

NÖVÉNYZETTÍPUSOK LEHATÁROLÁSA GÉPI TANULÁSSAL LÉGI LIDAR FELVÉTELEK ALAPJÁN EGY ALSÓ-TISZAI HULLÁMTÉRI MINTATERÜLETEN

FEHÉRVÁRY István¹, KISS Tímea²

¹ Alsó-Tisza-vidéki Vízügyi Igazgatóság
6720 Szeged, Stefánia 4., email: fehervaryi@ativizig.hu

² Szegedi Tudományegyetem, Természeti Földrajzi és Geoinformatikai Tanszék
6722 Szeged, Egyetem utca. 2-6., email: kisstimi@gmail.com

Kulcsszavak: döntési fa, scikit-learn, hullámtéri vegetáció, Python, LiDAR pontfelhő, kereszt-validálás

Összefoglalás: Az utóbbi évtizedekben végbemenő környezeti és társadalmi változások hatására a Tisza hullámterén jelentősen nőtt a növényzeti sűrűség, amely hozzájárult az emelkedő tetőző vízszintek kialakulásához. A probléma megoldásához elengedhetetlen a meder és hullámtér megfelelő kezelése. Célunk egy olyan gépi tanuláson alapuló osztályozási módszer kidolgozása, amellyel naprakész, pontos és térben folytonos adatok biztosíthatók a hullámtéri vegetációt kezelők számára. A tanulmányban a növényzetet reprezentáló légi LiDAR pontfelhő statisztikai paramétereit számítottunk ki 15*15 méteres felbontásban, majd egy döntési fa algoritmus segítségével osztályoztuk a kapott eredményeket. Az algoritmus pontosságának ellenőrzésére tízszeres kereszt-validálás (*10-fold cross-validation*) technikát alkalmaztunk. Az algoritmus terepi validációját is elvégeztük, mely során valós növényzeti kategóriákat a döntési fa által adott növényzeti kategóriákkal összesen 72 cella esetében vetettük össze. A kapott eredményeket egy keveredési mátrixban rögzítettük. A vizsgált hullámtéri területen az osztályozás pontossága 83% volt.

Bevezetés

Az ártereket és a folyómedreket az elmúlt közel másfél évszázad alatt számos olyan hatás érte, amelyek megváltoztatták a folyórendszer egészének korábbi jellemzőit. Ezen időszak éghajlatának és területhasználatának változásából eredő környezeti hatásokat azonban messze felülmúlták a folyószabályozási és árvízmentesítési munkák következményei. A folyamat eredményeképpen napjainkban a hullámtereken alapvetően megváltoztak a hidrológiai folyamatok, amelyekhez köthetően az ártéri növényzet is átalakult.

Az ártéri növényzet jelentősen befolyásolja a meder és az ártér kapcsolatát. Például a partok mentén szerepet játszik a partok megkötésében (Abernethy és Rutherford 1998), illetve az ártérre áramló víz sebességének csökkentésében (Kiss et al. 2019b). Míg a partok mentén elsősorban a folyóhátak kialakulását befolyásolja a vegetáció sűrűsége (Nagy et al. 2018), addig az ártér belsőbb részein a fás szárú növényzet alapvetően megváltoztatja az áramlás irányát és sebességét (Rátky és Farkas 2003, Zellei és Sziebert 2003, Brooks 2005, Corenblit et al. 2007, Geerling et al. 2008), és ezen keresztül az ártér vertikális akkumulációs mintázatát (Steiger et al. 2001, Kiss és Sándor 2009).

A fenti folyamatok elsősorban azzal hozhatók összefüggésbe, hogy a növényzet kulcsfontosságú szerepet tölt be az ártéri érdesség és ezzel összefüggésben a vízszintek magasságának alakításában (Jalonen et al. 2015; Kiss et al. 2019b). Az Alsó-Tisza mentén 2000-ben és 2006-ban is megdőlt az 1970-es árvízi rekord, így összességében a rekord vízállás 80 cm-rel emelkedett (Kiss et al. 2019a), miközben a vízhozam nem növekedett (Kovács és Váriné 2003). Mindez az ártéren bekövetkező változásokra hívja fel a figyelmet: jelenleg a sűrű ártéri növényzet miatt az árvíznek kb. 13%-a vezetődik le a hullámtéren, szemben a korábbi 23%-kal (Kovács és Váriné 2003). Véleményünk szerint ebben a helyzetben a megfelelő ártér- és mederkezelés, és az ezt támogató hatásvizsgálatok kiemelt hangsúlyt kell kapjanak. Mivel ennek első lépése az ártéri növényzet tipizálása, célunk ennek elvégzése a rendelkezésre álló legmodernebb módszerek segítségével: azaz légi LiDAR felmérések automatizált feldolgozása révén.

A növényzeti típusok elkülönítésére az utóbbi évtizedekben több olyan módszer került kifejlesztésre, amelyek LiDAR felmérések pontfelhőinek statisztikai elemzéséhez köthetők. Ezeket elsősorban erdészeti és ökológiai kutatásokban alkalmazták. Például Hudak et al. (2008) különböző statisztikai módszerek alapján különítették el a fafajokat az egyes erdőrészekben LiDAR pontfelhő alapján. Heurich és Thoma (2008) LiDAR adatok felhasználásával határozták meg az erdőt alkotó fák bizonyos paramétereit (pl. famagasság és lombkorona átmérő). Naeset et al. (2004) légi LiDAR adatok és terepi felmérések kombinációjával egyes erdőrészek erdészeti paramétereit (pl. törzsek száma, kitermelhető famennyiség térfogata) számították ki. Jung et al. (2011) légi és terepi LiDAR mérésekkel nemcsak a fák paramétereit (magasság, lomb alap magasság, lomb térfogat, törzsátmérő) számították ki, de elemezték a két felmérési mód közötti statisztikai kapcsolatot. Waldhauser et al. (2014) légi LiDAR adatok hatékony osztályozását mutatja be gépi tanulással, illetve az adatok sokszínű ipari felhasználását. Madsen et al. (2020) nagyfelbontású drónos légi LiDAR felméréseket (>1000 pont/m²) ötvözte terepi felmérésekkel a különböző fajtájú bokrok elkülönítéséhez és a biomassza kiszámításához. A fenti kutatások bár más és más módon közelítik meg az növényzeti típusok lehatárolását és parametrizálását, azonban általánosan elmondható, hogy vannak módszerek, amelyek csak kisebb területeken, lokálisan alkalmazhatóak nagy pontossággal, ugyanakkor a nagyobb területre kiterjedő mérési módok csak korlátozott felbontásban kivitelezhetők.

Az utóbbi években a LiDAR alapú növényzet-elemzést már a vízrajzi vizsgálatok során is elkezdték használni. Például Vetter et al. (2011) egy nagy visszaverődési pont-sűrűségű (>25 pont/m²) légi LiDAR felmérés alapján határozták meg a voxelek (3D pixel) térbeli kapcsolata és a voxelekben mért visszaverődések aránya alapján a növényzeti érdességet, azaz, hogy a növényzet milyen mértékben lassítja a vízáramlást. Az eredményeket 2D hidraulikai modellben használták fel, majd a modellezett adatokat összehasonlították egy hagyományos területhasználati térkép alapján számított modell eredményeivel, megállapítva, hogy a LiDAR felmérés alapján valósághűbb eredményeket kaptak. Manners et al. (2011) tamariska (*Tamarix sp.*) cserjék ártér-érdességben betöltött szerepét határozták meg terepi LiDAR mérésekkel.

