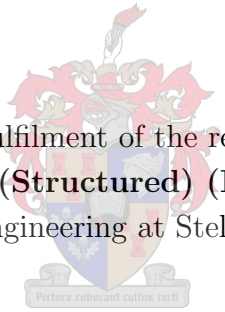
A predictive model for precision tree measurements using applied machine learning

by

Johannes Lodewyk Pieterse



Thesis presented in partial fulfilment of the requirements for the degree of
Master of Engineering (Structured) (Industrial Engineering)
in the Faculty of Engineering at Stellenbosch University

Supervisor: Dr GS Nel
Co-supervisor: Dr D Drew

April 2022

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: April 2022

Dedicated to my parents, Nicolaas Wilhelmus and Anne-Marie Pieterse for the hope and inspiration that they have afforded me throughout my life.

Acknowledgements

- My loving wife, Melissa for her unwavering support and commitment throughout the entire master's degree process. My family members for their patience and positive frame of mind towards advancing my academic qualifications.

- I would like to acknowledge and thank Dr Stephan Nel for his guidance as the main project supervisor. His comprehensive and consistent direction was key to many of the successful deliverables in my thesis. The relationship was mutually beneficial as I gained specific knowledge and further understanding into the subject matter from our discussions.

- I am also grateful for Dr David Drew in his capacity as co-supervisor on the project. His domain expertise and academic knowledge was instrumental for the identification of research opportunities within the forestry sector.

- The industry partnership with the South African Forestry Company SOC Limited (SAFCOL) deserves special thanks directed especially toward Johan Erwee, Christiaan Smit and Madeleen Algera for their constant professional support and expert guidance during the project.

- My group of colleagues at the Stellenbosch Unit for Operations Research in Engineering (SUnORE) for their exposure to interesting content during our sessions, discussing topics in the field of industrial engineering.

- The Department of Industrial Engineering at Stellenbosch University.

- Philip Bester van Niekerk for his general research advice and keen insights into the latest technological advancements with regards to the total wood products value chain. His quick validation of certain aspects within forestry domains brought about critical reflection and allowed me to work at a greater rate of progress.

Abstract

Accurately determining biological asset values is of great importance for forestry enterprises — the process ought to be characterised by the proper collection of tree data by means of utilising appropriate enumeration practices conducted at managed forest compartments. Currently, only between 5–20% of forest areas are enumerated which serve as a representative sample for the entire enclosing compartment. For forestry companies, timber volume estimations and future growth projections are based on these statistics, which may be accompanied by numerous unintentional errors during the data collection process.

Many alternative methods towards estimating and inferring tree data accurately are available in the literature — the most popular characteristic is the so-called *diameter at breast height* (DBH), which can also be measured by means of remote sensing techniques. The advancements in laser scanning measurement apparatuses are significant in recent decades, however, these approaches are notably expensive and require specialised and technical skills for their operation. One of the main drawbacks associated with the measurement of DBH by means of laser scanning is the lack of scalability — equipment setup and data capture are arduous processes that take a significant amount of time to complete.

Algorithmic breakthroughs in the domain of data science, predominantly spanning *machine learning* (ML) and *deep learning* (DL) approaches, warrant the selection and practical application of *computer vision* (CV) procedures. More specifically, an algorithmic approach towards *monocular depth estimation* (MDE) techniques was employed for the extraction of tree data features from video recordings (captured using no more than an ordinary smartphone device) and are investigated in this thesis. Towards this end, a suitable forest study area was identified to conduct the experiment and the industry partner of the project, *i.e.* the South African Forestry Company SOC Limited (SAFCOL) granted the necessary plantation access.

The research methodology adopted for this thesis includes fieldwork at the given site, which involved first performing data collection steps according to accepted and standardised operating procedures developed for tree enumerations. This data set is regarded as the "ground truth" and comprises the target feature (*i.e.* actual DBH

measurements) later used for modelling purposes. The video files were processed in a structured manner in order to extract tree segment patterns from the corresponding imagery. Various ML models are then trained and tested in respect of the basic input feature data file, which produced a *relative root mean squared error* (RMSE %) between 14.1 and 18.3% for the study. The *relative bias* yields a score between −0.08% and 1.13% indicating that the proposed workflow solution exhibits a consistent prediction result, but at an undesirable error rate (*i.e.* RMSE) deviation from the target output.

