

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
INFORMATIKAI KAR

Geológiai képződmények térképezése műholdképek alapján a Hátszegi- medencében

DIPLOMAMUNKA
TÉRKÉPÉSZ MESTERSZAK

Készítette:

Árvai Valentin

Témavezető:

Dr. Albert Gáspár

Egyetemi docens

ELTE Térképtudományi és Geoinformatikai Tanszék



EÖTVÖS LORÁND TUDOMÁNYEGYETEM
INFORMATIKAI KAR
TÉRKÉPTUDOMÁNYI ÉS GEOINFORMATIKAI TANSZÉK

DIPLOMAMUNKA TÉMABEJELENTŐ

Hallgató adatai:

Név: Árvai Valentin

Neptun kód: I4HK6U

Képzési adatok:

Szak: térképész, mesterképzés (MA/MSc)

Tagozat: Nappali

Belső témavezetővel rendelkezem

Témavezető neve: Dr. Albert Gáspár

munkahelyének neve: ELTE IK Térképtudományi és Geoinformatikai Tanszék

munkahelyének címe: 1117, Budapest, Pázmány Péter sétány 1/A.

beosztás és iskolai végzettsége: Egyetemi docens, PhD

A diplomamunka címe: Geológiai képződmények térképezése műholdképek alapján a Hátszegi-medencében

A diplomamunka témája:

(A témavezetővel konzultálva adja meg 1/2 - 1 oldal terjedelemben diplomamunka témájának leírását)

A műholdas távérzékeléssel nyert adatok ingyenes hozzáférhetősége (Landsat 2009), illetve spektrális felbontásának növekedése (Sentinel) nyomán megnőtt a tudományos közösség érdeklődése ezen adatok felhasználása iránt. Többek között a földtudományokban is olyan úttörő cikkek születtek, amelyek multispektrális felvételeket használtak adatforrásként (Van der Werff, H., & Van der Meer, 2016), ugyanis ezekkel lehetőség nyílik különböző földtani jellemzők kinyerésére. A kinyert adatok felhasználhatók geológiai formák, például réteg határok vagy tektonikus törések kiemelésére, illetve a felszínen található anyagok meghatározására is. Ez utóbbi főként a növényzeti fedettséggel nem rendelkező területeken használható eredményesen. Növényzettel fedett területeken a vegetációs időszakon kívüli időpontokban készített felvételek összesített elemzése és képjavító eljárások alkalmazása vezethet eredményre.

Jelen diplomamunka célja a Hátszegi medencében található Válioara település környékének földtani vizsgálata multispektrális felvételek alapján. A munka során elsősorban a növényzeti fedettség kiküszöbölésének lehetőségét kell megvizsgálni olyan paraméterek (indexek) számításakor, amelyek korábbi terepi felmérés során feltárt rétegek kimutatására alkalmasak. A paraméterek számításához elsősorban az európai Copernicus programban jelenleg is aktív Sentinel-2, valamint az amerikai Landsat földmegfigyelő műholdak felvételei lennének felhasználva.

Források:

Van der Werff, H., & Van der Meer, F.: Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing. Remote sensing, 8(11), 883 (2016).

Budapest, 2020.11.27.

Tartalomjegyzék

1. BEVEZETÉS	3
1.1. A mintaterület	4
2. ADATFORRÁSOK, SPEKTRÁLIS ALAPOK, OSZTÁLYOZÁS	5
2.1. Spektrális alapok és egyéb paraméterek.....	5
2.2. Műholdfelvételek osztályozásának alapjai.....	8
2.3. Geológiai alap.....	9
2.4. Sentinel-2	10
3. ADATOK FELDOLGOZÁSA	13
3.1. Folyamatsor.....	13
3.2. Geológiai térkép digitalizálás.....	13
3.3. Műholdképek beszerzése, előfeldolgozása	16
3.4. NDVI.....	18
3.5. Sáv kalkulációk és NDVI maszkok.....	18
3.6. Osztályozás paraméterei	23
4. EREDMÉNYEK	26
4.1. Legkisebb távolság.....	27
4.2. Legnagyobb valószínűség	28
4.3. Véletlen erdő	29
5. EREDMÉNYEK KIÉRTÉKELÉSE	30
5.1. NDVI 0,3.....	31
5.2. NDVI 0,35	33
5.3. Összesítés	35
6. ÖSSZEGZÉS	37
7. SUMMARY	38
HIVATKOZÁSOK	39
KÖSZÖNETNYILVÁNÍTÁS	41
ÁBRAJEGYZÉK	42
MELLÉKLETEK	43

1. Bevezetés

A geológiai felmérések - a tudományág kifejlődése óta - elsődleges adatforrásként használatosak az természeti erőforrások kutatása terén. A jó geológiai térkép alap elvárás egy terület nyersanyagainak kiaknázása céljából. A terepi felmérés azonban nem minden esetben kivitelezhető. A sűrű növényzettel fedett, illetve sivatagi területek megközelítése nagy kihívást okoz, gyakran kivitelezhetetlen. Gyakori jelenség a nehéz terepen végzett geológiai térképek esetében az rétegek futásának extrapolálási pontatlansága is.

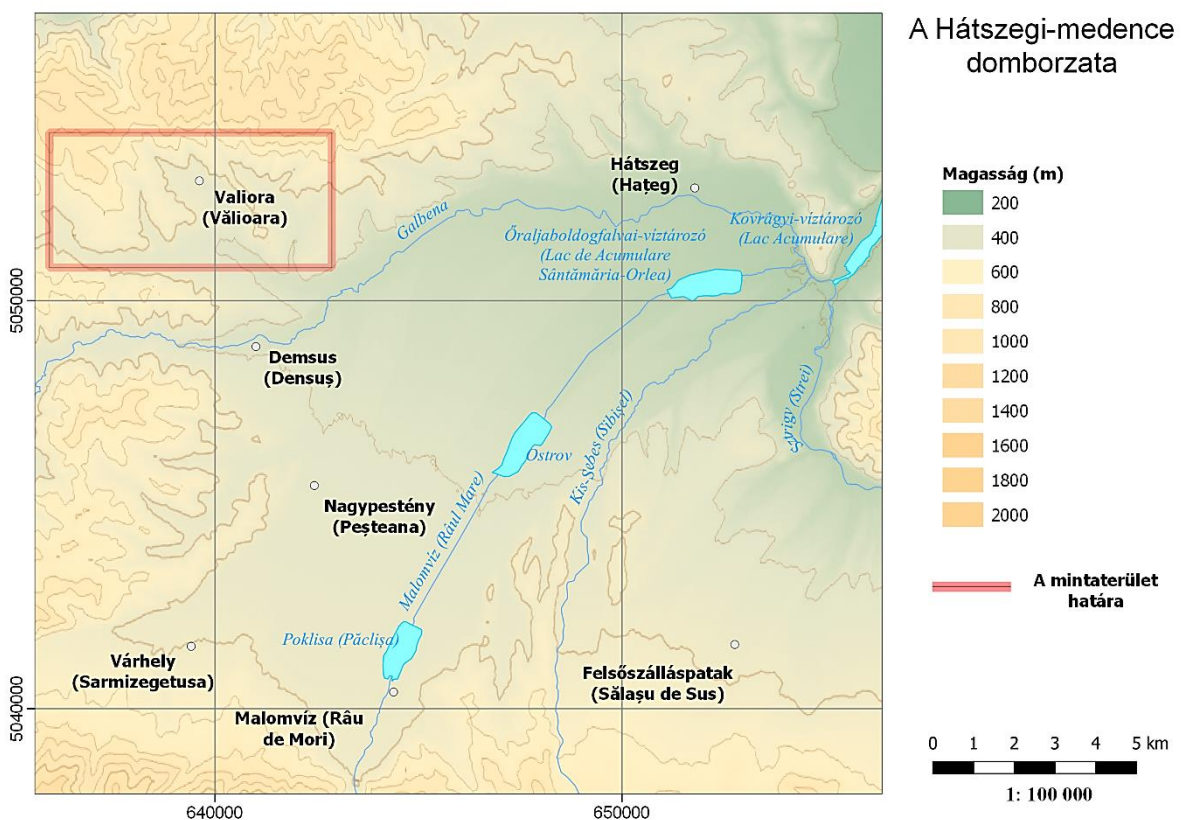
A szakterületen nagy változást hozott a műholdak megjelenése, amelyek a nehezen megközelíthető helyekről is tudnak információt szolgáltatni. Míg az első műholdak egy sávon tudtak szürkeárnyalatos képet alkotni, egy évtizeddel később megjelentek a multispektrális szenzorral felszerelt műholdak is (Landsat; ERTS-1). A kezdeti Landsat műholdak négy spektrális sávot érzékelő szenzorral operáltak. A látható zöld (0.5-0.6 μm), látható vörös (0.6-0.7 μm), közeli infravörös (0.7-0.8 μm), illetve egy további közeli infravörös (0.8-1.1 μm) sávban tudtak felvételeket készíteni (Kramer, 1994). A szenzorhoz mérten először a közeli infravörös tartományban észlelhető ásványokat (réz) tartalmazó kőzetek kutatása kezdődött (Schmidt, 1976, Smith, Green, Robinson, & Honey, 1978). Azonban a nagy áttörést a műhold felvétel alapú ásvány kutatásban a „Terra Mission” kódnevű projekt eredményezte. A projekt keretében az EOS/AM-1 (felbocsátás – 1999.12.18) műholdra többek közt felkerült az ASTER nevezetű szenzor is, amely 14 spektrális tartományban képes egyidejűleg felvételt készíteni.

Míg kezdetben a földmegfigyelő műholdak tervezése, gyártása, pályára állítása, üzemeltetése az állami szférába tartozott (Landsat), később fokozatosan megjelentek a kommerciális (Spot, WorldView-2-4) és az internacionális finanszírozású műhold programok (Kopernikusz program). A Kopernikusz program az Európai Bizottság által irányított, és az Európai Unió tagállamok által finanszírozott földmegfigyelési program. A program alapvetései közt szerepel az adatok ingyenes hozzáférhetősége, az életminőségbéli, gazdasági és tudományos előrehaladás céljából. A program két adatgyűjtő komponensből áll. Az operáló műholdakból, illetve a helyi-érzékelőkből (földi). Ennek keretében működik a Sentinel műholdcsalád, melyek közül multispektrális szenzorral a Sentinel-2 (A-B) jelű műholdak rendelkeznek. Emellett a család 4 további indexel ellátott taggal rendelkezik. Az 1, (2), 3, 5p, 6 jelű műholdak operálnak jelenleg, típusonként 1, illetve 2 (A-B) műholdból álló konstellációval a jobb időbeli lefedettség érdekében. A Kopernikusz program keretében az Európai Unió célkitűzése alapján 2030-ra 30, köztük több különböző műszerrel (indexel) ellátott műhold lesz földkörüli pályán (A Kopernikusz program, 2021).

Jelen dolgozatban a Sentinel-2 műholdak adatainak felhasználásával készült osztályozó modell létrehozását mutatom be, amellyel a későbbiekben is viszonylag jó pontossággal lehet osztályozni a nehezen térképezhető területeket. A távérzékelte adatok legmegfelelőbb kiértékelése érdekében több osztályozó algoritmust hasonlítottam össze, amelyeket a következő fejezetekben mutatok be. Az adatok feldolgozásához a QGIS, illetve SNAP (ESA) szoftvereket használtam. A térképekhez az OpenStreetMap vektoros anyagait használtam (tavak, folyók), a domborzati réteg SRTM alapú.

1.1.A mintaterület

A dolgozat mintaterületét az Erdély Nyugati határvidékén található Hátszegi-medence (Țara Hațegului) lett kiválasztva. Az ELTE TTK-IK jelenleg is részt vesz a területen zajló nemzetközi kutatásban, főként a Valiora-völgyben (*Vălioara*) található gazdag dinoszaurusz lelőhelyek feltárása céljából. Az elemzésre kijelölt szűkebb értelemben vett Hátszegi-medence geológiai és vegetáció tekintetében is változatos.



1. ábra. A mintaterület (UTM N34Z)(OSM adatok alapján)

A körülbelül 20km * 20km-es terület medence jellegéből fakadóan a központi területek folyóvízi üledék töltötték fel. A medence keleti oldalát a Sztrigy (Strei) folyó formálja, amelybe a Kis-Sebes, Malomvíz (Râu Mare), valamint a Galbena folyók torkoltnak. A területre esik

néhány nagyobb tó is, amelyeken a Malomvíz folyik keresztül. Ezek az Ostrov, a Poklisa (Päclişa), Óraljaboldogralvai-víztározó, Kovrágyi-víztározó (1.ábra).

A medencét körülvevő hegységek kétféleképp oszthatók fel. Míg keleti, illetve déli irányból a Déli-Kárpátok határolják, nyugat, észak-nyugati irányból a használt felosztástól függően Déli vagy Bánsági-hegyvidék. Az egyik gyakran használt felosztás a Hátszegi-medencét nyugatról határoló Ruzska-havas hegycsoportot a Bánsági-hegyvidék részeként a Déli-Kárpátoktól különállónak tekintik (Magyar Nemzeti Atlasz), (Mezei, és mtsai., 2018). Más felosztás szerint a Ruzska-havas csoport a Bánsági-hegyvidékkel a Déli-Kárpátokhoz sorolható (Gábris, és mtsai., 2014).

A medencét főként üledékes kőzetek alkotják, amelyek a perm-holocén közti, valamint a késő-kréta időszakokban rakódtak le. Ez utóbbi nagy mennyiségben tartalmaz dinoszaurusz, illetve emlős maradványokat. A környező főként kristályos alapkőzetű hegységek még az üledék lerakódás előtt emelkedtek ki (Botfalvai, és mtsai., 2021). Az üledékes kőzetek mellett a medencében különböző magmás, illetve metamorf kőzetek is megtalálhatók.

2. Adatforrások, spektrális alapok, osztályozás

Az elemzéshez a Sentinel-2 műholdak által készített felvételek, valamint a területről készült korábbi geológiai felmérések szolgálták alapul. A fejezetben a források ismertetésén túl a távérzékelés spektrális alapjai, a használt algoritmusok, illetve szoftverek is bemutatásra kerülnek.

2.1. Spektrális alapok és egyéb paraméterek

A távérzékelte adatokat szenzorok által érzékelt jel forrása alapján két fő csoportba soroljuk. Így léteznek aktív, illetve passzív típusú szenzorok (2. ábra). Az aktív csoportba tartozó felvételt készítő egységek az általuk kibocsátott meghatározott frekvenciájú és hullámhosszú jel visszaverődését érzékelik. Ebbe a csoportba tartoznak a szintetikus apertúrájú radarok (SAR – Synthetic Aperture Radar). A Kopernikusz program Sentinel-1, 3, illetve 6-os sorszámmal ellátott tagjai rendelkeznek valamilyen aktív távérzékelésen alapuló szenzorral.

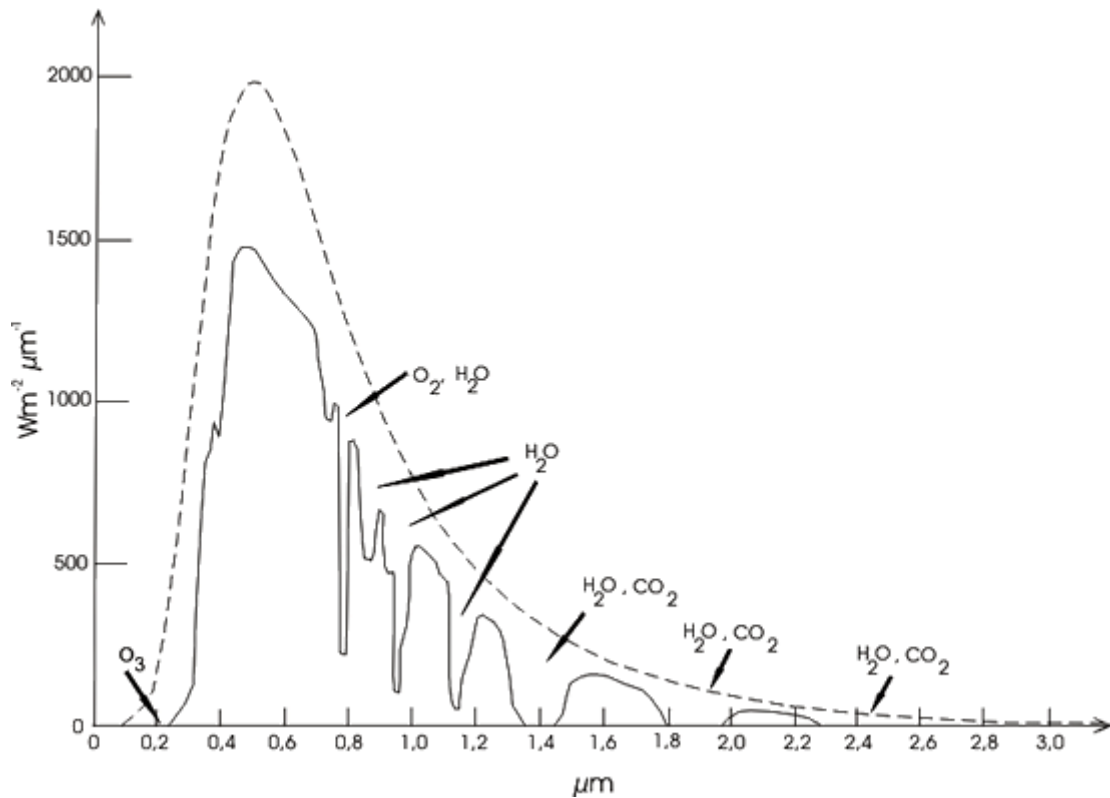
A passzív típusú szenzorok a napfény visszaverődését érzékelik különböző hullámhosszokon. Az elektromágneses spektrum minél több sávjában érzékel az adott műszer, annál nagyobb spektrális felbontása. Így léteznek multispektrális, illetve az ezeket is meghaladó spektrális felbontású hiperspektrális eszközök (Vágó, Seres, & Hegedűs, 2011).



