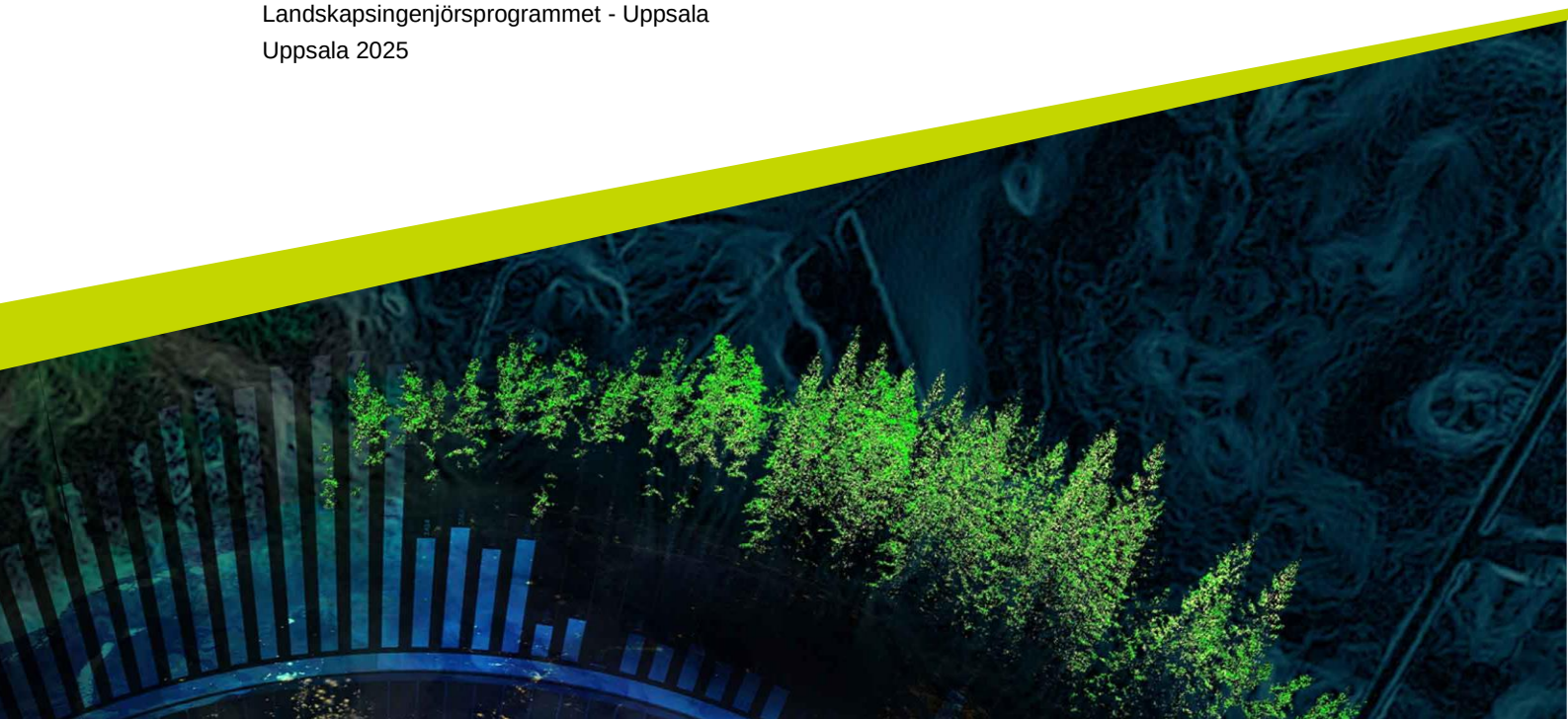
I jakten på jättarna

En studie om automatisk identifiering av jätteträd.

Karl Wanhainen

Examensarbete/Självständigt arbete • 15 hp
Sveriges lantbruksuniversitet, SLU
Fakulteten för naturresurser och jordbruksvetenskap
Landskapsingenjörsprogrammet - Uppsala
Uppsala 2025



I jakten på jättarna. En studie om automatisk identifiering av jätteträd.

In search of the giants: A study on automatic identification of giant trees.

Karl Wanhainen

Handledare:	Roger Elg, Sveriges lantbruksuniversitet, Institutionen för stad och land
Bitr. handledare:	Johan Östberg, Trädkontoret
Examinator:	Göran Thor, Sveriges lantbruksuniversitet, Institutionen för ekologi
Bitr. examinator	Vera Vicenzotti, Sveriges lantbruksuniversitet, Institutionen för stad och land

Omfattning:	15 hp
Nivå och fördjupning:	Grundnivå, G2E
Kurstitel:	Självständigt arbete i landskapsarkitektur
Kurskod:	EX1004
Program/utbildning:	Landskapsingenjörsprogrammet - Uppsala
Kursansvarig inst.:	Institutionen för stad och land
Utgivningsort:	Uppsala
Utgivningsår:	2025
Upphovsrätt:	Alla tabeller är författarens egna
Elektronisk publicering:	https://stud.epsilon.slu.se

Nyckelord: jätteträd, inventering, maskininlärning, stamdiameter, krondiameter, trädkroneavgränsning, segmentering

Sveriges lantbruksuniversitet
Fakulteten för naturresurser och jordbruksvetenskap
Institution för stad och land
Avdelningen för landskapsarkitektur

Sammanfattning

Jätteträd är träd vars stamdiameter är större än 100 centimeter. Dessa träd besitter stora biologiska, kulturhistoriska samt landskapsarkitektoniska värden och det har ofta tagit flera hundra år att nå sin imponerande storlek. Sverige har ett unikt stort bestånd av jätteträd ur en europeisk kontext och därför enligt Naturvårdsverket också ett särskilt ansvar att vårda och bevara dessa. För att göra det måste man veta var träden finns, men enligt Länsstyrelsen finns ett stort mörkertal över antalet jätteträd i Sverige.

I denna rapport har metoder för automatisk identifiering av jätteträd undersökts. För detta krävs en modell som automatiskt kan identifiera och segmentera trädkronor från ortofoton och en modell som kan undersöka sambandet mellan kron- och stamdiameter.

Metoder för individuell trädkroneavgränsning (ITCD) har undersökts genom en narrativ litteraturstudie. Resultatet från litteraturstudien visar att det finns lovande djupinlärningsmodeller för individuell trädkroneavgränsning och segmentering. Den modell med högst precision har tränat på 20 000 individuellt segmenterade träd vilket är ett tidskrävande arbetsmoment.

Däremot saknas modeller för att korrelera krondiameter med stamdiameter. Därför har, som konceptvalidering, en maskininlärningsmodell tagits fram i programspråket Python. Modellen tränades med data från Malmös träddatabas. Resultatet från konceptvalidering visar ett generellt starkt linjärt samband mellan kron- och stamdiameter för flera arter, med R^2 värden upp mot 0,72. Detta samband blir dock mindre tydligt för de grövsta träden för samtliga arter. Därför kan modellen i sin nuvarande form inte användas för att identifiera jätteträd med tillräckligt hög precision.

Vidare studier kan med fördel utföras för att vidareutveckla konceptvalideringsmodellen. Dels för att träna nuvarande modell med mer och bättre data, och dels för att förfina modellen för att kunna påvisa ett tydligare samband mellan kron- och stamdiameter för jätteträd.

Nyckelord: jätteträd, maskininlärning, trädinventering stamdiameter, krondiameter, trädkroneavgränsning,

Abstract

Giant trees are trees whose trunk diameter exceeds 100 centimeter. These trees hold great biological, cultural and landscape architectural values and has often taken several centuries to reach their impressive size. Sweden has an unusual large number of giant trees in a European context and according to Naturvårdsverket also a special responsibility to care for and preserve these. To be able to do that you have to know where the trees are and according to Länstyrelsen there is a large dark figure regarding the number of giant trees in Sweden.

In this report, methods for automatic identification of giant trees has been explored. The method requires two models in order for this to work. One model that automatically can identify and segment individual tree crowns based on orthophotos. Another model that can explore the relation

between trunk size and crown size. Methods for segmenting individual tree crowns has been explored through a literature study.

The results from the literature study shows that there are promising deep learning models for individual crown detection and segmentation. The models showing the highest precision was trained using 20 000 manually segmented trees, a very labor-intensive task.

Models that correlates the relationship between crown- and trunk diameter does not exist however. Therefore, as a concept validation, such a model was developed using the programming language Python. The model was trained using data from Malmö´s tree data base. The results from the concept validation shows a generally strong linear correlation between crown- and trunk diameter for several tree species. With R^2 values reaching up towards 0.72. However, this correlation grows weaker for the trees with the biggest trunk diameter for all species. Therefor the model in its current form can not be used to identify giant trees with a satisfactory precision.

Further studies can be performed to further develop the concept validation model. Partly to train it with more and better data, and partly to refine the model to show a clearer correlation between crown- and trunk diameter for giant trees.

Keywords: giant trees, machine learning, tree inventory, trunk diameter, crown diameter, tree crown segmentation.

Innehållsförteckning

1. Inledning 6
1.1 Vad är jätteträd? 6
1.2 Behovet av inventering 6
1.3 Hur AI och maskininlärning kan hjälpa till i jakten 7

2. Syfte och frågeställning 8

3. Avgränsningar 9

4. Metod 10
4.1 Metod för litteraturstudie 10
4.2 Metod för konceptvalidering 10

5. Teoretisk bakgrund 12
5.1 Svenska jätteträd - kulturella och biologiska värden 12
5.2 Supervised learning (väglett lärande) 13
5.3 Linjär regression 13
5.4 Logistisk regression 13
5.5 Random Forest 14
5.6 SMOTE (Synthetic Minority Over-sampling Technique) 14
5.7 Confusion matrix 15
5.8 ROC-värde & AUC-kurva 15
5.9 Andra parametrar för att utvärdera en modells prestanda 15

6. Resultat – Undersökande studie 17
6.1 Fältinventering 17
6.2 Systematisk fältinventering 17
6.3 Fjärranalys 18
6.4 Användning av maskininlärning 19
6.5 Användning av ortofoton 19
6.6 Användning av LiDAR 20
6.7 Användning av CNN (Convolutional neural networks) 21
6.8 Kombination av olika metoder 22

7. Resultat - Konceptvalideringsmodell 23
7.1 Resultat av konceptvalidering 23
7.1.1 Det linjära sambandet mellan kron- och stamdiameter för 10 utvalda arter 24
7.1.2 Den linjära regressionsmodellens prestanda 29
7.1.3 Logistisk regression resultat 29
7.1.4 Random Forest resultat 30

8. Diskussion 32

9. Konklusion och vidare studier. 36

Referenser 37

10. Tack 40

1. Inledning

Träd är några av de största levande organismerna som finns på vår planet. De allra största finns inte i Sverige, men även i vårt kalla och mörka land kan träd nå en imponerande storlek. Ju större ett träd blir desto större blir dess bidrag i form av ekosystemtjänster och upplevelsevärden. En helt central aspekt i att kunna förvalta våra svenska jättar är att veta var de finns.