Mivel a Tisza mentén a 20. század végén elterjedtek az erdők és bennük egyre gyakoribbá válnak az invazív növények, a hullámtér növényzeti érdessége jelentősen megnőtt a hagyományos térképezési adatok alapján (Kiss et al. 2019b). Mivel a növényzeti érdesség alapvetően lassítja és magasítja az árvizeket, ezért a hullámtér teljes területéről pontos és naprakész adatok lennének szükségesek a precíz árvízi modellezéshez. Ezért fő célunk a hullámtéri ártéri növényzet osztályozása légi LiDAR felmérések automatizált feldolgozása révén. Hangsúlyozni szeretnénk, hogy a módszerrel nem lehet cönológiai pontosságú felmérést végezni, estünkben az osztályok a főbb növényzeti típusokra vonatkoznak (pl. gyepek, nemesnyaras, gyalogakácos, ültetett nyaras, ártéri füzes). Ezen cikkben célunk az automatizált osztályozás módszertanának részletes bemutatása, illetve egy alsó-tiszai mintaterületen az osztályozás elvégzése és az eredmények értékelése. Összeségében a hullámtér kezelői számára szeretnénk egy olyan módszert ajánlani, amellyel naprakész információkat kaphatnak adott helyen a növényzet állapotáról, és amelynek segítségével az aktuális beavatkozások könnyebben megtervezhetők.

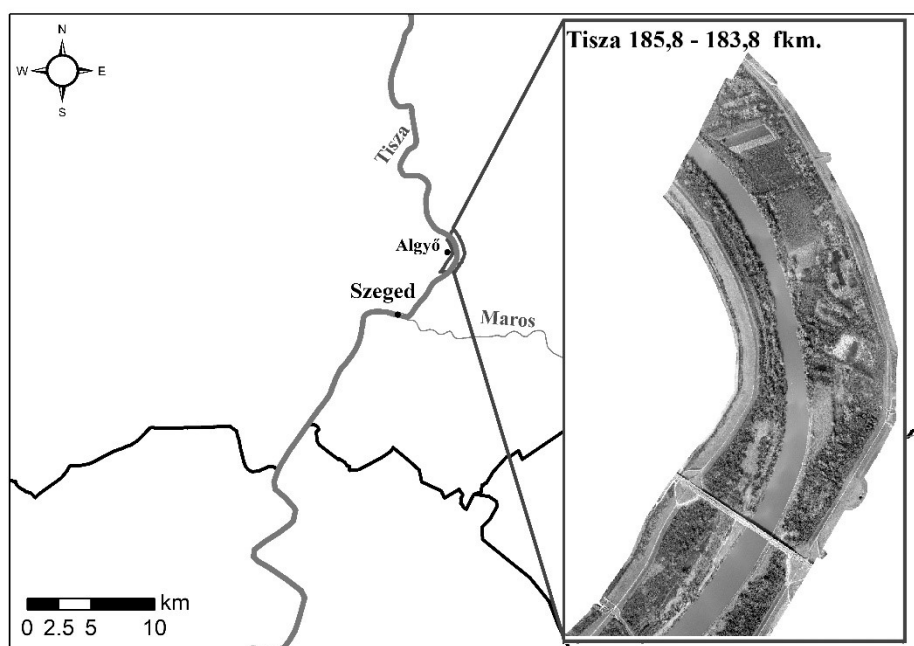
Anyag és módszer

A mintaterület leírása

A kiválasztott hullámtéri terület Szeged és Algyő között (185,8-183,8 fkm) terül el, nagysága 3 km² (1. ábra). Itt a hullámtér átlagos szélessége 800 m, míg a nagyvízi meder átlagos szélessége 130 m. A mintaterülethez közel lévő algyői vízmércén észlelt legmagasabb vízállás 1065 cm, míg a legalacsonyabb -245 cm a vízmérce nullpontjához viszonyítva, azaz a vízjáték 13,1 méter. A hullámtéren 600 cm körüli vízállásnál indul meg a vízmozgás, amely a kicsi esés

miatt nagyon lassú (max. 0,1-0,2 m/s). Árvízkor a töltéskorona szintjéig érhet a víz, ami 5-6 m magas vízoszlop kialakulását jelenti a hullámtéren.

A 19. század közepén kezdődő szabályozási munkálatok (kanyarulatok átvágása és árvízvédelmi töltések építése) jelentős változásokat idéztek elő a mintaterület területhasználatában (Kiss et al. 2019b). Az egykori mocsaras és vizenyős területeket rétek, legelők és szántók váltották fel a 20. század első felében, miközben az erdőterületek aránya alacsony maradt. Azonban az 1980-as években az alsó-tiszai hullámtéren az erdőterületek aránya 70 % fölé nőtt, míg napjainkban arányuk meghaladja a 80%-ot. Összességében négyszeresére nőtt (0,02-ről 0,08-ra) a hullámtér növényzeti érdekessége, amit tovább növelnek (0,13) az erdőkben és a parlagokon az utóbbi évtizedekben jelentősen elterjedt invazív növényfajok. Közülük a mintaterületen kiemelendő a gyalogakác (aránya 11 %), ami árvízvédelmi és ökológiai kockázatot is okoz, mivel legutóbbi vizsgálataink alapján a gyalogakác hatására a hullámtéren az áramló víz sebessége harmadára csökkent, ami közel 20-30 cm-es vízszintemelkedést eredményez a mintaterületen (Kiss et al. 2019b).



1. ábra A mintaterület elhelyezkedése
Figure 1. Location of the study area

Adatforrás

Az elemzéshez használt pontfelhő légi LiDAR (full-waveform) felméréssel készült 2015 kora tavaszán (vízállás: 110 cm), amikor még nem volt lombzat az ágakon, így a fák ágszerkezete szépen kirajzolódik a felvételen. A LiDAR felméréssel egyidőben 10 cm felbontású ortofotó is készült a vizsgált területről. A légi LiDAR felmérés során 22,5 millió visszaverődő pontot rögzítettek a mintaterületen, összesen 8 darab .las formátumú file-ban. A legalacsonyabb pontokat a folyó vízszintjében lévő felszín jelenti (75,1 m Bf.) míg a legmagasabbak a 30-35 m magas fák lombkoronájának tetejét jelzik. A LiDAR működési elvéből adódóan az erősen vizenyős területről és a nyílt vízfelszínéről nem érkezik visszaverődés. Ezeket a területeket leszűrve az átlagos pontsűrűség a mintaterületen 9 pont/m², ami elégséges a vizsgálatokhoz, hiszen az erdőterületek elemzéséhez minimum 4 pont/m² pontsűrűsége van szükség Laes et al. (2008) szerint. A vizenyős területeken kívül a vizsgált terület csupán 3%-a nem tett eleget ennek a feltételnek, ami tehát nem rontotta az elemzés megvalósíthatóságát és minőségét.

Vizsgálatainkhoz a pontfelhő alapján készített a 0,5 m felbontású digitális domborzatmodellt használtuk fel. A pontfelhőt, a domborzatmodellét és az ortofotót a EuroSense Kft. készítette.

Elemzési módszer

Az elemzések elkészítéséhez a Fusion 3.8 és az ArcMap 10.6.1 szoftvereket használtuk, az osztályozó algoritmust – döntési fát – pedig Python program nyelven a scikit-learn (0.22.1) könyvtár (Pedreagosa et al. 2011) felhasználásával írtuk meg. A döntési fát az sklearn.tree modul *DecisionTreeClassifier* osztályával a *Gini-index* alapján meghatároztuk meg. A döntési fa algoritmus ideális paramétereinek megtalálásához a *GridsearchCV* osztályt használtuk többszörös kereszt-validációs (*K-fold cross validation*) módszerrel. A döntési fa részletes leírását az Eredmények fejezetben ismertetjük. A döntési fa alapján, a teljes mintaterületen meghatároztuk az egyes pixelek növényzeti típusát, majd terepi mérések alapján validáltuk az eredményeket.

Az adatok előkészítése az elemzésekhez

Az előkészítési fázisban a *.flt* formátumban lévő domborzat modellt *.dtm* formátumra alakítottuk át, hogy a Fusion programban a további számításokhoz használni tudjuk. Ehhez először ArcMap szoftver segítségével a domborzatmodellét a *RasterToAscii* eszközzel *.ascii* formátumra alakítottuk, majd a Fusion programban az *ASCII2DTM* paranccsal *.dtm* formátumra konvertáltuk.

