

**Česká zemědělská univerzita v Praze**

**Fakulta lesnická a dřevařská**

**Katedra hospodářské úpravy lesů**



**Metody klasifikace dřevin v laserových bodových  
mračnecích**

**Diplomová práce**

**Autor: Bc. Martin Koudelka**  
**Vedoucí Práce: Ing. Karel Kuželka Ph.D.**

© 2022 ČZU v Praze

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Martin Koudelka

Lesní inženýrství

Lesní inženýrství

Název práce

**Metody klasifikace dřevin v laserových bodových mračnách**

Název anglicky

**Methods of species classification in LiDAR point clouds**

---

### Cíle práce

Ověřit možnosti automatické klasifikace dřevin nebo skupin dřevin na základě strukturálních charakteristik a spektrálních vlastností v trojrozměrném bodovém mračnu lesního porostu získaného z laserového skenování.

Vyhodnotit úspěšnost klasifikace.

### Metodika

V součinnosti s katedrou hospodářské úpravy lesů budou získána trojrozměrná bodová mračna lesních porostů.

Z bodových mračen budou separovány jednotlivé stromy pro vytvoření trénovacích setů (LS 2021).

Na trénovacích datech budou odvozeny charakteristické strukturální a spektrální vlastnosti skupin dřevin; ty budou následně využity v klasifikátorech (ZS 2022).

Bude vyhodnocena úspěšnost klasifikace a zhodnocena použitelnost v praxi.

Harmonogram:

LS 2021: Příprava dat, identifikace charakteristik

ZS 2022: kvantifikace charakteristik, vytvoření klasifikačních modelů

LS 2022: statistická analýza, vyhodnocení

## Doporučený rozsah práce

50 stran

## Klíčová slova

klasifikace, laserové skenování, bodová mračna, strojové učení

---

## Doporučené zdroje informací

- Gollob, C., Ritter, T., & Nothdurft, A. (2020). Forest inventory with long range and high-speed Personal Laser Scanning (PLS) and Simultaneous Localization and Mapping (SLAM) technology.
- Hunčaga, M.; Chudá, J.; Tomašík, J.; Slámová, M.; Koreň, M.; Chudý, F. The Comparison of Stem Curve Accuracy Determined from Point Clouds Acquired by Different Terrestrial Remote Sensing Methods. *Remote Sens.* 2020, 12, 2739.
- Kukko, A., Kaijaluoto, R., Kaartinen, H., Lehtola, V. V., Jaakkola, A., & Hyyppä, J. (2017). Graph SLAM correction for single scanner MLS forest data under boreal forest canopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 199–209.
- Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2005). UTILIZING AIRBORNE LASER INTENSITY FOR TREE SPECIES CLASSIFICATION Objectives Characterize and analyze ALS-derived intensity. *Natural Resource Management*, 1, 1–8.
- Pierzchała, M., Giguère, P., & Astrup, R. (2018). Mapping forests using an unmanned ground vehicle with 3D LiDAR and graph-SLAM. *Computers and Electronics in Agriculture*, 145(December 2017), 217–225.
- SUROVÝ, P. – KUŽELKA, K. – ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE. LESNICKÁ A DŘEVAŘSKÁ FAKULTA. *Aplikace dálkového průzkumu Země v lesnictví*. Praha: Česká zemědělská univerzita, 2019. ISBN 978-80-213-3008-5.
- Terryn, L.; Calders, K.; Disney, M.; Origo, N.; Malhi, Y.; Newnham, G.; Raunonen, P.; Åkerblom, M.; Verbeeck, H. Tree species classification using structural features derived from terrestrial laser scanning. *ISPRS J. Photogramm. Remote Sens.* 2020, 168, 170–181.
- 

## Předběžný termín obhajoby

2021/22 LS – FLD

## Vedoucí práce

Ing. Karel Kuželka, Ph.D.

## Garantující pracoviště

Katedra hospodářské úpravy lesů

Elektronicky schváleno dne 16. 3. 2022

**doc. Ing. Peter Surový, PhD.**

Vedoucí katedry

Elektronicky schváleno dne 29. 3. 2022

**prof. Ing. Róbert Marušák, PhD.**

Děkan

V Praze dne 30. 03. 2022

### **Čestné prohlášení**

Prohlašuji, že svou diplomovou práci "Metody klasifikace dřevin v laserových bodových mračnecích" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 10.4.2022

---

**Martin Koudelka**

## **Poděkování**

Rád bych touto cestou poděkoval vedoucímu práce Ing. Karlu Kuželkovi Ph.D. za vstřícnost a příkladné vedení diplomové práce, kterou bych bez jeho mentoringu jen těžko dovedl k cíli. Dále můj vděk směřuje rodině a blízkým přátelům za nedocenitelnou podporu a motivaci.

# Metody klasifikace dřevin v laserových bodových mračnec

## Abstrakt

Cílem této práce ověřit možnosti automatizované klasifikace 2 druhů dřevin pomocí výpočetní techniky na základě strukturálních či spektrálních charakteristik vstupních dat v podobě bodových mračen vzniklých ručním pozemním laserovým skenováním. Pro řešení projektu byly zvoleny dva přístupy: matematicko-statistický popis intenzit odrazů laserových paprsků a klasifikace na základě strukturálních charakteristik analyzovaných bodových mračen, které byly následně matematicky popsány pomocí Houghovy liniové transformace. Vlastní proces klasifikace proběhl skrze model strojového učení Random forest, který vycházel z číselných proměnných poskytnutých jednotlivými postupy. První metoda vycházela z celkem 11 proměnných a umožnila přesnost klasifikace 98,5 až 100 %, zatímco sekundární přístup postavený na 24 vytipovaných proměnných dosahoval přesností okolo 85 %. Každý z uvažovaných postupů představoval určité výhody i možné nedostatky, nicméně perspektivní výsledky pokládají solidní základ budoucímu zkoumání této problematiky rozšířením experimentu, kombinací metod či rozšířením klasifikačního modelu.

**Klíčová slova:** klasifikace, laserové skenování, strojové učení, bodová mračna

# **Methods of tree species classification in laser point clouds**

## **Abstract**

The goal of this thesis is to evaluate the possibilities of automated classification of 2 tree species using computing techniques based on structural and spectral characteristics of input point clouds created by handheld laser scanning. As the two solutions for project handling were selected: mathematical and statistical description of laser return signals intensities and classification based on structural characteristics of analysed point clouds, that were additionally mathematically described by Hough line transformation. Classification process was realized by Random Forest model of machine learning that was based on numerical variables supplied by individual methods. Classification based on 11 variables originating from the first method performed on 98,5 – 100 % overall accuracy, meanwhile the second set of 24 variables allowed classification accuracy about 85 %. Both suggested approaches represented certain advantages as well as possible drawbacks, although perspective outcomes are laying solid base for future research of this problematic with larger experiment, through combination of methods of with upscaling the classification model.

**Keywords:** classification, laser scanning, machine learning, point clouds

# Obsah

<b>1 Úvod</b> .....	<b>12</b>
<b>2 Cíl práce</b> .....	<b>13</b>
<b>3 Rozbor problematiky</b> .....	<b>14</b>
3.1 Digitální modelování v lesnictví .....	14
3.2 Nástroje a technika DPZ: přehled .....	15
3.3 Fotogrammetrie v lesní inventarizaci.....	17
3.3.1 Letecká UAV fotogrammetrie .....	17
3.3.2 Pozemní fotogrammetrie.....	17
3.3.3 Srovnání fotogrammetrie a LiDAR .....	18
3.3.4 Pozice fotogrammetrie v rámci inventarizace .....	18
3.4 Laserové skenování porostů pomocí LiDAR senzorů .....	19
3.4.1 Letecké UAV skenování systémem LiDAR.....	20
3.4.2 Pozemní skenování systémem LiDAR .....	21
3.5 Klasifikační modely využívající principů strojového učení.....	24
3.5.1 Lineární klasifikátory a Logistická regrese .....	25
3.5.2 K-means a KNN: příbuzní jménem .....	25
3.5.3 Random forest.....	27
3.5.4 Support vector machine (SVM).....	28
3.5.5 Naive Bayes klasifikátor .....	29
3.5.6 Klasifikace pomocí hlukového učení a neuronových sítí .....	29
3.6 Metody klasifikace dřevin na základě LiDAR skenů .....	31
3.6.1 Metody výpočetní a statistické klasifikace .....	32
3.6.2 Metody tvarové klasifikace a metody hlubokého učení .....	33
<b>4 Metodika</b> .....	<b>35</b>
4.1 Studovaná oblast a výchozí data .....	35
4.1.1 Technika získávání dat.....	35
4.1.2 Datová charakteristika .....	36
4.2 Návrh klasifikačních modelů .....	37
4.2.1 Matematicko-statistický popis bodových mračen.....	38
4.2.2 Klasifikace na základě strukturálních charakteristik .....	39
4.2.3 Random forest klasifikátor.....	41
<b>5 Výsledky</b> .....	<b>43</b>
5.1 Matematicko-statistická klasifikace .....	43
5.1.1 Parametry modelu Random Forest .....	44
5.2 Klasifikace na základě strukturálních charakteristik.....	46
5.2.1 Voxelizace a příprava grafických dat .....	46



5.2.2	Hough transformace a výstupní hodnoty .....	48
5.2.3	Výsledky klasifikace .....	49
5.2.4	Parametry modelu Random forest .....	50
<b>6</b>	<b>Diskuse .....</b>	<b>53</b>
6.1	Hodnocení modelu matematicko-statistické klasifikace .....	53
6.2	Hodnocení modelu obrazové klasifikace .....	54
6.3	Předpoklady pro integraci do praxe .....	56
<b>7</b>	<b>Závěr.....</b>	<b>57</b>
<b>8</b>	<b>Seznam použitých zdrojů .....</b>	<b>58</b>

## **Seznam obrázků**

Obr. 1. Graf funkce KNN klasifikátoru

Obr. 2. Schéma modelu Random forest.

Obr. 3. Schéma modelu SVM.

Obr. 4. Schéma ANN sítě.

Obr. 5. Schéma CNN sítě.

Obr. 6. Trajektorie chůze porostem č.3 při skenování.

Obr. 7. Graf závislosti počtu klasifikačních stromů na ose X a chybovosti na ose Y pro jednotlivé třídy a jejich průměr.

Obr. 8. Graf hodnot Giniho koeficientu značícího podíl parametrů na klas. model.

Obr. 9. Výsledek automatického ořezu vstupních dat na datové pruhy.

Obr. 10. Demonstrace rozdílu efektu velikosti voxelu 0,25 m nahoře a 0,1 m dole.

Obr. 11. Snímek obrazovky prostředí RStudio: zobrazení ilustračního výseku výsledné voxelové matice do výšky 1,3 m (sloupec V13).

Obr. 12. Výstupní snímek aplikace Houghovy transformace na smrku.

Obr. 13. Výstupní snímek aplikace Houghovy transformace na buku.

Obr. 14. Graf hodnot Giniho koeficientu značícího podíl parametrů na klas. model.

Obr. 15. Graf závislosti počtu klasifikačních stromů na ose X a chybovosti na ose Y pro jednotlivé třídy a jejich průměr.

Obr. 16. Nezahrnutí celého kmenu do datového výřezu, které se projevilo na výstupu.

## **Seznam tabulek**

Tab. 1. Charakteristika dřevinné skladby jednotlivých ploch s vyznačením počtu zahrnutých a vyřazených jedinců.

Tab. 2. Popis veličin vybraných pro výpočetní klasifikační modely na základě Lin a Hyypä (2016).

Tab. 3. Seznam proměnných použitých v klasifikačním modelu obrazové klasifikace.

Tab. 4. Matice záměn – hodnocení klasifikátoru na klasifikačním vzorku.

Tab. 5. Matice záměn – hodnocení klasifikátoru na zkušebním vzorku.

Tab. 6. Matice záměn – hodnocení klasifikátoru na klasifikačním vzorku.

Tab. 7. Matice záměn – hodnocení klasifikátoru na zkušebním vzorku.

## Seznam použitých zkratek

UAV – Unmanned aerial vehicle (bezpilotní letecký prostředek)

DPZ – Dálkový průzkum Země

SLAM – Simultaneous localization and mapping (současná lokalizace a mapování)

GNSS – Global Navigation Satellite System (Globální satelitní navigační systém)

LiDAR – Light detection and ranging – (detekce světla a škálování)

RMSE – Root mean square error – (střední kvadratická chyba)

PLS – Personal laser scanning (LiDAR skenování člověkem neseným zařízením)

MLS – Mobile laser scanning (mobilní LiDAR skenování)

TLS – Terrestrial laser scanning (pozemní LiDAR skenování)

HMLS – Handheld Mobile Laser Scanning (ruční LiDAR skenování)

CNN – Convolutional Neural Network – konvoluční neuronová síť

SMV – Support Vector Machine (metoda klasifikace dat strojovým učením)

ASCII – American Standard Code for Information Interchange – (Americký standard kódu pro datovou výměnu)

RF – Random forest (metoda klasifikace dat strojovým učením)

PNG – Portable Network Graphics (formát grafických souborů)

MB – Mega byte

BMLR – Bayesian multinomial logistic regression

# 1 Úvod

S rozvojem moderních technologií a postupů dochází v posledních letech k řadě úspěchů na poli inovací zavedených procesů v duchu automatizace, robotizace a zpřesňování získaných terénních dat za současného snižování nákladů i zvyšování bezpečnosti práce. Obecný trend lze sledovat napříč odvětvími a nevyhýbá se ani lesnictví, bez výjimky hospodářské úpravy lesů a lesní inventarizace.

Četné snahy o využití letecké techniky při mapování lesních porostů lze sledovat při aplikaci prostředků Dálkového průzkumu Země (DPZ). Typicky jde o využití letounů, letadel či družic, které skrze osazené senzory vzdáleně mapují růstové veličiny, zdravotní stav i jiné parametry (Dainelli et al. 2021). Výzkumy této problematiky nicméně identifikují omezenou přesnost danou systémem určování polohy (Chen et al. 2019) i charakteristikou procesu snímání dat omezující schopnost snímání podrostu (Warfield a Leon 2019), zejména při aplikacích fotogrammetrie (Li et al. 2020). Dále také naráží na problém současného legislativního rámce leteckého provozu při užití bezpilotních prostředků, který vyžaduje neustálý oční kontakt s pilotovaným strojem, což je v lesním provozu problematické.

Alternativou k leteckým prostředkům jsou pozemní metody DPZ. Technologie fotogrammetrie je výhodná zejména díky nízké vstupní ceně přístrojů, nicméně s rozvojem laserových senzorů má značnou konkurenci zejména co se týče přesnosti (Hunčaga et al. 2020). Robotikou a automatizací hnaný průmysl dokázal vyprodukovat utilizovaná řešení či modulární základ produktů, která bylo možné nasadit v rámci DPZ s požadovanou přesností, mobilitou a relativně dosažitelnou pořizovací cenou (Raj et al. 2020).

Jedním ze slibných zařízení je GeoSLAM ZEB Horizon, jde o ruční laserový skener s vysokou rychlostí záznamu až 300 000 bodů za sekundu, vysokou výslednou přesností, který je navíc vybaven algoritmem souběžné lokalizace a mapování (SLAM). V praxi nepotřebuje pro skenování informace ze systému určování polohy GNSS (Global Navigation Satellite System), což je jinak známý nedostatek při aplikaci DPZ v lesnictví (Gollob et al. 2020). Oproti statickým skenerům nevyžaduje ani referenční odrazová tělesa, jejichž instalace je zdlouhavá a značně prodlužuje celkovou dobu měření (Hunčaga et al. 2020). I přestože je zmíněný přístroj novinkou roku 2020, citované výzkumy již ukazují úspěšnou aplikaci při odvozování stromových veličin s vysokou přesností.

## 2 Cíl práce

Cílem této práce je ověřit možnosti automatické klasifikace druhů dřevin z laserových bodových mračen vzniklých ručním pozemním skenováním. Předpoklad vysoké přesnosti a rozlišení získaného bodového mračna slibuje takovou úroveň zachycení skutečnosti, ze které bude teoreticky množné určení jehličnatých a listnatých druhů stromů automatizovaným způsobem v podobě počítačového algoritmu s možným budoucím rozšířením na identifikaci konkrétního druhu dřeviny. Rozlišovací klasifikační atributy spočívají ve spektrálních a strukturálních charakteristikách stromů zachycených skenováním. Pro klasifikaci bude využito modelů strojového učení postavených na vstupních datech a následně bude vyhodnocena její celková přesnost. V neposlední řadě je cílem práce také literární rešerše tématu s průmětem do techniky vlastní práce. Literární opora je z většiny založena na publikovaných výsledcích výzkumu vzhledem k rychlému vývoji technologií a absencí relevantní tištěné literatury, která se stává méně aktuální v pouhých několika letech.

## 3 Rozbor problematiky

### 3.1 Digitální modelování v lesnictví

Tvorba digitálního modelu se stala nástrojem celé řady inovativních směrů v lesnictví s využitím nových technologií pořizování obrazu skutečnosti a jeho následného vyhodnocování pro účely hospodaření, výzkumu či vzdělávání (Marsel et al. 2021). Digitální model v různých variacích, slouží pro sledování veličin v krajinném měřítku i v rozměru jednotlivých stromů pro efektivní zhodnocení zdravotního stavu, růstových veličin nebo například monitoring fenologie dřevin (Dainelli et al. 2021). Při zpracování rešerše jsem se nejčastěji setkal s modely v podobě sady numerických dat, případně modelu terénu, nebo jednotlivých bodových mračen, dle potřeb řešené problematiky.

Digitální 3-Dimenzionální modely terénu i jednotlivých prvků jsou zajímavé zejména proto, že umožňují odvození jinak fyzicky měřených veličin přímo z modelu s využitím počítačové technologie (Chen et al. 2019; Hunčaga et al. 2020; Marsel et al. 2021). Jinými slovy počítačový program na základě získaných dat „vidí“ zkoumané prostředí, a dokáže provádět automatizované analýzy a měření. Samostatné strojové vyhodnocení je ale až druhým krokem procesu zpracování vstupních dat, na jehož počátku stojí model s požadovanou přesností a úrovní detailu umožňující efektivní následnou analýzu (Corte et al. 2020). Použití určitých typů senzorů (například hyperspektrální) navíc poskytují přidanou hodnotu, dokážou totiž zaznamenat skutečnost mimo spektrum lidského vnímání, což umožní zcela nový pohled na zkoumanou problematiku (Näsi et al. 2015). Avšak, co se týče počtu studií v roce 2021, patří k méně populárním směrům vývoje s nicméně vysokým potenciálem (Dainelli et al. 2021).

V praxi používané přístupy terénního sběru dat jsou často spojeny nejen s chybami při záznamu či samotném měření (Larjavaara a Muller-Landau 2013), ale také s vysokou finanční náročností terénních prací (Fankhauser et al. 2018). Digitalizace tohoto procesu by mohla odstranit možné chyby zaviněné lidským faktorem, ale také proces nejen zefektivnit, zejména při nasazení leteckých prostředků (Fankhauser et al. 2018). I přes nespornou výhodu rychlého a bezpečného pohybu létacích prostředků, pozemní digitalizační metody nabízí možnou vyšší úroveň přesnosti v podrostu i další výhody (Warfield a Leon 2019) a mohou tak být možnou konkurencí současným praktickým metodám lesní inventarizace.

### 3.2 Nástroje a technika DPZ: přehled

Jak jsem již zmínil v textu dříve, Dálkový průzkum Země umožňuje letecké snímání i záznam dat z pozemních senzorů. V současné praxi lze předpokládat, že možnost letecké aplikace vzhledem k hmotnosti senzorů a nosnosti kompaktní letecké techniky je ve řadě případů indiferentní/záměnné s možnostmi pozemní aplikace (Dainelli et al. 2021). Co se týče hlavních rozdílů, letecké snímání poskytuje velké plošné pokrytí a lépe mapuje korunový prostor, oproti tomu pozemní snímání lépe zachytí spodní porostní patra a etáže včetně přesnějších strukturálních charakteristik (Mikita et al. 2016). Potenciální zvýšení efektivity při rozsáhlých operacích může spočívat ve spojení obou technik, tedy v kalibraci leteckých dat pozemními, čímž dojde k zásadnímu zpřesnění modelu (Levick et al. 2021).

Tomuto přímému porovnání se vymykají družice s odlišnou výbavou, násobně nižším rozlišením a diametrální vyšším akčním rádiusem (Surový a Kuželka 2019). Typicky je taková technika užitečná při sledování změn pokryvu Země v krajinném měřítku, v lesnictví jsou takovéto analýzy již s praxí využívány, například pro monitoring zdravotního stavu porostů prostřednictvím mapy (Ústav pro hospodářskou úpravu lesů a Výzkumný ústav lesního hospodářství a myslivosti 2018).

Vhodnou poznámkou v tomto bodě je zmínka o dvou základních typech dat, které lze skrze senzory pořizovat a dále použít pro vyhodnocení cílových veličin. Jak lze pozorovat na práci Fauvel (2008), strukturální či prostorová data popisují rozměry a dimenze snímaných objektů, neboli tvoří následný model struktury. Oproti tomu data spektrální slouží k popisu určité vlastnosti cílových prvků, která vyplývá z odrazové charakteristiky určitého spektra světla (Fauvel 2008). Řada materiálů může být v tomto směru poměrně specifických, čehož je možné využít při snímání různých povrchů (Fauvel 2008), případně také při identifikaci růstových procesů rostlin (White 2012). Při určitých aplikacích je vhodná také kombinace a vzájemné doplnění obou přístupů pro zpřístupnění jinak obtížně zjistitelných charakteristik (Kestur et al. 2018; Mikita et al. 2016).