1.1 Vad är jätteträd?

Jätteträd är ett begrepp som beskriver ett träd med en stamdiameter större än en meter. Tillsammans med väldigt gamla träd och grova hålträd utgör jätteträd en kategori träd som av Länsstyrelsen definieras som särskilt skyddsvärda (Holst 2008). Dessa jätteträd har höga ekologiska värden och det finns hundratals hotade arter som är knutna till dem. I trädens varierande livsmiljöer finns livsviktiga miljöer för fåglar, insekter, lavar, mossor och svampar (ibid). Arbetet med att bevara och vårda särskilt skyddsvärda träd är en viktig faktor för att uppnå flera av miljökvalitetsmålen som staten siktar mot.

Utöver detta bidrar dessa träd också med ett flertal ekosystemtjänster för oss människor, till exempel temperatur- och bullerreglering i urbana miljöer samt renare luft (*Träd som biologiskt kulturarv* 2019). De utgör också ett särskilt imponerande landskapsarkitektoniskt element i landskapet, något som tagit hundratals år att uppnå och som i stort sett är omöjligt att ersätta med andra strukturer (ibid.).

Vidare hyser dessa jättar också stora kulturhistoriska värden (*Träd i Sverige: kulturhistoriska och samtida perspektiv* 2024). I denna bok beskrivs träd som biologiska kulturarv. Här nämns flera exempel på kulturellt modifierade träd som skvallrar om olika användningsområden och förhållningssätt under historien, till exempel hamling, råvaruutvinning ur stammar och katning (ibid.). Dessa spår ger värdefulla insikter om historiska och kulturella metoder som utan dessa levande påminnelser riskerar att falla i glömska.

1.2 Behovet av inventering

Enligt Naturvårdsverket har Sverige ett av Europas största bestånd av jätteträd, och därmed också ett särskilt ansvar för bevarandet av dessa. Nyrekryteringen, av nya jätteträd är av sin natur långsam och ibland saknas den helt Naturvårdsverket skrivet därför i rapporten "*Samråd om åtgärder på särskilt skyddsvärda träd*" att bevara så många av dessa träd som möjligt har hög prioritet. Vidare skriver länsstyrelsen i en rapport att det finns ett stort mörkertal när det kommer till mängden skyddsvärda träd i vårt landskap (Samråd om åtgärder på särskilt skyddsvärda träd u.å.).

En central aspekt i att kunna arbeta med strategier för Sveriges bestånd av jätteträd är att veta var de finns. Idag har en del inventeringar med fokus på särskilt skyddsvärda, och jätteträd, skett på beställning av bland annat länsstyrelsen. Dessa inventeringar har skett i fält och är på grund av sin natur inte möjliga att genomföra över större ytor, det skulle vara orealistiskt att manuellt inventera hela Sveriges område.

1.3 Hur AI och maskininlärning kan hjälpa till i jakten

AI-verktyg och modeller baserade på maskininlärning blir en allt större del av allt fler branscher, och landskapsarkitekturen kommer inte att vara ett undantag. Med verktyg som automatiskt kan hantera och analysera stora mängder data kommer nya möjligheter att öppnas för inventering av träd. Detta sker i varierande skala redan idag, och flera pilotprojekt har bevisat möjligheterna med denna teknik. Huruvida dessa verktyg kan användas för storskalig inventering av jätteträd behöver utredas.

2. Syfte och frågeställning

Syftet med denna rapport är att undersöka metoder som möjliggör automatisk identifiering av jätteträd i en större skala. Rapporten undersöker dels metoder för individuell kronidentifieting och segmentering samt presenterar för- och nackdelar med dessa. Vidare utforskas möjligheten att ta fram en konceptvaliderings-modell för att undersöka sambandet mellan kron- och stamdiameter som ett steg i att kunna identifiera möjliga jätteträd via flygfoto.

Fokus ligger på metoder som använder modern teknik som möjliggör automatisk identifiering i större skala. Metoderna ska vara rimliga att genomföra ur ett tids- och kostnadsperspektiv.

Frågeställningarna som rapporten ämnar besvara är:

Vilka moderna tekniker kan användas för att automatiskt identifiera trädkronor i stor skala?

Kan man med publikt tillgänglig data träna en modell som ser samband mellan krondiameter och stamdiameter och på så vis kunna uppskatta jätteträd genom flygfoton?

3. Avgränsningar

Den litteraturstudie som ligger till grund för detta arbete är en narrativ litteraturstudie där relevanta källor identifierades genom sökningar på Google Scholar. Urvalet skedde subjektivt baserat på studiens syfte och tillgängliga sökresultat, vilket innebär att viss relevant forskning kan ha förbisetts

Rapporten har Sverige som utgångspunkt men har använt sig av studier från andra länder, ett antagande att liknande modeller skulle kunna appliceras/tas fram för Sverige har gjorts.

Konceptvalideringsmodellen är anpassad för att fungera i Sverige.

Konceptvalideringsmodellen är tränad på data innehållande stadsträd från en specifik stad.

4. Metod

4.1 Metod för litteraturstudie

För att undersöka inventeringsmetoder som används idag och som kan användas för att identifiera och segmentera trädkronor har en narrativ litteraturstudie genomförts. Relevanta källor identifierades genom sökningar på Google Scholar. Urvalet skedde baserat på studiens syfte och tillgängliga sökresultat.

Litteraturen som använts i arbetet inkluderar vetenskapliga artiklar, forskningsrapporter, antologier, myndighetsrapporter och av branschen vedertaget accepterade standarder.

4.2 Metod för konceptvalidering

För att undersöka sambandet mellan kron- och stamdiameter har en konceptvalideringsmodell tagits fram. Denna maskininlärningsmodell är skriven med Python som kodspråk.

Modellen tränades med data från Digitaliseringsenheten på Fastighets- och Gatukontoret på Malmö stad. Härifrån mottogs en shape-fil med ett utdrag ur Malmös träddatabas. Filen innehöll över 80 000 inventerade stadsträd, filen innefattade tre utvalda parametrar; artnamn, krondiameter (m) och stamdiameter (cm). Koordinatsystem SWEREF99 1330.

För att ladda ner datan som en CSV-fil användes QGIS.

Python användes för att städa och strukturera datan. Alla träd där stamdiameter och/eller krondiameter saknas eller är inskrivet som 0 exkluderades från datasetet.

I datasetet definieras ett ”jätteträd” som stamdiameter ≥ 100 centimeter, övriga träd klassificeras som icke-jätteträd.

Konceptvalideringen innefattar totalt tre metoder, samtliga faller under maskininlärningskategorin *supervised learning* (väglett lärande). De modeller som testas är linjär regression, logistisk regression och Random Forest, samt undermetoder där bland annat SMOTE används. Läs mer om dessa begrepp under rubrik 5 (Teoretisk bakgrund). Totalt tränades modeller för 10 arter, för varje art tränades sex olika modeller med följande kod:

1. Linjär regression.

from sklearn.linear_model import LinearRegression
"Linear Regression (Thresholding)": LinearRegression()

2. Logistisk regression

from sklearn.linear_model import LogisticRegression

LogisticRegression(max_iter=1000),

3. Logistik regression (Blanced)

"Logistic Regression (Balanced)":
LogisticRegression(class_weight='balanced', max_iter=1000),

4. Logistisk regression (SMOTE)

from imblearn.over_sampling import SMOTE
"Logistic Regression (SMOTE)": LogisticRegression(max_iter=1000),

5. Random Forest (Balanced)

from sklearn.ensemble import RandomForestClassifier
"Random Forest (Balanced)":
RandomForestClassifier(class_weight='balanced', random_state=42),

6. Random Forest (SMOTE)

from sklearn.ensemble import RandomForestClassifier
from imblearn.over_sampling import SMOTE
"Random Forest (SMOTE)": RandomForestClassifier(random_state=42),

Sedan definieras träningsdatan och vad modellen ska försöka förutspå. Input är krondiameter och Output är Giant; sant eller falskt?

80% av datan valdes slumpmässigt ut för att träna modellen, och resterande 20% användes för testning.

Modellen använder sig av biblioteket Scikit-learn, vilket är ett etablerat bibliotek med kod som används för dessa modeller. Scikit-learn används för att definiera och träna modellerna.

Därefter testas modellernas prestanda genom att beräkna; Accuracy, precisicon, recall, F1-score och ROC AUC. Läs mer om dessa parametrar under rubrik 5 (Teoretisk bakgrund).

5. Teoretisk bakgrund

I detta avsnitt presenteras en fördjupad förståelse av jätteträd och deras betydelse. Därefter introduceras grundläggande begrepp inom maskininlärning som är viktiga för att förstå metoden bakom utvecklingen av konceptvalideringsmodellen.

5.1 Svenska jätteträd - kulturella och biologiska värden

Ett jätteträd är ett träd som har en stamdiameter större än 100 centimeter, motsvarande 3.14 meter i omkrets. Trädet kan vara både levande eller dött och omkretsen mäts på det smalaste stället under brösthöjd (Samråd om åtgärder på särskilt skyddsvärda träd u.å.). För att få betecknas som ett jätteträd måste trädet vara enstammigt. Flerstammiga träd eller träd med två stammar som vuxit ihop godkänns inte, men även utan att tillhöra denna definition av jätteträd besitter även dessa träd höga värden (Vad är ett jätteträd? u.å.).

Historiskt har jätteträd ofta gått att hitta på åkermark, åkerholmar och i vägrenar, träden var vanliga i slåtter- och beteslandskap (Holst 2008). Ett av de vanligaste jätteträden i Sverige, och kanske det mest karaktäristiska, eken, har under perioder i historien varit skyddad genom lag. Under 1200-talet ansågs det lika allvarligt att ta ner en ollonbärande ek som att mjölka någon annans kossa, detta för att skydda ollonproduktionen som var en viktig inkomstkälla. På 1500-talet var det virket som värdesattes, bland annat för produktion av skepp, detta ledde till att Kungen fick ensamrätt till alla landets ekar. När detta skydd försvann på 1800-talet minskade beståndet av stora ekar i landet. Ett undantag var ekar på adlig mark, där flera individer sparades (ibid.).