A következő lépésben LiDAR adatokat tartalmazó *.las* fájlokban tárolt pontfelhő minőségi ellenőrzését végeztük el, a kiugró értékek és a visszaverődések számának elemzése alapján. A műveletet a Fusion programban a *Catalog* paranccsal végeztük. A szoftver eredményként minden *.las* fájlról elkészített egy jelentést, amely a pontok számát, a minimum és maximum magasság értéket, valamint a visszaverődések sűrűségét (db/pixel) tartalmazza. A kiugró értékeket a *FilterData* paranccsal töröltük a pontfelhőből. Kiugró adatnak azon pontokat tekintettük, amelyek 100*100 méteres cellafelbontásban a kiszámolt átlagmagasság értéktől negatív vagy pozitív irányba a szórás értékének tízszeresénél nagyobb értékben tértek el.

A térbeli felbontás meghatározása

A felbontás megválasztása fontos kérdés mivel, ha túl alacsony értéket választunk, akkor a pontfelhő túlságosan fel lesz darabolva és az egyes növénytípusokra jellemző paraméterek nem fognak elkülönülni. Azonban minél nagyobb térbeli felbontást választunk, annál kisebb térbeli változékonyságot lehet kimutatni, és annál nagyobb esély van az osztályok keveredésére. A szakirodalom a térbeli felbontás megválasztásához a mintaterületen lévő átlagos lombkorona átmérőt ajánlja (Laes et al. 2008). Ezért választottuk ki a 15x15 méteres felbontás értéket. Ez azt jelenti, hogy a LiDAR pontfelhőt 15x15 méteres alapterületű cellákra bontottuk, amelyek magassága a cellában lévő legmagasabb növény magasságával egyezik meg.

A cellák statisztikai paramétereinek kiszámítása

A következő lépésben a növényzetet reprezentáló pontfelhő leíró statisztikai paramétereit számítottuk ki 15x15 méteres felbontásban (cellákban), a Fusion program *GridMetrics* eszköze segítségével. A bemenő fedvényeket a szűrt pontfelhő és a *.dtm* formátumú domborzat jelentette. A programban lehetőség van egy *heightbreak*, azaz magassági töréspont beállítására is, amely alapján a szoftver újra számítja a cellákban (néhány paraméter esetében) a statisztikai változó értékét (pl. első visszaverődések aránya és első visszaverődések aránya egy általunk megadott magassági töréspont fölött). Az erdők osztályozásához jelen esetben 6 méteres értéket választottuk, elsősorban azért, mert ebben a magasságban különül el az aljnövényzet a lombkorona szinttől, és célunk a gyalogakác által nem befolyásolt erdőtípusok elkülönítése volt. Az eredményfájlban *.csv* formátumban tároltuk el az eredményül kapott pixelekre vonatkozó statisztikai adatokat. A program a magassági adatokból 55 statisztikai változót

számol ki. Ebben a cikkben nem részletezzük az összes változót, a következőkben csak azokat mutatjuk be részletesen, amelyeket a terület növényzeti típusainak elkülönítésére megalkotott döntési fához használtunk fel.

A *lombkorona relief-arány* (*canopy relief ratio*; CRR) számítása során a pixel pontjainak átlag és a legkisebb magassági értékeinek a különbségét osztottuk el a legnagyobb és minimum magasság különbségével $[(\text{átlag}-\text{min})/(\text{max}-\text{min})]$ McGaughey (2018) alapján. Tehát minél nagyobb a magasságkülönbség az átlag és a maximum értéke között, annál kisebb a CRR értéke. Ez a paraméter utal a lombkorona terebélyességére, hiszen minél magasabb és szélesebb a lombkorona, annál közelebb kerül egymáshoz az átlag és a maximum értéke. A mintaterületen az idős, tagolt lombkoronájú fehér nyarasokra 0,2-0,3 körüli CRR értékek a jellemzőek, míg a fiatal, sudár lombkoronájú nemes nyarasok alig érik el a 0,03-0,04 értéket. A nyílt felszínek és a rövidfűvű gyepek esetében a legnagyobbak az értékek (0,4-0,5), hiszen az átlag és a maximum magasság szinte megegyezik ezen osztály esetében.

A *cellák magassági értékeinek szórását* (*Elev_stddev*) a cellákban lévő összes pont szórásának kiszámolásával kapjuk meg. A szórás utal a cella pontjainak magassági diverzitására, azaz a lombkorona vertikális tagoltságára és sűrűségére: minél laposabb és adott szinten tömöttebb a lombkorona, annál homogénebb a pontok eloszlása a cellában, tehát annál kisebb szórásértéket kapunk. A mintaterületen a nyílt felszínek esetében, amikor szinte minden pont a terepet vagy a gyepet reprezentálja, a szórás értéke 0,03 m-nek adódott, míg az ártéri füzesek esetében a szórás értéke két nagyságrenddel nagyobb (3-4 m). A legnagyobb szórás értékek (8-10 m) a kiterjedt lombkoronájú fehér nyarasokra jellemzőek.

A *cella magassági értékének 99%-a* (*Elev_P99*) azt jelenti, hogy hol van az a magassági érték, ahol a cellát reprezentáló pontok aránya a fölfelszíntől számítva eléri a 99%-ot. Kiszámítása során az adott cellába tartozó pontfelhő magassági értékeit sorba rendezzük és kiválasztjuk a 99%-hoz tartozó magasság értéket, ami közel a maximumot jelenti. A mintaterületen például az ártéri füzesekre jellemző értéket (~15 m) jelentősen meghaladja a foltszerűen megjelenő fehér nyarasok átlagos értéke (~25 m).

A *cella magassági értékének 95%-a* (*Elev_P95*): a változó értékét a 95%-os percentiliszhez tartozó magassági érték jelenti. Kiszámítása során a magasságértékeket sorba rendezzük és kiválasztjuk a 95%-hoz tartozó magasság értéket.

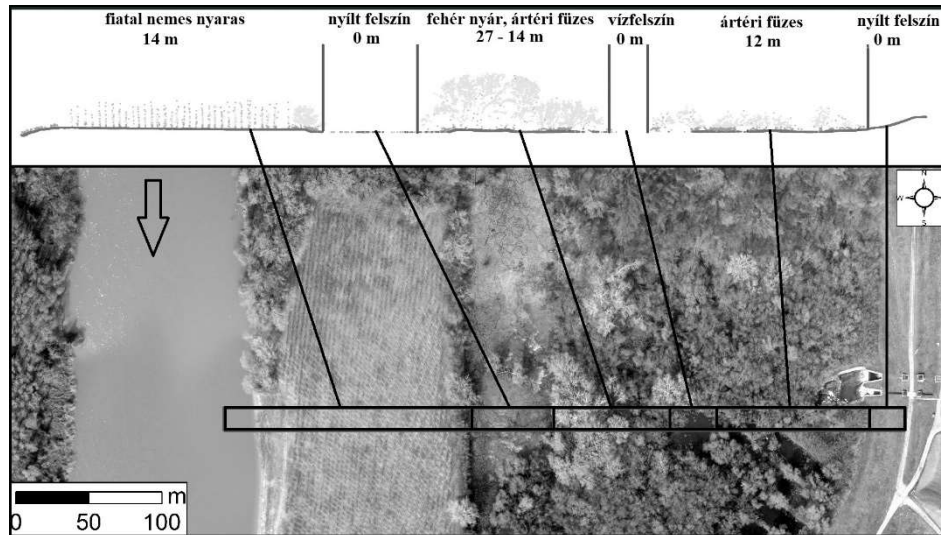
A *cella magassági pontjainak eloszlásgörbéjének ferdesége* (*Elev_skewness*) a cellában lévő pontok magasságát reprezentáló eloszlási görbe alakjára utal, annak szimmetriáját adja meg. Szimmetrikus eloszlásnak tekintjük például a standard normál eloszlást, amelynek a ferdesége nulla. A ponteloszlás ferdeségét (aszimmetriáját) legnagyobb mértékben a kiugró értékek határozzák meg. Ha a kiugró értékek az alacsony értéktartományban vannak, akkor a ferdeség értéke negatív lesz, míg, ha a magas értéktartományban tartoznak a ponteloszlás szélső értékei, akkor a ferdeség értéke pozitív lesz. A mintaterületen magas értékű (4-6) pozitív ferdeség jellemző a fiatal nemes nyarasokra és a sudár, vékony fákra: esetükben a lombkorona nem záródott tökéletesen, így a felszínről is nagy számban verődnek vissza pontok és a fák csúcsajtásairól visszaverődő adat is kiugró értéknek fog számítani. A nyílt felszínek esetében a ferdeség értéke alacsony (0,1-0,4) hiszen nem jelentkeznek kiugró pontok, a visszaverődések az egyenletes felszínt adják vissza.