Additionally, the suggested CV/ML workflow model is capable of generating a discernibly similar spatial representation upon visual inspection (when compared with the ground truth data set — *i.e.* tree coordinates captured during fieldwork). In the pursuit of precision forestry, the proposed predictive model developed for accurate tree measurements produce DBH estimations that approximate real-world values with a fair degree of accuracy.

Opsomming

Die akkurate bepaling van biologiese batewaardes is baie belangrik vir groot bosbou ondernemings — die proses word gekenmerk deur die korrekte versameling van boomdata, deur gebruik te maak van gepaste opsommingspraktyke wat in bestuurde bosbou kompartemente uitgevoer word. Tans word slegs tussen 5 en 20% van die bosareas opgesom wat as 'n verteenwoordigende steekproef van die hele omhulde kompartement dien. Vir bosbou ondernemings is die beraming van houtvolumes en toekomstige groeiprojeksies gebaseer op hierdie statistieke, wat moontlik gepaard gaan met talle onbedoelde foute tydens die data-insamelingsproses.

Baie alternatiewe metodes om boomdata akkuraat te bereken is in die literatuur beskikbaar — die gewildste data punt (kenmerkend in bosbou) is die sogenaamde *diameter op borshoogte* (DBH), wat selfs ook gemeet kan word deur middel van afstandswaarnemings tegnieke. Die vooruitgang in meetapparate vir laserskandering is die afgelope dekades aansienlik verbeter, maar hierdie benaderings is veral duur en vereis gespesialiseerde en tegniese vaardighede vir die werking daarvan. Een van die belangrikste nadele verbonde aan die meting van DBH deur middel van hierdie laserskandering is die gebrek aan skaalbaarheid — die opstel van toerusting en die opneem van data is moeisame prosesse wat aansienlik baie tyd neem om te voltooi.

Algoritmiese deurbrake op die gebied van data wetenskap, wat oorwegend *masjien leer* (ML) en *diep leer* (DL) benaderings bevat, regverdig die keuse en praktiese toepassing van *rekenaarvisie* (CV) prosedures. Meer spesifiek is die algoritmiese benadering ten opsigte van *monokulêre diepte skatting* (MDE) tegnieke vir die onttrekking van boomdatafunksies vanuit video opnames (met nie meer as 'n gewone slimfoonapparaat nie) en word in hierdie tesis deeglik ondersoek. Hiervoor is 'n geskikte bosstudiegebied geïdentifiseer om die eksperiment uit te voer en die bedryfs vennoot van die projek, South African Forestry Company SOC Limited (SAFCOL) het die nodige toegang tot die plantasie verleen.

Die navorsingsmetodologie wat vir hierdie proefskrif aangeneem is, bevat veldwerk op die gegewe terrein en die eerste stap van die uitgevoerde data insameling was volgens aanvaarde en gestandaardiseerde werkingsprosedures wat vir boomtellings

ontwikkel is. Hierdie opgawe en datastel word beskou as die "grondwaarheid" en bevat die teikenfunksie (werklike DBH metings), wat later vir modelleringsdoeleindes gebruik is. Die videolêers is op 'n gestruktureerde manier verwerk om boomsegment patrone uit die ooreenstemmende beelde te onttrek. Verskeie ML modelle word dan opgelei en getoets ten opsigte van die basiese invoerfunksiedatalêer, wat 'n *relatiewe wortel gemiddelde kwadraatfout* (RMSE %) tussen 14.1% en 18.3% vir die studie opgelewer het. Die *relatiewe vooroordeel* lewer 'n telling tussen -0.08% en 1.13% wat aandui dat die voorgestelde werkstroom oplossing 'n konstante voorspellings resultaat toon, maar met 'n ongewenste foutkoers (RMSE) afwyking vanaf die teikenuitset wat verlang word.

