



1949

**GÉPI ÉS MÉLY TANULÁSI ELJÁRÁSOK ALKALMAZÁSA A VÁROSI
KÖRNYEZET VIZSGÁLATÁBAN, NAGYFELBONTÁSÚ ÉS KÜLÖNBÖZŐ
DIMENZIÓJÚ TÁVÉRZÉKELT ADATOK ALAPJÁN**

Egyetemi doktori (PhD) értekezés

Abriha Dávid

Témavezető:
Prof. Dr. Szabó Szilárd

DEBRECENI EGYETEM
Természettudományi és Informatikai Doktori Tanács
Földtudományok Doktori Iskola
Debrecen, 2023

Ezen értekezést a Debreceni Egyetem Természettudományi és Informatikai Doktori Tanács Földtudományok Doktori Iskola „A lito- és hidroszféra természetes és antropogén folyamatai” programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából.

Nyilatkozom arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.

Debrecen, 2023. 05. 16.

*Abriha Dávid
a jelölt aláírása*

Tanúsítom, hogy Abriha Dávid doktorjelölt 2019-2023 között a fent megnevezett Doktori Iskola „A lito- és hidroszféra természetes és antropogén folyamatai” programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Nyilatkozom továbbá arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.

Az értekezés elfogadását javasolom.

Debrecen, 2023. 05. 16.

*Prof. Dr. Szabó Szilárd
a témavezető aláírása*

**GÉPI ÉS MÉLY TANULÁSI ELJÁRÁSOK ALKALMAZÁSA A VÁROSI
KÖRNYEZET VIZSGÁLATÁBAN, NAGYFELBONTÁSÚ ÉS KÜLÖNBÖZŐ
DIMENZIÓJÚ TÁVÉRZÉKELT ADATOK ALAPJÁN**

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a Földtudományok tudományágban

Írta: **Abriha Dávid** okleveles geográfus

Készült a Debreceni Egyetem Földtudományok Doktori Iskolája (A lito- és
hidroszféra természetes és antropogén folyamatai programja) keretében

Témavezető: **Prof. Dr. Szabó Szilárd**, egyetemi tanár

Az értekezés bírálói:

Dr.

Dr.

A bírálóbizottság:

elnök: Dr.

tagok: Dr.

Dr.

Dr.

Dr.

Az értekezés védésének időpontja: 2023.

Tartalomjegyzék

Rövidítések jegyzéke	9
1. Bevezetés, célkitűzések	10
2. Szakirodalmi áttekintés	12
2.1. A távérzékelés fejlődése	12
2.2. A távérzékelte adatok alkalmazási lehetőségei	12
2.3. Képosztályozás tetőfedő anyagok alapján	14
2.4. Modellek pontosságának értékelése a képosztályozásban	15
2.4.1. Területi autokorreláció, mint a modellek pontosságát torzító tényező....	17
2.5. Mély tanulás a távérzékelte képfeldolgozásban	18
2.6. A U-Net architektúra áttekintése	19
2.7. Épületek szegmentációja távérzékelte adatok alapján	20
3. Anyag és módszer	22
3.1. A mintaterületek bemutatása	22
3.2. A felhasznált távérzékelte adatok	24
3.2.1. Műholdas adatbázisok	24
3.2.2. Légi felvételezéssel nyert adatbázisok	25
3.3. Képosztályozás	25
3.3.1. Referenciaadatok gyűjtése	25
3.3.2. Osztályozási módszerek alkalmazása	26
3.3.3. Modellvalidáció	27
3.3.4. Objektum alapú pixel homogenitás	28
3.4. A területi autokorreláció hatásának elemzése	30
3.4.1. Vizsgálatok különböző mintavételezési módszerekkel	30
3.5. Épület szegmentálás mély konvolúciós neurális hálózat alapján	34
3.5.1. Az input adatok előállítása, adat augmentáció	35
3.5.2. Modell implementáció és értékelés	37
4. Eredmények és értékelésük	42

4.1.	Tetőfedő anyagok osztályozhatóságának az elemzése	42
4.1.1.	<i>A pansharpening, valamint a kategóriák számának hatása a tematikus pontosságra.....</i>	<i>42</i>
4.1.2.	<i>A területi autokorreláció hatásának vizsgálata.....</i>	<i>45</i>
4.2.	A U-Net mély konvolúciós neurális hálózat értékelése épület szegmentáción keresztül	52
4.2.1.	<i>Az egyes modellek pontosságának értékelése ugyanazon a felvételen történő predikció során</i>	<i>53</i>
4.2.2.	<i>Az ortofotó szegmentálása a WV2 és a WV3 tanító adatai alapján</i>	<i>56</i>
4.2.3.	<i>A WV2 szegmentálása a WV3 és az ortofotó tanító adatai alapján</i>	<i>56</i>
4.2.4.	<i>A WV3 szegmentálása a WV2 és az ortofotó tanító adatai alapján</i>	<i>57</i>
4.2.5.	<i>A WV3 szegmentálása a WV2, az ortofotó és a WV3 tanító adatai alapján</i>	<i>58</i>
5.	Összefoglalás.....	64
6.	Summary.....	67
	Köszönetnyilvánítás	70
	Irodalomjegyzék	71

Rövidítések jegyzéke

CNN	konvolúciós neurális hálózat (<i>Convolutional Neural Network</i>)
DL	mély tanulás (<i>deep learning</i>)
FCN	<i>Fully Convolutional Network</i>
FT	független tesztelés
GAN	<i>Generative Adversarial Network</i>
ISPRS	<i>International Society for Photogrammetry and Remote Sensing</i>
KCV	k-szoros keresztvalidáció (<i>k-fold cross validation</i>)
LDFA	lineáris diszkriminancia-analízis (<i>Linear Discriminant Function Analysis</i>)
MI	Moran-féle I (<i>Moran's I</i>)
NDFM	normalizált digitális felszínmodell
NDVI	<i>Normalized Difference Vegetation Index</i>
OA	általános pontosság (<i>Overall Accuracy</i>)
OPP	objektum alapú pixel homogenitás (<i>object-based pixel purity</i>)
QDFA	kvadratikus diszkriminancia-analízis (<i>Quadratic Discriminant Function Analysis</i>)
RF	<i>Random Forest</i>
RGB	<i>red-green-blue</i>
RKCV	ismételt k-szoros keresztvalidáció (<i>repeated k-fold cross validation</i>)
SA	területi autokorreláció (<i>spatial autocorrelation</i>)
SCV	rétegzett k-szoros keresztvalidáció (<i>stratified k-fold cross validation</i>)
SCV2	SCV 2 m-es leválogatással
SCV10	SCV 10m-es leválogatással
SCVCent	SCV centroidokkal
SCVSeg	SCV szegmensekkel
SWIR	<i>short-wave infrared</i>
TP	tényleges pontosság
TT	tanító és tesztelő területek
UAS	pilóta nélküli légijármű-rendszerek (<i>Unmanned Aerial Systems</i>)
ViT	<i>Vision Transformer</i>
VNIR	<i>visible and near-infrared</i>
WV	<i>WorldView</i>
WV2	<i>WorldView-2</i>
WV3	<i>WorldView-3</i>

1. Bevezetés, célkitűzések

Az elmúlt évtizedekben növekvő tendenciát mutató urbanizáció hatására a természetes környezet egyre jelentősebb mértékben alakul át. Az ENSZ által 2018-ban kiadott, a világ urbanizációs folyamatait elemző jelentés szerint a városi területeken élők aránya 2050-re közel 70%-ra fog növekedni, emellett a globális népességnövekedés és a megavárosok (10 millió főt meghaladó lakosságszám) gyarapodó száma egyre nagyobb nyomást gyakorolnak a Föld ökoszisztémáira (UN, 2019). Az ilyen léptékű változások számos gazdasági, társadalmi és környezeti kihívásokat generálnak (Beaverstock et al., 1999; Riffat et al., 2016; Taubenböck et al., 2009), amelyeknek a kezelése egyre több városban kap kiemelt prioritást, és ennek eredményeként a fenntarthatóság és a fenntartható urbanizáció lett a városfejlesztési stratégiák egyik legfontosabb alapelve (Kadhim et al., 2016; Roy, 2009; Shen et al., 2012).

Bár a városi régiók rendkívül összetettek, a távérzékelés széles körben kínál lehetőségeket a döntéshozóknak ezen kihívások leküzdéséhez. Doktori értekezésem fókuszában az épített környezet, azon belül is az épületek nagyfelbontású és különböző dimenziójú távérzékelte adatokkal végzett elemzése állt, elsődlegesen Debrecen példáján. Kutatásom során kiemelt szerepet kapott a tradicionális *machine learning* (gépi tanulás) módszerekkel végzett képosztályzás, illetve az ehhez szorosan kapcsolódó pontosságellenőrzés, valamint napjaink legkorszerűbb technológiájával, *deep learning* (mély tanulás) alapú megközelítéssel végzett épületszegmentáció is.

Munkám során a céloom elsőként különböző tetőfedő anyagok osztályozhatóságának a vizsgálata volt, eltérő körülmények között. Ennek során egy kiválasztott mintaterületen a három leggyakrabban előforduló tetőfedő típust vontam be az osztályozásba, melyet WorldView-2 műholdfelvétel alapján végeztem el. A gépi és mély tanulási algoritmusok egyre népszerűbbé válásával a modellezés mind inkább gyakoribb feladattá vált, melynek elengedhetetlen részét képezi a pontosságvizsgálat, így következő lépésként megvizsgáltam, hogy az osztályozások során a területi autokorreláció milyen hatást gyakorol a modellek pontosságára. A tetőfedő anyagok osztályozása előtt célszerű egy olyan maszk előállítását és a felvételen történő alkalmazását, amely csak ezeket a célobjektumokat tartalmazza, ezzel csökkentve a szükséges feldolgozási kapacitást. Bár az épületek kinyerésére számos megoldás létezik, ezek sokszor igen drága adatokon (pl. *Light*

Detection and Ranging; LiDAR) alapulnak. A kutatásom során szintén célom volt megvizsgálni, hogy a relatívan olcsón előállítható vagy beszerezhető felvételeken végzett, modern, mély konvolúciós neurális hálózatokon alapuló objektum detektálás milyen pontossággal alkalmazható épületek kinyerésére, különös tekintettel kevés, több forrásból származó adaton történő tanítás során, melyhez a U-Net architektúrát alkalmaztam.

Kutatásaim során az alábbi kérdésekre kerestem a választ:

1. Hogyan befolyásolja a WorldView-2 műholdfelvételen végzett tetőfedő anyagok osztályozását a térbeli felbontást javító *pansharpening* eljárás, valamint milyen hatást gyakorol a tematikus pontosságra az egyes tetőfelületek napsütötte és árnyékos kategóriákra történő szétválogatása?
2. Milyen mértékben járul hozzá a magas pozitív területi autokorreláció az osztályozási pontosság túlbecsléséhez, továbbá hogyan lehet kiküszöbölni a területi autokorreláció torzító hatását a k-szoros keresztvalidációval történő pontosságbecslés megbízhatóságának biztosítása érdekében?
3. Milyen hatékonysággal alkalmazható a U-Net konvolúciós neurális hálózat különböző forrásból származó multispektrális távérzékelte adatokon alapuló épületszegmentáció esetében, ha csak kevés tanító adat áll rendelkezésre?

2. Szakirodalmi áttekintés

2.1. A távérzékelés fejlődése

Napjainkra az élet számos területén váltak megkerülhetetlenné a különböző távérzékeléshez kapcsolódó alkalmazások. A technológia egyik legfőbb előnye az adatgyűjtés forradalmasítása: segítségével lehetőség nyílt nagy területekről hatalmas mennyiségű adat gyűjtésére, viszonylag rövid idő alatt. Emellett a távérzékelésnek köszönhetően elérhetővé vált az adatgyűjtés olyan térségekből is, amelyeket a földről nem, vagy csak nagyon nehezen lehet megközelíteni, ilyenek például a magashegységi régiók, a kiterjedt, sűrű esőerdők, vagy a katasztrófa sújtotta területek.

A távérzékelés korai céljai leginkább a katonai igényekhez igazodtak, és a felvételezéshez szükséges kamerákat repülőgépekre szerelték (Moore 1979). Az 1960-as évektől a CORONA programmal (Dashora et al., 2007) kezdődően megjelentek a műholdak, mint új hordozó platformok, azonban az igazi áttörést a Landsat program (Zhu et al., 2019) jelentette. A repülőgépes adatgyűjtéssel ellentétben a műholdak ugyanazon terület rendszeres, újbóli felvételezésével jelentősen egyszerűbbé tették az idősoros elemzések elvégzését (Lulla et al., 2021; Shalaby és Tateishi, 2007; Wulder et al., 2019). A relatívan olcsó, vagy ingyenes képeket szolgáltató műholdak száma a 2000-es évek elejétől kezdve ugrásszerűen megnőtt, ami fokozta ezen felvételek közcélú és tudományos felhasználását. A műholdas távérzékelés mellett a légi felvételezés is nagymértékben fejlődött. A repülőgépek korai fekete-fehér analóg kameráit mára felváltották a digitális multispektrális, hiperspektrális berendezések. A távérzékelésben a legújabb, igen dinamikusan fejlődő szegmensnek, a pilóta nélküli légijármű-rendszerek (*Unmanned Aerial Systems*, UAS) tekinthetők (Szabó et al., 2018). Ezek az eszközök elsősorban viszonylag olcsó áruk és sokoldalú felhasználhatóságuk (pl. több szenzor együttes használata; Noor et al., 2018; Kucharczyk és Hugenholtz, 2021) miatt váltak népszerűvé. A távérzékelte felvételek fejlődése során nem csak a spektrális, hanem a térbeli felbontás is javult: a műholdak esetében mára elérhetővé váltak a 30-50 cm-es felbontással rendelkező felvételek, míg a légi térképezés esetében néhány cm-ről beszélhetünk.

2.2. A távérzékelte adatok alkalmazási lehetőségei

A távérzékelte adatok alkalmazhatósági területeit alapvetően azok tulajdonságai határozzák meg, melyek közül a két legalapvetőbb a felvételek térbeli és spektrális felbontása. A nagy térbeli felbontású adatok részletesebb információt szolgáltatnak a felszín jellegzetességeiről, objektumairól, ezért azok felhasználása lokálisabb: Phinzi és munkatársai (2020) nagy felbontású műholdfelvétel (SPOT-7) alapján térképeztek árkos eróziós területeket, Bertalan és munkatársai (2022) pedig UAS alapú adatgyűjtéssel végeztek talajnedvességre vonatkozó előrejelzéseket. Ezzel szemben az alacsonyabb térbeli felbontású adatok jellemzően nagyobb területeket fednek le, de ezekre vonatkozóan kevesebb részletet tartalmaznak. Az ilyen felvételek elsősorban regionális és globális elemzésekhez hasznosak: Friedl és munkatársai (2002) globális felszínborítási katalógust készítettek 1 km felbontású MODIS adatok alapján.

A spektrális felbontás a távérzékelte adatok esetében a felvételek spektrális sávjainak mennyiségére, azok spektrumon belüli elhelyezkedésére és az általuk lefedett hullámhossz tartományra utal. Nagyobb spektrális felbontás mellett az egyes anyagok nagyobb pontossággal különíthetők el egymástól, ugyanis a különböző vizsgált anyagok egyedi spektrális jellemzői jobban szétválnak: Burai és munkatársai (2015) légi hiperspektrális felvétel alapján (128 spektrális csatorna) nagy pontossággal tudtak elkülöníteni 20 lágyszárú növényzeti kategóriát. Alacsonyabb spektrális felbontás (multispektrális felvételek) esetében az egyes csatornák szélesebb hullámhossz tartományt fednek le, ezáltal egy adott területegységről egységnyi idő alatt több foton jut el az érzékelőbe, így magasabb térbeli felbontást érhetünk el. Azoknál az alkalmazásoknál, ahol fontosabb az objektumszintű, minél részletesebb feltérképezés, a multispektrális felvételek kerülhetnek előtérbe.

A távérzékeléssel nyert adatok egyik legjelentősebb alkalmazási területének az épített környezet vizsgálata tekinthető. A különböző típusú szenzorokkal (LiDAR, radar, multispektrális, hiperspektrális stb.) gyűjtött adatok széleskörűen alkalmazhatók a városi területek tanulmányozására. Mucsi és munkatársai (2017) 8 különböző, városi felszínborítási kategóriát vizsgáltak, és megállapították, hogy a magas spektrális felbontású hiperspektrális (359 csatorna, 400-2500 nm), valamint a nagy térbeli felbontással rendelkező multispektrális (<1 m) felvételek fúziója magasabb

osztályozási pontosságokhoz vezethet. Schlosser és munkatársai (2020) kutatásukban az épület szegmentáció hatékonyságát vizsgálták ortofotók, valamint légifelvételekből származtatott pontfelhő alapján. Megállapították, hogy a szimpla, RGB (*red-green-blue*) csatornák alapján végzett szegmentáció pontossága (80%) jelentősen javítható egyéb változók (spektrális indexek, morфомetriai mutatók stb.) bevonásával (>95%).

Összességében elmondható, hogy a legfontosabb alkalmazási, kutatási területeknek a földhasználat/felszínborítás térképezése, a változásvizsgálat, a káresemények detektálása, valamint az egyes objektumok azonosítása és kinyerése tekinthetők (Goldblatt et al., 2018; Szabó et al., 2016). A távérzékelte adatokból nyert eredményeket a mezőgazdaságban, a természetvédelemben, valamint a táj- és várostervezésben is széles körben hasznosítják (Díaz-Delgado et al., 2017; Kennedy et al., 2009; Weiss et al., 2020; Weng, 2012).

2.3. Képosztályozás tetőfedő anyagok alapján

A képosztályozás során az egyes szenzorok által rögzített felvételek pixeleit különböző kategóriákba soroljuk (Al-doski et al., 2013). Attól függően, hogy az osztályba sorolást a pixel mely tulajdonságai alapján végezzük el, beszélhetünk pixel-, és objektum-alapú képosztályozásról. Előbbi esetben a pixelek egyedül a spektrális jellemzői alapján kerülnek kategorizálásra, míg utóbbinál a felvételt első lépésként meghatározott homogenitási kritériumok alapján objektumokra bontjuk, majd ezekhez az objektumokhoz rendeljük hozzá az egyes osztályokat (Zhou és Troy, 2008).

A különböző tetőfedő anyagok osztályozása a távérzékelés egyik fontos kutatási területévé vált, ugyanis az épített környezetre vonatkozó pontos és aktuális információk meglétére egyre nagyobb igény mutatkozik. Korábbi tanulmányok mind hiperspektrális (Chisense 2012; Fiumi et al., 2012; Szabó et al., 2014), mind pedig multispektrális (Raczko et al., 2022; Taherzadeh és Shafri, 2013) adatforrásokat nagy pontossággal alkalmaztak tetőfedő anyagok azonosítására.

Kutatómunkám során pixel-alapú képosztályozási módszerekkel végeztem el városi területen a leggyakrabban előforduló tetőfedő anyagok osztályozását. A tetők, mint célobjektumok osztályozása egyedi kihívást jelentenek, ugyanis referenciaadat gyűjtésére relatívan kis felületek állnak rendelkezésünkre, amelyek megjelenését számos tényező teszi igen

változatossá: a tetősíkok különböző aspektusai (napsütés hatása), az anyagok öregedése, a tetőablakok, kémények, valamint akár moha és zuzmó jelenléte. A téma több szempontból is releváns: az anyagok lehetnek gyúlékonyak (fa, szalma), vagy egészségügyi szempontból jelenthetnek kockázati tényezőt, mint például az azbeszt-cement tartalmú tetőfedők (Cili et al., 2015; Gibril et al., 2017; Szabó et al., 2014). Az azbeszt rostok az anyag fizikai és kémiai mállásának következtében a levegőbe, onnan pedig a tüdőbe jutva olyan súlyos megbetegedéseket okozhatnak, mint az azbesztózis, mezotelióma, vagy éppen a tüdőrák (Kottek és Yuen, 2022; Mándi et al., 2000). Bár hazánkban 2005 óta tilos az azbeszt építőipari felhasználása, jó hőszigetelő tulajdonsága és alacsony ára miatt korábban nagyon népszerű választás volt, így egyes becslések alapján Magyarországon az azbeszttartalmú tetőfelületek nagyjából 100-150 millió négyzetmétert tesznek ki. Az önkormányzatoknak a közeljövőben szembe kell nézniük az azbeszt okozta környezeti problémákkal, ugyanis ezen tetők maximális élettartama átlagosan – környezeti viszonyoktól függően – 60-80 év. Probléma azonban, hogy még települési szinten sem léteznek átfogó tetőkataszteri nyilvántartások, mely hiányosságokra kínál elérhető alternatívát a megfelelő pontosságellenőrzésnek alávetett képosztályozás.

2.4. Modellek pontosságának értékelése a képosztályozásban

A modellezés gyakori feladatként jelenik meg az adattudomány (*data science*) gyakorlati és tudományos szegmensében egyaránt, amely a gépi és mély tanulási algoritmusok révén vált népszerűvé. A modellek célja, hogy segítsenek megérteni a környezetet, annak jellemzőit, folyamatait, változásait és a változások következményeit. Számos tudományterületen, így a földtudományokban is igen széles körben alkalmaznak modelleket bizonyos célobjektumok, például felszínborítás, tetőfedő anyagok, fafajok stb. azonosítására (Abdollahi et al., 2022; Likó et al., 2022; Mohajane et al., 2021; Raczkó et al., 2022). A modellek pontosságának mérése azok hatékonyságát mutatja, ami korlátozhatja az adott módszer alkalmazhatóságát (Varga et al., 2021). A pontosságvizsgálat a képosztályozási munkafolyamat elengedhetetlen része (Congalton, 2001; Foody, 2002; Lunetta és Lyon, 2004), elvégzésére többféle megközelítés létezik.

(i) Az egyik legkorábbi – mindmáig széleskörben alkalmazott – gyakorlat a referenciaadatok tanító- és tesztterületekké történő felosztása.

Ennek során általában a referenciaadatok 50-80%-át használják a modell tanítására, a maradékot pedig az eredmény tesztelésére. Bár ez a módszer értékes információkat szolgáltat a pontosságról, a felosztás esetleges, mivel a referenciaadatok tanító- és tesztterületekké történő elkülönítése véletlenszerű kiválasztáson alapul. Következésképpen, egy újabb véletlenszerű felosztás másképp betanított modellt, ezáltal más eredményt adhat: ha egy adott adathalmazzal osztályozást végzünk, a végeredmény nagymértékben függ a tanító adatok megbízhatóságától; azaz, a nem reprezentatív pixelek (amelyek az adott osztályt képviselik, de amelyek értékei a begyűjtött referenciaadatok halmazában kiugró értéknek számítanak) relevánsan torzíthatják az osztályozás eredményét, és a pontossági mutatók (általános pontosság, felhasználói pontosság, előállítói pontosság; Kappa index) csak ennek az egyetlen modellnek a pontosságát fogják tükrözni (Congalton, 1991). Amennyiben feltételezzük, hogy a tanító adatok inkonzisztensek lehetnek, és egyetlen modell csak egy lehetséges eredményt tükröz, a véletlenszerű kiválasztás többször is megismételhető, és több modell futtatásával megítélhetjük, hogy az osztályozott térkép véletlenszerűen keletkezett-e, vagy a statisztikai paramétereket figyelembe véve megbízhatónak számít.

(ii) Egy lehetséges megoldás az előbbieken ismertetett problémára az egyre népszerűbbé váló k -szoros keresztvalidáció (*k-fold cross validation*, KCV) (James et al., 2021; Lyons et al., 2018). Ennek során az adatokat k egyenlő méretű részhalmazra (általában 5 vagy 10) osztjuk fel, és ezekből a részhalmazokból $k-1$ részt használunk a modell tanítására, a fennmaradó részt pedig a tesztelésre (öt egyenlő részhalmazt feltételezve az arány 4:1). Az eljárást k -szor megismételjük, és addig tart, ameddig minden egyedi részhalmazt egyszer fel nem használtunk a teszteléshez. A modellekhez tartozó pontosságokat k -szor számoljuk ki, amelyből különböző statisztikai mérőszámok (átlag, medián, szórás, konfidenciaintervallum vagy kvartilisek) számolhatók (Yadav és Shukla, 2016).