Tanulóterületek kijelölése és a növényzeti típusok definíciója

A tanulóterületek kijelölésének első lépéseként meghatároztuk az azonosítandó növényzeti típusokat: nyílt felszín, gyalogakácós bozót, fiatal nemes nyaras, nemes nyaras, ártéri füzes és fehér nyaras. A kategóriák előzetes meghatározását korábbi terepi bejárásaink és Magyarország Erdészeti Webtérképe (<http1>) alapján végeztük.

Terepi bejárások és a rendelkezésre álló ortofotó alapján homogén növényzetű, 15x15 m pixelméretű tanulóterületeket jelöltünk ki minden osztályhoz kapcsolódóan (osztályonként 40-

50 darabot) az algoritmus számára. A tanulóterületek kijelölése során törekedtünk arra, hogy a lehető legtisztábban válasszuk ki a betanításra a pixeleket, azaz a kiválasztott 15x15 m cella homogén legyen felszínfedettség szempontjából, azaz peremhatástól mentes. A homogén cellák kiválasztásában segítségünkre volt az ortofotó (ellenőrizhető a növényzeti folt jellege), a cellák (növényzet) magassági értéke, valamint a növényzetet reprezentáló pontfelhő alakja. A különböző növényzetű cellákra eltérő ponteloszlások és így eltérő statisztikai paraméterek a jellemzők, melyek jó kiindulási alapot nyújtottak a további munkához (2. ábra).



2. ábra Különböző növényzeti kategóriák megjelenése a LiDAR pontfelhő oldalnézeti képén és az ortofotón
 Figure 2. Various vegetation types represented by the vertical view of the LiDAR point-cloud, and its appearance on the ortho-photo

A nyílt felszín típusába az alacsony fűvű és lágyszárúakkal borított területek, illetve a csupasz felszínek (pl. szántó) tartoznak. Gyepesek a mintaterületen nagyrészt az árvízvédelmi töltéseken fordulnak elő. Mivel a LiDAR felvétel kora tavasszal készült, ezért a visszaverődő pontok gyakorlatilag a felszínnel egyeznek meg, így felszíntől számított növényzeti magasság érték 0 m körüli. A gyalogakácos területeken a sűrű növényzet pontfelhője 1-6 m közötti magasságra koncentrálódik, a lombkorona teteje csaknem homogén, és a pontok sűrű rétegben helyezkednek el a gyalogakácos felső szintjében. Az ártéri fűzesekre a 16-20 méteres magasság jellemző, az ortofotón pedig a barnás-narancssárgás árnyalatok jelzik a hamarosan rügyező fűzfákat. A nemes nyarasokat egyenlő sorközben telepítik, így a sorközök miatt megfigyelhető volt ezeken területeken az első visszaverődések szórásának megnövekedése (azaz sok olyan pont volt, ami vagy a nyárfák közötti felszínről verődött vissza, vagy a lombkorona tetejéről). A LiDAR pontfelhőn jól látható gyér ágszerkezetük (gyakorlatilag a törzs közvetlen közelében lévő pontok), míg az ortofotón homogén lombkoronájuk árulkodó. A nemes nyarasokon belül elkülönítettük a fiatal nemes nyaras ültetvényeket, amelyekre az alacsony magasság (> 5,6 m), fejletlen ágszerkezet és nagy sortávolság jellemző. Ezen kategóriába esnek azok a pixelek is, amelyek egyedülálló alacsony fákat vagy bokrokat reprezentálnak. A fehér nyarasok általában a fűzesek közé ékelődnek be elszórtan, kis egyedszámmal. A fehérynár fehér-szürkés árnyalatokkal és nagyon jellegzetes ágszerkezettel jelenik meg az ortofotón és a LiDAR pontfelhőn is.

A döntési fa létrehozása

A kiválasztott tanulóterületek statisztikai paramétereit .csv formátumban mentettük el, amelyben megadtuk a növényzeti típusok neveit és a kiszámolt statisztikai paraméterek (55) neveit is.

A következő lépésben a döntési fa algoritmust parametrizáltuk. A döntési fák az elemek, esetünkben a cellák attribútumain alapuló döntések sorozatával osztályoznak. Az attribútumok kiválasztása és az elválasztó érték meghatározása egy kiválasztott számítási algoritmus alapján történik. A szétválasztás célja, hogy a legtisztább, leghomogénebb osztályokat kapjuk. Az általunk használt döntési fa a Gini-index alapján határozza meg az elválasztó értékeket, mert ez gyorsabban futtatható, mint az entrópia alapú számítás, ráadásul nincs jelentős minőségbeli különbség az osztályozás pontosságát tekintve. A Gini-index azt mutatja meg, milyen valószínűséggel rakunk egy populációból véletlen kiválasztott elemet helytelen osztályba (Grabmeier és Lambe 2007). Amennyiben értéke nulla, az azt jelenti, hogy a kritérium teljesen tisztán elválasztotta a létre jövő osztályt a teljes populációtól, míg az 1,0-es érték egy teljes mértékben diverz osztályt jelent.

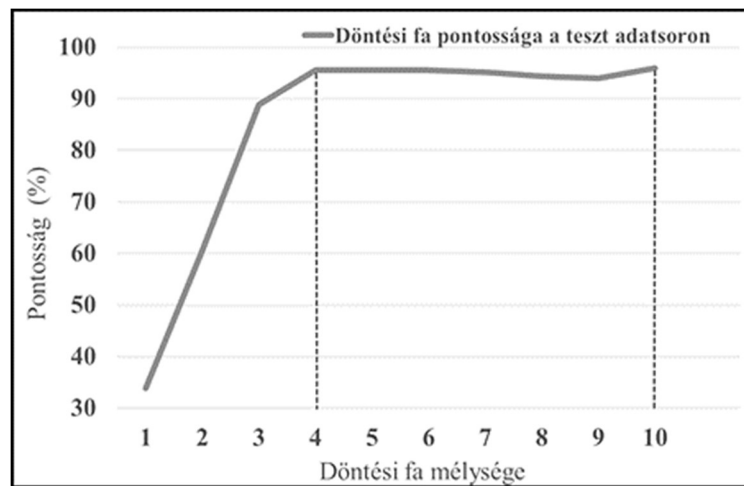
A döntési fa paramétereinek beállítását automatizáltan végeztük el a *GridsearchCV* modul segítségével, figyelembe véve a (1) döntési fa maximális mélységét; (2) a döntési fa leveleinek minimális elemszámát; és a (3) döntési fa leveleinek további felosztását meghatározó minimális elemszámot. A döntési fa maximális mélysége azt adja meg, hogy hány szintből álljon a döntési fa. A döntési fa leveleinek minimális elemszáma azt szabályozza, hogy hány elem maradhat minimálisan a döntési fa levelein, a felosztásra vonatkozó minimális elemszám pedig azt határozza meg, hogy mi az a minimális elemszám, ami még a levél további osztályozását vonja maga után. Az ideális paraméterek kiválasztásához minden változónak megadtunk egy értékintervallumot (maximális mélység 1–10, minimális elemszám 2–10, minimális felosztási érték 2–20), és ezen intervallumokból került ki az a paraméter együttes, ami a legpontosabb döntési fát hozta létre a tesztadatokra vonatkozóan.