Z nejčastěji využívaných senzorů pro účely DPZ v lesnictví je vhodné vylišit fotografickou techniku včetně multi/hyper-spektrálních snímačů, laserové senzory fungující na technologii Light Detection and Ranging (LiDAR) a termální kamery, jak zmiňují Surový a Kuželka (2019). Obecný princip snímání ostatních jmenovaných snímačů je oproti laserové technologii zásadně odlišný. Senzor zachycuje obraz na základě pasivního záření z okolí,

LiDAR oproti tomu signály aktivně vysílá. Jak vysvětluje Fankhauser et al.(2018), praktický rozdíl spočívá ve schopnosti prostupnosti signálu porostem i větší míře nezávislosti na okolních podmínkách ve prospěch laseru. Každá kategorie představuje výhody i obtíže, a to nejen z hlediska přesnosti a povahy dat, ale také při zohlednění pořizovací ceny či možností praktického nasazení skrze dostupné komerční produkty. Obecně lze konstatovat, že fotogrammetrie je stále bezkonkurečně nejlevnější a nejuniverzálnější technologií, LiDAR se hodí díky vysoké přesnosti na měření struktury a rozměrů zájmových objektů, a nakonec snímače s dosahem nad rámec lidského vnímání vyniknou při sledování zdravotního stavu a fenologie (Dainelli et al. 2021).

V analýzách rozměrů stromů a dalších atributů z hospodářsko-úpravnického hlediska bude cílem digitalizace a počítačové analýzy především co nejdůvěrnější zachycení skutečnosti pro odvození stěžejních rozměrů stromů, tedy výšky, výčetní tloušťky či průměru koruny (Dainelli et al. 2021). Do tohoto odvětví spadá také tato práce, jelikož druh dřevin je možné určit právě na základě strukturálních a spektrálních charakteristik (Axelsson et al. 2018). Pro aplikace v rámci inventarizace se nabízí zejména využití laserových snímačů, či alternativně fotogrammetrických metod (Surový a Kuželka 2019). Jelikož je ale určení druhu dřevin pouze jeden z praxí požadovaných atributů, považuji za důležité nejprve porovnat perspektivu fotogrammetrických a LiDAR aplikací, co se týče možností praktického nasazení. Pro srovnání budou relevantní technologie detailněji rozebrány v následujících částech včetně kontextu.

Rád bych také úvodem této teoretické báze zmínil, že vystihnout při porovnávání všechny aspekty použití té či jiné metody není ambicí ani cílem této práce. Zmíněné hodnoty a dosažené výsledky jsou pouze referenčními rámci pro obecnou představu o schopnostech dané technologie. Důležitým faktorem je ale také poměr ceny a výkonu zvolené metody, a to především při aplikaci LiDAR senzorů, při kterých se, co se týče zařízení, lze pohybovat výší nákladů napříč řády. Přístupy se liší také v náročnosti zpracování dat, která může být u výkonných senzorů velmi bohatá a náročná na vyhodnocení, jak uvádí například Hunčaga et al. (2020). V rámci možností jsem se v relevantních případech snažil finanční stránku reflektovat s cílem zvýšení příkladnosti rozboru.



### 3.3 Fotogrammetrie v lesní inventarizaci

V rámci prostorových analýz pracuje fotogrammetrie na základě softwarů schopných skládat obraz pořízený z různých úhlů a v různém překryvu do 3D modelu (González-Jorge et al. 2012). Důležitým aspektem ovlivňujícím výslednou přesnost dat vytvořených pomocí této metody je nastavení dopředného i bočního překryvu snímků (Ni et al. 2018).

Jak uvádí Ramalho de Oliveira et al. (2020) nejvhodnější konfigurace snímací techniky se jeví digitální zrcadlový fotoaparát (DSLR) ve spojení s objektivem konstruovaným s jednak fixní ohniskovou vzdáleností, ale také se standardní šířkou záběru. Záznam obrazu je tak věrnější ve srovnání s širokoúhlým snímačem, který může při okrajích snímků zkreslovat. Z dat letecké aplikace fotogrammetrického měření DSLR kamerou se standardním neširokoúhlým snímačem dokáže software správně identifikovat až 92 % stromů, širokoúhlý snímač potom 83 % (Ramalho de Oliveira et al. 2020). Ještě znatelnější rozdíl publikuje Hunčaga et. al. (2020), označuje spolehlivost detekce stromu pomocí pozemní fotogrammetrie hodnotou 76 %, přičemž LiDAR technologie dosáhly v tomto ohledu spolehlivosti až 100 %.

#### 3.3.1 Letecká UAV fotogrammetrie

Při letecké aplikaci je fotogrammetrie velice závislá na povětrnostních vlivech a pozici slunce, zatímco laserové letecké snímání je v těchto ohledech méně náročnou metodou, další hlavní nevýhodou je slabá schopnost měřit výšky stromů a obtížná až nemožná je detekce překrytých stromů (Fankhauser et al. 2018). Nutno však podotknout, že při ideálních podmínkách (klimatických i strukturálních vlastnostech porostů) je fotogrammetrie s chybou měření v řádu vyšších desítek centimetrů při určování výšek kvalitativně srovnatelná s LiDAR senzory o nižším rozlišení (Ramalho de Oliveira et al. 2020). Pokud ale nebude otázkou cena, pak jsou nejmodernější laserové senzory schopny měřit s Root mean square error (RMSE) v řádech jednotek centimetrů, střední odchylkou dokonce v milimetrovém řádu a s extrémně vysokým rozlišením (Kuželka et al. 2020). Pro praxi je ale cena technologie stále velmi vysoká, při srovnání s fotogrammetrií se v průměru jedná o desetinásobek celkových nákladů na leteckou misi a vyhodnocení dat, které může být znatelně náročnější (Dainelli et al. 2021).

#### 3.3.2 Pozemní fotogrammetrie

Pokud je ale řeč o využití fotogrammetrie v rámci pozemních metod DPZ, je možné nalézt velice slibné aplikace. Například Mikita et al. (2016) publikuje výsledky pozemní

fotogrammetrie s využitím běžné fotografické DSLR techniky s chybou okolo 1 cm RMSE při stanovení výčetní prsní tloušťky. Díky své schopnosti přesného určení průměru byla pozemní data kombinována s leteckou fotogrammetrií, přičemž se obě metody navzájem doplňovaly. Nicméně dále také Mikita et al. (2016) podotýká, že podíl na takovém výsledku může mít i absence husté vegetace v rámci zkoumané plochy. Piermattei et al. (2019) uvádí podobnou přesnost určení výčetní tloušťky RMSE přibližně 1 cm. Navíc v rámci výzkumu bylo zjištěno, že až do výšky 3 metrů bylo možné určit růstovou křivku stromu s RMSE 2,45 cm.

### **3.3.3 Srovnání fotogrammetrie a LiDAR**

Studie Gollob et al. (2020) udává při použití ruční LiDAR technologie (Personal laser scanning – PLS) při pozemním měření RMSE přibližně 2,5 – 4 cm při stanovení výčetní prsní tloušťky. To částečně v odlišném poměru výsledků potvrzuje Hunčaga et al. (2020), při měření prsní tloušťky srovnával se svým týmem přesnost TLS, PLS a fotogrammetrie, přičemž nejmenší RMSE ve vztahu k referenčním datům podává technologie statického LiDAR senzoru a právě fotogrammetrie. Dále ale také uvádí, že ve výšce 8 metrů je již, co se týče RMSE, nejlepší metodou PLS, průměrné RMSE jsou potom 1,03 cm – TLS, 1,26 cm – PLS, a 1,90 cm – fotogrammetrie.

Co se týče leteckých aplikací, většina recentně vydaných prací se nevěnuje přímému odvozování výšek stromů a jejich tloušťek, přímé porovnání s výsledky lidarových senzorů proto chybí. Nabízí se ale více publikací využívaných v rámci výzkumu kombinaci pozemních a leteckých přístupů s velice slibnými výsledky (Li et al. 2020; Mikita et al. 2016).

### **3.3.4 Pozice fotogrammetrie v rámci inventarizace**

Při hodnocení nasazení fotogrammetrie v praktické inventarizaci, a ještě specifičtěji v kontextu dendrometrických měření, lze na základě prezentovaných výzkumů nahlížet na fotogrammetrii jakožto na parciální či vícekrokovou metodu, která doplní nebo musí být doplněna dalším měřením, což ovšem může vést k cenově efektivnímu a přesnému řešení, jak lze vidět na konkrétních příkladech výzkumů Shimizu et al. (2022) či Mikita et al. (2016). Na kvalitních senzorech LiDAR postavené metody nemívají potíže s určením obou stěžejních rozměrových parametru v rámci inventarizace – výšek i tloušťek stromů – ve srovnatelné a vyšší přesnosti oproti soudobým praktickým inventarizačním technikám (Ko et al. 2021). Nicméně lze vidět z výše v odstavci citovaných pracích, pokud jsou vhodnou kombinací v rámci metodiky eliminovány nevýhody fotogrammetrie, výsledky takových aplikací jsou

úctyhodné, a to navzdory poměrně jednoduché a levé technice při srovnání s LiDAR senzory. Snad právě proto je stále předmětem řady studií, její potenciál může být nadále využit a dle poskytnutých příkladů je vhodné tuto metodu považovat za efektivní v rámci praktické inventarizace v momentě zpracování této rešerše. Otázkou však zůstává, jak bude vývoj snímací zejména laserové techniky pokračovat, a zda si fotogrammetrie udrží konkurenceschopnost, jejíž důležitým atributem je, jak uvádí Shimizu et al. (2022), mimo dalších především nižší finanční náročnost (Dainelli et al. 2021).

### **3.4 Laserové skenování porostů pomocí LiDAR senzorů**

Základem technologie LiDAR senzorů je disperze laserových paprsků směrem k zájmovému objektu. Jak popisují Jupp a Lovell (2007), senzor vysílá větší množství pulzů a následně zachycuje zpětné odrazy, díky čemuž je možné zaznamenat informaci jednak o vzdálenosti, ale také o povaze povrchu objektu. Celá řada materiálů navíc zpravidla neodrazí jednotně signál zpět, ale dovolí jeho průchod. Po prostoupení části paprsku se proces opakuje v nižších úrovních a senzor poté zachycuje odraz částečný, díky čemuž je teoreticky možné získat určitou informaci o stavu a povaze nižších pater vegetace (Jupp a Lovell 2007). Již se prokázalo, že LiDAR technologie poskytuje kvalitnější informace o spodních vegetačních patrech než fotogrammetrie, přičemž použitá technologie náleží stejným cenovým řádům (Wallace et al. 2016). Další značný benefit spočívá ve vyšší toleranci vůči zejména světelným podmínkám okolního prostředí (Fankhauser et al. 2018). Jokela et al. (2019) nicméně popisuje sníženou funkci senzoru za přítomnosti vodních částic v okolí senzoru, překážkou je tedy déšť, sníh, ale problém může způsobit také mlha, nebo teploty hlouběji pod bodem mrazu. Celkový výsledek měření může být značně zkreslený (Jokela et al. 2019).

Řadu specifík využití laserových senzorů při zjišťování prostorových atributů lesních porostů jsem již zmínil v předchozí kapitole pro bezprostřední a věcné porovnání s metodami založenými na fotogrammetrii. Shrnu tedy pro přehlednost nejdůležitější poznatky. LiDAR snímače jsou poměrně finančně nákladné (Dainelli et al. 2021), mohou být velmi přesné při určování rozměrových charakteristik (Slavík et al. 2020; Kuželka et al. 2020) a přesné z hlediska spolehlivosti detekce stromů, tedy eliminace duplicit i falešně pozitivních či nedetekovaných jedinců (Hunčaga et al. 2020; Ramalho de Oliveira et al. 2020).

Pro představu o schopnostech a úměrné cenové hladině použitého systému je vhodné sledovat následující parametry senzorů: počet odrazů, počet pulzů i průměrnou odchylku

stanovenou výrobcem. Rámcové srovnání nabízí Lin et al. (2021), nejde sice o lesnickou aplikaci, nicméně je zde patrné, že senzory s podporou více odrazů patří mezi nejdražší techniku. S úrovní detailu roste také celková hmotnost senzoru, což může být limitující faktor zejména při leteckém snímání pomocí UAV. U výsledných dat je poté příkladná hodnota průměrné hustoty bodového pokrytí, která závisí na konkrétní metodice, vysoká míra překryvu jednotlivých snímků a nižší letová rychlost umožní vysoce detailní výsledek (Slavík et al. 2020). Ze zmíněných letových atributů je možné pomocí vyššího dopředného překryvu dosáhnout vyšší přesnosti dat bez vlivu na výsledné operativní náklady (Leberl et al. 2010). V následujících podkapitolách přiblížím specifika leteckých i pozemních aplikací a na závěr nabídnu srovnání obou přístupů.

### **3.4.1 Letecké UAV skenování systémem LiDAR**

Možnost provádět měření lesních majetků přeletem nad zájmovým územím je časově méně náročnější oproti pozemním metodám, zvláště pokud se jedná o obtížně přístupný terén (Warfield a Leon 2019). Letecké mise ale mohou vyžadovat komplexní přípravu i náročné následné zpracování výstupů (Dainelli et al. 2021), což může mít dopad na výslednou efektivitu. Jak jsem již zmínil výše, LiDAR snímače mají lepší předpoklady pro skenování uvnitř porostu včetně průniku korunovým zápojem či dílčích etážích. Zda bude ale model úspěšný záleží na metodice měření, ale především také na použitém snímači.

Zásadní rozdíly lze pozorovat například na pracích autorů Wallace et al. (2016) a Kuželka et al. (2020). První zmíněná publikace aplikuje cenově efektivní konfiguraci řadového UAV osazeného levnějším LiDAR senzorem s podporou jednoho odrazu a celkovou hustotou výsledného mračna 174 bodů/m<sup>2</sup> v korunovém zápoji při 95% dopředném a 66% bočním překryvu. Zvolená sestava byla schopna zachytit s výšky stromů s poměrně nízkou chybou RMSE 92 cm, nicméně autoři se nezaměřili na odvození průměrů stromů, což je vzhledem k nízké schopnosti zachytit strukturu v nízkých stromových patrech touto metodou neproveditelné (Wallace et al. 2016). Oproti tomu druhý příklad popisuje nasazení špičkového senzoru na speciálním UAV kvůli požadované nosnosti, LiDAR podporuje až 7 odrazů a měření poskytl celkovou hustotu 200 bodů/m<sup>2</sup> každého jednotlivého přeletu, po sestavení výsledného modelu se tato hodnota vzhledem k překryvům snímků navýší, přesnou hodnotu však autoři neuvádí. V této konfiguraci se podařilo stanovit průměry stromů v prsní výšce s chybou RMSE v rozsahu 0,1 cm – 5 cm. (Kuželka et al. 2020) Vzhledem k těmto výsledkům lze předpokládat také spolehlivé odvození výšek stromů, i přestože se autoři se

touto problematikou nezabývali. Díky prokázaným schopnostem odvození průměru a zřejmě i výšek by taková metoda dokázala plně nahradit klasickou lesní inventarizaci a to v režimu plného měření každého stromu, zejména, jak uvádí Kuželka et al. (2020), na ploše do 100 ha souvislých porostů vzhledem k náročnosti plánování a celkové souvislé náročnosti leteckých misí.

Zmíněné extrémy ukazují dva póly současné problematiky nasazení LiDAR senzorů, respektive reflektují nedostatečnou efektivitu cenově přijatelných řešení při samostatném nasazení a potenciál vysoce přesných senzorů, kde je ale stále limitním faktorem cena použité techniky. Jak ale popisuje v rozboru tematiky Dainelli et al. (2021), technologie LiDAR snímačů se vyvíjí rychlým tempem, a je tedy snad otázkou času, zda dojde k postupné utilizaci těchto senzorů i snížení jejich ceny.

### **3.4.2 Pozemní skenování systémem LiDAR**

Jak vyplývá z předchozí kapitoly, zejména nákladové i technologické bariéry nutí výzkumníky hledat nová řešení a odlišné metody nasazení LiDAR technologie v rámci lesní inventarizace. Jedním z takových přístupů je pozemní skenování porostů (Dainelli et al. 2021). Vzhledem k měření pod korunovým zápojem nebrání těmto metodám detailní zachycení povahy spodních pater lesa (Gollob et al. 2020), což by mohlo umožnit spolehlivé odvození prsních tlouštěk. Mohlo by tak být možné nasazení levnějších senzorů, které, jak je popsáno v předchozí kapitole, dosahují vysoké přesnosti měření, ale nepodporují registraci většího počtu odrazů.

#### **Vylišení přístupů pozemního skenování**

Neexistuje normalizované názvosloví jednotlivých metod, nicméně z publikovaných článků autorů Gollob et al. (2020) Hunčaga et al. (2020) lze definovat následující. Základní dělením lze vymežit statické měření označované jako Terrestrial Laser Scanning (TLS), mobilní skenování senzorem umístěným na dopravním prostředku nazývané Mobile Laser Scanning (MLS) a Peronal Laser Scanning (PLS), kde je senzor nesen člověkem na batohu. V poslední době je navíc vylišován systém Handheld Mobile Laser Scanning (HMLS), při kterém je skener nesen v ruce. Specifika lze nalézt v použitých přístroji, respektive v postupu měření, jaký technika vyžaduje. TLS je typicky spojováno aplikací statického senzoru a sady referenčních bodů (sfér) v terénu. Oproti tomu PLS, MLS a HMLS spočívá v nasazení mobilního skeneru, který je v průběhu měření přesunován zájmovým územím (Gollob et al. 2020; Hunčaga et al. 2020). V dalších odstavcích této kapitoly představím bližší porovnání metod TLS a HMLS.

### **Problematika systému určování polohy**

Určité ruční mobilní skenery využívané v rámci HMLS nabízí zásadní výhodu v podobě absence nutnosti přesného určení polohy skrze totální stanici či přesný GNSS přijímač, pokud není zapotřebí akurátní georeference snímané plochy (Balenović et al. 2020). Jako příklad lze uvést recentně na trh uvedený ruční mobilní skener GeoSLAM ZEB Horizon, který disponuje technologií současného snímání a mapování SLAM, s jejíž pomocí je v rámci skenování autonomně zachycena prostorová orientace zájmové oblasti a vytvořen lokální souřadnicový systém (Bailey a Durrant-Whyte 2006). V lesnickém prostředí je podmínka přístupu k silnému GNSS signálu známým nedostatkem zejména kvůli časté nedostupnosti na požadované úrovni (Chen et al. 2019). Systém určování polohy navíc obvykle není integrován v TLS přístrojích, je proto vyžadováno použití další techniky, kterou je nutné a přenášet mezi lokacemi, což má dopad na výslednou efektivitu a je označeno za známou a esenciální překážku metody TLS (Thrun a Montemerlo 2006). Kvůli svým aplikačním charakteristikám je metoda HMLS v konfiguraci se SLAM technologií považována za velice perspektivní pro užití v lesní inventarizaci (Balenović et al. 2020).

### **Časová náročnost terénních prací**

Další podstatný benefit užití HMLS technologie spočívá v rychlosti terénních prací. Gollob et al. (2020) i Hunčaga et al. (2020) dospěli přibližně ke stejnému závěru, že pomocí HMLS snímače se SLAM technologií dokázali dokončit měření přibližně 5x rychleji ve srovnání s metodou TLS, a to zejména kvůli absenci GNSS zaměření a umístění referenčních objektů. Použití mobilního ručního skeneru může být zároveň datově až 10násobně datově náročné, co se týče čistého počtu bodů (Hunčaga et al. 2020). Při použití skeneru ZEB Horizon vzniká ale stále příliš dat pro vyhodnocení v reálném čase, což by jinak bylo pro operátory velice přínosné z hlediska kontroly správnosti skenů (Gollob et al. 2020). Nicméně nejspíš je jen otázkou času, kdy bude tato funkce dostupná pro tento přístroj, vzhledem ke skutečnosti, že starší modely ji podporují (Chen et al. 2019).

### **Přesnost určení tloušťek stromů**

Výchozí technologie LiDAR je sice principálně identická ve všech zmíněných metodách, z výstupů odvozené veličiny jsou určeny s různou přesností. Gollob et al. (2020) uvádí, že při použití HMLS skeneru ZEB Horizon byla míra správné detekce stromů 96 % s chybou stanovení výčetního průměru RMSE 2,32 cm , zatímco data metody TLS poskytla data pro

determinaci pouze 78,5 % jedinců s chybou průměru RMSE 2,52 cm na identických zkušných plochách (Gollob et al. 2020). S odlišnými závěry publikoval výsledky výzkumu Hunčaga et al. (2020), přesnost detekce obou přístupů stanovil až na 100 % a chyby měření výčetní tloušťky na RMSE 0,8 cm v případě TLS a RMSE 1,62 cm za použití HMLS. Dodává také, že přesnost stanovení tlouštěk záleží také na metodě datové analýzy a tvarového pasování referenčních objektů při stanovování průměrů, obě metody totiž generují odlišnou úroveň šumu a uspořádanosti bodů. Data pocházející z HMLS mývají více rozptýlených bodů, může tedy být problém s určením zejména malých průměrů stromů (Gollob et al. 2020; Hyypä et al. 2020). Alternativní studie zaměřující se na volbu optimální reprezentativního tělesa jiného, než je typický válec, který je běžně pro tento účel volen. Pomocí speciální funkce dokázal snížit chybu určení tloušťky až o RMSE 7,3 mm při aplikaci na dat pořízených za pomoci TLS skeneru (Pitkänen et al. 2019).

### **Přesnost určení výšek stromů**

Hyypä et al. (2020) v rámci výzkumu porovnával různé aplikace LiDAR odlišných senzorů při letecké i pozemní aplikaci. Publikovaná práce naznačuje, že HMLS senzory jsou efektivně schopny také tvořit data pro spolehlivé určení výšek stromů, konkrétně senzor ZEB Horizon se umístil na předních příčkách porovnání s chybou RMSE 0,4 m ve volnějším zápoji a 1,4 m v hustším porostu. Porosty s výškou nad 25 metrů byly systematicky více podhodnoceny oproti těm s výškou pod touto hranicí. Autoři dále zmiňují mimořádnou přesnost stanovení kmenového profilu metodami MLS a HMLS, což opět naznačuje velký potenciál těchto přístupů. Alternativní studie uvádí chybu určení výšek RMSE 1,11 m, přičemž zkoumané porosty se pohybovaly v rozmezí 9 – 33 m výšky (Jurjević et al. 2020). Sofia et al. (2021) uvádí chybu určení výšky RMSE v rozmezí 2,1 – 6 m, která je ale pro běžnou inventarizační praxi příliš velká. Nicméně je třeba zmínit, že autoři také určili chybu odvození prsní tloušťky RMSE 2 – 4,6 cm, což je znatelně více oproti zdrojům uvedeným výše v této kapitole. Pro srovnání, TLS skener dokáže určit výšky s chybou RMSE 1,23 m (Pitkänen et al. 2019).