Verder kan die voorgestelde CV/ML werkstroom model ook 'n waarneembare en soortgelyke ruimtelike voorstelling genereer onder meer visuele inspeksie (in vergelyking met die grondwaarheids data stel — *m.a.w.* boomkoördinate wat tydens veldwerk vasgelê is). In die strewe na presiese bosbou lewer hierdie voorspellingsmodel wat ontwikkel is vir boommetings (i.t.m. DBH beramings), die werklike waardes verteenwoordigend tot 'n redelike mate van akkuraatheid.

Contents

List of Abbreviations

Acronyms

AGBD	Above-ground Biomass Density
AI	Artificial Intelligence
ALS	Airborne Laser Scanning
ANN	Artificial Neural Network
APO	Annual Plan of Operations
BOP	Best Operating Procedure
BPMN	Business Process Modelling Notation
CCD	Charge-coupled Device
CI	Computational Intelligence
CNN	Convolutional Neural Network
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRP	Close-range Photogrammetry
CV	Computer Vision
DBH	Diameter at Breast Height
DL	Deep Learning
DNN	Deep Neural Network
DPE	Department of Public Enterprises
DPT	Dense Prediction Transformers
DT	Decision Tree
DTP	Digital Terrestrial Photogrammetry

FNN	Feed-forward Neural Network
FOV	Field of View
GAN	Generative Adversarial Network
GEDI	Global Ecosystem Dynamics Investigation
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
H-PLS	Hand-held Personal Laser Scanner
IoT	Internet of Things
ITD	Individual Tree Detection
k-NN	k-Nearest Neighbours
KLF	Komatiland Forests
LiDAR	Light Detection and Ranging
LR	Linear Regression
MC	Mannequin Challenge
MDE	Monocular Depth Estimation
ML	Machine Learning
MLP	Multilayer Perceptron
MLS	Mobile Laser Scanning
MSFA	Management of State Forests Act
NFL	No Free Lunch
NLP	Natural Language Processing
NN	Neural Network
PLS	Personal Laser Scanning
QMD	Quadratic Mean Diameter
R&D	Research & Development
RF	Random Forest

RGB	Red Green Blue
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RSA	Republic of South Africa
RSE	Residual Standard Error
RTK	Real-time Kinematic
SAFCOL	South African Forestry Company SOC Limited
SI	Site Index
SLAM	Simultaneous Localisation and Mapping
SOC	Stated Owned Company
SVR	Support Vector Regression
TLS	Terrestrial Laser Scanning
TSP	Temporary Sample Plot
UAV	Unmanned Aerial Vehicle
UMAI	Utilisable Mean Annual Increment
ViT	Vision Transformer

List of Figures

List of Tables

Chapter 1

Introduction

Recent advancements in mobile storage capacity, information transfer technologies, graphical processing units and high-performance computing have resulted in improved capabilities for the collection of an abundance of data in various formats. Increased data availability and the progress of data-driven algorithmic approaches — *i.e.* from the domain of ML — present the broader forestry industry with a significant value proposition. More specifically, the process of determining biological asset values stand to benefit considerably by this potential automation (*i.e.* enabled by improvements in data collection in conjunction with the noteworthy algorithmic breakthroughs made in the domain of ML).

Performing an inventory of forest compartments is certainly not a new operation within the forestry domain, however, improving the effectiveness and efficiency of gathering this data in an automated fashion and scaling the solution for larger areas remains an important challenge. The aim in this thesis is to investigate a new approach towards estimating basic external tree dimensions and more specifically DBH which is specified at 1.30m above the base of trees. In conjunction with this process it is also possible to infer total tree counts at larger scales of entire landholdings for companies that have forests under management.

The general convention adopted by commercial forestry enterprises in Southern Africa is to capture location data by means of manually sampling multiple smaller plots from isolated sub-areas of a designated stand. Statistical modelling methods are applied to the gathered sample plots, followed by an inferencing step in order to estimate — *a posteriori* — various other parameters in respect of the greater enclosed compartment. Even when digitising some of these aspects, the current standard practice remains markedly tedious, often neglecting key vital statistics (such as a complete tree count census for the encompassed area or exact DBH-height pair data), and is considerably labour intensive — which may contribute to recording inaccurate and/or erroneous data.

