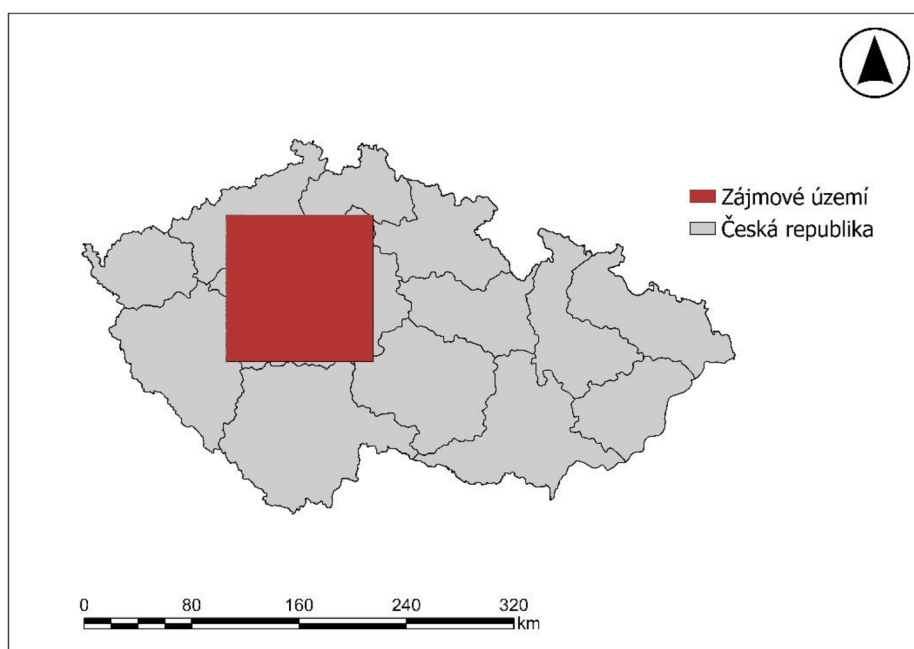
Tyto nové technologie ale přináší i nové výzvy. Pro nejefektivnější fungování těchto klasifikátorů je potřeba velké množství trénovacích dat. Jedním z problémů však je samotná tvorba těchto dat, neboť například pro klasifikace krajinného pokryvu a využití krajiny může být obtížné dostatečné množství spolehlivých trénovacích vzorků zajistit. Tyto problémy plynou i z již zmiňovaných výzev spojených se

současným sběrem informací o krajinném pokryvu (Comber et al., 2005). Když už se povede získat dostatečně rozsáhlé množství trénovacích dat, vlastnosti samotných dat DPZ činí modely komplexnějšími, a to ztěžuje jejich optimalizaci (Alem & Kumar, 2020).

Metody hlubokého učení obvykle v oblasti dálkového průzkumu země nejsou využívány experty na umělou inteligenci.(Vali et al., 2020). Přesto však některé studie naznačují jejich potenciál pro klasifikaci krajinného pokryvu (Karra et al., 2021; Jo et al., n.d.), či pro detekci jeho změn(Saha et al., 2019).V současné době se však rozvíjí možnosti použití těchto metod v rámci komerčních programů zaměřených právě na problematiku DPZ. Společnost Esri nabízí v rámci svého programu Living Atlas katalog již vytvořených a natrénovaných modelů pro různé aplikace. Zároveň poskytuje nástroje pro tvorbu těchto modelů a další práci s nimi. Software ArcGIS Pro využívá pro své modely knihovnu Tensor Flow a její přidružené knihovny, které jsou zároveň volně dostupné pro programovací jazyk Python. ArcGIS Pro v současné době podporuje objektové a pixelové klasifikace snímků a klasifikace bodových mračen (*Deep Learning in ArcGIS Pro—ArcGIS Pro | Documentation*, n.d.).

4 Charakteristika studijního území

Zájmové území diplomové práce představuje rozsah jednoho snímku Sentinel-2, zachycujícího Prahu a většinu Středočeského kraje, s přesahem do Karlovarského kraje v severozápadní části (viz Obr. 4). Toto území jsem volil tak, aby pokrývalo pouze území České republiky, a zároveň aby zde byla zastoupena většina druhů krajinného pokryvu. Díky využití celého tile (dlaždice) nebylo nutné vytvářet mozaiky z více snímků, což považuji za praktické pro účely této práce, protože tak nemohla být vstupní data ovlivněna případnými neurčitostmi vzniklými právě procesem mozaikování. Celková rozloha tohoto území je 12 056 km².



Obrázek 4 - Umístění zájmového území v rámci ČR.

Území je z velké části pokryto zemědělskou půdou a lesy a v jeho středu se nachází hlavní město Praha. Terén v této oblasti je členitý a kopcovitý. Centrum města se vyznačuje hustou zástavbou a pouze malými ploškami zeleně. V okrajových částech se pak nachází modernější obytná zástavba a průmyslové zóny. Středem města protéká řeka Vltava, v jižní části se do ní vlévá řeka Berounka. V severní části zájmového území se nachází Český kras, známý především vápencovými skalními útvary. V této oblasti se nachází větší podíl zemědělské půdy. V jižní části území se pak nachází rozsáhlé zalesněné oblasti jako je Karlštejsko, Brdy, Křivoklátsko a další.

5 Metodika

Metodická část diplomové práce se skládá z kapitol popisujících vstupní data, která jsem v průběhu práce používal, jejich prvotnímu zpracování, které bylo pro provedení dalších kroků nutné, kapitoly věnující se vytvoření vlastního klasifikačního modelu, samotné klasifikace dat a validace výsledků. Pro veškerou práci s daty jsem používal software ArcGIS Pro, zejména jeho nástroje deep learningu.

5.1 Vstupní data

Veškeré snímky z družic Sentinel-2 poskytl program Copernicus prostřednictvím služby Copernicus Dataspace. Snímky jsou úrovně zpracování Level-2A (L2). Volil jsem snímek, pokud možno bezoblačný nebo s minimální oblačností, primárně z jarních a letních měsíců. Konkrétně se jednalo o snímek pořízený družicí Sentinel-2A dne 3. 7. 2018.

Snímky družic Sentinel-2 pořizovány ve třinácti pásmech označených B01 – B12, kde B představuje zkratku pro spektrální pásmo (z angl. slova band), v prostorovém rozlišení 10, 20 a 60 m. Vysoké rozlišení 10 m je poskytováno pouze pásmy viditelného světla (RGB) B02-B04 a pásmem blízkého infračerveného záření (NIR) B08.

Jako referenční data jsem používal data Corine Land Cover (CLC) 2018, která rovněž poskytuje program Copernicus v rámci tematického projektu Copernicus Land Monitoring service (CORINE Land Cover — Copernicus Land Monitoring Service, n.d., dostupné z: <https://land.copernicus.eu/en/products/corine-land-cover>) Data je možné zdarma stáhnout jak ve vektorovém, tak rastrovém formátu. Pro potřeby diplomové práce byla použita data v podobě vektorů.

Pro účely porovnání různých klasifikačních modelů, jsem se rozhodl pro využití předtrénovaného modelu dostupných v prostředí ArcGIS Living Atlas. konkrétně jsem zvolil model “*Land Cover Classification (Sentinel-2)*” (Esri, n.d., dostupné z: <https://www.arcgis.com/home/item.html?id=afd124844ba84da69c2c533d4af10a58>).

Vstupem modelu jsou snímky Sentinel-2 v úrovni zpracování 1C nebo 2A, výstupem pak rastr krajinného pokryvu kategorizovaný do 15ti tříd (jsou-li v daném snímku zastoupeny). Model by měl dosahovat kvalitních výsledků pro zájmová území blízká se svými podmínkami prostředí státům USA a EU (Esri, n.d.), viz Tab. 2. Pro účely

této práce jsem používal kvůli vyšší deklarované přesnosti klasifikace pouze snímky úrovně zpracování 2A.

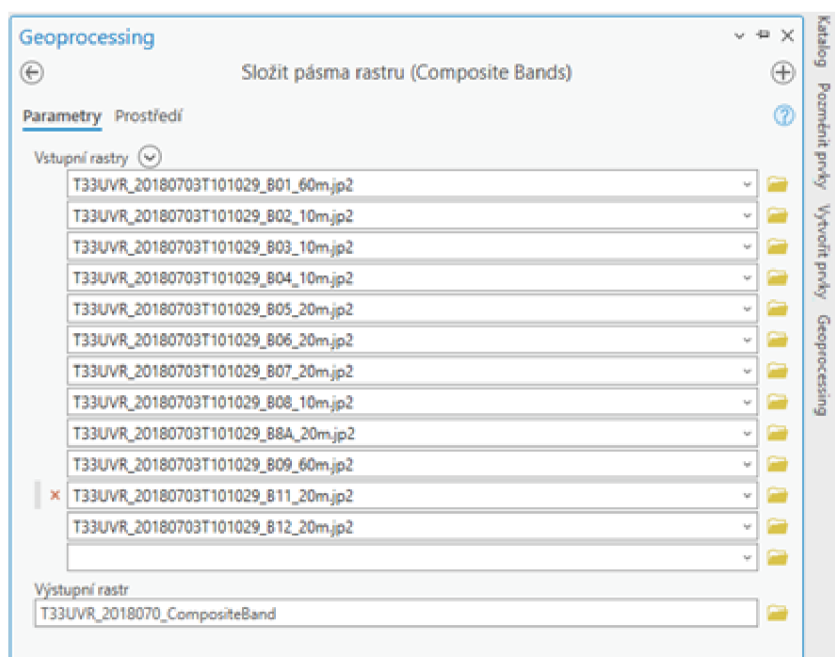
Class	Level-2A Imagery			Level-1C Imagery		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Urban fabric	0.81	0.83	0.82	0.82	0.84	0.83
Industrial, commercial and transport units	0.74	0.65	0.69	0.73	0.66	0.7
Mine, dump and construction sites	0.63	0.52	0.57	0.69	0.55	0.61
Artificial, non-agricultural vegetated areas	0.70	0.46	0.55	0.67	0.47	0.55
Arable land	0.86	0.90	0.88	0.86	0.89	0.87
Permanent crops	0.76	0.73	0.74	0.75	0.71	0.73
Pastures	0.75	0.71	0.73	0.74	0.71	0.73
Heterogeneous agricultural areas	0.61	0.56	0.58	0.62	0.51	0.56
Forests	0.88	0.93	0.90	0.88	0.92	0.9
Scrub and/or herbaceous vegetation associations	0.74	0.69	0.72	0.73	0.67	0.7
Open spaces with little or no vegetation	0.87	0.84	0.85	0.85	0.82	0.84
Inland wetlands	0.81	0.78	0.80	0.82	0.77	0.79
Maritime wetlands	0.74	0.76	0.75	0.87	0.89	0.88
Inland waters	0.94	0.92	0.93	0.94	0.91	0.92
Marine waters	0.98	0.99	0.98	0.97	0.98	0.98

Tabulka 2 - Tabulka klasifikačních tříd a deklarované přesnosti pro model Land Cover Classification (Sentinel-2) (Land Cover Classification (Sentinel-2) - Přehled, *n.d.*).