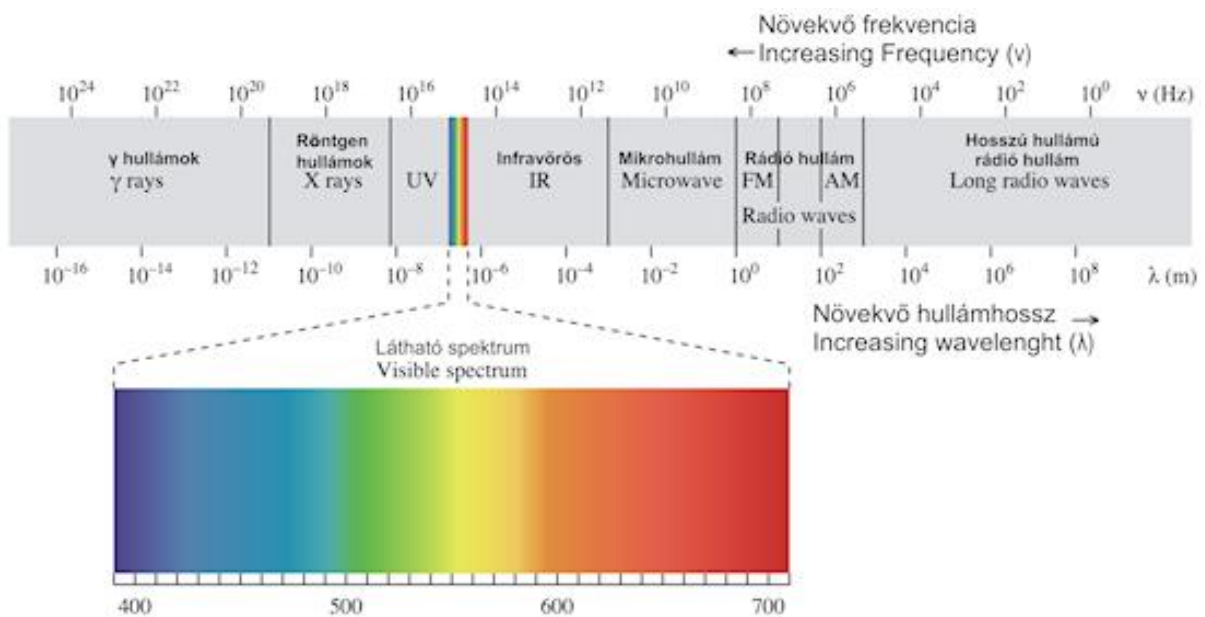
2. ábra Passzív és aktív távérzékelés (A to Z About Active and Passive Remote Sensing, 2017)

A passzív távérzékelő szenzor az adott területről visszaverődő elektromágneses sugárzás intenzitását érzékelve alkotja a képet. Az intenzitás értéket adott hullámhosszon a különböző hőmérsékletű abszolút fekete test hőmérsékleti sugárzásával modellezzük. A távérzékelés esetén az 6000K hőmérsékletű (nap) fekete test sugárzását vesszük alapul. Az abszolút fekete test olyan ideális objektum, ami bármely hullámhosszú elektromágneses sugárzást elnyel. (Mucsi, 2013)

A föld légköre azonban különböző hullámhossz tartományokban elnyeli a nap által sugárzott energiát (3. ábra). A legnagyobb elnyelők a sztratoszférában található ózon alacsonyabb hullámhosszokon, illetve troposzférai vízgőz a közeli infravörös tartományban. Ennek okán az abszorbeált energiájú hullámhosszokon az intenzitás alacsony lenne, így a távérzékelő műholdak spektrális sávjait az intenzitás függvényében választják meg. A legtöbb sávot a látható fény ($0,4 \mu\text{m} - 0,8 \mu\text{m}$), valamint a közeli infravörös tartomány első harmadában ($0,8 \mu\text{m} - 1,0 \mu\text{m}$) határozzák meg (4. ábra). Az adott tartományokban a visszaverődés mértékét a felszín tulajdonságai is befolyásolják. Erre épül a műholdfelvételek felhasználása az ásványi nyersanyag kutatásban is (Mucsi, 2013).



3. ábra Földi energia spektrum, illetve a referencia abszolút fekete test sugárzás a légkör felső részén (szaggatott) (Vágó, Seres, & Hegedűs, 2011)



4. ábra Az elektromágneses spektrum felosztása (Vágó, Seres, & Hegedűs, 2011)

A spektrális felbontáson túl a felvételek további fontos jellemzője a térbeli felbontás. A térbeli felbontás adja meg a felvétel egy pixele által lefedett terület méretét. Minél több pixel fed le egységnyi földdarabot, annál jobb a felvétel térbeli felbontása.

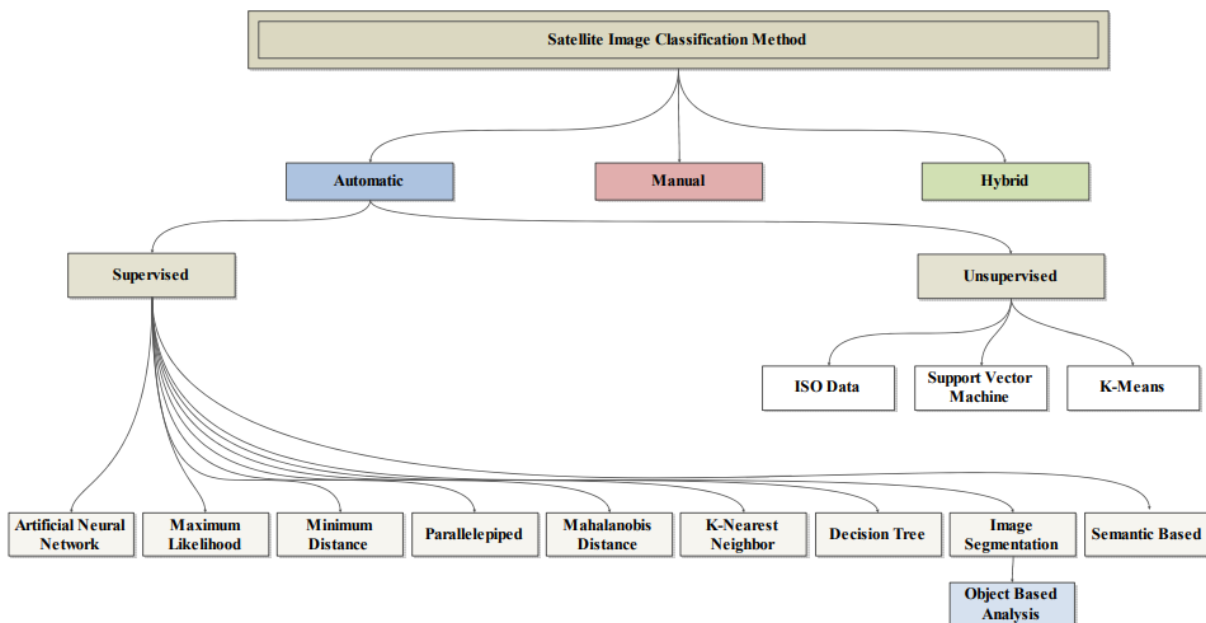
Szintén fontos paraméter a radiometrikus felbontás, ami a műszer intenzitás érzékenységét jelöli. A nagyobb radiometrikus felbontású szenzor kisebb intenzitású visszaverődést is képes érzékelni.

A negyedik felbontás, amivel a műholdas távérzékelési egységeket jellemezni lehet, az időbeli felbontás. Ez az érték adja meg azt az időt, ami eltelik addig, amíg a műhold az adott földdarabról újra felvételt készít (visszatérési idő).

2.2. Műholdfelvételek osztályozásának alapjai

A műholdfelvételek osztályozására különböző osztályozó metódusokat használhatunk (5. ábra). A csoportosításuk legfelső szintje alapján három kategóriába sorolhatók. Vannak automatikus, manuális, illetve hibrid megoldások. Manuális minden olyan módszer, amely nem egy előre meghatározott algoritmus alapján sorolja csoportokba az egyes pixeleket. Az automatikus módszerek esetében a pixeleket a meghatározott osztályozó eljárás sorolja csoportokba. A hibrid módszerek e kettő jó tulajdonságainak ötvözete.

A következő szint az automatikus csoport alatt az irányított, illetve irányítatlan osztályozás. Az irányított osztályozás alapja, hogy előre meghatározott számú, típusú csoportot használunk, amelyekhez külön tanító adatok is tartoznak. A tanító adatok két részre oszlanak, egy részük az



5. ábra Műholdképek osztályozási metódusainak hierarchiája (Abburu & Golla, 2015)

osztályozó modell tanítására szolgál, másik részük a modell ellenőrzésére/pontosságának javítására.

Az irányítatlan osztályozási módszereknél az osztályok „címkézése” az csoportosítás után történik. Így ebben az esetben nincsenek előre meghatározott csoportba tartozó tanító pixelek. Leggyakoribb módszerek az ISO Data, a Support Vector Machine (SVM), K-Means.

Az Irányított osztályozást további három típusra lehet bontani, léteznek a mesterséges neurális háló (Artificial Neural Network – ANN) alapú módszerek, amelyek az emberi agy tanulási módját utánozzák. Továbbá a „bináris döntési fa” (Binary Decision Tree – BDT) alapú módszerek, amelyek szabály alapú rétegek alapján határozzák meg az egyes pixelek csoportját. A harmadik kiemelt csoport a „kép szegmentálás” (Image Segmentation) alapú algoritmusok csoportja, amelyek nem elsősorban az osztályozással, hanem a minta felismeréssel kapcsolatosak. (Abburu & Golla, 2015)

Az irányított osztályozási eljárások esetében a létrehozott modell pontosságát a tanuló pixelek alapján különböző arányokkal lehet jellemezni. Ezek az arányok a tanító pixelek számából eredő

- helyes pozitív (TP) – Az adott pixelt a megfelelő csoportba sorolja,
- fals pozitív (FP) – Az adott pixelt a nem a megfelelő csoportba sorolja,
- helyes negatív (TN) – Az adott pixelt nem az adott csoportba sorolta helyesen,
- fals negatív (FN) – Az adott pixelt nem az adott csoportba sorolta helytelenül

értékekből számolhatók ki. Ezen az értékek osztályonként határozhatók meg. Ilyen arány a pontosság (accuracy), precizitás (precision), korreláció (correlation, Matthews correlation coefficient – MCC), hibaráta (error rate) (Braun, 2020).

Pontosság képlete (accuracy):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Precizitás képlete (precision):

$$PPV = \frac{TP}{TP + FP}$$

Korreláció (Matthews correlation coefficient):

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

2.3. Geológiai alap

A diplomamunka geológiai alapjául a Romániai Földtani Intézet által kiadott 1: 50 000-es méretarányú térképeket használtam (Harta Geologică a României la scara 1:50.000). A kutatási területre két szelvény esik, amelyek két különböző térképezési időszakban készültek. A

meghatározott terület nagyobb részét a Hátszegi (Hațeg) (Lupu, és mtsai., 1993) szelvény fedi le, míg a terület nyugati pár száz méterrel átlóg a Bautari (Băuțar) (Georgescu, Vlad, Lupu, & Maier, 1979) szelvényre. A Bautari szelvény 1979-ben készült az azonosítója „105a Băuțar L-34-94-A”, míg a Hátszegi szelvény 1993-ban, azonosítója „105b Hațeg L-34-094-B”. Ez utóbbi, további különbséggént egy felújított szelvény, így jelkulcsi elemekben is van eltérés. Mindkét szelvény több geológiai felmérési alaplól lett összeállítva. Ezek a 20. század második felében zajlottak.

A területről készült 1:200.000-es méretarányú szelvény elérhető online vektoros formában Román Geológiai Intézet oldalán¹, azonban az 50 ezres szelvények ilyen formában nem megtekinthetők. Ezen szelvényeket a Dr. Albert Gáspár tanár úr bocsátotta rendelkezésemre georeferált raszteres állományok formájában.

2.4.Sentinel-2

A műholdak egyenként 1,2 tonna súlyúak. A tervezett élettartamuk 87 hónap (7,25 év). Az „A” és „B” jelű eszközöket 2015-ben, illetve 2017-ben állították poláris pályára. Emellett mindkettő 786km körüli átlag magasságban kering napszinkron pályán. Ennek köszönhetően egy adott terület felett mindig azonos időben haladnak el. Így a fény–árnyék hatások mindig ugyanazok, megkönnyítve ezzel a felvételek elemzését. Az Európai űrügynökség három pontban határozta meg a széria céljait:

- Nagy felbontású, multispektrális adatgyűjtés globálisan, rövid visszatérési idővel (nagy időbeli felbontással),
- adat folytonosság biztosítása a SPOT és Landsat műholdak kiesésével,
- megfigyelési adatok az operatív termékek következő generációjához, mint például a felszínfedettség-térképek, a földváltozások kimutatására szolgáló térképek és a geofizikai változók.

A Sentinel-2 műholdrendszert az Astrium GmbH (Németország) által vezetett ipari konzorcium fejlesztette ki. Az Astrium SAS (Franciaország) felelős a többspektrumú műszerért (MSI). A felvételeket az MSI (Multi-spectral Instrument) szenzor készíti, összesen 13 spektrális sávon (1. táblázat) és 290x290 négyzetkilométeres területen.

¹ <https://geoportal.igr.ro/viewgeol200k.php>

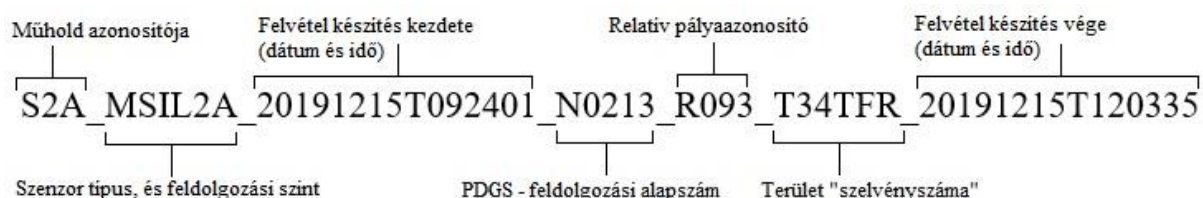
1. táblázat Sentinel-2 spektrális sávok

	Leírás	Közép hullámhossz (µm)	Térbeli felbontás (m)
Sáv 1	Parti aeroszol	0,443	60
Sáv 2	Kék	0,490	10
Sáv 3	Zöld	0,560	10
Sáv 4	Vörös	0,665	10
Sáv 5	Vegetáció, vörös él	0,705	20
Sáv 6	Vegetáció, vörös él	0,740	20
Sáv 7	Vegetáció, vörös él	0,783	20
Sáv 8	NIR	0,842	10
Sáv 8A	Vegetáció, vörös él	0,865	20
Sáv 9	vízgőz	0,945	60
Sáv 10	SWIR – Cirrus	1,375	60
Sáv 11	SWIR	1,610	20
Sáv 12	SWIR	2,190	20

A küldetés zavartalan működését és az adatok minőségét a projekt földi szegmense biztosítja, amely felelős az űrszondák vezérléséért, az adatok letöltéséért, feldolgozásáért, valamint közreadásáért.

A felvételek feldolgozottsági szintjét tekintve két adattípus elérhető, az „1C”, illetve a „2A”. Az „1C” típus a nyers adat, vagyis a légköri korrekció nélküli TOA (Top Of the Atmosphere) felvétel, míg a „2A” a BOA (Bottom Of the Atmosphere), vagyis a légköri korrekcióval ellátott, így a felszínre viszonyított adat (ESA, 2017).

Minden felvételhez tartozik egy egyedi azonosító, ami alapján egyértelműen kereshetők. Ennek mintája a 6. ábrán látható.



6. ábra Sentinel-2 felvétel egyedi azonosító minta

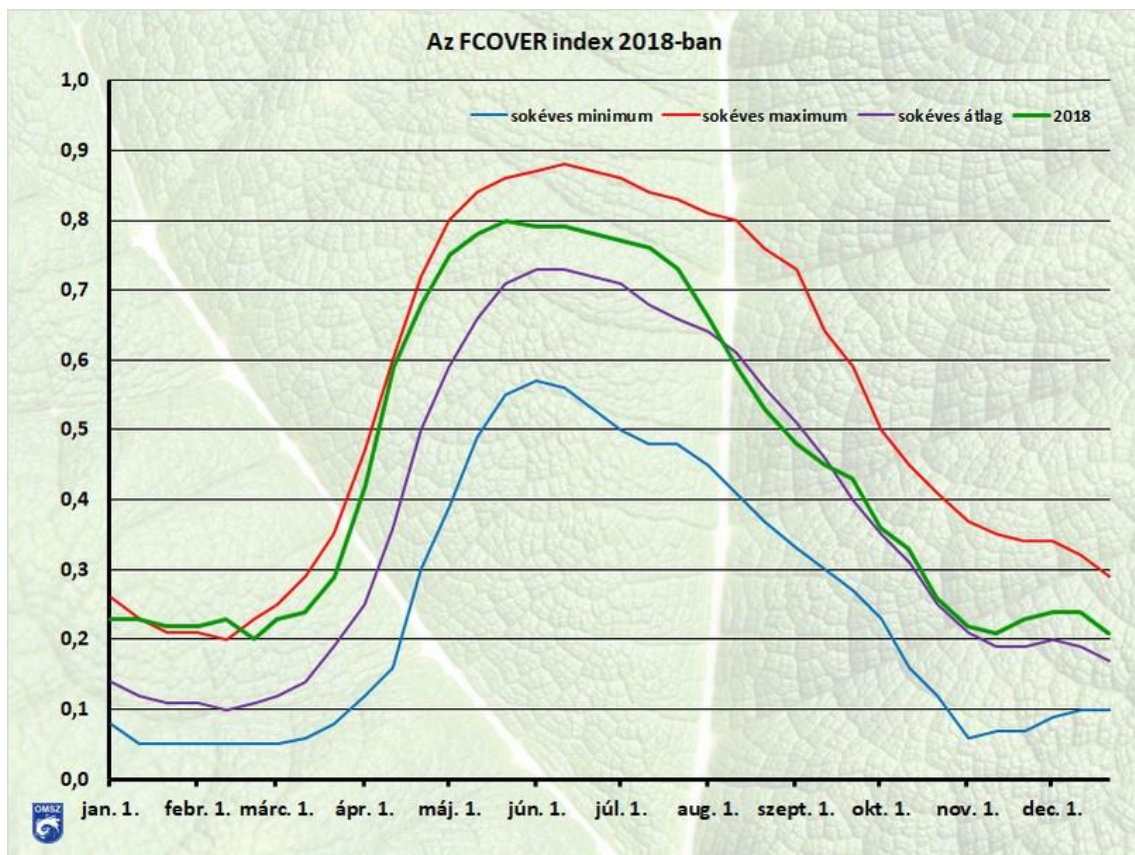
A műhold azonosítója jelzi, hogy a két sentinel-2 műholdból álló konstelláció melyik tagja (A-B) készítette a felvételt. A kódnév második tagja jelzi a szenzor típusát (MSI), illetve a

feldolgozottsági szintet (L2A – Level-2A). A következő tag a felvétel készítés kezdetének időpontja, az a másodperc, amelyben az adott műholdképet az aktuális szatellit elkezd felvenni. A PDGS címke (Payload Data Ground Segment – az adatáramlás földi szegmense) jelzi, a lesugárzás óta az adaton történt változtatásokat, elvégzett folyamatok sorát. A PDGS számot követi a relatív pályaaazonosító index, majd a felszín sajátos rendszerben vett, darabjának azonosítója (pld. T34TFR = Nyugat-Erdély egy része). Az azonosítószám utolsó eleme a felvétel elkészítésének a végét jelzi „ÉÉÉÉHHNN T ÓÓPPMM” formátumban, ahogy a harmadik tag is.