In conjunction with delineating the problem background and detailing the project scope and objectives, an appropriate literature review is conducted in order to provide a formalised reference to relevant research and a discussion thereof. The literature review aids the process of deriving the best practices of design and analysis techniques applicable to solving the problem addressed in this project — whilst grounding the approach in a good commercial understanding which takes business goals into account.

The research methodology adopted in the thesis at hand entails extensive fieldwork which was carried out by means of enumerating a large plot and performing standard measurements in respect of all the trees. This will represent the "ground truth" data — ultimately, serving as the training and evaluation sets of a comprehensive CV toolkit developed in this project. The basic idea is to employ a video recording device that moves above ground, but below tree canopies and captures observations at breast height. Collecting the data in the proposed structured manner, in conjunction with suitable processing steps, yields analytical insights and understanding of the problem at hand.

CV concepts are applied to the generated data sets which take the video feed as input, after which appropriate processing is performed by utilising *artificial neural network* (ANN) and DL based approaches in order to translate the data into information that contains meaningful insight. Central to this idea is MDE, which enables the virtual measurement of trees — the project involves devising a predictive model for the objective of generating accurate and reliable tree data — in pursuit of precision forestry.

1.1 Background

A concise definition of *artificial intelligence* (AI) as a concept is often described as machines (*i.e.* computers) that possess a level of intelligence or cognitive ability that emulates a human mind (to a limited extent) in respect of carrying out tasks such as learning, reasoning, and problem solving — a definition proffered by Russell and Norvig (2009) (1).

The aim of AI focussed studies is to enable a computer to "think" as a human would. In the current era, relevant technology is still considered as **weak AI** because machines are merely "acting intelligently" (1) and it is not used for general purpose tasks or activities. In order to reach the paradigm of **strong AI** computers will need to reason as people do and this is, arguably, achievable by "learning" without being explicitly programmed. As with intelligent human beings we acquire and refine our natural skills with learning from experience as opposed to clear instructions given to an individual.

Computational intelligence (CI) is characterised by operations performed within changing and complex environments — CI facilitates and enables the intelligent behaviour of networks or systems by means of adaptive mechanisms, as stated by Engelbrecht (2007) (2). The expressed networks include models that can learn to adapt to new scenarios, discover and classify situations in order to approximate and generalise the underlying abstraction. In Figure 1.1, a holistic taxonomy of AI and classification of subfields within the discipline are illustrated graphically.

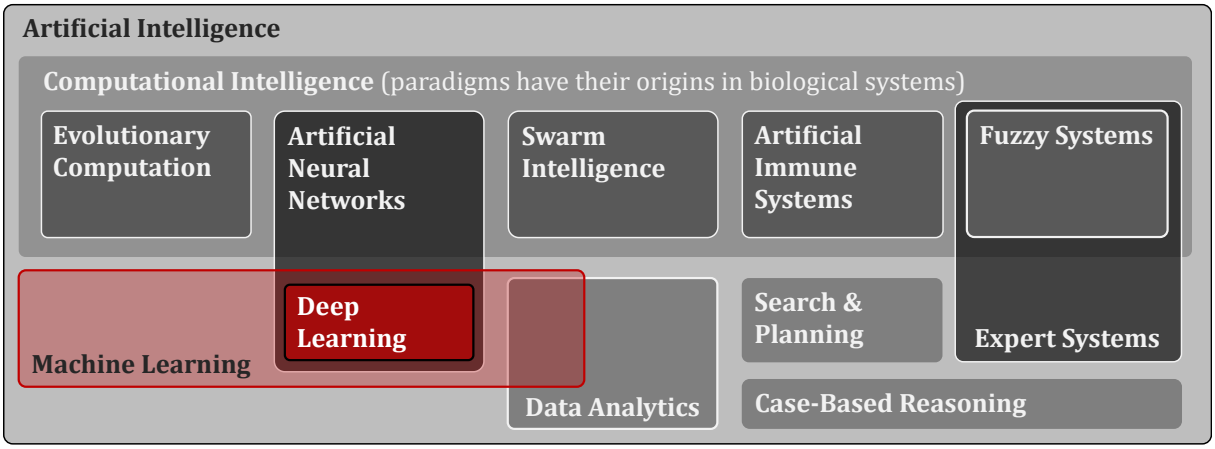


Figure 1.1: The taxonomy of artificial intelligence and associated sub-disciplines (2).