5.2 Předzpracování dat

Pro účely klasifikace pomocí hlubokého učení, bylo nejprve nutné “složit” jednotlivé rastry pásem snímků Sentinel-2 do jedné kompozitní datové sady, neboť software ArcGIS při načtení metadatového souboru družice Sentinel-2 pracuje pouze s pásmy RGB + NIR, čímž znemožňuje použití předtrénovaných modelů, které v práci porovnávám. Při skládání pásem by došlo k převzorkování na jednotné prostorové rozlišení 60 m. Toto rozlišení vychází z prvního pásma, které do nástroje vstupuje, v

mém případě tedy z pásma B01. Protože jsem se snažil o klasifikaci co možná nejpřesnější, bylo potřeba toto pásmo převzorkovat na prostorové rozlišení 10 m. Převzorkování jsem prováděl nástrojem *Resample* s nastavením vstupní velikosti buňky 60 m a výstupní velikosti 10 m. Složení pásem jsem prováděl pomocí nástroje *Composite bands* (viz Obr. 5). Při skládání kompozitu, bylo třeba věnovat pozornost v jakém pořadí jsou jednotlivá vstupní pásma do nástroje zadávána. Chybné pořadí pásem vrací v případě předtrénovaných modelů nepoužitelné výsledné rastry.



Obrázek 5 - Pořadí pásem používané pro kompozitní rastry.

Výsledkem použití nástroje *Composite Bands* byly rastry o dvanácti pásmech v pořadí odpovídajícím produktům Sentinel-2. Tyto nové rastry byly uloženy ve formátu TIFF (.tif).

5.3 Předtrénovaný model

Model jsem aplikoval na předpřipravený rastrový snímek. Při použití tohoto modelu jsou uživatelé nabízeny různé možnosti jeho nastavení, tyto hodnoty jsem však ponechal výchozí, aby nedošlo k ovlivnění výsledku mým vstupem. rámci prvotní kontroly vizuálně porovnal s produktem Corine Land Cover. Čas zpracování snímku se pohyboval okolo dvou hodin.

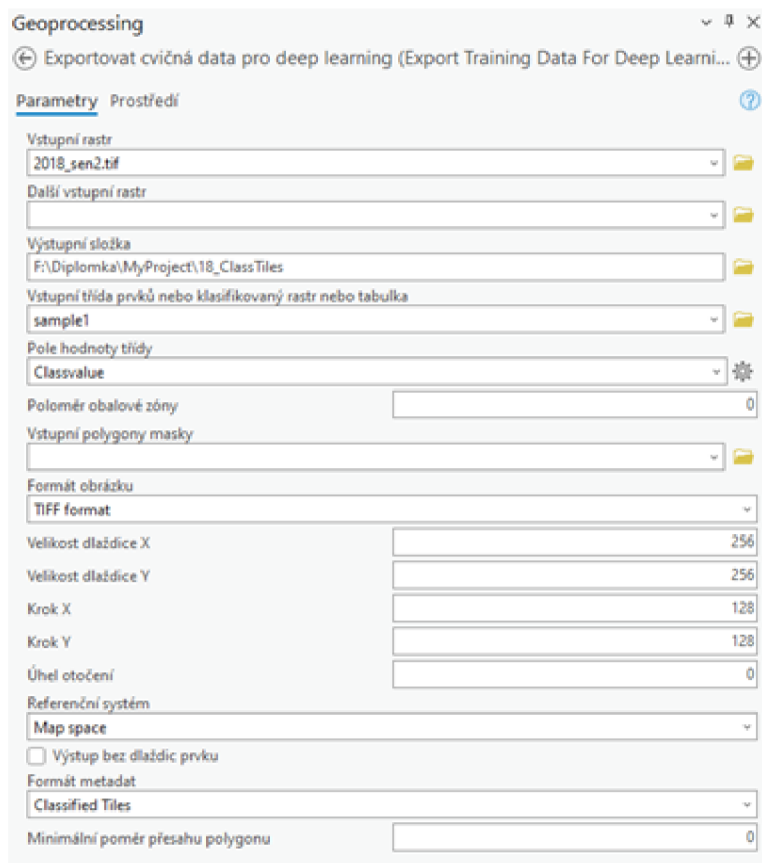
5.4 Tvorba vlastního modelu

Nejprve jsem vytvořil trénovací data z produktu CORINE Land Cover. Pro tyto účely jsem musel použít vektorovou verzi tohoto produktu, neboť umožňuje jednodušší práci se svými atributy. Vrstvu CLC jsem ořízl na plochu zájmového území a jednotlivé polygony sloučil do multipolygonů podle jednotlivých tříd CLC (atribut CODE_18). V atributové tabulce této nové sjednocené vrstvy jsem vytvořil sloupce “classname” a “classvalue”, které jsou nutné pro optimální kompatibilitu s nástroji pro klasifikaci. Sloupec “classvalue” obsahuje odpovídající hodnoty ze sloupce “CODE_18”, “classname” pak obsahuje názvy jednotlivých tříd CLC.

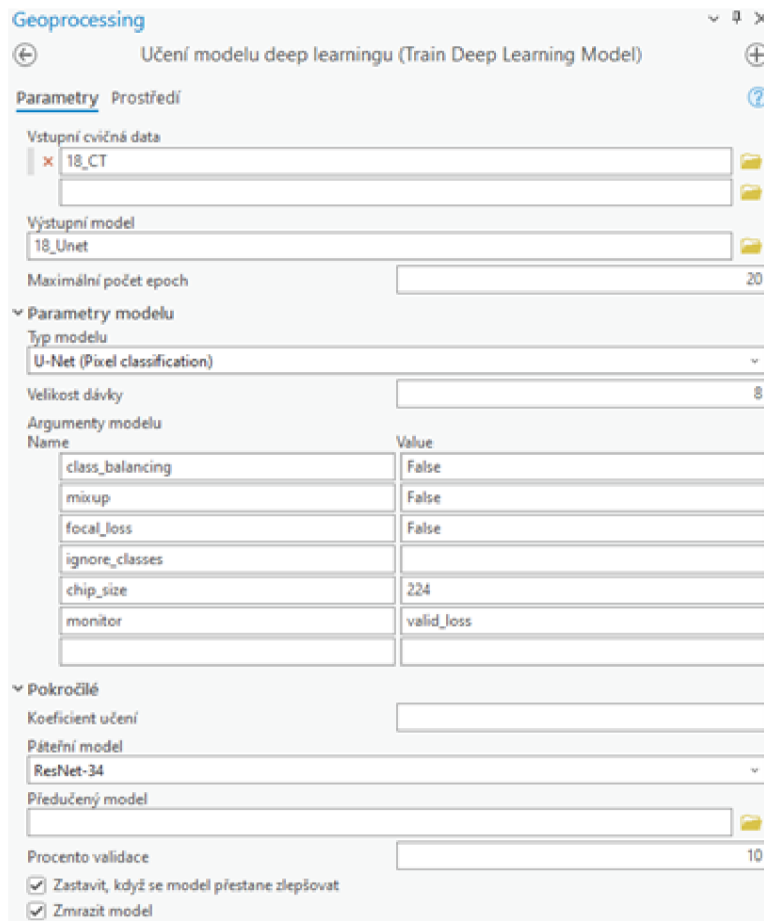
Pomocí nástroje *Create random points* jsem uvnitř polygonů jednotlivých tříd vytvořil náhodné body. Bodů bylo vytvořeno 50 pro každou třídu a byly rozmístěny v minimální vzdálenosti 50 m od sebe. Pomocí nástroje *Generate training samples from seed points* jsem poté na základě vrstvy náhodných bodů vygeneroval trénovací plochy. Tato bodová vrstva byla uložena jako shapefile trénovacích ploch.

Tato data jsem používal pro vstup do nástroje *export training data for deep learning* (viz Obr. 4). společně s již připraveným rastrem družicového snímku. V tomto nástroji je mnoho nastavení, pro mou práci bylo nejdůležitější nastavit správný formát metadat, neboť ten určuje, které modely bude možné z daných dat natrénovat. Pro tuto práci jsem volil formát Classified tiles, protože tento formát je kompatibilní s většinou modelů pro pixelovou klasifikaci. Rovněž je třeba věnovat pozornost formátu vrstvy s referenčními daty. ArcGIS Pro sice uvádí jako možnost pro tyto účely použít jakoukoliv sadu prvků (feature class), či klasifikovaný rastr, avšak pouze některé datové typy atributů mohou do nástroje vstupovat.

Po exportu trénovacích vzorků (viz Obr. 6) následovalo učení modelu pomocí nástroje *Train deep learning model* (viz Obr. 7). Zde je nutno vybrat architekturu modelu kompatibilní s metadaty trénovacích vzorků (v mém případě Classified tiles). Pro účely práce jsem volil architekturu U-net. Tato architektura je používána pro pixelovou klasifikaci (*Deep Learning Model Architectures—ArcGIS Pro | Documentation*, n.d.) a je využita i v předučeném modelu, z předchozí kapitoly.

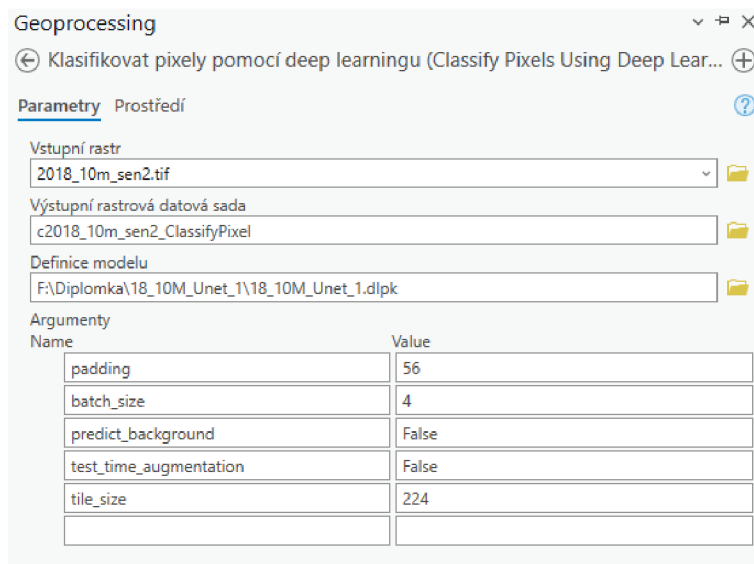


Obrázek 6 - Nastavení nástroje Export Training Data For Deep Learning.