Osztályozandó adatként a Sentinel-2 műholdak több felvételét használtam, amelyek közt szerepel az A, illetve a B jelű műhold kép is. Az idősoros adatok több mintát jelentenek az osztályozás során, így jobb eredményt lehet elérni kisebb területek feldolgozásánál is. Címke szerint a

- S2A_MSIL2A_20191215T092401_N0213_R093_T34TFR_20191215T120335,
- S2A_MSIL2A_20210101T093411_N0214_R136_T34TFR_20210101T122900,
- S2B_MSIL2A_20210225T093029_N0214_R136_T34TFR_20210225T121212

felvételeket.



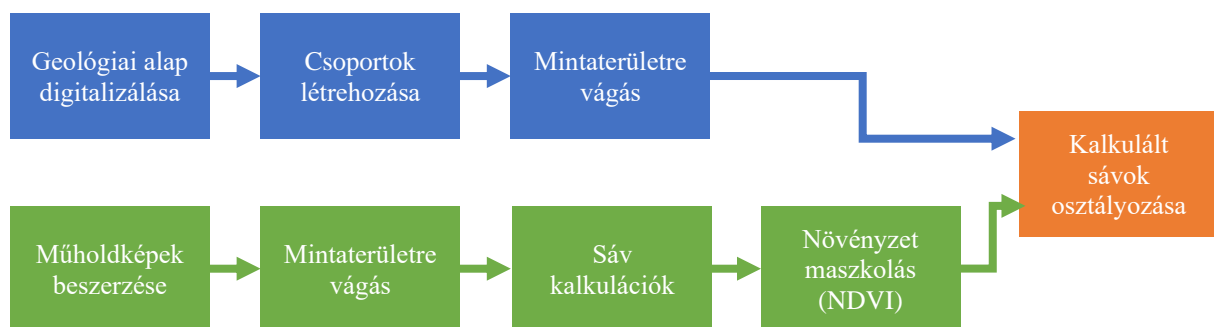
7. ábra Az FCOVER index sokéves átlaga Magyarországra viszonyítva (Kovács & Erdődiné Molnár, 2019)

A vegetációs időszakon kívüli képeket használtam, a lehető legkisebb növényzeti fedettség okozta „zaj” elkerülése érdekében. Többéves átlag alapján a legkisebb zöld növényzet által fedett terület (a mintaterületen) november, illetve március közötti időszakra esik. A zöld növényzet által fedett terület indikátora az „FCOVER” index. A 7. ábrán az OMSZ (Országos Meteorológiai Szolgálat) által készített sokéves FCOVER statisztika Magyarországra viszonyítva. Mivel a mintaterület azonos klímazónába tartozik, így az ábrán látható értékek jelen esetben relevánsak. (Kovács & Erdődiné Molnár, 2019)

3. Adatok feldolgozása

3.1. Folyamatsor

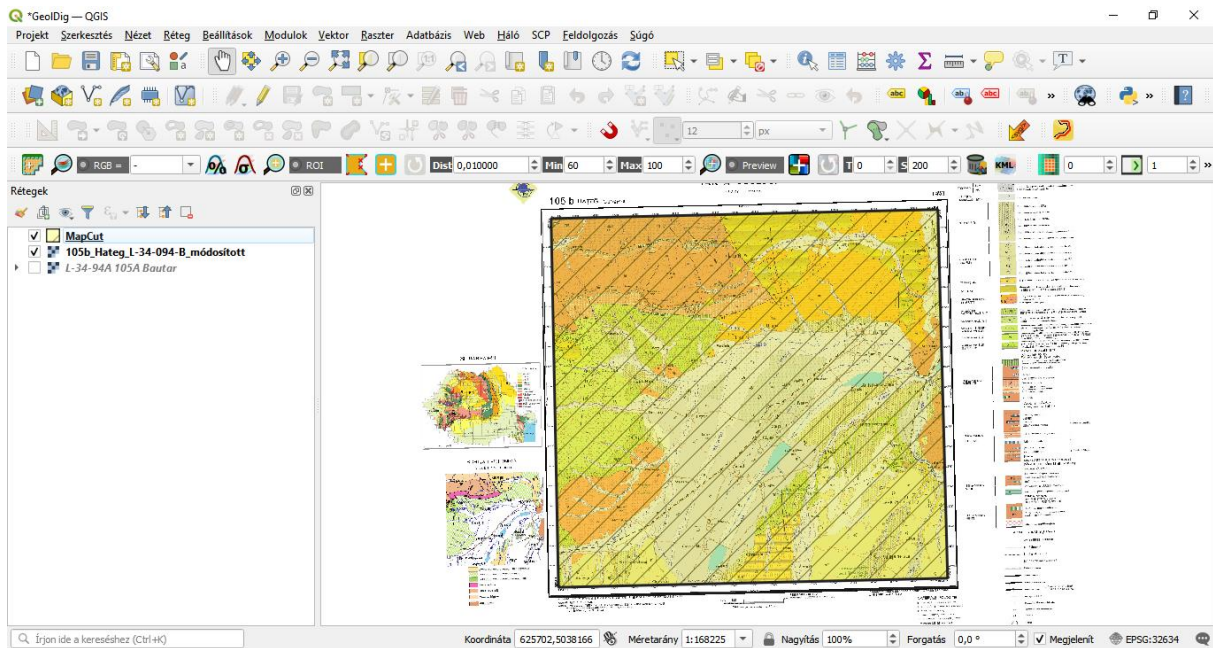
Az adatok feldolgozása két szálon történt (8. ábra). A tanulóterületek digitalizálásával egyidejűleg készítettem el a letöltött felvételek sávjaiból a különböző indexeket. A két szál az utolsó lépésben fonódik össze, az osztályozásnál.



8. ábra A folyamatsor

3.2. Geológiai térkép digitalizálás

Az osztályozáshoz szükséges vektoros anyagot a bemutatott geológiai térképek digitalizálásával készítettem el. Ehhez a QGIS térinformatikai szoftver 3.14-es („pi”) verzióját használtam. Első lépésként importáltam a georeferált geológiai térképeket, majd raszter menüpont kivonat almenüjének „Raszter vágása terjedelemmel” eszközzel levágtam a térképek térképkereten kívüli részeit. Ehhez a projekt koordinátarendszerét a területnek megfelelő UTM zónára állítottam, valamint készítettem egy vektoros réteget, amivel kivágtam a szelvény térképkereten belül található részét (9. ábra – „Hátszeg”). Ezt a folyamatsort végrehajtottam a „Bautar” című szelvényre is.



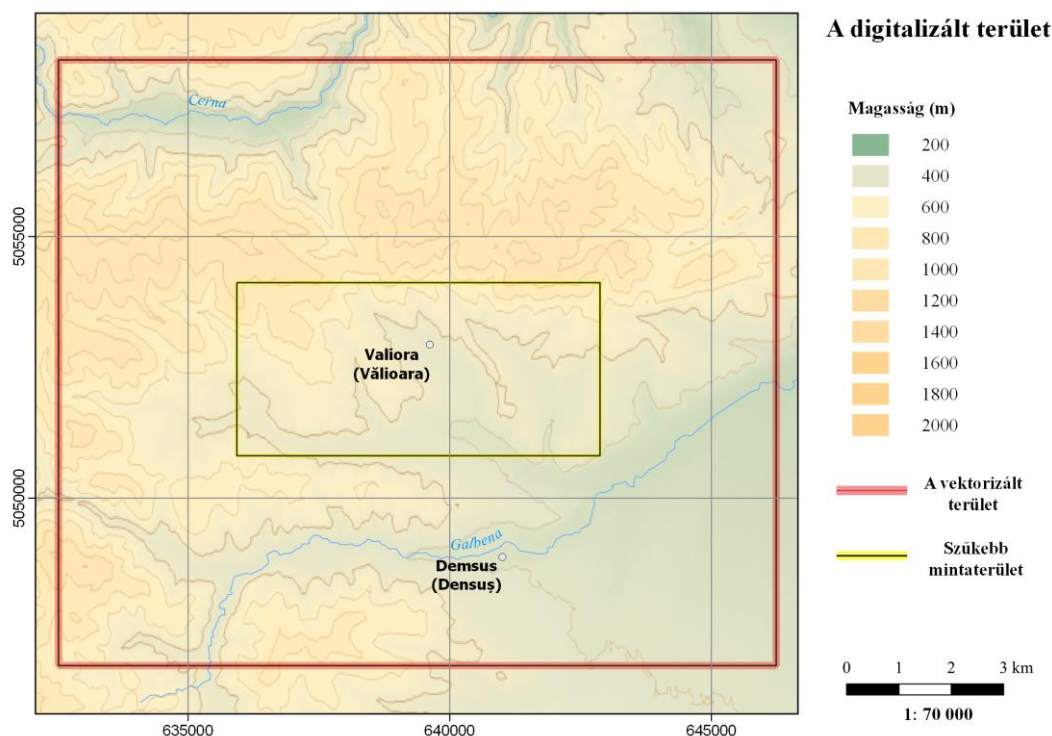
9. ábra Raszter vágása a térkép szélénél

A felvételek vágása után a következő lépés az összevonásuk volt. Mivel a két szelvény két különböző szín tárolási móddal kaptam (Bautar – PCT, Hátszeg – RGB).

A „Bautar” című szelvényt PCT-ből RGB típusú rendszerbe konvertáltam a raszter menü Konverzió menüpontjának „PCT-ből RGB-be” funkcióval. Az „Input réteg” lenyíló menüben kell a megfelelő réteget kiválasztani, majd az alatta lévő lenyíló menüben a sávot.

Amint mindkét szelvény azonos szín tárolási rendszerbe került, összevontam őket a QGIS „Összevon” funkciójával. Paraméterként a két szelvényt az „Input rétegek” alatt lehet kiválasztani, míg az adattípust egy lenyíló menüben a panel közepén. Minden feldolgozási lépésnél lehetőség van exportálni az elkészült raszteres állományt. A térképek összevonását követően kiválasztottam a bővebb értelemben vett mintaterületet, majd ezt a területet kivágtam az összevont térképlapokról.

A térképi alap előkészítésével a következő lépés az osztályok meghatározása volt. Témavezetőm, Dr. Albert Gáspár tanár úr javaslata alapján 18 osztályt alakítottam ki a felszíni kőzetek alapján. A negyedidőszaki folyóvízi üledékeket összevontam, így ezek egy kategóriát alkotnak. Ezen túl minden kőzettípus külön kategóriába került (2. táblázat).

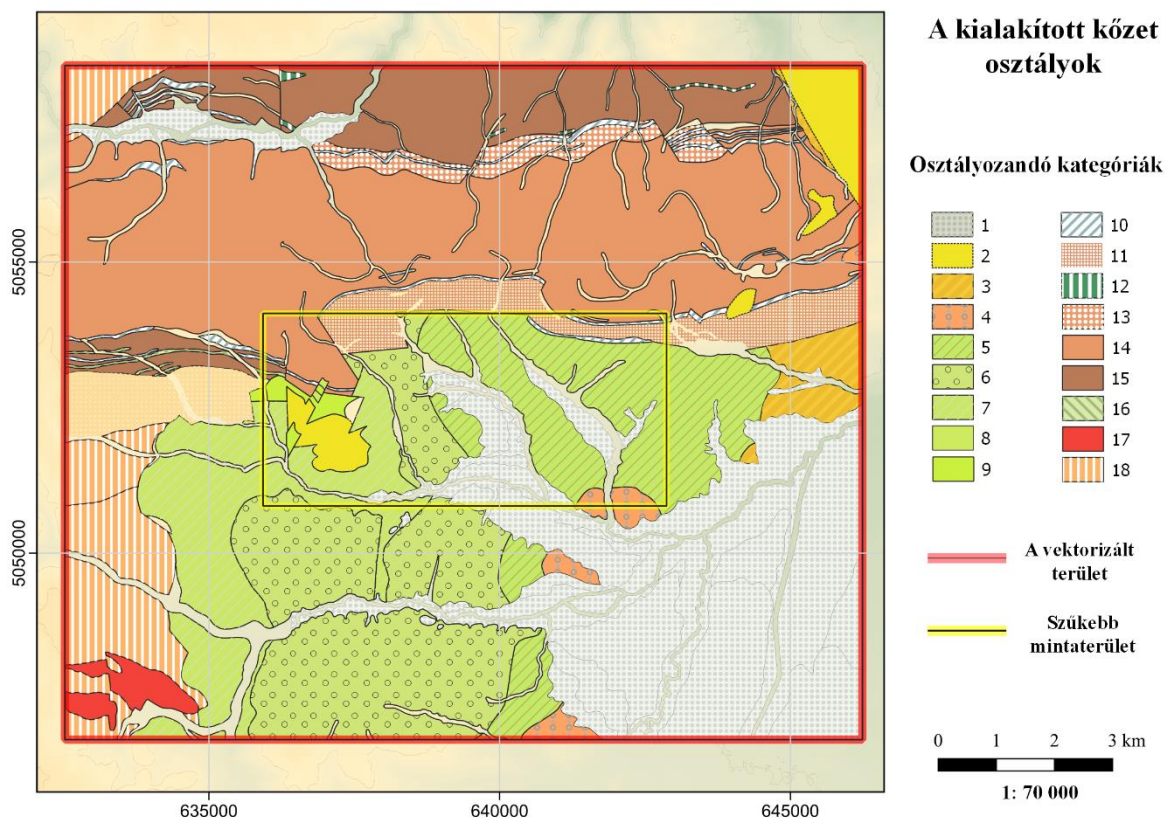


10. ábra Bővebb és szűkebb mintaterület (UTM N34Z) (OSM adatok alapján)

2. táblázat Kialakított kőzet osztályok indexek és közettípus alapján

Index	Közzettípus	Osztály
qh ₂	lejtőtörmelék, hordalékkúp törmelék	1
qh ₁	terasz lerakódás 2-4 méteres magasságban	1
qh ₁ ^{pr}	proluviális törmelék	1
qp ₃ -qh	árvíz üledékek	1
qpf ₇	terasz lerakódás 15-20 méteres magasságban	1
m ₂ ^l	márgák, homokok és kövületes meszes homokkövek, tufa közberétegzések – Delinești formáció	2
Pg-m ₁	szürke agyagok, vörös homok és konglomerátumok	3
Pg	homok és kavics	4
ma ₂ Pg ₁	konglomerátum, breccsa, durva homokkövek, lila és fekete aleurit széntartalmú közberétegzésekkel	5
ma ₂	konglomerátum, homokkő, aleurit, vulkanoklasztitok, lávák és andezit tufák	6
cp ₂ -ma ₁	márgák/agyagok + - aleurit, homokkövek, homokos kalkarenitek és mikrokonglomerátumok. flis fácies, uralkodóan disztális. (Pui Formáció)	7
cp ₂	márga / agyag +- aleurit, homokkő és meszes homokkő, finom- és középszemcsés rétegekben. disztális flis üledék	8
st-cp ₂	márgák / agyagok, mészmárga és márgás, aleuritos, szürke, fehér és helyileg kvarcos homokkő. Selfperemi és hemipelágikus fáciesek	9
c	karbonátpala + dolomit	10
gn	gneisz	11
a	amfibolit	12
sq _t	kvarcit/kvarcit rendszerek biotittal	13
sscl	szericites-kloritos palák	14
msgr	gránátos (alamandinos) csillámpala	15
msst	gránát és sztaurolit	15
mtb	bázikus metatufit	16
Pc ₃ A	granitoid	17
pgnbi	paragneisz és biotit	18

A csoportok kialakítása után elkezdtem a térkép digitalizálást. Ehhez külön vektoros réteget hoztam létre, amelyben három különböző mezőt különítettem el. Így lett egy oszlop, amiben a kőzettípusok indexei szerepelnek, egy, amiben a kőzettípusok sorszámait találhatók koruk alapján, illetve az utolsó, amiben a csoportosítási indexek szerepelnek.



11. ábra A kialakított kőzet osztályok a mintaterületeken (megnevezések lásd 2. táblázat)(UTM N34Z)

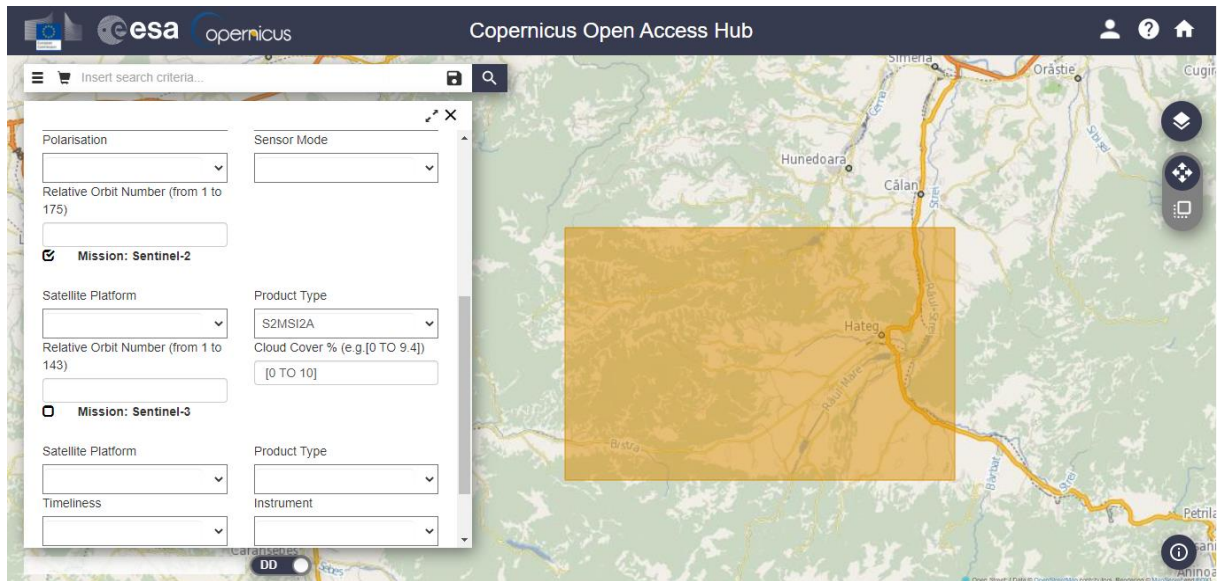
Az elkészült vektoros állomány a 11. ábrán látható. A jelmagyarázatban az „Osztályozandó kategóriák,” alatt a csoportosítási osztályok jelkulcsa található. Ezt a 2. táblázatban lehet párosítani a megfelelő kőzettípussal, illetve a táblázatban láthatók azok a kőzettípusok is, amelyek nem képeznek külön osztályozási kategóriát.