Emerging from brain modelling studies, an ANN is inspired by the neuronal structure of a human's biological brain (2). In simplified terms, this approach comprises of a large collection of activation nodes (*i.e.* neurons) that are inter-connected *via* an intricate communication network (*i.e.* synaptic links). DL, regarded as a subset of ML, adopts the architecture of ANNs comprising multiple layers of abstraction that can process and then functionally approximate input-output mappings with computational models designed to learn its representations, as described by LeCun *et al.* (2015) (3). This generalised approach enables an ANN to "learn" elaborate structures that have manifested in (big) data sets, according to which parameters of the network are adjusted in order to closely approximate a functional mapping of the desired input-to-output representation.

Shalev-Shwartz and Ben-David (2014) (4) report that the process of learning converts experience into knowledge and expertise — most ML models therefore represent manifestations of human, animal or environmental behaviours. The availability of a superabundance of data and significantly increased computational processing power are presenting opportunities for the application of different algorithms in order to build powerful predictive models which may aid decision making in various contexts, such as fraud detection, medical diagnosis, and quality control, to name but a few.

The limitations of effective ML models in production systems come at the cost of hidden technical debt, as reported by Sculley *et al.* (2015) (5). Even though the final ML code typically constitute a small part of an overall system, several risk factors need to be considered during the initial scheme design. In Figure 1.2, the focus is placed on the various supporting components that can be regarded with proportionate relevance toward an operational ML system — these aspects where considered during the initiation and development of this thesis.

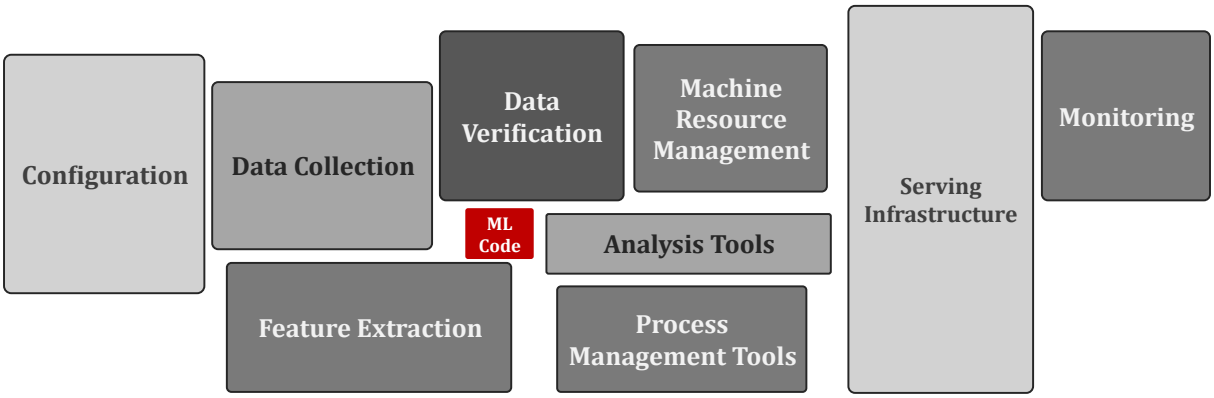


Figure 1.2: Hidden Technical Debt in ML Systems, as reported by Sculley *et al.* (2015) (5).

The proposed model in this thesis attempts to reproduce real-world DBH measurements using techniques in the realm of data science, which include ML and DL algorithms. In order to ensure data quality, the collection and validation of input features are done independently and do not rely on extracting the required data from other systems. Appropriate computing resources are employed which are suitable given the processing and analysis requirements.