(iii) A megközelítés továbbfejlesztése az ismételt KCV: a teljes eljárást valahányszor (általában 3-5 alkalommal), további véletlenszerűen kiválasztott mintával megismételjük (azaz az eredeti referenciaadatokból teljesen új adatkészleteket generálunk). Ezt a megközelítést ismételt KCV-nek (*repeated KCV*, RKCV; Brownlee, 2016; Phinzi et al., 2021a) nevezzük, és ezáltal az olyan statisztikai mérőszámok, mint az általános pontosság átlaga, mediánja és szórása 30-50 modell alapján kerülnek meghatározásra. Bár az RKCV

széleskörben ismert, a távérzékelésben viszonylag új technikának számít (Phinzi et al., 2021b), épp ezért a különböző szoftverekben való implementálása korlátozott (pl. KCV-re egyedül az EnMap-Box 3, Python alapú QGIS bővítményben van lehetőség, ismétlések nélkül).

2.4.1. *Területi autokorreláció, mint a modellek pontosságát torzító tényező*

A pontosságértékelés legfontosabb eleme maga a referencia adathalmaz: az általános követelmények (reprezentativitás, megfelelő adatmennyiség stb.) mellett a tanító és teszt adatoknak egymástól függetleneknek kell lenniük. A nagy térbeli felbontás azonban jelentős területi autokorrelációhoz (*spatial autocorrelation*, SA) vezet (Woodcock és Strahler, 1987): az egymás melletti pixelek hasonlóak lesznek, pl. míg egy ház teteje 10 m-es felbontás mellett 1 pixelnyi, addig 0.1 m esetében 10000 pixelnyi területet fed le. Az SA koncepciója összefüggésben van Tobler, amerikai geográfus úgynevezett „első földrajzi törvényével”, mely szerint az egyes objektumok közötti hasonlóság a távolság csökkenésével nő (Miller, 2004; Mucsi, 2006); vagyis a közeli, szomszédos objektumok nem függetlenek, hanem általában hatással vannak egymásra.

Az SA különösen gyakori a távérzékelte felvételek esetében (Strahler és Woodcock, 1986), ahol a határos képpontok általában kapcsolatban állnak egymással. A modell kiértékelése, amennyiben a teszteléshez használt adathalmaz nem független a tanító adathalmaztól, az osztályozási pontosságok túlbecsléséhez vezet, mivel az egymást követő vagy egymás melletti adatsorokból hiányzik a véletlenszerűség, azaz az egyik adat értéke függ a másik értékétől, ezért a referenciaadatokra különös figyelmet kell fordítani a hamisan jó eredmények elkerülése érdekében. Amennyiben a referenciaadatokat a műholdképek vagy légifotók alapján poligonokként határoljuk le, a véletlenszerű kiválasztás ugyanazon poligon – akár egymás szomszédságában lévő – képpontjait választhatja ki mind a tanításhoz, mind a teszteléshez, amely megsérti az adatok függetlenségének elvét (Carranza-García et al., 2019; Drzewiecki et al., 2016). Míg az egyszerűbb pontosságértékelési esetekben a tanító és tesztelő adatok függetlensége könnyen biztosítható egy-egy különálló adathalmaz előállításával, vagy poligonok helyett pontok használatával, addig az RKCV, mint modern pontosságértékelő technika torzítani fog az eredményeken. Tekintettel arra, hogy ebben az esetben nincs külön tanító és tesztelő adathalmaz, az eljárás a

teljes referencia adathalmazból történő véletlenszerű kiválasztáson alapul, ezért az adatok függetlenségét más módon kell biztosítani.

Számos tanulmány vizsgálta az RKCV alkalmazhatóságát erősen autokorrerált adatokon, többek között a hiperparaméterek finomhangolása, a talajkárosodás, vagy az alluviális területek ásványi potenciáljának feltérképezése kapcsán (Ibrahim és Bennett, 2014; Mannel et al., 2011; Pohjankukka et al., 2017; Schratz et al., 2019), és arra a következtetésre jutottak, hogy az autokorreláció jelentősen torzítja a pontossági mérőszámokat, ezért a tanító és teszt adatokat térben el kell különíteni a tematikus pontosság túlbecslésének elkerülése érdekében. Az RKCV esetében, annak működési elve miatt azonban a teljesen független tesztelés nem lehetséges, így ahhoz, hogy mint megbízható pontosságbecslő módszert alkalmazhassuk, az SA-t közvetlenül a referenciaadatokban kell megszüntetni.

2.5. Mély tanulás a távérzékelte képfeldolgozásban

A mély tanulási (*deep learning*, DL) módszerek egyre népszerűbbek a különböző távérzékelési alkalmazásokban. A rendelkezésre álló adatok minőségének és mennyiségének növekedése olyan DL alapú módszerek kifejlesztéséhez vezetett, amelyek a hagyományosnak tekinthető gépi tanulási eljárásoknál hatékonyabban képesek azonosítani és kinyerni a felvételek komplex jellemzőit, ami nagyobb pontosságot eredményez (He et al., 2018; Ma et al., 2019; Yuan et al., 2020; Zhang et al., 2016).

A számítási teljesítmény folyamatos növekedése megalapozta olyan modern módszerek létrejöttét, mint például a konvolúciós neurális hálózatok (*Convolutional Neural Network*, CNN). A távérzékelésben a CNN-eket sikeresen alkalmazták számos területen, például változásvizsgálatban (Mou et al., 2019; Wang et al., 2019), objektumdetektálásban (Hoeser et al., 2020; Lovász et al., 2023), szemantikus szegmentálásban (Kajári et al., 2023; Kampffmeyer et al., 2016; Panboonyuen et al., 2019; Yuan et al., 2021) vagy képjavításban (Hoque et al., 2019; Hu et al., 2021). Mindezek következtében rohamos léptékűvé vált a különböző, CNN alapú architektúrák megjelenése.

Mérföldkönek számít a Krizhevsky és munkatársai (2012) által bemutatott AlexNet, amely már képes volt kihasználni a számítógépek grafikus feldolgozó egységét (GPU), így biztosítva hatalmas mennyiségű adat viszonylag gyors feldolgozását. Egy másik nagyon jelentős mérföldkö a Long

és munkatársai (2015) által megalkotott, úgynevezett *Fully Convolutional Networks* (FCNs) volt. Az FCN architektúra minden egyes pixel klasszifikációjához egy kódoló-dekódoló struktúrát használ, és mivel nem tartalmaz teljesen összekötött (*fully connected*) rétegeket, a bemeneti kép tetszőleges méretű lehet. A U-Net létrejötté Ronneberger és munkatársai (2015) munkájához kötődik, akik a fejlesztéshez az FCN architektúrára támaszkodtak. A két architektúra közötti egyik legfontosabb különbség, hogy a U-Net struktúra összeköttetések kihagyása (*skip connection*) révén képes jobb eredmények elérésére, továbbá a módosítások miatt a U-Net kevesebb tanító adat mellett is képes viszonylag pontos szegmentációra (Bardis et al., 2020).

A mély neurális hálózatok esetében a modellek hajlamosak a túlillesztésre (*overfitting*), melynek számos oka lehet, azonban a probléma leggyakrabban a tanító adatokban rejlik (Antoniou et al., 2017; Srivastava et al., 2014). Ha a rendelkezésre álló adatmennyiség nem elégséges, nem reprezentatív vagy túl zajos, akkor a modell a tanító adatok egyedi sajátosságait is megtanulja, így bár ezek esetében jól működik, azonban a korábban nem látott tesztadatokra rosszul fog általánosítani (Brownlee, 2018; Rice et al., 2020; Ying, 2019). A modern CNN-ek optimális működésük eléréséhez hatalmas mennyiségű adatra támaszkodnak. Tekintettel azonban arra, hogy számos alkalmazási területen meglehetősen korlátozott referenciaadatok rendelkezésre állása, a túlillesztés elkerülése érdekében szükség van az adatok bővítésére (*data augmentation*). Az adathővítés legegyszerűbb esetben szimpla képtranszformációkon alapul, hagyományos módszerekkel: tükrözés, forgatás, levágás, eltolás stb. (Harangi, 2018; Shorten és Khoshgoftaar, 2019; Perez és Wang, 2017). Egy másik, egyre népszerűbb augmentációs technika a GAN-ok alkalmazása (*Generative Adversarial Networks*) (Goodfellow et al. 2014), amelyek a meglévő tanító adatok mesterséges másolatait hozzák létre, megtartva azok főbb jellemzőit (Antoniou et al., 2017; Chlap et al., 2021).

2.6. A U-Net architektúra áttekintése

Bár a U-Net architektúrát eredetileg orvosbiológiai kép szegmentációs feladatok elvégzésére tervezték, hatékonyságát azóta számos más területen igazolták, beleértve a távérzékelést.

A U-Net nevét az architektúra jellegzetes, „U” alakjáról kapta. Ez két fő ágból áll, melyeket kódoló (*encoder*) és dekódoló (*decoder*) szakasznak hívunk. A kódoló szakasz konvolúciós és úgynevezett *max-pooling* rétegekből áll. Ennek során a képek mérete csökken, miközben a dimenziók (sávok) száma nő. Ez azt eredményezi, hogy a modell rögzíti a képek fő jellemzőit, ugyanakkor a képek pozíciójára vonatkozó információk elvesznek. Mivel a szemantikus szegmentálás célja az eredeti kép összes pixelének valamilyen kategóriába sorolása, szükséges lesz a lokalizációra vonatkozó információk kinyerésére, valamint az eredeti képméret visszaállítására is egyaránt. Ezt a U-Net dekódoló szakasza biztosítja, amely transzponált konvolúciót használ az úgynevezett *upsampling*-hez (Abdollahi et al. 2020; Du et al. 2020; Fan et al. 2022; Ronneberger et al. 2015; Yan et al. 2022).

Korábbi tanulmányok már igazolták a U-Net architektúra hatékonyságát az épület detektáció területén. Pan és munkatársai (2020) nagy felbontású műholdfelvétel alapján végeztek épület detektálást, melynek során 86% feletti általános pontosságot (*overall accuracy*; OA) értek el. Hasonlóan jó eredményekről számotlak be Irwansyah és munkatársai (2020), akik légifelvétel alapján szegmentáltak sűrűn beépített városi területeken épületeket, és kaptak 92%-os pontosságot. Li és munkatársai (2019) tanulmányukban épületekre vonatkozó változásvizsgálatot végeztek Sentinel-1 radar adatok alapján, és értek el 73-92% közötti OA-kat. Ezek az eredmények azt sugallják, hogy a U-Net architektúra ígéretes módszer az épület detektációs feladatokban, és alkalmazása kiterjeszthető különböző távérzékelési adattípusokra magas pontossággal.

2.7. Épületek szegmentációja távérzékelte adatok alapján

A távérzékelés egyik leggyakoribb városi alkalmazási területe a különböző objektumok detektálásához köthető. Mind az államigazgatási szervek, mind pedig a magánszektor egyre nagyobb igényt támaszt a pontos és megbízható térbeli adatok iránt (Eslami és Mohammadzadeh, 2016; Rose et al., 2015). A városi objektumdetektálás célja az adatok minőségétől függően igen eltérő lehet: az UAS-ok vagy a térfigyelő kamerák által szolgáltatott nagyon nagy felbontású képek vagy videók felhasználhatók például a különböző közlekedési eszközök (autók, buszok, kerékpárok stb.) megkülönböztetésére (Leung et al., 2019; Peppas et al., 2018; Zhou et al., 2017), míg a durvább felbontású adatokat – leggyakrabban légi vagy műholdas

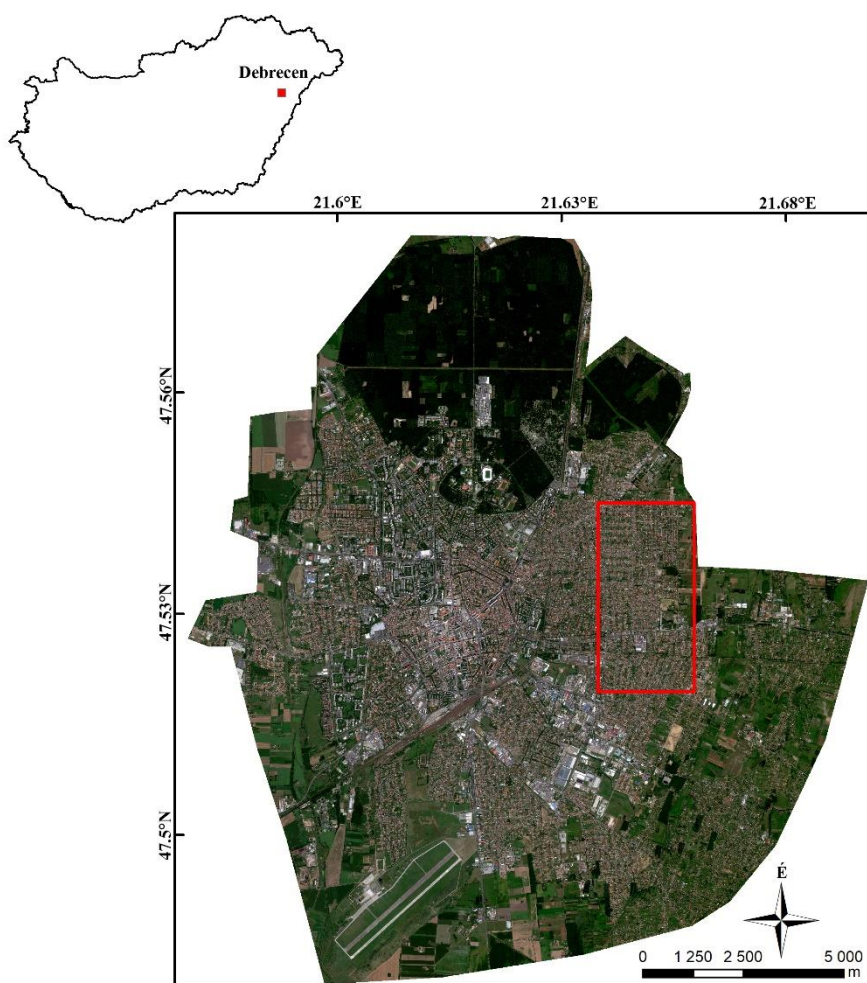
felvételeket – egész városok feltérképezésére használják különböző célokra: épületek, úthálózatok vagy épp növényzettel borított területek felmérésére (Fischer et al., 1998; Gavankar and Ghosh, 2018; Rottensteiner et al., 2014; Wang et al., 2016).

Az automatikus épületkinyerést számos tényező nehezíti. A városi felületek rendkívül heterogének, sok különböző anyag található egymás szomszédságában, így az épületek és más objektumok közötti kontraszt viszonylag alacsony (Jaynes et al., 2003; Xu et al., 2018). Ráadásul egy adott területen az épületek általában nagyon változatosak a tetőfedő anyagok, a formák, a tetők állapota és kora stb. tekintetében. Emellett a különböző objektumok által vetett árnyékok félrevezethetik a szegmentáló algoritmusokat, és torzíthatják az eredményt (Han et al., 2022; Jung et al., 2022; Schlosser et al., 2020; You et al., 2018). Ezáltal a hagyományos pixelalapú osztályozási technikák alkalmazása esetén sokszor tematikus pontossági problémákkal találkozhatunk.

3. Anyag és módszer

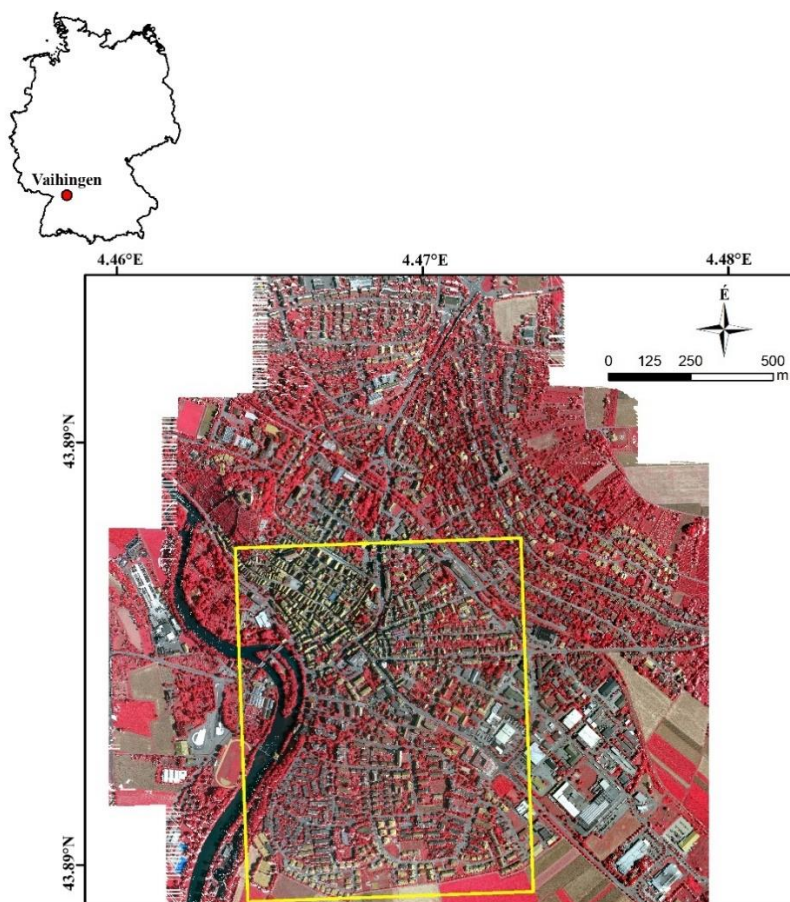
3.1. A mintaterületek bemutatása

Kutatásom során a vizsgálataim középpontjában Debrecen, Magyarország második legnépesebb települése (199.725 fő; 2022) állt. A városon belüli konkrét mintaterület lehatárolásánál fontos kritérium volt egyrészt a családi házas övezet, másrészt pedig az azbeszt-cement tartalmú tetőfedő anyagok nagyszámú jelenléte. Így esett a választás a keleti városrész egy 5 km² méretű területére (*1. ábra*), ahol túlnyomó többségben a 20. század második felében épített kertesházak találhatók, melyek fő tetőfedő anyagai az azbeszt és a különböző színű cserepek.



1. ábra. A debreceni mintaterület elhelyezkedése

A kutatás szempontjából fontosnak tartottam, hogy az SA-hoz kapcsolódó vizsgálatokat ne csak helyi, Debrecenhez kötődő adatokon végezzem el, ezért az *International Society for Photogrammetry and Remote Sensing* (ISPRS) nemzetközi szervezet által biztosított, széles körben használt referencia-adatbázist is felhasználtam, ezzel is rámutatva, hogy a felmerülő probléma független a lokális tényezőktől. Az így kiválasztott másik mintaterület a németországi Vaihingen an der Enz városában található (2. ábra). A település lakosságszámát tekintve jelentősen elmarad Debrecentől (29.696 fő; 2022). Bár itt is próbáltam a debreceni kritériumoknak megfelelően kijelölni a konkrét mintaterületet (családi- illetve sorházas területek), azonban a felvételen nem azonosítottam azbeszt-cement tartalmú tetőket. A cserepek mellett itt a leggyakoribb tetőfedő anyag még a bitumen volt.



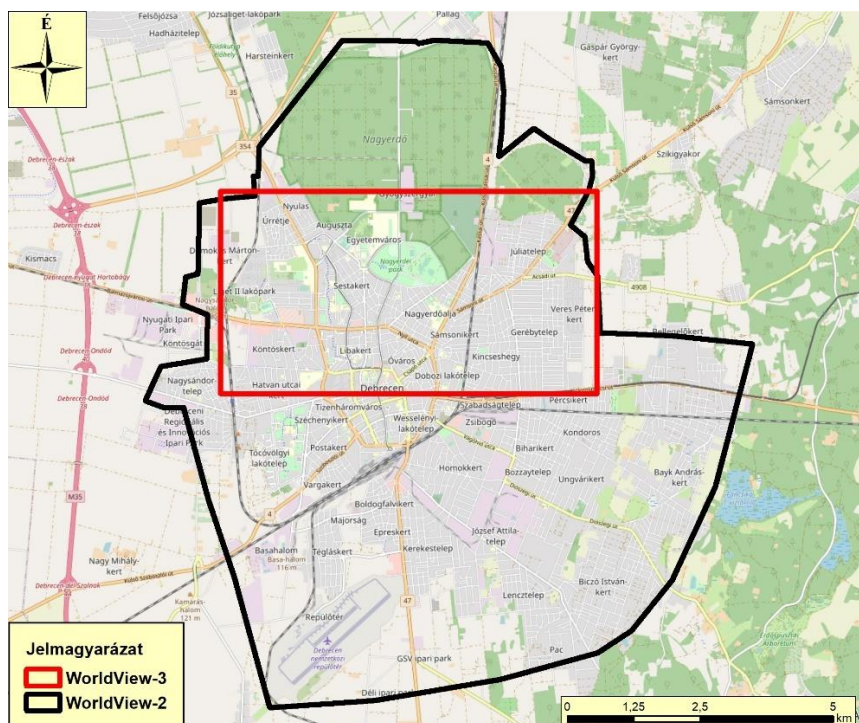
2. ábra. A vaihingeni mintaterület elhelyezkedése

3.2. A felhasznált távérzékelt adatok

3.2.1. Műholdas adatbázisok

Kutatásom során Debrecen területéről két különböző multispektrális műholdas távérzékelt adattal dolgoztam, egy WorldView-2 (WV2) és egy WorldView-3 (WV3) felvétellel. A WV2 felvétel 2016.07.24-én készült, 8 spektrális sávból áll, melyek térbeli felbontása 2 m, kiterjedését tekintve a teljes várost lefedi. Mivel a felvétel tartalmazott egy 350 nm széles, 0,5 m felbontású pankromatikus csatornát is, a Gram-Schmidt féle *pansharpening* módszert alkalmazva a multispektrális sávok térbeli felbontását feljavítottam (Maurer, 2013).

A WV3 felvétel 2019.09.16-án készült. Bár eredetileg a WV3 felvételek a WV2-höz képest további csatornákkal rendelkeznek (8 db SWIR, és 12 db CAVIS), az általam elérhető kép csak pankromatikus és VNIR sávokból állt. Bár ezek a sávok hasonló spektrális tulajdonságokkal rendelkeznek, mint a WV2, a pankromatikus csatorna térbeli felbontása jobb volt (0,3 m). Kiterjedését tekintve a WV3 elmarad a WV2-höz képest, csupán a város északi részét fedi le (3. ábra).



3. ábra. A WorldView-2 és WorldView-3 felvételek kiterjedése

3.2.2. Légi felvételezéssel nyert adatbázisok

A munkám során három multispektrális légifelvétellel dolgoztam, melyek közül egy Vaihingen, kettő pedig Debrecen területeiről készült. A vaihingeni felvétel az ISPRS által közzétett *WGII/4 2D* elnevezésű, szemantikus szegmentációs referencia-adatbázisból származó, nyilvánosan elérhető ortofotó. A felvétel térbeli felbontása 9 cm és 3 csatornából áll: zöld, vörös és közeli infravörös. A kép 2008 augusztusában készült az Intergraph *Z/I Imaging* kamerarendszerével.

A debreceni ortofotók közül az egyik, amellyel dolgoztam a Lechner Tudásközpont 2011-es lerepüléséből származik, mely felvételezéséhez Ultracam X digitális légi kamerarendszert használtak. Térbeli felbontása 40 cm, és 4 spektrális sávból áll: kék, zöld, vörös és közeli infravörös. A másik felvételt 2013. augusztus 12-én készítette az EnviroSense Hungary Kft., Leica RCD 30 RGBN 60 MP-es kamerával. A felvétel spektrális csatornái megegyeznek a 2011-es ortofotóéval, térbeli felbontása azonban jobb, 10 cm-es.