3.3. Műholdképek beszerzése, előfeldolgozása

A felvételeket az ESA által üzemeltetett, a Kopernikusz program számára létrehozott weboldalról²(Copernicus Open Access Hub) ingyenesen szereztem be. Ezen az oldalon lehetőség van kiválasztani a jelenleg operáló műholdak felvételeinek különböző tulajdonságait,

² <https://scihub.copernicus.eu/dhus/#/home>

úgy mint a műhold konstelláció tagjait, a felvétel feldolgozási szintet, felhő borítottságát, pályaszámot, érzékelési időszakot (12. ábra).



12. ábra A Kopernikusz program adatelérési felülete (Copernicus Open Acces Hub)

A mintaterület az oldal központi részén található térképen választható ki. Jobb kattintással felület pontok tehetők le, a jobb egér gombot nyomva tartva téglalap alakú terület választható ki. A paramétereket megadva a keresés gombra kattintva az oldal kilistázza az eredményeket. Amiket közvetlenül a letöltés gombra kattintva lehet beszerezni. A 2.4-es fejezetben található a letöltött felvételek listája.

A fájlokat .zip formátumban kapjuk meg, amikben a felvételeken kívül különböző metaadatok találhatóak. A felvételek .tiff formátumban a könyvtárstruktúra „Granule”- Felvétel azonosító – „IMG_DATA” mappában találhatóak. A műhold különböző sávjain különböző térbeli felbontások miatt a sávok három különböző mappába kerülnek ez alapján. Ezek az „R10m”, „R20m”, „R60m” mappák. Az „R” betű az angol „Resolution” (felbontás) szóra utal, az utána következő számok pedig a térbeli felbontásra méterben kifejezve. Az egyre rosszabb felbontású mappákban a jobb felbontású sávok is megtalálhatók a mappa nevének megfelelő felbontással. A műholdképek letöltése és egységes formában hozása után ezeket a geológiai térképhez hasonlóan a mintaterületre vágtam. Azzal a különbséggel, hogy a terület ez esetben nem lógott túl a kijelölt zónán, így nem volt szükség összevonásokra. A kivágást a korábban leírt módon a QGIS szoftverben hajtottam végre

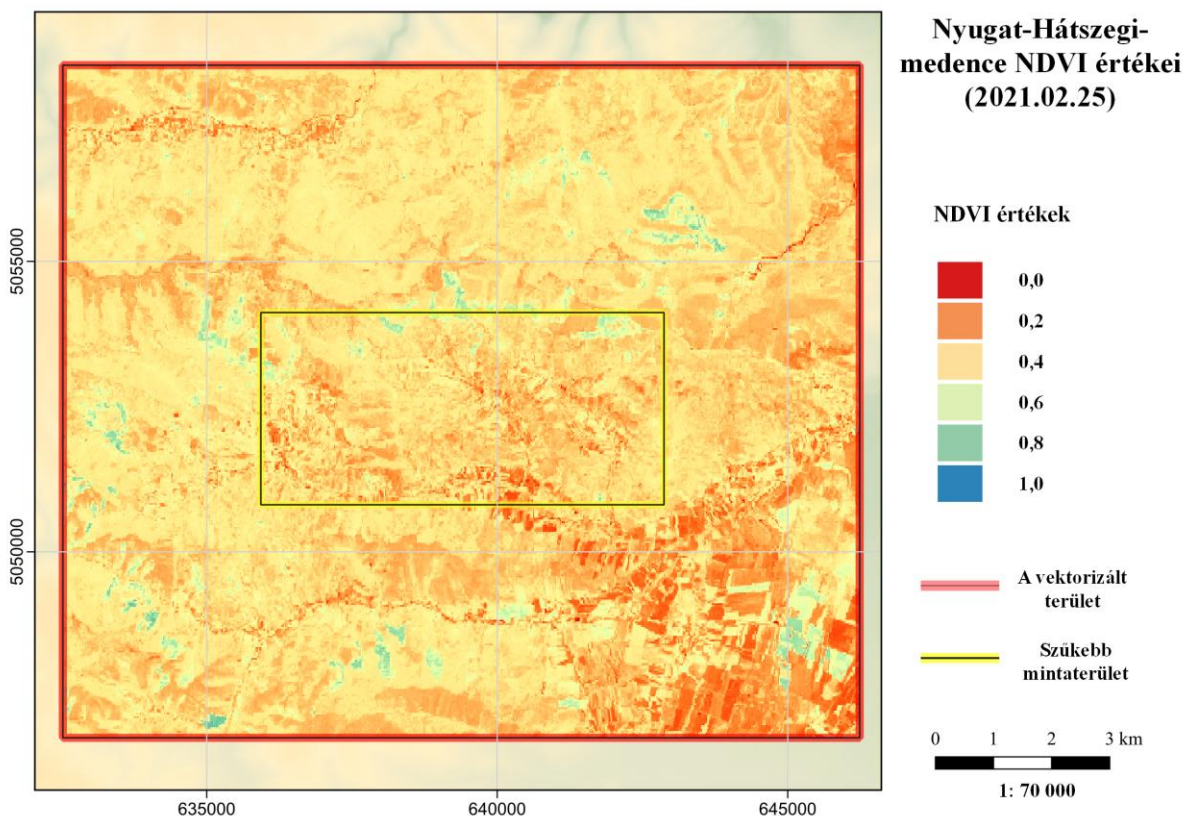
Minden esetben a legjobb térbeli felbontású 10m-es felvételeket használtam. A kalkulációhoz szükséges 20m felbontású sávokat (11-es, illetve 12-es) pedig a QGIS programban található „Raszterek igazítása” függvényvel interpoláltam az azonos felbontás érdekében. Az interpoláláshoz a beépített „Bilinear” interpolációs módszert használtam.

3.4. NDVI

A távérzékelés szempontjából fontos tényező, a vegetáció eloszlása a 13. ábrán látható. A vegetáció felméréséhez az úgynevezett NDVI-t (Normalized Difference Vegetation Index; Normalizált Vegetációs Index) használtam. Az NDVI a növényzet fotoszintetikus aktivitását írja le. Dimenzió nélküli arány, így nincs mértékegysége. Minden pixel -1 és 1 közötti értéket vehet fel. Kiszámításához szükséges a közeli infravörös, illetve a vörös spektrális sávon készült felvétel (Kovács & Erdődiné Molnár, 2019).

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Így a felvételek feldolgozásának következő (párhuzamos) lépése a letöltött, kivágott felvételek területére az NDVI kiszámítása volt. A Sentinel-2 esetében a szükséges közeli infravörös sávnak 8-as felel meg, míg a vörösnek a 4-es. Így a képletbe behelyettesítve a 13. ábrán látható eredményhez körülbelül hasonlókat kaptam minden kiválasztott felvételen.

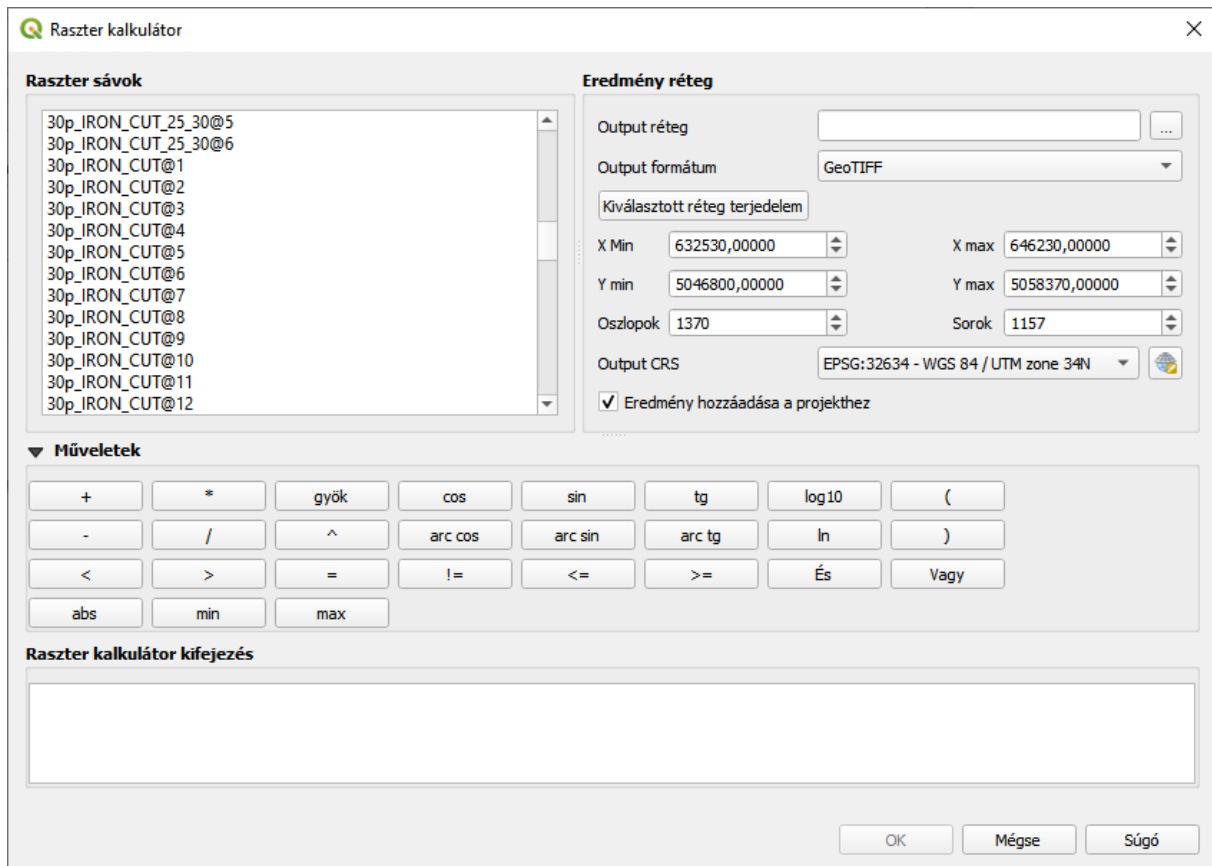


13. ábra A Hátszegi-medence nyugati részének NDVI indexe(2021.02.25) (UTM N34Z)

3.5. Sáv kalkulációk és NDVI maszkok

A műholdképek feldolgozásának első lépésében a geológiai térképekhez hasonlóan a bővebb mintaterületre kellett vágni a kiválasztott felvételeket. Ezt műveletet szintén a QGIS vágó

eszközével hajtottam végre. A különböző sáv kalkulációkat is ebben a szoftverben, ennek „raszter kalkulátor” nevezetű eszközével készítettem el, ami a raszter menüben található. Ezzel a funkcióval készítettem el az előző alfejezetben bemutatott NDVI értékeket is.



14. ábra Raszter kalkulátor

A raszter kalkulátor négy részből áll (14. ábra). A ball felső sarokban található a projekt fájlban található rétegek sávonként szétválasztva. Mellette jobbra található az exportálási lehetőség és annak beállításai. Itt is van opció terület vágásra, illetve a felbontás állítására az oszlopok (pixelek száma horizontálisan) és a sorok (pixelek száma vertikálisan) számának átírásával. Vetület választás (váltás) is lehetséges (Output CRS). Ezek alatt helyezkedik el műveletek panel, amelyen azok az alap matematikai műveletek található, amiket ez a funkció kezelni képes. A raszter kalkulátor kifejezés panelen kell megadni a kalkulálandó képletet a megfelelő változókkal (sávokkal) kiegészítve.

A geológiai térképről digitalizált felszíni kőzetek különböző mértékben tartalmaznak ásványokat. A diplomamunka szempontjából a vas ásványok érdekesek. Főként azért, mert ideális környezetben a vas ásványokban gazdag kőzetek jól kimutathatók és elkülöníthetők a multispektrális műholdfelvételeken. Az általam kiválasztott vas ásványok a (van der Werff & van der Meer, 2016) tanulmányban, az ASTER, illetve Landsat műholdak adat kontinuitás

vizsgálatában meghatározott vas alkotók voltak. Az általuk meghatározott indikátor sávok képleteit használtam (15. ábra).

Feature	ASTER	Landsat 5 TM	Landsat 8 OLI	Sentinel-2A MSI
TM Ratios				
Hydroxyl bearing alteration	4/{5,6,7}	5/7	6/7	11/12
All iron oxides	–	3/1	4/2	4/2
Ferrous iron oxides	2/4	3/5	4/6	4/11
ASTER Iron				
Ferric Iron, Fe ³⁺	2/1	3/2	4/3	4/3
Ferrous Iron, Fe ²⁺	5/3 + 1/2	7/4 + 2/3	7/5 + 3/4	12/8 + 3/4
Laterite	4/5	5/7	6/7	11/12 [†]
Gossan	4/2	5/3	6/4	11/4
Ferrous silicates ‡	5/4	7/5	7/6	12/11 [†]
Ferric oxides	4/3	5/4	6/5	11/8
ASTER Silicates				
Alteration	4/5	5/7	6/7	11/12 [†]
ASTER Other				
Vegetation	3/2	4/3	5/4	8/4
NDVI *	(3 – 2)/(3 + 2)	(4 – 3)/(4 + 3)	(5 – 4)/(5 + 4)	(8 – 4)/(8 + 4)

15. ábra Kalkulált vas ásványok főbb összetevőinek képletei (van der Werff & van der Meer, 2016)

A táblában látható ASTER Iron fül alatti Sentinel-2A MSI képleteket. Kiolvasható a táblázatból, hogy a „vas3” (Ferric Iron, Fe³⁺) legkifejezőbben a 4-es és a 3-as sávok arányában mutatkozik meg. Így a „vas2” (Ferrous Iron, Fe²⁺) a $\frac{12}{8} + \frac{3}{4}$ képlettel, a laterit a $\frac{11}{12}$ sávok arányából, a gossan a $\frac{11}{4}$ arányból, a vas szilikátok a $\frac{12}{11}$ arányból, illetve a vas-oxid a $\frac{11}{8}$ arányból számítható ki (15. ábra).

Ezeket az arányokat a korábban spektroszkópiával meghatározott vizsgáltból nyert ASTER sávok alapján határozták meg. Hozzá kell tenni azonban, hogy a Sentinel sávok a legtöbb esetben nagyobb tartományt fednek le, így egyértelmű hozzárendelés csak néhány esetben lehetséges (van der Werff & van der Meer, 2016).

A sávok kiszámítása után a létrejövő 3x6 sávból (három felvétel, hat ásvány alkotó) összevontam a már bemutatott „összevonás” eszközzel. Így egy vas ásvány összetevőkből álló, idősoros raszter lett az eredmény, amit az osztályozó algoritmusok egyben képesek feldolgozni.

Még az osztályozás előtt indokolt a műholdképenként elkészített NDVI – növényzeti index – alapján történő maszkolás. Erre azért van szükség, mert a magas/sűrű növényzetről visszaverődő jelek a geológiai osztályozás szempontjából zajként jelennek meg. Ennek okán az ideális felszín alacsony, ritka növényzettel rendelkezik vagy növényzet nélküli.

Több osztályozási módszer létezik az NDVI értékek csoportosításához. A legtöbb esetben a nulla alatti értékek víz, illetve fedetlen talaj besorolást kapnak. Az általam használt csoportosítást a 16. ábrán látható, két tizedes érték közökkel sorolja a növényzetet kategóriákba.

Class	Classification criterion
Bare soil and/or water	$NDVI \leq 0$
Very Low	$0 < NDVI \leq 0.2$
Low	$0.2 < NDVI \leq 0.4$
Moderately Low	$0.4 < NDVI \leq 0.6$
Moderately High	$0.6 < NDVI \leq 0.8$
High	$0.8 < NDVI \leq 1$

16. ábra A használt NDVI osztályok (do Nascimento Aquino, Coimbra da Rocha Neto, Alves Moreira, dos Santos Teixeira, & Maia de Andrade, 2018)

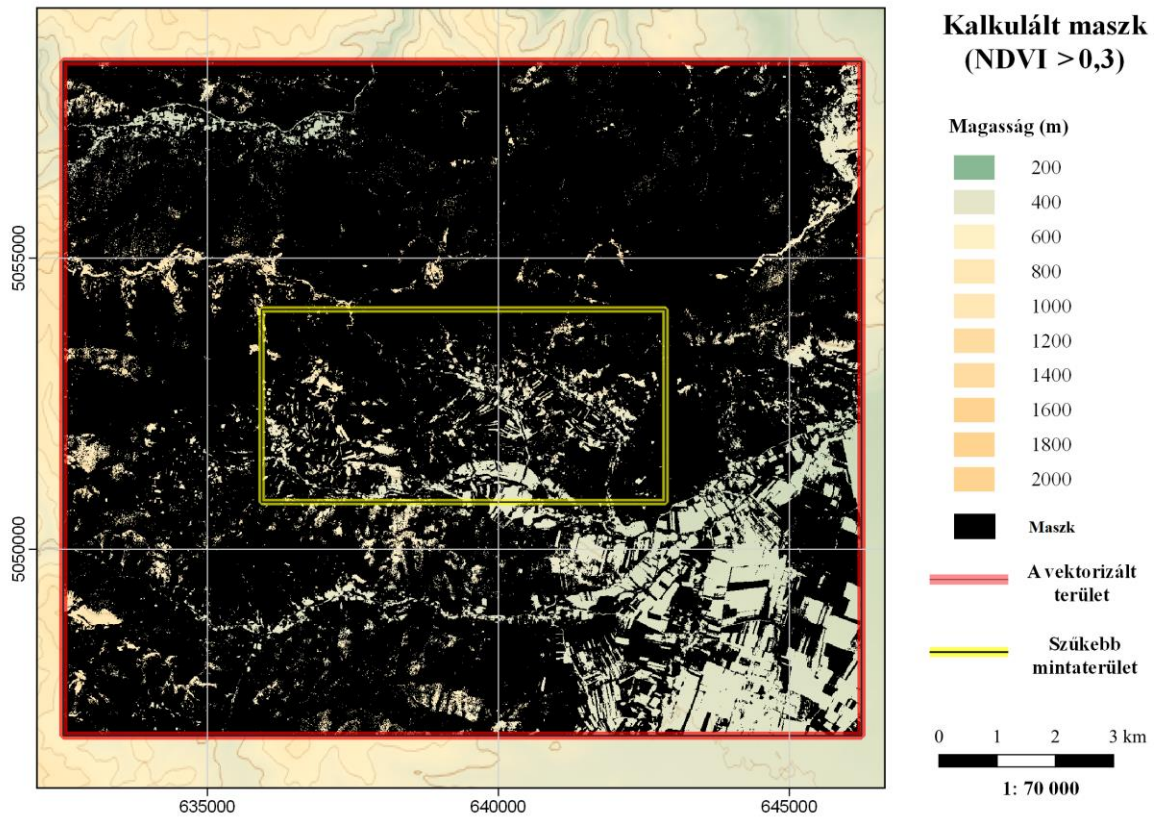
A táblázat alapján a 0-0,2 érték kategória lenne ideális, ám a mintaterület sűrűbb növényzete miatt a 0,2-0,4-es kategóriában készítettem el a növényzeti maszkokat. A növényzet okozta zaj felmérése érdekében két különböző maszkkal is végrehajtottam az osztályozást. Az első maszkkal a 0,3 feletti, a másodikba a 0,35 feletti indexű növényzetet fedtem ki.