Az algoritmus pontosságának ellenőrzésére egy kereszt-validációs technikát alkalmaztunk, melyet széleskörűen használnak a gépi tanulós problémák megoldásában. Többfajta kereszt-validációs technika is létezik, az általunk használt tízszeres kereszt-validálás (*10-fold cross-validation*) módszer lényege, hogy felosztjuk a tanulóterületeket reprezentáló adatsort 10 részre, majd az algoritmus az egyik részt félreteszi a validációra. A kereszt-validálás 10 ilyen iteráció után fejeződik be, ahol minden létrejövő (10 elemű) adatsor-részlet pontosan egyszer validációs adathalmazként lesz felhasználva (Bengio és Grandvalet 2004). A pontosság becslésére 10 db eredmény átlagolása után kerül sor. A módszer előnye, hogy az adathalmaz minden pontja fel lesz használva tanításra és validációra is. A módszer hátránya, hogy számításigényes, hiszen jelen esetben 10-szeres tanítást végeztünk. Ezt a módszert kiegészítésként használtuk a terepi validáció mellett. Az osztályozás pontosságát kategóriákra lebontva a terepi felmérés adta meg.

A létrejövő osztályozás pontosságát az adja meg, hogy a teszt adatsor hány százalékát sikerült jó osztályba tenni, ugyanakkor a számérték önmagában nem adja meg, hogy mennyire sikerült hatékony osztályozást létrehozunk. A döntési fa algoritmus alkalmazásánál kétfajta módszertani hiba követhető el. Az egyik, hogy létrejövő döntési szabályok nem szűrik le elég tisztán az osztályokat és alacsony lesz (underfit) a döntési fa osztályozási pontossága. Amennyiben nagy adatmennyiséggel és sok változóval rendelkezünk, ez a ritkább eset. Sokkal gyakrabban előfordul, hogy a döntési fa „túltanul” (overfit) és olyan szabályokat hoz létre, melyek egyedileg csak a teszt adatsor néhány elemére igazak. Ebben az esetben a pontosság nagyon magas lesz (>95%), ugyanakkor fa nem lesz hatékonyan alkalmazható más adatsorokon (Schaffer 1993).

A paraméterek kiválasztásának automatizálása és a tízszeres kereszt-validálás használata is arra irányul, hogy azt a döntési fát találjuk meg, amely a „túltanulás”-ból és a pontatlanságból

adódó hibákat a lehető legjobban kiküszöböli. A futtatások vizsgálata során látható volt, hogy a fa maximális mélysége az a paraméter, ami a legnagyobb hatással van a döntési fa működésére. Az eredmények azt mutatták, hogy ha a döntési fa mélysége nagyobb lesz, mint 4, a pontosság már nem javul jelentősen, azonban a „túlfitelés” kockázata jelentős lesz (3. ábra). Ezért a vizsgálat során a döntési fa mélységét 4-ben határoztuk meg. A minimális elemszámra és a minimális felosztási elemszámra vonatkozó paraméterek estében a futtatás során kapott 2-es értéket használtuk. A tanulóterületre az általunk létrehozott döntési fa pontossága a tízszeres kereszt-validálás alapján 92%.



3. ábra A döntési fa pontossága különböző maximális döntési fa mélységeknél

Figure 3. Relationship between the depth of the decision tree and the accuracy of the classification

Az döntési fa kiterjesztése a teljes mintaterületre és validálása

A döntési fát követve növényzeti kategóriákba soroltuk a mintaterületet reprezentáló 11656 pixelt (*sklearn.fit*). Az osztályozás eredményességét terepi mérésekkel ellenőriztük 2019 telén. A terepi validációhoz DJI Phantom III Pro drónnal készített légifotókat használtunk fel. A fotókat 30-60 m magasságból készítettük el ortogonális pozícióból (90°-kal megdőntött kameraállásból). A drónos fotók mellett a terepen is készítettünk felvételeket a növényzeti kategória meghatározáshoz. A képek alapján és a terepbejáráskor meghatároztuk a cellák növényzeti típusát. Összesen 72 ponton készítettünk képeket a mintaterületen, törekedve arra, hogy osztályonként egyenlő legyen a kontrollpontok száma, bár az osztályok egyenlőtlen területi aránya ezt megnehezítette. A nyílt felszíneken, gyalogakácosokban és fehér nyarasokban 12-12 felvételezési pontunk volt, míg a nemes nyarasokban 11. A fiatal nemes nyarasok kis területe miatt csak 6 pontot tudtunk rögzíteni a mintaterületen, míg az áltéri füzesek nagy kiterjedése miatt 19 felvételezési pontot rögzítettünk. Meg kell jegyeznünk, hogy az egyenlőtlen kontrollpont szám a fiatal nemes nyarasok esetében a pontosság túlbecslését eredményezheti. A képek készítésének koordinátáit kinyertük, majd a döntési fa által létrehozott növényzeti kategória térképről kiolvastuk a ponthoz tartozó növényzeti típust. A cellák automatizált módon meghatározott növényzeti kategóriáját összevetettük a légifotók és a terepbejárás alapján meghatározott vegetáció típusal. Az összehasonlítás eredményeit egy keveredési mátrixban (confusion matrix) foglaltuk össze. A mátrixban lévő sorok és oszlopok száma megegyezik az osztályok számával. A főatlóban található a helyesen osztályozott pontok aránya, míg az adott sor x-edik eleme adja meg azoknak a pontoknak az arányát, amelyeket az osztályozó algoritmus helytelenül osztályoz.

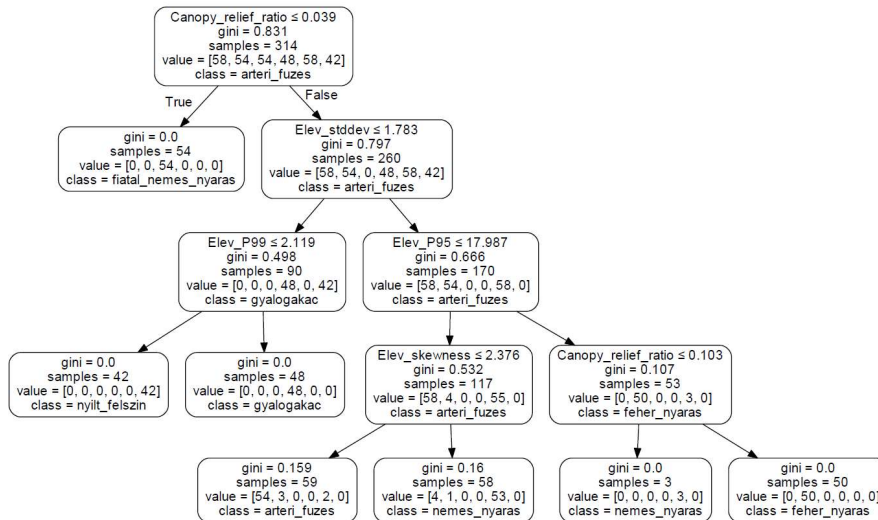
Eredmények

A döntési fa

A döntési fa a fent leírt folyamat alapján választotta ki az osztályozó paramétereket. Az algoritmus első lépésben a fiatal nemes nyarasokat szűrte le a $CRR \leq 0,039$ alapján. Ez a feltétel jól elkülöníti a fiatal, alacsonyabb és fejletlen lombkoronával rendelkező fákat a magasabb és fejlett lombkoronával rendelkező fáktól és a nyílt felszínektől is. A következő lépcsőként a hamis ág esetében ($CRR \geq 0,039$) a cella pontjainak szórása alapján (Elev_std) lettek leszűrve a gyalogakácos bozótok és a nyílt felszín. Ezen felszín növényzeti típusok pontjainak eloszlása homogénebb, mint a fásszárú típusoké, így a szórás alapján jól elkülönülnek. A nyílt felszínnek és a gyalogakácos bozótok magasságuk alapján jól elkülöníthetők, így a két növényzeti típus szétválasztására az (Elev_P99) magassági paramétert használta az algoritmus. A nyílt felszíneket azok a cellák képviselték, ahol a $CRR \geq 0,039$, és $Elev_std \leq 1,783$, és $Elev_P99 \leq 2,119$ feltételek teljesültek. A gyalogakácos, bozótos felszínfedettségű osztályba azon cellák kerültek, melyeken az alábbi szűrési paraméterek teljesültek: $CRR \geq 0,039$, és $Elev_std \leq 1,783$, és $Elev_P99 > 2,119$. A tesztadatok alapján a fiatal nemes nyarasok, a nyílt felszínnek és a gyalogakácos bozótok kategóriákat teljesen tisztán (Gini-index = 0) ki lehetett szűrni (4. ábra).