Sedan 1830-talet har beståndet av jätteträd minskat och dagens fokus på homogen produktionsskog gynnar inte jätteträden. Hoten är inte så ofta avverkning som att de tidigare öppna landskap där träden står växer igen av sly och buskar, samt konkurrens med andra träd (ibid.)

Riksantikvarieämbetet skriver på sin hemsida om hur träd har en särställning som biologiska kulturarv (Träd som biologiskt kulturarv 2019).

Här beskrivs hur träd genom att de fått växa under en lång tid fungerar som levande arkiv som berättar om tidigare miljöer. Exempel på detta är tecken på hamling. Vidare beskrivs hur träd alltid haft en central roll i människors liv, dels ur ett praktiskt perspektiv som virkesproduktion, men också som kulturella element, som till exempel vårdträd (ibid.).

Genom att studera trädens växtsätt kan vi människor få insikt i hur landskapet har använts historiskt, det gäller speciellt för träd som växer i anknytning till gammal odlingsmark, fäbodar eller gamla jaktmarker (ibid.).

Biologiskt bidar träden med livsmiljöer för en mängd organismer. Grova, gamla och ihåliga träd har en stor betydelse för dessa olika arter och livsmiljöerna i dessa träd är ofta helt avgörande för många hotade arters överlevnad brösthöjd (Samråd om åtgärder på särskilt skyddsvärda träd u.å.). En tumregel är att ju äldre och grövre ett träd blir desto fler livsmiljöer erbjuder de. Exempel på livsmiljöer knuta till gamla och grova träd är grov bark, solexponerad ved och håligheter. I Sverige är det eken som imponerar särskilt med antal arter som är knutna till den, över 1500 arter varav mer än 500 av dem är hotade (ibid.).

5.2 Supervised learning (väglett lärande)

Supervised learning är en typ av maskininlärning där modellen tränas med indata och motsvarande korrekt utdata. Datan som modellen tränar på är med andra ord redan klassificerad. Det modellen gör att lära sig sambandet mellan denna in- och utdata genom att träna på den, och minimera skillnaden mellan förutsägelser och riktiga värden. Supervised learning är ett passande verktyg för exempelvis klassificeringar och regressioner där man har ett dataset som redan är klassificerat att träna på, exempelvis en träddatabas från Malmö stad (Jordan & Mitchell 2015).

5.3 Linjär regression

Linjär regression är en grundläggande statistisk modell som undersöker linjära samband mellan olika variabler. Modellen undersöker sambandet mellan en beroende- och en oberoende variabel. I denna rapport är det sambandet mellan variablerna stamdiameter respektive krondiameter som undersöks (läs mer under metod samt resultat). Linjär regression är en etablerad statistisk modell som ger ett resultat som är enkelt att tyda. Eftersom linjär regression är ett etablerat verktyg inom statistik och maskininlärning är kunskapen kring modellerna hög och de går att träna modellerna snabbt med ny data. (Jordan & Mitchell 2015).

En regressionsmodell genererar en determinationskoefficient (R2), som är en koefficient som anger hur stor del av variationerna i den beroende variabeln (y) som kan förklaras av variationer i den oberoende variabeln (x) under förutsättning att sambandet mellan x och y är linjärt. Determinationskoefficienten kallas ofta förklaringsgradvarianskoefficient.

R2-värdet sträcker sig från 0,00 (0%) till 1,00 (100%), där 0,00 betyder att modellen inte förklarar något av resultatet, och 1,00 indikerar en perfekt passform. Vad som kan anses vara ett bra R2-värde beror på kontexten och varierar mellan olika vetenskapsområden. Inom biologi brukar ett R2-värde >0,5 ofta anses tyda på ett bra eller mycket bra samband (Piepho 2018).

5.4 Logistisk regression

Logistisk regression är en statistisk metod som kan modellera oddset för utfall som är binära, det vill säga kan delas upp i två kategorier, till exempel ja/nej eller

stamdiameter större/mindre än 100 centimeter. Metoden bygger på den logistiska funktionen vilket betyder att den räknar ut oddskvoten för att en observation tillhör en av de två klasserna, och ligger alltid mellan noll och ett. Denna metod används ofta för maskininlärning för binära klassificeringsproblem (Jordan & Mitchell 2015).

5.5 Random Forest

Random Forest är en vanlig maskininlärningsalgoritm som kan användas till både klassificering och regression. Modellen använder sig av flertalet så kallade beslutsträd och kombinerar resultatet av dessa till ett enda resultat för att förbättra noggrannheten i prediktionerna.

Varje beslutsträd delar upp data vid flera klyvningspunkter baserade på de valda variablerna, vid varje klyvningstillfälle väljs ett antal slumpmässigt utvalda variabler för att identifiera den optimala uppdelningen. Random Forest använder sig också av så kallad ensemblemetod, en kombination av flera modeller för att förbättra prestandan jämfört med enskilda modeller.

Denna kombination av slumpmässiga variabler och flera olika beslutsträd gör Random Forest mer robust och mindre benägen att överanpassa sig till träningsdatan. Överanpassning innebär att modellen lär sig mönster som är specifika för träningsdatan istället för generella mönster som kan användas på ny data. Modellen tror då att den är bra på att göra förutsägelser men i praktiken presterar modellen dåligt (Mansour & Schain 2001).

5.6 SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE (Synthetic Minority Over-sampling Technique) är en metod inom maskininlärning som används för att hantera obalanserade dataset. Obalanserade dataset kan leda till att modeller blir partiska mot de dominerande kategorierna, vilket leder till att modellen har svårare att korrekt klassificera de underrepresenterade klasserna.

Konceptvalideringen i denna rapport bygger på dataunderlag från Malmö stad. Av dessa över 80 000 inventerade stadsträd är endast en mycket liten andel jätteträd. Det innebär att när modellen tränar på dessa arter och sedan ska avgöra om trädet tillhör kategorin icke-jätteträd eller kategorin jätteträd kan den felaktigt presentera en hög noggrannhet i sina förutsägelser. I själva verket skulle modellen ha kunnat säga att inget av träden är jätteträd och fortfarande presentera en hög noggrannhet, eftersom den på grund av den obalanserade datan i de flesta fallen får rätt.

Metoden innebär att syntetiska exempel skapas genom interpolation mellan befintliga datapunkter i minoritetsklassen, vilket ökar dess representation i datasetet. Detta kallas översampling och används för att förbättra modellens förmåga att känna igen och korrekt klassificera den underrepresenterade klassen.

Genom att använda SMOTE kan fler syntetiska exempel av jätteträd skapas, vilket hjälper modellen att bättre identifiera dessa i framtida prediktioner (Chawla et al. 2002).

5.7 Confusion matrix

Confusion matrix är en tabell som används när man utvärderar prestandan hos en klassificeringsmodell. Tabellen är uppdelad i fyra kategorier: true positives (TP), false positives (FP), true negatives (TN) och false negatives (FN). Alla förutsägelser modellen gör delas in i någon av dessa kategorier, till exempel hamnar en korrekt förutsägelse att ett träd är ett jätteträd i TP, medan ett träd som antogs vara ett jätteträd men inte var det hamnar i FP. Dessa förutsägelser presenteras i matrixen och gör det möjligt att beräkna precisionen och noggrannheten hos modellen (Fawcett 2006).

5.8 ROC-värde & AUC-kurva

ROC-kurva (Receiver Operating Characteristic) och AUC-kurva (Area Under the Curve) är mått på klassificeringsmodellens prestanda. ROC-kurvan används för att visa sambandet mellan följande två mått:

1. True Positive Rate (TPR), som är andelen verkliga positiva fall (TP) som modellen har identifierat korrekt, kallas även för sensitivitet.
2. False Positive Rate (FPR), som är andelen negativa fall som felaktigt klassificerats som positiva.

AUC är arean som faller in under ROC-kurvan och ger ett mått på modellens prestanda, modellens förmåga att skilja mellan de olika klasserna. Här betyder ett AUC-värde på 1.0 en perfekt modell som gör en 100% korrekt bedömning medan ett värde på 0.5 betyder att modellen presterar lika bra som en slumpmässig gissning. Desto högre värde, desto bättre modell (Fawcett 2006).

5.9 Andra parametrar för att utvärdera en modells prestanda

Nedan beskrivs kortfattad en del vanliga parametrar som används för att utvärdera en klassificeringsmodells prestanda. Dessa mått är särskilt viktiga att analysera när man studerar prestandan på klassificeringsmodeller som tränat på obalanserad data (Powers 2020).

Noggrannhet (Accuracy): Beskriver andelen korrekta förutsägelser, både true positives och true negatives, i förhållande till det totala antalet förutsägelser.

Precision: Beskriver andelen true positives resultat i förhållande till den totala mängden förutsägelser.

Återkallning (Recall): Beskriver andelen true positive resultat i förhållande till alla resultat som faktiskt var positiva.

F1-poöng (F1-score): Ett medelvärde av precision och återkallning vilket ger en balanserad bedömning av modellens prestanda.

6. Resultat – Undersökande studie

I detta avsnitt presenteras ett urval av inventeringsmetoder som används idag och som skulle kunna användas i framtiden när tekniken utvecklats vidare. Fokus ligger på metoder för individuell trädkroneavgränsning (ITCD). Anledningen till att fokuset är på just ITCD är att det är genom sambandet mellan kron- och stamdiameter denna rapports konceptvalideringsmodell bygger på. Vidare presenteras för- och nackdelar med dessa metoder.

6.1 Fältinventering

En grundläggande metod för inventering är fältinventering, där man i fält och med enklare verktyg läser av trädets dimensioner, uppskattad ålder, vitalitet samt andra parametrar. Vanligt är att använda sig av måttband för att mäta omkrets och diameter på stam, samt för att mäta krondiameter. För inventering av stadsträd följs vanligtvis Standard för trädinventering i urban miljö 3.0 (författad av Johan Östberg och Elin Rowicki), i denna standard tas en stor mängd parametrar upp men vilka parametrar man väljer att inventera beror på syftet med inventeringen. (Östberg et al 2022).