A használni kívánt algoritmusok (pl. véletlen erdő) nem támogatják, ha különböző az értékkel rendelkező, illetve értékkel nem rendelkező pixelek száma a sávokon a többsávós raszter feldolgozásakor. Ezért a maszkoláshoz olyan egybites réteget készítettem, ami tartalmazza kategóriánként (0,3;0,35) az összes pixelt, ami értékkel rendelkezik az összevont idősoros ásványalkotókat tartalmazó állományban. Így minden olyan pixel, ami legalább egy sávon (a 18ból) tartalmaz értéket az aktív, vagyis 1-es értéket kapott, minden további 0-ást. Ehhez elég volt minden felvétel első levezetett sávját (képletben „1S”) megvizsgálni. A raszter kalkulátorban az alábbi képletet használtam erre³.

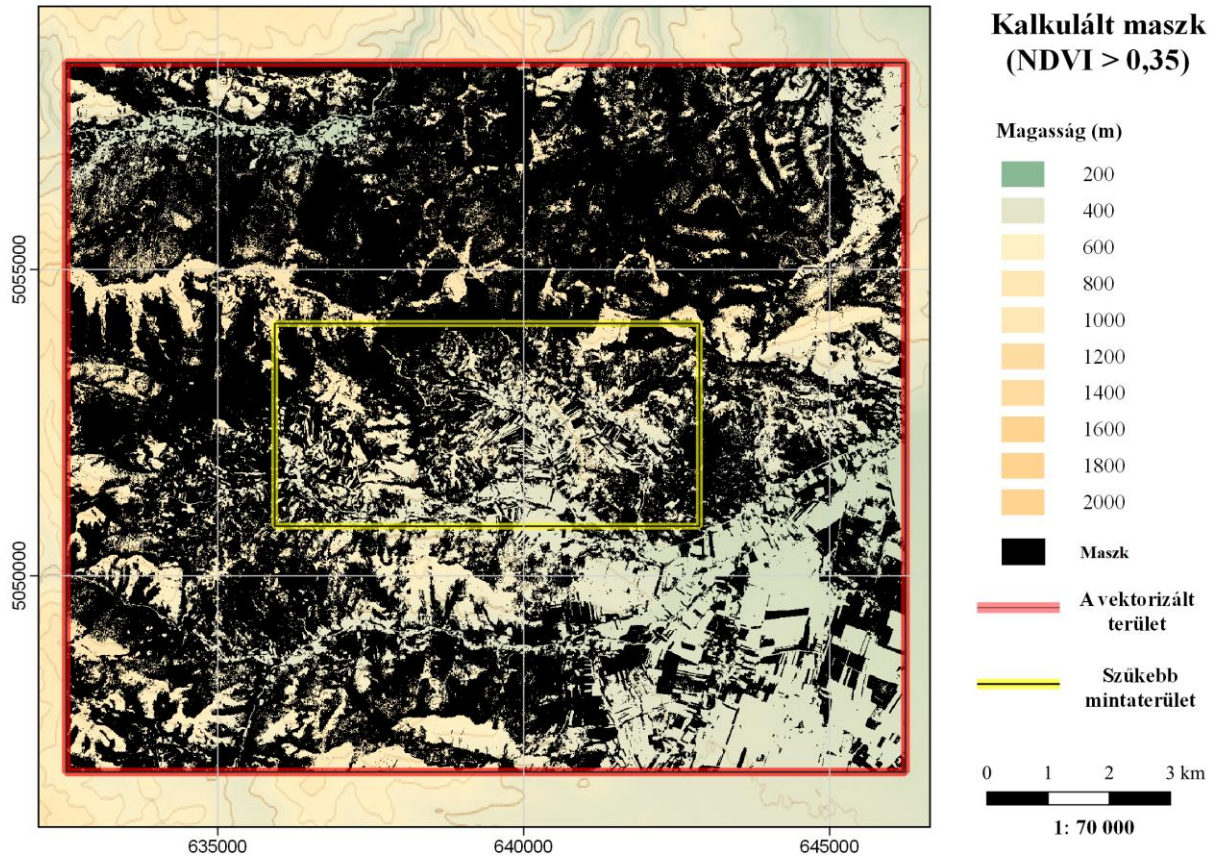
$$(A \text{ felvétel } (1S) + B \text{ felvétel } (1S) + C \text{ felvétel } (1S)) \geq 1$$

Így a következő térképeken látható maszk eredményeket kaptam a különböző kategóriákban. A maszkok a 0,3, illetve 0,35-ös értékek feletti területet fedik le. Ezek eredményei a 17-es, illetve 18-as ábrán láthatók.

³ A raszter kalkulátorban nincs HA elágazás, így hasonló alternatív megoldásokat kell keresni.



17. ábra A kalkulált NDVI alapú maszk 0,3 értékkel (UTM N34Z)



18. ábra A kalkulált NDVI alapú maszk 0,35 értékkel (UTM N34Z)

3.6. Osztályozás paramétere

A feldolgozás következő lépése a műholdképek, illetve a geológiai térképek digitalizálása után az ezek felhasználásával történő osztályozás. Az előző fejezetben bemutatott módszertípusok közül jelen diplomamunkában három szimpla irányított (supervised), illetve két gépi tanulás alapú irányított osztályozási módszert használtam. Ebben az alfejezetben röviden bemutatom a kiválasztott eljárásokat, valamint a lefuttatáshoz általam használt paramétereiket. Ezeket az eljárásokat mindkét – különböző NDVI értékhatárokkal elkészített – maszkolt többsávós felvételre lefuttattam.

Az osztályozásokhoz a SagaGIS szoftvert használtam, azon belül a 7.9.0-ás verziót. A választás azért erre a programra esett, mert funkciók és eszközök terén kompakt létére az egyik legbővebb az ismert térinformatikai szoftverek közül. Többek közt a QGIS-is a SagaGIS-ben implementált algoritmusokat használja. A következőben bemutatott klaszterező módszereken túl több egyéb módszer is megtalálható benne.

A három „szimpla” irányított osztályozó algoritmusként a gyakran használt „legnagyobb valószínűség” (Maximum likelihood), a „legkisebb távolság” (Minimum distance) módszereket választottam. A gépi tanulás alapú módszerek közül a szintén gyakran használt „véletlen erdő” (Random forest) algoritmust teszteltem. Azért esett ezekre az algoritmusokra a választás, mert a tesztelések során ezekről kaptam valamilyen szinten értékelhető eredményeket, illetve, több tanulmány, szakkönyv mutatott rá, hogy a műholdkép osztályozás terén ezek az algoritmusok jó közelítéssel sorolják be az egyes pixeleket a megfelelő kategóriákba (FAL, MAANAN, BAIDDER, & Rhinane, 2019), (A. Richards & Jia, 2005), (Albert & Ammar, 2020).

Az algoritmust a „szimpla” módszerekhez a SagaGIS-ben „Supervised Classification for Grids” néven lehet megtalálni. Az csoportosítandó adathalmaz (Grid-System) kiválasztása után a megfelelő objektumot (Features) kell meghatározni. Lehetőség van az objektum sávjainak pixel értékeit normalizálni. Készíthetünk kereséstitőt (Look-up-Table) is, amivel az értékek térképezését gyorsíthatjuk későbbi kielemezés során. Ezek a paraméterek (leszámítva az adathalmazokat) az összes további esetben hasonló módon lettek meghatározva, így nem kerülnek bemutatásra. Ezek mellett lehetőség van az osztályozási statisztikák elmentésére is az opciók fül első kitölthető sorában. Ezután következnek az algoritmusra szabott paraméterek.

A következő lefuttatott algoritmus a **legkisebb távolság** (Minimum distance) módszere volt. Ennek lényege, hogy Euklideszi távolságot használva számolja ki az aktuális pixel (2 dimenzióban 2 sáv érték alapján), illetve a különböző csoportok átlagos középtávolságának

nagyságát. Az aktuális pixel abba csoportba kerül, amelynek középpontjához a legközelebb van. Spektrális távolság módszerének is szokták nevezni.

Az Euklideszi távolság képlete (Pouncey, Swanson, & Hart, 1999):

$$SD_{xyz} = \sqrt{\sum_{i=1}^n (\mu_{ci} - X_{xyi})^2}$$

SD_{xyz} = az „x,y” koordinátájú pixel távolsága a „z” középponttól

μ_{ci} = a „c” osztály átlag értéke az „i” sávon

X_{xyi} = az aktuális pixel „x,y” koordinátája az „i” sávon

n = a sávok száma

i = az aktuális sáv

Ezen módszer előnye, hogy nincsenek osztályozatlan pixelek, ugyanis minden pixelhez van egy legközelebb eső csoport. Valamint a gyorsasága a parallelepipedon módszert leszámítva a legjobb. Legnagyobb hátránya, ami egyben előnyként is szerepelt, hogy minden pixel csoportba kerül, így a kiugró értékek is. Ezért nagy szórású értékek esetében érdemes távolság küszöbértéket (Distance Threshold) megadni (Pouncey, Swanson, & Hart, 1999).

Options	
Save Statistics to File...	C:\Users\Valentin\Desktop\diplomamunka\adatok\Hátszegi_medence\Sepa
Method	Minimum Distance
Distance Threshold	0
Update Colors from Features	<input type="checkbox"/>

19. ábra A legkisebb távolság módszer paraméterei

Az előzőhöz hasonlóan, egyszerűsége révén itt sem sok beállítás áll rendelkezésre, leszámítva a távolság küszöbértéket (19. ábra). Mivel az általam használt adatforrás viszonylag homogén eloszlású, ezért nem használtam ezt az opciót.

Az harmadik, egyben utolsó ebben a funkcióban implementált osztályozó algoritmus, amit használtam, a **legnagyobb valószínűség** (Maximum likelihood). Ez a módszer a pixelek hovatartozását illetően valószínűség értékeket ad meg. Az alapfelvetés az, hogy egy adott pixel bármely csoportba azonos valószínűséggel tartozhat. A pontos valószínűségek meghatározására a Bayes-i valószínűség (Bayesian probability) képletét használja (Pouncey, Swanson, & Hart, 1999).

$$D = \ln(a_c) - [0.5 \ln(|Cov_c|)] - [0.5 (X - M_c)^T (Cov_c - 1) (X - M_c)]$$

D = a súlyozott távolság

c = az adott osztály

X = az aktuális pixel mérési vektora

M_c = a „ c ” osztály átlag vektora

a_c = százalékos valószínűsége annak, hogy az adott pixel a „ c ” osztály része (alapesetben 1)

Cov_c = a „ c ” osztály pixeleinek kovariancia mátrixa

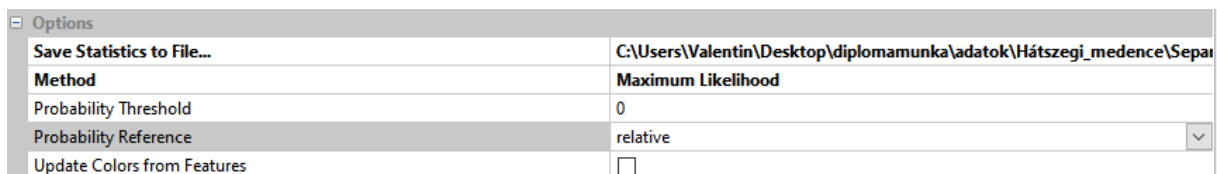
$|Cov_c|$ = a Cov_c determinánsa

Cov_c^{-1} = a Cov_c inverze

\ln = természetes logaritmus

T = Transzponálás

Előnye, hogy az egyik legpontosabb osztályozási módszer, mert a szimpla módszerek közül a legtöbb változót veszi figyelembe. A kovariancia mátrix használatával figyelembe veszi a csoportok változékonyságát. Hátránya, hogy lassú, és minél több sávot dolgozunk fel annál több időt vesz igénybe a kalkuláció. További hátrányok, hogy az pixelértékeknek szükségszerűen normális eloszlásúnak kell lenniük az eredményesség érdekében, illetve a hajlamos a túlosztályozásra (Pouncey, Swanson, & Hart, 1999).



Options	
Save Statistics to File...	C:\Users\Valentin\Desktop\diplomamunka\adatok\Hátszegi_medence\Separ
Method	Maximum Likelihood
Probability Threshold	0
Probability Reference	relative
Update Colors from Features	<input type="checkbox"/>

20. ábra A legnagyobb valószínűség módszer paraméterei

A legkisebb távolság módszer opcióit kiegészíti egy további paraméterrel, amivel meg tudjuk adni a valószínűségek referenciáját, ami lehet abszolút vagy relatív az általunk kívánt kimeneti rendszer értékektől függően, illetve a távolság küszöb helyett valószínűségi küszöb van (20. ábra). Az előzőhöz hasonlóan itt sem adtam meg osztályozási küszöbértéket, a referenciát pedig relatív értékkel használtam.

Utolsóként a **véletlen erdő** (Random forest) osztályozási eljárást. Nevéből következtethetően a döntési fa típusú osztályozó módszerek közé tartozik. Ezen módszerek alapja, hogy minden levél (igaz-hamis) értéket ad, illetve bemeneti adatként az előző kapcsolt levél kimenetét használja. Minél mélyebb fa, annál több információ nyerhető ki az adatokból. A mélység mellett több párhuzamos fa szintén növeli a pontosságot.

A módszer előnye, hogy viszonylag nehezen lehet túltanítani. Szintén előny, hogy ún. OOB adatsorokat véletlenszerűen kihagyja az osztályozásból, amelyeket aztán a létrehozott döntési fák tesztelésére használ fel, így nincs szükség keresztvalidációra. Hátránya, hogy több fa/nagy mélység esetén a struktúrája átláthatatlan. Emellett minél több fát használunk az osztályozásra az adott modellben, annál lassabb az algoritmus (Burján & Gönczy, 2017).

Model Training	
>> Training Areas	01. Hátszeg_vektor_BGG
Class Identifier	ClassID
Save Model	C:\Users\Valentin\Desktop\diplomamunka\adatok\Hátszegi_medence\Separated
Maximum Tree Depth	10
Minimum Sample Count	2
Maximum Categories	18
Use 1SE Rule	<input checked="" type="checkbox"/>
Truncate Pruned Trees	<input checked="" type="checkbox"/>
Regression Accuracy	0.01
Active Variable Count	0

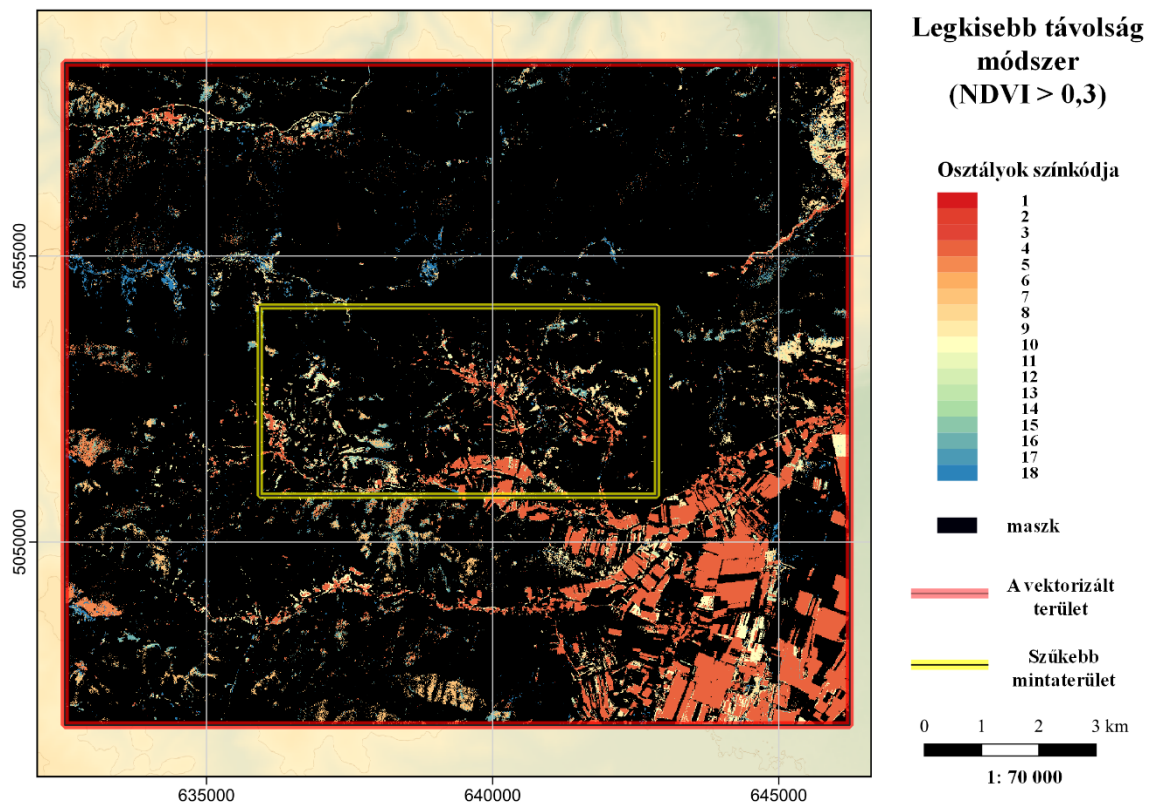
21. ábra A véletlen erdő módszer paraméterei

A SagaGIS-ben implementált algoritmus fontosabb paraméterei közt megadhatjuk a fa mélységét, a minimum minta mennyiséget, illetve a csoportok maximális számát. A minimum minták száma jelzi, hogy hány minta esetén kezd új szintet a fa (minimum 2, alap beállítás). A fa mélység pedig a csoportok száma/adatok tulajdonságai (dimenzió) alapján kerül megállapításra. A használt adatok függvényében az 21. ábrán látható paramétereket használtam. A további opciókat ez esetben is alap beállításokon használtam.