3.3. Képosztályozás

A különböző tetőfedő anyagok alapján végzett képosztályozáshoz a 3.2. fejezetben ismertetett felvételek közül mind légi, mind pedig műholdas adatokat felhasználtam (*1. táblázat*).

1. táblázat. A képosztályozáshoz felhasznált távérzékelte felvételek

Felvétel	Spektrális csatornák száma	Térbeli felbontás (cm)	Felvételezés időpontja
WorldView-2 (Debrecen)	8	50	2016.07.24
Ortofotó (Debrecen)	4	10	2013.08.12
Ortofotó (Vaihingen)	3	9	2008.08

3.3.1. Referenciaadatok gyűjtése

A debreceni mintaterület esetében a képosztályozáshoz szükséges referenciaadatokat terepi bejárások során gyűjtöttem egy Stonex S9 RTK GNSS készülékkel. A vizsgált területen összesen 350 ház tetőfedő anyagának típusát azonosítottam. Az így gyűjtött adatokból állt össze a modellek tanításához szükséges adatbázis, melyet ENVI 5.3 szoftver segítségével vektorizáltam. Fontos megemlíteni, hogy a terepi adatgyűjtés és a felvételek készítésének időpontja között 4, illetve 7 év telt el. Bár az említett intervallumban jelentős változás nem történt a mintaterületen, megfigyelhető volt néhány eltérés, például egyes épületeket lebontottak vagy azóta felépítettek, továbbá előfordult az is, hogy néhány ház felújítása révén annak tetőfedő anyagát cserélték le. A pontatlan vagy megbízhatatlan tanító adatok felvételének elkerülése érdekében, annak vektorizálásakor a Google Street View és a Google Earth képeit használtam a helyszíni megfigyelések kiegészítéséhez és a változások ellenőrzéséhez. A területen alapvetően 3 tetőfedő dominált: vörös cserép, sötét cserép (szürke, barna, fekete cserepek) illetve az azbeszt-cement tartalmú panelek, így az osztályozásokhoz is ezt a 3 kategóriát használtam fel. Bár az adatgyűjtés alatt azonosítottam egyéb anyagokat is (pl. fém vagy kátrány), ezeket a típusokat alacsony előfordulási gyakoriságuk miatt kihagytam az elemzésből.

A németországi mintaterület esetében a már említett, Google-s adatbázisokat használtam fel a különböző tetőfedők azonosítására. Szembetűnő különbség volt e terület esetében az azbeszt-tartalmú tetők teljes hiánya, azonban bitument nagy számban tudtam azonosítani, így a sötét és vörös cserép mellett ez lett a harmadik kategória.

3.3.2. Osztályozási módszerek alkalmazása

Kutatásom során a tetőfedő anyagokat kétféle megközelítéssel osztályoztam. Elsőként hagyományos, többváltozós statisztikai osztályozási módszert, a lineáris diszkriminancia-analízist (*Linear Discriminant Function Analysis*; LDFA) és a kvadratikus diszkriminancia-analízist (*Quadratic Discriminant Function Analysis*; QDFA) alkalmaztam. Az LDFA és a QDFA az általuk generált diszkriminanciafüggvény-pontszámokkal helyettesíti az eredeti változókat (csatornákat), és olyan diszkriminanciafüggvényt hoznak létre, amely a pixeleket azok spektrális jellemzői alapján igyekezik elkülöníteni (Podani, 2000; Tharwat, 2016).

A másik alkalmazott eljárás a Random Forest (RF), egy széles körben elterjedt, robusztus gépi tanulási technika (Breiman, 2001). Az RF az együttes tanulók (*ensemble learners*) csoportjába tartozik, melynek alapgondolata, hogy az osztályozók csoportosítva jobb és megbízhatóbb eredményeket adnak, mintha önállóan alkalmaznánk őket (Rodriguez-Galiano et al., 2012). Az RF modellt számos paraméterén keresztül lehet optimalizálni, én a munkám során a generált döntési fák számát (*ntree*) és a döntési fák egyes csomópontjainál véletlenszerűen kiválasztott változók számát (*mtry*) optimalizáltam, melyet Python programozási nyelven, grid-keresés (*grid-search*) alapján végeztem el. Az *mtry* tekinthető az RF egyik legfontosabb paraméterének, ugyanis a fák csomópontjainál a véletlenszerűen kiválasztott változók teremtik meg a szükséges változatosságot az egyes különálló döntési fák között, ezzel megakadályozva, hogy azok túlzottan korreráljanak. Az *mtry* optimális értéke függ többek között magától az adathalmaztól, illetve a vizsgálat problémától is. A paraméternek a széleskörben elfogadott alapértéke a modellezéshez felhasznált független változók (a felvétel csatornái) számának négyzetgyökével, maximális értéke pedig a változók számával azonos. A grid-keresést követően a kiválasztott értéke a paraméternek a WV2 felvételnél 4, az ortofotóknál pedig 2-2 volt. Az RF több száz döntési fát foglal magába (általában 300 és 1500 között), ahol minden fa egy különálló predikciót képvisel: az az osztály nyer, amelyik a legtöbb "szavazatot" kapja. Bár az *ntree*-nek a modell teljesítménye szempontjából kisebb jelentősége van, mint az *mtry*-nak, a fák számának helytelen kiválasztása túlillesztéshez vezethet. A modelleket 300 fával építettem fel. A képosztályozást Python 3.7 programozási nyelv segítségével végeztem, a *scikit-learn* csomaggal (Pedregosa et al., 2011).

3.3.3. Modellvalidáció

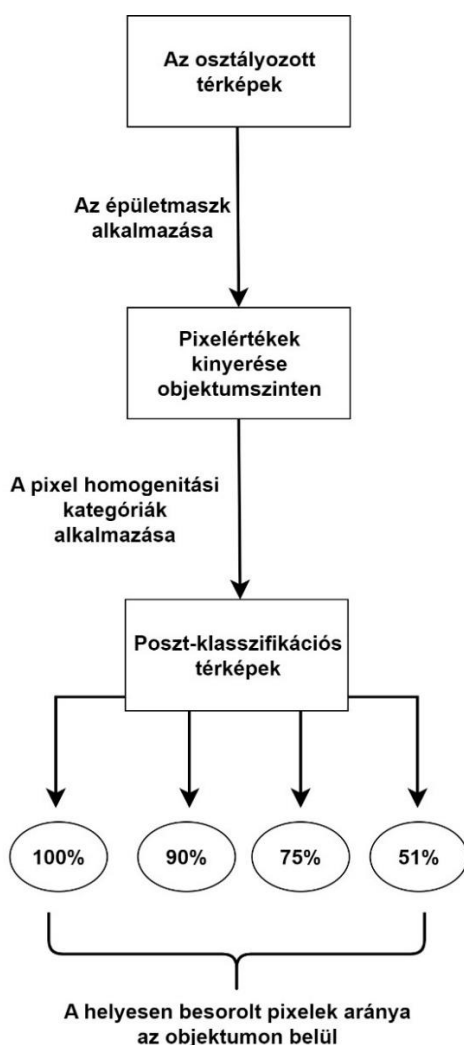
A kutatásom során alapvető fontosságú volt az egyes modellek teljesítményének a számszerűsítése, ezért meghatároztam az OA-kat, mint a modellek kiértékelésének egyik leg többet használt mérőszámát (Burai et al., 2015; Congalton, 1991; Phinzi et al., 2021). Az OA-t a helyesen osztályozott pixelek összegének és a teszteléshez használt pixelek teljes számának a hányadosaként kapjuk meg. A különböző pontosságok kifejezésére két fogalmat használtam: (i) a keresztvalidációból számított becsült pontosságot, és (ii) a független tesztadatokkal történő validálással kapott tényleges pontosságot.

- (i) Az ismételt k -szoros keresztvalidáció (RKCV) alapvetően egy robusztus módszer, mivel számos modell alapján értékeli az osztályozó teljesítményét. Ha azonban a tanító adatok mennyiségében az egyes kategóriákon belül jelentős eltérés van, akkor az RKCV egy változata, a rétegzett k -szoros keresztvalidáció (SCV) használható az egyenlőtlen eloszlásnak a pontosságokra gyakorolt esetleges negatív hatásának minimalizálása érdekében: Az SCV biztosítja, hogy a tanító adatok felosztása során minden részhalmazon belül a kategóriákhoz tartozó pixelek számának eloszlása egyenlő legyen. Az eredményeim validálásához 5-szörös SCV-t alkalmaztam, 10-szer ismételve, így minden pontossági mérőszámot 50 modellből származtattam, továbbá kiszámítottam a modellek mediánját és kvartiliseit.
- (ii) A modellek tényleges pontosságának (TP) kinyerése céljából a már meglévő referenciaadatok mellé, azoktól függetlenül – a szomszédsági viszonyok figyelembevétele mellett – további adatokat gyűjtöttem a teszteléshez. Mivel ebben az esetben a tanítás az összes rendelkezésre álló referenciaadattal, a validáció pedig az utólag gyűjtött független tesztadattal történt, ezért itt elegendő volt az egyszeri modellezés, ugyanis – az SCV-vel ellentétben – a tanítás során a teszteléshez használt adatkészlet nem változott.

3.3.4. *Objektum alapú pixel homogenitás*

A képpont alapú validáció hatékony módot biztosít az eredmények igazolásához, azonban ezek a teszt adatbázisok pixeleire korlátozódnak. A munkám során céлом volt továbbá az objektumszintű pontosság feltárása is, azaz, hogy hány pixelt sikerült helyesen beosztályozni az egyes háztetőkön belül. Ehhez első lépésként épületmaszkot hoztam létre, mely segítségével kinyertem az egyes épületek által lefedett pixeleket (4. ábra). A maszk generálása két réteg felhasználásával történt. Az egyik egy 2013-ban készített LiDAR-os lerepülésből származtatott digitális felszín- és domborzatmodell különbségéből generált normalizált digitális felszínmodell (NDFM), míg a másik a felvételek vörös és közeli infravörös csatornái alapján számolt NDVI (*Normalized Difference Vegetation Index*) volt. Az elkészült épületmaszk az NDFM 3 m-nél magasabb, valamint az NDVI 0,1-es értékénél alacsonyabb pixeleit tartalmazta.

Ezután objektumszinten kiértékeltem a különböző osztályozások eredményeit, összegezve a helyesen osztályozott pixelek arányát egy adott objektum (azaz épület) összes pixeléhez viszonyítva. Négy, objektum alapú pixel homogenitási (*object-based pixel purity; OPP*) kategóriát határoztam meg: 100, 90, 75 és 51%. A 100%-os homogenitás azt jelenti, hogy az algoritmus egy adott tető összes pixelét a helyes kategóriába sorolta, míg az 51% azt jelzi, hogy a pixelek többsége a helyes osztályba tartozik. Ha a pixelek legalább 51%-a abba az osztályba került besorolásra, amelybe az adott tető ténylegesen tartozott, akkor az helyes predikciónak tekinthető. Az OPP-hez kapcsolódó elemzéseket csak a debreceni felvételeken végeztem el.



4. ábra. Az objektum alapú pixel homogenitás meghatározásának lépései

3.4. A területi autokorreláció hatásának elemzése

Az SA szintjének kvantitatív mérésére a Moran-féle I (MI) mutatót használtam (Moran, 1948). Az MI az egyik legismertebb statisztikai mérőszám az SA számszerűsített formájában történő kifejezésére, melyet széles körben használnak a különböző térinformatikai kutatások során (Fu et al., 2014; Kumari et al., 2019; Nikparvar és Thill, 2021; Zhang et al., 2008). Az MI -1 és +1 között vehet fel értéket, ahol a negatív értékek diszperziós, a pozitívak pedig klaszteresedési tendenciára utalnak (Tu és Xia, 2008).

Az MI-hez kapcsolódó számításokat ArcGIS 10.8.1 programban végeztem el. Mivel a célom a pozitív SA vizsgálata volt, és a távérzékelési adatokban a szomszédos pixelek általában erősen korrelálnak, az alacsony pozitív SA értékeket összevontam a negatív értékekkel. A kapott értékeket 3 kategóriába osztottam azok klaszteresedési tendenciái (magas pozitív SA) alapján:

- -1 – 0,4: alacsony,
- 0,4 – 0,7: közepes,
- 0,7 – 1: magas pozitív SA.

3.4.1. Vizsgálatok különböző mintavételezési módszerekkel

Első lépésként a modellvalidáció általános módjaként a referencia adathalmazt felosztottam tanító és teszt területekre (Lyons et al., 2018; Pawluszek-Filipiak és Borkowski, 2020). Ebben az esetben a modellek tanítása és értékelése csak egyetlen eseményen (azaz egyetlen véletlenszerű felosztáson) alapul. Az adathalmazt – az 5-szörös keresztvalidációhoz hasonló feltételek biztosítása végett – 80-20% arányban osztottam fel tanító és tesztelő adatokra (5. ábra). Annak érdekében, hogy ez a módszer statisztikailag összehasonlítható legyen az SCV-vel (ahol az eredmények 50 modellből származnak), az eljárást 50-szer megismételtem (TT), mely során minden ismétléshez egy új, véletlenszerűen felosztott tanító-tesztelő részhalmazt generáltam.

A keresztvalidációs módszerek jó becslést adnak a modell tényleges pontosságáról amennyiben a modellezéshez felhasznált adatok függetlenek. A távérzékelelt felvételeket azonban erősen terheltek az SA-val, mivel a szomszédos képpontok értékei egymáshoz igen hasonlóak (Wulder és Boots, 1998). A távérzékelelt adatoknak ez egy általános jellemzője, azonban ez

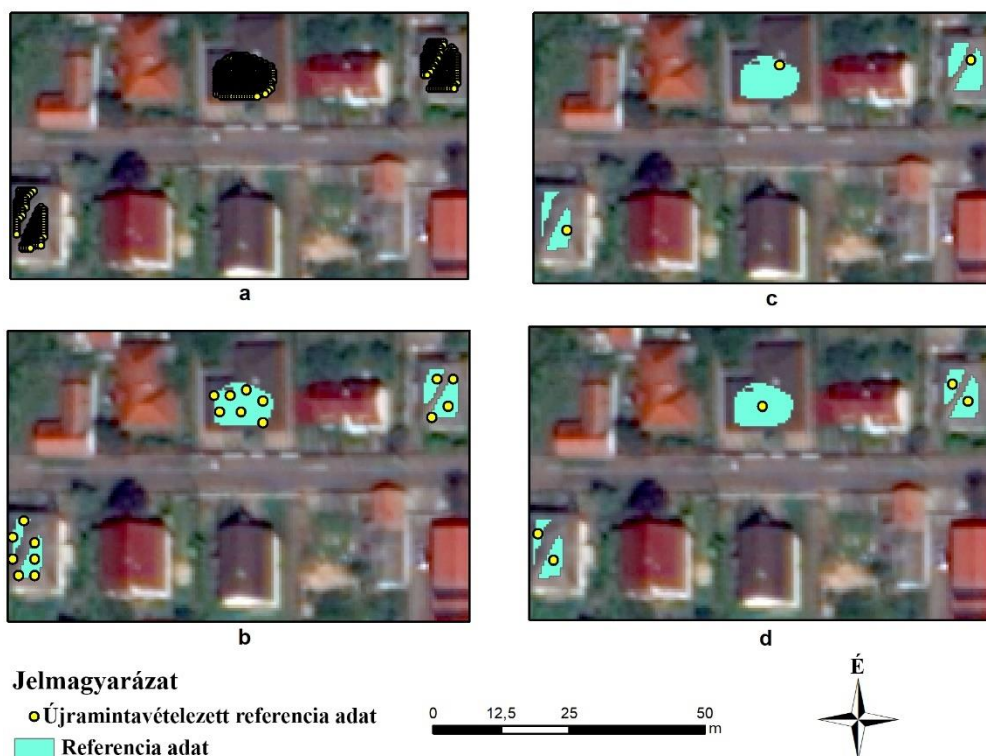
egyúttal megkérdőjelezheti a keresztvalidációval kapott eredmények hitelességét: a keresztvalidáció-alapú pontossági mérések tévesen jó eredményeket adnak, ha a referenciaadatokat nem megfelelően választjuk ki. Ennek megfelelően az esetben a referenciaadatokból történő egyszerű véletlenszerű kiválasztás nem alkalmazható, mivel a tanításra és a tesztelésre kiválasztott pontok hasonló értékekkel rendelkező szomszédos pixelek lesznek, így a tesztelési adathalmaz nem lesz független a tanító adatoktól.



5. ábra. A tanító és teszt adatok vizualizációja 80-20%-os véletlenszerű felosztás esetén, a WorldView-2 egy kivágata alapján

A tetőfedő anyagok osztályozására alkalmazott szimpla, 10-szer ismételt 5-szörös SCV esetében is felmerül a fent említett probléma, mivel ennek során az adatok véletlenszerűen kerülnek felosztásra, a térbeli elhelyezkedésük figyelembevétele nélkül (6/a. ábra). Annak érdekében, hogy az SCV alkalmas legyen a modellek tényleges pontosságának becslésére, Pythonban fejlesztettem egy módszert, amely a referenciaadatokat a térbeli

elhelyezkedésük alapján szelektálja: csak a paraméterként meghatározott távolságon kívül eső pontok kerülnek kiválasztásra a modellezéshez. Kutatásom során három különböző távolsági paramétert (méterben) használtam: 0 (SCV; 6/a. ábra), 2 (SCV2; 6/b. ábra) és 10 (SCV10; 6/c. ábra).



6. ábra. Térbeli mintavételezés különböző távolsági küszöbértékek mellett (a: küszöbérték nélkül – SCV; b: 2 m – SCV2; c: 10 m – SCV10; d: centroid – SCVCent)

Az adatpontok közötti minimális távolság növelésével (6/b. és c. ábra) csökkent a kiválasztott képpontok száma. Előfordultak olyan esetek, amikor egyes poligonokból nem kerültek kiválasztásra adatok, ezért azt az esetet is megvizsgáltam, amikor a poligonokból egy-egy adatpont (a centroid) kerül kijelölésre, így az osztályozás 1 pixel/poligonnal (SCVCent, 6/d. ábra) történt.

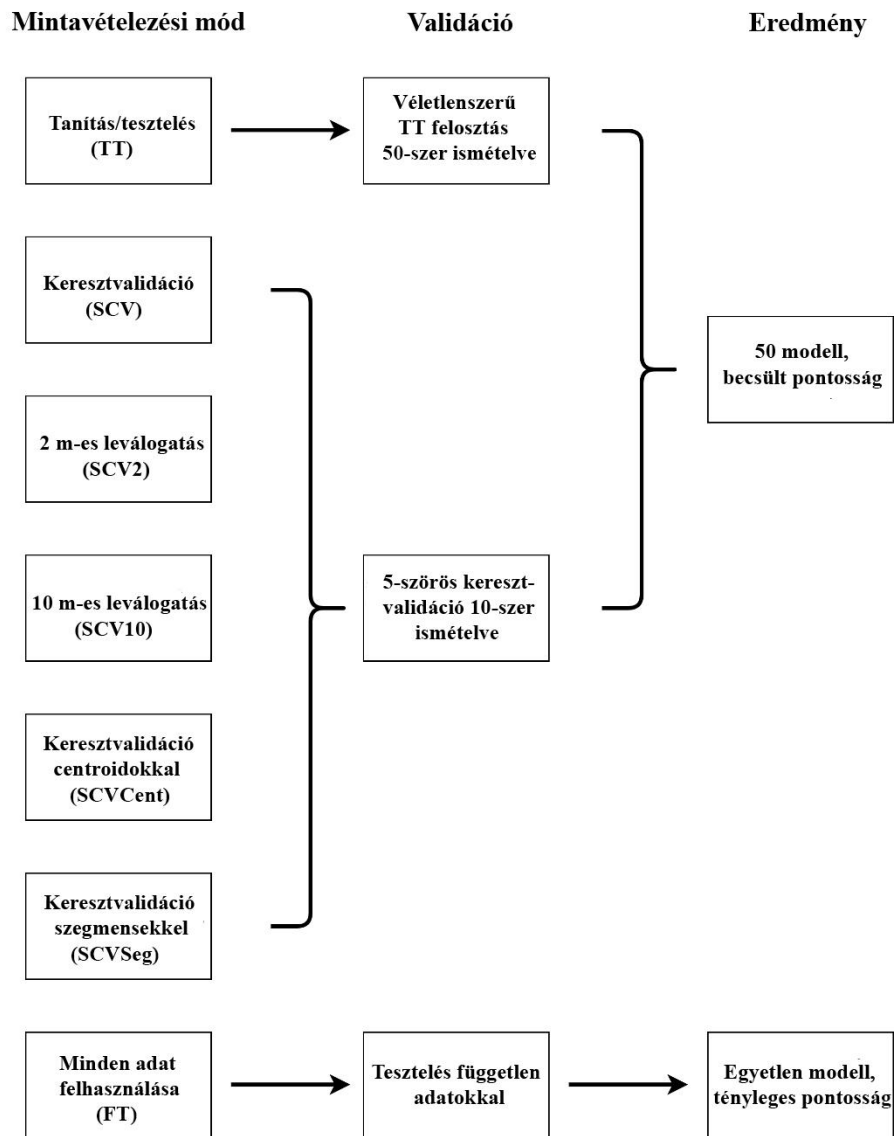
A mintavételezési módok során objektumalapú megközelítést is teszteltem, melyhez SAGA GIS 7.9-et (Conrad et al., 2015) használtam. A szegmentáláshoz az említett szoftverben integrált, Adams és Bischof (1994)

által bemutatott algoritmust használtam, amely egy gyors, robusztus módszer, mely nem igényel paraméteroptimalizációt. A szegmentáció során a szomszédsági viszonyok a Neumann-féle módszer alapján kerültek figyelembevételre, ugyanis Lemenkova (2020) megállapította, hogy a heterogén városi területeken ez a módszer biztosít jó eredményeket. Az eljárás egy központi cella körüli, a négy ortogonális irányban elhelyezkedő szomszédos pixeleket veszi figyelembe a lehatárolás során. Az eredeti felvételek szegmentálása után a létrejött szegmenseket a referenciaadatok poligonjai alapján vágtam ki. Végül minden szegmensből egy pontot választottam ki a 10-szer ismételt 5-szörös SCV alapján történő osztályozáshoz (SCVSeg).

Összesen 7 féle mintavételezési módszerrel teszteltem a modelleket. A validáció során a pontossági értékek minden esetben 50 futtatás alapján kerültek meghatározásra, kivéve a kontroll mérésnek számító, TP céljából végzett független tesztelést (FT; 7. ábra). Meghatároztam továbbá a tanító pixelek számát is, mivel a területi szelekció jelentősen csökkentette az érintett adatokhalmazok méretét (2. táblázat).

2. táblázat. A modellek tanításához használt pixelszám, felvételenként

Mintavételezés módja	WorldView-2	Ortofotó (Debrecen)	Ortofotó (Vaihingen)
FT	6200	15805	55216
TT és SCV	4340	11064	38651
SCV2	268	253	122
SCV10	47	83	38
SCVCent	98	88	90
SCVSeg	357	1970	3301



7. ábra. A modellértékelések a mintavételezési és a validációs módszerek típusa szerint

3.5. Épület szegmentálás mély konvolúciós neurális hálózat alapján

Kutatásom során a U-Net alapú épület szegmentálást három különböző, Debrecen területéről készült távérzékelte felvételen végeztem el (3. táblázat).

3. táblázat. A szemantikus szegmentációhoz felhasznált távérzékelt felvételek

Felvétel	Spektrális csatornák száma	Térbeli felbontás (cm)	Felvételezés időpontja
WorldView-2	8	50	2016.07.24
WorldView-3	8	30	2019.09.16
Ortofotó	4	40	2011

3.5.1. Az input adatok előállítása, adat augmentáció

A U-Net konvolúciós neurális hálózattal történő modellezés kétféle input adattal történik: egyrészt szükség van az eredeti raszterre, valamint az ehhez tartozó, – a vizsgálat szempontjából releváns – maszkra (8. ábra). Mivel a kutatásom során az épületek szegmentációját tűztem ki célul, ezért esetemben az eredeti raszterek mellé 1-1 bináris maszkot kellett előállítani, mely az épületeket különíti el az összes többi objektumtól. A modellezés előtt az input adatokat 512 x 512 pixel méretű darabokra vágtam.