Fördelar:
- Enkelt att genomföra och kräver lite hårdvara.
- Noggrannhet med möjlighet att komma nära träden och analysera en stor mängd parametrar.

Nackdelar:
- Mänsklig faktor. Mäter alla likadant?
- Svårt/omöjligt att genomföra på stora ytor.

6.2 Systematisk fältinventering

Under denna metod faller även systematisk fältinventering. Det är en metod där man begränsar fältinventeringens ytor till för projektets syfte relevanta områden (Inventering av skyddsvärda träd i kulturlandskapet 2021).

Urvalen av inventeringsområden kan se olika ut beroende på inventeringsprojetets syfte, beroende på projektets natur kan man använda sig av slumpmässiga ytor, förutbestämda ytor baserat på olika biotoper eller av GPS och GIS-integration, genom att använda GIS kan man koppla fältdata till kartor och ortofoton.

Naturvårdsverket sammanställde 2021 version 3.0 av Inventering av skyddsvärda träd i kulturlandskapet (ibid.). Metoden bygger på fältarbete där träd inventeras enligt en given standard, i detta fall bland annat stamdiameter, art, vitalitet och andra eventuella indikationer på höga naturvärden. För att säkerställa att

inventeringsresultaten kan följas upp över tid noteras trädets position med GPS-punkter.

Urvalet på områden man prioriterar att inventera kan bygga på olika faktorer. Ett sätt är att välja prioriterade landskapstyper där man genom historiska data vet att skyddsvärda träd vanligen förekommer, nämligen kulturlandskap som gamla betesmarker, parker, kyrkogårdar, alléer och gamla skogsbestånd (ibid.).

Fördelar:
- Enkelt att genomföra. Kräver lite hårdvara.
- Noggrannhet. Möjlighet att komma nära träden och analysera en stor mängd parametrar.
- Effektivare inventering genom avgränsade inventeringsområden.

Nackdelar:
- Mänsklig faktor. Mäter alla likadant?
- Stor risk att missa träd. Resultat avhängt på val av sökområde.

6.3 Fjärranalys

Fjärranalys är ett samlingsbegrepp på metoder där inventeringen sker från avstånd. Det finns flera olika tekniker som kan samla information om träd på detta vis. I grunden handlar det om olika bildalstrande sensorer monterade på flygplan, satelliter eller andra fordon. Signaler i det elektromagnetiska spektrumet (strålning) skickas ut från sensorerna och omvandlas sedan till bilder. Dessa bilder kan sedan med stöd av fältobservationer omvandlas till kartor (Skoglig fjärranalys 2016).

Analysen av den insamlade datan sker vanligtvis med datorbaserade modeller, ett vanligt exempel är att analysera färgskalan och/eller avstånd mellan strålningspunkterna. Med färganalys har flera studier visat på möjligheten att urskilja enskilda trädarter ur ett samlat trädbestånd (ibid.).

Fördelar:
- Möjlighet att läsa av stora områden.
- Relativt kostnadseffektivt beroende på val av data och hårdvara.

Nackdelar:
- Resultat avhängt på val av data.
- Låg detaljrikedom i inventeringsparametrar.
- Resultat beroende av data. Ofta svårt med individuell kronsegmentering.

6.4 Användning av maskininlärning

Traditionella inventeringsmetoder är tidskrävande och resursintensiva. Det finns flera internationella och svenska projekt där man har nyttjat maskininlärning och AI för att automatisera delar av dessa processer för att mer effektivt kunna inventera träd (Vos, Christian 2024).

Att använda AI och maskininlärning för trädanalys och automatisk ITCD innebär en del utmaningar. Noggrannheten i resultatet beror på vilket dataunderlag modellen har att tillgå. (ibid.)

6.5 Användning av ortofoton

Genom att träna maskininlärningsmodeller med RGB-ortofoton kan man teoretiskt sätt urskilja enskilda trädkronor. Modellerna använder sig av bilddatans färgskala som går i Rött, Grönt och Blått och kan genom skillnader i dessa lära sig vilken färgskala som representerar vilket träd (Besker et al 2024).

Att enbart träna en modell med RGB-flygfoton innebär dock en del utmaningar. Frånvaron av annan data i bildunderlaget, till exempel höjddata, gör det svårt för modellerna att lära sig skillnaden mellan ett trädens skugga och det faktiska trädet. Det blir särskilt utmanande i täta trädbestånd då trädkronor klumpas ihop. Ofta är dessa bilder också insamlade under den bladfria säsongen vilket förvisso ger mer information om trädets struktur men det gör också svårare att använda sig av färgskalorna för att urskilja enskilda träd under bladsäsong (ibid.).

Idag blir användning av digitala ortofoton allt vanligare i och med att kvalité och upplösning på dessa blir bättre. Ett exempel på detta är hur Lantmäteriet tagit fram en produkt kallad “Ytmodeller från flygbilder” där de bearbetar sterofotograferade flygbilder så att de liknas ett 3D punktmoln som liknar dem från laserdata (Ytmodell från flygbilder u.å.).

Fördelar:
- Möjlighet att läsa av stora områden.
- Relativt kostnadseffektivt beroende på val av data och hårdvara.
- Möjligt att genomföra med publikt tillgängliga data.

Nackdelar:
- Resultat avhängt på val av data.
- Ofta lågt upplösta foton vilket gör individuell kronsegmentering svårare.
- Låg detaljrikedom i inventeringsparametrar.

6.6 Användning av LiDAR

LiDAR (Ligt Detection and Ranging), även kallat laserhöjdmätning, är en etablerad metod för fjärranalysteknik. Tekniken fungerar genom att en sensor monterad på till exempel en drönare eller flygplan skickar ut strålar av laserenergi mot marken och mäter tiden det tar för strålarna att studsa tillbaka. Beroende på hur högt ett objekt är från marken varierar tidsintervallet det tar för laserstrålen att studsa tillbaka till sensorn, dessa skillnader i tid ligger sedan till grund för en tredimensionell punktmolnsmätning av området under lasersensorn. Resultatet blir detaljerade 3D-modeller av landskapet (Lidar 101: An Introduction to Lidar Technology, Data, and Applications 2012).

Ett exempel på framgångsrik tillämpning av LiDAR är en studie från 2016 där forskarna framgångsrikt tog fram en tredimensionell modell av träden i en regnskog och kunde uppskatta biomassan för respektive träd (Ferraz et al. 2016).

I en artikel från 2023 där forskarna använder LiDAR som en del i sin metod för att segmenteras enskilda träd (se mer under rubrik 6.8.) beskrivs några begränsningar med att använda LiDAR på stor skala. I artikeln beskrivs hur LiDAR förvisso ger ett starkt dataunderlag med hög noggrannhet, men begränsningen ligger i att detta dataunderlag ofta samlas in under begränsade och mindre kampanjer. Det kan därför bli dyrt att genomföra på nationell skala.

Fördelar:
- Möjlighet att läsa av stora områden.
- Hög detaljrikedom.
- Möjlighet att läsa av flera parametrar med hjälp av maskininlärning.

Nackdelar:
- Dyr hårdvara.
- Dyrt att genomföra på stora ytor.
- Hög teknisk kompetens krävs.
- Finns limiterad publikt tillgänglig data.

6.7 Användning av CNN (Convolutional neural networks)

CNN är en maskinlärningsalgoritm som automatiskt lär sig tolka bilder. Dessa neurala nätverk använder sig av en mängd lager som analyserar olika aspekter av bilden, i början väldigt grundläggande aspekter för att senare i processen analysera allt mindre detaljer i bilden. Tekniken är väl lämpad för bildigenkänning och baserat på vilken in-data som modellen har tränat på kan den lära sig urskilja specifika kategorier av objekt, till exempel fordon, ansikten eller trädarter (Krizhevsky et al. 2017).

I en studie från 2023 användes tekniken för att lära sig känna igen specifika trädarter baserat på RGB-bilder, bilder bestående av tre färgkanaler Röd, Grön och Blå som tillsammans utgör bilden (Beloiu et al. 2023).

Studien beskriver hur CNN visat sig vara väldigt användbart när det kommer till att analysera ekosystem på stor skala. I studien beskrivs fyra huvudsakliga metoder för bildklassificering: semantisk segmentering, objektigenkänning, instanssegmentering och panoptisk segmentering. I studien visade författarna hur de lyckades urskilja trädarter som Picea abies, Abies alba, Pinus sylvestris och Fagus sylvatica baserat på ovan beskrivna RGB-bilder (ibid.)

I en annan studie från 2017 utvecklade forskare en metod för att hitta enskilda träd, dess storlek och höjd med hjälp av flygbilder (Fisher et al. 2017). Bildunderlaget i studien kom från Lanstat eller SPOT5, vilket är satellitbilder med hög upplösning som båda är vanliga inom miljöanalys och forskning. Genom att kombinera dessa högupplösta bilder med höjddata kunde modellen identifiera enskilda trädkronor samt trädets höjd. Studien visade att det är möjligt att urskilja enskilda träd även fast just denna metod hade en del svårigheter med lövträd i täta bestånd då kronorna klumpas ihop. För barrträd var resultaten desto bättre (ibid.)

Balansen mellan att utvärdera krontäckning och identifiera enskilda trädkronor var också en utmaning, när modellen fokuserar på enskilda träd kunde delar av kronan missas och när modellen fokuserar på att få med hela kronan hände det att trädkronor slogs ihop. Sammanfattningsvis visade modellen lovande resultat i att identifiera enskilda trädkronor, ju större träd desto mer tillförlitligt resultat (ibid).

I ett samarbete mellan RISE och Trollhättan kommun genomfördes mellan 2023 och 2024 projektet T.AI.TRÄ. Här beskrivs implementeringen av CNN för att genomföra individuell trädkroneavgränsning med hjälp av kombinerade ortofoton och LiDAR, rapporten beskriver hur detta medför en del utmaningar (Besker et al 2024).

Fördelar:
- Tidseffektivt.
- Hög detaljrikedom.

- Möjlighet att läsa av flera parametrar med hjälp av maskininlärning.
- Möjlighet att hantera stora mängder data.

Nackdelar:
- Hög teknisk kompetens krävs.
- Resultat beroende av träningsdatan.

6.8 Kombination av olika metoder

I en forskningsrapport från 2023 skriver skribenterna att de tagit fram en modell för automatisk bildbaserad trädräkning, segmentering och höjdestimering på nationell skala (Li et al. 2023).