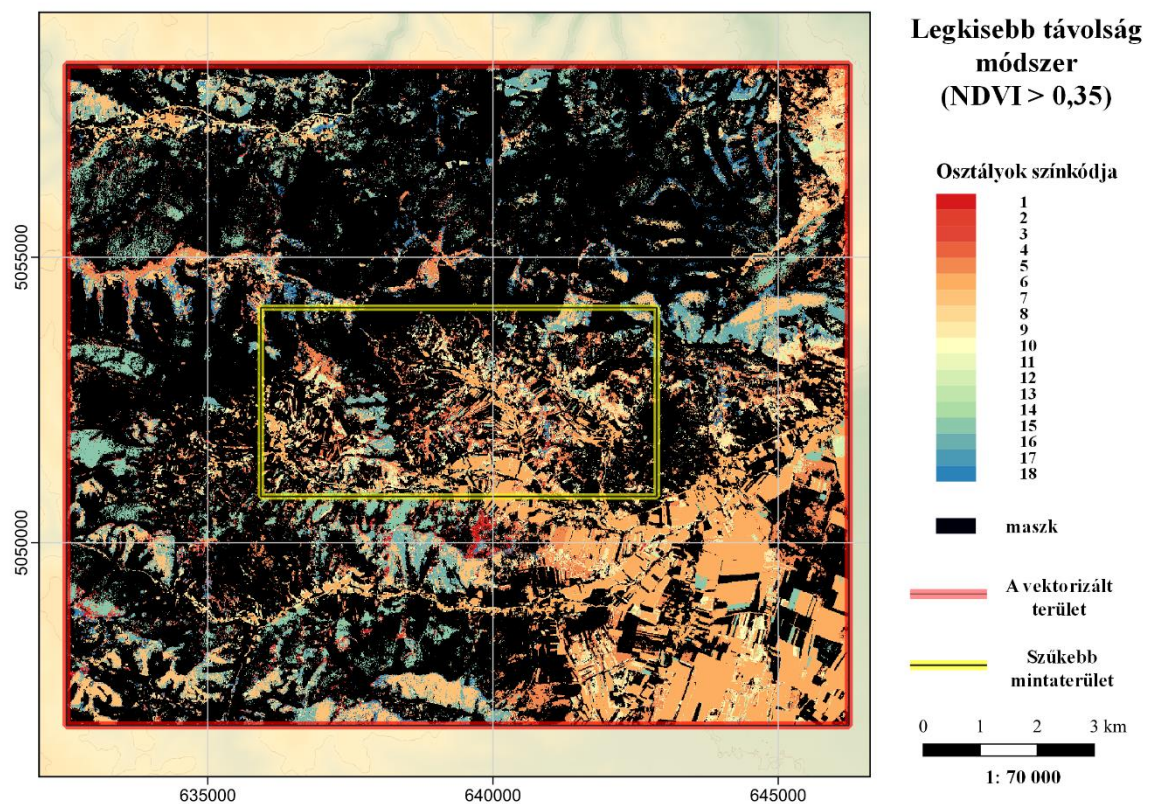
4. Eredmények

Ebben a fejezetben a feldolgozott adatokból, az ismertetett osztályozási eljárásokkal készült raszteres állományokat mutatom be. Minden különböző eljárás esetében két eredmény van, ez a kétféle alkalmazott NDVI maszkból következik. Az vizualizált eredmények NDVI határértékei minden esetben a 0,3, illetve 0,35 feletti növényzet kifestését jelentik. Így a címben látható $NDVI > 0,3/0,35$ a kifestett területre utalnak. Az eredmény térképeken a fekete szín jelöli a maszkolt területeket, a színek pedig a jelmagyarázatban található értékeknek felelnek meg. Ezek a kőzetek korát is jellemzik, minél nagyobb sorszámú az adott pixel, annál idősebb kőzetnek felel meg. A kategóriákhoz kapcsolható kőzeteket a 2. táblázatban találhatók.

4.1. Legkisebb távolság

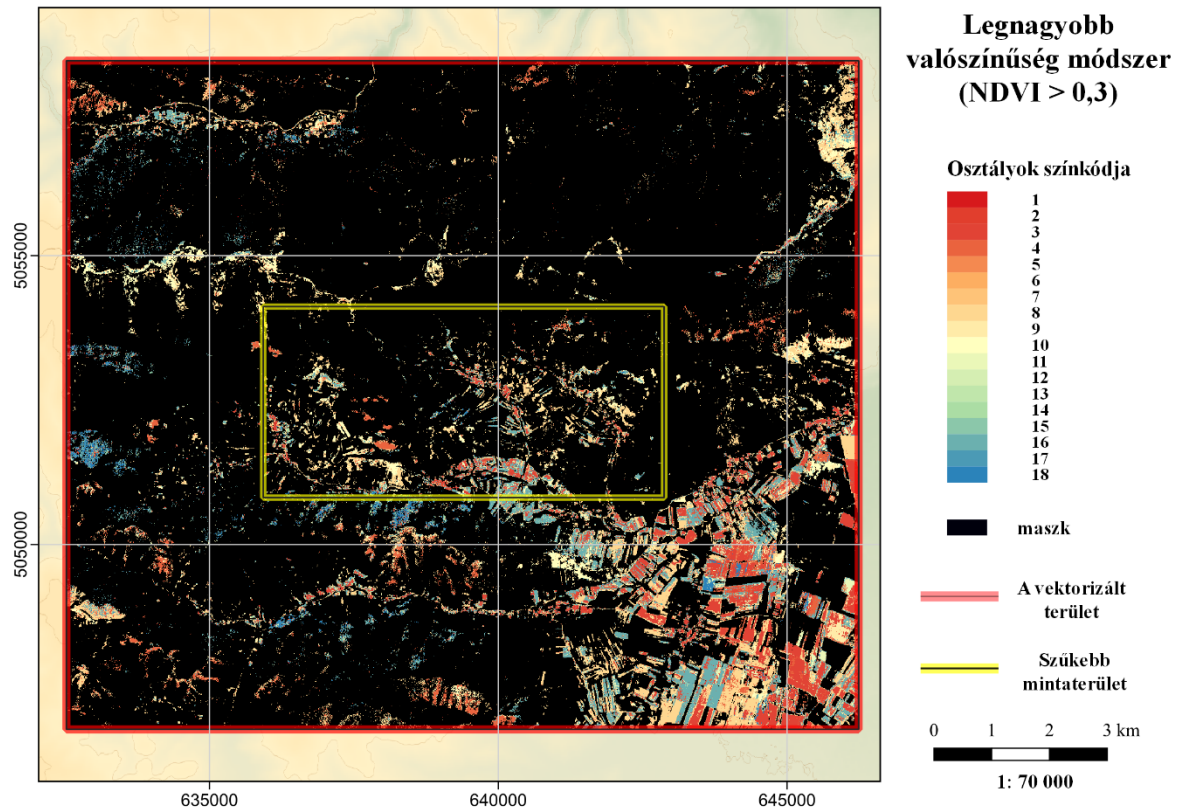


22. ábra A legkisebb távolság módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z)

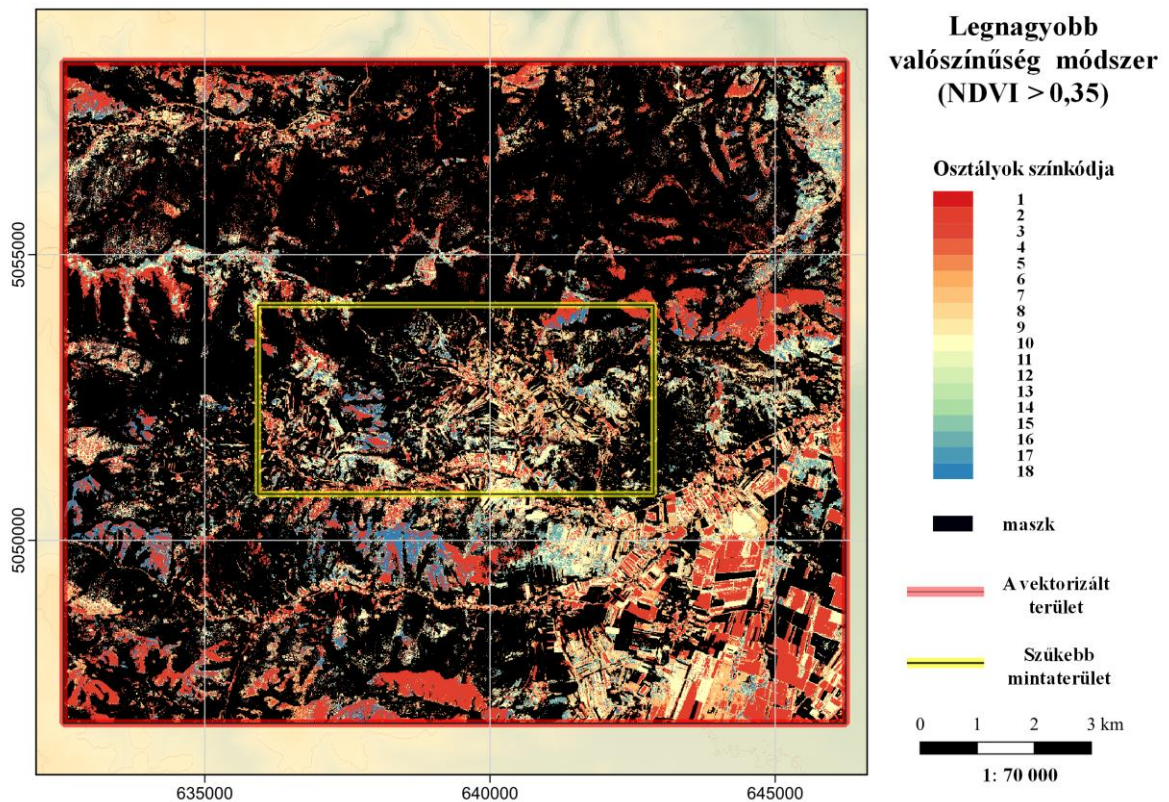


23. ábra A legkisebb távolság módszer a 0,35-ös NDVI maszk kategóriában (UTM N34Z)

4.2. Legnagyobb valószínűség

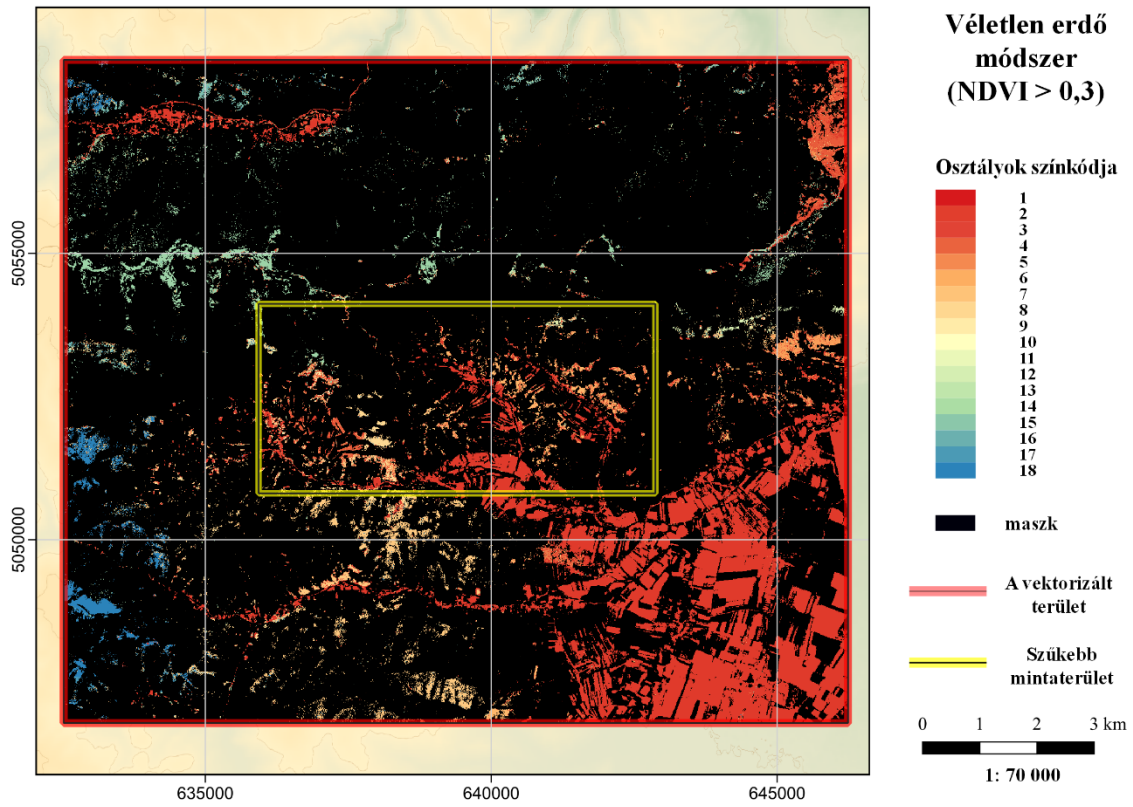


24. ábra Legnagyobb valószínűség módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z)

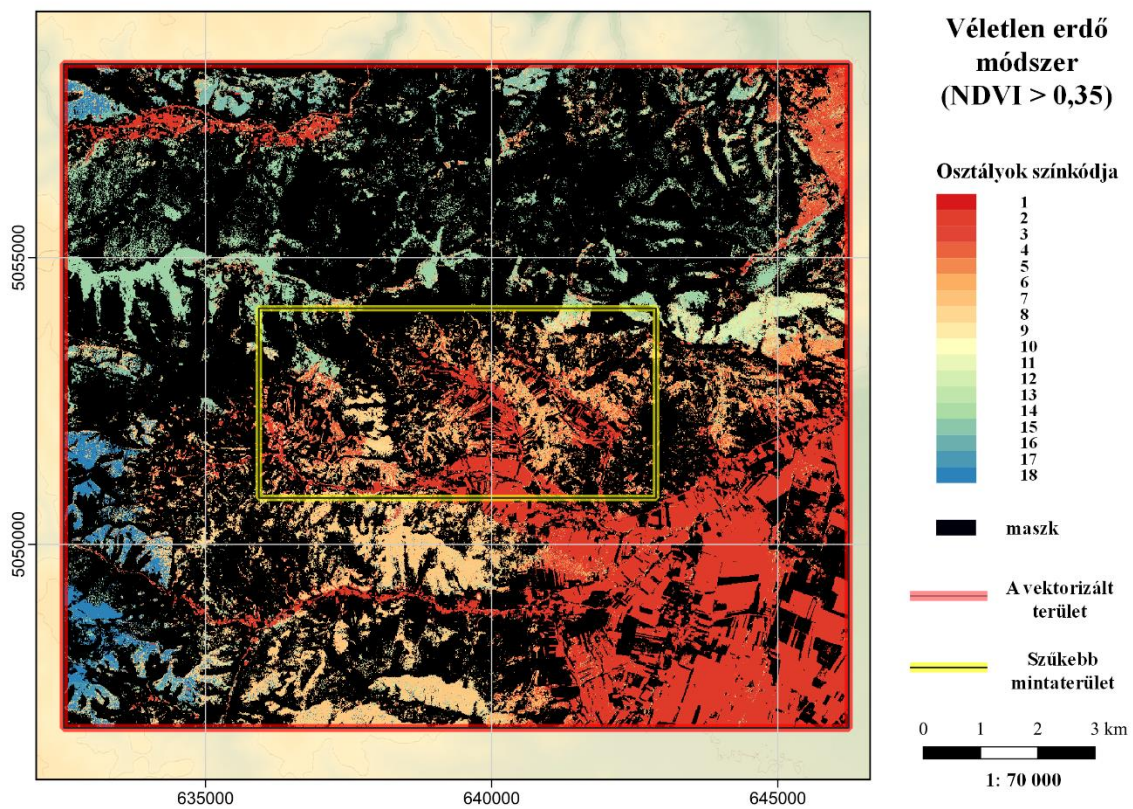


25. ábra Legnagyobb valószínűség módszer a 0,35-ös NDVI maszk kategóriában (UTM N34Z)

4.3. Véletlen erdő



26. ábra A véletlen erdő módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z)



27. ábra A véletlen erdő módszer a 0,35-ös NDVI maszk kategóriában (UTM N34Z)

5. Eredmények kiértékelése

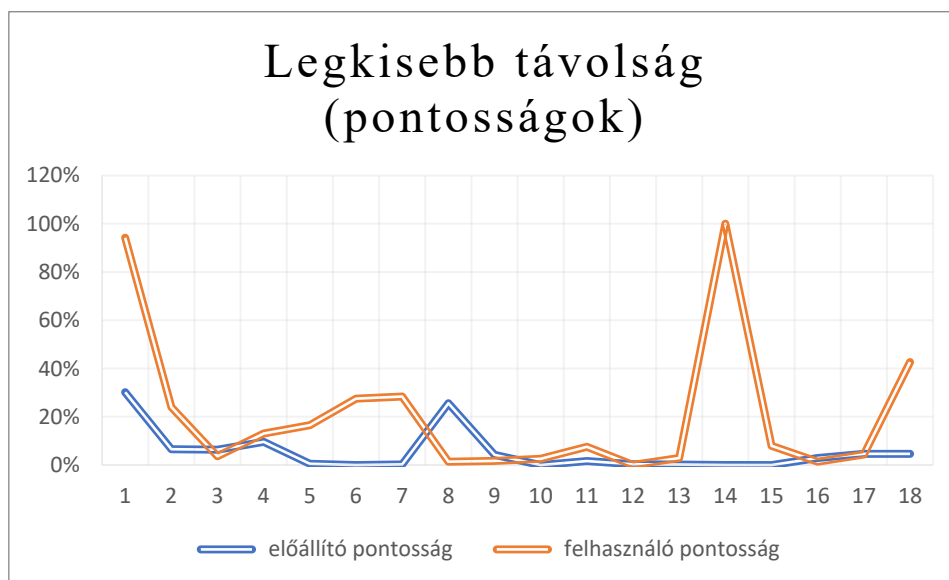
Szabad szemmel is megfigyelhető az eredmény térképeken az osztályozási eljárások pontossági sorrendje. Általánosságban sok területre, így a jelenlegi mintaterület különböző részeire is elmondható, hogy a kőzetek korára lehet következtetni az egyes formációk relatív magassága alapján (magasabb terület idősebb), emellett jelen mintaterület dél-keleti részéről nyugat, észak-nyugatra haladva egyre idősebb kőzetek találhatók (folyó völgyeket leszámítva). A vizuális kiértékelés során ezt a két szempontot vettem figyelembe. Ily módon legjobbnak, mindkét kategóriában (NDVI – 0,3;0,35) legalább egy nagyságrenddel a véletlen erdő módszer találtam. Második legjobbnak a legkisebb távolság módszer látszik, bár nagy különbség a legnagyobb valószínűség módszerhez viszonyítva nem, csak az osztályozott terület jobb alsó felén látható. Ez a terület főként szántóföldeket foglal magába, így ebből a szempontból az etalonnak a két szimpla módszer közül a legkisebb távolság mondható.