A szórás feltétel ($Elev_std \geq 1,783$) hamis ágán a nemes nyaras, a fehér nyaras és az ártéri füzes kategóriák maradtak. A mintaterületen az idős fehér nyarasok magasságukat tekintve kiemelkednek az hullámtéri vegetációból, ezért az ártéri füzesektől és a nemes nyarasoktól az $Elev_P95 > 17,987$ paraméter alapján lettek leszűrve. Ez a szűrési feltétel nem lett teljesen tiszta, mivel néhány magasabb nemes nyárfát tartalmazó cella is eleget tett ennek kritériumnak. Ezeket a nemes nyarasokat a CRR paraméter alapján lehetett elkülöníteni a fehér nyarasoktól. A nemes nyarasokra a kisebb kiterjedésű lombkorona miatt alacsonyabb CRR érték jellemző, így a $CRR \leq 0,103$ feltétel jól elkülönítette az idős nemes nyarasokat a fehér nyarasoktól. A fehér nyarasokat az alábbi feltétel-rendszerrel lehetett leszűrni a mintaterületen: $Elev_std > 1,783$ és $Elev_P95 > 17,987$ és $CRR > 0,103$. A nemes nyarasok egy kisebb hányada (idősebb, magasabb egyedek) is ezen az ágon szűrődtek le, csak a CRR paraméterben volt különbség ($Elev_std \geq 1,783$ és $Elev_P95 > 17,987$ és $0,039 < CRR \leq 0,103$). A tesztadatok alapján a magas (≤ 18 m) és idős nemes nyarasokat és a fehér nyaras kategóriákat teljesen tisztán (Gini-index = 0) ki lehetett szűrni a fent említett feltételek alapján (4. ábra).

Az $Elev \leq 17,987$ feltétel igaz ágán az ártéri füzesek és a kevésbé idős és alacsonyabb nemes nyarasok maradtak. A nemes nyarasok karcsú, oszlopszerű lombja és a sorközeik nagy eséllyel aszimmetrikus ponteloszlást eredményeznek a cellákban, és arányaiban kevesebb visszaverődés érkezik a lombkoronáról, mint a füzesek esetében. Mivel a ferdeség paraméter jól számszerűsíti a pontok aszimmetrikus eloszlását, így az $Elev_skewness > 2,376$ megbízhatóan elkülöníti a nemes nyarasokat a füzesektől. A nemes nyarasok elkülönítését az alábbi feltételek írják le: $CRR > 0,039$ és $Elev_std > 1,783$ és $Elev_P95 \leq 17,987$ és $Elev_skewness > 2,376$. Az ártéri füzesek szűrési paraméterei szinte megegyeznek a nemes nyarasok feltételrendszerével, csak a ferdeség paraméterben különböznek ($CRR > 0,039$ és $Elev_std > 1,783$ és $Elev_P95 \leq 17,987$ és $Elev_skewness \leq 2,376$). A természetben akár egy 15x15 méteres felszínen belül is megfigyelhető a nyarasok és füzesek keveredése, és ez a jelenség a tesztadatok osztályozásán is megfigyelhető. Az ártéri füzesek és nemes nyarasok osztályába is került néhány más osztályból származó adat, azonban az osztályozás még így is effektív volt (Gini-index < 0,16; 4. ábra).

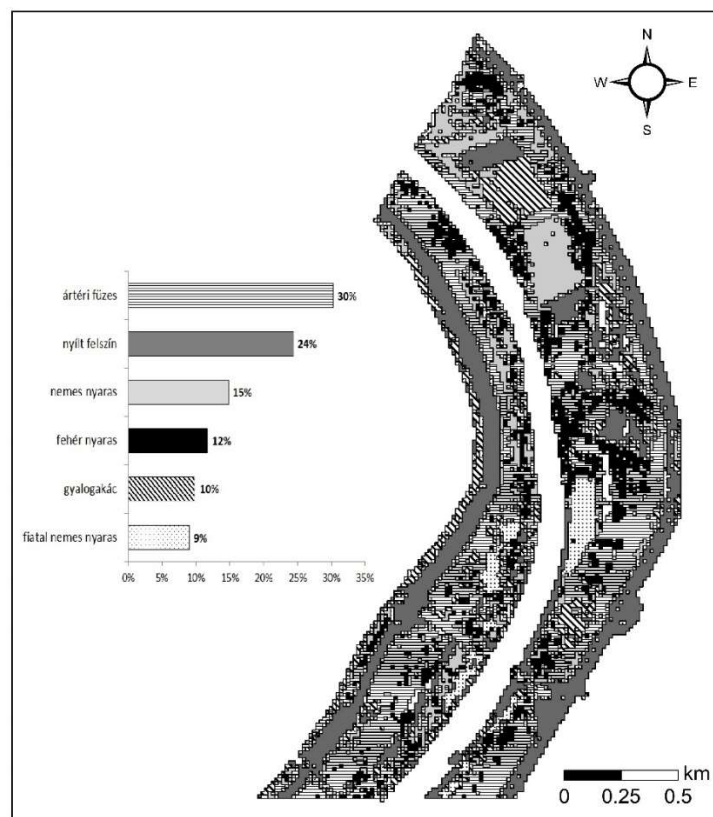


4. ábra A mintaterület tanulóterületei alapján felépített döntési fa, amely lehetővé teszi az egyes növényzeti típusok elkülönítését

Figure 4. Decision tree built up based on training plots of the study area; the decision tree was applied to classify the vegetation types on the entire study area

A mintaterület növényzeti típusai a döntési fa alapján

A tanulóterületek alapján létrehozott döntési fát alkalmaztuk a teljes vizsgált mintaterületre, azaz a 11656 darab pixelre is. Így elkészítettük a 3 km²-es terület növényzeti térképét (5. ábra).



5. ábra A mintaterületen a döntési fa alapján azonosított növényzeti típusok térbeli kiterjedése

Figure 5. Vegetation type map based on the classes of the decision tree

A mintaterületen legnagyobb területet az ártéri füzesek foglalják el (30%, 80 ha). Ezek a füzesek elsősorban a gátak előterében lévő kubikgödrök közötti gerinceken és a mélyfekvésű térszíneken fordulnak elő. A nyílt felszín kategóriáját (24%, 63 ha) elsősorban a bal és jobb oldali árvízvédelmi töltések gyeptakarója képviseli, de ide tartozik még néhány nagy kiterjedésű szántó és a takarás nélküli kubikgödrök (amelyekben a felvételezéskor elpusztult lágyszárúak hevertek a felszínen). A mintaterületen viszonylag kis arányban (15%, 40 ha) fordulnak elő ültetett nemes nyarasok, elsősorban nagyobb foltokban. A lefolyás szempontjából kritikus gyalogakácos, bozótos növényzeti típus a terület 10 %-án (25 ha) fordul elő, elsősorban más növényzeti típusok peremén, illetve egykori parlaggá vált parcellákon. A legkisebb kiterjedésű (9%, 23 ha) felszín-fedettségi kategória a fiatal nemes nyarasok kategóriája, ami tartalmazza a fiatal, fejletlen fákkal és bokrokkal borított gyér növényzetű foltokat is (5. ábra).