8. ábra. Példa a U-Net tanításához előállított bináris maszkra

Az ortofotó és a WV2 felvétel esetében egy nagyjából 2 km²-es területen került vektorizálásra az összes ház, míg a WV3-nál ez a terület 0,3 km² volt. A két terület közötti eltérés oka, hogy a WV3 esetében a kevesebb tanító adattal történő modellezés értékelése is részét képezte a kutatásnak.

A tanító adatok manuális gyűjtése (kézi vektorizálása) nagyon időigényes feladat. A kutatás során a 3 felvételen összesen több, mint 3000 épület lett vektorizálva. Bár a U-Net architektúra alapvetően jól működik viszonylag kis számú adattal is, az augmentáció (adatbővítés) elengedhetetlen lépés a jobb tematikus pontosság elérése érdekében. A kutatásom során a tradicionálisnak tekinthető augmentációs módszereket alkalmaztam: az eredeti képeket vízszintesen és függőlegesen tükröztem, valamint 90, 180 és 270 fokkal elforgattam azokat. Így az eredeti kép mellett az augmentációnak köszönhetően 5 további képet kaptam (9. *ábra*).

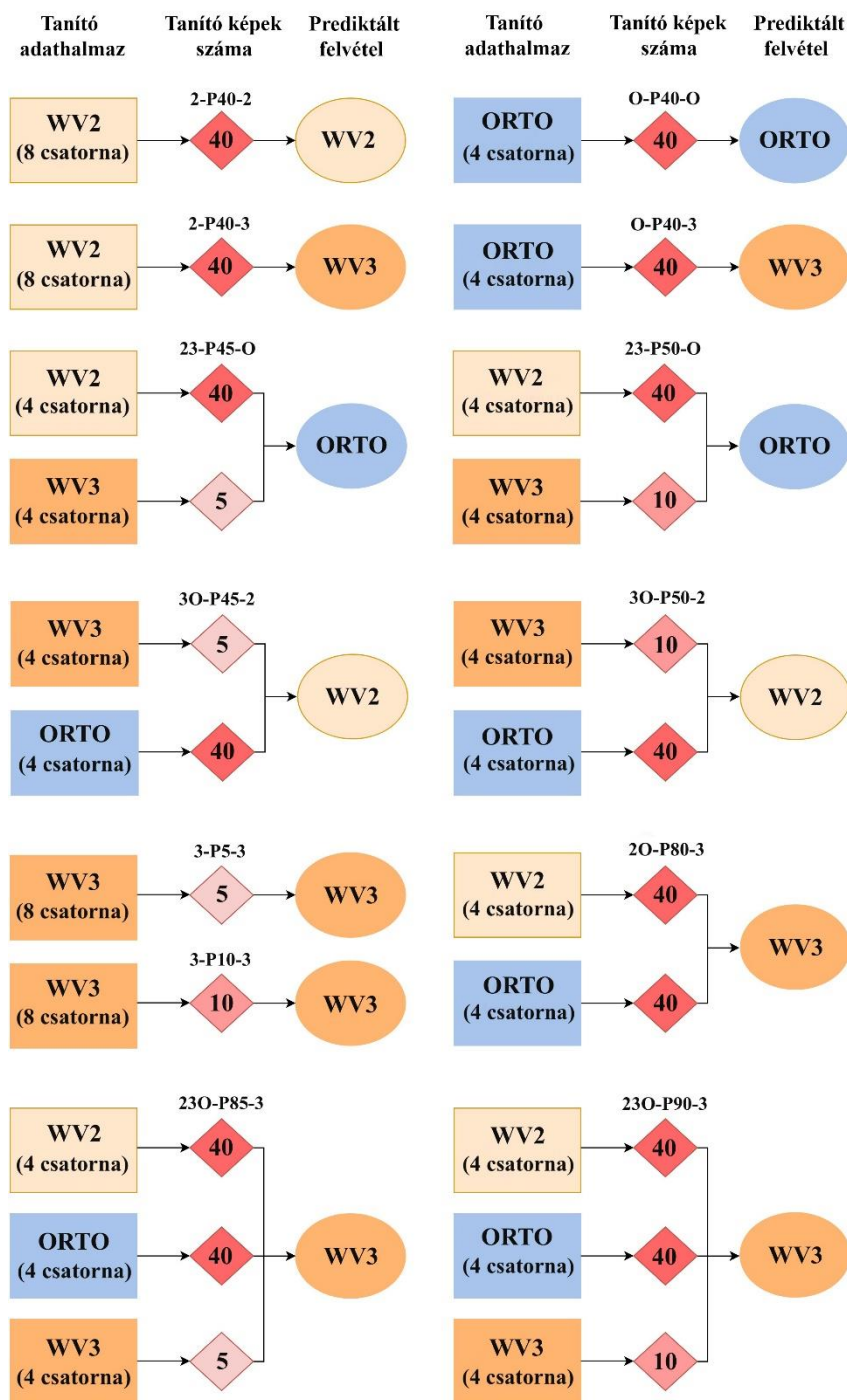


9. ábra. A hagyományos módszerekkel végzett adat augmentáció (a: eredeti kivágat; b: 90 fokos forgatás; c: 180 fokos forgatás; d: 270 fokos forgatás; e: vízszintes tükrözés; f: horizontális tükrözés)

3.5.2. Modell implementáció és értékelés

Az egyes (U-Net) modelleket (10. ábra) legfeljebb 50 *epoch-on* (a tanítási iterációk száma) keresztül tanítottam. Implementáltam egy olyan, korai leállító mechanizmust, amely megszakította a tanítási folyamatot, ha a modell teljesítménye nem javult 15 *epoch* alatt. Minden modell esetében 0,001-es kezdeti tanulási rátát (*learning rate*) választottam. Optimalizációs algoritmusként az Adam-módszert (Kingma és Ba 2014) használtam, 20-as batch-mérettel. A tanításhoz használt képek mérete 512 x 512 pixel volt. Összesen 40 „eredeti”, azaz augmentáció előtti képet használtam a WV2 és a légi ortofotó esetében, és 5, valamint 10-et a WV3 esetében (5. táblázat). A két különböző adatforrás (légi és műholdas felvételek) kombinálása problémát

okozott a modellezési folyamatban, mivel az adatok spektrális felbontása eltérő (4, illetve 8 csatorna), ezért a WV képek 8 sávjából kiválasztottam azt a négyet, amelyek legközelebb esnek az ortofotó 4 sávjának hullámhossz tartományához (kék, zöld, vörös, közeli infravörös). A modellezés Python programozási nyelven történt a *TensorFlow* (Abadi et al. 2016) csomag segítségével. Minden modellfuttatáshoz 24 GB VRAM-mal rendelkező Nvidia RTX 3090 videokártyát használtam.



10. ábra. A modellezési munkafolyamat a modellek rövidített neveivel. Első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

A referenciaadatokat 80-20%-os arányban osztottam fel tanító és validáló adatokra. A modellezés során a validálási pontosságot (*validation accuracy*) és veszteséget (*validation loss*) minden egyes *epoch-on* rögzítettem. A referenciaadatokon kívül a WV3 felvételtől egy további, 0,3 km²-es teszterületet is kiválasztottam. A területen található házakat a referenciaadatokhoz hasonlóan vektorizáltam, az így kapott adatbázis 600 épületből állt. Azokban az esetekben, amikor a prediktált felvétel a WV3 volt (2. ábra), a szegmentáció eredményét ezen új tesztadathalmaz felhasználásával, hibamátrix segítségével is értékeltem. A modellek teljesítményét a hibamátrix alapján számolt precizitás (*precision*) és szenzitivitás (*recall*) alapján hasonlítottam össze. Meghatároztam továbbá az F1-értéket (*F1-score*), amely a precizitás és a szenzitivitás harmonikus közepe (4. táblázat). A precizitás megmutatja, hogy a pozitív predikciók között mekkora a valójában pozitív esetek (jelen esetben pixelek) aránya, míg a szenzitivitás arra utal, hogy a valós állapot szerint pozitív esetek milyen arányban lettek ténylegesen pozitívként prediktálva. Mivel sok esetben mindkét metrika egyaránt fontos, és az egyikre történő optimalizáció a másik rovására mehet, ezért ilyen esetekben célszerű a kiegyensúlyozott értékelést biztosító F1-értéket használni a pontosságok kifejezésére.

4. táblázat. Hibamátrix magyarázatokkal

		Valós állapot	
		Negatív	Pozitív
Prediktált állapot	Negatív	Valós negatív (VN)	Fals negatív (FN)
	Pozitív	Fals pozitív (FP)	Valós pozitív (VP)

$$\text{Precizitás} = \frac{VP}{VP + FP}$$

$$\text{Szenzitivitás} = \frac{VP}{VP + FN}$$

$$F1 = \frac{2VP}{2VP + FP + FN}$$

5. táblázat. A tanító és validáló képek száma az egyes modellek esetében Első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

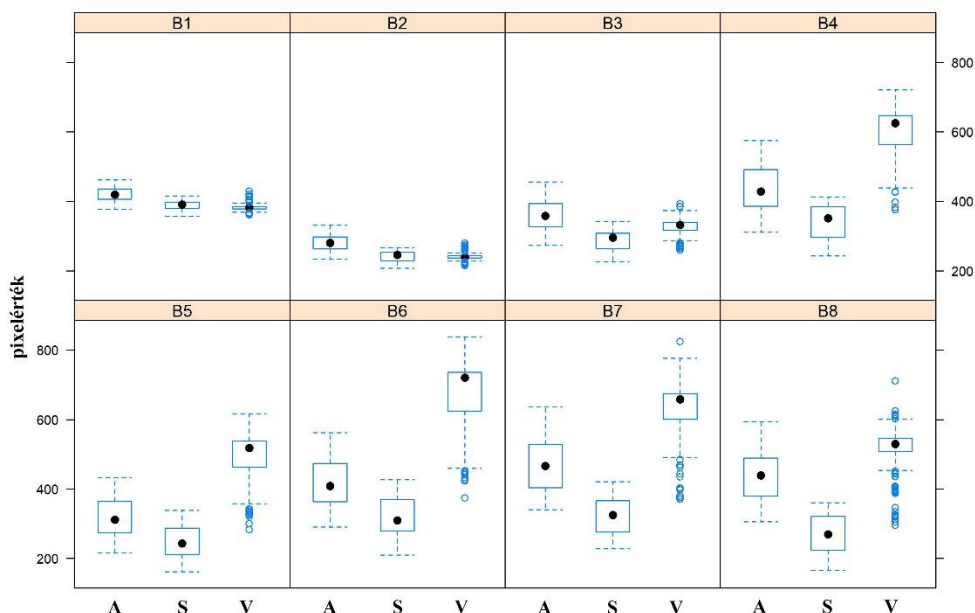
Modell	Tanító képek száma	Validáló képek száma
3-P5-3	24	6
3-P10-3	48	12
2-P40-2	192	48
O-P40-O	192	48
2O-P80-3	384	96
23O-P85-3	408	102
23O-P90-3	432	108
23-P45-O	216	54
23-P50-O	240	60
3O-P45-2	216	54
3O-P50-2	240	60

4. Eredmények és értékelésük

4.1. Tetőfedő anyagok osztályozhatóságának az elemzése

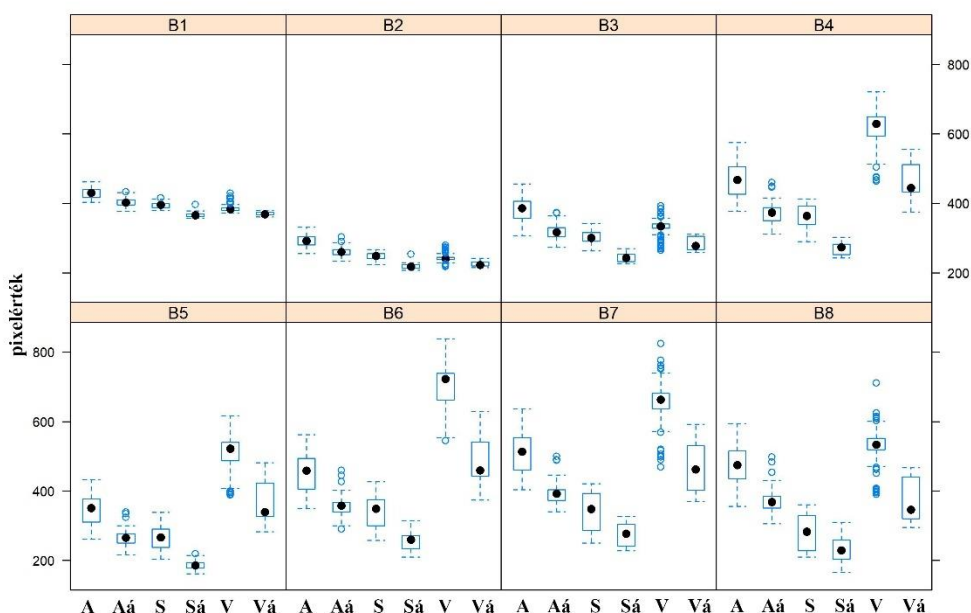
4.1.1. *A pansharpening, valamint a kategóriák számának hatása a tematikus pontosságra*

A tetőfedő anyagok osztályozása 3 kategórián keresztül történt: vörös cserép, sötét cserép és azbeszt. A referenciaadatok alapján a WorldView-2 (WV2) felvételen végzett pixelérték-eloszlásból látszik, hogy a WV2 első két csatornája (B1 [„coastal”, B2 [kék]) nem alkalmas a különböző osztályok azonosítására, ugyanis a hozzájuk tartozó pixelérték-eloszlások nagyban átfednek. A zöld sáv (B3) volt az első, amely korlátozott megkülönböztetést tett lehetővé, de az azbeszt és a vörös cserép pixeleinek értékei hasonlóak voltak. Az 585-625 nm-ig terjedő elektromágneses tartományt lefedő sárga (B4) csatornától kezdve azonban mindhárom kategória pixelintenzitás-értékeinek tartománya jól elkülöníthető volt (11. ábra).



11. ábra. A tetőfedő anyagok pixelérték-eloszlásai a WorldView-2 műholdfelvétel csatornáin alapján, három osztály esetében. A = azbeszt; S = sötét cserép; V = vörös cserép

A tetőfelületek napos és árnyékos kategóriákra történő szétválogatása nagyobb átfedést eredményezett a pixelintenzitás-tartományok között (12. ábra). A B1 és B2 csatornákon rögzített hasonlóság itt is nagyon jelentős volt. Kiemelendő, hogy az azbeszt árnyékos szegmenseinek intenzitás-eloszlásai szinte minden sávban egybeestek a sötét cserép napsütötte szegmenseinek értékeivel. Bár kisebb mértékben, de hasonló tendencia figyelhető meg a B4 csatornától kezdődően a napsütötte azbeszt és az árnyékos vörös cserép kategóriák között.

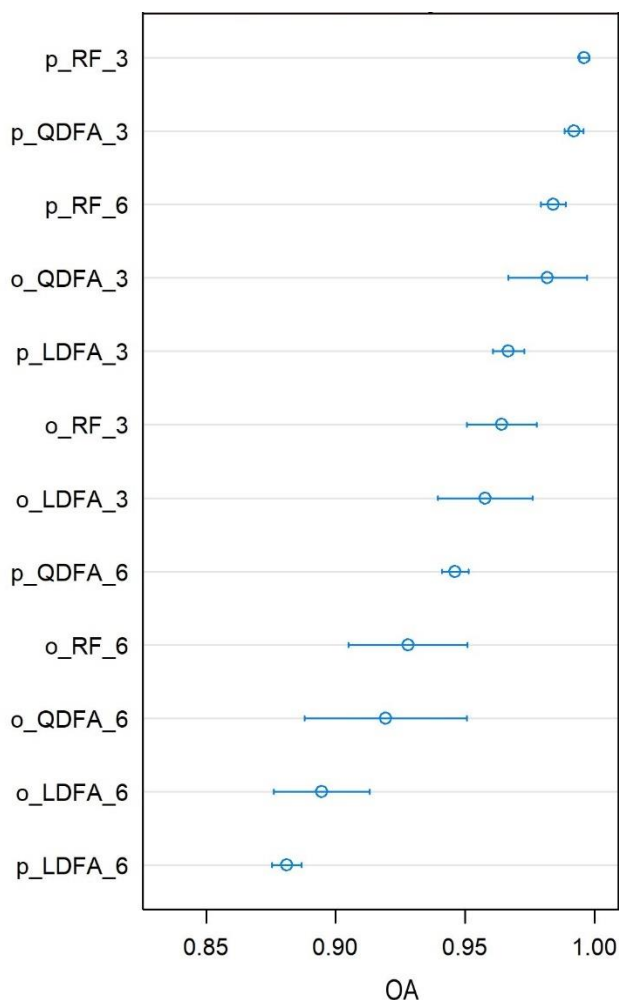


12. ábra. A tetőfedő anyagok pixelérték-eloszlásai a WorldView-2 műholdfelvétel csatornáin alapján, hat osztály esetében. A = azbeszt; S = sötét cserép; V = vörös cserép; Aá, Sá, Vá = A, S és V árnyékos verziói

Az eredeti és a *pansharpening* módszerrel élesített felvételek, valamint a kategóriák megvilágítottság szerint történő megkülönböztetése alapján összesen 12 modellt készítettem (13. ábra). Az osztályozásokhoz használt algoritmusokkal kapcsolatban elmondható, hogy az LDFA teljesített a legrosszabbul, míg a QDFA és az RF hasonlóan jó eredményeket adott. Ugyanakkor az összes modell kapcsán általánosan elmondható, hogy azok sikeresek voltak, az OA-k alapján a legrosszabb eredmény is 87% feletti volt.

A *pansharpening* eljárással feljavított felvétel az eredeti csatornákon végzett osztályozáshoz képest általánosságban 2-3%-os javulást

eredményezett az OA értékek kapcsán. Nagyobb különbségek figyelhetők meg abban az esetben, amikor a tetőket a megvilágítottságuk alapján különíttem el: a magasabb osztályszám jelentősen alacsonyabb OA értékeket eredményezett, esetenként 6-7%-os csökkenés figyelhető meg. Minden modellpár kapcsán elmondható, hogy az összevont kategóriák alkalmazása helyett a napsütötte és árnyékos szegmensekre történő szétosztás rontott a pontosságokon. Kiemelendő azonban az RF osztályozó, amellyel a feljavított felvételen végzett, 6 kategóriás klasszifikáció (p_RF_6) az egyik legjobb eredményt hozta: a 12 modellből a harmadik legmagasabb OA-t kaptam.

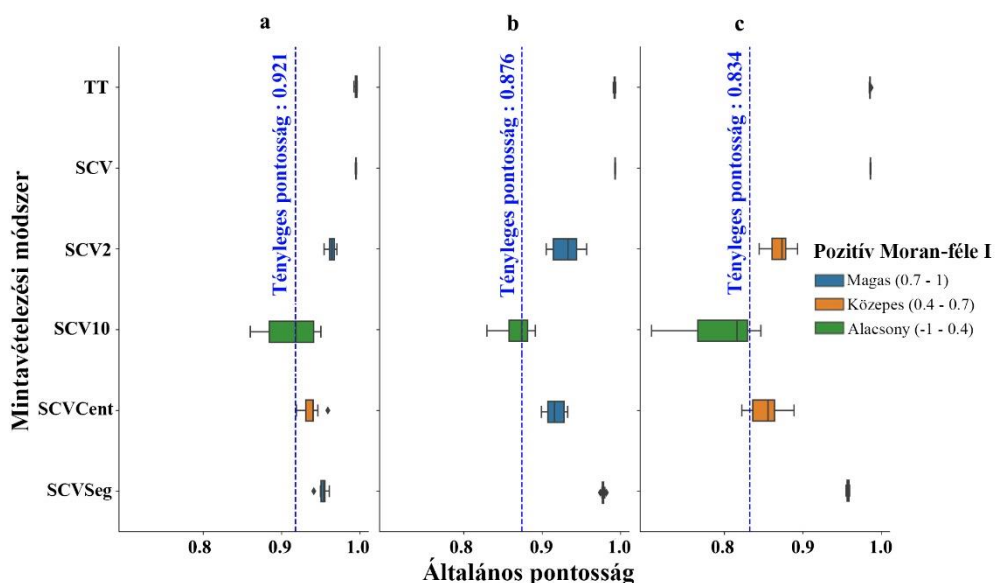


13. ábra. Az egyes modellváltozatok az általános pontosságuk (OA) alapján rangsorolva. o = eredeti felvétel; p = pansharpening eljárással feljavított felvétel; LDFA = lineáris diszkriminancia-analízis; QDFA = kvadratus diszkriminancia-analízis; RF = Random Forest; 3 vagy 6 = az osztályok száma

A diszkriminancia-analízis gyakori osztályozónak számított a távérzékelésben, melynek különböző változatait számos kutatás során alkalmazták (Chhikara és Odell, 1973; Du és Nekovei, 2005; Du és Younan, 2008; Wina és Isa, 2014), mára azonban ezek használatát háttérbe szorították a robusztus gépi és mély tanulási eljárások. Bár az LDFA jól teljesített, a QDFA jobb tematikus pontosságot biztosított, mind az eredeti, mind a *pansharpening* eljárással feljavított sávok esetében, amely eredményről más szerzők is beszámoltak (Tharwat, 2016; Siqueira et al., 2017). Előfordulnak olyan esetek, amikor a kétféle diszkriminancia-analízis szinte azonos teljesítményt nyújt (Hallouche et al., 1993; Manickavasagan et al., 2008; Vadivambal et al., 2010), eredményeim viszont a QDFA jobb teljesítményét igazoló tanulmányok megállapításait támasztják alá.

4.1.2. A területi autokorreláció hatásának vizsgálata

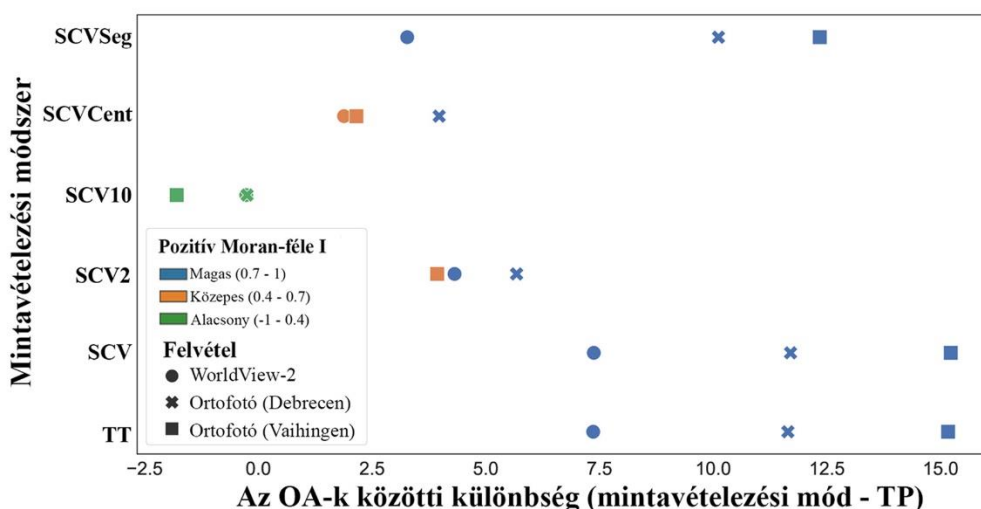
Az SA-hoz kötődő vizsgálatokat a 3.3. fejezetben bemutatott felvételek alapján végeztem el. A különböző mintavételezési módszerek közül a TT és SCV mind a három felvétel esetében hasonló eredményeket adott: az OA-k átlagértéke 98,5% feletti volt, melyekhez szűk interkvartilis tartomány tartozott (14. ábra).