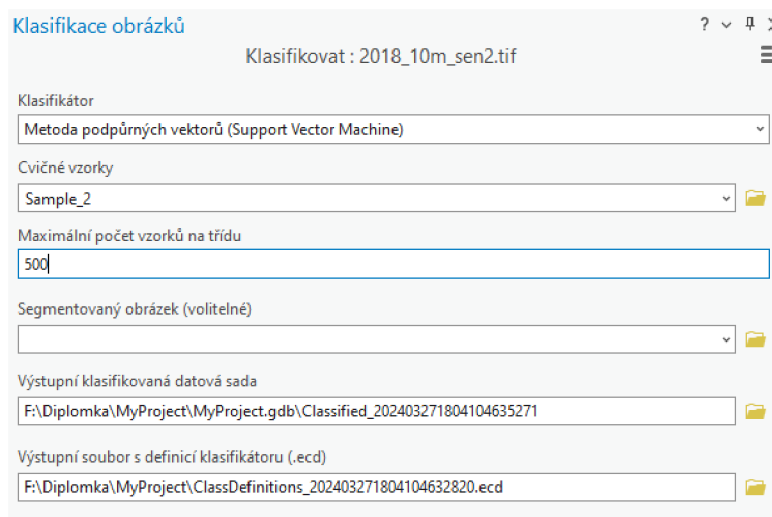
Obrázek 7 - Nastavení nástroje Train Deep Learning Model.

Po učení modelu následovaly již samotné klasifikace. Ty jsem prováděl nástrojem *Classify pixels using deep learning*. Do nástroje vstupoval snímek pro klasifikaci a již natrénovaný model z předchozího kroku. Je zde možné nastavit parametry modelu, rozsah klasifikované plochy a procesor, který bude pro výpočty používán. Tato nastavení jsem ponechával výchozí (viz Obr. 8). Nejprve na menším výřezu snímku abych si ověřil, že postup funguje bez zbytečného čekání, poté již na celém snímku. Celkový čas klasifikace se pohyboval okolo 3 hodin.



Obrázek 8 - Nastavení nástroje *Classify Pixels Using Deep Learning*.

Pro účely porovnání jsem snímek také klasifikoval algoritmem Support Vector Machine (SVM), který je pro řízenou pixelovou klasifikaci běžně používán. Tyto klasifikace jsem prováděl s použitím stejného schématu klasifikace vytvořeného na základě dat CLC. Používal jsem trénovací vzorky identické se vzorky používanými pro učení modelu hlubokého učení. Pro tuto klasifikaci jsem používal nástroje pro klasifikaci snímků (viz. Obr. 9) dostupné v záložce *Obrazová data*.



Obrázek 9 - Nastavení nástroje pro klasifikaci snímku klasifikátorem SVM.

5.5 Validace výstupů klasifikace

Validaci vytvořeného modelu jsem prováděl nástrojem *Accuracy assessment* který porovnává klasifikovaný rastr s referenčními daty, v mém případě daty CLC na náhodně generovaných bodech. Toto porovnání probíhalo na 10 000 náhodně

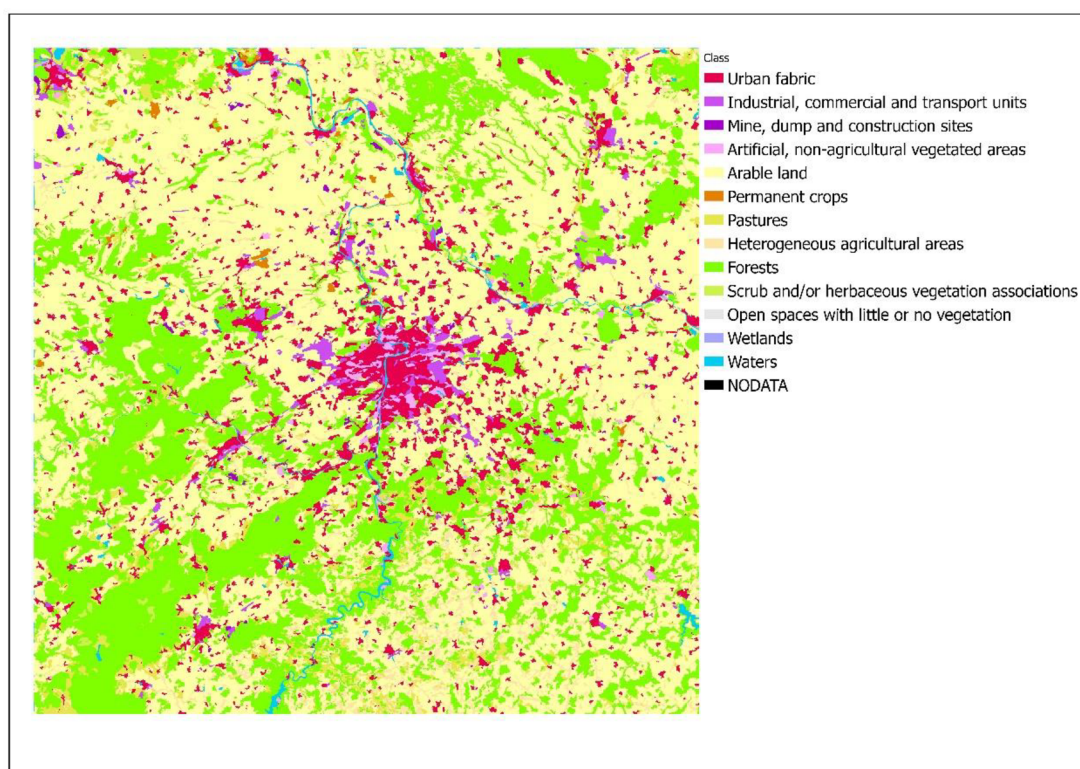
generovaných bodech, jejichž počet v každé třídě byl úměrný poměru plochy dané třídy ku celkové ploše území. Výsledkem je pak matice záměn.

Pro porovnání přesnosti s předtrénovaným modelem, jsem rovněž u všech výsledných rastrů provedl reklasifikaci na nadřazené třídy CLC. Z trojmístného kódu třídy jsem zachoval pouze první dvojčíslí, vzniklo tím 12 kategorií. Po provedení reklasifikace již ArcGIS Pro neumožňoval použití nástroje *Accuracy assessment*, volil jsem proto náhradní postup. Pro tyto nové rastry jsem vygeneroval 10 000 náhodných kontrolních bodů pomocí funkce *Create accuracy assessment points*, do atributů těchto bodů byla uložena referenční data vrstvy CLC a hodnoty klasifikovaného rastru pomocí funkce *update accuracy assessment points*. Z výsledných bodových vrstev jsem pak vypočítal matice záměn pomocí funkce *compute confusion matrix*.

6 Výsledky

6.1 Předtrénovaný model

Předtrénovaný model Land Cover Classification (Sentinel-2) dosáhl uspokojivého výsledku klasifikace (viz Obr. 10). Model poskytuje klasifikaci na druhé úrovni dělení tříd dle CLC, ovšem dané třídy klasifikuje kvalitně a je relativně konzistentní v přesnosti klasifikace jednotlivých tříd. Celková přesnost klasifikace v mém případě dosahovala 74 % (viz Tab. 3) Z tabulky 3 je také patrné, že nejvyšších hodnot producer accuracy (PA) a user accuracy (UA) dosahovala třída orné půdy (Arable land) a to 81 % PA a 83% UA, přičemž nejčastěji u ní docházelo k záměnám s heterogenními zemědělskými oblastmi (Heterogenous agricultural areas).



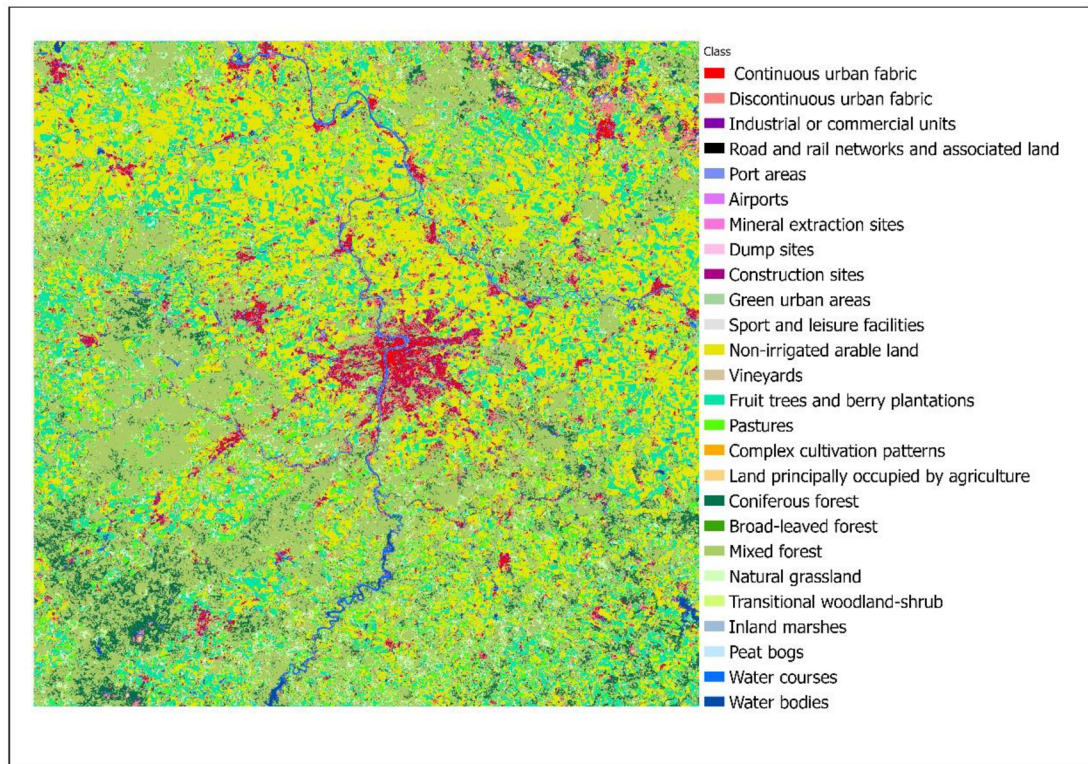
Obrázek 10 - Výsledek klasifikace předtrénovaným modelem.