Studien genomfördes i Danmark och metoden bygger på två separata modeller som samtidigt utför tre separata uppgifter, den första modellen kombinerar trädidentifiering och kronsegmentering med hjälp av flygbilder i flera olika färgspektrum. Den andra modellen tränar på LiDAR-data och uppskattar trädets höjd. För att uppnå segmenteringen av trädkronor, något som i många andra studier visat sig vara en utmaning, använde modellen över 20 000 färdigsegmenterade träd att träna på. Dessa färdigsegmenterade träd gjordes manuellt (ibid.).

Fördelar:
- Tidseffektivt.
- Hög detaljrikedom.
- Möjlighet att läsa av flera parametrar med hjälp av maskininlärning.
- Möjlighet att hantera stora mängder data.
- Möjlighet att framgångsrikt segmentera enskilda trädkronor.

Nackdelar:
- Hög teknisk kompetens krävs.
- Resultat beroende av träningsdatan.
- Tidskrävande process att ta fram bra träningsdata.

7. Resultat - Konceptvalideringsmodell

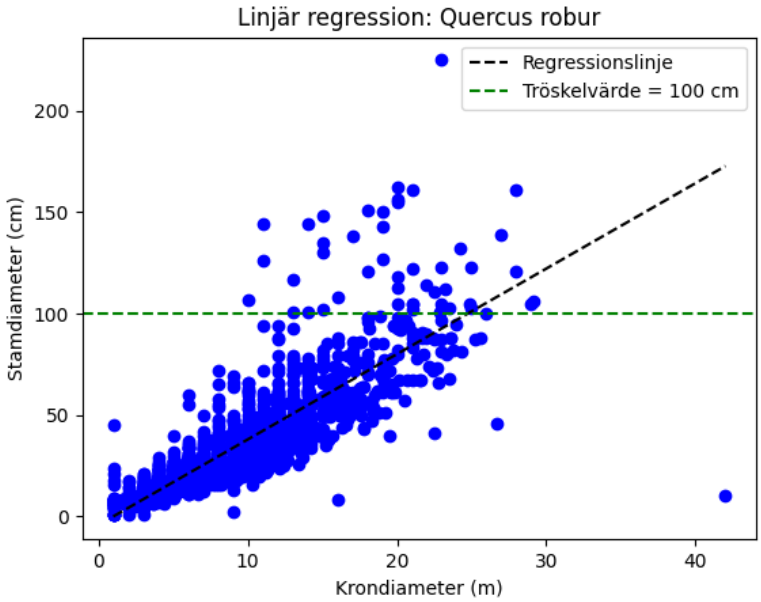
7.1 Resultat av konceptvalidering

I detta avsnitt presenteras resultatet för modellen som har undersökt sambandet mellan kron- och stamdiameter. 10 av rapportförfattaren utvalda arter presenteras i regressionsdiagram som visar sambandet mellan krondiameter (m) på X-axeln och stamdiameter (cm) på Y-axeln. De enskilda datapunkterna är utplottade som blåa punkter och regressionslinjen illustrerar det linjära mönstret i datan.

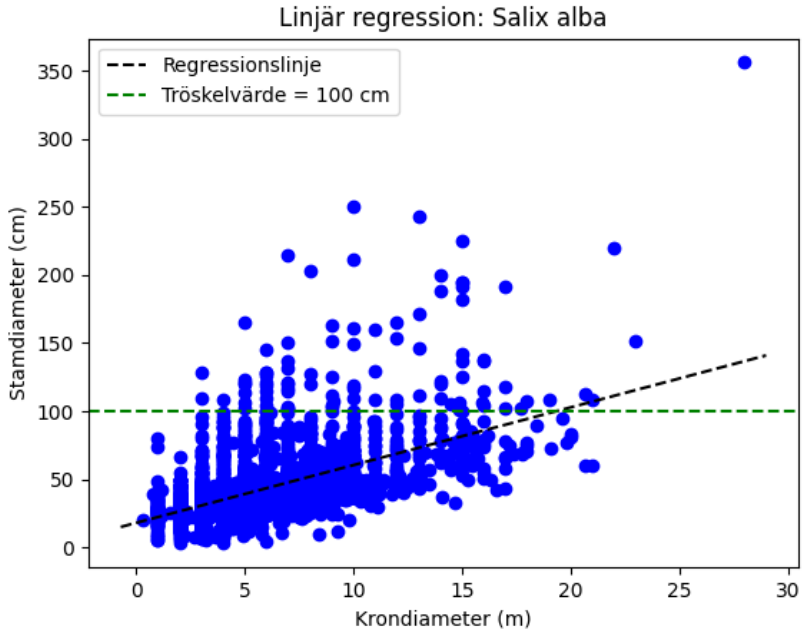
Figurerna visar att det generellt finns ett relativt starkt linjärt samband mellan kron- och stamdiameter för flertalet arter, med R2-värden upp mot drygt 0,70. Det går också tydligt att urskilja att sambandet i de flesta fallen förloras för jätteträden.

Det går också att dra slutsatsen att det linjära sambandet ser olika ut för olika arter och att olika arter har olika enhetligt linjärt samband.

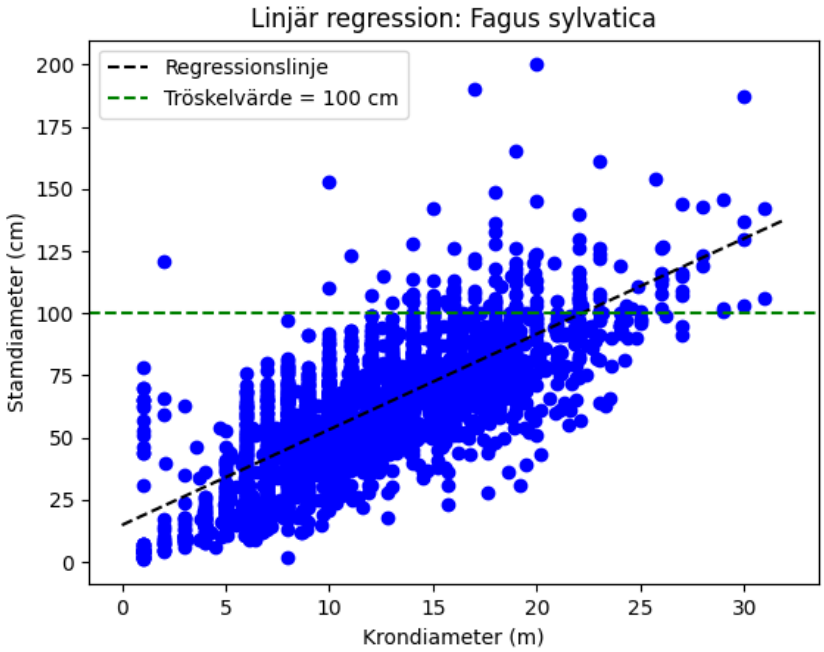
7.1.1 Det linjära sambandet mellan kron- och stamdiameter för 10 utvalda arter



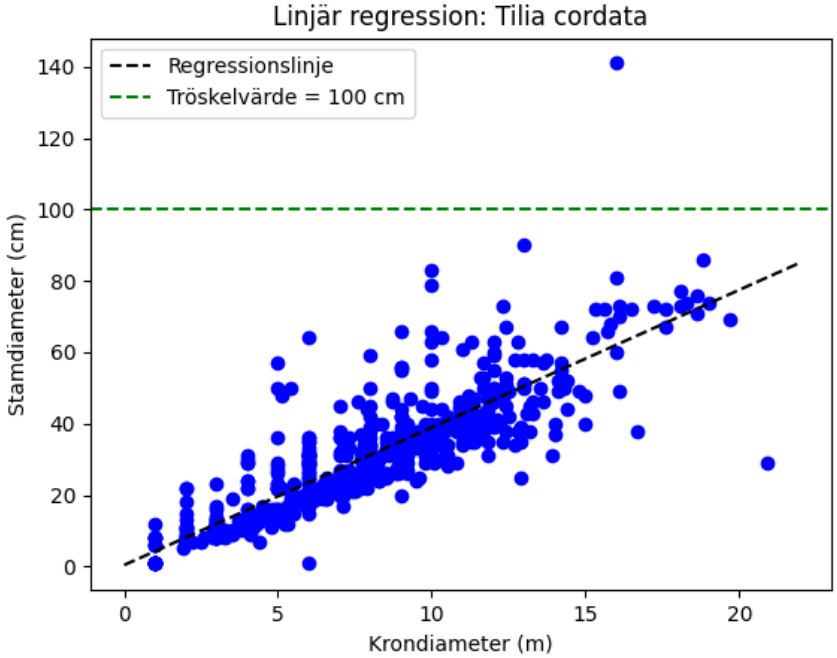
Figur 1. Det linjära sambandet mellan kron- och stamdiameter för ek (Quercus robur). R2=0,70.



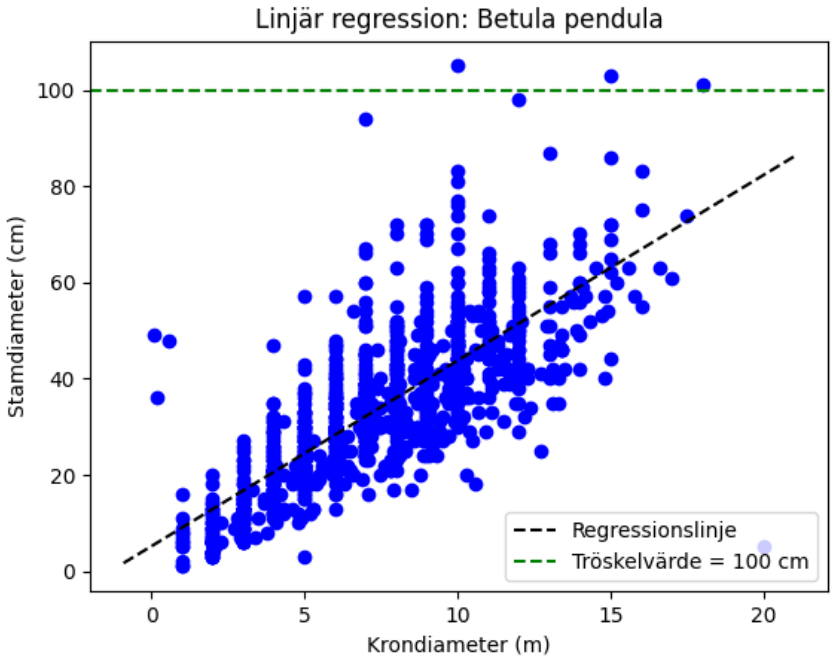
Figur 2. Det linjära sambandet mellan kron- och stamdiameter för vitpil (Salix alba). R2=0,33.