14. ábra. Az egyes mintavételezési módszerek alapján végzett osztályozások eredményei 50 modell alapján, a Moran-féle I vonatkozásában (a = WorldView-2; b = Debreceni ortofotó; c = VaiHINGENI ortofotó)

A TP-hez viszonyítva az SCVSeg módszeren alapuló OA-k szintén túlbecsülnek (95-97%). A fejlesztett Python script alapján készült, távolsági paraméterek megadásával leválogatott modellek bevonása (SCV2 és SCV10) az OA-k egészen a TP szintjéig tartó csökkenését eredményezte, azonban a kevesebb tanító adat az interkvartilis terjedelmének növekedését okozta. A centroidok segítségével végzett osztályozások az SCV2 és az SCV10 modellek közötti pontosságot eredményeztek.

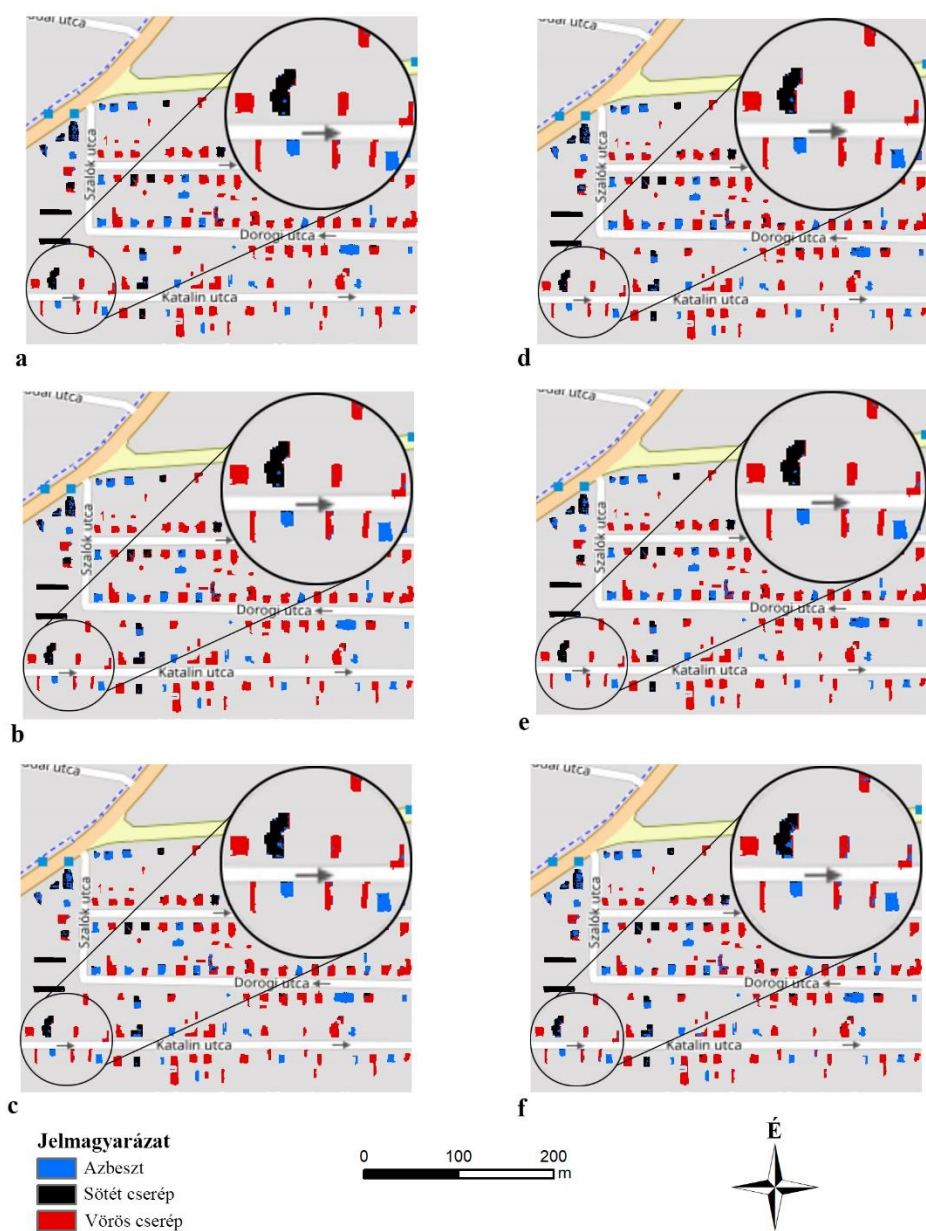
A modellekhez tartozó MI-k az egyes felvételek, az OA-k és a mintavételezési technikák függvényében változtak (15. ábra).



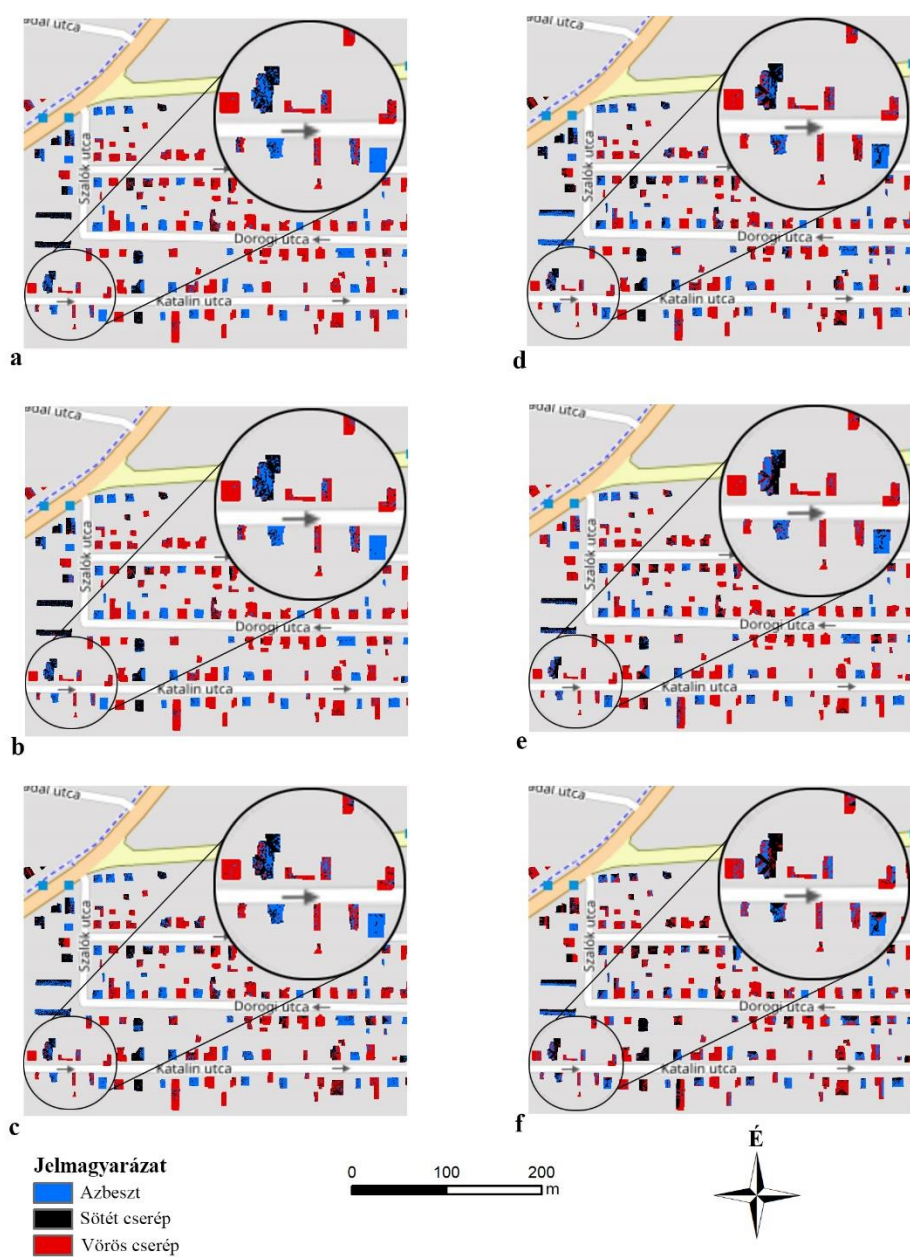
15. ábra. A modellek átlagos általános pontosságai és a független teszteléssel nyert tényleges pontosságok közötti különbségek

Mindhárom felvételen hasonló mintázat rajzolódott ki a TT, SCV és SCVSeg módszereknél: magas pozitív MI figyelhető meg a nagyon magas OA-k (95% felett) esetében. A legalacsonyabb (de még mindig erősen autokorrerált) értékeket a WV2 műholdfelvételnél kaptam. Alacsony MI értékeket csak az SCV10 mintavételezési technikával értem el, továbbá minden felvétel esetében ez volt az a módszer, amely a leghatékonyabban minimalizálta a becslt és a TP értékek közötti különbségeket. Az SCV2 és az SCVCent mintavételezés alkalmazásával a pozitív MI értékek csak a közepes kategóriáig (0,4 – 0,7) csökkentek.

Az egyes tető objektumok pixel homogenitása (OPP) az alkalmazott mintavételezési módszerek, a felvételek és a tetőfedő típusok szerint változott (16-17. ábrák).



16. ábra. A tetőfedő anyagok osztályozási eredményei a WorldView-2 műholdkép alapján, a különböző mintavételezési módszerekkel (a: FT; b: TT; c: SCVSeg; d: SCVCent; e: SCV2; f: SCV10)

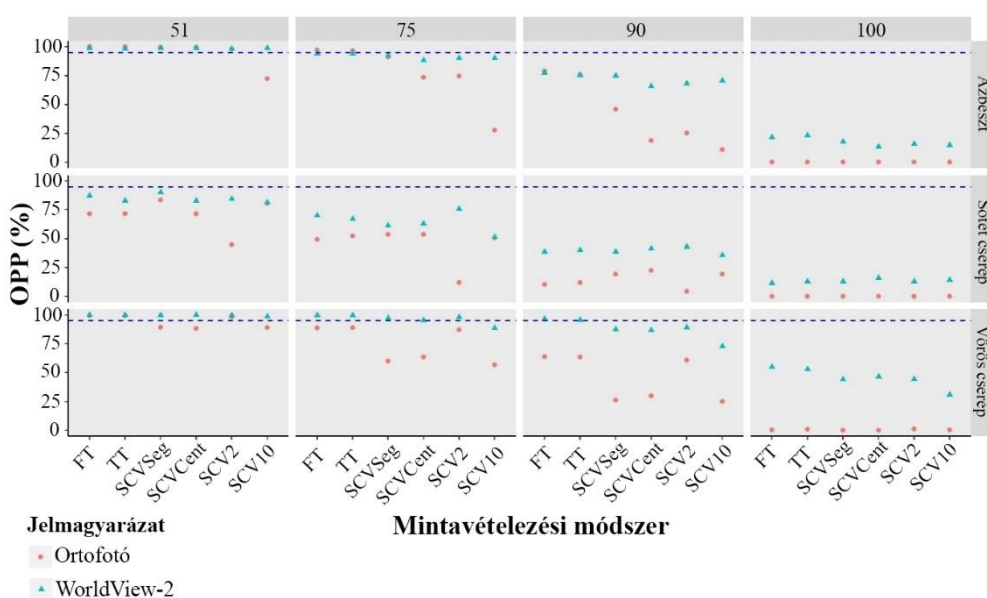


17. ábra. A tetőfedő anyagok osztályozási eredményei a debreceni ortofotó alapján, a különböző mintavételezési módszerekkel (a: FT; b: TT; c: SCVSeg; d: SCVCent; e: SCV2; f: SCV10)

A vizuális kiértékelés alapján a legjobb módszernek mind a WV2, mind pedig az ortofotó esetében az FT modell bizonyult, de a többi is elfogadható eredményt adott. Néhány épület esetében az osztályozott pixelek között

nagyfokú keveredés figyelhető meg, mely általában a sötét cserép és az azbeszt kategóriák között volt jellemző, de előfordult olyan is, hogy mind a három osztály pixelei előfordultak egy tetőn. Objektum alapon vizsgálva a hibás osztályozás szintje többnyire 50% alatt volt, és a végső osztály az objektumalapú többségi szűrő (*majority filter*) alkalmazása után helyes lett.

Az ortofotó esetében a 100%-os homogenitási kategóriát tekintve (azaz amikor egy adott tető minden pixele helyesen volt osztályozva) a különböző mintavételezési módszerek gyengén teljesítettek, általában nem került tető ebbe a kategóriába. Az egyetlen kivételnek a vörös cserép számított, azonban ezek OPP-je 1% alatt volt (18. ábra).



18. ábra. A pixel homogenitási kategóriák (51-75-90-100%), a mintavételezési módszerek (FT; TT; SCVSeg; SCVCent; SCV2; SCV10), a tetőfedő anyagok (azbeszt; sötét cserép; vörös cserép) és a felhasznált felvételek (ortofotó; WorldView-2) szerepének vizsgálata az OPP (%) -re. (kék szaggatott vonal: a 95%-os határérték)

Összehasonlításképp, a WV2 felvétel esetében jobb eredményeket kaptam, az OPP 30-55% között mozgott a vörös cserép, és 10-24% között a sötét cserép és az azbeszt esetében, függően a különböző mintavételezési módszerektől. A 75%-os homogenitási kategóriában az átlagos OPP értékek az ortofotó esetében 60% felett, a WV2-nél pedig 84% felett voltak, míg az 51%-os kategóriát nézve ezek az értékek 86% és 94% fölé tehetőek. Mind a négy

homogenitási osztályban felülmúlta a műholdkép a légi ortofotót, a különbségek azonban jelentősen csökkentek az alacsonyabb homogenitási kategóriák felé haladva: míg 90%-nál 32-36% volt a különbség a két felvétel között, addig 51%-nál ez már csupán 5-8% volt.

A sötét cserépek általában gyengébben teljesítettek, mint a másik két kategória. A 144 modell (4 homogenitási szint $\times$ 3 tetőfedő anyag $\times$ 6 klasszifikációs eljárás $\times$ 2 felvétel) rangsorát tekintve elmondható, hogy az első 5 modell (2 azbeszt és 3 vörös cserép) 100%-os OPP-vel rendelkezik, melyek az 51%-os homogenitási kategóriába, az FT és TT módszerekhez tartoznak. 29 modell érte el a 95%-os OPP határértéket, melyekből 13 azbeszt és 16 vörös cserép volt: 2 modell a 90%-os, 6 a 75%-os és 21 az 51%-os homogenitási kategóriába esett. Ezen modellek között a mintavételezési módszerek az alábbiak szerint alakultak: 7 esetben FT és TT, 5 esetben SCV2, 4 esetben SCVCent és SCVSeg, valamint 2 esetben SCV10. A sötét cserép osztály első előfordulása a 36. helyre tehető.

Bár az SCV hatékony módszernek számít a modellek teljesítményének értékelésére, megbízható eredményt csak az autokorreláció nélküli adathalmazok esetén fog adni (Karasiak et al., 2021). A távérzékelésben az SA a felvételek természetes velejárója, amely hatására – bizonyos esetekben – problémaként merülhet fel az adatok egymáshoz viszonyított függetlenségének hiánya. A függetlenség problémája különösen akkor merül fel, ha a képosztályozáshoz a referenciaadatok gyűjtése klaszteres mintavételezésen (pl. poligonok) alapul, így az SCV-vel értékelt modellek túlbecsülik azok valós teljesítményét (Bahn és McGill, 2013; Hampe, 2004; Mannel et al., 2011; Meyer et al., 2019).

Esetemben az osztályozás célobjektumai – a tetők – olyan speciális esetet képviseltek, amelyben a referenciaadatok gyűjtésének lehetséges területei meglehetősen korlátozottak. Az SA hatására az OA-k 7-17%-kal haladták meg a független tesztterülettel kapott TP értékeket. Megállapítottam, hogy a térbelileg független adathalmazok alkalmazása megfelelő módszer az SA pontosságtorzító hatásának elkerülésére, mely tekintetében a kapott eredményeim összhangban vannak más kutatók korábbi munkáival (Bahn és McGill, 2013; Ibrahim és Bennett, 2014; Ploton et al., 2020). A mintavételezés során a távolsági küszöbérték paraméterének kellően nagy értékre (esetemben 10 m) állítása megszüntette a pozitív SA hatásait, így az SCV nem torzított a

tematikus pontosságok validálása során. Ez a fajta térbeli mintavételezés azonban relevánsan csökkenti a referenciaadatok számát, ami növeli a modell bizonytalanságát, azaz szélesíti a konfidenciaintervallumokat (14. ábra; Nakagawa és Hauber, 2011).

Az összes referenciaadat (FT) felhasználásával végzett osztályozás esetén a becsült pontosság és a TP közötti különbség a WV2 felvétel esetében volt a legkisebb (7,5%), míg a vaihingeni ortofotó esetében elérte a 15%-ot. Erre a magyarázat a térbeli felbontások közti különbségben rejlik. A WV2 esetében a felbontás 50 cm, míg az ortofotónál 9 cm, amely azt jelenti, hogy az adatmennyiségek közötti különbség közel 31-szeres volt, azaz a tetők jelentősen több pixelen voltak reprezentálva az ortofotó esetében. A különböző mintavételezéssel nyert adathalmazokon mért pontosságok (14. ábra) alapján elmondható, hogy a nagyobb térbeli felbontás negatívan hat a tematikus pontosságra, ugyanis a felbontással nő az egyes osztályokon belüli pixelértékek szórása. A kapott eredmények konzisztensek Underwood és munkatársai (2007) megállapításaival, mely szerint bizonyos esetekben a spektrális felbontás szerepe jelentősebb a jobb eredmény elérésében. Korábbi munkájuk során Marceau és munkatársai (1994), Tran és munkatársai (2011), valamint Sun és munkatársai (2016) is kimutatták, hogy a nagy osztályon belüli szórás negatív hatásai csökkenthetők alacsonyabb térbeli felbontású felvételek használatával, így különösen a nagy heterogenitású városi területek esetében érhető el magasabb osztályozási pontosság.

Az OPP-hez kapcsolódó vizsgálatok számszerűsítették a pixelalapú képosztályozás egyik általános velejáróját, az úgynevezett só-bors hatást, melynek során az egy osztályhoz tartozó objektumok félreosztályozott pixeleket is tartalmaznak. Azok az objektumok, amelyek a 100%-os homogenitási kategóriákba estek, mentesek voltak a só-bors hatástól, tehát nem tartalmaztak félreosztályozást, míg az 51%-os küszöbérték esetében többségben jól osztályozott képpontok alkották az adott tetőt. Bár az egyes tetők azonos anyagból épülnek fel, a felületüket különböző inhomogenitások törik meg, mint például tetőablakok, napelemek, árnyékok, zuzmók, mohák stb., ezért a szórványosan más osztályba sorolt pixelek előfordulása normálisnak tekinthető, azonban a helyesen besorolt pixelek abszolút többsége meghatározhatja a valódi osztályt. Bár az OA-k az SA miatt jelentősen túlbecsültek, az OPP ezzel szemben más eredményeket mutatott. Fontos azonban, hogy míg az OA értékek a pixelek, addig az OPP-k az objektumok

alapján, posztklasszifikációt (generalizáció) követően kerültek kiszámításra. A térbeli felbontás növelésével az OPP csökken, ezzel a só-bors hatás hangsúlyosabbá válik az osztályozott képeken, ami végső soron alacsonyabb pontosságokat, és zajosabb tematikus térképeket eredményez.

A képosztályozás során a sötét cserép esetében kaptam a legalacsonyabb pontosságot, ugyanis – különösen a debreceni ortofotó esetében – ennél a kategóriánál volt a legmagasabb az osztályon belüli pixelértékek szórása, mivel ez a kategória több különböző árnyalatú (fekete, barna és szürke) cserepekből áll. A vaihingeni ortofotó alacsonyabb pontossága betudható annak, hogy a felvétel minősége alapvetően rosszabb volt (magas zaj szint, jelentős geometriai torzulások, adathiány stb.) valamint a napelemek nagy száma miatt a tetők heterogénebbek voltak, mint a debreceni területen.

A modellezéshez használt tanító pixelek számának csökkentésével a pontosság, valamint az OPP is csökkent. Ezek alapján két fő megállapítás tehető: (i) bár az SA jelentősen torzítja a pontossági metrikákat, objektumszinten a végeredmény ettől függetlenül pontos lesz; (ii) az SCV10 esetében a tanításhoz használt képpontok száma jelentősen lecsökkent, ezért nem ez adta a legpontosabb térképet, azonban ennek ellenére mégis hatékonynak bizonyult: egyrészt minimálisra csökkentette az SA hatását, így jó becslést adott a TP-re, másrészt az abszolút többségi pixel alapú vizsgálatot (51%-os homogenitás) tekintve az OPP értékek alig különböztek a legpontosabb (FT) modelltől. A térbeli mintavételezéssel kapcsolatos megközelítés egyik korlátja, hogy sokszor nem biztosítható megfelelő mennyiségű adat a modellezéshez. Különösen kisebb objektumok (pl. tetők) esetében az ilyen típusú térbeli mintavételezés jelentősen csökkentheti az adatmennyiséget, ami növeli a modellek bizonytalanságát (azaz szélesebb konfidenciaintervallumokat eredményez), így előfordulhat olyan helyzet, hogy bár a modell mentes az SA negatív hatásaitól, mégsem az fogja adni a legjobb osztályozási térképet.

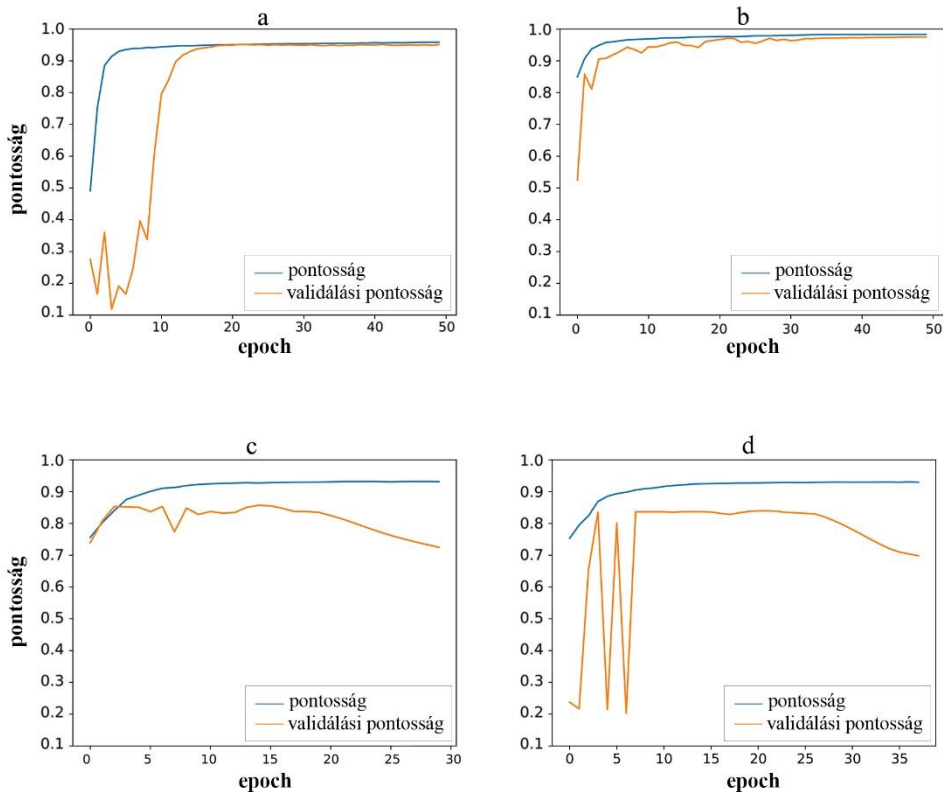
4.2. A U-Net mély konvolúciós neurális hálózat értékelése épület szegmentáción keresztül

A U-Net hálózat értékelését a modellezéshez felhasznált három különböző felvételtől (3. táblázat) gyűjtött tanító adatok alapján az alábbi megközelítések szerint vizsgáltam:

- egy adott felvételtől gyűjtött adatokkal tanítottam a modellt, és ugyanazon felvételnél egy másik területét szegmentáltam,
- a szegmentált felvételtől nem használtam tanító adatot a modellezés során,
- az előző kiegészítése úgy, hogy bevontam a célfelvételtől gyűjtött adatokat is a tanításba.