Class/Value	C_11	C_12	C_13	C_14	C_21	C_22	C_23	C_24	C_31	C_32	C_41	C_51	Total	Uj_Accuracy	Kappa
C_11 - Urban fabric	541	31	3	19	122	0	10	85	38	4	0	1	854	0.633	0
C_12 - Industrial, commercial and transport units	6	84	3	1	25	0	2	6	9	1	1	0	138	0.609	0
C_13 - Mine, dump and construction sites	1	0	14	0	5	0	1	3	1	2	0	1	28	0.500	0
C_14 - Artificial, non-agricultural areas	3	2	0	32	13	0	0	10	30	1	0	0	91	0.352	0
C_21 - Arable land	90	20	6	1	4017	48	137	284	242	10	2	0	4865	0.826	0
C_22 - Permanent crops	0	1	0	0	6	8	2	3	1	0	0	0	21	0.381	0
C_23 - Pastures	17	0	0	6	228	9	184	90	39	1	0	1	578	0.318	0
C_24 - Heterogenous agricultural areas	9	1	0	6	185	0	48	220	150	3	0	6	628	0.351	0
C_31 - Forests	9	10	2	7	192	3	23	118	2237	69	4	3	2677	0.836	0
C_32 - Shrub and/or herbaceous vegetation associations	0	1	0	0	14	0	2	3	9	13	0	0	42	0.310	0
C_41 - Inland wetlands	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000	0
C_51 - Inland waters	3	1	1	1	8	0	2	9	5	0	3	55	88	0.625	0
Total	687	151	29	73	4815	68	409	834	2761	104	10	67	10008	0.000	0
P_Accuracy	0.787	0.556	0.483	0.438	0.834	0.118	0.450	0.264	0.810	0.125	0.000	0.821	0	0.74	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.62

Tabulka 3 - Matice záměn předtrénovaného modelu.

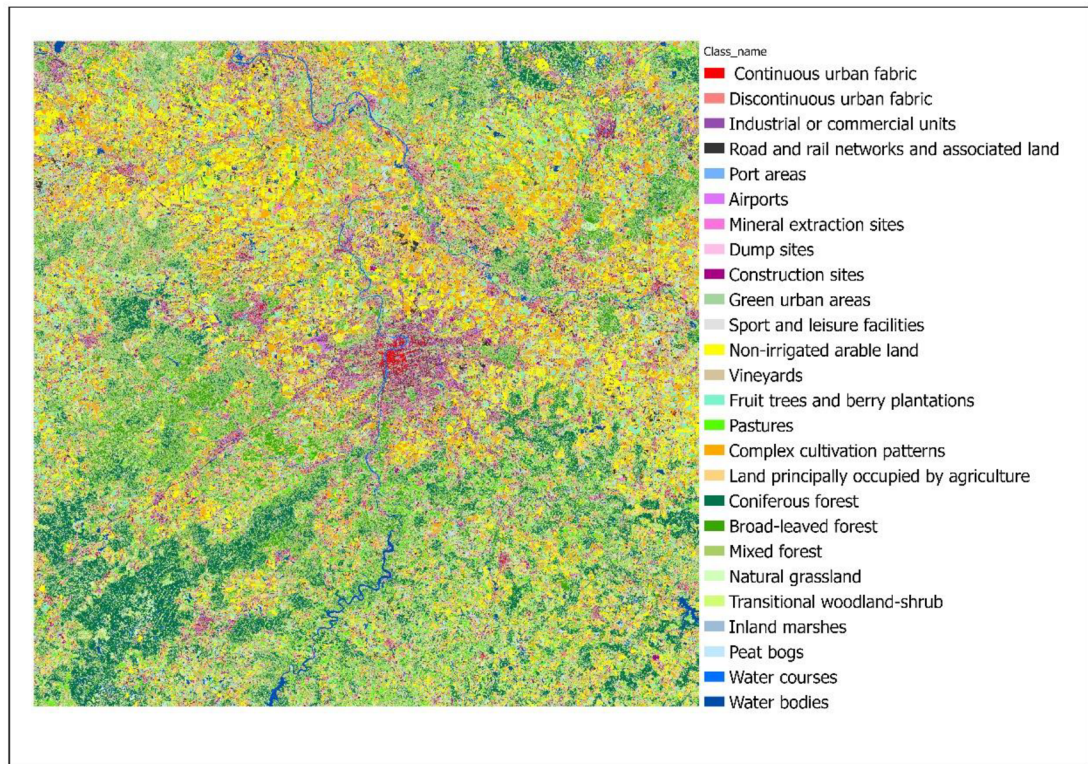
6.2 Vlastní model

Model vytvořený na základě trénovacích dat CLC poskytl rozpoznatelné výsledky (viz Obr. 11). Na klasifikovaném rastru je již na první pohled jasně rozpoznatelný krajinný pokryv zájmového území, ovšem přesnost zařazení do jednotlivých tříd není vysoká (viz příloha 1). Model například klasifikuje většinu břehů řek jako přístavní oblasti (Port areas), nedokáže efektivně rozlišit letiště, dopravní infrastrukturu ani oblasti pro sportovní aktivity. Má rovněž problém rozlišit hustou zástavbu (Continuous urban fabric) od zástavby roztroušené (Discontinuous urban fabric). Celková přesnost byla 34 %, Nejvyšších hodnot PA dosahovala třída husté zástavby (Continuous urban fabric) a to 80 %, nejvyšších hodnot UA pak třída nezavlažované orné půdy (Non-irrigated arable land) a to 83 %. Rašeliniště (Peat bogs), bažiny, (inland marshes), staveniště (construction areas) a skládky (dump sites), model nedokázal klasifikovat vůbec.



Obrázek 11 - Výsledek klasifikace modelem U-net na základě trénovacích vzorků generovaných z produktu CLC.

Pro účely porovnání jsem snímek klasifikoval rovněž pomocí algoritmu SVM (viz Obr. 12). Tento algoritmus poskytl výsledky ještě horší, než model hlubokého učení (viz příloha 2). Algoritmus SVM sice některé třídy klasifikuje s vyšší přesností, ovšem celková přesnost je pouhých 21 %. Na rozdíl od mého modelu však dokázal zastoupit všechny třídy.

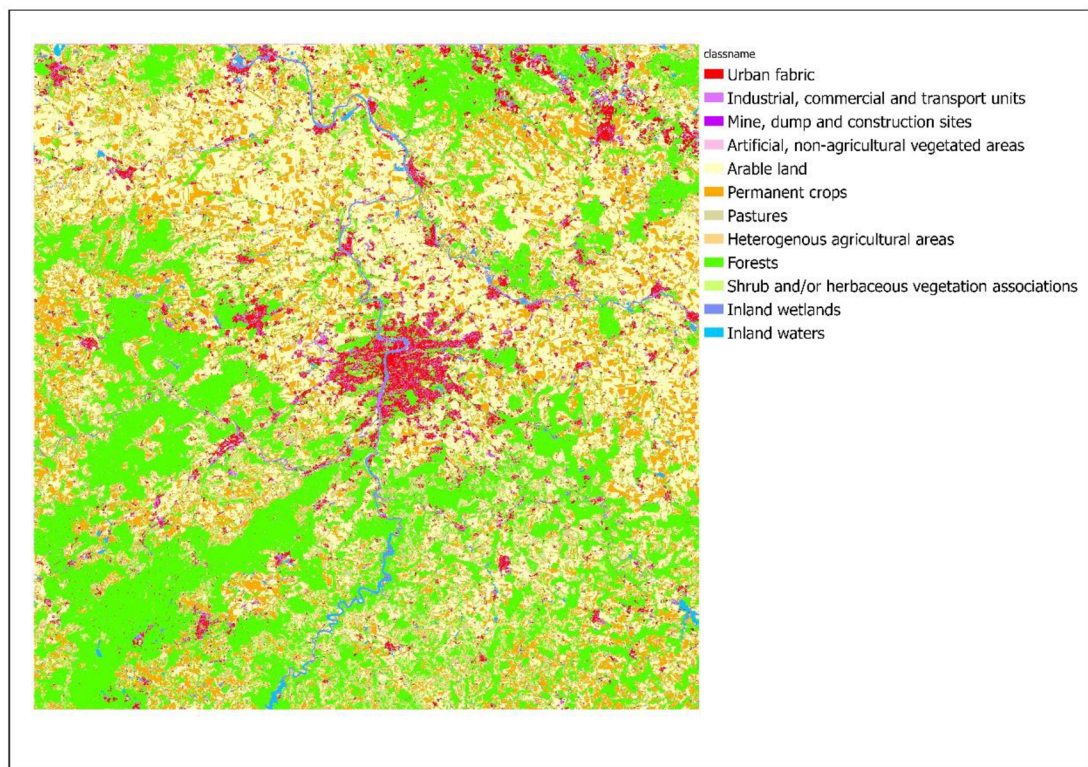


Obrázek 12 - Výsledek klasifikace klasifikátorem SVM

Po překlasifikování výsledku mnou vytvořeného modelu hlubokého učení na 12 tříd (viz Obr. 13), dle druhé úrovně CLC se zvýšila přesnost klasifikace, neboť se eliminovaly záměny výrazně podobných tříd (například Continuous urban fabric-Discontinuous urban fabric). Celková přesnost vzrostla o 12 %, tedy na 46 % (viz Tab. 4). Z této tabulky je také patrné, že můj model dosahuje nejvyšších hodnot PA pro lesní oblasti a to 69 %. Nejvyšších hodnot UA pak dosahuje zemědělská půda a to 82 %. Naopak mokřadní oblasti (Inland wetlands) model nerozpoznal vůbec.