### **Problematika systému určování polohy**

Jak dále uvádí Liang et al. (2018), při určování přesnosti dat z TLS skenování, hrají roli četné faktory, jako je například typ skeneru, podmínky skenování, druh dřevin, zakmenění, metodika skenování nebo velikost plochy (Liang et al. 2018), pokud tedy nejsou výsledky porovnávány za stejných podmínek, nemusí být komparace konkrétních hodnot vždy na místě. Podobně jako u leteckých aplikací volba překryvů, má na výslednou přesnost a efektivitu specificky metody HMLS zásadní vliv volba trasy procházení měřené plochy (Balenić et al.

2020). Vzhledem ke technickým limitům procesu snímání pomocí HMLS ZEB Horizon se také nedoporučuje delší snímání 30 minut (Gollob et al. 2020), aby nedošlo k nežádoucímu posunu bodů výsledného bodového mračka, starší model ZEB-REVO-RT je vhodné užívat bez přestávky pouze 20 minut (Chen et al. 2019).

Na základě rozboru problematiky lze tedy usoudit vhodnost a potenciál využití metody HMLS, a kvůli technologickým benefitům konkrétně skeneru ZEB Horizon, pro potřeby lesní inventarizace. Výsledné chyby měření jsou v porovnání s metodikou tradičních metod měření tloušťek průměrkou, jelikož při zařazení do tloušťkových tříd je záměrně připsána chyba 0-2 cm. Výzvou pro budoucí bádání spočívá v utilizaci metod, jak vyplývá z literárních rozboru, prostřednictvím tvorby ideální techniky pořízení i následného zpracování dat a stanovení dalších metodických charakteristik pro efektivní vyhodnocení strukturálních charakteristik lesních porostů.

### **3.5 Klasifikační modely využívající principů strojového učení**

V rámci teoretického základu je také účelné zmínit použité klasifikační metody s přiblížením jejich principu a charakteristik. Nutno také poznamenat, že v rámci řešené problematiky klasifikace dřevin se jedná pouze o dva často skloňované postupy – Random forest (RF) a Support vector machine (SVM), které ovšem nejsou limitními, co se týče možných a v praxi používaných klasifikačních nástrojů (Supervised learning algorithms). Alternativami jsou také metody K-průměrů, Logistická regrese, KNN – K nejbližších sousedů, či Naive Bayes (Gong 2022; Osisanwo et al. 2017). Volba bližšího zaměření na RF a SVM je opodstatněna nejčastějším zastoupením těchto klasifikátorů ve studované literatuře, jak je naznačeno dále v textu. Stranou těchto výpočetních klasifikačních metod stojí postupy využívající pro obrazovou klasifikaci hluboké učení prostřednictvím neuronové či konvoluční sítě (Lecun et al. 2015), které bývají v rámci klasifikace laserových dat často využívány (Surový a Kuželka 2019).

V praxi jsou na základě rozdílnosti úrovně vstupu analytika do učícího procesu rozlišovány na elementární rovině 2 přístupy – učení dozorované a nedozorované. Jak zmiňuje Batta (2020) nedozorované učení postupuje zcela samostatně a cílové třídy jsou vylišeny zcela automaticky na základě testovaných datových charakteristik. Oproti tomu dozorované učení staví na vstupu člověka, který stanoví cíle rozhodování, které jsou reprezentovány v rámci trénovacího setu dat (Kotsiantis 2007).



Při snaze o porozumění tématu je potřebné separovat poznatky samotných principů některých klasifikačních metod a metod samotných. Rozdíl spočívá v principu shlukování (clustering) dat, které klasifikátory využívají (Erman et al. 2006), nicméně definice metod shlukování s definicemi výsledné klasifikace nemusí být vždy shodné například v úrovni systematiky dělní jednotlivých principů a definicí úrovní úrovně dozoru (Osisanwo et al. 2017; Batta 2020), což může být v počátcích matoucí.

### **3.5.1 Lineární klasifikátory a Logistická regrese**

Cílem klasifikačních metod na základě lineárních charakteristik spočívá v tvorbě lineárních vztahů na základě vztahu jednotlivých vlastností či parametrů vstupujících do modelu ze strany klasifikovaných dat. Seskupením charakteristických dat za předpokladu správné volby proměnné a nastavení modelu je dosaženo segregace datových vstupů do více či méně izolovaných skupin (Verlinde a Mayoraz 1998), tedy je účelně provedena klasifikace.

Na podobném principu je postavena také metoda Logistické regrese s rozdílem v použitém statistickém nástroji postupu. Namísto lineárních vztahů je implementována právě logistická regrese, stanovena je potom hranice dvou klasifikačních tříd a jednotlivé prvky klasifikovaného datasetu obdrží hodnocení vzdálenosti od této hranice (Osisanwo et al. 2017), což reprezentuje pravděpodobnost příslušnosti jednotek jednotlivým třídám.

Lineární vztahy jsou dostatečně elementární pro velice rychlé a efektivní vyhodnocování i velkého objemu dat, přičemž míra závislosti klasifikačních parametrů přímo závisí na výsledném výkonu (Setiono a Leow 2000), hodí se proto pro více až převážně izolované veličiny. Rychlá a efektivní je také metoda Logistické regrese a lze využít její jednoduchosti a nízké komplexity jakožto pomůcky při vývoji a kalibraci příbuzných klasifikačních metod, například neuronových sítí, jelikož základní princip je založený na stejném statistickém základu (Dreiseitl a Ohno-Machado 2002). Jak lze odvodit z metodického přehledu lesnických aplikací strojového učení při klasifikaci dřevin, tyto metody nejsou v současné praxi primárně užívány (Michałowska a Rapiński 2021). Nesporně mají své výhody a opodstatnění, avšak v rámci zachování věrnosti tématu nebudou dále podrobněji rozebrány.

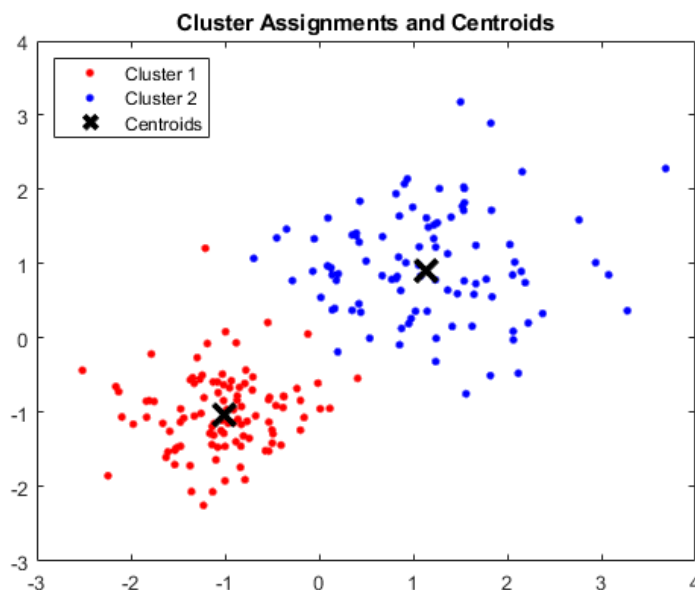
### **3.5.2 K-means a KNN: příbuzní jménem**

Problematiku orientace v rámci dozorované a nedozorované klasifikace, respektive rozdílnost shlukování a klasifikace, zmíněné v předchozím odstavci je možné demonstrovat při srovnání metod klasifikace na základě algoritmů K-průměrů (K-means) a K-nejbližších

sousedů (K-neraest neighbors). Ačkoliv je název intuitivně naznačuje příbuznost obou postupů, ve skutečnosti není podobnost zcela jednoznačná.

Jak uvádí Batta (2020), K-means algoritmus je nedozorovanou metodou clusterizace dat založenou na hledání shluků co nejpodobnějších výsledků v rámci datové báze a konkrétních veličin, ze kterých jsou poté automaticky a nedozorovaně izolovány výstupní třídy bez vstupu analytika. KNN oproti tomu spočívá v klasifikaci datového souboru s přihlédnutím k designovanému rozhodnutí analytického pracovníka, který vyčlení chtěné cílové stavy veličin pro konkrétní třídy, přičemž klasifikátor poté přiřazuje řešené jednotky zmíněným třídám na základě pozice hodnot vůči zvolenému cílovému stavu. Jedná se tedy o dozorovanou klasifikační metodu, kterou je možné pomocí K-means clusterizace řešit, ačkoliv se nejedná o výlučně jediný postup (Muja a Lowe 2009). Tato rozpolcenost jen podtrhuje problematiku orientace v prostoru moderních výpočetních technologií, které se neustále vyvíjí a často je obtížné při vstupu do řešené problematiky prostředí efektivně rozpoznat, jelikož definice se mohou postupem vývoje měnit, rozporovat i doplňovat.

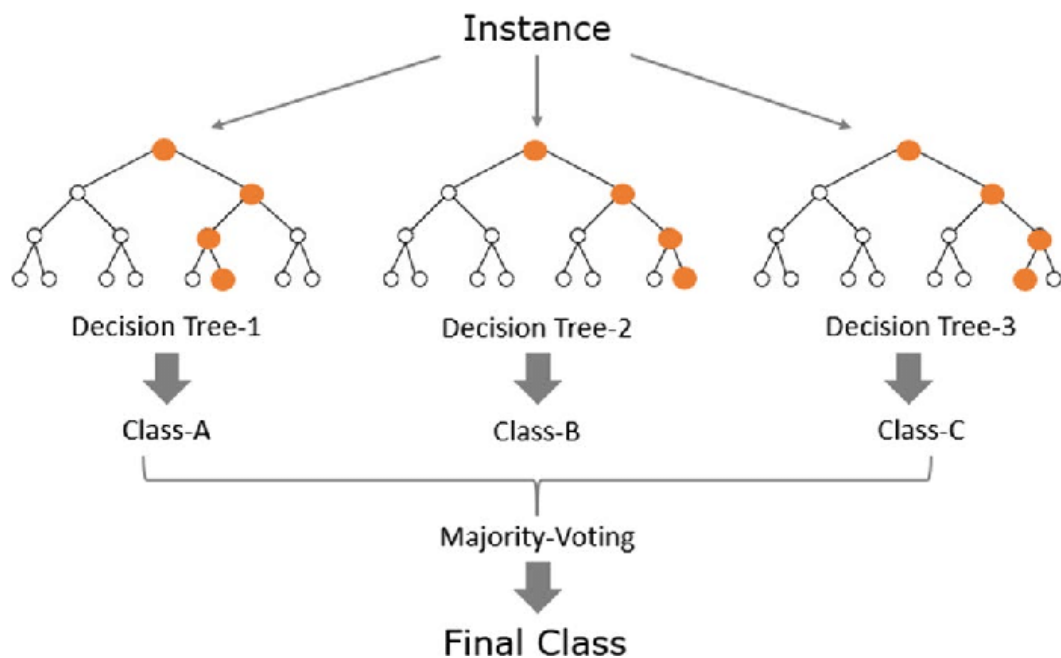
Abych nezapomněl stručně charakterizovat fungování obou metod, souhrnně lze označit KNN a K-means za jednoduché metody se snadnou implementací, které jsou ale neefektivní při zpracování velkého množství dat (Batta 2020). Princip KNN je blíže znázorněn na Obr. 1, K-means algoritmus by pracoval podobně s rozdílem absence vytyčených bodů, které by byly rozpoznány automaticky.



Obr. 1. Graf funkce KNN klasifikátoru s vymezením středových bodů a zařazení jednotlivých hodnot k jednomu či druhému řešení. Zdroj: (MathWorks 2022)

### 3.5.3 Random forest

Jedná se o konkrétní automatickou klasifikační metodu, při které dochází k sestavení sady rozhodovacích stromů do jednotného modelu na základě předem stanovených veličin, na jehož závěru je dosaženo určení výsledku (IBM Cloud Education 2022), v tomto případě klasifikace druhu dřeviny. Princip je graficky znázorněn na schématu Obrázku 2.

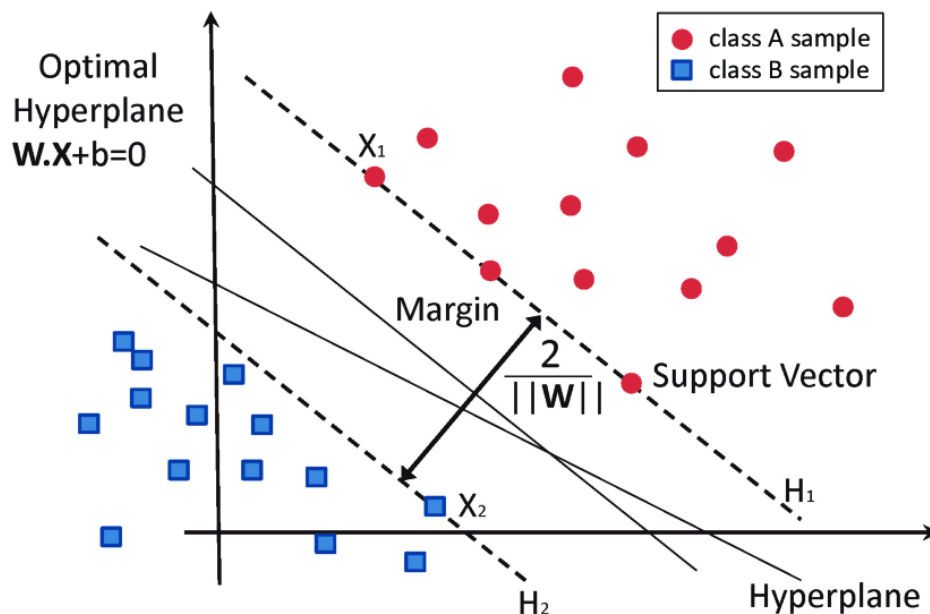


Obr. 2. Schéma modelu Random forest. Zhruba je proměnná testována v jednotlivých rozhodovacích stromech, na jejichž základě je určen výsledek každého jednotlivého stromu. Průměr hlasů všech stromů dává vzniknout finální klasifikaci. Zdroj: (Golze et al. 2020)

Značnou výhodou velká robustnost modelu a nízké riziko tzv. „overfittingu“ (IBM Cloud Education 2022), neboli přehlcení modelu vstupními parametry, což zapříčiní jeho nesprávnou funkci (Shea a Nash 2015). Ačkoliv to v případě výzkumných účelů není zcela patrné, RF model je schopný také zpracování velmi rozsáhlých dat o desítkách až stovkách milionů položek při zachování relativně vysoké efektivity (Genuer et al. 2017). Taková čísla však s největší pravděpodobností tento účel klasifikace nebude běžně dosahovat. Lze proto s jistotou konstatovat, že je zachován předpoklad využitelnosti modelu také při rozsáhlých analýzách celých porostů, ale také není vyloučena expanze a budování modelových parametrů při aplikacích zahrnujících více podobnějších dřevin, což může vyžadovat pro spolehlivou klasifikaci integraci dodatečných počítaných charakteristik.

### 3.5.4 Support vector machine (SVM)

Metoda dozorované klasifikace spočívá ve stanovení rozhodnutí na základě vkládání přímk, které v bodovém grafu oddělují skupiny hodnot za předpokladu, že jsou veličiny dostatečně specifické (Kotsiantis 2007). Jde o dozorovanou klasifikační metodu, která využívá techniky strojového učení (Noble a Street 2006). Důležitým pro popis principu je termín podíl (margin). Vlastní postup zmíněný v publikaci Kotsiantis (2007) je poměrně komplikovaný, nicméně základem tvorba rovin (hyperlane) v největší vzdálenosti od všech bodů ležících na jedné ze dvou polí, které tato nová rovina dělí. Spojnice protínající okrajové body je následně proložena okrajovými body jednotlivých tříd, čímž vzniká tzv. pomocný vektor (support vector) (Kotsiantis 2007). Model zobrazen na Obr. 3.



Obr. 3. Schéma modelu SVM. Dostatečná rozrůzněnost dvou porovnávaných veličin umožní oddělení support/pomocnými vektory. [Zdroj: García-Gonzalo et al. 2016]

Metoda funguje tedy správně pokud jsou veličiny dostatečně rozrůzněné, a tedy oddělitelné pomocí nástrojů SVM, což může být v reálném prostředí problematické a vyžaduje dodatečnou úpravu vstupujících proměnných (Kotsiantis 2007). Řešením může být využití kernelových funkcí, které zjednodušeně řečeno prvotně zjednodušují a ovlivňují rovinu vynášení bodů, což může jednak klasifikaci umožnit v případě nedostatečné či příliš složité podoby dat, kterou není možné jednoduše rovinami separovat (Scholkopf et al. 1999). Výhodou této metody je tím i jistá míra přizpůsobivosti vstupním veličinám, která umožní spolehlivou klasifikaci i při zdánlivě neoptimálním rozdělením hodnotám proměnných

(Zhang a Liu 2013), ačkoliv je oproti zvláště nedozorovaným metodám tímto procesem poměrně komplikovanější z hlediska nastavování a optimalizace modelu.

### **3.5.5 Naive Bayes klasifikátor**

Velmi jednoduchým klasifikačním nástrojem s nízkými nároky na paměť i výkon, který je navíc vybavený schopností škálování parametrů je právě Naive Bayes (Kotsiantis 2007). Principem vychází z předpokladu, že klasifikované parametry v rámci testovaných položek jsou v ideálním případě zcela nezávislé a mohou tak výsledné třídy od sebe diferencovat (Batta 2020). Tato podmínka však může být v praxi špatně stanovitelná, a proto může mít výsledná klasifikace určitou odchylku (Kotsiantis 2007). Pro klasifikaci spojitých dat bodových mračen tudíž se zdá být nevhodný, což potvrzuje absence výskytu tohoto přístupu při nahlížení výsledků výzkumů této problematiky.

### **3.5.6 Klasifikace pomocí hlukového učení a neuronových sítí**

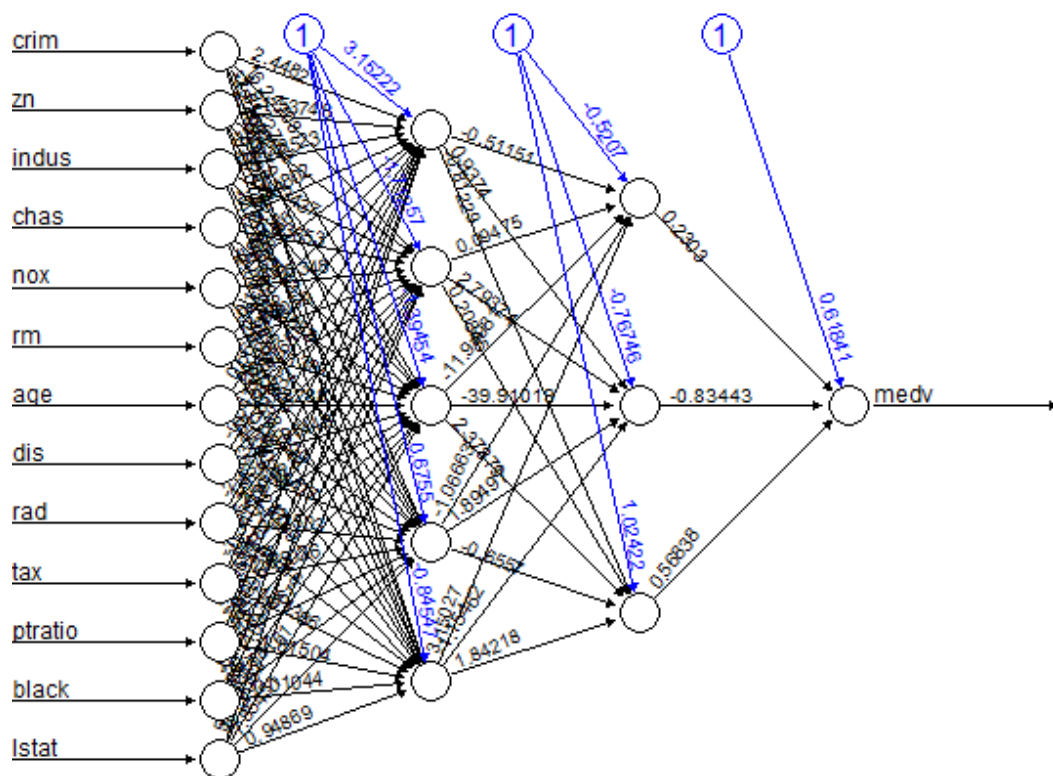
Jedná se alternativní metodu s již prezentovanými slibnými výsledky při řešení druhové klasifikace dřevin (Jin et al. 2018). Využívá neuronových sítí, pomocí jichž je možné automaticky stanovit velké množství klasifikačních veličin bez ohledu na nutnost přímé interakce analytika (Kattenborn et al. 2021), což je podstatnou nevýhodou výpočetní klasifikace. Základem technologie je stejně jako u klasifikačních modelů hluboké učení (deep learning), což je obecné označení vícevrstevné klasifikace pomocí umělé inteligence (Shea a Nash 2015). Oproti samotnému strojovému učení poskytuje zásadní výhodu ve schopnosti vyhodnocení nezpracovaných či základně upravených dat za předpokladu vhodné konstrukce modelu (Lecun et al. 2015).