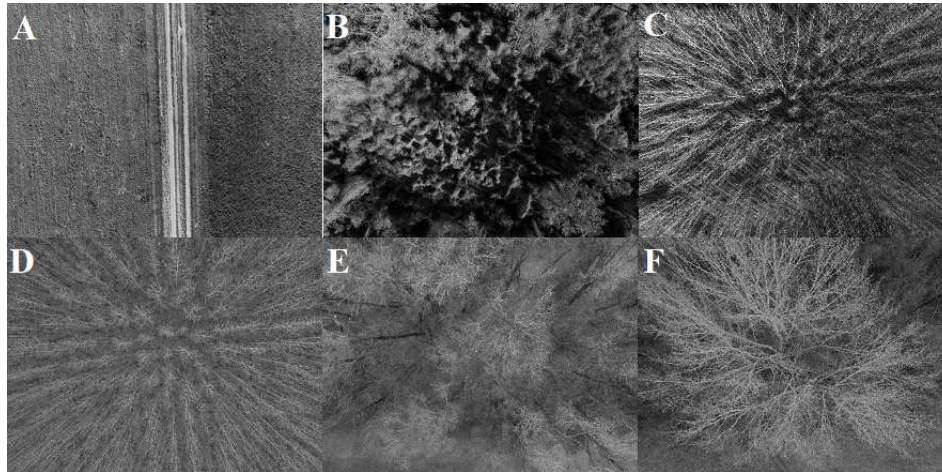
Az eredmények validálása

Az osztályozás pontosságát úgy kaptuk meg, hogy összehasonlítottuk a valós növényzeti kategóriákat (6. ábra) a döntési fa által adott növényzeti kategóriákkal összesen 72 cella esetében. A kapott eredményeket egy tévesztési vagy más néven keveredési mátrixban rögzítettük (1. táblázat).

A mátrix alapján a döntési fa osztályozási pontossága 83%-os. A nyílt felszínek osztályozási pontossága adódott a legalacsonyabbnak (75%). Ennek oka, hogy az általunk nyílt felszínként meghatározott területet a vizsgált cellák 8 %-ában gyalogakácosként osztályozta a döntési fa, míg 17%-ban fiatal nemes nyarasként. Ezek a hibák szinte minden esetben a töltés és az ártéri erdőfoltok határán jelentkeztek. Ezeken a határterületeken a valóságban is húzódik egy fiatal hajtásokból álló sáv, továbbá a belógó alacsony ágak is okozhatták ezt az osztályozási problémát. Az ártéri füzesek osztályozási pontossága a vizsgált esetek alapján 84 %-nak adódott. Egyenlő arányban (5-5 %-ban) az ártéri füzeset a döntési fa nemes nyarasként, gyalogakácosként vagy fehér nyarasként osztályozta. Ez a hiba több forrásból ered: (1) a LiDAR felvételezés (2015) és a validálás (2018) közötti vágtak ki füzeseket, amelyeken a terepbejáráskor már gyalogakác nőtt; (2) a természetben a 15x15 méteres cellákban keveredhetnek ezek a típusok; (3) a növényzeti folt korától függően a fák rendelkezhetnek nagyon hasonló magassággal és akár lombkorona paraméterekkel is. A gyalogakácos területeket osztályozta a legpontosabban (92%) a döntési fa. A terepi mérések alapján, mindössze 8%-ban sorolta az algoritmus az ártéri füzeseket gyalogakácos kategóriába. Azonban ez a tévesztés nem tekinthető komoly hibának, hiszen egy gyalogakácosval erősen fertőzött ártéri füzes területen jelentkezett. A fehér nyarasok osztályozási pontossága 83%-nak adódott, a valóságban fehér nyaras területet a döntési fa 8%-ban nemes nyarasnak és további 8 %-ban fiatal nemes nyarasnak osztályozta. A tévesztések mindkét esetben vegyes cellákban történtek, ahol a fehér nyarasok ültetett nyarasokkal keveredtek. A vizsgált esetek alapján a fiatal nemes nyarasok osztályozási pontossága is 83%-nak adódott, az esetek 17%-ban lettek fiatal nemes nyarasok gyalogakácosként osztályozva. Ez azzal magyarázható, hogy a fiatal nemes nyarasok magasságukban és sűrűségükben is hasonlítanak a gyalogakácosokra, ráadásul, ha nem gondolják őket, akkor elterjedhet bennük a gyalogakác. A terepi validáció szerint a döntési fa osztályozási pontossága nemes nyarasok esetében 82% volt. A nemes nyarasok fehér nyarasként való osztályozása volt a keveredési mátrixban a legnagyobb hibaarány, ami a nemes és fehér nyár hasonló magassági viszonyaival magyarázható.

Az osztályozás pontossága (83%) a szakirodalomban található adatokhoz illeszkedik, mivel sűrű bokros, illetve ártéri területeken hasonló eredmények születtek. Például ártéri erdőkben Saarinen et al. (2013) mobil lézer szkennerekkel adatok osztályozásával 72,6%-os pontosságot ért el, míg Michez et al. (2016) 79.5-84.1%-os pontosságot kapott drónnal felmért pontfelhő osztályozása során. Bokros területen Madsen et al. (2020) 86,9-95,2%-os osztályozási pontosságot ért el légi lidar adatok osztályozása során. Az eredmények a modellezési és

tervezési munkálatok során jól hasznosulhatnak, nagyfelbontású pontos képet adnak raszteres vagy vektoros formában a hullámtéri vegetációról. A jövőben tervezett, növényzeti kategóriánként megvalósuló sűrűség vizsgálatok további hasznos információkat szolgáltatnak majd a modellezés egyik alapparaméterének az érdesség meghatározásához, valamint az árvízi levezető sávok meghatározásához.



6. ábra Növényzeti típusok a mintaterületen a drónfelvételek alapján. A: nyílt felszín (gyep), B: gyalogakác, bozót, C: fiatal nemes nyaras, D: nemes nyaras, E: ártéri füzes, F: fehér nyaras

Figure 6. Vegetation types based on drone photos. A: open surface, B: Amorpha thicket, C: young poplar plantation, D: poplar plantation, E: riparian willow forest and F: riparian poplar forest with *Populus alba*

1. táblázat A validációs keveredési mátrix

Table 1. Confusion matrix

Terepbejárás alapján	Döntési fa alapján						
	/	nyílt felszín/ gyep	ártéri füzes	gyalogakác, bozót	fehér nyaras	fiatal nemes nyaras	nemes nyaras
nyílt felszín/gyep		0,75	0,00	0,08	0,00	0,17	0,00
ártéri füzes		0,00	0,84	0,05	0,05	0,00	0,05
gyalogakác, bozót		0,00	0,08	0,92	0,00	0,00	0,00
fehér nyaras		0,00	0,00	0,00	0,83	0,08	0,08
fiatal nemes nyaras		0,00	0,00	0,17	0,00	0,83	0,00
nemes nyaras		0,00	0,00	0,00	0,18	0,00	0,82

Összegzés

Megállapítható, hogy az alkalmazott gépi tanulós osztályozási módszer alkalmas arra, hogy légi LiDAR alapú felvétel alapján növényzeti típusokat határozzunk meg. Nem csupán felszínborítási típusok (pl. erdő), de azokon belüli különféle vegetáció- és erdőtípusok is azonosíthatók ezzel a megfelelő pontosságú módszerrel. Ezt bizonyítja, hogy a vizsgált hullámtéri területen az osztályozás pontossága 83% (72 pont alapján).

A légi LiDAR felmérések feldolgozásának kombinálása gépi tanulós módszerekkel lehetővé teszi nagy kiterjedésű, ugyanakkor nehezen bejárható területeken a növényzet gyors és hatékony felmérését, illetve osztályozását eleget téve a XXI. század elvárásainak. Ismételt felmérések esetén az algoritmus könnyen betanítható az új adathalmazra, így a változások gyorsan és automatizáltan nyomon követhetőek lesznek, amelyek a döntéshozók és felhasználók számára is nagy előnyt jelenthetnek.

Az eredményül kapott növényzeti kategória térkép felhasználása széleskörű lehet. Például a vízügyes gyakorlatban a naprakész növényzeti térképek az árvízi levezető sávok tervezésekor és karbantartásakor használhatók, illetve az árvizek modellezésekor a friss LiDAR felvételek gyorsan feldolgozhatók gépi tanulós módszerrel, és így a növényzeti érdesség és a Manning

koefficiens napra készen meghatározható. Az erdészetben is jól használható adatok szülehetnek így, és az erdőrészletek statisztikai paraméterei gyorsan és pontosan meghatározhatók, így az erdészeti munkák pontosan ütemezhetők.

Köszönetnyilvánítás

A kutatásban felhasznált LiDAR adatokat az ATIVIZIG bocsájtotta rendelkezésünkre. A kutatást az OTKA 119193 pályázata támogatta.