Class/Value	C_11	C_12	C_13	C_14	C_21	C_22	C_23	C_24	C_31	C_32	C_41	C_51	Total	U_Accuracy	Kappa
C_11 - Urban fabric	228	46	7	5	109	0	3	42	53	3	0	0	496	0.460	0
C_12 - Industrial, commercial and transport units	71	49	6	9	78	0	2	32	40	2	2	15	306	0.160	0
C_13 - Mine, dump and construction sites	3	0	0	0	5	0	0	2	1	0	0	0	11	0.000	0
C_14 - Artificial, non-agricultural areas	47	3	0	3	23	0	5	22	20	1	0	1	125	0.034	0
C_21 - Arable land	128	27	13	13	2162	16	49	121	87	9	0	0	2625	0.824	0
C_22 - Permanent crops	16	3	2	1	970	34	57	134	321	10	0	3	1551	0.022	0
C_23 - Pastures	4	0	0	6	388	7	131	33	15	1	0	0	585	0.224	0
C_24 - Heterogenous agricultural areas	57	3	0	9	116	1	12	54	49	0	0	1	302	0.179	0
C_31 - Forests	107	16	0	22	358	5	65	279	1903	60	8	0	2823	0.674	0
C_32 - Shrub and/or herbaceous vegetation associations	25	2	0	3	588	5	83	104	259	18	0	1	1068	0.017	0
C_41 - Inland wetlands	0	1	0	1	1	0	1	0	2	0	0	0	6	0.000	0
C_51 - Inland waters	1	1	1	1	17	0	1	11	11	0	0	45	90	0.511	0
Total	607	151	29	73	4615	68	409	834	2761	104	10	67	10008	0.000	0
P_Accuracy	0.332	0.325	0.000	0.041	0.449	0.500	0.320	0.065	0.689	0.173	0.000	0.687	0	0.46	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0.32	0

Tabulka 4 - Matice záměn pro mnou vytvořený model po reklasifikaci.

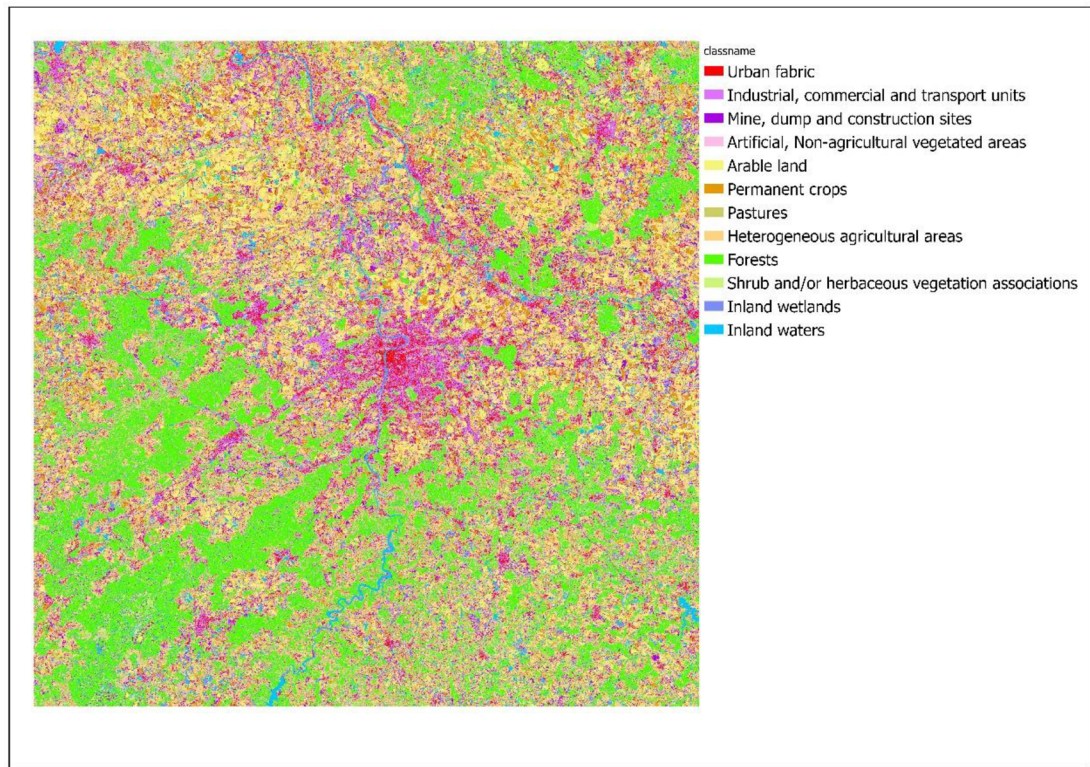


Obrázek 13 - Výsledek klasifikace pomocí mnou vytvořeného modelu po reklasifikaci na 2. úroveň CLC

Po stejné reklasifikaci v případě výsledku algoritmu SVM (viz Obr. 14) rovněž celková přesnost klasifikace vzrostla, ovšem pouze o 8 %, z 21 % na 29 % (viz tab. 5). Rozdíly v těchto výsledcích jsou na první pohled patrné, klasifikátor SVM například oproti mnou vytvořenému modelu často zaměňuje městskou zástavbu za industriální. Nejvyšších úrovní PA pak dosahuje třída mokřadů a to 80 %, naopak UA pro tuto třídu je pouhých 0,014 %. SVM totiž zaměňuje tyto oblasti převážně se zemědělskou půdou. Nejvyšších hodnot UA pak dosahuje zemědělská půda a to 83 %.

Class\Value	C_11	C_12	C_13	C_14	C_21	C_22	C_23	C_24	C_31	C_32	C_41	C_51	Total	U_Accuracy	Kappa
C_11 - Urban fabric	146	18	4	6	107	4	8	30	16	2	0	0	349	0.418	0
C_12 - Industrial, commercial and transport units	152	62	3	7	272	2	17	53	41	4	0	5	618	0.100	0
C_13 - Mine, dump and construction sites	79	19	4	5	488	5	18	44	74	2	0	3	741	0.005	0
C_14 - Artificial, non-agricultural areas	66	10	3	16	352	6	43	123	309	8	0	2	938	0.017	0
C_21 - Arable land	32	11	6	1	843	5	13	28	43	5	0	1	988	0.853	0
C_22 - Permanent crops	71	9	2	7	542	17	24	61	70	4	0	0	807	0.021	0
C_23 - Pastures	18	1	0	5	375	4	113	78	45	4	0	0	643	0.176	0
C_24 - Heterogenous agricultural areas	76	7	1	9	1059	15	62	137	192	15	0	4	1577	0.087	0
C_31 - Forests	8	5	0	7	227	0	39	117	1449	28	1	3	1884	0.769	0
C_32 - Shrub and/or herbaceous vegetation associations	19	5	2	4	231	9	53	68	284	22	1	1	699	0.031	0
C_41 - Inland wetlands	17	4	0	6	265	1	16	66	190	8	8	3	584	0.014	0
C_51 - Inland waters	3	0	4	0	54	0	3	21	48	2	0	45	180	0.250	0
Total	687	151	29	73	4815	68	409	834	2781	104	10	67	10008	0.000	0
P_Accuracy	0.213	0.411	0.130	0.219	0.175	0.250	0.276	0.164	0.525	0.212	0.800	0.672	0	0.29	0
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18

Tabulka 5- Matice záměn pro klasifikaci klasifikátorem SVM po reklasifikaci na 2. úroveň CLC



Obrázek 14 - Výsledek klasifikace klasifikátorem SVM, po reklasifikaci na 2. úroveň CLC.

7 Diskuse

Přestože metody klasifikace krajinného pokryvu s využitím technologie hlubokého učení mohou potenciálně poskytovat z pohledu běžného uživatele přesnější výsledky než doposud používané algoritmy (Kwan et al., n.d.). Naopak z hlediska uživatelské přívětivosti má aplikace těchto metod stále celou řadu překážek a dílčích problémů, které musí uživatel často řešit. Při zpracovávání této diplomové práce jsem narážel na nespočet překážek jak při používání předtrénovaného modelu, tak při samotné tvorbě modelu vlastního. Výsledky popsané v této práci jsou finálním výsledkem dlouhodobého snažení, které vyžadovalo neustálé přehodnocování používaných pracovních postupů, řešení problémů způsobených ne vždy kompletní a roztržitou dokumentací a nemalého množství nezdařených pokusů. Software ArcGIS Pro je jakožto nástupce softwaru ArcMap relativně nový a při práci s ním, je třeba počítat s problémy. Tyto problémy se projevovaly hlavně nestabilitou některých používaných nástrojů. Při mnohých pokusech jsem se setkal s neočekávanými pády aplikace. V případě učení modelu, které trvá samo o sobě několik hodin jsou takové situace velmi nepříjemné a způsobují výrazné zdržení.

Pro první pokusy o vytvoření vlastního modelu, jsem volil trénovací data, která neměla atributy identické s daty, která v ArcGIS Pro vytváří nástroj *Training sample manager*. Dle nápovědy programu se nemělo jednat o problém a nástroj použití takovýchto dat umožňuje bez problémů aplikovat. Ve většině případů však učení modelu vždy selhalo. Zde bych chtěl zdůraznit, že protokol o průběhu úlohy v softwaru ArcGIS Pro v těchto případech běžnému uživateli poskytuje velmi málo informací o důvodech selhání nástroje, v mnohých případech mi vypsal pouze informaci, že došlo k neočekávané chybě.

Po těchto nezdařených pokusech jsem vygeneroval trénovací data pomocí nástroje *Training sample manager*, s použitím výchozího klasifikačního schématu. Chtěl jsem ověřit, zda je chyba ve vstupních datech, či v samotném procesu učení. Učení i klasifikace poté proběhly bez problémů, ovšem s nepoužitelným výsledkem. Z tohoto výsledku jsem usoudil, že chyba samotného učení byla původně způsobena vstupními daty a v tomto směru jsem pátral dále. Vzhledem k tomu, že vytvářet nové klasifikační schéma dle CLC, by bylo časově náročné, modifikoval jsem atributy těchto dat, aby odpovídaly atributům z *Training samples manager*. Tento postup se nakonec ukázal jako správný, ovšem i zde jsem se setkal s pády aplikace.

Tyto problémy by byly řešitelné lepší dokumentací a odebráním možnosti takto nekompatibilní data pro vstup do nástroje použít. Vzhledem k tomu, že ArcGIS Pro je na trhu relativně krátkou dobu, jsem si jist, že na řešení společnost Esri pracuje. Přes všechny tyto počáteční problémy vidím v technologii hlubokého učení pro klasifikaci

krajinného pokryvu potenciál, zvláště pro práci s velkými objemy dat, časové analýzy s podrobným časovým rozlišením apod.