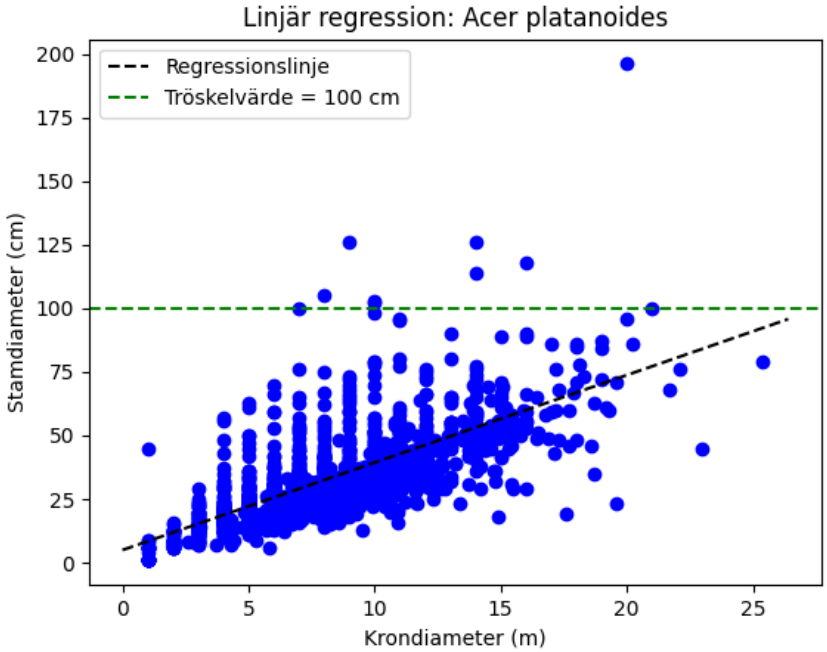
Figur 3. Det linjära sambandet mellan kron- och stamdiameter för bok (Fagus sylvatica). R2=0,56.



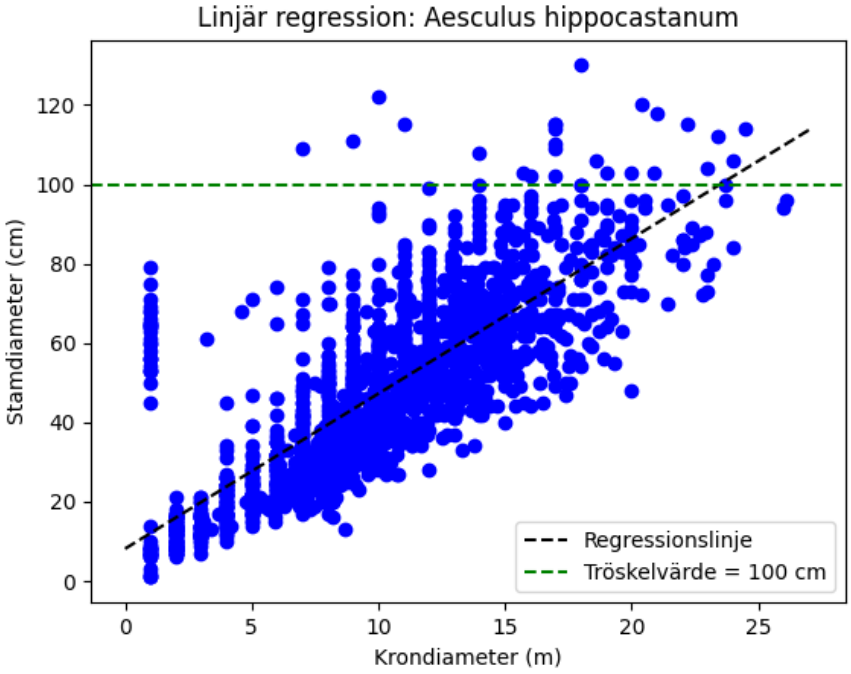
Figur 4. Det linjära sambandet mellan kron- och stamdiameter för skogslind (Tilia cordata). R2=0,72.



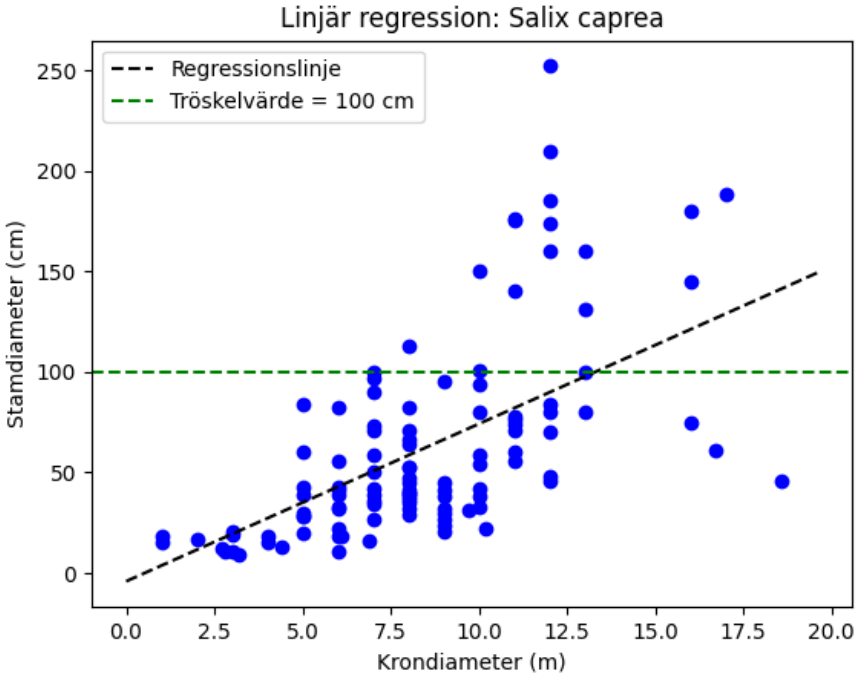
Figur 5. Det linjära sambandet mellan kron- och stamdiameter för vårtbjörk (Betula pendula). R2=0,60.



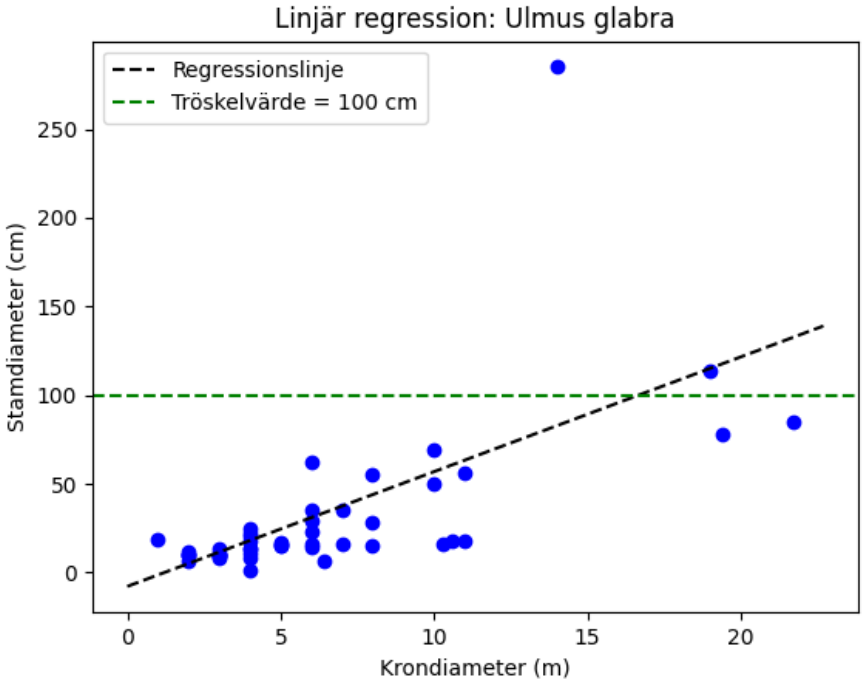
Figur 6. Det linjära sambandet mellan kron- och stamdiameter för skogslönn (Acer platanoides). R2=0,50.



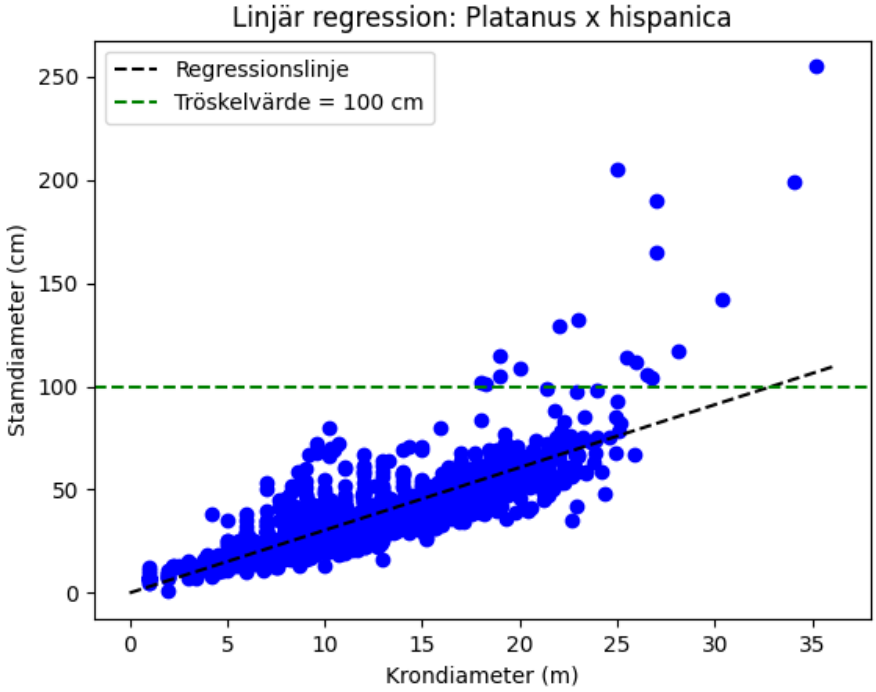
Figur 7. Det linjära sambandet mellan kron- och stamdiameter för hästkastanj (Aaesculus hippocastanum). R2=0,59.