Je vhodné také definovat rozdíl mezi neuronovými sítěmi (ANN) a konvolučními neuronovými sítěmi (CNN). Jak definuje Shea a Nash (2015), metoda ANN je základem dále rozvíjeným v konceptu CNN, v rámci kterého je klasifikace dále vylepšována. Zjednodušeně tedy platí že: typová klasifikace pomocí neuronových sítí tedy zahrnuje rozčlenění načteného vstupu do dílčích vrstev, dále pak do tzv. skrytých vrstev, ve kterých je poté učiněno klasifikační rozhodnutí na základě předchozí vstupní vrstvy. Stochastické změny rozhodnutí z vrstev předcházejících oproti následujícím ve formě váhy poté vylepšují či přímo určují výsledek klasifikace obrazu. Konvoluční sítě tento postup doplňují aplikací konvolučních vrstev, které zjednodušují či vylepšují klasifikační model a činí proces efektivnějším (Shea a Nash 2015). ANN dokonce nejsou ani ve své základní podobě vhodné pro obrazovou klasifikaci

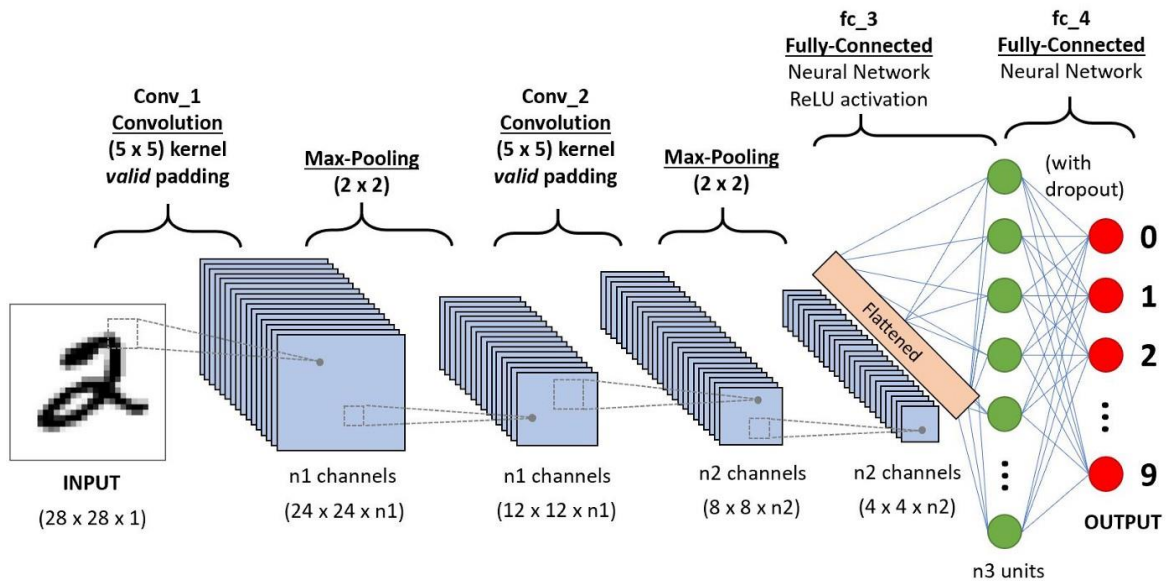
vzhledem k výslednému propojení všech skrytých vrstev s výstupními vrstvami (Obr. 4.), což je i případě nízkého rozlišení datově velmi náročné, CNN ve finální vrstvě vykazují podstatně nižší komplexitu (Kumar 2017).

Jak dále uvádí Shea a Nash (2015), fungování CNN sítí lze dělit do 4 dílčích činností, které svou návazností tvoří procesní řetězec celé metody. Graficky znázorněno na Obr. 5.

- 1) Vstupní obraz je načten ve formě vhodně reprezentující hodnoty jeho pixelů (často v podobě matice).
- 2) Dochází k aplikaci konvoluční vrstvy, kterou jsou identifikovány rozpoznatelné tvary stanoveny výstupní hodnoty jednotlivých neuronů.
- 3) Tzv. „pooling“ je dalším krokem, v rámci kterého dochází k redukci dat a zjednodušení modelu.
- 4) Plné propojení vrstev do jedné finální.



Obr. 4. Schéma ANN sítě demonstrující značný počet jednotlivých vah mezi vrstvami poukazující na složitost tohoto modelu. [Zdroj: ANON. 2016]



Obr. 5. Schéma CNN sítě při obrazové klasifikaci se znázorněním procesu aplikace konvoluční vrstvy a zjednodušení modelu. Zdroj: (Saha 2018)

### 3.6 Metody klasifikace dřevin na základě LiDAR skenů

Ve výzkumech aplikace laserových senzorů se objevují zmínky o odlišných vlastnostech jednotlivých druhů dřevin, které je možné pozorovat na výsledných datech, a to například na různých statistických charakteristikách výsledných bodových mračen (Lin a Hyypä 2016), na odlišném podílu šumu a distribuce bodů (Gollob et al. 2020), či na základě stavby a habitu stromů (Hastings et al. 2020; Kattenborn et al. 2021). Často je základem prací právě vizuální 2D reprezentace tvaru stromů či jejich částí a poměrů mezi jednotlivými sekcemi jakožto východiska pro druhovou determinaci, jak také popisují Michałowska a Rapiński (2021). Pro nejspolehlivější mezidruhovou klasifikaci se zdá být vhodná kombinace obou přístupů, tedy doplnění vizuální reprezentace statistickou deskriptivou (Axelsson et al. 2018).

Pro zvýšení spolehlivosti modelu je na místě také zařadit větší počet automaticky kalkulovaných charakteristik odvozených z LiDAR bodových mračen (Lin a Hyypä 2016). Postupy obrazové i statistické klasifikace jsou nejen v lesnickém odvětví v posledních letech zdokonalovány aplikací pokročilých metod strojového učení. Zatímco na základě klasických přístupů strojového učení je model nutné definovat na základě vlastních znalostí a předpokladů, metody užití konvolučních sítí (Convolutional Network – CNN) slibují větší míru samostatnosti i teoreticky vyšší přesnost (Kattenborn et al. 2021).

Laserem skenované porosty jsou popsány nejen polohou každého jednotlivého bodu, ale také hodnotou intenzity jejich odrazů, čímž mohou poskytnout informaci o stavu a povaze

snímané plochy s možnými rozdíly mezi dřevinami. Tyto intenzity se ale mohou významně lišit mezi jednotlivými stromy, a dokonce i mezi jednotlivými částmi koruny určitého jedince (Lin a Hyypä 2016). Přímé srovnání absolutních hodnot se proto zdá nevhodné, informace by mohla ale být užitečná v případě vyjádření pomocí poměru či indexů. Příklad udává Zhang a Liu (2013), při porovnání statistické klasifikace distribuce bodů a charakteristik jejich intenzit bylo dosaženo nejvyšší přesnosti až 92,8 % oproti klasifikaci pomocí těchto přístupů individuálně.

Při zpracovávání literární rešerše tématu jsem shledal nedostatek publikací se zaměřením na druhovou klasifikaci stromů s využitím dat pozemního laserového skenování v lesnickém kontextu. Prvenství v tomto ohledu patří leteckým přístupům pořízení laserových dat, jak lze odvodit z prací uvedených v následujících kapitolách. Zvolil jsem tedy také články shodující se použitou technologií TLS a MLS či LiDAR aplikace v příbuzných oborech.

### **3.6.1 Metody výpočetní a statistické klasifikace**

Jelikož je z terénního měření vyplývající mračno s desítkami tisíc bodů poměrně obsáhlým statistickým vzorkem, je možné jej postupy popisné statistiky a matematiky definovat. Pokud je v nějakém ohledu určitý parametr druhově specifický, stává jedním z aspektů pro model strojového učení, podle kterého jsou následně druhy dřevin určovány (Lin a Hyypä 2016). Následný postup spočívá v aplikaci klasifikátoru, jako je například Random forest či Support vector machine, což jsou populární techniky užívané při řešení problémů automatické klasifikace dřevin (Michałowska a Rapiński 2021)

Z hlediska nejvyšší možné přesnosti klasifikace se jako nejlepší přístup konstrukce klasifikačního schématu jeví stanovení druhově specifických metrik početního popisu intenzity a výšky všech bodů náležících jednomu stromu, poměrů intenzit mezi jednotlivými částmi stromu, a dalších poměrových a statistických parametrů ve spojení s SVM aplikací (Lin a Hyypä 2016). Maximální dosažená přesnost dosahovala v takovém případě 92,5 %, a to při selekci parametrů s nejvyšší individuální klasifikační přesností v rámci zmíněných kategorií, z celkového počtu 42 parametrů zvolili autoři celkem 15. Alternativní výsledky s aplikací Random forest a dalších klasifikačních přístupů prezentuje Zhou et. al. (2018), přičemž výsledná přesnost těchto postupů klasifikace dosahovala 74 % za použití RF modelu. Zároveň výzkumníci identifikují vhodné metriky pro automatické určování druhů stromů z leteckých LiDAR snímků mezi zhruba 80 parametry. Shledán byl možný neúspěch klasifikace při použití nízkého počtu veličin, nicméně specificky nebyl specificky stanoven. Při aplikaci SVM



dosahuje Terryn et al. (2020) téměř 80% přesnosti klasifikace 7 druhů dřevin na základě 19 různých strukturálních veličin odvozených z TLS skenování.

Na příkladech v předchozím odstavci je opět patrná zdatelně vyšší přesnost klasifikace spojená s větším počtem posuzovaných veličin klasifikačních modelů. Tato skutečnost je však poměrně velkou limitací metody i přes její relativně snadnou aplikaci. Nutnost předem stanovit větší počet klasifikačních veličin je často neefektivní a ne zcela objektivní (Kattenborn et al. 2021). Toto tvrzení podporuje i příklad uvedený v předchozím odstavci. Ve výsledném schématu byla, jak uvádí Lin a Hyyppä (2016), použita pro efektivní klasifikaci pouze asi třetina kalkulovaných atributů.

### **3.6.2 Metody tvarové klasifikace a metody hlubokého učení**

Výchozím výstupem pro další analýzy je v případě LiDAR technologie husté bodové mračno vyžadující následnou úpravu, aby bylo možné extrahovat z datového rámce informaci o tvaru, která je následně použita pro stanovení druhu dřeviny (Ko et al. 2018). Prvotně je nutné vytvořit obraz na základě bodového mračna, který vhodně reprezentuje tvar stromu, ten je následně klasifikátorem zařazen do určité třídy na základě podobnosti s trénovacími daty. Podobnost je poté jednak tvarová, ale také spektrální (Kattenborn et al. 2021), přičemž barevný gradient může sloužit pro vyjádření distribuce určité veličiny v rámci trénovacích dat, v případě analýzy laserového snímkování například intenzity odrazů paprsků v určitých oblastech (Ko et al. 2021).

Alwosheel et al. (2018) zmiňuje možnou překážkou v rámci aplikace metody, kterou představuje potřeba rozsáhlé databáze trénovacích vzorků pro plnou funkčnost této metody, která je závislá na počtu určovaných veličin a může čítat až tisíce vzorků, což může být v rámci výzkumu problematické. S tímto problémem se potýká také Ko et. al. (2018), který řeší nízký výchozí počet vzorků jejich rotací kolem osy stromu, čímž získal navíc 64 násobek základního počtu snímků. Úspěšně se v rámci výzkumu podařilo vytvořit klasifikační schéma pomocí konvoluční sítě s celkovou přesností určení druhu stromu 91 % při 3 různých druzích dřevin na základě leteckých laserových dat.

Literární rešerše tématu odhalila také nedostatek výzkumných prací zaměřujících se na determinaci druhů dřevin z pozemního laserového skenování oproti leteckým přístupům, což může mít za následek různou relevanci zjištěných výsledků. Při letecké aplikaci LiDAR snímače naznačuje Korpela et al. (2010), že například rozdílné průměry intenzit prvních, či

jediných, odrazů mohou sloužit pro druhou klasifikaci dřevin v lese i na nelesních plochách. Dále také vysvětluje, že stanovené trendy se jeví silnější při selekci dat z horních částí koruny oproti druhotným odrazům. Při pozemní klasifikaci může být ale realita jiná, silnější korelace je dle této logiky možné teoreticky odvodit ze spodních pater vegetace či kmenů stromů.

## 4 Metodika

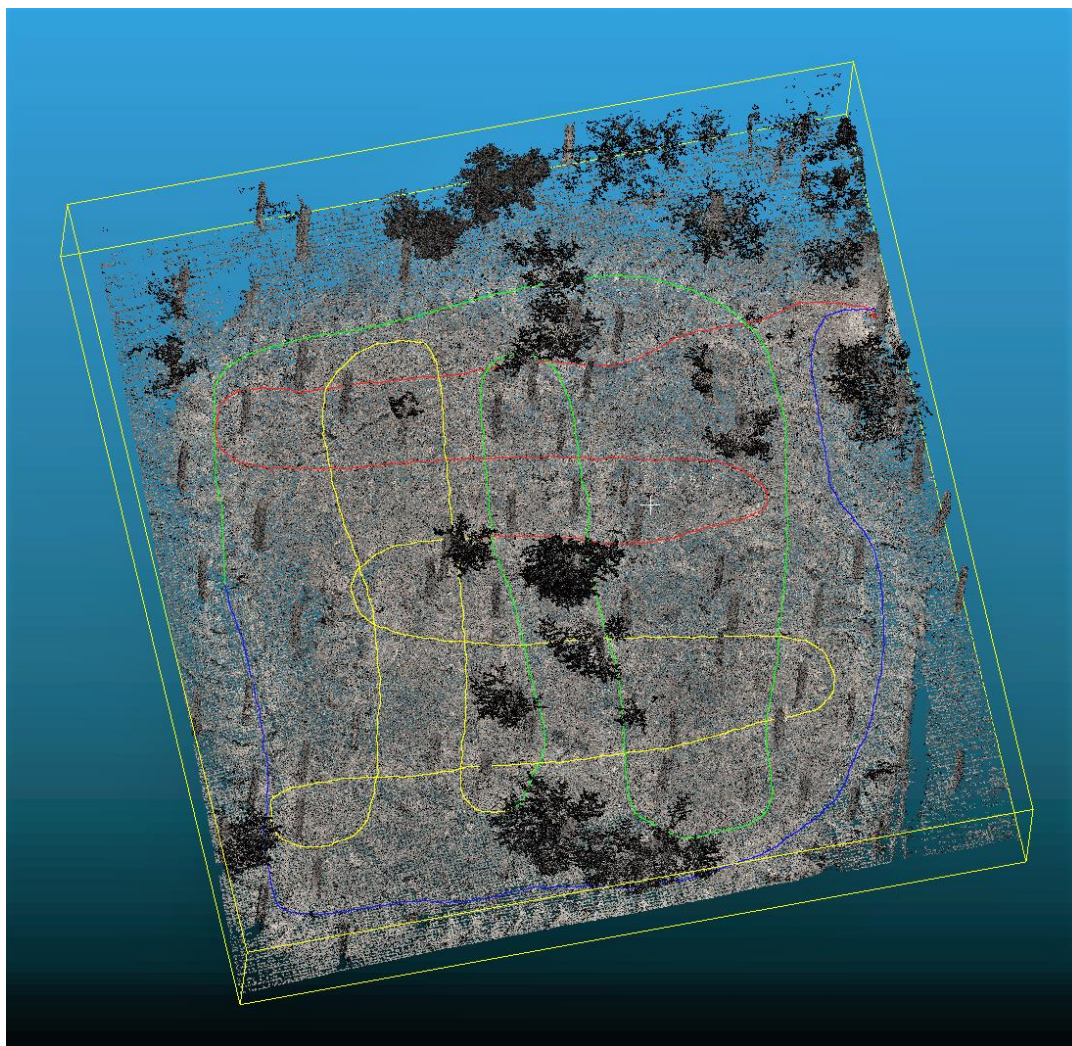
### 4.1 Studovaná oblast a výchozí data

V součinnosti s Katedrou hospodářské úpravy lesů při Fakultě lesnické a dřevařské byla získána data laserového skenu částí porostů nacházejících se na území Školního lesního podniku v Kostelci nad Černými lesy asi 30 km východně od hlavního města Prahy. Z celkem třech lokalit byly měřeny plochy s plným zastoupením Smrku ztepilého (*Picea abies* (L.) H. Karst), Buku lesního (*Fagus sylvatica* L.) a jednoho smíšeného porostu obou dřevin. Celkem datový vzorek čítá 188 stromů, z toho 120 smrků a 68 buků. Snímání bylo provedeno na jaře roku 2021. Z důvodů návaznosti procesů sběru dat byl nejdříve pořízen sken prvního smíšeného porostu, na kterém docházelo k testování technologie, z toho důvodu se jedná o méně kvalitní vzorek. Následující dva datové rámce plynou ze skenování dvou dalších jedno-druhových porostů s lépe optimalizovaným procesem skenování. Rozdíl je patrný již ve velikosti jednotlivých souborů, průměrný vzorek z plochy 3 je asi 40-100 MB velký, přičemž z plochy 1 lze pozorovat velikosti okolo 5-20 MB. Dá se tedy předpokládat nižší hustota bodů i nižší celková kvalita.

#### 4.1.1 Technika získávání dat

Pořízená data vznikala v rámci výzkumné činnosti Fakulty, proto je následující odstavec přejet z úryvku nepublikované metodiky poskytnuté vedoucím práce. Stručné informace jsou uvedeny pouze pro celistvost a celkovou srozumitelnost původu dat.

Snímky byly pořízeny pomocí HMLS LiDAR skeneru GeoSLAM ZEB Horizon. Skenování probíhalo chůzí pracovníka lesním porostem za současného nesení senzoru, který automaticky zaznamenával odrazy vyslaných laserových paprsků, ale také vyhodnocoval svou polohu. Rozhodná byla při činnosti trajektorie chůze porostem, která předurčovala výslednou kvalitu. Příklad trajektorie je zobrazen na Obr. 6. Na základě doporučení výrobce je vhodné do trajektorie zařadit také smyčky, které pak mohou svým průnikem poskytnout re-kalibrační bod a minimalizovat tak chybu v celé smyčce. Data jsou standartně zaznamenána v lokálním souřadnicovém systému, nicméně je možné v případě potřeby provést referenci na požadovaný souřadnicový systém pomocí minimálně 3 zemních kontrolních bodů. Výstupní soubor bodového mračka s příponou LAS byl zpracován pomocí softwaru GeoSLAM Hub (GeoSLAM ltd. 2021).



Obr. 6. Trajektorie chůze porostem č.3 při skenování. [Zdroj: Ing. Karel Kuželka Ph.D.]

#### 4.1.2 Datová charakteristika

Původní celistvé bodové mračno bylo předem v rámci činnosti Katedry hospodářské úpravy lesa při Fakultě lesnické a dřevařské segmentováno na jednotlivé stromy, výchozí data mají tedy podobu bodových mračen reprezentující samostatné jedince uložené v souborech uložených ve formátu LAS. Ten je speciálně navržen pro archivaci či následné zpracování dat laserového skenování a uchovává informaci nejen o poloze jednotlivých bodů v trojrozměrném prostoru, ale také o intenzitě odrazů laserových paprsků. Se souborů ale nebyl předem vyloučen šum, a data také nebyla selektována z hlediska správnosti segmentace jedinců ze vstupního komplexního bodového mračna. V prvotní fázi jsem tedy provedl ruční čištění a výběr reprezentativních vzorků pro trénink a testování klasifikátorů. Zejména v rámci plochy 1 byl ze vzorku vyloučen větší počet jedinců, kteří byli shledáni nekompletní nebo zavádějící. Tento jev se nevyskytoval v případě ploch 2 a 3. V případě všech 3 ploch byla důvodem pro vyřazení byla nesprávná segmentace jedinců.

Tab. 1 Charakteristika dřevinné skladby jednotlivých ploch s vyznačením počtu zahrnutých a vyřazených jedinců. SM: Smrk ztepilý (*Picea abies*), BK: Buk Lesní (*Fagus sylvatica*)

číslo plochy	počet stromů	z toho SM	z toho BK	vyřazené SM	vyřazené BK	zahrnuté SM	zahrnuté BK
č.1	95	75	20	22	5	53	15
č.2	47	0	47	0	2	0	45
č.3	46	45	1	5	0	40	1
<b>celkem</b>	<b>188</b>	<b>120</b>	<b>68</b>	<b>27</b>	<b>7</b>	<b>93</b>	<b>61</b>

Zobrazování a manipulace s daty byla realizována ve volně dostupném programu CloudCompare (GPL software 2022). Po manuální úpravě a třídění byla data uložena ve formátu textového souboru v souladu s formátem s Americkým standardem kódu pro datovou výměnu (ASCII), aby bylo možné je následně vyhodnocovat a upravovat v prostředí programovacího jazyku R, konkrétně v kompilátoru RStudio (RStudio Team 2021) za použití specifických balíčků při aplikaci jednotlivých metod. Díky užití univerzálního datového formátu je možné také užití alternativního programového prostředí, pokud by to budoucí postup vyžadoval.

## 4.2 Návrh klasifikačních modelů

V rámci této práce jsou po studiu literatury i zvážení předpokladů pro implementaci v předpokládaném rozsahu uvažovány celkem dvě hlavní metody automatické klasifikace druhů dřevin, jmenovitě metoda matematicko-statistického popisu číselných dat a metoda klasifikace na základě strukturální obrazové reprezentace. Oba zmíněné přístupy svým výstupem poskytnou číselné proměnné, které následně využije klasifikátor Random forest pro konečné určení druhu dřeviny. Postupy jsou v jádru principálně podobné a vycházejí ze stejného datového základu, na jehož bázi bude možnost užití jednotlivých metod testována. Určitá část vstupních dat je použita jako trénovací vzorek, přičemž zbytek slouží pro ověření přesnosti klasifikace a označuje se jako vzorek zkušební. Na základě studia dokumentace klasifikačního softwaru byl zvolen jednotný poměr 7:3 s převahou trénovacího vzorku. V rámci rozboru problematiky bylo možné ve výzkumné praxi pozorovat aplikaci uvažovaných i příbuzných technik, nicméně vzhledem k odlišnému způsobu získávání dat i rozdílným nárokům jednotlivých metod je vhodné možnosti i podmínky aplikací ověřit.

#### 4.2.1 Matematicko-statistický popis bodových mračen

Pro aplikaci vlastního klasifikátoru je nejprve zapotřebí zvolit a kalkulovat relevantní veličiny, které budou co nejsilněji korelovat s konkrétním druhem dřeviny. Indexy mohou mít různou podobu od prostých průměrů až po matematické funkce podle zkušeností a předpokladů analytika. Stanovují se přitom tak, aby co nejlépe vystihovaly největší možné rozdíly mezi dřevinami, tedy například ve výšce stromů, distribuce biomasy v podél profilu kmene či možných rozdílech v intenzitě odrazů laserových paprsků u listnáčů a jehličnanů. Postup takových operací je řešen v rámci vlastního kódu pomocí programovacího jazyku R snadno proveditelný bez nutnosti instalace dodatečných balíčků. Pro efektivní provedení těchto úkonů byl proto využito prostředí programovacího prostředí RStudio (RStudio Team 2021).