4.2.1. *Az egyes modellek pontosságának értékelése ugyanazon a felvételen történő predikció során*

Mind a WV2, mind pedig az ortofotó hasonlóan jó eredményeket adott amikor a szegmentálást ugyanazon felvételnél egy másik területén végeztem el (2-P40-2, O-P40-O; 10.ábra): a validálási pontosságok 94% és 97% felett voltak (19/a. és b. ábra). A tanulási görbéken mindkét modellenél jól látszódik, hogy néhány *epoch* után a validálási pontosságok a – tanítási – pontosságokkal együtt változtak, anélkül, hogy a tanulási fázisban meghaladták volna azt, így ezeknél a modelleknél nem figyelhető meg túlillesztés.

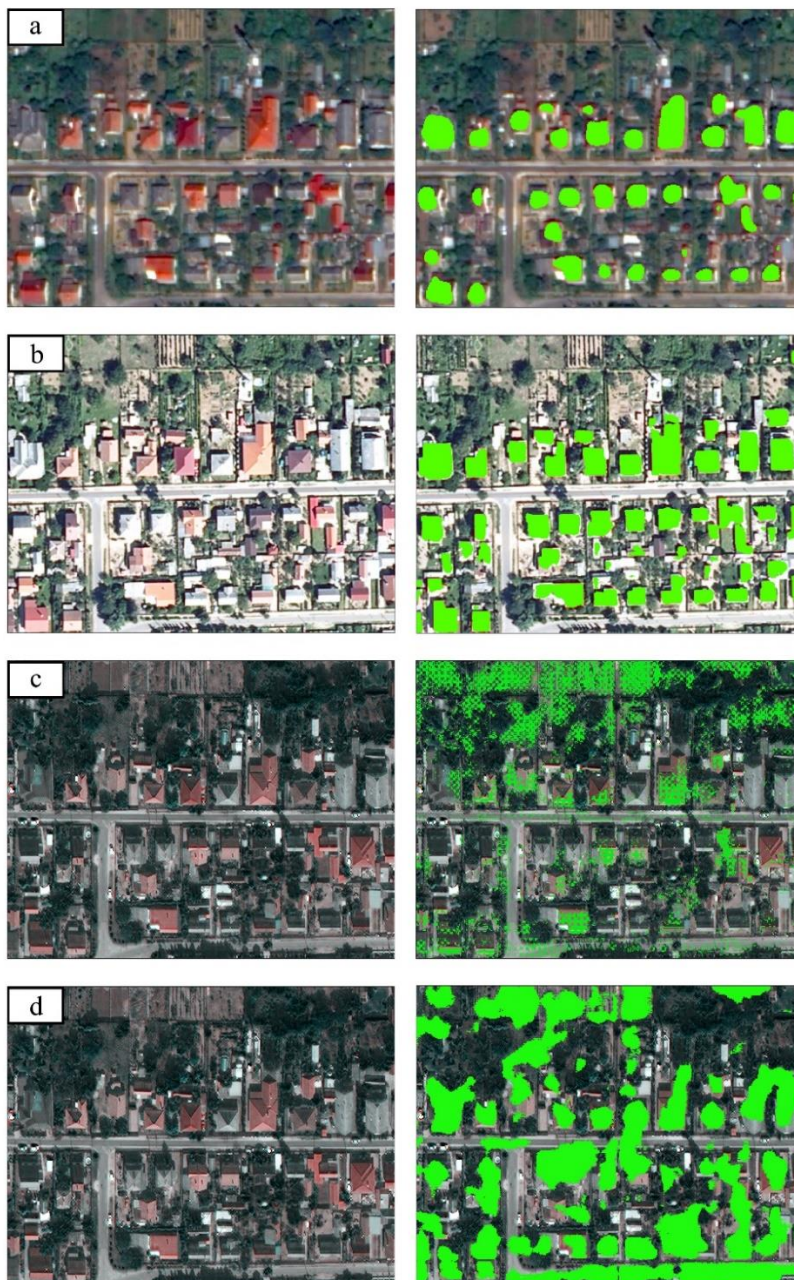


19. ábra. Tanulási görbék abban az esetben, amikor a tanító adatok a szegmentált felvételtől származtak. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

A WV3 esetében (19/c. és d. ábra) a modellek tanulási görbéi mind az 5, mind pedig a 10 tanító adattal történő tanítás során túlilleszkedést mutatnak az első néhány epoch-ot követően. Bár alapvetően az összes modell 50 epoch-on keresztül tanulási folyamattal lett implementálva, a korai leállítási mechanizmus megszakította a tanítási fázist a túlilleszkedés miatt. A négy modell alapján készült szegmentálás eredményei (20. ábra) összhangban voltak a tanulási görbék tendenciáival.

Eredeti felvétel

A szegmentáció eredménye



20. ábra. A szegmentáció eredménye abban az esetben, amikor a tanító adatok a szegmentált felvételből származtak. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Amennyiben a validálási pontosság magas volt, és a modell nem illeszkedett túl, a szegmentálás is sikeres volt (WV2, ortofotó; 20/a. és b. ábra): szinte minden épület azonosításra került, és csak néhány helyen jelent meg tévesen szegmentált pixel. Azok az épületek, amelyeket a modell tévesen háttérként (azaz nem épületként) azonosított (fals negatív eredmények), szinte kizárólag kisebb melléképületek (garázsok, fészerek stb.) voltak. A WV3 felvétel alapján betanított két modell esetében (3-P5-3, 3-P10-3) gyenge szegmentálási eredményeket kaptam. Amikor csak 24 WV3 tanító képet (5. táblázat) használtam a modellezéshez (20/c. ábra), a szegmentálás nem eredményezett összefüggő területeket, és az épületek nagy része kihagyásra került a végeredményből. Bár a 48 tanító képpel végzett elemzés során már néhány épület összefüggő foltokként jelent meg (20/d. ábra), a szegmentálás összességében még mindig gyenge volt, és sok hibát tartalmazott: a csupasz talajfelszín és az utak nagy része tévesen épületként került azonosításra (fals pozitív eredmények).

4.2.2. Az ortofotó szegmentálása a WV2 és a WV3 tanító adatai alapján

A WV2 és WV3 tanító adatai alapján betanított modellel végzett szegmentáció az ortofotó esetén gyenge eredményeket adott (6. táblázat). Az osztályszintű pontossági metrikák alapján elmondható, hogy bár a precizitás értékek viszonylag magasak voltak (0,62 és 0,66), a hozzájuk tartozó szenzitivitások viszont nagyon alacsonyak (0,05 és 0,09), így összességében gyenge F1-értékeket kaptam.

6. táblázat. A modellek pontossága az ortofotón végzett szegmentáció esetében a WV2 és a WV3 tanító adatainak a felhasználásával. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Modell	Precizitás	Szenzitivitás	F1-érték
23-P45-O	0,66	0,05	0,095
23-P50-O	0,62	0,09	0,154

4.2.3. A WV2 szegmentálása a WV3 és az ortofotó tanító adatai alapján

Mindkét modellekombináció esetében a WV2 kép szegmentálása 0,5 feletti F1-értékeket eredményezett (7. táblázat). Abban az esetben, amikor a

WV3-ból 10 képpel tanítottam a modellt (3O-P50-2) 5 helyett (3O-P45-2), a precizitás bár kissé csökkent, de a szenzitivitás növekedése majdnem 0.1 volt, így összességében a több képpel történő tanítás magasabb F1-értéket eredményezett.

7. táblázat. A modellek pontossága a WV2-n végzett szegmentáció esetében a WV3 és az ortofotó tanító adatainak felhasználásával. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Modell	Precizitás	Szenzitivitás	F1-érték
3O-P45-2	0,55	0,52	0,536
3O-P50-2	0,53	0,61	0,567

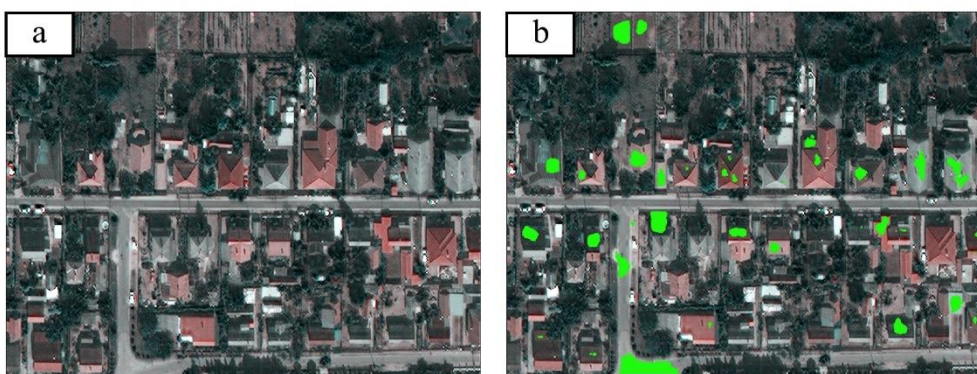
4.2.4. A WV3 szegmentálása a WV2 és az ortofotó tanító adatai alapján

Minden esetben, amikor csak a WV2 és az ortofotó tanító képeit használtam a WV3 szegmentálásához, alacsony F1-értékeket kaptam (8.táblázat).

8. táblázat. A modellek pontossága a WV3-on végzett szegmentáció esetében a WV2 és az ortofotó tanító adatainak felhasználásával. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Modell	Precizitás	Szenzitivitás	F1-érték
2-P40-3	0,23	0,003	0,0067
O-P40-3	0,03	0,00002	0,000034
2O-P80-3	0,69	0,0011	0,184

Bár a három modell közül a legmagasabb F1-értéket a WV2 és az ortofotó tanító adatainak a kombinálásával (2O-P80-3) kaptam, de még ez is nagyon alacsonynak tekinthető (0.184), ezáltal a szegmentálás eredménye sem volt megfelelő: a modell sok házat teljesen kihagyott, míg másokat csak részben szegmentált, továbbá a csupasz talajfelszíneket és az utakat is tévesen épületként azonosította (21. ábra).



21. ábra. A WV3-on végzett szegmentáció eredménye a WV2 és az ortofotó tanító adatainak felhasználásával (2O-P80-3). – a) eredeti felvétel; b) a szegmentálás eredménye. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Bár a 2O-P80-3 modell precizitása viszonylag magas (közel 0,7), a hozzá tartozó szenzitivitás értéke mindössze 0,001 volt, ami arra utal, hogy a modell sok, valójában épületként megjelenő pixelt hagyott ki a szegmentáció során, azaz a fals negatívok száma nagyon magas volt.

4.2.5. *A WV3 szegmentálása a WV2, az ortofotó és a WV3 tanító adatai alapján*

Abban az esetben, amikor a WV3 szegmentálásához az egyéb felvételeken gyűjtött tanító adatok mellé a modellépítés során a WV3-ból származó adatokat is bevontam, jelentősen magasabb F1-értékeket kaptam (9. táblázat). Amennyiben csak az ugyanarról a felvételtől származó tanító adatokat használtam a modellezéshez (3-P5-3; 3-P10-3), a kapott F1-érték bár magasabb volt, mint az ortofotó és a WV2 kombinált modelljének esetében (2O-P80-3; 8. táblázat), de még mindig alacsonynak tekinthető. A 3-P5-3 és a 3-P10-3 modellek összehasonlítása alapján elmondható, hogy mindössze 5 további, a WV3-ról származó tanító adat bevonása a szenzitivitás jelentős növekedését eredményezte (0,2-ről 0,72-re), azonban a precizitás szinte változatlan maradt; tehát a modell több épületet tudott helyesen beazonosítani, azonban emellett a fals pozitívok száma magas maradt.

9. táblázat. A modellek pontossága a WV3-on végzett szegmentáció esetében a WV2, az ortofotó és a WV3 tanító adatainak a felhasználásával. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Modell	Precizitás	Szenzitivitás	F1-érték
3-P5-3	0,22	0,20	0,213
3-P10-3	0,25	0,72	0,369
23O-P85-3	0,69	0,64	0,661
23O-P90-3	0,68	0,71	0,693

A legjobb eredményeket a WV2 és az ortofotó adatoknak a WV3-mal való összevonása hozta (23O-P85-3; 23O-P90-3). Az összes rendelkezésre álló tanító adat (5. táblázat) felhasználásával készült modell (23O-P90-3) F1-értéke 0,693 volt, amely egyúttal a legpontosabb szegmentációt is biztosította a WV3 felvételen (22. ábra). Bár itt is előfordultak téves kategorizálások, ezek aránya jelentősen csökkent, és az épületek többségét helyesen azonosította a modell.



22. ábra. A WV3-on végzett szegmentáció eredménye az összes rendelkezésre álló tanító adat felhasználásával (23O-P90-3). – a) eredeti felvétel; b) a szegmentálás eredménye. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Bár mára az épületek szegmentálására számos módszer került kidolgozásra, a magas tematikus pontosság elérése a lehető legkevesebb ráfordítással (pl. referenciaadatok gyűjtésével) a sűrűn beépített városi

területek inhomogenitása miatt még mindig nagy kihívást jelent: egyes korlátozó tényezők, mint például a növényzet vagy az árnyékok jelenléte nagyban torzíthatják az eredményeket. Továbbá a technológia fejlődésével a rendelkezésre álló, potenciális bementi adatok (felvételek) száma egyre nő. A különböző hordozó platformok (UAS, repülőgép, műhold) segítségével készített felvételek sajátos jellemzői (az alkalmazott szenzor térbeli és spektrális felbontása, a felvétel időpontja stb.) nagyban befolyásolják a kimenetek eredményét. A mély tanuláson alapuló szegmentációs módszerek igen népszerűek a képfeldolgozásban, azonban ezen módszerek egy jelentős limitáló tényezője a hatalmas tanító adat igényük, melyek elérhetősége gyakran korlátozott, előállításuk pedig rendkívül időigényes (Brigato és Iocchi 2020; Jia et al., 2021). Napjainkban az objektumok kinyerésében és azonosításában az egyik legnagyobb kihívást többek között a különböző adatforrások fúziója és együttes alkalmazása jelenti (Liu et al., 2020; Meng et al., 2020), ami egy fontos lépés, ha korábban gyűjtött adatok alapján szeretnénk egy új (független) felvételt elemezni. A munkám során a referenciaadatok fúziója kapcsán én is találkoztam az említett nehézségekkel: az adatok összevonása során Li és munkatársai (2021) ajánlását követve a WV felvételek spektrális csatornái közül úgy szelektáltam, hogy a megmaradó sávok száma és spektrális tartománya a lehető legközelebb essen az ortofotó csatornáihoz. A U-Net mély konvolúciós neurális hálózatnak az épület szegmentációban betöltött szerepével már számos korábbi tanulmány foglalkozott (Alsabhan et al., 2022; McGlinchy et al., 2019; Wang et al., 2020; Yu et al., 2022). Bár a U-Net architektúra alapján sok különböző modellváltozat került kifejlesztésre az elmúlt években a minél nagyobb pontosság elérése céljából (He et al., 2020; Hui et al., 2019; Rastogi et al., 2020; Wang és Miao 2022), a munkám során én a Ronneberger és munkatársai (2015) által először bemutatott, eredeti U-Net architektúrát alkalmaztam, mivel a kutatásom elsődleges célja a különböző forrásokból származó bemeneti adatok együttes alkalmazásának a vizsgálata volt az alap hálózaton keresztül. Mivel az eredeti hálózat megfelelő összehasonlítási alapot biztosított, így a munkám során nagyobb hangsúlyt tudtam helyezni a különböző tanítási stratégiáknak a szegmentálási teljesítményére gyakorolt hatásának vizsgálatára.

Elmondható, hogy amikor az adatgyűjtés és a szegmentáció is ugyanazon a felvételen történt, csekély számú tanító adat mellett is jó

eredmények születtek. A WV2 és az ortofotó esetében 94% feletti validálási pontosságot kaptam úgy, hogy mindössze 192 képet (beleértve az augmentáció során kapott adatokat is; 5. táblázat) használtam a tanításhoz. A tanító képek mérete mindkét felvétel esetében 512 x 512 pixel volt, míg a felbontás a WV2-nél 50 cm, az ortofotónál pedig 40 cm. Más kutatók hasonlóan magas pontosságot értek el a U-Net alapú modellekkel végzett épületszegmentálásban, de a legtöbb esetben jóval több képet használtak fel a tanítás során. Wang és munkatársai (2022) a kutatásukat két ingyenesen hozzáférhető adatbázis, a WHU (Ji et al. 2019) és az Inria (Maggiori et al. 2017) alapján végezték, melynek során az általam használt adatokkal azonos méretű (512 x 512 pixel), de eltérő térbeli felbontású (30 cm) felvételekkel dolgoztak. Ők 96% feletti validálási pontosságot értek el, azonban a tanításhoz használt felvételek száma 4736, illetve 9737 volt. Abdollahi és Pradhan (2021) az AIRS adatbázisban (Chen et al., 2019) közzétett felvételeket használták, melyek mérete 1536 x 1536 pixel, térbeli felbontásuk pedig 7,5 cm, és több, mint 220.000 épületet tartalmaznak: U-Net alapú hálózat alkalmazásával 95% feletti validálási pontosságot értek el.

A WV3 esetében csökkentettem a tanító adatok számát: az augmentáció után a modelleket 24 és 48 képpel tanítottam be (3-P5-3; 3-P10-3). Az ezekhez a modellekhez használt adatmennyiség elégtelennek bizonyult: a tanulási görbén túlilleszkedés volt megfigyelhető (19/c. és d. ábra), és a szegmentálási eredmények is jelentős mennyiségben tartalmaztak téves besorolásokat (20/c. és d. ábra). Az említett két modell hibái mellett azok F1-értékei is alacsonyok voltak (9. táblázat). Amennyiben csak a WV2 és az ortofotó tanító képeit használtam a WV3-on végzett predikcióhoz, még rosszabb eredményeket kaptam (0,184-es F1-érték; 8. táblázat): annak ellenére, hogy a tanításhoz felhasznált kombinált képek száma 384 volt (2O-P80-3), az F1-érték alacsonyabb volt, mint amikor a WV3 24 képének segítségével végeztem a szegmentációt (3-P5-3), tehát bár 16-szor több adat állt a modell rendelkezésére, az eltérő forrásokból gyűjtött adatokkal való tanítás alacsonyabb pontosságot eredményezett. Abban az esetben, ha a WV2 és az ortofotó mellé a WV3 adatait is bevontam a tanításba (23O-P85-3; 23O-P90-3) jelentős javulást tapasztaltam: mindössze 24 kép bevonásával az F1-érték közel 0,5-tel nőtt (0,184-ről 0,661-re; 8. és 9. táblázat). A WV3 felvételből 24 további kép bevonása még pontosabbá tette a predikciót, egyúttal ez a modell (23O-P90-3) eredményezte a legmagasabb F1-értéket (0,693). Emellett

elmondható, hogy ennek a modellnek a precizitási és szenzitivitási értékei hasonlóak voltak (0,68 és 0,71), ami arra utal, hogy a modell teljesítménye a tanítás során konzisztens maradt.

A különbség szembetűnő volt, amikor a WV2 alapján végeztem predikciót a WV3-ra (2O-P80-3), és fordítva (3O-P50-2): előbbi esetben (WV2 $\rightarrow$ WV3) az F1-érték 0,184 volt, míg az utóbbiban (WV3 $\rightarrow$ WV2) 0,567, annak ellenére, hogy itt kevesebb tanító adat állt rendelkezésre. A jelenségre a magyarázat a két felvétel minőségében rejlik: a WV2 nyáron készült, kisebb off-nadír szöggel (18°) és jobb fényviszonyokkal, míg a WV3 ősszel, nagyobb off-nadír mellett (30°). Ez azt eredményezi, hogy a WV3 felvételen az épületek torzabbak, ami egyúttal az árnyékok jelenlétének növekedésével is jár. Amennyiben a rosszabb minőségű WV3 felvételt használtam a tanításhoz, a U-Net jobban generalizált a WV2 predikciója esetében, azonban amikor a WV3 felvételt próbáltam szegmentálni a jobb minőségű WV2 tanító adataival, a modellezés sikertelen lett. Hasonló következtetésre jutottak Weir és munkatársai (2020), akik munkájuk során $7,8^\circ$ és 54° közötti off-nadír szögű felvételeket elemeztek, és az F1-értékek között 0,4-es különbséget kaptak.

Az épületek kinyerésére a U-Net mellett számos más, nagy pontosság elérésére képes mély tanulási módszer létezik. Shi és munkatársai (2020) különböző CNN-eket (pl. U-Net, SegNet, FCN architektúrák különböző változatai stb.) hasonlítottak össze mind közepes (PlanetScope; 3 m), mind pedig nagy felbontású (ISPRS Potsdam referencia-adatbázis; 5 cm) felvételeken, és megállapították, hogy a PlanetScope képek esetében a U-Net több hálózatot is felülmúlt (FCN-32s, FCN-16s, SegNet stb.). Közepes felbontásnál az FC-DenseNet esetében kapták a legmagasabb pontosságot, de a U-Net-hez képest a különbség az F1-értéket tekintve mindössze 3% volt. Megállapították továbbá, hogy nagy felbontás (5 cm) mellett a U-Net esetében az épületek határai bár élesen elválnak, az egyre finomodó részletek mellett a szegmentálás eredménye egyre heterogénebbé válik. Wang és munkatársai (2022) egy új, épületek kinyerésére specializált úgynevezett *Vision Transformer* (ViT) architektúrát dolgoztak ki, és munkájukban összehasonlították a teljesítményét más, korszerű, ugyanezre a feladatra széles körben használt CNN-ekkel (DeepLab V3+; különböző U-Net architektúra változatok stb.). Bár arra az eredményre jutottak, hogy a ViT alapú megközelítésük felülmulta a hagyományos CNN-eket az épületek

szegmentálásában, az egyes U-Net verziókkal összehasonlítva a különbség mindössze 2-3% volt az F1-értékek tekintetében. Fontos megemlíteni, hogy bár ezekben a tanulmányokban számos különböző mély tanuló modellt vizsgáltak, a modellek tanítását elsősorban a különböző referencia-adatbázisokban (WHU, Inria, Massachusetts) elérhető, hatalmas mennyiségű adatra alapozva végezték el. A közelmúltban megjelent tanulmányok eredményei azt mutatják, hogy bár a legújabb megoldások általában jobb szegmentálási eredményeket adnak, a különbségek általánosságban igen csekélyek, ezáltal a U-Net és annak változatai mind a mai napig relevánsak tudtak maradni.

5. Összefoglalás

Disszertációm középpontjában városi területek nagyfelbontású és különböző forrásból származó távérzékelt felvételekkel történő vizsgálata állt, különböző gépi és mély tanulási módszerek alapján. Az elemzések fő célterülete Debrecen volt. A mintaterület kijelölését követően terepi adatgyűjtés során meghatároztam a leggyakrabban előforduló tetőfedő anyagokat, melyek alapján előállt a képosztályozáshoz szükséges tanító adathalmaz. Megvizsgáltam, hogy az egyes tetőfelületek napsütötte és árnyékos kategóriákra történő szétválogatása, valamint a *pansharpening* eljárás hogyan hat a modellek tematikus pontosságra. Továbbá meghatároztam a magas pozitív területi autokorreláció pontosságra gyakorolt túlbecslésének mértékét különböző mintavételezési módszerek esetén, és egy olyan eljárást fejlesztettem, amely ezt a torzító hatást igyekszik megszüntetni. Ezenkívül megvizsgáltam a U-Net mély tanulási architektúra hatékonyságát limitált tanító adattal történő modellezés mellett épületek kinyerése céljából, több forrásból származó távérzékelt adatok felhasználásával. Az értekezésben megfogalmazott kutatási kérdésekre adott válaszokat az alábbiakban összegzem:

1. Hogyan befolyásolja a WorldView-2 műholdfelvételen végzett tetőfedő anyagok osztályozását a térbeli felbontást javító pansharpening eljárás, valamint milyen hatást gyakorol a tematikus pontosságra az egyes tetőfelületek napsütötte és árnyékos kategóriákra történő szétválogatása?