Figur 8. Det linjära sambandet mellan kron- och stamdiameter för sälg (Salix caprea). R2=0,37.



Figur 9. Det linjära sambandet mellan kron- och stamdiameter för alm (Ulmus glabra). R2=0,40.



Figur 10. . Det linjära sambandet mellan kron- och stamdiameter för platan (Platanus x hispanica). R2=0,66.

7.1.2 Den linjära regressionsmodellens prestanda

Tabell 1 visar den linjära regressionsmodellens prestanda att identifiera jätteträd. I tabellen lyfts fyra trädarter fram av artikelförfattaren, urvalet bygger på att nedan visade arter hade en förhållandevis stor andel jätteträd. Parametern F1-score visas här som 0.00 vilket beror på att även Precision och Recall är 0.00. Detta beror på den obalanserade träningsdatan.

Modell: **Linjär regression**	Accuracy	Precision	Recall	**F1-Score**	ROC AUC
Quercus robur	0.984252	0.000000	0.00000	**0.00000**	0.823778
Fagus sylvatica	0.951473	0.000000	0.00000	**0.00000**	0.815834
Salix alba	0.965612	0.000000	0.00000	**0.00000**	0.902279
Aesculus hippocastanum	0.994169	0.000000	0.00000	**0.00000**	0.562317

Tabell 1. Modellens prestanda. Linjär regression.

7.1.3 Logistisk regression resultat

Tabell 2 visar den logistiska regressionsmodellens prestanda att identifiera jätteträd. Denna modell är balanserad, vilket innebär att träningsdatan har balanserats för att väga upp för det lilla antalet jätteträd. I tabellen lyfts fyra trädarter fram av artikelförfattaren, urvalet bygger på att nedan visade arter hade en förhållandevis stor andel jätteträd. Parametern F1-score skiljer sig något mellan trädarter vilket beror på att det linjära sambandet ser olika ut för olika arter.

Modell: **Logistisk regression (balanserad)**	Accuracy	Precision	Recall	**F1-Score**	ROC AUC
Quercus robur	0.895013	0.130435	1.000000	**0.230769**	0.957778
Fagus sylvatica	0.811092	0.176000	0.785714	**0.287582**	0.871845
Salix alba	0.812930	0.111888	0.640000	**0.190476**	0.888205

Aesculus hippocastanum	0.819242	0.016129	0.500000	**0.031250**	0.857771

Tabell 2. Modellens prestanda. Logistisk regression (balanserad).

Tabell 3 visar den logistiska regressionsmodellens prestanda att identifiera jätteträd. Denna modell innefattar SMOTE, vilket innebär att träningsdatan har fyllts på med syntetiska jätteträd för att väga upp för det lilla antalet jätteträd. I tabellen lyfts fyra trädarter fram av artikelförfattaren, urvalet bygger på att nedan visade arter hade en förhållandevis stor andel jätteträd. Parametern F1-score skiljer sig något mellan trädarter vilket beror på att det linjära sambandet ser olika ut för olika arter

Modell: **Logistisk regression (SMOTE)**	Accuracy	Precision	Recall	**F1-Score**	ROC AUC
Quercus robur	0.895013	0.130435	1.000000	**0.230769**	0.957778
Fagus sylvatica	0.811092	0.176000	0.785714	**0.287582**	0.871845
Salix alba	0.811554	0.111111	0.640000	**0.189349**	0.888205
Aesculus hippocastanum	0.819242	0.016129	0.500000	**0.031250**	0.857771

Tabell 3. Modellens prestanda. Logistisk regression (SMOTE).

7.1.4 Random Forest resultat

Denna tabell visar prestandan på modellen Random Forest att identifiera jätteträd. Denna modell är balanserad, vilket innebär att träningsdatan har balanserats för att väga upp för det lilla antalet jätteträd. I tabellen lyfts fyra trädarter fram av artikelförfattaren, urvalet bygger på att nedan visade arter hade en förhållandevis stor andel jätteträd. Parametern F1-score är för denna modell lägre än den visad i tabell 5. Detta beror på att träningsdatan varit obalanserad.

Modell: **Random Forest (balanced)**	Accuracy	Precision	Recall	**F1-Score**	ROC AUC
Quercus robur	0.910761	0.062500	0.333333	**0.105263**	0.757778
Fagus sylvatica	0.811092	0.176000	0.785714	**0.287582**	0.871845
Salix alba	0.807428	0.138365	0.880000	**0.239130**	0.889003
Aesculus hippocastanum	0.819242	0.016129	0.500000	**0.031250**	0.857771

Tabell 4. Modellens prestanda. Random Forest (balanserad).

Tabell 5 visar prestandan på modellen Random Forest att identifiera jätteträd. Denna modell innefattar SMOTE, vilket innebär att träningsdatan har fyllts på med syntetiska jätteträd för att väga upp för det lilla antalet jätteträd. I tabellen lyfts fyra trädarter fram av artikelförfattaren, urvalet bygger på att nedan visade arter hade en förhållandevis stor andel jätteträd. Parametern F1-score är högre än för den modell som tränade på icke-balanserad data vilket betyder att träningsdatan är obalanserad.

Modell: **Random Forest (SMOTE)**	Accuracy	Precision	Recall	**F1-Score**	ROC AUC
Quercus robur	0.926509	0.076923	0.333333	**0.125000**	0.823778
Fagus sylvatica	0.831889	0.177570	0.678571	**0.281481**	0.815834
Salix alba	0.803301	0.140244	0.920000	**0.243386**	0.902279
Aesculus hippocastanum	0.976676	0.000000	0.000000	**0.000000**	0.562317

Tabell 5. Modellens prestanda. Random Forest (SMOTE).

8. Diskussion

Detta kandidatarbete föddes av idén att man automatiskt skulle kunna identifiera jätteträd med hjälp av publikt tillgängliga data i form av flygfoton. Idén bygger på att man genom att veta krondiametern skulle kunna uppskatta stamdiametern. För att detta ska fungera krävs två element:

Det första är en modell som kan segmentera individuella trädkronor automatiskt. Efter att ha genomfört en narrativ litteraturstudie som undersökt vetenskapliga studier och projekt som nyttjat maskininlärningsmodeller för detta syfte dras slutsatsen att det är en utmaning. Ofta är problemet att modellerna har svårt att urskilja enskilda trädkronor då skuggor och hopväxta trädkronor i täta trädbestånd är svåra för maskininlärningsmodellerna att lära sig urskilja från faktiska trädkronor.

En modell är bara så bra som datan den tränar på.

I grunden handlar det om att datan som modellerna har tränat på är undermålig. Den modell som identifierats som visat överlägset bäst resultat är också den som haft överlägset bäst träningsdata att tillgå.

Den modellen är den som beskrivs under rubrik 6.8. Modellen visade imponerande resultat när det kom till att korrekt identifiera och segmentera individuella trädkronor från en blandning av ortofoton och LiDAR-data, men egentligen är det endast ortofotografierna som skulle behövas för att urskilja och segmentera trädkronorna. Den ytterligare data som LiDAR bidrog med kan utifrån syftet med denna studie ses som ett plus, men inte något nödvändigt.

Denna modell visade goda resultat i att segmentera individuella trädkronor från ortofoton på blandskogar i Danmark, vårt eget grannland. Dessutom visade studien lovande resultat när den applicerades på finska skogar. Något försök att applicera modellen i Sverige har inte gjorts.

Datan som denna modell tränade på var över 20,000 manuellt segmenterade trädkronor. Det innebär att någon eller några har suttit och ritat ut över 20,000 trädkronor på ortofoton som sedan använts som träningsdata för modellen. Detta är såklart ett oerhört arbetsintensivt moment. Men det gav onekligen lön för mödan. Arbetsbördan till trots kan det argumenteras att det är ett rent av nödvändigt arbetsmoment för att kunna ta fram en så pass framgångsrik modell.

Att träna en modell likt den beskriven under rubrik 6.8 på svenska träd skulle vara av stort värde. Som nyss nämnt skulle det innebära en stor arbetsinsats att ta fram träningsdata likt den som användes för den danska modellen. Men samtidigt skulle det genera i en modell som skulle kunna användas för en mängd olika intressanta projekt, inte minst i den fortsatta jakten på jätteträd. Framtagandet av denna modell skulle kunna användas för att driva på utvecklingen i hela

branschen och bidra med insikter och verktyg som skulle kunna bidra till arbetet med träd i hela landet.

Det andra elementet som skulle krävas för att stödja idén som ligger till grund för detta arbete är att hitta ett samband mellan krondiameter och stamdiameter. Hypotesen var att ju grövre ett träd blir desto större blir trädkronan, och vice versa. I den narrativa litteraturstudien som genomfördes för detta projekt har ingen metod som använder maskininlärning för att undersöka detta samband kunnat identifieras. Därför togs det för rapporten fram en ny sådan.

Denna metod visade på att relativt starkt generellt samband mellan kron- och stamdiameter för flertalet arter. Däremot är sambandet för just jätteträd inte tillräckligt bra för att kunna använda i detta syfte. Studerar man det linjära sambandet i figur 1–10 kan man tydligt se det linjära sambandet, när stamdiametern växer följer krondiametern med. Detta stöds av att analysera R^{2-} talet som för en del arter är högt och för andra arter lägre.

Man ser också hur sambandet blir mindre enhetligt ju större stamdiametern blir. Här är de plottade trädpunkterna mer utspridda, detta betyder att grova träd har en större variation i krondiameter och det finns inte något direkt krondiametervärde man kan säga innebär ett jätteträd. Felmarginalen är stor.

Detta beror troligtvis på att jätteträd av sin natur också ofta är väldigt gamla. Under dess långa livsspann har kronorna ofta genomgått spontana kronförminskringar. Grenar torkar in och dör, trädet utsätts för angrepp av svamp och sjukdomar och mekaniska skador som tunga snöfall, åsknedslag och liknande har under de flera hundra år onekligen satt sin prägel på trädet. Med andra ord blir sambandet mellan kron- och stamdiameter mindre tydligt ju större ett träd blir. Detta skulle kunna påvisas tydligare om man enbart tränade modellen med små- till genomstora träd, då skulle troligtvis R^{2-} talet bli ännu högre.