Tab. 2. Popis veličin vybraných pro výpočetní klasifikační modely na základě Lin a Hyypä (2016)

označení v kódu	popis
SD	směrodatná odchylka intenzit celého stromu
iMean020	průměr intenzit v rozmezí 0-20% výšky stromu
iMean040	průměr intenzit v rozmezí 20-40% výšky stromu
iMean060	průměr intenzit v rozmezí 40-60% výšky stromu
iMean080	průměr intenzit v rozmezí 60-80% výšky stromu
iMean100	průměr intenzit v rozmezí 80-100% výšky stromu
iMed020	medián intenzit v rozmezí 0-20% výšky stromu
iMed040	medián intenzit v rozmezí 20-40% výšky stromu
iMed060	medián intenzit v rozmezí 40-60% výšky stromu
iMed080	medián intenzit v rozmezí 60-80% výšky stromu
iMed100	medián intenzit v rozmezí 80-100% výšky stromu

Na základě předpokladů, které uvádí Lin a Hyypä (2016), a dle vlastní iniciativy bylo pro tyto účely vybráno celkem 11 veličin (viz. Tab. 2.) odvozených z intenzity odrazů laserových paprsků, které v podobné sestavě v citované publikaci umožnily asi 80% přesnost určení druhu dřeviny za přítomnosti 2 jehličnatých a 2 listnatých druhů. Částečně přejatá východiska byla tedy ověřena na datech pořízených HMLS senzorem, což je v souladu s cílem této práce. Výsledné hodnoty byly zapisovány do datového rámce s odkazem na konkrétní strom, aby bylo možné v následujícím kroku aplikovat data v rámci klasifikátoru.

#### 4.2.2 Klasifikace na základě strukturálních charakteristik

Oproti početním charakteristikám je základem metod tvarové klasifikace vizuální reprezentace zájmového objektu ve formátu rozpoznatelného počítačovým algoritmem. Surová data laserových mračen nejsou pro tyto účely ideálním východiskem, jelikož jsou poměrně komplexní a obecná ve smyslu absence optimální reflexe charakteristického znaku důležitého pro následnou klasifikaci. Před samotnou aplikací vlastního klasifikátoru je proto nutná určitá transformace či alternativní vyobrazení dat za použití analytických nástrojů a výpočetní techniky. Svým způsobem se tedy jedná také o aplikaci matematických metod a podstata je v zásadě paralelní s předchozím postupem, nicméně důležité je základní východisko, které je zde skutečně vizuální reprezentací, jež je následně pomocí nástrojů a funkcí popisována a zkoumána.

Při hledání ideálního východiska byl zvolen následující postup tvorby klasifikačních vzorků. Bodová mračna jsou pro účely této práce dále segmentována na podmnožiny ve tvaru pruhů s šířkou 1 metr a středem shodným se středem stromu. Předpokládaným výsledkem je poté optimální reprezentace kmenového profilu včetně směru růstu větví. Dále je segment procesem voxelizace převeden do matice s proměnnou a dále optimalizovanou velikostí jednotlivé buňky dle požadavků grafického výstupu. V zásadě jde o tvorbu dvourozměrné matice kryjící svým rozsahem předem segmentovaný pruh bodového mračna, přičemž do každé buňky bude zapsána suma intenzit odrazů laserových paprsků všech bodů prostorově přílišných dané pozici v matici. Finální krok spočívá v generování grafické podoby matice v barvách stupňů šedi reprezentující gradient intenzit odrazů barevnou škálou ve výsledné, PNG souboru. Ta byla realizována za pomoci softwarového balíčku png (Urbanek 2013) v rámci prostředí RStudio (RStudio Team 2021).

V tomto bodě se nabízí dle poznatků uvedených v literární rešerši použití klasifikačního modelu na základě neuronových či konvolučních sítí. Nicméně je pro praktické využití byl shledán takový postup ve srovnání s alternativami značně složitější nehledě na nutnost značně rozměrného klasifikačního vzorku (viz. 3.5.6.). Pro řadu komplikací byl tedy tento postup zavrhnut, avšak jde jistě o zajímavé a perspektivní téma pro další výzkum problematiky klasifikace dřevin v bodových mračnech, který již v příbuzných aplikacích ukázal velice slibné výsledky (viz. 3.6.2.)

Pro extrakci tvarových charakteristik byl proto zvolen model Houghovy transformace liniových prvků grafických snímků na přímkou definované počátečním bodem, úhlem a

hodnotou skóre, tedy jakousi škálou významnosti konkrétní linie, kterou stanoví algoritmus. Předpokládaným výstupem je dostatečně specifická podmnožina linií pro dotyčnou dřevinu zejména na základě krytí větví vzniklými přímkami.

Samotný model je dostupný v prostředí programovacího jazyku R v podobě balíčku *imager* (Barthelme et al. 2022) s funkcí *hough\_transform*. Postup je opět poměrně jednoduchý a spočívá pouze v načtení grafického vstupu ve stupních šedi a možném nastavení limitních faktorů *theta* či *rho*. Výsledný datový rámeček čítá reprezentace jednotlivých přímek parametry vzdálenosti od středu *rho*, úhlem *theta* v radiánech a hodnocením přesnosti detekce linie *score*. Jak již bylo zmíněno výše, předpokládaným charakteristickým znakem jsou především větve, je proto na místě eliminovat linie jasně reprezentující jiný prvek, jako je například kmen či podrost pomocí matematických definicí. Linie byly filtrovány následujícím způsobem:

- 1) Výběr pouze linií s **kvantilem skóre vyšším než 0,99**
- 2) **Eliminace linií v blízkosti země**, které se tvořily na základě podrostu ve snímku
- 3) **Odstranění linií shodných s kmenem** z důvodu nespecifičnosti

V poslední řadě přípravného procesu stojí popis výsledných linií matematicko-statistickými metodami, aby byla výsledkem číselná vyjádření vhodná pro nasazení do klasifikačního modelu Random forest, jak tomu bylo v předchozí metodě. V tomto případě bylo zvolen zásadně větší počet veličin, jelikož již neexistovala opora v již publikovaných výsledcích, s touto metodou spojený výzkum nebyl nalezen. Přehled veličin je vyobrazen v Tab. 3.

Celkem bylo vybráno:

- **24 veličin v 8 sadách po 3 položkách** – průměr, medián a směrodatná odchylka.
- Jednotlivé sady pocházely z **2 datových rámečků** – základní a přímky z kvantilu skóre 0,7-1,0
- Pro účely jednoduššího strojového vyhodnocení byl úhel *theta* převeden z rozsahu  $(0 - 2\pi)$  do rozmezí  $(-\pi - \pi)$  s následnou transformací na *rotaceTheta*  $(0 - \pi)$ .
- **4 sady** v každém rámci testovaly veličiny *rho*, *rotaceTheta*, *rho x rotaceTheta*, *score*



Tab. 3. Seznam proměnných použitých v klasifikačním modelu obrazové klasifikace. Modře je vyobrazen kompletní datový rámec, oranžově je specifikován soubor dat s liniemi s hodnotou skóre v kvantilu 0,7-1,0 původních dat. Odstíny jsou odlišeny jednotlivé sady po 3 testovaných prvcích: průměr, medián, směrodatná odchylka.

označení v kódu	popis
meanThetaRho	průměr úhlů Theta transformovaných
medThetaRho	medián úhlů Theta transformovaných
sdThetaRho	směrodatná odchylka úhlů Theta transformovaných
meanThetaRot	průměr úhlů Theta x vzdálenost Rho
medThetaRot	medián úhlů Theta x vzdálenost Rho
sdThetaRot	směrodatná odchylka úhlů Theta x vzdálenost Rho
meanRho	průměr vzdáleností Rho
medRho	medián vzdáleností Rho
sdRho	Směrodatná odchylka vzdáleností Rho
meanScore	Průměr Skóre
medScore	Medián Skóre
sdScore	Směrodatná odchylka Skóre
meanThetaRho30	průměr úhlů Theta transformovaných
medThetaRho30	medián úhlů Theta transformovaných
sdThetaRho30	směrodatná odchylka úhlů Theta transformovaných
meanThetaRot30	průměr úhlů Theta x vzdálenost Rho
medThetaRot30	medián úhlů Theta x vzdálenost Rho
sdThetaRot30	směrodatná odchylka úhlů Theta x vzdálenost Rho
meanRho30	průměr vzdáleností Rho
medRho30	medián vzdáleností Rho
sdRho30	Směrodatná odchylka vzdáleností Rho
meanScore30	Průměr Skóre
medScore30	Medián Skóre
sdScore30	Směrodatná odchylka Skóre

#### 4.2.3 Random forest klasifikátor

Metoda klasifikace pomocí rozhodovacích stromů vyžaduje kromě vstupních dat také nastavení tří parametrů: počtu uzlů (*nodesize*), počtu rozhodovacích stromů (*ntrees*) a počtu posuzovaných veličin. Výsledná rozhodnutí jednotlivých stromů potom určují výsledek na základě většiny (IBM Cloud Education 2022). Z důvodu přesnosti klasifikace a možné tvorby nežádoucích odchylek je vhodné používat v rámci aplikace této metody výhradně číselných datových formátů (Strobl et al. 2007). Pokud je nutné na určité úrovni rozhodovacího stromu vylišovat například názvy, které mají standartně textovou podobu, je vhodné užití číselných indexů názvy reprezentujících, jak dále navrhuje Strobl et. al. (2007).

Při konstrukci klasifikátoru je zapotřebí také tréninkových dat, stejně tak jako zkušebního vzorku, kterým je model poté testován. Typicky se data rozdělují na základě empirických vztahů zhruba v poměru 2:1 s převahou trénovacích dat, jak lze pozorovat na publikacích Strobl et al. (2007) i Kruppa et al. (2013). V rámci této práce byl zvolen poměr 7:3 s cílem maximalizovat velikost trénovacího vzorku se současným zachováním dostatečně objemného zkušebního segmentu. Zároveň je tento poměr blízký rozdělní 2:1, a současně byl shledán jako vhodnější pro tuto konkrétní aplikaci a zástavbu v kódu.

Zpětnou evaluaci výsledků, přesnosti jednotlivých veličin a nastavení vstupních parametrů v rámci klasifikačního schématu je možné zvýšit celkovou přesnost (Biau a Scornet 2016). Po průběhu klasifikace je možné snadno vyhodnotit také důležitost jednotlivých volených veličin pomocí Giniho koeficientu (IBM Cloud Education 2022). Do výsledného modelu je tak možné zařadit pouze relevantní položky, pokud má tento krok za následek zvýšení přesnosti či efektivity práce celého softwaru. Kromě selekce klasifikačních veličin je také možné nastavovat parametry modelu Random forest, tedy zpětná úprava parametrů může mít na klasifikaci vliv, zejména pokud je třeba klasifikační model zjednodušit (Kruppa et al. 2013).

Implementace modelu byla realizována prostřednictvím softwarových balíčků *randomForest* (Breiman et al. 2022) a *caTools* (Tuszynski a Dietze 2021) v prostředí kompilačního prostředí RStudio (RStudio Team 2021). Výhodou byla zejména přítomnost analytických funkcí a nástrojů přímo prostřednictvím těchto softwarových doplňků. Nebylo tedy nutné konstruovat vlastní kód, který by matice a grafy sestavoval z hodnot používaných klasifikátorem.

Pro vyhodnocení přesnosti klasifikátoru a jeho parametrů byly použity následující nástroje:

- 1) **matice záměn** – zobrazení faktické funkčnosti klasifikátoru prostřednictvím konkrétních chyb při testování zkušebního vzorku a na vlastních trénovacích datech.
- 2) **graf vztahu chybovosti a počtu použitých stromů** pro optimalizaci počtu stromů a pro zobrazení závislosti chyby klasifikované třídy na celkové chybě klasifikátoru.
- 3) **Giniho koeficient** pro hodnocení významnosti jednotlivých parametrů na základě jejich užitečnosti při klasifikaci.

## 5 Výsledky

Cílem této práce je ověření možnosti klasifikačních metod z detailních bodových mračen s využitím výpočetní techniky. Mezi výsledky budou tedy uváděny parametry pro tyto účely sestavených modelů a v návaznosti i výsledky vlastní přesnosti klasifikátoru, ale také ověření přesnosti jednotlivých postupů na základě schémat uvedených v předcházející kapitole. Faktický rozdíl použitých metod, jak již bylo uvedeno dříve, spočívá pouze v rozdílných vstupních parametrech a základní myšlence odvození dat, jelikož byl v rámci účelnosti zvolen identický koncový prvek, tedy klasifikátor Random forest.

### 5.1 Matematicko-statistická klasifikace

Prvním testovaným postupem byla pro svou přímočarost metoda využívající čistě číselný popis jednotlivých prvků bodových mračen a již v této fázi poskytla velice slibné výsledky. Výsledná přesnost klasifikace pomocí algoritmu Random forest při zahrnutí kompletně všech 188 stromů ve formě datových mračen ze 3 výzkumných ploch bez předchozího výběru dosahovala zhruba 95-98,5 %, což lze pokládat za úspěch. Uváděný rozptyl má opodstatnění v náhodném sestavování trénovacího a zkušebního vzorku při opětovném spuštění klasifikátoru za účelem testování. Odlišné složení potom podává určitý rozptyl výsledných přesností. Po okulární selekci a vyřazení většiny evidentně zavádějících či nekompletních kusů byl počet redukován na 154 jedinců. Výsledná odhadovaná průměrná přesnost klasifikace dosahovala v tomto případě **98,7 %** se směrodatnou odchylkou chyb **0,8 %**. Kvůli podezření na možnou nekvalitu vstupních dat pocházejících ze zkusné plochy 1 byl za účelem testování tohoto předpokladu použit vzorek sestavený pouze ze dat ploch 2 a 3 s výsledkem **100%** úspěšnosti klasifikace, a to ve všech testovaných případech. Rozdíl ve výsledcích tedy naznačuje, že za použití optimální metodiky je možné dosáhnout vyšší kvality bodových mračen, která může mít přímý vliv na výslednou kvalitu odvozovaných veličin.

Jak je možné odvodit z Tab. 4 a 5, nedostatečná velikost vzorku zřejmě způsobuje nekonsistenci výsledků také chybovostí modelu. Ačkoliv odhad výsledné chyby dle klasifikačních dat v konkrétním případě náhodné distribuce vzorků do klasifikační a zkušební skupiny (viz. kapitola 4.2.3) nebylo na zkušebním vzorku dosaženo očekávané chybovosti.

Tab. 4. Matice záměn – hodnocení klasifikátoru na klasifikačním vzorku.

	bk	sm	chyba
bk	43	1	2,27 %
sm	1	61	1,16 %
<b>Odhad celkové chybovosti</b>			<b>1,89 %</b>

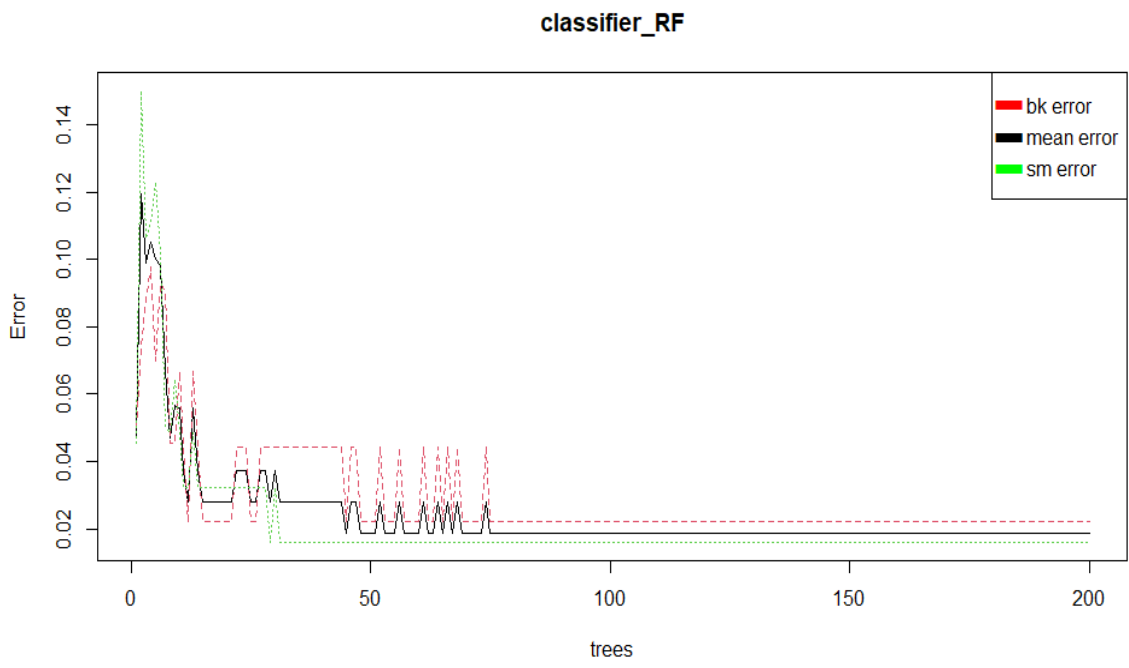
Tab. 5. Matice záměn – hodnocení klasifikátoru na zkušebním vzorku.

	bk	sm	chyba
bk	20	0	0 %
sm	0	28	0 %
<b>Celková chybovost</b>			<b>0 %</b>

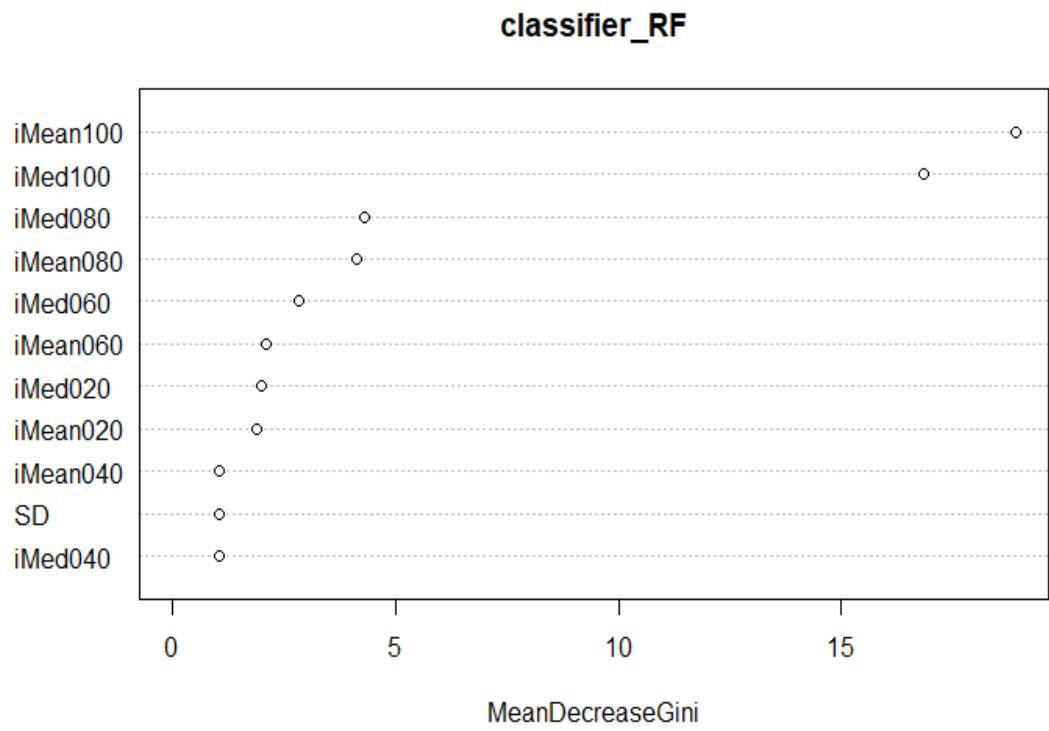
### 5.1.1 Parametry modelu Random Forest

Opakovaným spouštěním klasifikačního skriptu byl testován vliv nastavitelných parametrů modelu na výslednou klasifikaci. Ze základních se jedná o počet rozhodovacích stromů (*ntree*) a minimální velikost konečné buňky na konci větvi (*nodesize*). Parametr *nodesize* neprojevil při postupném snižování až na nulovou hodnotu významný vliv na průběh procesu, zřejmě je tento úkol příliš jednoduchý na to, aby byla nutná restrikce počtu úrovní stromů za účelem zvýšení rychlosti klasifikace. Počet stromů (*ntree*) opět neprojevil zásadní vliv při razantním navýšení jeho hodnoty, ačkoliv bylo nutné jej nastavit tak, aby nízkým počtem neomezoval přesnost klasifikace (viz. Obr. 7), při nízké hodnotě docházelo ke zvýšení chybovosti. Opět se jedná spíše o parametr vhodný k zjednodušení modelu při analýzách rozsáhlých datových rámců s mnoha položkami.

Naopak klíčovými faktory přesnosti klasifikace se dle očekávání prokázaly vlastní zvolené datové parametry. Zajímavým zjištěním byly rozdíly podílů na správnosti klasifikace mezi jednotlivými faktory. Z celkového počtu 11 proměnných se prokázaly pouze 2 jako mimořádně přínosné, a to konkrétně medián a průměr intenzit v korunách stromů, tedy v kvantilu 0,8-1,0 výšky stromu. Jak lze vyčíst z Obr. 8, ostatní veličiny nižších výškových kvantilů jsou z pohledu poklesu Giniho koeficientu podstatně vzdáleny a klesají na významnosti s klesající výškou stromu s již značně menšími rozdíly. V rámci evaluace a zpřesňování modelu by tedy bylo možné většinu parametrů vyřadit a ponechat pouze mediány a průměry výškových kvantilů 0,6-0,8 a 0,8-1, nicméně vzhledem k malému objemu dat, celkové nízké náročnosti a rychlé odezvě klasifikátoru taková úprava se nezdá být nutná.