Irodalom

- Abernethy, B., Rutherford, I.D. 1998: Where along a river's length will vegetation most effectively stabilise stream banks? *Geomorphology* 23: 55–75.
- Bengio, Y., Grandvalet, Y. 2004: No unbiased estimator of the variance of K-Fold cross-validation. *Journal of Machine Learning Research* 5: 1089–1105.
- Brooks, G.R. 2005: Overbank deposition along the concave side of the Red River meanders, Manitoba, and its geomorphic significance. *Earth Surface Processes and Landforms* 30: 1617–1632.
- Corenblit, D., Tabacchi, E., Steiger, J., Gurnell, A.M. 2007: Reciprocal interactions and adjustments between fluvial landforms and vegetation dynamics in river corridors: A review of complementary approaches. *Earth-Science Reviews* 84: 56–86.
- Geerling, G.W., Kater, E., van den Brink, C., Baptist, M.J., Regas, A.M.J., Smits, A.J.M. 2008: Nature rehabilitation by floodplain excavation: The hydraulic effect of 16 years of sedimentation and vegetation succession along the Waal River, NL. *Geomorphology* 99: 317–328.
- Grabmeier, J. L., Lambe, L. A. 2007: Decision trees for binary classification variables grow equally with the Gini impurity measure and Pearson's chi-square test. *International Journal of Business Intelligence and Data Mining* 2(2): 213.
- Heurich, M., Thoma, F. 2008: Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry*, 81: 645–661.
- Hudak, A., Crookston, N., Evans, J., Hall, D., Falkowski, M. 2008: Nearest neighbour imputation of species-level, plot-scale forest structure attributes from lidar data. *Remote Sensing of Environment* 112: 2232–2245.
- Jalonen, J., Järvelä, J., Virtanen, J.P., Vaaja, M., Kurkela, M., Hyyppä, H. 2015: Determining characteristic vegetation areas by terrestrial laser scanning for floodplain flow modelling. *Water* 7(2): 420–437.
- Jung, S.E., Kwak, S.A., Park, T., Lee, W.K., Yoo, S. 2011: Estimating crown variables of individual trees using airborne and terrestrial laser scanners. *Remote Sensing* 3: 2346–2363.
- Kiss, T., Fiala, K., Sipos, Gy., Szatmári, G. 2019a: Long-term hydrological changes after various river regulation measures: are we responsible for flow extremes? *Hydrology Research* 50(2): 417–430.
- Kiss, T., Nagy, J., Fehérvári, I., Vaszkó, Cs. 2019b: (Mis)management of floodplain vegetation: The effect of invasive species on vegetation roughness and flood levels. *Science of the Total Environment* 686: 931–945.
- Kiss, T., Sándor, A. 2009: Land-use changes and their effect on floodplain aggradation along the Middle-Tisza River, Hungary. *AGD Landscape and Environment* 3(1): 1–10.
- Kovács S., Váriné Szöllősi I. 2003: A Vásárhelyi Terv Továbbfejlesztését megalapozó hidrológiai és hullámtér hidraulikai vizsgálatok eredményei a Közép-Tiszán. *MHT XXI. 2/12*: 1–11.
- Laes, D., Reutebuch, S., McGaughey, B., Maus, P., Mellin, T., Wilcox, C., Anhold, J., Finco, M., Brewer, K. 2008: Practical lidar acquisition considerations for forestry applications. RSAC-0111-BRIEF1. U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center, Salt Lake City, UT. 32 p.
- Madsen, B., Treier, A. U., Zilinszky, A., Lucieer, A., Normand, S. 2020: Detecting shrub encroachment in seminatural grasslands using UAS LiDAR. *Ecology and Evolution* 10(11): 4876–4902.
- Manners, R., Schmidt, J., Wheaton, M.J. 2013: Multiscalar model for the determination of spatially explicit riparian vegetation roughness. *Journal of Geophysical Research: Earth Surface* 118: 65–83.
- McGaughey, R. 2018: Users Manual of Fusion/LDV: Software for LIDAR Data Analysis and Visualization. United States Department of Agriculture, Forest Service, Pacific Northwest Research Station
- Michez, A., Piégay, H., Lisein, J., Claessens, H., Lejeune, P. 2016: Classification of riparian forest species and health conditioning multi-temporal and hyperspatial imagery from unmanned aerial system. *Environmental Monitoring and Assessment* 188: 146.
- Naesset, E., Gobakken, T., Holmgren, J., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., Doderman, U. 2004: Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research* 19: 482–499.
- Nagy J., Kiss T., Fiala K. 2018: Hullámtér-feltöltődés vizsgálata az Alsó-Tisza mentén. II. Folyóhátak (parti hátak) feltöltődését befolyásoló tényezők. *Hidrológiai Közöny* 98(1): 33–40.

- 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, É. 2011: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12(85): 2825–2830.
- Rátky I., Farkas P. 2003: A növényzet hatása a hullámtér vízszállító képességére. *Vízügyi Közlemények* 85(2): 246–264.
- Saarinen, N., Vastaranta, M., Vaaja, M., Lotsari, E., Jaakkola, A., Kukko, A., Kaartinen, H., Holopainen, M., Hyppä, H., Alho, P. 2013: Area-Based Approach for Mapping and Monitoring Riverine Vegetation Using Mobile Laser Scanning. *Remote Sensing* 5: 5285–5303.
- Schaffer, C. 1993: Overfitting Avoidance as Bias. *Machine Learning* 10: 153–178.
- Steiger, J, Gurnell, A.M., Ergenzinger, P., Snelder, D.D. 2001: Sedimentation in the riparian zone of an incising river. *Earth Surface Processes and Landforms* 26: 91–108.
- Vetter, M., Höfle, B., Hollaus, M., Gschöpf, C., Mandlbürger, G., Pfeifer, N. 2011: Vertical vegetation structure analysis and hydraulic roughness determination using dense ALS point cloud data—a voxel based approach. *International archives of the photogrammetry, remote sensing and spatial information sciences* 38(5): 200–206.
- Zellei L., Sziebert J. 2003: Árvízi áramlásmérések tapasztalatai a Tiszán. In: Szlávik L. (szerk.) *Elemző és módszertani tanulmányok az 1998-2001. évi ár- és belvizekről*. *Vízügyi Közlemények* 4: 133–144.
- Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Korzeniowska, K., Wagner, G. 2014: Automated classification of airborne laser scanning point clouds, solving computationally expensive engineering problems. In: Koziel, S., Leifsson, L., Yang X. (eds.) *Springer Proceedings in Mathematics & Statistics* 97: 269–292.

http1: <https://erdoterkep.nebih.gov.hu>

AUTOMATISED IDENTIFICATION OF VEGETATION TYPES ON A FLOODPLAIN AREA BASED ON AIRBORNE LIDAR SURVEY

I. FEHÉRVÁRY¹, T. KISS²

¹ Lower Tisza Hydrological Directorate

6720 Szeged, Stefánia 4., email: fehevaryi@ativizig.hu

² University of Szeged, Department of Physical Geography and Geoinformatics

6722 Szeged, Egyetem str. 2-6, email: kisstim@gmail.com

Keywords: Decision Tree, scikit-learn, riparian vegetation, Python, point-cloud, cross-validation

In the last decades several environmental and anthropogenic effects altered the channels and floodplains of rivers, some of them resulted in increased riparian vegetation density and increased flood levels. Therefore, to provide sufficient flood safety proper floodplain management is needed, which must be based on up-to-date spatial data. The present paper aims to provide a method for the classification of riparian vegetation applying spatially continuous airborne LiDAR data and machine learning. The 3 km² large floodplain area was divided into 15x15 m pixels. The statistical parameters of the LiDAR point-cloud representing the vegetation of these pixels were calculated, and resulted data were classified applying a decision tree. Based on the data the following vegetation types were identified: open surface/grassland, *Amorpha* thicket, young and older poplar plantations, riparian willow forest, and riparian poplar forest dominated by white poplar. The accuracy of the decision tree was tested by 10-fold cross-validation, and also by field-survey checking the results of the decision tree on 72 points. Based on the resulted confusion matrix the overall accuracy of the classification was 83%.