7.1 Jednoduchost a uživatelská přívětivost

Pro zodpovězení otázky, zda je možné jednoduše vytvořit vlastní model pro klasifikaci krajinného pokryvu si musíme nejdříve definovat, co považujeme za “jednoduché”. Pro tuto definici považuji za dobrý příklad ostatní klasifikační nástroje ArcGIS Pro, které ačkoliv nejsou vždy naprosto spolehlivé nebo stabilní, poskytují relativně uživatelsky přívětivé rozhraní a uživatele provedou celým postupem. Naproti tomu nástroje hlubokého učení vyžadují o něco vyšší úroveň znalostí uživatele, hlavně v oblasti datových formátů a jejich atributů, které do daného nástroje mohou vstupovat. V kombinaci s nedokonalou dokumentací je tato skutečnost ještě zvýrazněna. Rovněž musíme brát v úvahu časovou náročnost veškeré práce s daty. Zde jsem v průběhu práce zaznamenal výrazné rozdíly. Při učení modelu hlubokého učení pomocí velmi jednoduchých trénovacích dat, trval proces několik hodin a výsledný model nepřinesl použitelné výsledky. Naproti tomu algoritmus SVM, který jsem používal pro porovnání finálních výsledků, poskytl během pouhých několika desítek vteřin výsledek na první pohled relativně kvalitní. Vzhledem k zanedbatelné přesnosti výsledku obou modelů, jsem v daném případě statistické porovnání klasifikací neprováděl. V případě obsáhlejších trénovacích dat, jaké byly použity v případě finálního výsledku už ale model hlubokého učení poskytl výsledky statisticky kvalitnější než model SVM, samotné učení a klasifikace přitom trvaly pouze o několik hodin déle, zhruba 14 hodin v případě modelu hlubokého učení a 10 hodin v případě algoritmu SVM. Na základě tohoto poznatku lze soudit, že pro velké objemy dat, jsou modely hlubokého učení efektivnější, a tedy v případě klasifikování rozsáhlých území se jejich nevýhody stávají méně významnými.

7.2 Porovnání výsledků

Přesnosti klasifikace pomocí mnou vytvořeného modelu a modelu předtrénovaného, poskytovaného společností Esri jsou výrazně odlišné. Tyto rozdíly je možné přikládat velkému množství faktorů, například objemu trénovacích dat. Pro tuto práci jsem používal trénovací data vytvořená na základě jednoho snímku Sentinel-2, objem trénovacích dat použitých společností Esri není uveden. Rovněž mohou být rozdíly způsobeny jinou volbou páteřního modelu pro učení, já jsem byl při svém výběru omezován výkonem svého zařízení a časem potřebným k výpočtům, v případě, že by tato omezení neplatila, dá se předpokládat výsledek výrazně lepší. Je také potřeba brát v úvahu, že data CLC nejsou vytvářena pouze automatizovanými procesy, ale jsou rovněž manuálně revidována, tyto vstupy nemusí umělá inteligence zatím efektivně nahradit. Například efektivně rozlišit sportovní areály od pastvin, zemědělské půdy či průmyslových hal vyžaduje znalost zkoumaného území spíše než

pouhé porovnání spektrálních příznaků a kontextu okolních pixelů. Už v rámci trénovacích dat tedy můžeme mluvit o zanesení jisté chyby. Co se týče okolních sledovaných pixelů, i to je parametr, který je možno uživatelsky nastavit, a i zde může být důvod výrazného rozdílu mezi výsledkem mého modelu a modelem předučeného. Toto velké množství proměnných by pro nalezení optimálního nastavení a získání výsledku s velmi vysokou přesností vyžadovalo rozsáhlé experimentování, velký výpočetní výkon a samozřejmě i dostatek času. Vzhledem k počátečním nezdarům při mých pokusech bohužel čas na opakované učení modelu s použitím různých kombinací nastavení již nebyl dostatečný.

Rousset et al., (2021) ve svém postupu využívá augmentace dat pomocí nově vytvořených pásem, např. s indexem NDVI, pro lepší odlišení některých tříd. Další rozdíl je v počtu pásem, Rousset et al. (2021) pracoval s daty družice SPOT6, v pásmech RGB+NIR, zde poskytují mnou využívaná data větší množství informací. Ne všechny jsou však užitečné. Roli hraje rovněž prostorové rozlišení dat. Senzor družice SPOT6 má například rozlišení 6 m v pásmech RGB+NIR a 1,5 m v pásmu panchromatickém. Po provedení pansharpeningu je tedy možné získat rozlišení až 1,5 m ve všech pásmech. V této studii pak při použití pouze pásem RGB+NIR dosahuje nejvyšší celkové přesnosti 61 %, při použití i dalších nově vytvořených pásem pak přesnost byla vyšší jen v jednotkách desetin procent.

Kwan et al. (n.d.) ve své studii experimentuje s klasifikací hyperspektrálních snímků pomocí Convolved Neural Networks (CNN) algoritmů. V této studii se zabývá různými kombinacemi pásem a také jejich augmentací za pomoci Multi-attributes profilů (EMAP), a dat z laserového skenování. V této studii pak dosáhl přesnosti až 87 % při použití 44 pásem vzniklých z augmentace pásem RGB+NIR pomocí EMAP. Klasifikátor SVM pak na stejných datech vykázal přesnost 81 %.

Mnou vytvořený model na první pohled nedosahuje tak dobrých výsledků. Faktorů, které to mohou způsobovat je nespočet. Zmiňované studie používají jiné třídy klasifikace, a tedy nelze přesně srovnávat. Rovněž to může být způsobeno jinými vstupními daty a do jisté míry i augmentacemi, které byly součástí těchto studií. Jak jsem již zmiňoval výše tato problematika je značně rozsáhlá a dosažení optimálních výsledků vyžaduje experimentování s různými kombinacemi všech parametrů a do jisté míry i dávkou štěstí.

8 Závěr a přínos práce

Klasifikace krajinného pokryvu pomocí metod hlubokého učení má využití hlavně při zpracování velkého objemu dat. Jedná se o komplexní problematiku, kde hraje roli velké množství proměnných a je ovlivněna kvalitou a objemem trénovacích dat. Software ArcGIS Pro od společnosti ESRI poskytuje efektivní nástroje pro vytvoření vlastních modelů klasifikace. Navíc tzv. Esri Living Atlas poskytuje různé předtrénované modely pro hluboké učení, které poskytují relativně vysoké přesnosti klasifikací, a navíc díky svému jednoduchému využití minimalizují chyby na straně uživatele. Nejvyšší celkové přesnosti klasifikace dosáhl v rámci této diplomové práce předtrénovaný model od společnosti ESRI, jehož celková přesnost byla 74 %. Vlastní model hlubokého učení vytvořený v rámci této práce pomocí data Corine Land Cover pak dosáhl přesnosti 46 %. Kontrolní klasifikace pomocí algoritmu SVM využívající totožná trénovací data pak dosáhla přesnosti pouhých 28 %.

Diplomová práce splňuje v zadání stanovené cíle a její výsledky odpovídají na předem stanovené výzkumné otázky, tedy že:

- (a) Pro klasifikace krajinného pokryvu na základě dat Sentinel-2 je možné využít předtrénované modely a tyto modely poskytují konzistentní a relativně kvalitní výsledky s minimální uživatelskou náročností na provedení těchto klasifikací.
- (b) V programu ArcGIS pro je možné vytvořit vlastní klasifikační model. Tato činnost však vyžaduje pokročilé znalosti na straně uživatele, velké množství trénovacích vzorků a výsledky klasifikace nedosahovaly v případě tohoto modelu kvalit modelu předtrénovaného, ovšem vykazují vyšší přesnost než v současné době často používaný klasifikátor SVM.
- (c) Volně dostupný předtrénovaný model dosahuje vyšší přesnosti klasifikace než model vytvořený uživatelem a také klasifikátor SVM.

9 Přehled literatury a použitých zdrojů

About | *Landsat Science*. (n.d.). Retrieved March 25, 2024, from <https://landsat.gsfc.nasa.gov/about/>

Ahmad, A., & Quegan, S. (2013). Comparative Analysis of Supervised and Unsupervised Classification on Multispectral Data. *Applied Mathematical Sciences*, 7(74), 3681–3694. <https://doi.org/10.12988/ams.2013.34214>

Alem, A., & Kumar, S. (2020). Deep Learning Methods for Land Cover and Land Use Classification in Remote Sensing: A Review. *ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, 903–908. <https://doi.org/10.1109/ICRITO48877.2020.9197824>

Blundell, J. S., & Opitz, D. W. (n.d.). *OBJECT RECOGNITION AND FEATURE EXTRACTION FROM IMAGERY: THE FEATURE ANALYST® APPROACH*.

Bojinski, S., Schaepman, M., Apfer, D. S., & Itten, K. (2003). SPECCHIO: a spectrum database for remote sensing applications. *Computers & Geosciences*, 29, 27–38.

Cieślak, I., Szuniewicz, K., Pawlewicz, K., & Czyza, S. (2017). Land Use Changes Monitoring with CORINE Land Cover Data. *IOP Conference Series: Materials Science and Engineering*, 245(5). <https://doi.org/10.1088/1757-899X/245/5/052049>

Comber, A., Fisher, P., & Wadsworth, R. (2005). What is land cover? *Environment and Planning B: Planning and Design*, 32(2), 199–209. <https://doi.org/10.1068/B31135>

CORINE Land Cover — Copernicus Land Monitoring Service. (n.d.-a). Retrieved March 25, 2024, from <https://land.copernicus.eu/en/products/corine-land-cover>

CORINE Land Cover — Copernicus Land Monitoring Service. (n.d.-b). Retrieved March 27, 2024, from <https://land.copernicus.eu/en/products/corine-land-cover>

De Jong, S. M., Hornstra, T., & Maas, H. G. (2001). An integrated spatial and spectral approach to the classification of Mediterranean land cover types: the SSC method. *International Journal of Applied Earth Observation and Geoinformation*, 3(2), 176–183. [https://doi.org/10.1016/S0303-2434\(01\)85009-1](https://doi.org/10.1016/S0303-2434(01)85009-1)

- Deep learning in ArcGIS Pro—ArcGIS Pro | Documentation.* (n.d.). Retrieved March 27, 2024, from <https://pro.arcgis.com/en/pro-app/latest/help/analysis/deep-learning/deep-learning-in-arcgis-pro.htm>
- Deep learning model architectures—ArcGIS Pro | Documentation.* (n.d.). Retrieved March 27, 2024, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/overview-of-the-deep-learning-models.htm>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/J.RSE.2011.11.026>
- Dwivedi, R. S., Kandrika, S., & Ramana, K. V. (2004). Comparison of classifiers of remote-sensing data for land-use/land-cover mapping. *Current Science*, 86(2), 328–335. <http://www.jstor.org/stable/24107878>
- Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (n.d.). *remote sensing Analysis Ready Data: Enabling Analysis of the Landsat Archive*. <https://doi.org/10.3390/rs10091363>
- Enhanced Thematic Mapper Plus (ETM+) | Landsat Science.* (n.d.). Retrieved March 25, 2024, from <https://landsat.gsfc.nasa.gov/etm-plus/>
- Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Remote Sensing*, 18(3), 679–685. <https://doi.org/10.1080/014311697219015>
- Forshaw, M. R. B., Haskell, A., Miller, P. F., Stanley, D. J., & Townshend, J. R. G. (1983). Spatial resolution of remotely sensed imagery A review paper. *International Journal of Remote Sensing*, 4(3), 497–520. <https://doi.org/10.1080/01431168308948568>
- Grekousis, G., Mountrakis, G., & Kavouras, M. (2015). An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, 36(21), 5309–5335. <https://doi.org/10.1080/01431161.2015.1093195>
- History | Landsat Science.* (n.d.). Retrieved March 25, 2024, from <https://landsat.gsfc.nasa.gov/satellites/history/>
- Inglada, J., Arias, M., Tardy, B., Morin, D., Valero, S., Hagolle, O., Dedieu, G., Sepulcre, G., Bontemps, S., & Defourny, P. (2015). Benchmarking of algorithms for crop type land-cover maps using Sentinel-2 image time series.