Obr. 7. Graf závislosti počtu klasifikačních stromů na ose X a chybovosti na ose Y pro jednotlivé třídy a jejich průměr.

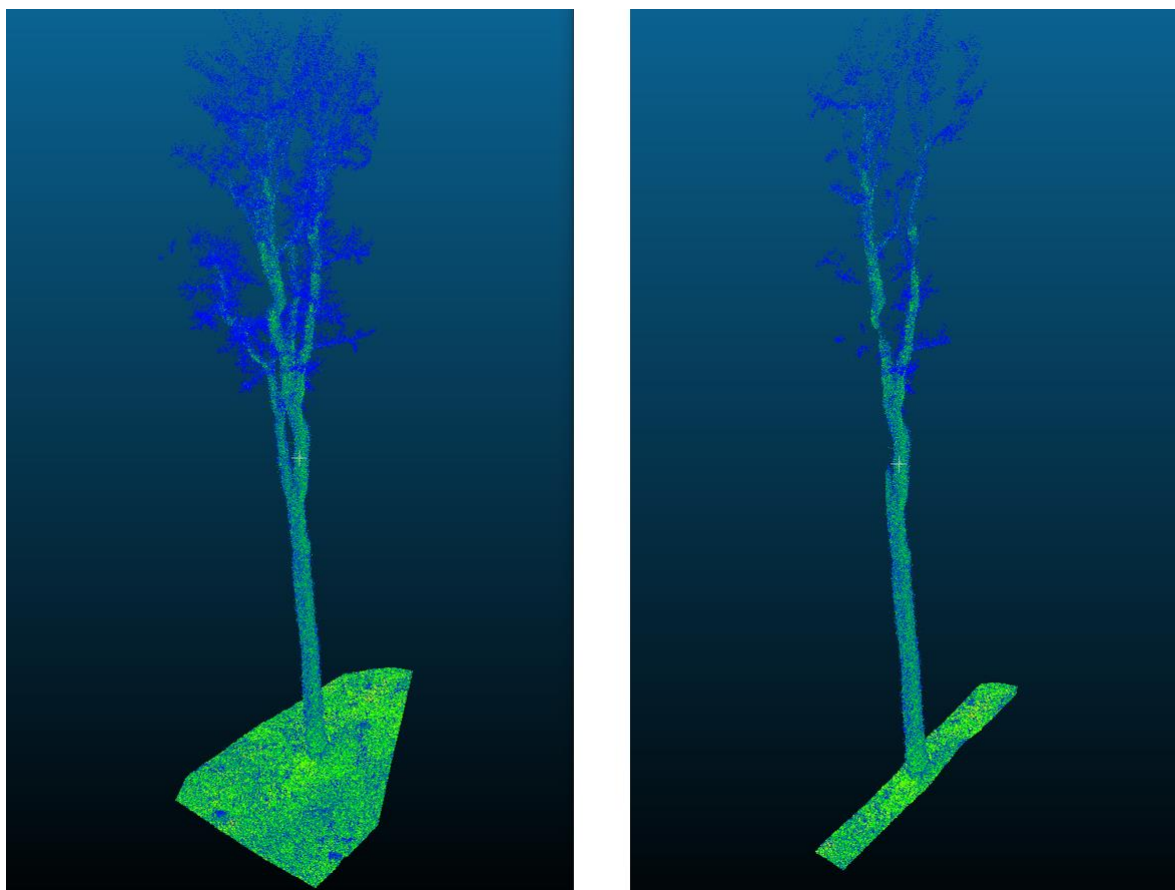


Obr. 8. Graf hodnot Giniho koeficientu značícího podíl parametrů na klas. model.

## 5.2 Klasifikace na základě strukturálních charakteristik

Klasifikace dřevin na grafickém základu byla poměrně složitější přípravou, celkovým časem zpracování i celkovým sestavením řešení. Ačkoliv samotný ořez dat na datové pruhy o šířce 1 metr (Obr. 9.) byl velmi rychlou záležitostí, zejména tvorba klasifikačních vzorků v podobě grafických souborů byla výpočetně velmi náročná a zabrala v průměru 4-6 minut na jeden snímek na běžném kancelářském počítači. Jedná se o bezkonkurenčně nejdéle trvající úkon v rámci celé práce, ostatní výpočetní operace byly provedeny v řádech desítek či jednotek pro jednotlivou zpracovávanou položku. Následně byl grafický výstup popsán pomocí přímek získaných funkcí Houghovy transformace. Posléze proběhla analýza a klasifikace číselných koeficientů modelem Random forest.

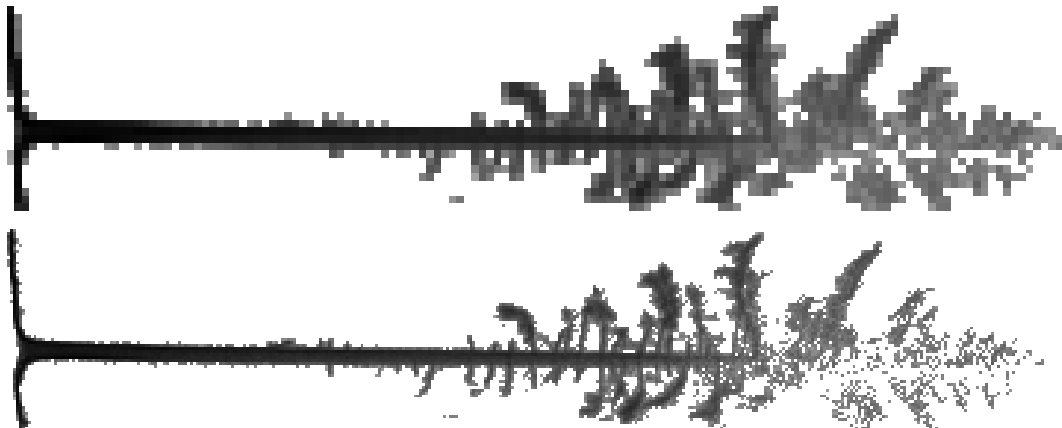
### 5.2.1 Voxelizace a příprava grafických dat



Obr. 9. Výsledek automatického ořezu vstupních dat na datové pruhy.

Byl pozorován pochopitelný, ale překvapivě významný, trend růstu času výpočtu s klesající velikostí jednoho voxelu, tedy jedné pozice v rámci matice. Finálně byla zvolena

velikost buňky  $cellsize = 0,1$  m, což je nižší hodnota oproti původnímu předpokladu s výchozí hodnotou  $0,25$  m. Bylo tak rozhodnuto zejména kvůli nízkému výslednému rozlišení. Jeden voxel totiž na výsledném PNG grafickém souboru reprezentuje právě jeden pixel, což znamená při výšce stromů průměrně asi  $30$  metrů a rádiusů korun okolo  $10$  metrů velikost obrázku zhruba  $300 \times 100$  pixelů. Ukázka rozdílu nastavení tohoto faktoru je zobrazena na Obr. 10. Část výsledné matice sumy intenzit je demonstrována na Obr. 11.



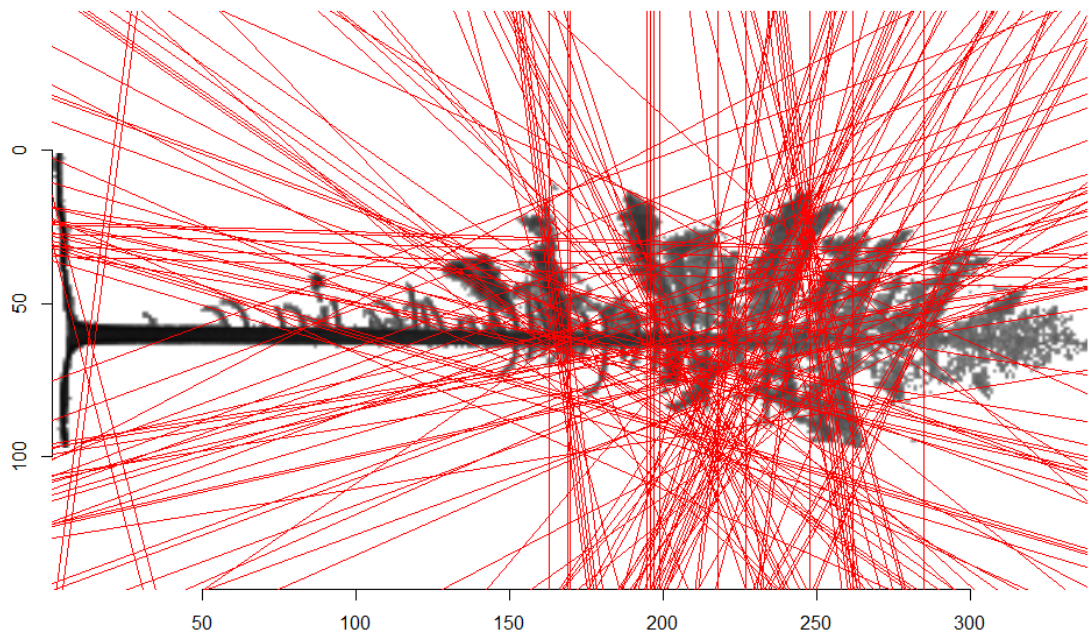
Obr. 10 Demonstrace rozdílu efektu velikosti voxelu  $0,25$  m nahoře a  $0,1$  m dole.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	0	3910512	140579	0	0	0	0	0	0	0	0	0	0
2	56283	28291331	0	0	0	0	0	0	0	0	0	0	0
3	145719	44111737	7710	0	0	0	0	0	0	0	0	0	0
4	4770434	34057897	2313	0	0	0	0	0	0	0	0	0	0
5	9302886	30839743	0	0	0	0	0	0	0	0	0	0	0
6	1280374	35128816	5447629	7655773	8489224	10892174	12706080	10348876	2223050	36751	15420	15420	2056
7	114622	39509381	23134369	28731829	29671678	29287720	30717154	30947426	36565446	38032402	36979216	34626895	33216479
8	11002684	21960136	0	0	0	0	0	0	12079	25700	51657	56283	86352
9	26448898	20265992	0	0	0	0	0	0	0	0	0	0	0
10	39179136	3254391	0	0	0	0	0	0	0	0	0	0	0
11	44015619	0	0	0	0	0	0	0	0	0	0	0	0
12	38501427	0	0	0	0	0	0	0	0	0	0	0	0
13	38154734	0	0	0	0	0	0	0	0	0	0	0	0
14	37469572	0	0	0	0	0	0	0	0	0	0	0	0
15	39243129	0	0	0	0	0	0	0	0	0	0	0	0
16	39090471	0	0	0	0	0	0	0	0	0	0	0	0

Obr. 11. Snímek obrazovky prostředí RStudio: zobrazení ilustračního výseku výsledné voxelové matice do výšky  $1,3$  m (sloupec V13).

### 5.2.2 Hough transformace a výstupní hodnoty

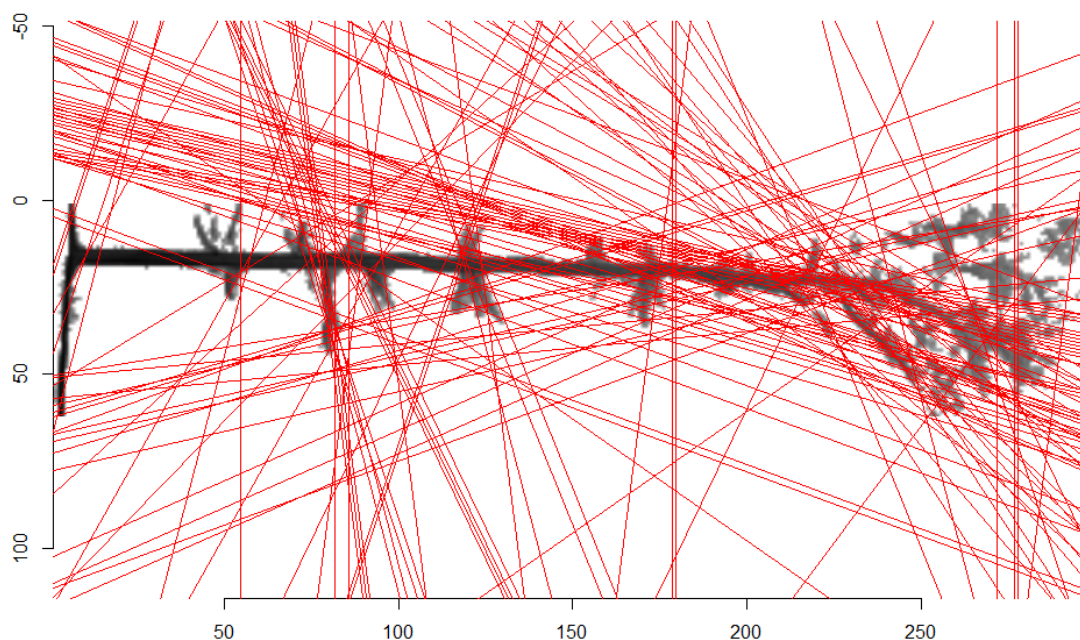
Po provedení Houghovy transformace obašoval základní soubor průměrně nižší desítky tisíc linií, přičemž úpravy uvedené v kapitole 4.2.2 zúžily výběrem počet na přibližně 100 – 200 výsledných čar. Kíčovými se při testování ukázal parametr modelu  $ntheta$ , tedy počet „bins“ neboli tříd či intervalů. V zásadě má faktor přímý vliv na celkový počet vytvořených linií. Pro účely této aplikace bylo rozhodnuto tento paramter snížit z výchozí hodnoty 100 na 50, čímž bylo dosaženo zásadní redukce počtu čar, kterýžto se zdál pro účely v zásadě jednoduchého obrazu účelný.



Obr. 12. Výstupní snímek aplikace Houghovy transformace na smrku.

Lépe byly v tomto ohledu popsány snímky smrků, zatímco buky vykazovaly nižší počet linií zachycujících větvení jednotlivých stromů. Nicméně i tento rozdíl podtrhuje určitou specifičnost popisu vzorků touto metodou a může být vhodným klasifikačním faktorem v následujícím postupu. Rozdíly fungování metody jsou patrné na příkladech Obr. 12 a 13.





*Obr. 13. Výstupní snímek aplikace Houghovy transformace na buku.*

### 5.2.3 Výsledky klasifikace

Při testování klasifikační metody založené na obrazové klasifikaci byla opět porovnávána přesnost postupu při zahrnutí dat ze všech 3 zkoumaných ploch se vzorkem zahrnujícím pouze kvalitnější plochy 2 a 3. V tomto případě však výsledky neprokázaly zásadní rozdíl těchto dvou vzorku, což může znamenat větší univerzálnost a lepší schopnost zpracovávat také data o horší kvalitě. Výsledky se při testování v omezeném rozsahu shodovaly v takové míře, že bylo možné předem vyloučit statisticky prokazatelnou odchylku testování obou datových rámců. Nebyl tedy důvod testovat zvlášť kompletní i selektovaný vzorek a analýza byla tedy provedena pouze na kompletním datovém rámci čítajícím 154 stromů.

Celková přesnost této metody je však znatelně nižší. Při opakovaných pokusech dosahovala včetně randomizované distribuce vzorků do testovací a zkušební skupiny (viz. 4.2.3.) přesnosti v průměru **85,15 %**. Odhadovaná chybovost je tedy v tomto případě dosahuje hodnoty **14,85 %** se směrodatnou odchylkou **2 %** při opakovaných pokusech.

Obdobně jako v předchozí kapitole se i zde projevila nekonzistence výsledků matic záměn v rámci jedné vybrané zkoušky klasifikačního a testovacího vzorku, jak je patrné

z Tabulek 6. a 7. V obou případech byl ale rozpoznán buk jako častěji chybový, při testování vlastního klasifikátoru dokonce zcela zásadně převažoval. Při ověření prostřednictvím zkušební vzorku se již tak značný rozdíl neprojevil, nicméně nutno brát v úvahu jeho velikost a tím i věcnost tohoto závěru.

Tab. 6. Matice záměn – hodnocení klasifikátoru na klasifikačním vzorku.

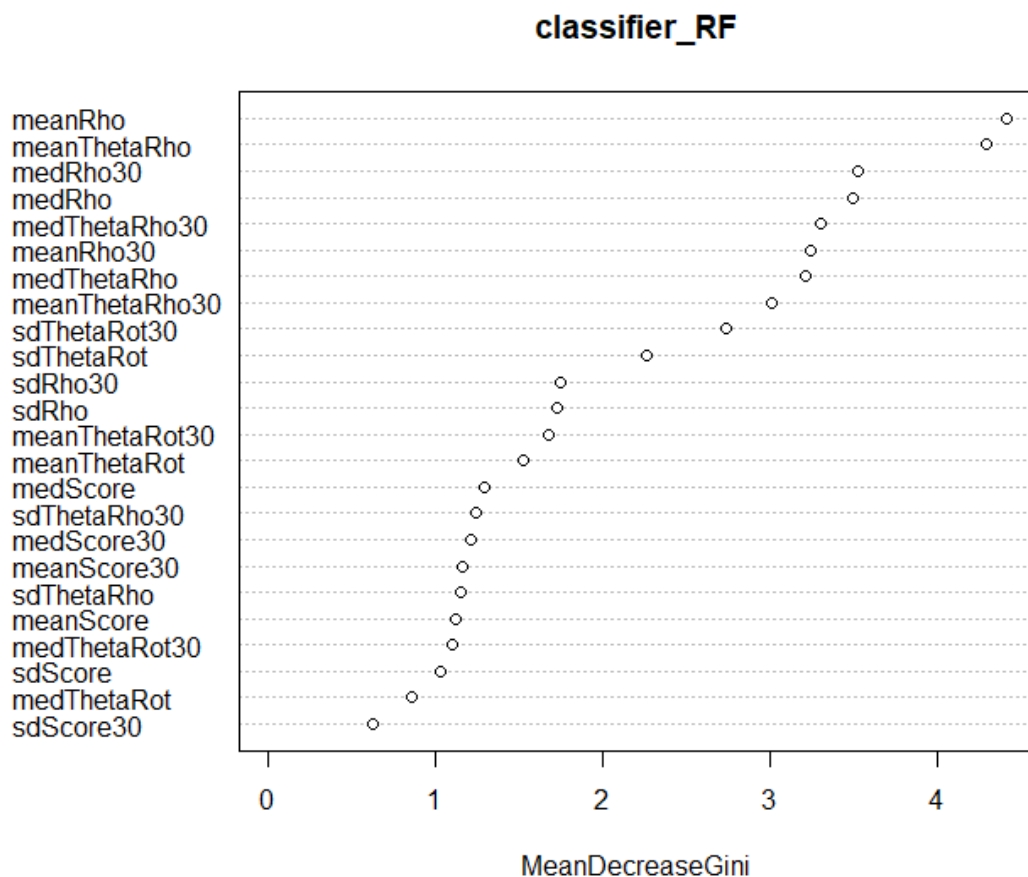
	<b>bk</b>	<b>sm</b>	<b>chyba</b>
<b>bk</b>	32	12	27,2 %
<b>sm</b>	4	59	6,34 %
<b>Odhad celkové chybovosti</b>			<b>14,95 %</b>

Tab. 7. Matice záměn – hodnocení klasifikátoru na zkušebním vzorku.

	<b>bk</b>	<b>sm</b>	<b>chyba</b>
<b>bk</b>	16	3	16 %
<b>sm</b>	3	26	10 %
<b>Celková chybovost</b>			<b>13 %</b>

#### 5.2.4 Parametry modelu Random forest

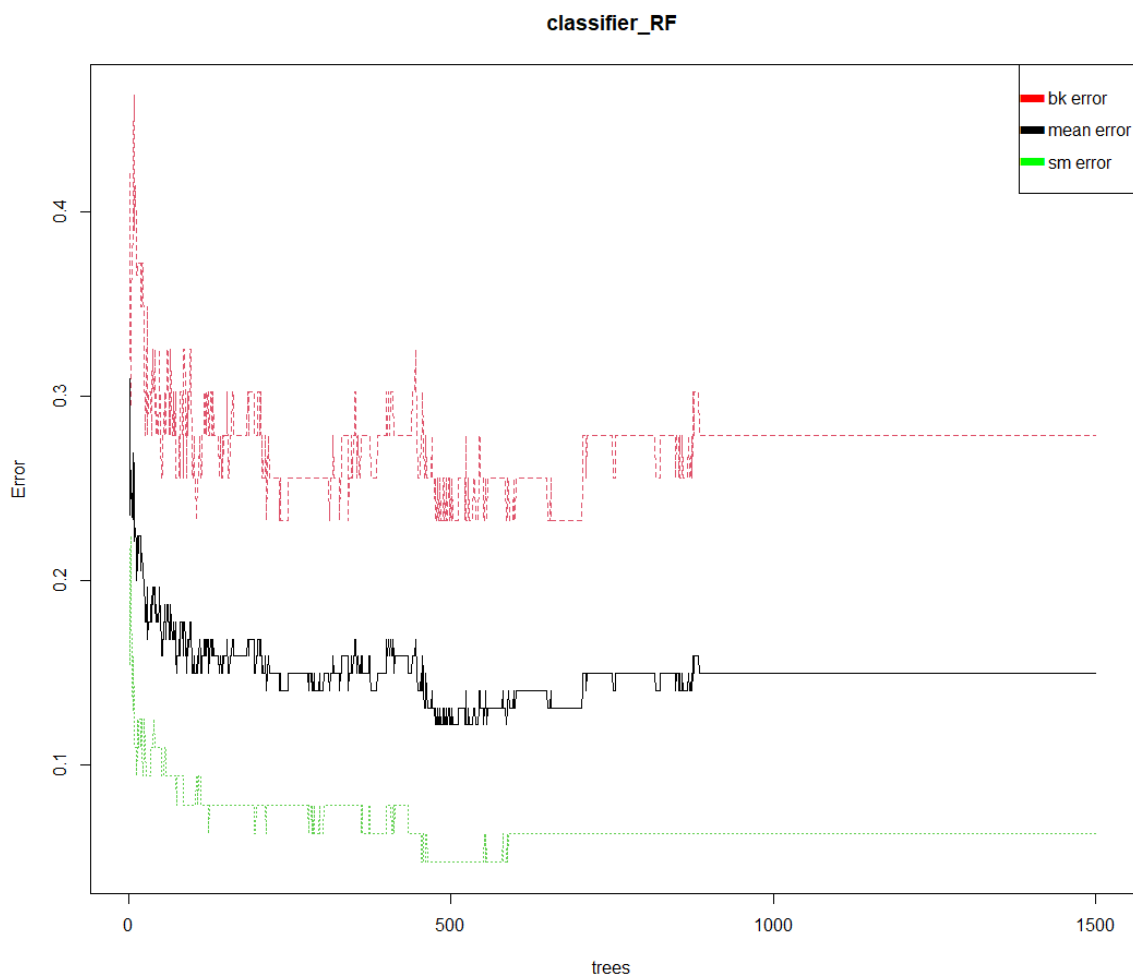
Stejně jako při ověřování metody RF v rámci výpočetní metody, pro ověření a popis fungování modely byl využit graf chybovosti klasifikace jednotlivých tříd a graf Giniho koeficientu pro jednotlivé klasifikační proměnné. V tomto případě však nedošlo k výrazné separaci určitých proměnných z hlediska Giniho koeficientu, nicméně i zde byly rozdíly patrné, jak je patrné z Obr. 14. Jako věcnější se ukázalo použití průměru oproti mediánům a směrodatným odchylkám. Stejně tak se jako vhodnější projevilo užití kompletního vzorku oproti aplikaci datového rámce s hodnotou skóre nacházejícím se v horním 0,3 kvantilu. Nutno však dodat, že v tomto ohledu se jednalo jen o drobné rozdíly. Obecně lze také konstatovat, že nejvýznamnějším užitým parametrem se zdál být průměr v kombinaci se vzdáleností *Rho*, naopak méně prospěšné se zdály proměnné postavené na *Score* či transformovaném úhlu *ThetaRot*.