Az eredmények statisztikai kiértékelésére a SagaGIS konfúziós mátrix funkcióját használtam, amely csoportonként, a 2. fejezetben bemutatott elvek alapján számítja ki a különböző értékeket. A funkció végrehajtásával három táblázatot kapunk eredményül, amelyek közt van az osztály alapú statisztikák táblája, a konfúziós tábla, illetve az összegzést tartalmazó tábla. Ebben a fejezetben ezek kivonatát mutatom be, a két NDVI maszk kategóriát külön-külön tárgyalva. A táblák tartalmazzák az előállító pontosságát (producer accuracy), a felhasználó pontosságát (user accuracy), illetve az általános pontosságot és a kappa értéket. Az előállító pontosság érték a referencia pixelek osztályozási pontosságát adja meg, a területre eső jól besorolt referencia pixelek és a területre eső összes referencia pixel arányával, százalékban kifejezve. A felhasználó pontossága a „valós” pontosságot mutatja meg. Az összes jól besorolt pixel és az összes aktuális területre eső pixel arányával, százalékban kifejezve. Ezek osztályonként határozhatók meg. Általánosan igaz, hogy ezen két változó nem azonos, a referencia arány nem feltétlenül felel meg a valós osztályozási aránynak. Az általános pontosság az összes osztály jó csoportba sorolt pixelének, és az összes besorolt pixelnek az aránya. A kappa együttható az osztályozási eljárás pontosságát mutatja meg, a véletlenszerű besoroláshoz viszonyítva. Az értéke 1 és -1 közé eshet. A 0 kappa együttható egyenértékű a teljesen véletlenszerű osztályozással. A negatív értékek ennél rosszabbak a pozitívok jobbak (Humboldt State University, 2019).

A kiértékelés során először a különböző kategóriák különböző módszereinek pontosságát mutatom be (releváns osztályonként), majd az ezeket követően az összefoglalás részben ezekre próbálok magyarázattal szolgálni, illetve összehasonlítani a módszereket.

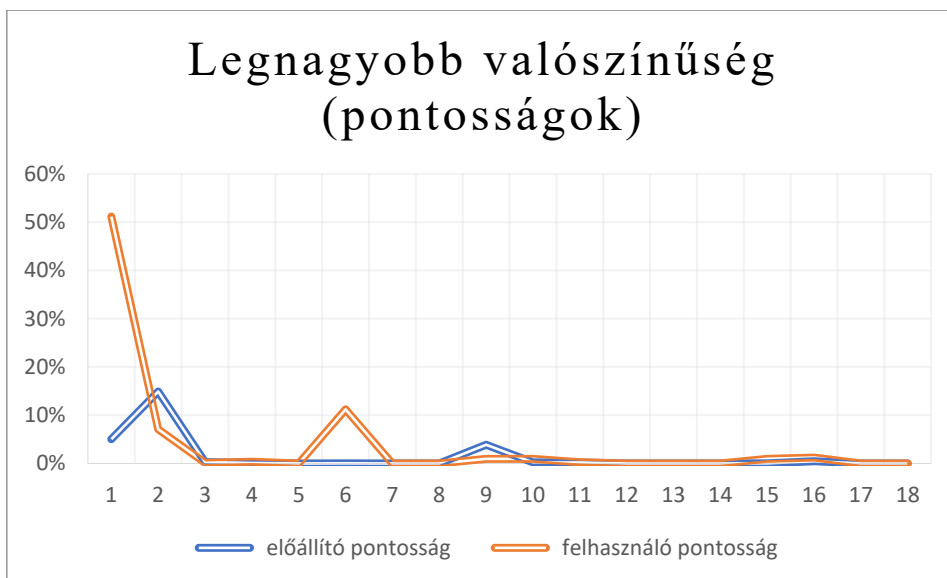
5.1.NDVI 0,3

A legkisebb távolság módszerével készített osztályozás felhasználó és előállító pontossága a különböző osztályokra viszonyítva a 28-as ábrán látható. Egyértelműen látszik, hogy a legtöbb osztály esetében kevés a jól besorolt pixel mind a felhasználói, mind az előállítói nézőpontból. Megfigyelhetők emellett kiugró értékek is, az 1-es, illetve a 14-es csoport esetében a felhasználó pontosságánál. Az előállító pontosság részéről, szintén megfigyelhetők kiugró értékek, főként itt is, az 1-es, illetve ebben az esetben az a 8-as osztályoknál. A többi csoport pontossági értékei nem érik el a 20%-ot sem, sőt van olyan csoport is (12-es), amelybe egyetlen pixel sem került. A 31. ábrán láthatók a kappa együttható, illetve az összesített pontosság, amik szintén megerősítik a vizuális kiértékelésben tett kijelentést a pontossági rangsorról. Ezekből az adatokból az is megállapítható, hogy a használt feldolgozási eljárással és adatokkal a legkisebb távolságok módszer nem sokkal (5%) jobb a random osztályozásnál. Összesített pontossága 7%.



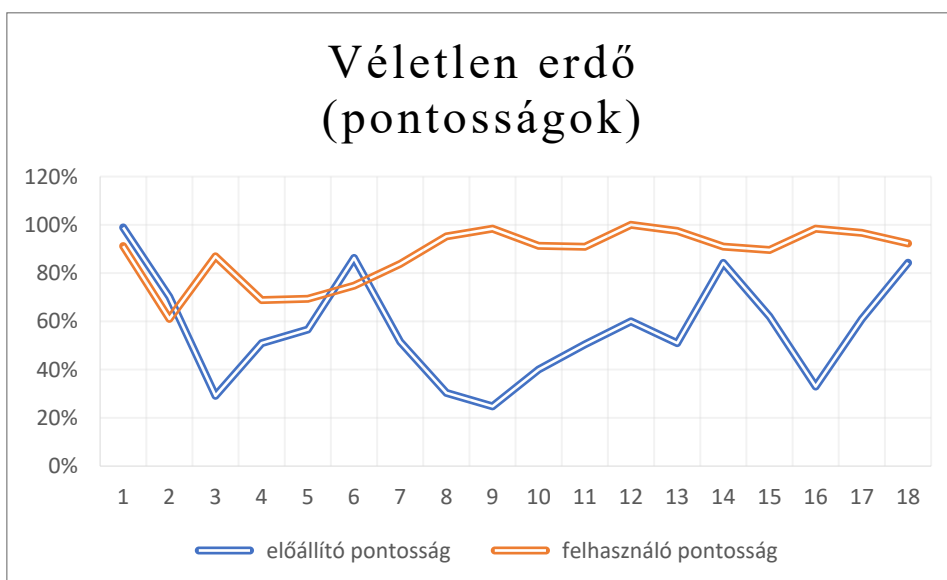
28. ábra A legkisebb távolság módszer pontosságai (0,3)

A legrosszabb pontosságú modell a legnagyobb valószínűségek módszer lett (29-es ábra). Az előző módszer 14-es és 18-as kiugró értékét (felhasználó pontossága) leszámítva nem sok különbség akad. Ennek diagramján is kivehető, az 1-es csoport kiugró értéke, a felhasználó pontosság tekintetében. Itt viszont a felhasználói pontosság részéről azt a csoportot, illetve a 6-os csoportot leszámítva egyik sem éri el a 10%-os pontosságot sem. A kappa együttható alapján jelen adatokkal/körülmények között ez a módszer 1 százalékkal teljesítene jobban a teljesen véletlenszerű besorolásnál. Az összesített pontossága szintén 1% körüli.



29. ábra A legnagyobb valószínűség pontosságai (0,3)

A 0,3-as NDVI értékhatárú maszkolt adatokon lefuttatott osztályozási eljárások közül a legjobb eredményt a véletlen erdő eljárás érte el. Az előállító pontosság hullámzó, legjobb értékeket az 1-es, 6-os, 14-es, illetve 18-as csoportokban adja. Ezek esetében a 14-es kategóriát leszámítva a minták száma 1000 pixel feletti (2. melléklet). Minél nagyobb a minták száma egy osztály esetén annál reprezentatívabb a pontossági érték. A felhasználó pontosság a legtöbb osztályban 70% feletti, több esetben megközelíti a 100%-ot. Az osztályozás kappa 78,2%, míg a modell összesített pontossága 86,9% százalék.



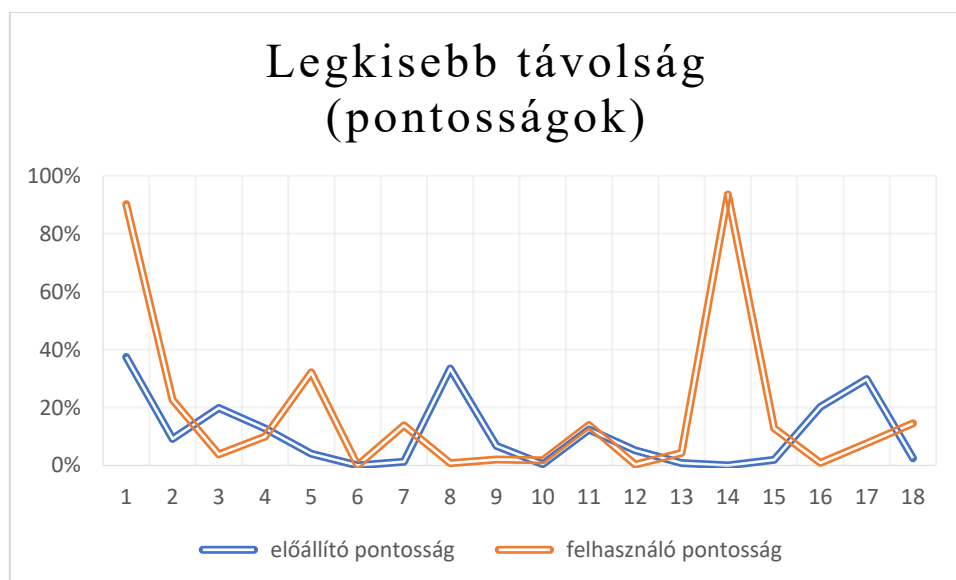
30. ábra A véletlen erdő módszer pontosságai (0,3)



31. ábra A 0,3-as NDVI kategória osztályozásainak kappa és összesített pontosság értékei

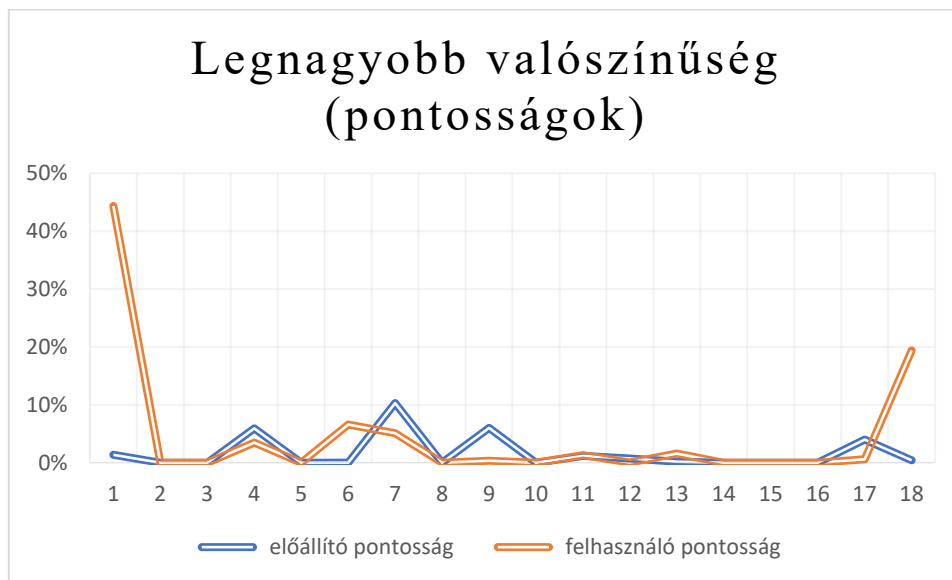
5.2. NDVI 0,35

A 0,35-ös NDVI értékhatárral kapott eredmények a legkisebb távolságok módszer esetében megfigyelhető az 1-es, 5-ös, illetve 14-es csoport kiugró értéke a felhasználói pontosság esetében (32. ábra). Ehhez hasonlóan az előállító pontosság értékek esetében is több szembetűnő érték van. Ezek főként az 1-es, 3-as 8-as, illetve 16-17-es osztályok, bár ezek értékei is épphogy elérik a 20%-os pontosságot. A kappa értéke 7,5%-os, az összesített pontossága pedig 9,6%-os. Ezzel a 0,35-ös kategóriában a használt módszerek közül a második „legjobb” mondható.



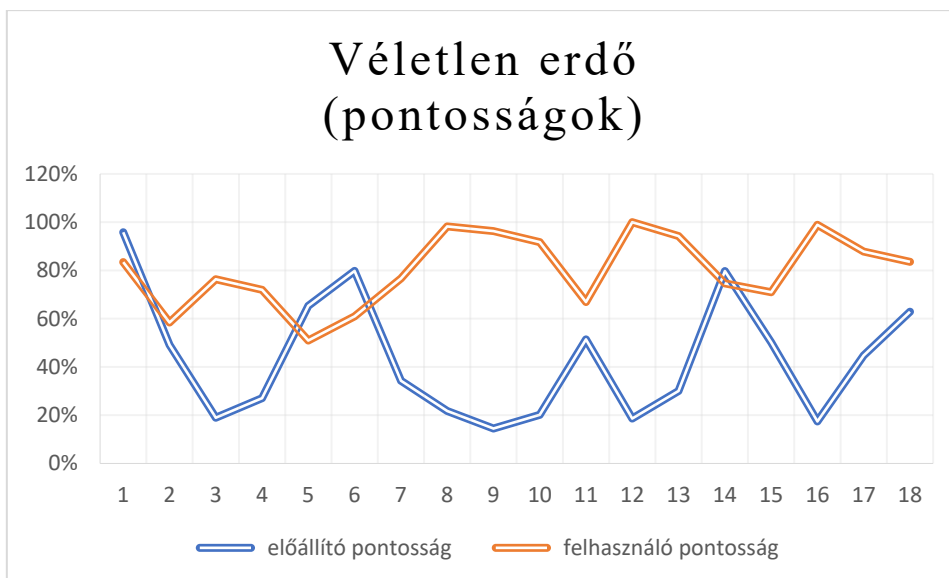
32. ábra A legkisebb távolság módszer pontosságai (0,35)

A legrosszabb pontosságú modell a 0,35-ös kategóriában szintén a legnagyobb valószínűségek módszer lett (35. ábra). Az első és utolsó osztály értékeit leszámítva a 10-es alatti osztályokban látható 10% alatti hullámzó pontosság mindkét érték alapján (33. ábra). Az 1-es, illetve 18-as és 7-es (előállító pontosság) csoportokat leszámítva mindkét értéket tekintve egy csoport sem éri el a 10%-ot. A kappa együtthatója alapján (-0,1%) ez a módszer rosszabbul teljesítene egy véletlenszerű osztályozásnál. Az összesített pontossága 1,2% körüli (35. ábra).



33. ábra A legnagyobb valószínűség módszer pontossága (0,35)

A 0,35-ös NDVI értékhatárú maszkolt adatokon lefuttatott osztályozási eljárások közül a legjobb eredményt a 0,3-as kategóriához hasonlóan, újra a véletlen erdő eljárás érte el. Ezek esetében több az elemzéshez ideális érték. Az előállítói pontosság értékek az 1-es, 5-6-os, 11-es 14-es, illetve 17-18-as csoportokban a legjobbak. Ennek megfelelően ezen értékek közt beesés tapasztalható. A felhasználó pontosságok tekintetében a mintázat ennek körülbelül a fordítottja/tükörképe, az 1-es kategóriát figyelmen kívül hagyva (34-es ábra). Ennek megfelelően viszonylag magas értékek figyelhetők meg az 1-3-4-es, a 8-9-10-es, a 12-13-as, illetve a 16-17-18-as osztályokban. Az osztályozás kappa értékét tekintve a használt módszerekhez mérten jónak mondható 66,6%-al. Összesített pontossága 73% (35-ös ábra).



34. ábra A véletlen erdő módszer pontossága (0,35).



35. ábra A kapa együtthatók és az összesített pontosságok a 0,35 NDVI kategóriában

5.3. Összesítés

A kapott értékek nagyrészt egybevágnak a szabad szemmel végzett kiértékeléssel, ugyanis az 1-es osztály főként az üledékes kőzeteket tartalmazza, ami a folyóvölgyekben, illetve a terület dél-keleti sarkában található nagyobb mennyiségben. Ez a legkisebb mértékben kifestett terület mindkét maszk kategóriába. Továbbá az 1-es csoportba tartozik a legtöbb besorolható pixel, illetve ez és 14-es csoport rendelkeznek a legnagyobb referencia területtel. Ennek megfelelően ezen osztályokban a kiugró értékeket ezekkel lehet magyarázni. A legkisebb távolság módszer esetében a további tüskék a konglomerátumok/durva homokkövek, illetve a szericites-kloritos

palák osztályaiban jelentősök. A legnagyobb valószínűség módszer, átlagot véve a két NDVI kategóriában csak az újabb, üledékes kőzetek tekintetében kiemelkedő a többi osztályhoz viszonyítva. A véletlen erdő módszer osztály értékei (felhasználó pontosság) a 0,3 felett kifestett NDVI értékek kategóriájában (30. ábra) nagyjából egyenletesen futnak a 80-100% közötti sávban. Enyhébb beesés a Delényes-formáció (Delinești) látható, ami a márgák, homokok és homokkövek csoportját takarja. A referencia pontosság értékei hullámzást mutatnak, ami referencia pixelek számával korrelál (3. melléklet). A 0,35 kategória véletlen erdő osztályozásánál szintén ez látható, a nagy referencia pixel számú osztályok (6-os) jobb eredményt mutatnak, mint az egy nagyságrenddel kevesebb referencia pixellel rendelkező osztályok (3-6-os csoportok) (6. melléklet). Szintén megemlítenő, hogy magas felhasználói pontosság értékek nem feltétlenül reprezentatívak, ugyanis kifestéstől függően néhány besorolt pixel is adhat nagy pontosság értéket.

A meghatározott NDVI kategóriák közt is látható különbség minden osztályozás esetén. Ezek főként a kappa, illetve összesített pontosság értékeken figyelhetők meg. A véletlen erdő módszer esetében mindkét érték legalább 10%-al csökkent a kisebb területek kifestésével, vagyis ahogy várható volt a besorolható pixelek számának nagysága ez esetben fordítottan arányos a kappa és az összesített pontosság értékeivel. Ez lényegében a növényzet okozta zajt jelenti. A másik két, alacsonyabb pontosságú kategóriában. Ennek ellenkezője lenne igaz, azonban ezek alacsony kappa értéke miatt a besorolás pontosságának kismértékű növekedése a legtöbb osztályra nem releváns, ugyanis a több besorolt pixellel nő a rossz csoportba sorolt pixelek száma is. Ez alátámasztható például a legkisebb távolság módszer esetében a viszonylag jó pontosságú 1-es csoport pontosság csökkenésével (94%-ról 90%-ra) (1. melléklet és 4. melléklet; 1-es osztály). Ebből következőleg a jelen bemenő adatok növényzeti zaja lehet a fő oka ezen osztályozási eljárások alacsony pontosságának.