En intressant slutsats man kan dra är också att det linjära sambandet är olika tydligt för olika arter. En av arterna som sticker ut genom dess konsekventa linjära samband är platan (*Platanus x hispanica).* Jämför man denna art med till exempel sälg (*Salix caprea)* är det en tydlig skillnad i hur det linjära sambandet mellan kron- och stamdiameter ser ut. Vad detta beror på och vilka slutsatser man kan dra av det skulle vara intressant att göra vidare studier på.

I denna rapport presenteras endast det linjära sambandet för 10 utvalda arter, samtidigt har konceptvalideringsmodellen genererat sambandet för alla arter planterade i Malmö stad. Denna helt nya data kan generera viktig och intressant kunskap samt besvara frågeställningar som inte lyfts i denna rapport.

Även modellerna Logistisk regression och Random Forest resulterade i en så pass låg F1-score att modellerna inte kan anses vara användbara för syftet att identifiera jätteträd. Modellerna i denna rapport har haft ett begränsat dataset att träna på. Studerar man F1-score så kan man se att de modeller som är balanserade presterar bättre, detta indikerar att det dataset modellerna tränat på ofta är

obalanserat, med andra ord är andelen jätteträd av den totala mängden träd så pass låg att modellen blir obalanserad och sämre på att identifiera jätteträden.

I denna rapport har Malmös träddatabas använts som träningsdata för maskininlärnignsmodellerna. Att modellen tränats på stadsträd från Malmö bör tas i beaktning. Sambandet mellan kron- och stamdiameter bör se snarlikt ut i hela landet, men faktumet att modellen tränats på just stadsträd är en begränsning om modellen ska kunna appliceras utanför urbana områden. I städer beskärs träd redan från en ung ålder, träden genomgår flera omgångar av uppbyggnadsbeskärning och säkerhetsbeskärningar samt diverse andra skötselinsatser vilket påverkar storleken på kronan. Dessutom växer många av träden i oförlåtande och hårda ståndorter vilket kan påverka trädets växtsätt.

Det bör även noteras att modellerna i denna rapport är grundläggande modeller som är förhållandevis enkla att genomföra. Det finns en mängd andra modeller som kan appliceras på samma data, vilket skulle kunna generera andra ännu mer intressanta och relevanta resultat. Att träna andra modeller på träningsdatan har inte rymts inom ramen för detta projekt men är något som bjuder in till vidare studier på området.

Vidare studier kan med fördel ta del av andra städers träddatabaser för att ytterligare utvidga den mängd data som modellen kan träna på, och öka generaliserbarheten. Det skulle vara särskilt intressant att träna modellen på enbart jätteträd. Sådan data kan med en liten arbetsinsats tas fram. Artportalen har ett register över särskilt skyddsvärda träd inrapporterade från fältinventeringar. Detta register bjuder in till möjligheten att sålla för att identifiera de träd som är jätteträd och sedan genomföra en fältinventering där man mäter krondiametern. För denna studie fanns inte utrymme att manuellt sålla bland dessa och sedan i fält inventera jätteträden, det skulle dock ha inneburit mycket värdefull träningsdata för modellen.

Datan som tagits fram för denna rapport, sambandet mellan kron- och stamdiameter för olika arter verkar inte vara lämpligt för att identifiera jätteträd. Däremot finns det en mängd olika användningsområden för denna helt nya data. För flera arter är det linjära sambandet mellan kron- och stamdiameter och så pass starkt (förutom för jätteträd) att modellen som tagits fram skulle kunna användas för inventering av trädstorlek i stor skala.

Framtida studier bör göras på området, dels för att utveckla modeller som kan identifiera och segmentera trädkronor, dels för att fortsatt utveckla modellen som presenteras i denna rapport som identifierar sambandet mellan kron- och stamdiameter.

Denna rapport är väldigt tekniktung och lyfter fram värdet av maskininlärning och andra moderna tekniker. För visst finns det ett stort värde i dessa, det är bara fantasin som sätter gränser för vad som skulle kunna vara möjligt.

Teknikens möjligheter till trots bör inte värdet av människor underskattas eller förringas.

Även om denna rapport hade generat en helt perfekt modell för att identifiera jätteträd behövs det fortfarande människor för att vårda och vårda våra svenska jättar för framtida generationer.

9. Konklusion och vidare studier.

Detta arbete har generat intressant och värdefull data som kan appliceras på andra områden än i jakten på jätteträd. Sambandet mellan kron- och stamdiameter är starkt linjärt men blir mindre så för jätteträd. Det innebär samtidigt att sambandet för mindre och medelstora träd är än tydligare. Här kan denna data till exempel användas för att uppskatta ekosystemtjänster enbart baserat på krondiameter.

Vidare studier bör göras på området. Dessa studier kan med fördel ta del av mer träningsdata. Denna modell har enbart tränat på stadsträd från Malmö, att få tillgång till ännu fler träddatabaser skulle generera i en modell med ännu mer värdefull data.

Artportalen har ett register över särskilt skyddsvärda träd inrapporterade från fältinventeringar. Detta register innebär att det finns en relativt stor mängd jätteträd som manuellt kan inventeras och mätas in för att på så sätt generera ännu mer högkvalitativ träningsdata.

Vidare bör urvalet av modeller som undersökt sambandet utvidgas från denna rapports grundläggande metoder. Modeller som bör appliceras på ett större dataset är till exempel SVR (Support Vector Regression), Generalized Additive Models (GAM), polynomregression samt andra mer avancerade djupinlärningsmodeller.

För att applicera mer komplexa modeller bör dock träningsdatan först kompletteras med ännu fler träd, både urbana och icke-urbana och från olika delar av landet, samt andra datapunkter.

Vidare studier på området samt tolkning av den existerande datan välkomnas av artikelförfattaren.

Referenser

Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27 (8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferraz, A., Saatchi, S., Mallet, C. & Meyer, V. (2016). Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment*, 183, 318–333. https://doi.org/10.1016/j.rse.2016.05.028

Fisher, A., Danaher, T. & Gill, T. (2017). Mapping trees in high resolution imagery across large areas using locally variable thresholds guided by medium resolution tree maps. *International Journal of Applied Earth Observation and Geoinformation*, 58, 86–96. https://doi.org/10.1016/j.jag.2017.02.004

Holst, C. (u.å.). Giant trees – inventory, conservation, management and recruiting. https://stud.epsilon.slu.se/11441/1/holst_c_171005.pdf

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60 (6), 84–90. https://doi.org/10.1145/3065386

Li, S., Brandt, M., Fensholt, R., Kariryaa, A., Igel, C., Gieseke, F., Nord-Larsen, T., Oehmcke, S., Carlsen, A.H., Junttila, S., Tong, X., d'Aspremont, A. & Ciais, P. (2023). Deep learning enables image-based tree counting, crown segmentation, and height prediction at national scale. Yortsos, Y. (red.) (Yortsos, Y., red.) *PNAS Nexus*, 2 (4), pgad076. https://doi.org/10.1093/pnasnexus/pgad076

Lidar 101: An Introduction to Lidar Technology, Data, and Applications (2012). NOAA Coastal Services Center. https://coast.noaa.gov/data/digitalcoast/pdf/lidar-101.pdf

Träd i Sverige: kulturhistoriska och samtida perspektiv (2024). Institutet för språk och folkminnen.

Mansour, Y. & Schain, M. (2001). [Random Forests]. *Machine Learning*, 45 (2), 123–145. https://doi.org/10.1023/A:1010950718922

Powers, D.M.W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. https://doi.org/10.48550/ARXIV.2010.16061

Samråd om åtgärder på särskilt skyddsvärda träd (u.å.). *Naturvårdsverket*. https://www.naturvardsverket.se/vagledning-och-stod/samhallsplanering/samrad-om-atgarder-pa-sarskilt-skyddsvarda-trad/ [2025-03-11]

Jordan, M.I. & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349 (6245), 255–260. https://doi.org/10.1126/science.aaa8415

Terese Besker; Pontus Svenson; Zohreh Ranjbar; Christian Vos; Jörgen Einarsson (2024). *T.AI.TRÄ Trollhättan stads - AI-baserade Trädinventering*. (RISE Rapport). RISE Research Institutes of Sweden. https://ri.diva-portal.org/smash/record.jsf?pid=diva2%3A1906949&dswid=-2790 [2025-03-11]

Vad är ett jätteträd? (u.å.). *Naturskyddsföreningen*. https://skane.naturskyddsforeningen.se/jattetrad/vad-ar-ett-jattetrad/ [2025-03-11]

Piepho, H.-P. (2018). A Coefficient of Determination (R2) for Linear Mixed Models. arXiv. https://doi.org/10.48550/ARXIV.1805.01124

Vos, Christian (2024). *Data Fusion of RGB Imagery and LiDAR Point Cloud to Study Tree Inventory in an Urban Forest using Deep Learning*. Kungliga Tekniska Högskolan. https://kth.diva-portal.org/smash/get/diva2:1941072/FULLTEXT01.pdf

Ytmodell från flygbilder (u.å.). *Lantmäteriet*. https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/ytmodell-fran-flygbilder/ [2025-03-11]

Östberg, J. & Rowicki, E. (2022). Standard för trädinventering i urban miljö Version 3.0. Svenska Trädföreningen.

Publicering och arkivering

Godkända självständiga arbeten (examensarbeten) vid SLU kan publiceras elektroniskt. Som student äger du upphovsrätten till ditt arbete och behöver i sådana fall godkänna publiceringen. I samband med att du godkänner publicering kommer SLU även att behandla dina personuppgifter (namn) för att göra arbetet sökbart på internet. Du kan närsomhelst återkalla ditt godkännande genom att kontakta biblioteket.

Även om du väljer att inte publicera arbetet eller återkallar ditt godkännande så kommer det arkiveras digitalt enligt arkivlagstiftningen.

Du hittar länkar till SLU:s publiceringsavtal och SLU:s behandling av personuppgifter och dina rättigheter på den här sidan:

- https://libanswers.slu.se/sv/faq/228316

☒ JA, jag, Karl Wanhainen har läst och godkänner avtalet för publicering samt den personuppgiftsbehandling som sker i samband med detta

10. Tack

Tack Roger Elg, SLU, för trygg och engagerad handledning.

Tack Johan Östberg, Kalle Ågren, och alla vänner på Trädkontoret för inspiration och handledning.

Tack Digitaliseringsenheten på Fastighets- och Gatukontoret på Malmö stad för data.

Tack min kära bror, Erik Wanhainen, för ovärderlig hjälp och teknisk support.