Obr. 14. Graf hodnot Giniho koeficientu značícího podíl parametrů na klas. model.

Ve snaze vylepšit celkovou přesnost klasifikace a odstranit možný efekt přílišného počtu klasifikačních proměnných bylo vybráno 9 prvků s nejvyšší hodnou Giniho koeficientu a opětovně spuštěna klasifikace. Tímto postupem však nebylo dosaženo razantně lepších výsledků, konkrétně přesnost klasifikace dosahovala **85,65 %** se směrodatnou odchylkou **3 %**. Model byl tedy zjednodušen, avšak pouhým odstraněním nechtěných proměnných nebyl proces zásadně vylepšen. Je otázkou, zda by byl rozdíl statisticky prokázán, avšak v rámci této práce bylo od takových snah upuštěno vzhledem k naznačené nízké rozdílnosti výsledků a celkově mizivému přínosu.

Odlišný se ukázal v postupu obrazové klasifikace proti čistě početním metodám také graf chybovosti modelu, který poukázal na jasnou rozdílnost přesnosti klasifikace buku a smrku. Zatímco jehličnan byl v tomto ohledu převážně spolehlivý, výsledky klasifikace listnatého druhu byly zatíženy značnou chybou, což způsobovalo celkovou průměrnou nepřesnost modelu. Distribuce chybovosti je znázorněna na Obr. 14.



Obr. 15. Graf závislosti počtu klasifikačních stromů na ose  $X$  a chybovosti na ose  $Y$  pro jednotlivé třídy a jejich průměr. Zahrnuto je všech 24 klasifikačních proměnných.

Co se týče vlastních nastavení modelu Random forest, byl v tomto případě použit větší počet rozhodovacích stromů, konkrétní 1 500, což ale nevedlo k rozpoznatelnému zpomalení klasifikačního procesu. Ostatní nastavení velikosti uzlu a počtu testovaných veličin zůstaly shodné s modelem použitým při výpočetní klasifikaci. Při různých konfiguracích neprojevily vliv na celkovou přesnost ani rychlost, což potvrzuje předcházející předpoklad většího vlivu při klasifikaci značně rozměrnějších souborů dat.

## 6 Diskuse

Prezentované výsledky prokázaly použitelnost obou vytyčených metod při klasifikaci druhu dřeviny z laserových bodových mračen pomocí umělé inteligenci, ale také poukázaly na možný prostor pro zlepšení při návrhu klasifikačních modelů. Zarážející je ale rozdílnost dosažené přesnosti, cílem diskuse bude tedy snaha o porovnání s výsledky výzkumů jiných autorů, ale také pokus o objasnění původu chyb i okolností působících zásadní rozdíly chybovosti modelů.

### 6.1 Hodnocení modelu matematicko-statistické klasifikace

Klasifikační metoda výpočtu popisných proměnných přímo z dat bodového mračen poskytl výjimečně dobré výsledky s dosaženou přesností až 100 %, což je úctyhodné, ale také svým způsobem vyvolává otázky. Při porovnání s podobnou prací, kterou prezentovali Lin a Hyypä (2016), a na které se princip tohoto zkoumání částečně zakládá, bylo dosaženo vyšší přesnosti. Citovaný výzkum prokázal celkovou přesnost klasifikace 80 % při aplikaci příbuzné SVM metody strojového učení při využití podobných proměnných, nicméně je nutné zmínit, že byla v tomto případě klasifikována čtveřice druhů dřevin, a to dvou listnatých a dvou jehličnatých. Důvodem odlišnosti výsledků může být tedy větší rozdílnost mezi dvěma druhy dřevin zkoumaných v této práci, jelikož se jedná o jeden jehličnan a jeden listnáč. Celková dosažitelná přesnost metody ale při kombinaci nejspolehlivějších proměnných z obsáhlé testované sady dokáže vylepšit výslednou přesnost, a to ze zmíněných 80 % až na 92 % (Lin a Hyypä 2016), což značí prostor pro zlepšení při výzkumu této problematiky. Otázkou je také vhodnost modelu Random forest jako takového, předchozí výzkum již naznačil kvalitnější výstupy za použití alternativní metody postavené na Bayesově regresi, a to konkrétně přesnosti 74 % RF oproti 81 % dosažené modelem Bayesian multinomial logistic regression (BMLR) při klasifikaci 3 dřevin.

Pozoruhodná je také identifikovaná proměnná nejvíce přispívající přesnosti klasifikace v rámci této metody. Navzdory původním osobním předpokladům se interval intenzit nejvíce specifických pro danou dřevinu nachází v posledním výškovém kvantilu 0,8-1,0. Důvodem může být jednak větší rozdílnost korun a morfologie obou druhů, ale také mohlo hrát určitou roli také brzké jarní období, při kterém snímkování probíhalo. Na začátku vegetačního období totiž buk ještě nedisponuje olistěním, což může určitým způsobem ovlivnit distribuci intenzit odrazů zejména v korunách, kde se nachází nejvíce listové plochy. Z tohoto důvodu není ani

možné porovnání rozdílnosti s výše zmíněnou prací, jelikož snímkování zde probíhalo dokonce v období zimním. Tento jev by bylo vhodné dále prověřit prostřednictvím dalšího výzkumu.

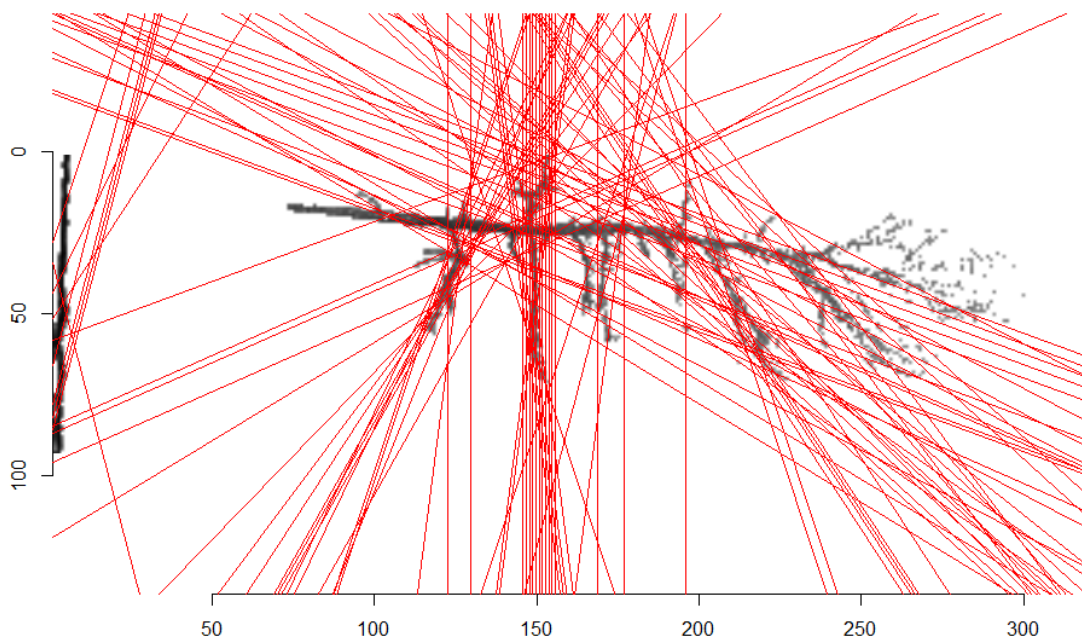
Vhodné by bylo také zařazení většího počtu zkoumaných ploch včetně větší palety zkoumaných dřevin, aby došlo k silnější obhajobě aplikace metody v rámci praktické lesní inventarizace. Ačkoliv v rámci této práce nebylo zapotřebí dalšího zpřesňování klasifikace, po zařazení dalších dřevinných druhů je možné a pravděpodobné, že by chyba klasifikace vzrostla. V takovém případě by bylo na místě zařazení a otestování dalších proměnných postavených na matematicko-statistickém popisu intenzit odrazů, ale třeba také prostorové distribuce bodů v rámci bodového mračna. Celkový potenciál metody je nicméně dostatečně reflektován.

## **6.2 Hodnocení modelu obrazové klasifikace**

Větší paletu otázek vzbuzuje oproti první metodě bezesporu klasifikace na obrazovém základu. Při vypracování této práce nebyl autorovi znám publikovaný výzkum klasifikace druhů dřevin na základě popisu objektů pomocí Houghovy transformace s následnou klasifikací prostřednictvím modelu Random forest. Nicméně v konečném důsledku se jedná taktéž o početní metodu, i přestože je postavena na obrazovém základu. Nabízí se tedy porovnání s výsledky čistě početní klasifikace, ale také s výsledky obrazové klasifikace pomocí neuronových sítí, protože se zabývá obdobnou problematikou.

V rámci této metody byl relativně velmi přesně rozpoznán smrk, přičemž bukové vzorky vykazovaly mnohem častěji chybné určení druhu (viz. Obr. 15.). Jedním z důvodů může být neuniverzálnost nastavení Houghovy transformace, která na bukových vzorcích ne zcela ideálně kopírovala charakteristické větve a výsledný obraz obsahoval častěji nespecifických linií. Důvodem může být například nejednotnost a větvení kmene v horní části stromu, což je jev, který se u smrkových snímků v případě zkoumaných ploch nevyskytoval (viz. Obr. 11. a 12). Pro optimalizaci klasifikace by bylo proto zřejmě vhodné alternativní nastavení generování a popisu linií, nebo například sestavení dalších proměnných s úhlovou i vzdálenostní specifikací podél celého kmene. Předpokládaným zlepšením fungování by také mohlo být dosaženo sestavením dalších proměnných lépe vystihujících úhel nasedání větví na strom, což je očividný optický rozdíl, který bohužel nebyl tímto postupem dostatečně reflektován. S jistotou je zde stále ukrytý nevyužitý potenciál této metody.

Další možný zádrhel spočívá v křivosti bukových kmenů, které se v několika případech v oříznutém datovém pruhu objevily jen částečně či vůbec, což mohlo mít určitý vliv na následné umístění linií, jak je naznačeno na Obr. 16. Možné řešení lze předpokládat v rozšíření rozměru ořezu snímků, případně v důslednější přípravě a čištění bodových mračen. Odchylku mohl způsobit totiž také šum v těsně nadzemních patrech. Tento efekt by bylo možné odstranit odříznutím spodních patret do výše například 1 metru, jelikož ta nejsou v rámci testovaných metod pro klasifikaci rozhodující. Separace přízemního patra se zdá vhodnější, jelikož ruční úprava dat a čištění je velmi časově náročné a je v rozporu s filosofií automatizace procesů.



*Obr. 16. Nezahrnutí celého kmenu do datového výřezu, které se projevilo na výstupu.*

Výsledná přesnost klasifikace okolo 85 % je nicméně úctyhodná, a to včetně vyloučení vlivu absence olistění, která se zde neprojevuje jako charakteristická vlastnost. Dokonce je možné, že větve bez listů nelze popsat tak spolehlivě, pokud jsou ve vzorku společně se smrky. Tuto hypotézu by bylo opět vhodné prověřit dalším výzkumem. Pokud by se podařilo lépe specifikovat také listnaté dřeviny, jedná se jistě o perspektivní metodu, kterou je navíc možné také kombinovat s faktory získanými z výpočetního popisu. Výběr nejvýznamnějších proměnných z odlišných souborů totiž prokazatelně dokáže zlepšit výslednou klasifikaci (Lin a Hyypä 2016).

Zůstává však otázkou, zda tento přístup dokáže konkurovat v rámci obrazové klasifikace druhů dřevin aplikacím neuronových a konvolučních sítí, což je běžně se vyskytující perspektivní metoda klasifikace obrazových dat při vyhodnocování charakteristik vegetace (Kattenborn et al. 2021). Již publikované výsledky klasifikace při použití konvolučních sítí naznačují dosažitelnou přesnost až 91 %, avšak je nutné splnit podmínku obsáhlého vzorku trénovacích dat, aby metoda správně fungovala (Ko et al. 2018). Srovnatelnost kvality výstupů potvrzuje také již zmíněná publikace výpočetních metod a SVM klasifikátoru s dosaženou přesností až 92 % (Lin a Hyypä 2016). Širší pojetí výsledků pro porovnání nabízí kapitola 4.2.

### **6.3 Předpoklady pro integraci do praxe**

Výstupy této práce naznačují konkurenceschopnost postupů postavených na modelech Random forest a pravděpodobně také SVM vzhledem k podobnosti jejich principů a shodu s výstupy dalších autorů. Překážkou může být ale například jistá míra individuality, jakou vyžaduje nastavení klasifikačního modelu pro určitý soubor dat (Lin a Hyypä 2016). Pro přenos do praxe by bylo zřejmě nutné vytvořit obsáhlou databázi s co možná největší kvalitou vstupních dat, která by poskytovala východiska pro aplikaci metody v různorodém lesním prostředí.

Pokud by ale došlo k tvorbě rozsáhlé datové knihovny, tak se nabízí otázka, zda by v konečném důsledku nebylo efektivnější provést klasifikaci pomocí neuronových sítí, které se v praxi již osvědčili i v jiných odvětvích (Kumar 2017). Vycházím při této úvaze čistě z logiky stanovených trendů v odvětví strojového učení a klasifikace dat. Jelikož se nepovažuji za odborníka na tuto problematiku, do jisté míry bych se spoléhal na závěry výzkumníků aplikace metod strojového učení, kteří v poslední době stále více pozornosti věnují právě neuronovým sítím z důvodu většího potenciálu novějších technologií (Kattenborn et al. 2021). Pro potvrzení této hypotézy by bylo vhodné provést komparační výzkum obou metod na základě objemnějšího vzorku porostů v diverzitou druhů dřevin v testované databázi. Tím ale není nutně výzkum metod jednodušších metod strojového učení zcela odepsán. Neuronové sítě lze považovat za komplexnější klasifikační schéma vhodné pro definitivní aplikace, zatímco modely jako Random forest mohou být nejučelnějším řešením jednorázových či výzkumných úkolů. Jak totiž dokazuje i tato práce, první klasifikační model strojového učení prokázal excelentní výsledky při minimální komplexitě řešení.



## 7 Závěr

Při zpracování této diplomové práce došlo k ověření dvou metod automatické klasifikace druhu dřevin z bodových mračen za pomoci strojového učení se slibnými výsledky. Na jedné straně se zde jednoznačně projevil potenciál a možná budoucí aplikovatelnost inovativních postupů a využitelnost ručních mobilních skenerů při analýze lesních porostů či přímo při jejich inventarizaci. Druhý pohled ovšem ukázal stále částečně omezené poznání těchto postupů, které praktické nasazení zásadně omezuje. Ačkoliv literární rešerše poukazuje na celou řadu příkladů poměrně zdařilých aplikací, pro skutečnou funkčnost celistvých inventarizačních postupů zde obor stále postrádá konzistentnost výsledků a jasně identifikovatelný vývojový směr, který zahrnuje předpoklad úspěšného užití v reálném praktickém prostředí.

V rámci této práce byly představeny jednak konkrétní klasifikační modely implementované v jednoduchém a dostupném programovacím jazyce R spolu s nadějnými výsledky a z nich vyplývajícími závěry pro další studium dané problematiky. Dva prezentované modely prokázaly schopnost klasifikovat dva druhy dřevin (smrk a buk) s přesností až 85 % a 100 %, což jsou vynikající předpoklady pro další vývoj metod. Ověřit by bylo vhodné zejména schopnost rozpoznání většího počtu druhů dřevin, ale také hlouběji zkoumat problematiku tvorby a selekce klasifikačních proměnných. Částečně omezujícím se ukázal také soubor dat o 154 použitelných vzorcích stromů, který by mohl být z důvodu snahy o konzistenci výsledků rozměrnější. Další výzva spočívá v automatizaci časově poměrně náročné přípravy, ořezu a čištění dat, což je téma samo o sobě složité a pro praktickou implementaci neméně potřebné.

Jistou překážkou ve snaze zavést postupy do reálného provozního prostředí je ale stále finanční stránka pořízení potřebných prostředků i náročný proces vývoje potřebných aplikací a počítačových programů. Problémem může být ale také nedostatek kvalifikovaných multioborových pracovníků, což mohou pomoci změnit právě takováto témata závěrečných prací na univerzitách, které nejen poskytují určité závěry, ale také vědu popularizují skrze motivaci studentů. Naději vyvolávají značné rozdíly výsledků prací v uplynulých několika letech, zvláště pak v poslední době při aplikacích nejmodernějších senzorů. V následujícím desetiletí by tedy pravděpodobně bylo možné očekávat představení konkrétních, komplexních a funkčních aplikací vzhledem k tempu, jakým celý obor DPZ v poslední době postupuje.

## 8 Seznam použitých zdrojů

- ALWOSHEEL, Ahmad, Sander Van CRANENBURGH a Caspar G CHORUS, 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling* [online]. **28**(July), 167–182. ISSN 1755-5345. Dostupné z: doi:10.1016/j.jocm.2018.07.002
- ANON., 2016. Fitting a Neural Network in R; neuralnet package [graf]. *Analytics & IIoT* [online]. Dostupné z: <https://analyticks.wordpress.com/2016/09/22/fitting-a-neural-network-in-r-neuralnet-package/>
- AXELSSON, Arvid, Eva LINDBERG a Håkan OLSSON, 2018. Exploring Multispectral ALS Data for Tree Species Classification. *Remote Sensing* [online]. **10**(183). Dostupné z: doi:10.3390/rs10020183
- BAILEY, Tim a Hugh DURRANT-WHYTE, 2006. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics and Automation Magazine* [online]. **13**(3), 108–117. ISSN 10709932. Dostupné z: doi:10.1109/MRA.2006.1678144
- BALENOVIĆ, Ivan, Xinlian LIANG, Luka JURJEVIĆ, Juha HYYPPÄ, Ante SELETKOVIĆ a Antero KUKKO, 2020. Hand-held personal laser scanning – current status and perspectives for forest inventory application. *Croatian Journal of Forest Engineering* [online]. **42**(1), 165–183. ISSN 18455719. Dostupné z: doi:10.5552/crojfe.2021.858
- BARTHELME, Simon, David TSCHUMPERLE, Jan WIJFFELS, Haz ASSEMLAL a Shota OCHI, 2022. *Image Processing Library Based on „CImg”* [online]. 2022. Dostupné z: <http://dahtah.github.io/imager/>
- BATTA, Mahesh, 2020. Machine Learning Algorithms - A Review. *International Journal of Science and Research* [online]. **9**(1), 381–387. Dostupné z: doi:10.21275/ART20203995
- BIAU, Gérard a Erwan SCORNET, 2016. A random forest guided tour. *TEST* [online]. **25**(2), 197–227. ISSN 1863-8260. Dostupné z: doi:10.1007/s11749-016-0481-7
- BREIMAN, Leo, Adele CUTLER, Andy LIAW a Matthew WIENER, 2022. *Breiman and Cutler's Random Forests for Classification and Regression* [online]. 2022. Dostupné z: <https://www.stat.berkeley.edu/~breiman/RandomForests/>
- CHEN, Shilin, Haiyang LIU, Zhongke FENG, Chaoyong SHEN a Panpan CHEN, 2019. Applicability of personal laser scanning in forestry inventory. *PLoS ONE* [online]. **14**(2), 1–22. ISSN 19326203. Dostupné z: doi:10.1371/journal.pone.0211392
- CORTE, Ana Paula Dalla, Deivison Venicio SOUZA, Franciel Eduardo REX, Carlos Roberto SANQUETTA, Midhun MOHAN, Carlos Alberto SILVA, Angelica Maria Almeyda ZAMBRANO, Gabriel PRATA, Danilo Roberti ALVES DE ALMEIDA, Jonathan William TRAUTENMÜLLER, Carine KLAUBERG, Anibal DE MORAES, Mateus N. SANQUETTA, Ben WILKINSON a Eben North BROADBENT, 2020. Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes. *Computers and Electronics in Agriculture* [online]. **179**(September), 105815. ISSN 01681699. Dostupné z: doi:10.1016/j.compag.2020.105815