6. Összegzés

Diplomamunkámban az Hátszegi-medence műholdkép alapú geológiai osztályozását tűztem ki célul. Ehhez a Sentinel-2 földmegfigyelő műhold multispektrális felvételeit használtam fel. Geológiai alapként a területről a 19. század második felében készült felmérésekből szerkesztett (1:50.000-es méretarányú) térképeket használtam. Az osztályozáshoz a Sentinel-2 felvételek natív sávjai helyett az azokból kalkulált vas ásvány összetevőket vettem alapul, így egy idősoros 18 sávós rászteres állományt kapva. A terület sűrű növényzete miatt két különböző értékhatárral NDVI maszkokat (0,3;0,35) készítettem, majd ezekkel kifedtem a kalkulált rászteres állományokat. Az osztályozás során több algoritmust is teszteltem. Ezek a véletlen erdő, a legnagyobb valószínűség, illetve a legkisebb távolság módszerek voltak. Ezek futtatása közben a maszkokkal kapcsolatos nem várt hibával találkoztam, amit sikerült kiküszöbölni.

Az osztályozás után kiértékeltem a röviden bemutatott módszerek pontosságát a konfúziós mátrixal készített eljárás pontossági vizsgálatok segítségével. Ebből azt a következtetést vontam le, hogy „szimpla” módszerek jelen dolgozatban ismertetett feldolgozással összességében nem hatékonyak, míg a gépi tanulás alapú véletlen erdő módszer ezeknél sokkal jobban teljesített. A meghatározott 0,3-as kategóriában 86,9%-os összesített besorolási pontossággal. Hasonló alapanyagokkal/ feldolgozási eljárással történő osztályozás esetén előremutató lehet további gépi tanulás alapú módszerek tesztelése.

7. Summary

In my diploma thesis, I aimed at a satellite-based geological classification of the Hátaszegi Basin using multispectral images from the Sentinel-2 Earth observation satellite. As a geological basis, I used maps (scale 1:50.000) of the area, which were compiled from surveys made in the second half of the 19th century. For the classification, I used the calculated iron mineral compositions instead of the native bands of the Sentinel-2 images, resulting in a time-series raster dataset of 18 bands. Due to the dense vegetation of the area, NDVI masks (0.3;0.35) were created with two different thresholds and then used to mask the calculated raster inventories. Several algorithms were tested during the classification. These were the random forest, maximum likelihood, and minimum distance methods. While running these, I encountered an unexpected error with the masks, which I managed to eliminate.

After classification, I evaluated the accuracy of the methods briefly presented using the accuracy tests of the procedure with the confusion matrix. From this, I concluded that "simple" methods with the processing described in this thesis are inefficient overall, while the machine learning-based random forest method performed much better than these. With an overall classification accuracy of 86.9% in the defined category 0.3. In the case of classification with similar raw materials/processing methods, it may be beneficial to test further machine learning-based methods.

Hivatkozások

- A Kopernikusz program.* (2021). Forrás: Copernicus: Europe's eyes on Earth:
<https://www.copernicus.eu/hu/kopernikusz-program>
- A to Z About Active and Passive Remote Sensing.* (2017. május 2017.05.16). Forrás:
GrindGIS: <https://grindgis.com/remote-sensing/active-and-passive-remote-sensing>
- A. Richards, J., & Jia, X. (2005). *Remote sensing digital image analysis (Fourth edition)*.
Canberra: Springer-Verlag.
- Abburu, S., & Golla, S. (2015). Satellite Image Classification Methods and Techniques: A
review. *International Journal of Computer Applications*, 20-25.
- Albert, G., & Ammar, S. (2020. november 4). Application of machine learning and remotely
sensed data in geological mapping on the Jebel Meloussi area (Tunisia). Tunisia.
- Botfalvai, G., Csiki-Sava, Z., Kocsis, L., Albert, G., Magyar, J., Bodor, E. R., . . . Makádi, L.
(2021). 'X' marks the spot! Sedimentological, geochemical and palaeontological
investigations of Upper Cretaceous (Maastrichtian) vertebrate fossil localities from the
Vălioara valley (Densuş-Ciula Formation, Haţeg Basin, Romania). *Cretaceous
Research*, 123.
- Braun, A. (2020). *SAR-based landcover classification with Sentinel-1 GRD products*.
SkyWatch Space Applications (ESA).
- Burján, D., & Gönczy, L. (2017). Döntéstámogatási módszerek alkalmazása
teljesítménymodellek adatvezérelt szintézisében. Budapest: Budapesti Műszaki és
Gazdaságtudományi Egyetem.
- do Nascimento Aquino, D., Coimbra da Rocha Neto, O., Alves Moreira, M., dos Santos
Teixeira, A., & Maia de Andrade, E. (2018). Use of remote sensing to identify areas at
risk of degradation in the semi-arid region. *Revista Ciência Agronômica*, 420-429.
- ESA. (2017). *Satellite description*. Forrás: The European Space Agency: Sentinel:
<https://sentinel.esa.int/web/sentinel/missions/sentinel-2/satellite-description>
- FAL, S., MAANAN, M., BAIDDER, L., & Rhinane, H. (2019). The contribution of Sentinel-
2 satellite images for geological mapping in the south of Tafilalet basin (Eastern Anti-
Atlas, Morocco). *The International Archives of the Photogrammetry, Remote Sensing
and Spatial Information Sciences*, 75-82.
- Gábris, G., Horváth, E., Horváth, G., Kéri, A., Móga, J., Nagy, B., . . . Telbisz, T. (2014).
Európa regionális földrajza 1. Természetföldraj. Budapest: ELTE Eötvös Kiadó.
- Georgescu, M., Vlad, V., Lupu, M., & Maier, O. (1979). Harta geologică a României scara
1:50.000, foaia Băuţar. România: Institutul Geologic al României.
- Humboldt State University. (2019). *Accuracy Metrics*. Forrás: Humboldt State Geospatial
Online, Humboldt State University:
http://gis.humboldt.edu/OLM/Courses/GSP_216_Online/lesson6-2/metrics.html
- Kovács, A., & Erdődiné Molnár, Z. (2019. február 21). A növényzet műholdas megfigyelése
– vegetációs indexek. Magyarország.
- Kramer, H. J. (1994). *Observation of the Earth and Its Environment: Survey of Missions and
Sensors*. Springer-Verlag.
- Lupu, M., Popescu, G., Munteanu, T., Pop, G., Bindea, G., Stelea, I., & Munteanu, E. (1993).
Harta geologică a României scara 1:50.000, foaia Haţeg. Institutul Geologic al
României.
- Mezei, G., Kocsis, K., Csorba, P., Horváth, G., Bartos-Elekes, Z., Imecs, Z., & Poszet, S.
(2018). A Kárpát-Pannon térség természeti tájainak rendszertani felosztása. In K.
Kocsis, G. Horváth, Z. Keresztesi, & Z. Nemerkenyi, *Magyarország nemzeti atlasza*
(old.: 126). Budapest: Magyar Tudományos Akadémia Csillagászati és Földtudományi
Kutatóközpont.

- Mucsi, L. (2013). *Műholdas távérzékelés (elmélet és gyakorlat)*. Szeged: Szegedi Tudományegyetem.
- Pouncey, R., Swanson, K., & Hart, K. (1999). *ERDAS Field Guide (Fifth edition, revised and expanded)*. Atlanta.
- Schmidt, R. G. (1976. január-február). EXPLORATION FOR PORPHYRY COPPER DEPOSITS IN PAKISTAN USING DIGITAL PROCESSING OF LANDSAT-1 DATA. *Journal of Research of the U.S. Geological Survey*, old.: 27-34.
- Smith, R. E., Green, A. A., Robinson, G., & Honey, F. R. (1978. július). Use of LANDSAT-1 Imagery in Exploration for Keweenaw-type Copper Deposits. *Remote Sensing of Environment*, old.: 129-144.
- Vágó, J., Seres, A., & Hegedűs, A. (2011). *Alkalmazott térinformatika*. Miskolc: Miskolci Egyetem Földtudományi Kar.
- van der Werff, H., & van der Meer, F. (2016). Sentinel-2A MSI and Landsat 8 OLI Provide Data Continuity for Geological Remote Sensing. *Remote Sensing*.

Köszönetnyilvánítás

Köszönettel tartozom tanáromnak és témavezetőmnek, Dr. Albert Gáspárnak a diplomamunkám témájának kiválasztásában nyújtott segítségért, valamint a konzultációk alkalmával adott gyakorlati útmutatásért, illetve, hogy rendelkezéseimre bocsátotta az alapanyagul szolgáló geológiai térképeket.

Hálás vagyok a családomnak, akik a diplomamunka elkészítése során minden formában támogattak.

Ábrajegyzék

1. ábra. A mintaterület (UTM N34Z).....	4
2. ábra Passzív és aktív távérzékelés (A to Z About Active and Passive Remote Sensing, 2017).....	6
3. ábra Földi energia spektrum, illetve a referencia abszolút fekete test sugárzás a légkör felső részén (szaggatott) (Vágó, Seres, & Hegedűs, 2011).....	7
4. ábra Az elektromágneses spektrum felosztása (Vágó, Seres, & Hegedűs, 2011).....	7
5. ábra Műholdképek osztályozási metódusainak hierarchiája (Abburu & Golla, 2015)	8
6. ábra Sentinel-2 felvétel egyedi azonosító minta	11
7. ábra Az FCOVER index sokéves átlaga Magyarországra viszonyítva (Kovács & Erdődiné Molnár, 2019).....	12
8. ábra A folyamatsor.....	13
9. ábra Raszter vágása a térkép szélénél	14
10. ábra Bővebb és szűkebb mintaterület (UTM N34Z).....	15
11. ábra A kialakított közet osztályok a mintaterületeken (megnevezések lásd 2. táblázat)(UTM N34Z)	16
12. ábra A Kopernikusz program adatelérési felülete (Copernicus Open Acces Hub).....	17
13. ábra A Hátszegi-medence nyugati részének NDVI indexe(2021.02.25) (UTM N34Z)	18
14. ábra Raszter kalkulátor.....	19
15. ábra Kalkulált vas ásványok főbb összetevőinek képletei (van der Werff & van der Meer, 2016).....	20
16. ábra A használt NDVI osztályok (do Nascimento Aquino, Coimbra da Rocha Neto, Alves Moreira, dos Santos Teixeira, & Maia de Andrade, 2018)	21
17. ábra A kalkulált NDVI alapú maszk 0,3 értékkel (UTM N34Z)	22
18. ábra A kalkulált NDVI alapú maszk 0,35 értékkel (UTM N34Z)	22
19. ábra A legkisebb távolság módszer paraméterei	24
20. ábra A legnagyobb valószínűség módszer paraméterei	25
21. ábra A véletlen erdő módszer paraméterei	26
22. ábra A legkisebb távolság módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z)....	27
23. ábra A legkisebb távolság módszer a 0,35-ös NDVI maszk kategóriában (UTM N34Z) .	27
24. ábra Legnagyobb valószínűség módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z)	28
25. ábra Legnagyobb valószínűség módszer a 0,35-ös NDVI maszk kategóriában(UTM N34Z)	28
26. ábra A véletlen erdő módszer a 0,3-as NDVI maszk kategóriában (UTM N34Z).....	29
27. ábra A véletlen erdő módszer a 0,35-ös NDVI maszk kategóriában (UTM N34Z)	29
28. ábra A legkisebb távolság módszer pontosságai (0,3)	31
29. ábra A legnagyobb valószínűség pontosságai (0,3)	32
30. ábra A véletlen erdő módszer pontosságai (0,3)	32
31. ábra A 0,3-as NDVI kategória osztályozásainak kappa és összesített pontosság értékei ..	33
32. ábra A legkisebb távolság módszer pontosságai (0,35)	33
33. ábra A legnagyobb valószínűség módszer pontossága (0,35).....	34
34. ábra A véletlen erdő módszer pontossága (0,35).	35
35. ábra A kappa együtthatók és az összesített pontosságok a 0,35 NDVI kategóriában	35

Mellékletek

1. melléklet. Legkisebb távolság módszer pontossági táblázat (0,3)

<i>Osztály</i>	Referencia pixelek száma	Előállító pontosság (arány)	Besorolt pixelek száma	Felhasználói pontosság (arány)
1	293985	0.301655	94172	0.941702
2	33456	0.065698	9166	0.239799
3	14283	0.063222	24812	0.036394
4	21850	0.095698	16046	0.130313
5	127377	0.004946	3822	0.164835
6	178719	0.000106	69	0.275362
7	103524	0.002289	836	0.283493
8	344	0.255814	6343	0.013874
9	5539	0.042246	13930	0.016798
10	21804	0.000046	39	0.025641
11	50191	0.016138	10688	0.075786
12	1657	0.003018	10438	0.000479
13	21928	0.001961	1489	0.028878
14	348364	0.000014	5	1.000000
15	158413	0.000101	202	0.079208
16	2124	0.028719	4430	0.013770
17	13424	0.045962	14056	0.043896
18	83283	0.046348	9059	0.426096

2. melléklet. Legnagyobb valószínűség módszer pontossági táblázat (0,3)

Osztály	Referencia pixelek száma	Előállító pontosság (arány)	Besorolt pixelek száma	Felhasználói pontosság (arány)
1	293985	0.050492	29021	0.511492
2	33456	0.149002	70159	0.071053
3	14283	0.002801	32673	0.001224
4	21850	0.000961	6127	0.003427
5	127377	0.000000	0	0
6	178719	0.000173	277	0.111913
7	103524	0.000000	0	0
8	344	0.000000	0	0
9	5539	0.038635	22071	0.009696
10	21804	0.002293	5170	0.009671
11	50191	0.002052	40735	0.002529
12	1657	0.000000	8167	0.000000
13	21928	0.000000	0	0
14	348364	0.000000	0	0
15	158413	0.000259	4274	0.009593
16	2124	0.004708	820	0.012195
17	13424	0.000000	0	0
18	83283	0.000000	0	0

3. melléklet. A véletlen erdő módszer pontossági táblázat (0,3)

<i>Osztály</i>	Referencia pixelek száma	Előállító pontosság (arány)	Besorolt pixelek száma	Felhasználói pontosság (arány)
1	129886	0.989044	140880	0.911861
2	10400	0.701442	11954	0.610256
3	2286	0.291339	766	0.869452
4	5116	0.509382	3786	0.688325
5	9967	0.565968	8132	0.693679
6	15238	0.862384	17566	0.748093
7	8887	0.515697	5469	0.837996
8	198	0.303030	63	0.952381
9	755	0.247682	190	0.984211
10	1256	0.400478	551	0.912886
11	2325	0.504086	1289	0.909232
12	5	0.600000	3	1.000000
13	536	0.509328	280	0.975000
14	15756	0.842028	14580	0.909945
15	5046	0.620095	3493	0.895792
16	185	0.329730	62	0.983871
17	759	0.610013	479	0.966597
18	10893	0.842927	9951	0.922721

4. melléklet A legkisebb távolság módszer pontossági táblázat (0,35 maszk)

<i>Osztály</i>	Referencia pixelek száma	Előállító pontosság (arány)	Besorolt pixelek száma	Felhasználói pontosság (arány)
1	293985	0.374335	121946	0.902440
2	33456	0.091613	13513	0.226819
3	14283	0.198348	73934	0.038318
4	21850	0.127735	28062	0.099458
5	127377	0.040125	15924	0.320962
6	178719	0.000000	0	0
7	103524	0.012693	9518	0.138054
8	344	0.334302	15439	0.007449
9	5539	0.067521	18581	0.020128
10	21804	0.004265	5254	0.017701
11	50191	0.123588	44505	0.139378
12	1657	0.051901	32036	0.002684
13	21928	0.008209	4158	0.043290
14	348364	0.000083	31	0.935484
15	158413	0.019386	23933	0.128317
16	2124	0.202448	49247	0.008731
17	13424	0.297974	52648	0.075976
18	83283	0.024759	14193	0.145283

5. melléklet A legnagyobb valószínűség módszer pontossági táblázat (0,35 maszk)

<i>Osztály</i>	Referencia pixelek száma	Előállító pontosság (arány)	Besorolt pixelek száma	Felhasználói pontosság (arány)
1	293985	0.013729	9102	0.443419
2	33456	0.000000	0	0
3	14283	0.000000	0	0
4	21850	0.059222	37289	0.034702
5	127377	0.000000	0	0
6	178719	0.000129	349	0.065903
7	103524	0.102991	208905	0.051038
8	344	0.000000	0	0
9	5539	0.060119	94166	0.003536
10	21804	0.000000	0	0
11	50191	0.011795	44975	0.013163
12	1657	0.007242	31514	0.000381
13	21928	0.001961	2722	0.015797
14	348364	0.000000	0	0
15	158413	0.000000	0	0
16	2124	0.000000	0	0
17	13424	0.040077	92081	0.005843
18	83283	0.004239	1819	0.194063

6. melléklet A véletlen erdő módszer pontossági táblázat (0,35 maszk)

<i>Osztály</i>	Referencia pixelek száma	Előállító pontosság	Besorolt pixelek száma	Felhasználói pontosság
1	179100	0.958146	205665	0.834386
2	19590	0.491730	16481	0.584491
3	6306	0.189502	1565	0.763578
4	12189	0.270736	4585	0.719738
5	44508	0.653186	57054	0.509552
6	64862	0.797755	84889	0.609549
7	28518	0.342941	12750	0.767059
8	253	0.217391	56	0.982143
9	1623	0.144794	244	0.963115
10	6055	0.201486	1330	0.917293
11	17820	0.513805	13679	0.669347
12	177	0.186441	33	1.000000
13	3492	0.300115	1112	0.942446
14	70593	0.797459	75441	0.746212
15	31131	0.502072	22031	0.709455
16	522	0.174330	92	0.989130
17	5326	0.447615	2714	0.878408
18	30857	0.628609	23201	0.836042



SZAKDOLGOZAT / DIPLOMAMUNKA

EREDETISÉG NYILATKOZAT

AlulírottÁrvai Valentin.....Neptun-kód:I4HK6U.....

ezennel kijelentem és aláírással megerősítem, hogy az Eötvös Loránd Tudományegyetem
Informatikai Karának, Térképtudományi és Geoinformatikai Intézetében írt,
..Geológiai képződmények térképezése műholdképek alapján a Hátszegi-medencében

.....
című diplomamunkám saját, önálló szellemi termékem; az abban hivatkozott szakirodalom
felhasználása a szerzői jogok általános szabályainak megfelelően történt.

Tudomásul veszem, hogy szakdolgozat/diplomamunka esetén plágiumnak számít:

- szó szerinti idézet közlése idézőjel és hivatkozás megjelölése nélkül;
- tartalmi idézet hivatkozás megjelölése nélkül;
- más publikált gondolatainak saját gondolatként való feltüntetése.

A témavezető által benyújtásra elfogadott szakdolgozat PDF formátumban való
elektronikus publikálásához a tanszéki honlapon

HOZZÁJÁRULOK

NEM JÁRULOK HOZZÁ

Budapest, 2021. 05. 14.

.....Árvai Valentin s. k.

hallgató aláírása