A mintaterület bejárást követően a gyűjtött adatok alapján a képosztályozáshoz három kategóriát jelöltem ki: azbeszt, vörös cserép és sötét cserép. A klasszifikációkhoz három algoritmust alkalmaztam, melyből kettő (LDFA, QDFA) a klasszikus, többváltozós statisztikai megközelítések közé sorolható, míg a harmadik (RF) egy robusztus gépi tanulási eljárás volt. Mivel az osztályozásokat egyrészt elvégeztem a tetőfedő anyagok kategóriáinak összevont, valamint napsütötte és árnyékos területekre történő szétválogatása alapján, másrészt a WV2 felvétel eredeti, illetve *pansharpening* eljárással feljavított változatai tekintetében is, így összesen 12 modellt elemeztem. Megállapítható, hogy a kategóriáknak a megvilágítottságuk alapján történő felosztása a pontosságok csökkenését eredményezte: OA értékek 6-7%-kal alacsonyabbak voltak abban az esetben, amikor a napsütötte és árnyékos pixelek külön osztályt alkottak. A *pansharpening* eljárás hatékony

technikáknak bizonyult a tematikus pontosságok növelésére: az OA értékek rendszerint 2-3%-os növekedést mutattak, de előfordult olyan eset is (6 kategóriás RF), amikor az eredeti és feljavított modellváltozatok között 6%-os különbség volt megfigyelhető.

2. Milyen mértékben járul hozzá a magas pozitív területi autokorreláció az osztályozási pontosság túlbecsléséhez, továbbá hogyan lehet kiküszöbölni a területi autokorreláció torzító hatását a k-szoros keresztvalidációval történő pontosságbecslés megbízhatóságának biztosítása érdekében?

A magas SA következtében, a pontosságvizsgálat során fellépő torzító hatást három különböző távérzékelt felvételen vizsgáltam, melyből kettő debreceni területekről készült, egy pedig az ISPRS által közzétett referencia-adatbázisból származott. Az említett hatás nagyon jelentős volt azokban az esetekben, amikor a referenciaadatokat véletlenszerűen osztottam fel tanító és tesztelő területekké, valamint, ha a napjainkban egyre népszerűbbé váló SCV módszert alkalmaztam a pontosságok becslésére: a becsült pontosságok és a független tesztadatokkal végzett TP értékek közötti különbségek elérték a 17%-ot, így a modellek 99% feletti OA-t eredményeztek. A szegmensek használata sem volt elegendő az SA okozta torzító hatás kiküszöbölésére, ugyanis egyrészt a tetőfelületek meglehetősen heterogének, másrészt pedig relatívan kicsi a kiterjedésük, így az osztályozáshoz kiválasztott pixelek között továbbra is magas volt az autokorreláció. Az általam Pythonban fejlesztett módszer, azaz a referenciaadatok térbeli elhelyezkedésük alapján történő szelekciója sikeresen megszüntette az említett torzító hatást: a modellek SCV-vel becsült pontossága, és a független tesztadatokkal kapott TP értékek jól közelítették egymást. A kisebb térbeli felbontású WV2 (50 cm) felvétel 5-9%-kal teljesített jobban, mint a két ortofotó (9 és 10 cm felbontás), melynek legfőbb oka, hogy a WV2 esetében a tetőt reprezentáló képpontok száma jóval kisebb volt, ezáltal az osztályozáshoz használt kategóriákon belüli pixelértékek szórása is csökkent. Elmondható, hogy objektum szinten vizsgálva a pontosságokat minden mintavételezési technika alkalmazása 95% feletti OPP-ket eredményezett az 51%-os homogenitási kategória esetén.

3. Mennyire hatékony a U-Net konvolúciós neurális hálózat a különböző forrásból származó multispektrális távérzékelt adatokon alapuló épületszegmentálásban, kevés tanító adat felhasználásával?

Az épületek szegmentálásához az eredeti, alap U-Net architektúrát használtam, melynek hatékonyságát három különböző spektrális és térbeli felbontású távérzékelt felvétel esetében vizsgáltam. Amennyiben a tanításhoz használt, és a szegmentálni kívánt adatok forrása – azaz a szenzor – megegyezett, a validálási pontosságok már kis mennyiségű tanító adat (192 darab, 512 x 512 pixeles kivágat) esetén is 94% felettiék voltak (2-P40-2; O-P40-O). A szegmentálások eredményei vizuálisan is megfelelőek voltak az említett modellek esetében, a tévesen besorolt pixelek száma csekély volt. A WV3 esetében a tanító adatok további csökkentése (3-P5-3; 3-P10-3; 24 és 48 darab kivágat) a tanulási görbék túlilleszkedését eredményezték, és a szegmentálás eredménye is elégtelennek bizonyult. A WV3 felvétel szegmentálása abban az esetben is gyenge eredményt adott, ha a modellezéshez csak a másik két felvétel (WV2, ortofotó) tanító adatait használtam (2O-P80-3): bár a betanítás során a felhasznált kivágatok száma jelentősen nőtt (384 darab), az F1-érték mindössze 0,184 volt. Az előző modellt mindössze 24, a WV3-ből származó kivágattal bővítve (23O-P85-3) a pontosság is jelentősen emelkedett: a szegmentáció eredménye vizuálisan is sokat javult, az F1-érték pedig 0,661-re nőtt. További 24 kivágat bevonása (23O-P90-3) bár tovább emelte a pontosságot (F1-érték: 0,693), ennek mértéke már jóval kisebb volt.

Új tudományos eredmények

- A tetőfedő anyagok osztályozását tekintve a *pansharpening* eljárás hatékony eszköz a pontosság növelésére. Az egyes kategóriák napsütötte és árnyékos területekre történő felosztása negatívan hat a tematikus pontosságra.
- A pontosságellenőrzés során a magas pozitív területi autokorreláció jelentősen, akár 17%-kal is torzíthatja az eredményeket. A referenciaadatok térbeli elhelyezkedésük alapján történő újra mintavételezése hatékony módszerként alkalmazható az autokorreláció megszüntetése érdekében.
- A U-Net architektúra alacsony számú tanító adat mellett, eltérő forrásokból származó távérzékelt felvételek esetében is nagy pontossággal alkalmazható épületszegmentálási feladatokra, azonban fontos, hogy legalább kis mennyiségben minden felvételtől legyen felhasználva tanító adat a modellezés során.

6. Summary

The main focus of my dissertation was the examination of urban areas through the use of high-resolution remote sensing images obtained from various sources, employing different machine and deep learning methodologies. Debrecen was chosen as the primary area for the analysis. Following the selection of the study area, the most commonly occurring roofing materials were determined during field surveys, which served as the training dataset for image classification. The impact of discriminating the roofs into sunny and shadowed categories, as well as the effect of the pansharpening process on the thematic accuracy of the models were also investigated. Additionally, the degree of overestimation due to high positive spatial autocorrelation on accuracy for different sampling techniques was determined, and a technique was developed to mitigate this biasing effect. The effectiveness of the U-Net deep learning architecture for building extraction with limited training data was also assessed, utilizing remote sensing data from various sources. The answers to the research questions presented in the introduction of the dissertation are summarized as follows:

1. What is the effect of the pansharpening process on the classification of roofing materials in WorldView-2 satellite imagery, and how does the separation of individual roof surfaces into sunny and shadowed segments affect thematic accuracy?

Following data collection from the study area, three categories were designated for image classification: asbestos, red tiles, and dark tiles. Three algorithms were used for classification, two of which (LDFA, QDFA) belong to the category of traditional multivariate statistical approaches, while the third (RF) is a robust machine learning method. The classifications were performed based on the aggregated categories of roofing materials and on the basis of sunny and shadowed areas, resulting in a total of 12 models, including both the original and pansharpened versions of the WV2 imagery. The results showed that the division of categories based on irradiation led to a decrease in accuracy, with OA values being 6-7% lower when sunny and shadowed pixels were separated into different categories. Pansharpening proved to be an effective technique for improving thematic accuracy, with OA values generally increasing by 2-3%, although in some cases (such as the 6 category RF), a

difference of up to 6% was observed between the original and pansharpened model versions.

2. To what extent does high positive spatial autocorrelation contribute to the overestimation of classification accuracy, and how can the biasing effect of spatial autocorrelation be eliminated to ensure reliability in accuracy estimation through k-fold cross-validation?

To investigate the effect of high spatial autocorrelation (SA) on classification accuracy, I analyzed three different remote sensing images, two of which were acquired over Debrecen one from a reference database provided by ISPRS. The observed effect was significant when the reference data were randomly split into training and testing sets, as well as when the popular SCV method was applied to estimate the accuracy. In these cases, the discrepancies between the estimated and the true accuracies reached up to 17%, falsely resulting in models with OAs above 99%. The use of image segmentation was not sufficient to eliminate the basing effect of SA because roof surfaces are heterogeneous and relatively small in size, resulting in high autocorrelation among the selected pixels for classification. However, the method I developed in Python, which selects reference data based on their spatial location, successfully eliminated the distorting effect of SA. The accuracy of the models estimated with SCV and the true accuracies obtained with the independent test data converged well. The lower spatial resolution WV2 image (50 cm) performed 5-9% better than the two orthophotos (9 and 10 cm resolution), mainly because there were fewer pixels representing the roofs in the WV2 image, resulting in reduced intra-class variance within the categories. The object-based accuracy analysis showed that all sampling techniques achieved over 95% OPP in the 51% purity category.

3. How effective is the U-Net convolutional neural network for building segmentation using multispectral remote sensing data from different sources with limited training data?

For building segmentation, I used the original U-Net architecture and examined its effectiveness using three different multispectral images with varying spatial and spectral resolutions. When the trainings and predictions were conducted on the same image (2-P40-2; O-P40-O), even with a small amount of training data (192 images, 512 x 512 pixels), the validation accuracy

exceeded 94%. The segmentation results for these models were visually satisfactory with a low number of misclassified pixels. However, reducing the amount of training data for WV3 (3-P5-3; 3-P10-3; 24 and 48 images) resulted in overfitting of the learning curves and poor segmentation results. The segmentation of the WV3 image also gave poor results when only the training data of the other two images (WV2, orthophoto) were used for modelling (2O-P80-3): although the number of images used for training increased significantly (384 images), the F1-score was only 0,184. Incorporating 24 images from WV3 (23O-P85-3) significantly improved the accuracy, with an F1-score of 0,661 and visually improved segmentation results. The addition of another 24 images (23O-P90-3) further increased the accuracy (F1-score: 0,693), but to a lesser extent.

New scientific findings

- Pansharpening is an effective technique for improving accuracy in classifying roofing materials. However, dividing each category into sunny and shadowed segments have a negative impact on thematic accuracy.
- High positive spatial autocorrelation can bias accuracy assessment results by up to 17%. Resampling the reference data based on their spatial location is an effective method for eliminating autocorrelation.
- The U-Net architecture can achieve high accuracy for building segmentation tasks using multispectral remote sensing data from different sources even with a low number of training data. However, it is important to use at least a small amount of training data from the target image.

Köszönetnyilvánítás

Mindenekelőtt köszönettel tartozom témavezetőmnek, Prof. Dr. Szabó Szilárdnak, aki az egyetemi éveim alatt szakértelmével és támogatásával végig segítette a munkámat.

Köszönet illeti Enyedi Pétert, aki adatokkal, és szakmai tanácsaival hozzájárult az értekezés létrejöttéhez.

Köszönöm az értekezés bírálóinak, Dr. Hunyadi Gergelynek és Dr. Szabó Gergelynek, akik konstruktív kritikájuknak és megjegyzéseiknek hála nagyban hozzájárultak a disszertációm minőségének javításához.

Köszönöm Papp Melittának, Kovács Lillának és Miklós Viktornak a terepi adatgyűjtések és az adatok vektorizálása során nyújtott segítségét.

Köszönetet szeretnék mondani a családomnak, barátaimnak, munkatársaimnak, valamint volt PhD hallgatótársaimnak az elmúlt évek során nyújtott támogatásaikért.

Az értekezés az Innovációs és Technológiai Minisztérium Kooperatív Doktori Program doktori hallgatói ösztöndíj programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült. A kutatáshoz hozzájárult a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal OTKA K138079 és TKP2020-NKA-04 kódszámú pályázata is.

Irodalomjegyzék

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. arXiv:1603.04467.
2. Abdollahi, A., Liu, Y., Pradhan, B., Huete, A., Dikshit, A., & Nguyen Tran, N. (2022). Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture. *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 673–685. <https://doi.org/10.1016/j.ejrs.2022.06.002>
3. Abdollahi, A., & Pradhan, B. (2021). Integrating semantic edges and segmentation information for building extraction from aerial images using UNet. *Machine Learning with Applications*, 6. <https://doi.org/10.1016/j.mlwa.2021.100194>
4. Abdollahi, A., Pradhan, B., & Alamri, A. M. (2020). An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto International*, 37(12), 3355–3370. <https://doi.org/10.1080/10106049.2020.1856199>
5. Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641–647. <https://doi.org/10.1109/34.295913>
6. Al-doski, J., Mansor, S. B., & Shafri, H. Z. M. (2013). *Image Classification in Remote Sensing*. 3.
7. Alsabhan, W., Alotaiby, T., & Chaudhary, G. (2022). Automatic Building Extraction on Satellite Images Using Unet and ResNet50. *Computational Intelligence and Neuroscience*, 2022, 1–12. <https://doi.org/10.1155/2022/5008854>
8. Antoniou, A., Storkey, A., & Edwards, H. (2017). *Data Augmentation Generative Adversarial Networks*. arXiv:1711.04340.
9. Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122(3), 321–331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>
10. Bardis, M., Houshyar, R., Chantaduly, C., Ushinsky, A., Glavis-Bloom, J., Shaver, M., Chow, D., Uchio, E., & Chang, P. (2020). Deep

- Learning with Limited Data: Organ Segmentation Performance by U-Net. *Electronics*, 9(8). <https://doi.org/ARTN119910.3390/electronics9081199>
11. Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities*, 16(6), 445–458. [https://doi.org/Doi10.1016/S0264-2751\(99\)00042-6](https://doi.org/Doi10.1016/S0264-2751(99)00042-6)
 12. Bertalan, L., Holb, I., Pataki, A., Négyesi, G., Szabó, G., Kupásné Szalóki, A., & Szabó, S. (2022). UAV-based multispectral and thermal cameras to predict soil water content – A machine learning approach. *Computers and Electronics in Agriculture*, 200, 107262. <https://doi.org/10.1016/j.compag.2022.107262>
 13. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/Doi10.1023/A:1010933404324>
 14. Brigato, L., & Iocchi, L. (2020). *A Close Look at Deep Learning with Small Data*. arXiv:2003.12843.
 15. Brownlee, J. (2016). Feature Selection For Machine Learning in Python. *MachineLearningMastery.Com*. <https://machinelearningmastery.com/feature-selection-machine-learning-python/>
 16. Brownlee, J. (2018). *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery.
 17. Burai, P., Deák, B., Valkó, O., & Tomor, T. (2015). Classification of Herbaceous Vegetation Using Airborne Hyperspectral Imagery. *Remote Sensing*, 7(2), Article 2. <https://doi.org/10.3390/rs70202046>
 18. Carranza-García, M., García-Gutiérrez, J., & Riquelme, J. C. (2019). A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks. *Remote Sensing*, 11(3), Article 3. <https://doi.org/10.3390/rs11030274>
 19. Chen, Q., Wang, L., Wu, Y. F., Wu, G. M., Guo, Z. L., & Waslander, S. L. (2019). Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings (Withdrawn Publication). *Isprs Journal of Photogrammetry and Remote Sensing*, 147, 42–55. <https://doi.org/10.1016/j.isprsjprs.2018.11.011>
 20. Chhikara, R. S., & Odell, P. L. (1973). Discriminant analysis using certain normed exponential densities with emphasis on remote sensing application. *Pattern Recognition*, 5(3), 259–272. [https://doi.org/10.1016/0031-3203\(73\)90047-2](https://doi.org/10.1016/0031-3203(73)90047-2)

21. Chisense, C. (2012). Classification of roof materials using hyperspectral data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B7, 103–107. <https://doi.org/10.5194/isprsarchives-XXXIX-B7-103-2012>
22. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5), 545–563. <https://doi.org/10.1111/1754-9485.13261>
23. Cilia, C., Panigada, C., Rossini, M., Candiani, G., Pepe, M., & Colombo, R. (2015). Mapping of Asbestos Cement Roofs and Their Weathering Status Using Hyperspectral Aerial Images. *ISPRS International Journal of Geo-Information*, 4(2), Article 2. <https://doi.org/10.3390/ijgi4020928>
24. Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
25. Congalton, R. G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 10(4), 321–328. <https://doi.org/10.1071/wf01031>
26. Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
27. Dashora, A., Lohani, B., & Malik, J. N. (2007). A repository of earth resource information – CORONA satellite programme. *Current Science*, 92(7), 926–932.
28. Díaz-Delgado, R., Lucas, R., & Hurford, C. (Eds.). (2017). *The Roles of Remote Sensing in Nature Conservation*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-64332-8>
29. Drzewiecki, W., Wawrzaszek, A., Krupiński, M., Aleksandrowicz, S., & Bernat, K. (2016). Applicability of multifractal features as global characteristics of WorldView-2 panchromatic satellite images. *European Journal of Remote Sensing*, 49(1), 809–834. <https://doi.org/10.5721/EuJRS20164943>

30. Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical Image Segmentation based on U-Net: A Review. *Journal of Imaging Science and Technology*, 64(2), 20508-1-20508–20512. <https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508>
31. Du, Q., & Nekovei, R. (2005). Implementation of real-time constrained linear discriminant analysis to remote sensing image classification. *Pattern Recognition*, 38(4), 459–471. <https://doi.org/10.1016/j.patcog.2004.09.008>
32. Du, Q., & Younan, N. H. (2008). Dimensionality Reduction and Linear Discriminant Analysis for Hyperspectral Image Classification. In I. Lovrek, R. J. Howlett, & L. C. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 392–399). Springer. https://doi.org/10.1007/978-3-540-85567-5_49
33. Eslami, M., & Mohammadzadeh, A. (2016). Developing a Spectral-Based Strategy for Urban Object Detection From Airborne Hyperspectral TIR and Visible Data. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 1808–1816. <https://doi.org/10.1109/Jstars.2015.2489838>
34. Fan, X., Yan, C., Fan, J., & Wang, N. (2022). Improved U-Net Remote Sensing Classification Algorithm Fusing Attention and Multiscale Features. *Remote Sensing*, 14(15). <https://doi.org/10.3390/rs14153591>
35. Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Forstner, W., Plumer, L., & Steinhage, V. (1998). Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, 72(2), 185–203. <https://doi.org/DOI 10.1006/cviu.1998.0721>
36. Fiumi, L., Campopiano, A., Casciardi, S., & Ramires, D. (2012). Method validation for the identification of asbestos–cement roofing. *Applied Geomatics*, 4. <https://doi.org/10.1007/s12518-012-0078-0>
37. Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4)
38. Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., & Schaaf, C. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote*

- Sensing of Environment*, 83(1), 287–302.
[https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0)
39. Fu, W. J., Jiang, P. K., Zhou, G. M., & Zhao, K. L. (2014). Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences*, 11(8), 2401–2409. <https://doi.org/10.5194/bg-11-2401-2014>
 40. Gavankar, N. L., & Ghosh, S. K. (2018). Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing*, 51(1), 182–193. <https://doi.org/10.1080/22797254.2017.1416676>
 41. Gibril, M. B. A., Shafri, H. Z. M., & Hamedianfar, A. (2017). New semi-automated mapping of asbestos cement roofs using rule-based object-based image analysis and Taguchi optimization technique from WorldView-2 images. *International Journal of Remote Sensing*, 38(2), 467–491. <https://doi.org/10.1080/01431161.2016.1266109>
 42. Goldblatt, R., Stuhlmacher, M. F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A. K., Cheng, W.-H., & Balling, R. C. (2018). Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remote Sensing of Environment*, 205, 253–275. <https://doi.org/10.1016/j.rse.2017.11.026>
 43. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks (*arXiv:1406.2661*). *arXiv*. <https://doi.org/10.48550/arXiv.1406.2661>
 44. Hallouche, F., Adams, A. E., Hinton, O. R., Surtees, D. P., Wadehra, V., & Sherbet, G. V. (1993). Discriminant analysis for classification of murine melanomas and human cervical epithelial cells. *Analytical and Quantitative Cytology and Histology*, 15(1), 50–60.
 45. Hampe, A. (2004). Bioclimate envelope models: What they detect and what they hide. *Global Ecology and Biogeography*, 13(5), 469–471. <https://doi.org/DOI 10.1111/j.1466-822X.2004.00090.x>
 46. Han, Q. Z., Yin, Q., Zheng, X., & Chen, Z. Y. (2022). Remote sensing image building detection method based on Mask R-CNN. *Complex & Intelligent Systems*, 8(3), 1847–1855. <https://doi.org/10.1007/s40747-021-00322-z>

47. Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics*, 86, 25–32. <https://doi.org/10.1016/j.jbi.2018.08.006>
48. He, N., Fang, L., & Plaza, A. (2020). Hybrid first and second order attention Unet for building segmentation in remote sensing images. *Science China Information Sciences*, 63(4). <https://doi.org/10.1007/s11432-019-2791-7>
49. He, N. J., Fang, L. Y., Li, S. T., Plaza, A., & Plaza, J. (2018). Remote Sensing Scene Classification Using Multilayer Stacked Covariance Pooling. *Ieee Transactions on Geoscience and Remote Sensing*, 56(12), 6899–6910. <https://doi.org/10.1109/Tgrs.2018.2845668>
50. Hoeser, T., Bachofer, F., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications. *Remote Sensing*, 12(18), Article 18. <https://doi.org/10.3390/rs12183053>
51. Hoque UI, M. R., Burks, R., Kwan, C., & Li, J. (2019). Deep Learning for Remote Sensing Image Super-Resolution. 2019. *Ieee 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (Uemcon)*, 286–292.
52. Hu, L. S., Qin, M. J., Zhang, F., Du, Z. H., & Liu, R. Y. (2021). RSCNN: A CNN-Based Method to Enhance Low-Light Remote-Sensing Images. *Remote Sensing*, 13(1). <https://doi.org/ARTN 6210.3390/rs13010062>
53. Hui, J., Du, M., Ye, X., Qin, Q., & Sui, J. (2019). Effective Building Extraction From High-Resolution Remote Sensing Images With Multitask Driven Deep Neural Network. *IEEE Geoscience and Remote Sensing Letters*, 16(5), 786–790. <https://doi.org/10.1109/LGRS.2018.2880986>
54. Ibrahim, A. M., & Bennett, B. (2014). The Assessment of Machine Learning Model Performance for Predicting Alluvial Deposits Distribution. *Procedia Computer Science*, 36, 637–642. <https://doi.org/10.1016/j.procs.2014.09.067>
55. Irwansyah, E., Heryadi, Y., & Santoso Gunawan, A. A. (2020). Semantic Image Segmentation for Building Detection in Urban Area with Aerial Photograph Image using U-Net Models. *2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, 48–51. <https://doi.org/10.1109/AGERS51788.2020.9452773>

56. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
57. Jaynes, C., Riseman, E., & Hanson, A. (2003). Recognition and reconstruction of buildings from multiple aerial images. *Computer Vision and Image Understanding*, 90(1), 68–98. [https://doi.org/10.1016/S1077-3142\(03\)00027-4](https://doi.org/10.1016/S1077-3142(03)00027-4)
58. Ji, S., Wei, S., & Lu, M. (2019). Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *Ieee Transactions on Geoscience and Remote Sensing*, 57(1), 574–586. <https://doi.org/10.1109/tgrs.2018.2858817>
59. Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., & Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, 448, 179–204. <https://doi.org/10.1016/j.neucom.2021.03.035>
60. Jung, S. J., Lee, K., & Lee, W. H. (2022). Object-Based High-Rise Building Detection Using Morphological Building Index and Digital Map. *Remote Sensing*, 14(2). <https://doi.org/ARTN 33010.3390/rs14020330>
61. Kadhim, N., Mourshed, M., & Bray, M. (2016). Advances in remote sensing applications for urban sustainability. *Euro-Mediterranean Journal for Environmental Integration*, 1(1). <https://doi.org/10.1007/s41207-016-0007-4>
62. Kajári, B., Bozán, C., & Van Leeuwen, B. (2023). Monitoring of Inland Excess Water Inundations Using Machine Learning Algorithms. *Land*, 12(1), Article 1. <https://doi.org/10.3390/land12010036>
63. Kampffmeyer, M., Salberg, A. B., & Jenssen, R. (2016). Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. *Proceedings of 29th Ieee Conference on Computer Vision and Pattern Recognition Workshops, (Cvprw 2016)*, 680–688. <https://doi.org/10.1109/Cvprw.2016.90>
64. Karasiak, N., Dejoux, J. F., Monteil, C., & Sheeren, D. (2021). Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*. <https://doi.org/10.1007/s10994-021-05972-1>
65. Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., & Adams, P. (2009). Remote sensing change

- detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote Sensing of Environment*, 113(7), 1382–1396. <https://doi.org/10.1016/j.rse.2008.07.018>
66. Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. arXiv:1412.6980.
 67. Kottek, M., & Yuen, M. L. (2022). Public health risks from asbestos cement roofing. *American Journal of Industrial Medicine*, 65(3), 157–161. <https://doi.org/10.1002/ajim.23321>
 68. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25. <https://doi.org/10.1145/3065386>
 69. Kucharczyk, M., & Hugenholtz, C. H. (2021). Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities. *Remote Sensing of Environment*, 264, 112577. <https://doi.org/10.1016/j.rse.2021.112577>
 70. Kumari, M., Sarma, K., & Sharma, R. (2019). Using Moran's I and GIS to study the spatial pattern of land surface temperature in relation to land use/cover around a thermal power plant in Singrauli district, Madhya Pradesh, India. *Remote Sensing Applications-Society and Environment*, 15. <https://doi.org/ARTN10023910.1016/j.rsase.2019.100239>
 71. Lemenkova, P. (2020). Object Based Image Segmentation Algorithm of SAGA GIS for Detecting Urban Spaces in Yaoundé, Cameroon. *Central European Journal of Geography and Sustainable Development*, 2, 38. <https://doi.org/10.47246/CEJGSD.2020.2.2.4>
 72. Leung, H. K., Chen, X. Z., Yu, C. W., Liang, H. Y., Wu, J. Y., & Chen, Y. L. (2019). A Deep-Learning-Based Vehicle Detection Approach for Insufficient and Nighttime Illumination Conditions. *Applied Sciences-Basel*, 9(22). <https://doi.org/ARTN476910.3390/app9224769>
 73. Li, M., Wu, P., Wang, B., Park, H., Hui, Y., & Yanlan, W. (2021). A Deep Learning Method of Water Body Extraction From High Resolution Remote Sensing Images With Multisensors. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3120–3132. <https://doi.org/10.1109/jstars.2021.3060769>
 74. Likó, S. B., Bekő, L., Burai, P., Holb, I. J., & Szabó, S. (2022). Tree species composition mapping with dimension reduction and post-

- classification using very high-resolution hyperspectral imaging. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-25404-x>
75. Liu, J., Li, T., Xie, P., Du, S., Teng, F., & Yang, X. (2020). Urban big data fusion based on deep learning: An overview. *Information Fusion*, 53, 123–133. <https://doi.org/10.1016/j.inffus.2019.06.016>
 76. Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
 77. Lovász, V., Halász, A., Molnár, P., Karsa, R., & Halmai, Á. (2023). *Application of a CNN to the Boda Claystone Formation for high-level radioactive waste disposal*. *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-31564-1>
 78. Lulla, K., Nellis, M. D., Rundquist, B., Srivastava, P. K., & Szabó, S. (2021). Mission to earth: LANDSAT 9 will continue to view the world. *Geocarto International*, 36(20), 2261–2263. <https://doi.org/10.1080/10106049.2021.1991634>
 79. Lunetta, R. S., & Lyon, J. G. (2004). *Remote Sensing and GIS Accuracy Assessment*. CRC Press.
 80. Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J., & Elith, J. (2018). A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sensing of Environment*, 208, 145–153. <https://doi.org/10.1016/j.rse.2018.02.026>
 81. Ma, L., Liu, Y., Zhang, X. L., Ye, Y. X., Yin, G. F., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *Isprs Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprs.2019.04.015>
 82. Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). *Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark*. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).
 83. Mándi, A., Posgay, M., Vadász, P., Major, K., Rödelberger, K., Tossavainen, A., Ungváry, G., Woitowitz, H. J., Galambos, E., Németh, L., Soltész, I., Egerváry, M., & Böszörményi Nagy, G. (2000). Role of occupational asbestos exposure in Hungarian lung cancer patients. *International Archives of Occupational and*

- Environmental Health*, 73(8), 555–560.
<https://doi.org/10.1007/s004200000172>
84. Manickavasagan, A., Jayas, D. S., & White, N. D. G. (2008). Thermal imaging to detect infestation by *Cryptolestes ferrugineus* inside wheat kernels. *Journal of Stored Products Research*, 44(2), 186–192.
<https://doi.org/10.1016/j.jspr.2007.10.006>
 85. Mannel, S., Price, M., & Hua, D. (2011). Impact of reference datasets and autocorrelation on classification accuracy. *International Journal of Remote Sensing*, 32(19), 5321–5330.
<https://doi.org/10.1080/01431161.2010.498841>
 86. Marceau, D. J., Howarth, P. J., & Gratton, D. J. (1994). Remote-Sensing and the Measurement of Geographical Entities in a Forested Environment .1. The Scale and Spatial Aggregation Problem. *Remote Sensing of Environment*, 49(2), 93–104. [https://doi.org/10.1016/0034-4257\(94\)90046-9](https://doi.org/10.1016/0034-4257(94)90046-9)
 87. Maurer, T. (2013). How to Pan-Sharpen Images Using the Gram-Schmidt Pan-Sharpen Method—A Recipe. *Isprs Hannover Workshop 2013*, 40–1(W-1), 239–244.
 88. McGlinchy, J., Johnson, B., Muller, B., Joseph, M., & Diaz, J. (2019). *Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery*. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium.
 89. Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57, 115–129.
<https://doi.org/10.1016/j.inffus.2019.12.001>
 90. Meyer, H., Reudenbach, C., Wollauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
 91. Miller, H. J. (2004). Tobler’s First Law and Spatial Analysis. *Annals of the Association of American Geographers*, 94(2), 284–289.
<https://doi.org/10.1111/j.1467-8306.2004.09402005.x>
 92. Mohajane, M., Costache, R., Karimi, F., Bao Pham, Q., Essahlaoui, A., Nguyen, H., Laneve, G., & Oudija, F. (2021). Application of remote sensing and machine learning algorithms for forest fire mapping in a

- Mediterranean area. *Ecological Indicators*, 129, 107869. <https://doi.org/10.1016/j.ecolind.2021.107869>
93. Moore, G. K. (1979). What is a picture worth? A history of remote sensing / Quelle est la valeur d'une image? Un tour d'horizon de télédétection. *Hydrological Sciences Bulletin*, 24(4), 477–485. <https://doi.org/10.1080/02626667909491887>
 94. Moran, P. a. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2), 243–251. <https://doi.org/10.1111/j.2517-6161.1948.tb00012.x>
 95. Mou, L. C., Bruzzone, L., & Zhu, X. X. (2019). Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *Ieee Transactions on Geoscience and Remote Sensing*, 57(2), 924–935. <https://doi.org/10.1109/Tgrs.2018.2863224>
 96. Mucsi, L. (2006). A geoinformatika tudomány és a földrajz kapcsolata = Relationship between giscience and geography. In *Táj, környezet és társadalom: Ünnepi tanulmányok Keveiné Bárány Ilona professzor asszony tiszteletére = Landscape, environment and society: Studies in honour of professor Ilona Bárány-Kevei on the occasion of her birthday* (pp. 519–527). <http://acta.bibl.u-szeged.hu/62424/>
 97. Mucsi, L., Liska, C., Henits, L., Tobak, Z., Csendes, B., & Nagy, L. (2017). The evaluation and application of an urban land cover map with image data fusion and laboratory measurements. *Hungarian Geographical Bulletin*, 66, 145–156. <https://doi.org/10.15201/hungeobull.66.2.4>
 98. Nakagawa, S., & Hauber, M. E. (2011). Great challenges with few subjects: Statistical strategies for neuroscientists. *Neuroscience and Biobehavioral Reviews*, 35(3), 462–473. <https://doi.org/10.1016/j.neubiorev.2010.06.003>
 99. Nikparvar, B., & Thill, J. C. (2021). Machine Learning of Spatial Data. *Isprs International Journal of Geo-Information*, 10(9). <https://doi.org/ARTN 600 10.3390/ijgi10090600>
 100. Noor, N. M., Abdullah, A., & Hashim, M. (2018). Remote sensing UAV/drones and its applications for urban areas: A review. *IOP Conference Series: Earth and Environmental Science*, 169(1), 012003. <https://doi.org/10.1088/1755-1315/169/1/012003>
 101. Pan, Z., Xu, J., Guo, Y., Hu, Y., & Wang, G. (2020). Deep Learning Segmentation and Classification for Urban Village Using a

- Worldview Satellite Image Based on U-Net. *Remote Sensing*, 12(10), Article 10. <https://doi.org/10.3390/rs12101574>
102. Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S., Srestasathien, P., & Vateekul, P. (2019). Semantic Segmentation on Remotely Sensed Images Using an Enhanced Global Convolutional Network with Channel Attention and Domain Specific Transfer Learning. *Remote Sensing*, 11(1). <https://doi.org/ARTN 8310.3390/rs11010083>
 103. Pawluszek-Filipiak, K., & Borkowski, A. (2020). On the Importance of Train-Test Split Ratio of Datasets in Automatic Landslide Detection by Supervised Classification. *Remote Sensing*, 12(18). <https://doi.org/ARTN 305410.3390/rs12183054>
 104. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
 105. Peppas, M. V., Bell, D., Komar, T., & Xiao, W. (2018). Urban Traffic Flow Analysis Based on Deep Learning Car Detection from Cctv Image Series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII–4, 499–506. <https://doi.org/10.5194/isprs-archives-XLII-4-499-2018>
 106. Perez, L., & Wang, J. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*. arXiv:1712.04621.
 107. Phinzi, K., Abriha, D., Bertalan, L., Holb, I., & Szabó, S. (2020). Machine Learning for Gully Feature Extraction Based on a Pan-Sharpener Multispectral Image: Multiclass vs. Binary Approach. *ISPRS International Journal of Geo-Information*, 9(4), Article 4. <https://doi.org/10.3390/ijgi9040252>
 108. Phinzi, K., Abriha, D., & Szabó, S. (2021a). Classification Efficacy Using K-Fold Cross-Validation and Bootstrapping Resampling Techniques on the Example of Mapping Complex Gully Systems. *Remote Sensing*, 13(15), Article 15. <https://doi.org/10.3390/rs13152980>
 109. Phinzi, K., Holb, I., & Szabó, S. (2021b). Mapping Permanent Gullies in an Agricultural Area Using Satellite Images: Efficacy of Machine Learning Algorithms. *Agronomy*, 11(2), Article 2. <https://doi.org/10.3390/agronomy11020333>

110. Ploton, P., Mortier, F., Rejou-Mechain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pelissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11(1). <https://doi.org/ARTN 4540 10.1038/s41467-020-18321-y>
111. Podani, J. (2000). *Introduction to the Exploration of Multivariate Biological Data*. Backhuys Publishers.
112. Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, 31(10), 2001–2019. <https://doi.org/10.1080/13658816.2017.1346255>
113. Raczko, E., Krówczyńska, M., & Wilk, E. (2022). Asbestos roofing recognition by use of convolutional neural networks and high-resolution aerial imagery. Testing different scenarios. *Building and Environment*, 217, 109092. <https://doi.org/10.1016/j.buildenv.2022.109092>
114. Rastogi, K., Bodani, P., & Sharma, S. A. (2020). Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*, 37(5), 1501–1513. <https://doi.org/10.1080/10106049.2020.1778100>
115. Rice, L., Wong, E., & Zico Kolter, J. (2020). *Overfitting in adversarially robust deep learning*. arXiv:2002.11569.
116. Riffat, S., Powell, R., & Aydin, D. (2016). Future cities and environmental sustainability. *Future Cities and Environment*, 2(0). <https://doi.org/10.1186/s40984-016-0014-2>
117. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *Isprs Journal of Photogrammetry and Remote Sensing*, 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
118. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
119. Rose, R. A., Byler, D., Eastman, J. R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J.,

- Horning, N., Kaplin, B. A., Laporte, N., Leidner, A., Leinagruber, P., Morissette, J., Musinsky, J., Pintea, L., ... Wilkie, D. (2015). Ten ways remote sensing can contribute to conservation. *Conservation Biology*, 29(2), 350–359. <https://doi.org/10.1111/cobi.12397>
120. Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., & Jung, J. (2014). Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *Isprs Journal of Photogrammetry and Remote Sensing*, 93, 256–271. <https://doi.org/10.1016/j.isprsjprs.2013.10.004>
 121. Roy, M. (2009). Planning for sustainable urbanisation in fast growing cities: Mitigation and adaptation issues addressed in Dhaka, Bangladesh. *Habitat International*, 33(3), 276–286. <https://doi.org/10.1016/j.habitatint.2008.10.022>
 122. Schlosser, A. D., Szabó, G., Bertalan, L., Varga, Z., Enyedi, P., & Szabó, S. (2020). Building Extraction Using Orthophotos and Dense Point Cloud Derived from Visual Band Aerial Imagery Based on Machine Learning and Segmentation. *Remote Sensing*, 12(15). <https://doi.org/ARTN 2397 10.3390/rs12152397>
 123. Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, 109–120. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>
 124. Shalaby, A., & Tateishi, R. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt, Appl. *Applied Geography*, 27, 28–41. <https://doi.org/10.1016/j.apgeog.2006.09.004>
 125. Shen, L. Y., Peng, Y., Zhang, X. L., & Wu, Y. Z. (2012). An alternative model for evaluating sustainable urbanization. *Cities*, 29(1), 32–39. <https://doi.org/10.1016/j.cities.2011.06.008>
 126. Shi, Y., Li, Q., & Zhu, X. X. (2020). Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 184–197. <https://doi.org/10.1016/j.isprsjprs.2019.11.004>
 127. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). <https://doi.org/ARTN 60 10.1186/s40537-019-0197-0>

128. Siqueira, L. F. S., Araújo Júnior, R. F., de Araújo, A. A., Morais, C. L. M., & Lima, K. M. G. (2017). LDA vs. QDA for FT-MIR prostate cancer tissue classification. *Chemometrics and Intelligent Laboratory Systems*, 162, 123–129. <https://doi.org/10.1016/j.chemolab.2017.01.021>
129. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
130. Strahler, A. H., Woodcock, C. E., & Smith, J. A. (1986). On the nature of models in remote sensing. *Remote Sensing of Environment*, 20(2), 121–139. [https://doi.org/10.1016/0034-4257\(86\)90018-0](https://doi.org/10.1016/0034-4257(86)90018-0)
131. Sun, H. S., Wu, J. Q., & Xu, X. C. (2016). A Method to Determine Appropriate Spatial Resolution for Hard Image Classification. *Journal of the Indian Society of Remote Sensing*, 44(1), 11–19. <https://doi.org/10.1007/s12524-015-0471-3>
132. Szabó, G., Bertalan, L., Barkóczi, N., Kovács, Z., Burai, P., & Lénárt, C. (2018). Zooming on Aerial Survey. In G. Casagrande, A. Sik, & G. Szabó (Eds.), *Small Flying Drones: Applications for Geographic Observation* (pp. 91–126). Springer International Publishing. https://doi.org/10.1007/978-3-319-66577-1_4
133. Szabó, S., Burai, P., Kovács, Z., Szabó, G., Kerényi, A., Fazekas, I., Paládi, M., Buday, T., & Szabó, G. (2014). Testing algorithms for the identification of asbestos roofing based on hyperspectral data. *Environmental Engineering and Management Journal*, 13. <https://doi.org/10.30638/eemj.2014.323>
134. Szabó, S., Gácsi, Z., & Balázs, B. (2016). Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Landscape & Environment*, 10(3–4), Article 3–4. <https://doi.org/10.21120/LE/10/3-4/13>
135. Taherzadeh, E., & Shafri, H. Z. M. (2013). Development of a Generic Model for the Detection of Roof Materials Based on an Object-Based Approach Using WorldView-2 Satellite Imagery. *Advances in Remote Sensing*, 2(4), Article 4. <https://doi.org/10.4236/ars.2013.24034>
136. Taubenböck, H., Wegmann, M., Roth, A., Mehl, H., & Dech, S. (2009). Urbanization in India—Spatiotemporal analysis using remote sensing

- data. *Computers Environment and Urban Systems*, 33(3), 179–188. <https://doi.org/10.1016/j.compenvurbsys.2008.09.003>
137. Tharwat, A. (2016). Linear vs. quadratic discriminant analysis classifier: A tutorial. *International Journal of Applied Pattern Recognition*, 3, 145. <https://doi.org/10.1504/IJAPR.2016.079050>
 138. Tran, T. D. B., Puissant, A., Badariotti, D., & Weber, C. (2011). Optimizing Spatial Resolution of Imagery for Urban Form Detection-The Cases of France and Vietnam. *Remote Sensing*, 3(10), 2128–2147. <https://doi.org/10.3390/rs3102128>
 139. Tu, J., & Xia, Z.-G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of The Total Environment*, 407(1), 358–378. <https://doi.org/10.1016/j.scitotenv.2008.09.031>
 140. United Nations, Department of Economic and Social Affairs, Population Division (2019). World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). New York: United Nations
 141. Underwood, E. C., Ustin, S. L., & Ramirez, C. M. (2007). A comparison of spatial and spectral image resolution for mapping invasive plants in coastal California. *Environmental Management*, 39(1), 63–83. <https://doi.org/10.1007/s00267-005-0228-9>
 142. Vadivambal, R., Vellaichamy, C., Jayas, D., & White, N. (2010). Detection of Sprout-Damaged Wheat Using Thermal Imaging. *Applied Engineering in Agriculture*, 26, 999–1004. <https://doi.org/10.13031/2013.35900>
 143. Varga, O. G., Kovács, Z., Bekő, L., Burai, P., Csátráiné Szabó, Z., Holb, I., Ninsawat, S., & Szabó, S. (2021). Validation of Visually Interpreted Corine Land Cover Classes with Spectral Values of Satellite Images and Machine Learning. *Remote Sensing*, 13(5), Article 5. <https://doi.org/10.3390/rs13050857>
 144. Wang, H., & Miao, F. (2022). Building extraction from remote sensing images using deep residual U-Net. *European Journal of Remote Sensing*, 55(1), 71–85. <https://doi.org/10.1080/22797254.2021.2018944>
 145. Wang, L., Fang, S., Meng, X., & Li, R. (2022). Building Extraction With Vision Transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11. <https://doi.org/10.1109/TGRS.2022.3186634>

146. Wang, Q., Yuan, Z. H., Du, Q., & Li, X. L. (2019). GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *Ieee Transactions on Geoscience and Remote Sensing*, 57(1), 3–13. <https://doi.org/10.1109/Tgrs.2018.2849692>
147. Wang, S., Hou, X., & Zhao, X. (2020). Automatic Building Extraction From High-Resolution Aerial Imagery via Fully Convolutional Encoder-Decoder Network With Non-Local Block. *IEEE Access*, 8, 7313–7322. <https://doi.org/10.1109/access.2020.2964043>
148. Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., & Eklund, P. (2016). A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(3), 271–282. <https://doi.org/10.1016/j.jtte.2016.05.005>
149. Weir, N., Lindenbaum, D., Bastidas, A., Etten, A., Mcpherson, S., Shermeyer, J., Kumar, V., & Tang, H. (2020). SpaceNet MVOI: A Multi-View Overhead Imagery Dataset. *Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision*.
150. Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>
151. Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117, 34–49. <https://doi.org/10.1016/j.rse.2011.02.030>
152. Wina, Herwindiati, D. E., & Isa, S. M. (2014). Robust discriminant analysis for classification of remote sensing data. *2014 International Conference on Advanced Computer Science and Information System*, 454–458. <https://doi.org/10.1109/ICACISIS.2014.7065892>
153. 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)
154. Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., ... Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127–147. <https://doi.org/10.1016/j.rse.2019.02.015>

155. Wulder, M., & Boots, B. (1998). Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, 19(11), 2223–2231. <https://doi.org/Doi 10.1080/014311698214983>
156. Xu, Y. Y., Wu, L., Xie, Z., & Chen, Z. L. (2018). Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sensing*, 10(1). <https://doi.org/ARTN 144 10.3390/rs10010144>
157. Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83. <https://doi.org/10.1109/IACC.2016.25>
158. Yan, C., Fan, X., Fan, J., & Wang, N. (2022). Improved U-Net Remote Sensing Classification Algorithm Based on Multi-Feature Fusion Perception. *Remote Sensing*, 14(5). <https://doi.org/10.3390/rs14051118>
159. Ying, X. (2019). An Overview of Overfitting and its Solutions. *2018 International Conference on Computer Information Science and Application Technology*, 1168. <https://doi.org/Artn 022022 10.1088/1742-6596/1168/2/022022>
160. You, Y. F., Wang, S. Y., Ma, Y. X., Chen, G. S., Wang, B., Shen, M., & Liu, W. H. (2018). Building Detection from VHR Remote Sensing Imagery Based on the Morphological Building Index. *Remote Sensing*, 10(8). <https://doi.org/ARTN 1287 10.3390/rs10081287>
161. Yu, M., Chen, X., Zhang, W., & Liu, Y. (2022). AGs-Unet: Building Extraction Model for High Resolution Remote Sensing Images Based on Attention Gates U Network. *Sensors*, 22(8). <https://doi.org/10.3390/s22082932>
162. Yuan, Q. Q., Shen, H. F., Li, T. W., Li, Z. W., Li, S. W., Jiang, Y., Xu, H. Z., Tan, W. W., Yang, Q. Q., Wang, J. W., Gao, J. H., & Zhang, L. P. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241. <https://doi.org/ARTN 111716 10.1016/j.rse.2020.111716>
163. Yuan, X. H., Shi, J. F., & Gu, L. C. (2021). A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169. <https://doi.org/ARTN 114417 10.1016/j.eswa.2020.114417>

164. Zhang, C. S., Luo, L., Xu, W. L., & Ledwith, V. (2008). Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of the Total Environment*, 398(1–3), 212–221. <https://doi.org/10.1016/j.scitotenv.2008.03.011>
165. Zhang, L. P., Zhang, L. F., & Du, B. (2016). Deep Learning for Remote Sensing Data A technical tutorial on the state of the art. *Ieee Geoscience and Remote Sensing Magazine*, 4(2), 22–40. <https://doi.org/10.1109/Mgrs.2016.2540798>
166. Zhou, D. F., Fremont, V., Quost, B., Dai, Y. C., & Li, H. D. (2017). Moving object detection and segmentation in urban environments from a moving platform. *Image and Vision Computing*, 68, 76–87. <https://doi.org/10.1016/j.imavis.2017.07.006>
167. Zhou, W., & Troy, A. (2008). An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing*, 29(11), 3119–3135. <https://doi.org/10.1080/01431160701469065>
168. 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>