- DAINELLI, Riccardo, Piero TOSCANO, Salvatore Filippo DI GENNARO a Alessandro MATESE, 2021. Recent advances in unmanned aerial vehicles forest remote sensing—a systematic review. Part ii: Research applications. *Forests* [online]. **12**(4). ISSN 19994907. Dostupné z: doi:10.3390/f12040397
- DREISEITL, Stephan a Lucila OHNO-MACHADO, 2002. Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics* [online]. **35**(5–6), 352–359. ISSN 15320464. Dostupné z: doi:10.1016/S1532-0464(03)00034-0
- ERMAN, Jeffrey, Anirban MAHANTI a Martin ARLITT, 2006. Traffic Classification Using Clustering Algorithms. In: *The Scope for Online Social Network Aided Caching in Web CDNs View project Cyber Risk Management View project Traffic Classification Using Clustering Algorithms* [online]. [vid. 2022-03-22]. Dostupné z: doi:10.1145/1162678.1162679
- FANKHAUSER, Kathryn E., Nikolay S. STRIGUL a Demetrios GATZIOLIS, 2018. Augmentation of traditional forest inventory and Airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring. *Remote Sensing* [online]. **10**(10), 1–17. ISSN 20724292. Dostupné z: doi:10.3390/rs10101562
- FAUVEL, Mathieu, 2008. Spectral and Spatial Methods for the Classification of To cite this version : HAL Id : tel-00258717 Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data [online]. B.m. Institut National Polytechnique de Grenoble-INPG; Université d'Islande. Dostupné z: https://tel.archives-ouvertes.fr/tel-00258717
- GARCÍA-GONZALO, Esperanza, Zulima FERNÁNDEZ-MUÑIZ, Paulino José García NIETO, Antonio Bernardo SÁNCHEZ a Marta Menéndez FERNÁNDEZ, 2016. Hard-rock stability analysis for span design in entry-type excavations with learning classifiers. *Materials* [online]. **9**(7), 1–19. ISSN 19961944. Dostupné z: doi:10.3390/ma9070531
- GENUER, Robin, Jean Michel POGGI, Christine TULEAU-MALOT a Nathalie VILLA-VIALANEIX, 2017. Random Forests for Big Data. *Big Data Research* [online]. **9**, 28–46. ISSN 22145796. Dostupné z: doi:10.1016/j.bdr.2017.07.003
- GEOSLAM LTD., 2021. *GeoSLAM Hub* [online]. 2021. Dostupné z: https://geoslam.com/hub/
- GOLLOB, Christoph, Tim RITTER a Arne NOTHDURFT, 2020. Forest inventory with long range and high-speed Personal Laser Scanning (PLS) and Simultaneous Localization and Mapping (SLAM) technology. *Remote Sensing* [online]. **12**(9). ISSN 20724292. Dostupné z: doi:10.3390/RS12091509
- GOLZE, Jens, Stefania ZOURLIDOU a Monika SESTER, 2020. Traffic Regulator Detection Using GPS Trajectories. *KN - Journal of Cartography and Geographic Information* [online]. **70**(3), 95–105. ISSN 25244965. Dostupné z: doi:10.1007/s42489-020-00048-x
- GONG, Destin, 2022. *Top Machine Learning Algorithms for Classification* [online]. Dostupné z: https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501

GONZÁLEZ-JORGE, H., B. RIVEIRO, P. ARIAS a J. ARMESTO, 2012. Photogrammetry and laser scanner technology applied to length measurements in car testing laboratories. *Measurement: Journal of the International Measurement Confederation* [online]. **45**(3), 354–363. ISSN 02632241. Dostupné z: doi:10.1016/j.measurement.2011.11.010  
GPL SOFTWARE, 2022. *CloudCompare* [online]. 2022. Dostupné z: <http://www.cloudcompare.org/>

HASTINGS, Jack H., Scott V. OLLINGER, Andrew P. OUIMETTE, Rebecca SANDERS-DEMOTT, Michael W. PALACE, Mark J. DUCEY, Franklin B. SULLIVAN, David BASLER a David A. ORWIG, 2020. Tree Species Traits Determine the Success of LiDAR-Based Crown Mapping in a Mixed Temperate Forest. *Remote Sensing* [online]. **12**(2), 309 [vid. 2022-02-16]. ISSN 20724292. Dostupné z: doi:10.3390/RS12020309

HUNČAGA, Milan, Juliána CHUDÁ, Julián TOMAŠTÍK, Martina SLÁMOVÁ, Milan KOREŇ a František CHUDÝ, 2020. The comparison of stem curve accuracy determined from point clouds acquired by different terrestrial remote sensing methods. *Remote Sensing* [online]. **12**(17). ISSN 20724292. Dostupné z: doi:10.3390/RS12172739

HYYPPÄ, Eric, Xiaowei YU, Harri KAARTINEN, Teemu HAKALA, Antero KUKKO, Mikko VASTARANTA a Juha HYYPPÄ, 2020. Comparison of Backpack, Handheld, Under-Canopy UAV, and Above-Canopy UAV Laser Scanning for Field Reference Data Collection in Boreal Forests. *Remote Sensing* [online]. **12**(20), 3327 [vid. 2022-02-14]. ISSN 20724292. Dostupné z: doi:10.3390/RS12203327

IBM CLOUD EDUCATION, 2022. *What is random forest?* [online]. Dostupné z: <https://www.ibm.com/cloud/learn/random-forest>

JIN, Shichao, Yanjun SU, Shang GAO, Fangfang WU, Tianyu HU a Jin LIU, 2018. Deep Learning : Individual Maize Segmentation From Terrestrial Lidar Data Using Faster R-CNN and Regional Growth Algorithms. *Frontiers in Plant Science* [online]. **9**(June), 1–10. Dostupné z: doi:10.3389/fpls.2018.00866

JOKELA, Maria, Pasi PYYKONEN, Matti KUTILA a Kimmo KAUVUO, 2019. LiDAR Performance Review in Arctic Conditions. In: *Proceedings - 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing, ICCP 2019* [online]. s. 27–31. ISBN 9781728149141. Dostupné z: doi:10.1109/ICCP48234.2019.8959554

JUPP, David L B a Jenny L LOVELL, 2007. *Airborne and Ground-Based Lidar Systems for Forest Measurement : Background and Principles*. ISBN 9781921232688.

JURJEVIĆ, Luka, Xinlian LIANG, Mateo GAŠPAROVIĆ a Ivan BALENOVIĆ, 2020. Is field-measured tree height as reliable as believed – Part II, A comparison study of tree height estimates from conventional field measurement and low-cost close-range remote sensing in a deciduous forest. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **169**(April), 227–241. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2020.09.014

KATTENBORN, Teja, Jens LEITLOFF, Felix SCHIEFER a Stefan HINZ, 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **173**(January), 24–49. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2020.12.010

- KESTUR, Ramesh, Akanksha ANGURAL, Bazila BASHIR, S. N. OMKAR, Gautham ANAND a M. B. MEENAVATHI, 2018. Tree Crown Detection, Delineation and Counting in UAV Remote Sensed Images: A Neural Network Based Spectral–Spatial Method. *Journal of the Indian Society of Remote Sensing* [online]. **46**(6), 991–1004. ISSN 09743006. Dostupné z: doi:10.1007/s12524-018-0756-4
- KO, C., J. KANG a G. SOHN, 2018. DEEP MULTI-TASK LEARNING for TREE GENERA CLASSIFICATION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* [online]. **4**(2), 153–159. ISSN 21949050. Dostupné z: doi:10.5194/isprs-annals-IV-2-153-2018
- KO, Chiung, Seunghyun LEE, Jongsu YIM, Donggeun KIM a Jintaek KANG, 2021. Comparison of Forest Inventory Methods at Plot-Level between a Backpack Personal Laser Scanning (BPLS) and Conventional Equipment in Jeju Island, South Korea [online]. Dostupné z: doi:10.3390/f12030308
- KORPELA, Ilkka, Hans Ole ØRKA, Matti MALTAMO, Timo TOKOLA a Juha HYYPPÄ, 2010. LiDAR – Effects of Stand and Tree Parameters , Downsizing of Training Set , Intensity Normalization , and Sensor Type. *Silva Fennica*. **44**(1432), 319–339.
- KOTSIANTIS, S. B., 2007. Supervised Machine Learning: A Review of Classification Techniques. *Informatica*. **31**(3), 294–268.
- KRUPPA, Jochen, Alexandra SCHWARZ, Gerhard ARMINGER a Andreas ZIEGLER, 2013. Expert Systems with Applications Consumer credit risk : Individual probability estimates using machine learning. *Expert Systems With Applications* [online]. **40**(13), 5125–5131. ISSN 0957-4174. Dostupné z: doi:10.1016/j.eswa.2013.03.019
- KUMAR, Sathiesh, 2017. Flower Species Recognition System using Convolution Neural Networks and Transfer Learning. In: *4th International Conference on Signal Processing, Communications and Networking* [online]. ISBN 9781509047406. Dostupné z: doi:10.1109/ICSCN.2017.8085675
- KUŽELKA, Karel, Martin SLAVÍK a Peter SUROVÝ, 2020. Very high density point clouds from UAV laser scanning for automatic tree stem detection and direct diameter measurement. *Remote Sensing* [online]. **12**(8). ISSN 20724292. Dostupné z: doi:10.3390/RS12081236
- LARJAVAARA, Markku a Helene C MULLER-LANDAU, 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. *Methods in Ecology and Evolution* [online]. **4**, 793–801. Dostupné z: doi:10.1111/2041-210X.12071
- LEBERL, F, A IRSCHARA, T POCK, P MEIXNER, M GRUBER, S SCHOLZ a A WIECHERT, 2010. Point Clouds : Lidar versus 3D Vision. *Photogrammetric Engineering & Remote Sensing*. **76**(10), 1123–1134.
- LECUN, Yann, Yoshua BENGIO a Geoffrey HINTON, 2015. Deep learning. *Nature* [online]. **521**, 436–445. Dostupné z: doi:10.1038/nature14539

- LEVICK, Shaun R., Tim WHITESIDE, David A. LOEWENSTEINER, Mitchel RUDGE a Renee BARTOLO, 2021. Leveraging tls as a calibration and validation tool for mls and uls mapping of savanna structure and biomass at landscape-scales. *Remote Sensing* [online]. **13**(2), 1–19. ISSN 20724292. Dostupné z: doi:10.3390/rs13020257
- LI, Linyuan, Jun CHEN, Xihan MU, Weihua LI, Guangjian YAN, Donghui XIE a Wuming ZHANG, 2020. Quantifying understory and overstory vegetation cover using UAV-based RGB imagery in forest plantation. *Remote Sensing* [online]. **12**(2), 1–18. ISSN 20724292. Dostupné z: doi:10.3390/rs12020298
- LIANG, Xinlian, Juha HYYPPÄ, Harri KAARTINEN, Matti LEHTOMÄKI, Jiri PYÖRÄLÄ, Norbert PFEIFER, Markus HOLOPAINEN, Gábor BROLLY, Pirotti FRANCESCO, Jan HACKENBERG, Huabing HUANG, Hyun Woo JO, Masato KATOH, Luxia LIU, Martin MOKROŠ, Jules MOREL, Kenneth OLOFSSON, Jose POVEDA-LOPEZ, Jan TROCHTA, Di WANG, Jinhua WANG, Zhouxi XI, Bisheng YANG, Guang ZHENG, Ville KANKARE, Ville LUOMA, Xiaowei YU, Liang CHEN, Mikko VASTARANTA, Ninni SAARINEN a Yunsheng WANG, 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **144**(June), 137–179. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2018.06.021
- LIN, Yi-Chun, Raja MANISH, Darcy BULLOCK a Ayman HABIB, 2021. Comparative Analysis of Different Mobile LiDAR Mapping Systems for Ditch Line Characterization. *Remote Sensing* [online]. **13**(2485). Dostupné z: doi:10.3390/rs13132485
- LIN, Yi a Juha HYYPPÄ, 2016. A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *International Journal of Applied Earth Observation and Geoinformation* [online]. **46**, 45–55. ISSN 1872826X. Dostupné z: doi:10.1016/j.jag.2015.11.010
- MARSEL, Vagizov R., Istomin P. EUGENIE, Miheev L. VALERIE, Potapov P. ARTEM a Yagotinceva V. NATALYA, 2021. Visual digital forest model based on a remote sensing data and forest inventory data. *Remote Sensing* [online]. **13**(20). ISSN 20724292. Dostupné z: doi:10.3390/rs13204092
- MATHWORKS, 2022. k-means clustering - MATLAB kmeans. *Help Center Documentation* [online] [vid. 2022-03-22]. Dostupné z: <https://www.mathworks.com/help/stats/kmeans.html>
- MICHAŁOWSKA, Maja a Jacek RAPIŃSKI, 2021. A review of tree species classification based on airborne lidar data and applied classifiers. *Remote Sensing* [online]. **13**(3), 1–27. ISSN 20724292. Dostupné z: doi:10.3390/rs13030353
- MIKITA, Tomáš, Přemysl JANATA, Peter SUROVÝ, Juha HYYPPÄ, Xinlian LIANG a Eetu PUTTONEN, 2016. Forest Stand Inventory Based on Combined Aerial and Terrestrial Close-Range Photogrammetry. *Forests* [online]. **7**(165). Dostupné z: doi:10.3390/f7080165
- MUJA, Marius a David G. LOWE, 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: *VISAPP 2009 - Proceedings of the 4th International Conference on Computer Vision Theory and Applications* [online]. s. 331–340. ISBN 9789898111692. Dostupné z: doi:10.5220/0001787803310340

- NÄSI, Roope, Eija HONKAVAARA, Päivi LYYTIKÄINEN-SAARENMAA, Minna BLOMQVIST, Paula LITKEY, Teemu HAKALA, Niko VILJANEN, Tuula KANTOLA, Topi TANHUANPÄÄ a Markus HOLOPAINEN, 2015. Using UAV-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level. *Remote Sensing* [online]. **7**(11), 15467–15493. ISSN 20724292. Dostupné z: doi:10.3390/RS71115467
- NI, Wenjian, Guoqing SUN, Yong PANG, Zhiyu ZHANG, Jianli LIU, Aqiang YANG, Yao WANG a Dafeng ZHANG, 2018. Mapping Three-Dimensional Structures of Forest Canopy Using UAV Stereo Imagery: Evaluating Impacts of Forward Overlaps and Image Resolutions with LiDAR Data as Reference. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* [online]. **11**(10), 3578–3589. ISSN 21511535. Dostupné z: doi:10.1109/JSTARS.2018.2867945
- NOBLE, William S a Pacific STREET, 2006. What is a support vector machine ? *Computational Biology*. **24**(12), 1565–1567.
- OSISANWO, F.Y., J.E.T. AKINSOLA, O. AWODELE, J.O. HINMIKAIYE, O. OLAKANMI a J. AKINJOBI, 2017. Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology* [online]. **48**(3), 128–138. Dostupné z: doi:10.14445/22312803/ijctt-v48p126
- PIERMATTEI, Livia, Wilfried KAREL, Di WANG, Martin WIESER, Martin MOKROŠ, Peter SUROVÝ, Milan KOREŇ, Julián TOMAŠTÍK, Norbert PFEIFER a Markus HOLLAUS, 2019. Terrestrial Structure from Motion Photogrammetry for Deriving Forest Inventory Data. *Remote Sensing* [online]. **11**(8), 950. Dostupné z: doi:10.3390/rs11080950
- PITKÄNEN, Timo P., Pasi RAUMONEN a Annika KANGAS, 2019. Measuring stem diameters with TLS in boreal forests by complementary fitting procedure. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **147**(March 2018), 294–306. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2018.11.027
- RAJ, Thinal, Fazida Hanim HASHIM, Aqilah Baseri HUDDIN, Mohd Faisal IBRAHIM a Aini HUSSAIN, 2020. A survey on LiDAR scanning mechanisms. *Electronics (Switzerland)* [online]. **9**(5). ISSN 20799292. Dostupné z: doi:10.3390/electronics9050741
- RAMALHO DE OLIVEIRA, Luiz, H ANDREW LASSITER, Ben WILKINSON, Travis WHITLEY, Peter IFJU, Stephen R LOGAN, Gary F PETER, Jason G VOGEL a Timothy A MARTIN, 2020. Moving to Automated Tree Inventory: Comparison of UAS-Derived Lidar and Photogrammetric Data with Manual Ground Estimates. *Remote Sensing* [online]. **13**(72). Dostupné z: doi:10.3390/rs13010072
- RSTUDIO TEAM, 2021. *RStudio: Integrated Development for R* [online]. 2021. Dostupné z: <http://www.rstudio.com/>
- SAHA, Sumit, 2018. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. *Towards Data Science* [online]. Dostupné z: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

- SCHOLKOPF, B., C.J.C. BURGESS a A.J. SMOLA, 1999. *Advances in Kernel Methods: Support Vector Learning*. B.m.: The MIT Press. ISBN 0-262-19416-3.
- SETIONO, Rudy a Wee Kheng LEOW, 2000. FERNN: An algorithm for fast extraction of rules from neural networks. *Applied Intelligence* [online]. **12**(1–2), 15–25. ISSN 0924669X. Dostupné z: doi:10.1023/A:1008307919726
- SHEA, Keiron O a Ryan NASH, 2015. *An Introduction to Convolutional Neural Networks*. 2015.
- SHIMIZU, Katsuto, Tomohiro NISHIZONO, Fumiaki KITAHARA, Keiko FUKUMOTO a Hideki SAITO, 2022. Integrating terrestrial laser scanning and unmanned aerial vehicle photogrammetry to estimate individual tree attributes in managed coniferous forests in Japan. *International Journal of Applied Earth Observation and Geoinformation* [online]. **106**(October 2021), 102658. ISSN 1872826X. Dostupné z: doi:10.1016/j.jag.2021.102658
- SLAVÍK, Martin, Karel KUŽELKA, Roman MODLINGER, Ivana TOMÁŠKOVÁ a Peter SUROVÝ, 2020. Uav laser scans allow detection of morphological changes in tree canopy. *Remote Sensing* [online]. **12**(22), 1–15. ISSN 20724292. Dostupné z: doi:10.3390/rs12223829
- SOFIA, S, S SFERLAZZA, A MARIOTTINI, M NICCOLINI, T COPPI, M MIOZZO, T La MANTIA a F MAETZKE, 2021. A CASE STUDY OF THE APPLICATION OF HAND-HELD MOBILE LASER SCANNING IN THE PLANNING OF AN ITALIAN FOREST ( ALPE DI CATENAIA , TUSCANY ). *Remote Sensing and Spatial Information Sciences*. **XLIII**, 763–770.
- STROBL, Carolin, Anne Laure BOULESTEIX, Achim ZEILEIS a Torsten HOTHORN, 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* [online]. **8**. ISSN 14712105. Dostupné z: doi:10.1186/1471-2105-8-25
- SUROVÝ, Peter a Karel KUŽELKA, 2019. Acquisition of Forest Attributes for Decision Support at the Forest Enterprise Level Using Remote-Sensing Techniques-A Review. *Remote Sensing* [online]. **10**(273). Dostupné z: doi:10.3390/f10030273
- SUROVÝ, Peter a Karel KUŽELKA, 2019. Acquisition of forest attributes for decision support at the forest enterprise level using remote-sensing techniques-a review. *Forests* [online]. **10**(3). ISSN 19994907. Dostupné z: doi:10.3390/f10030273
- TERRY, Louise, Kim CALDERS, Mathias DISNEY, Niall ORIGO, Yadvinder MALHI, Glenn NEWMHAM, Pasi RAUMONEN, Markku Å KERBLÖM a Hans VERBEECK, 2020. Tree species classification using structural features derived from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **168**(February), 170–181. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2020.08.009
- THRUN, Sebastian a Michael MONTEMERLO, 2006. The graph SLAM algorithm with applications to large-scale mapping of urban structures. *International Journal of Robotics Research* [online]. **25**(5–6), 403–429. ISSN 02783649. Dostupné z: doi:10.1177/0278364906065387



TUSZYNSKI, Jarek a Michael DIETZE, 2021. *Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc.* 2021.

URBANEK, Simon, 2013. *Read and write PNG images* [online]. 2013. Dostupné z: <http://www.rforge.net/png/>

ÚSTAV PRO HOSPODÁŘSKOU ÚPRAVU LESŮ a VÝZKUMNÝ ÚSTAV LESNÍHO HOSPODÁŘSTVÍ A MYSLIVOSTI, 2018. *Kůrovcová mapa* [online]. Dostupné z: <https://www.kurovcovamapa.cz/>

VERLINDE, Patrick a Eddy MAYORAZ, 1998. Decision fusion using a multi-linear classifier. In: *International Conference on Multisource-Multisensor Information Fusion*. WALLACE, Luke, Arko LUCIEER, Zbyněk MALENOVSKÝ, Darren TURNER, Petr VOPĚNKA, Juha HYYPPÄ, Xinlian LIANG a Eetu PUTTONEN, 2016. Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds. *Forests* [online]. **7**(62). Dostupné z: [doi:10.3390/f7030062](https://doi.org/10.3390/f7030062)

WARFIELD, Angus D. a Javier X. LEON, 2019. Estimating mangrove forest volume using terrestrial laser scanning and UAV-derived structure-from-motion. *Drones* [online]. **3**(2), 1–17. ISSN 2504446X. Dostupné z: [doi:10.3390/drones3020032](https://doi.org/10.3390/drones3020032)

WHITE, Joanne C., 2012. Digital high spatial resolution aerial imagery to support forest health monitoring: the mountain pine beetle context. *Journal of Applied Remote Sensing* [online]. **6**(1), 062527. ISSN 1931-3195. Dostupné z: [doi:10.1117/1.jrs.6.062527](https://doi.org/10.1117/1.jrs.6.062527)

ZHANG, Zhenyu a Xiaoye LIU, 2013. Support vector machines for tree species identification using LiDAR-derived structure and intensity variables. *Geocarto International* [online]. **6049**. Dostupné z: [doi:10.1080/10106049.2012.710653](https://doi.org/10.1080/10106049.2012.710653)

ZHOU, Tan, Sorin C POPESCU, A Michelle LAWING, Marian ERIKSSON, Bogdan M Strimbu ID a Paul C Bürkner ID, 2018. Bayesian and Classical Machine Learning Methods : A Comparison for Tree Species Classification with LiDAR Waveform Signatures. *Remote Sensing* [online]. **10**(39), 1–27. Dostupné z: [doi:10.3390/rs10010039](https://doi.org/10.3390/rs10010039)