International Geoscience and Remote Sensing Symposium (IGARSS), 2015-November, 3993–3996. <https://doi.org/10.1109/IGARSS.2015.7326700>

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). GLOBAL LAND USE/LAND COVER WITH SENTINEL 2 AND DEEP LEARNING. *International Geoscience and Remote Sensing Symposium (IGARSS), 2021-July*, 4704–4707. <https://doi.org/10.1109/IGARSS47720.2021.9553499>

Kwan, C., Ayhan, B., Budavari, B., Lu, Y., Perez, D., Li, J., Bernabe, S., & Plaza, A. (n.d.). *remote sensing Deep Learning for Land Cover Classification Using Only a Few Bands*. <https://doi.org/10.3390/rs12122000>

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., Coomes, O. T., Dirzo, R., Unther Fischer, G. U., Folke, C., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., ... Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11, 261–269.

land cover — *European Environment Agency*. (n.d.). Retrieved March 25, 2024, from <https://www.eea.europa.eu/help/glossary/eea-glossary/land-cover>

Land Cover Classification (Sentinel-2) - Přehled. (n.d.). Retrieved March 27, 2024, from <https://www.arcgis.com/home/item.html?id=afd124844ba84da69c2c533d4af10a58>

Martimort, P., Arino, O., Berger, M., Biasutti, R., Carnicero, B., Del Bello, U., Fernandez, V., Gascon, F., Greco, B., Silvestrin, P., Spoto, F., & Sy, O. (2007). Sentinel-2 optical high resolution mission for GMES operational services. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2677–2680. <https://doi.org/10.1109/IGARSS.2007.4423394>

Miranda, E., Mutiara, A. B., Ernastuti, & Wibowo, W. C. (2018). Classification of Land Cover from Sentinel-2 Imagery Using Supervised Classification Technique (Preliminary Study). *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, 69–74. <https://doi.org/10.1109/ICIMTECH.2018.8528122>

Moss, D., & Wyatt, B. K. (1994). The CORINE biotopes project: a database for conservation of nature and wildlife in the European community. *Applied Geography*, 14(4), 327–349. [https://doi.org/10.1016/0143-6228\(94\)90026-4](https://doi.org/10.1016/0143-6228(94)90026-4)

- Multispectral Scanner (MSS) | Landsat Science.* (n.d.). Retrieved March 25, 2024, from <https://landsat.gsfc.nasa.gov/multispectral-scanner/>
- Noi, P. T., & Kappas, M. (n.d.). *Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery.* <https://doi.org/10.3390/s18010018>
- O'ahel', J. (2000). *Corine Land Cover-Technical Guide.* www.eea.eu.int
- Paek, S. W., & Kim, S. (2018). Space-based Earth remote sensing: Part 1. Satellite orbit theory. *Satellite Oceanography and Meteorology*, 3(1). <https://doi.org/10.18063/SOM.V3I1.646>
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (n.d.). *remote sensing Sentinel-2 Data for Land Cover/Use Mapping: A Review.* 12, 2291. <https://doi.org/10.3390/rs12142291>
- Rousset, G., Despinoy, M., Schindler, K., Mangeas, M., & Assessment, M. M. (2021). *Assessment of Deep Learning Techniques for Land Use Land Cover Classification in Southern New Caledonia remote sensing Assessment of Deep Learning Techniques for Land Use Land Cover Classification in Southern New Caledonia.* 13(12). <https://doi.org/10.3390/rs13122257i>
- Saha, S., Solano-Correa, Y. T., Bovolo, F., & Bruzzone, L. (2019). Unsupervised deep learning based change detection in Sentinel-2 images. *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2019.* <https://doi.org/10.1109/MULTI-TEMP.2019.8866899>
- Sci-Hub | CORINE Land Cover and Land Cover Change Products. Remote Sensing and Digital Image Processing, 55–74 | 10.1007/978-94-007-7969-3_5.* (n.d.). Retrieved March 25, 2024, from https://sci-hub.se/10.1007/978-94-007-7969-3_5
- Sentinel-2 - Mission Objectives - Sentinel Online - Sentinel Online.* (n.d.). Retrieved March 25, 2024, from <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/mission-objectives>
- Sentinel-2 - Missions - Instrument Payload - Sentinel Handbook - Sentinel Online.* (n.d.). Retrieved March 25, 2024, from <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/instrument-payload>

- Sentinel-2 - Missions - Sentinel Online - Sentinel Online.* (n.d.). Retrieved March 25, 2024, from <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2>
- Sentinel-2 - Satellite Description - Sentinel Online - Sentinel Online.* (n.d.). Retrieved March 25, 2024, from <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/satellite-description>
- Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.-A., & Rahman, A. (n.d.). *remote sensing Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations-A Review.* 12, 1135. <https://doi.org/10.3390/rs12071135>
- The difference between “land use” and “land cover” - MSU Extension.* (n.d.). Retrieved March 25, 2024, from https://www.canr.msu.edu/news/the_difference_between_land_use_and_land_cover
- Thematic Mapper (TM) | Landsat Science.* (n.d.). Retrieved March 25, 2024, from <https://landsat.gsfc.nasa.gov/thematic-mapper/>
- Thomson, A. G., Fuller, R. M., & Eastwood, J. A. (1998). Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing. *International Journal of Remote Sensing*, 19(17), 3423–3431. <https://doi.org/10.1080/014311698214091>
- Topaloğlu, R. H., & Sertel, E. (n.d.). *ASSESSMENT OF CLASSIFICATION ACCURACIES OF SENTINEL-2 AND LANDSAT-8 DATA FOR LAND COVER / USE MAPPING.* <https://doi.org/10.5194/isprsarchives-XLI-B8-1055-2016>
- Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X. P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *International Journal of Digital Earth*, 5(5), 373–397. <https://doi.org/10.1080/17538947.2012.713190>
- Vali, A., Comai, S., & Matteucci, M. (2020). *Remote Sens.* 12. <https://doi.org/10.3390/rs12152495>
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311–332. [https://doi.org/10.1016/0034-4257\(87\)90015-0](https://doi.org/10.1016/0034-4257(87)90015-0)

- XI, Y., Thinh, N. X., & LI, C. (2019). Preliminary comparative assessment of various spectral indices for built-up land derived from Landsat-8 OLI and Sentinel-2A MSI imageries. *European Journal of Remote Sensing*, 52(1), 240–252. <https://doi.org/10.1080/22797254.2019.1584737>
- Yılmaz, R., & Yılmaz, R. (2010). Monitoring land use/land cover changes using CORINE land cover data: a case study of Silivri coastal zone in Metropolitan Istanbul. *Environ Monit Assess*, 165, 603–615. <https://doi.org/10.1007/s10661-009-0972-z>
- Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J. F., Lymburner, L., Pahlevan, N., & Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224, 382–385. <https://doi.org/10.1016/J.RSE.2019.02.016>
- 임용호조원호 *, Jo, W., Lim, Y., & Key-Ho Park, . (n.d.). *합성곱 신경망을 이용한 딥러닝 기반의 토지피복 분류: 한국 토지피복을 대상으로 Deep learning based Land Cover Classification Using Convolutional Neural Network: a case study of Korea*. Retrieved March 27, 2024, from <https://egis.me.go.kr/intro/land.do>

10 Seznam obrázků a tabulek

Obrázek 1 - Časová osa programu Landsat (Wulder et al., 2019)	10
Obrázek 2- Skladba pásem družic Sentine-2 v porovnání s ostatními družicemi používanými pro sledování zemského povrchu (Drusch et al., 2012).	11
Obrázek 3 - Spektrální profily základních druhů krajinného pokryvu dle dat Landsat 8 a Sentinel-2 (XI et al., 2019).	14
Obrázek 4 - Umístění zájmového území v rámci ČR.	16
Obrázek 5 - Pořadí pásem používané pro kompozitní rastry.	19
Obrázek 6 - Nastavení nástroje Export Training Data For Deep Learning.	21
Obrázek 7 - Nastavení nástroje Train Deep Learning Model.	22
Obrázek 8 - Nastavení nástroje Classify Pixels Using Deep Learning.	23
Obrázek 9 - Nastavení nástroje pro klasifikaci snímku klasifikátorem SVM.	23
Obrázek 10 - Výsledek klasifikace předtrénovaným modelem.	25
Obrázek 11 - Výsledek klasifikace modelem U-net na základě trénovacích vzorků generovaných z produktu CLC.	27
Obrázek 12 - Výsledek klasifikace klasifikátorem SVM.	28
Obrázek 13 - Výsledek klasifikace pomocí mnou vytvořeného modelu po reklasifikaci na 2. úroveň CLC	29
Obrázek 14 - Výsledek klasifikace klasifikátorem SVM, po reklasifikaci na 2. úroveň CLC.	30
Tabulka 1 - Dělení tříd Land Cover/Land Use dle CORINE Land Cover (Bütner, 2014).	7
Tabulka 2 - Tabulka klasifikačních tříd a deklarované přesnosti pro model Land Cover Classification (Sentinel-2) (Land Cover Classification (Sentinel-2) - Přehled, n.d.).	18
Tabulka 3 - Matice záměn předtrénovaného modelu.	26
Tabulka 4 - Matice záměn pro mnou vytvořený model po reklasifikaci.	28
Tabulka 5- Matice záměň pro klasifikaci klasifikátorem SVM po reklasifikaci na 2. úroveň CLC	29

11 Přílohy

Priloha č. 1 – Matice záměn pro klasifikaci mnou vytvořeným modelem dle 3. úrovně CLC

Class/Value	C_111	C_112	C_121	C_122	C_123	C_124	C_131	C_132	C_133	C_141	C_142	C_211	C_221	C_222	C_231	C_242	C_243	C_310	C_311	C_313	C_321	C_324	C_411	C_412	C_511	C_512	Total	U_Accuracy	Kappa		
C_111 - Continuous urban fabric	53	615	415	37	1	5	44	3	5	7	9	172	0	1	11	4	32	5	4	0	2	3	0	0	0	0	1425	0.0371	0		
C_112 - Discontinuous urban fabric	1	1591	110	16	0	3	11	3	1	37	24	916	1	5	51	52	266	140	44	122	6	20	0	0	0	0	3421	0.4651	0		
C_121 - Industrial or commercial units	5	432	213	23	0	5	12	3	5	24	34	330	0	3	19	20	131	49	8	54	3	18	0	1	2	10	1404	0.1517	0		
C_122 - Road and rail networks and associated land	1	103	20	5	0	3	2	2	1	2	3	59	0	1	5	0	22	3	1	2	2	2	0	2	0	0	325	0.0153	0		
C_123 - Port areas	2	102	132	11	4	10	19	5	1	11	33	375	1	2	20	4	199	63	70	62	1	12	5	0	54	69	1267	0.0032	0		
C_124 - Airports	0	10	6	0	0	1	1	0	0	0	0	13	0	0	0	1	4	1	1	1	1	1	0	0	0	0	46	0.0208	0		
C_131 - Mineral extraction sites	0	7	4	0	0	0	0	0	0	0	0	4	0	0	0	0	3	1	0	0	0	3	0	0	0	0	23	0.0000	0		
C_132 - Dump sites	0	13	3	1	0	0	3	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.0000	0		
C_133 - Construction sites	0	426	3	0	0	0	0	0	0	14	17	247	0	2	19	27	154	102	14	70	2	9	0	0	0	0	67	0.0000	0		
C_141 - Green urban areas	0	31	6	0	0	0	0	0	0	1	1	34	0	0	1	0	20	9	2	15	1	1	0	0	0	0	1105	0.0126	0		
C_142 - Sport and leisure facilities	0	1280	143	30	0	92	74	22	10	5	73	2182	11	159	523	143	1060	327	68	204	13	54	0	0	0	2	15	25573	0.8314	0	
C_211 - Non-irrigated arable land	0	14	2	0	0	2	0	0	0	0	0	1	375	2	1	54	3	41	7	0	6	1	2	0	0	0	511	0.0039	0		
C_222 - Fruit trees and berry plantations	0	135	14	2	0	0	13	5	9	1	5	19	9152	6	257	691	63	1133	590	424	506	16	80	0	0	11	4	13230	0.0194	0	
C_231 - Pastures	0	43	5	0	0	13	1	1	0	0	32	4083	13	81	1259	18	370	74	14	41	15	13	0	0	2	0	6075	0.2071	0		
C_243 - Land principally occupied by agriculture	0	672	21	2	0	1	7	2	0	14	36	973	2	7	111	36	374	131	23	84	2	15	0	0	0	0	2513	0.0143	0		
C_310 - Coniferous forest	1	269	39	5	0	1	9	2	1	37	47	964	0	14	86	34	562	2996	112	940	11	111	1	2	2	12	6255	0.1952	0		
C_311 - Broad-leaved forest	0	18	1	0	0	0	0	0	0	2	1	443	0	3	71	6	132	42	71	90	1	16	0	0	2	1	900	0.0769	0		
C_313 - Mixed forest	0	662	45	4	0	3	7	1	0	128	84	2363	0	42	365	94	1925	9257	1811	6703	73	361	3	2	0	6	23969	0.2797	0		
C_321 - Natural grassland	0	219	3	2	0	0	4	1	0	1	10	5481	3	56	662	50	857	754	307	507	56	76	0	1	1	1	9085	0.0065	0		
C_324 - Transitional woodland-shrub	0	28	9	1	0	0	2	0	0	1	3	452	0	2	37	9	189	104	92	156	0	21	1	0	5	2	1114	0.0189	0		
C_411 - Inland marshes	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4	0.0000	0		
C_412 - Peat bogs	0	1	1	0	0	0	0	0	0	1	2	9	0	0	0	0	0	5	2	1	4	0	0	0	0	0	27	0.0000	0		
C_511 - Water courses	0	7	10	1	5	1	7	1	0	2	8	101	0	3	9	4	104	7	18	6	0	1	0	0	0	173	655	0.2641	0		
C_512 - Water bodies	0	6	2	0	0	0	2	0	1	0	1	63	0	1	3	0	45	10	4	5	0	1	0	0	2	101	247	0.4989	0		
Total	66	6802	1210	141	10	153	211	55	27	294	440	48151	41	641	4095	573	7765	14732	3126	9754	211	824	10	10	259	411	100012	0	0		
P_Accuracy	0.8030	0.2339	0.1760	0.0355	0.4000	0.0065	0.0000	0.0000	0.0000	0.0476	0.0023	0.4416	0.0023	0.4009	0.3074	0.0208	0.0156	0.2034	0.0227	0.6872	0.6872	0.2796	0.0255	0.0000	0.0000	0.6680	0.2457	0.0000	0.364	0	
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2229	0

Priloha č. 2 – Matice záměn pro klasifikaci pomocí SVM

Class/Value	C_111	C_112	C_121	C_122	C_123	C_124	C_131	C_132	C_133	C_141	C_142	C_211	C_221	C_222	C_231	C_242	C_243	C_310	C_311	C_313	C_321	C_324	C_411	C_412	C_511	C_512	Total	U_Accuracy	Kapita	
C_111 - Continuous urban fabric	7	8	4	0	1	0	0	0	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	24	0.2917	0	
C_112 - Discontinuous urban fabric	1	140	9	2	0	0	0	0	0	0	1	110	1	6	5	4	38	6	2	7	0	0	0	0	0	0	333	0.4204	0	
C_121 - Industrial or commercial units	0	59	25	1	1	1	4	1	1	2	1	74	0	7	2	20	6	2	9	0	1	0	0	0	0	0	217	0.1152	0	
C_122 - Road and rail networks and associated land	2	55	13	3	2	0	2	1	1	3	0	131	0	0	1	14	7	1	6	0	0	0	0	0	0	1	244	0.0123	0	
C_123 - Port areas	0	10	13	2	3	0	0	1	0	0	0	8	0	1	2	0	4	1	0	0	0	2	0	2	2	2	51	0.0588	0	
C_124 - Airports	0	18	8	1	0	4	1	0	0	0	3	69	0	1	11	1	11	0	1	3	0	1	0	0	0	0	133	0.0301	0	
C_131 - Mineral extraction sites	0	13	5	2	1	0	1	0	0	1	0	172	0	0	5	0	16	11	5	3	0	3	0	0	0	2	240	0.0042	0	
C_132 - Dump sites	0	26	3	0	0	1	2	0	0	0	118	0	5	11	3	10	17	0	4	0	2	0	0	0	0	2	204	0.0098	0	
C_133 - Construction sites	0	29	8	0	0	2	4	0	3	0	4	141	0	1	6	0	8	3	1	4	1	2	0	0	0	0	217	0.0138	0	
C_141 - Green urban areas	0	20	6	1	0	0	0	1	7	5	157	0	3	15	6	61	72	33	94	1	10	1	0	1	0	0	494	0.0142	0	
C_142 - Sport and leisure facilities	0	46	5	0	0	2	2	0	0	1	4	191	1	3	35	4	45	45	10	30	0	3	0	0	0	0	427	0.0094	0	
C_211 - Non-irrigated arable land	0	33	7	1	0	0	3	0	1	0	1	833	0	3	21	2	24	10	2	6	0	1	0	0	0	1	949	0.8778	0	
C_221 - Vineyards	0	35	4	0	0	0	1	0	1	1	1	167	5	5	13	8	20	3	2	4	0	3	1	0	0	0	274	0.0192	0	
C_222 - Fruit trees and berry plantations	0	37	0	0	0	0	0	1	1	0	0	375	0	17	13	1	28	16	3	18	1	5	0	0	0	0	516	0.0329	0	
C_231 - Pastures	0	9	1	0	0	2	0	0	0	8	391	0	2	113	1	81	15	3	12	2	1	0	0	1	0	0	642	0.1760	0	
C_242 - Complex cultivation patterns	0	48	5	1	0	0	1	0	0	3	2	540	1	4	23	10	29	12	5	9	2	2	0	0	0	0	697	0.0143	0	
C_243 - Land principally occupied by agriculture	0	46	2	0	0	0	0	1	0	1	4	435	0	4	40	4	86	52	12	51	1	4	0	0	2	0	745	0.1154	0	
C_310 - Coniferous forest	0	4	0	0	0	0	0	0	0	1	0	54	0	0	3	1	33	461	9	121	2	8	1	0	0	0	688	0.6605	0	
C_311 - Broad-leaved forest	0	5	0	0	0	1	0	0	1	3	2	159	0	0	18	2	51	118	95	196	0	6	0	0	1	0	657	0.1446	0	
C_313 - Mixed forest	0	2	1	0	0	0	0	0	0	1	0	50	0	0	9	0	35	339	74	241	1	6	0	0	0	1	760	0.3171	0	
C_321 - Natural grassland	0	6	0	0	0	2	0	1	0	1	1	140	0	3	28	1	35	13	2	7	6	3	0	0	0	1	250	0.0240	0	
C_324 - Transitional woodland-shrub	0	6	0	0	0	0	0	0	0	0	4	137	2	3	14	2	51	184	27	87	3	12	0	0	0	0	532	0.0226	0	
C_411 - Inland marshes	0	6	0	0	0	1	0	0	0	0	0	39	0	0	6	0	13	35	7	24	1	4	6	0	0	1	137	0.0438	0	
C_412 - Peat bogs	0	18	1	0	0	0	1	0	2	1	263	0	3	7	1	41	26	12	23	0	1	0	10	0	0	0	410	0.0244	0	
C_511 - Water courses	0	7	1	0	2	0	1	1	0	1	14	0	0	1	0	13	5	3	6	0	2	0	0	12	2	72	0.1667	0		
C_512 - Water bodies	0	0	0	0	0	0	0	0	0	0	1	46	0	0	2	2	10	16	2	10	0	0	1	0	6	28	124	0.2258	0	
Total	10	680	121	14	10	15	21	10	10	29	44	4815	10	64	409	57	777	1473	313	975	21	82	10	10	26	41	10047	0.0000	0	
P_Accuracy	0.7000	0.2059	0.2066	0.2143	0.3000	0.2667	0.0476	0.2000	0.3000	0.2414	0.0909	0.1730	0.5000	0.2556	0.2763	0.1754	0.1107	0.3130	0.3035	0.2472	0.2857	0.1463	0.6000	1.0000	0.4615	0.6829	0.0000	0.2124	0	
Kapita	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1462	0