1.2 Project scope and objectives

The proposed approach entailed conducting an experiment on a small area of trees, measuring tree characteristics and investigating whether it is, ultimately, possible to recreate approximately equal results but *via* alternative methods (from the realm of ML). The following four objectives are pursued in this project:

I Conduct an extensive academic review of the literature pertaining to this topic by:

- i Exploring algorithmic solutions in the domain of data science that encompasses ML and DL techniques (more specifically, CV with a keen focus on MDE procedures),
- ii study the field of forestry research to develop a clear understanding of the practical application domain at hand and

iii employ the same performance metrics utilised for remote sensing in forestry in order to evaluate the proposed solution on the same basis of measurement.

II Investigate the appropriate operational procedures for compartment enumerations and perform the research methodology requirements by means of fieldwork by:

i Gaining access to an appropriate forest compartment which is granted by the industry partner (SAFCOL) of this project,
ii record accurate tree measurements in-person from a forest compartment by means of a standard enumeration procedures and
iii in combination with information from the forest management system of the business, conduct an exploratory data analysis on the collected features.

III Record video files in a structured manner and then model accordingly in order to extract alternative feature data and tree segment patterns by:

i processing imagery in an ordered sequence through an DL network to produce MDE mappings,
ii develop bespoke algorithms to extract DBH line pattern data that corresponds to each picture frame,
iii employ clustering techniques to isolate data patterns that are relevant to single trees and
iv represent the results visually according to the spatial characteristics of the research study area.

IV Evaluate the suggested solution approach and report on the findings by means of:

i isolating a training data set proportionately that will be used by various ML models for an accurate input-output representation on the basis of least error and
ii test these models according to a standard evaluation criteria utilised for remote sensing proposals in forestry.

This project therefore involves a controlled study with labelled tree data used to identify only the DBH measurements, and connecting it to output information from the CV toolkit. The different facets of the project includes employing CV techniques, processing data sets offline and utilising applied ML practices towards estimating the DBH measurements — the accuracy of the proposed approach can be obtained by comparing the estimations from the ML approach with the real-world measurements. Accurate forest data is of upmost importance for good biological asset management — the value resides within the information that can be utilised for better decision-making. Proper silviculture, harvesting and transport activities can be derived from the availability of accurate information for forest compartments.

1.3 Contributions to data science

The views from Skiena (2017, 2020) (6; 7) are that data science and algorithmic design are emerging fields of study that amalgamates domain expertise, computer science and statistics so as to produce a meaningful value proposition. Ultimately, data science as a practice can help analysts develop intelligent computational models that have been verified and validated in respect of certain performance measures (6). The process will often lead to programs or applications that are of value to relevant stakeholders which are useful in respect of informed and/or automated decision making.

Businesses are currently collecting massive volumes of data and, together with the aid of analytical tools in the domain of data science, it is possible to utilise the value embedded within the data — standard analytical techniques are not suitable for extracting value from all the data. A widely accepted data science process, as reported by Zietsman (2021) (8), begins with an initiation of data mining, then modelling and its evaluation. Success from this approach has been proven and the deployment of these models in production systems can provide good solutions when dealing with data-driven projects.

The report by Zietsman (2021) (8) contains an insightful formula derived from a written comprehensive overview by Cao (2017) (9) for the process of *data-to-knowledge-to-wisdom*. This notion closely reverberates with the stated views of Skiena (2017, 2020) (6; 7) as a definition for data science. Accordingly, given data availability, suitable environmental factors and an appropriate high-level approach towards *problem thinking*, data science can be described as the combination (or summation) of management, sociology, communication, computing, informatics and statistics.

1.4 Thesis document outline

The Cross-Industry Standard Process for Data Mining (CRISP-DM) is a reference framework for handling a data project and is illustrated in Figure 1.3. The diagrammatic illustration shows the phases that constitute a data mining project — the life cycle encapsulates six main phases and is broken down in the flow diagram, with the directions of arrows in the figure representing the interaction and interdependence between the distinct phases. The diagram indicates a process that is continuous in nature, because a particular project might not necessarily be completed once a solution is deployed. In this case, the project lead might hand over the model to someone with a basic understanding of its workings in order to continually maintain the product.

The benefits of adopting the widely-known CRISP-DM framework is its independence from the reviewed industry data or technology, as stated by Wirth and Hipp (2000) (10). The specific

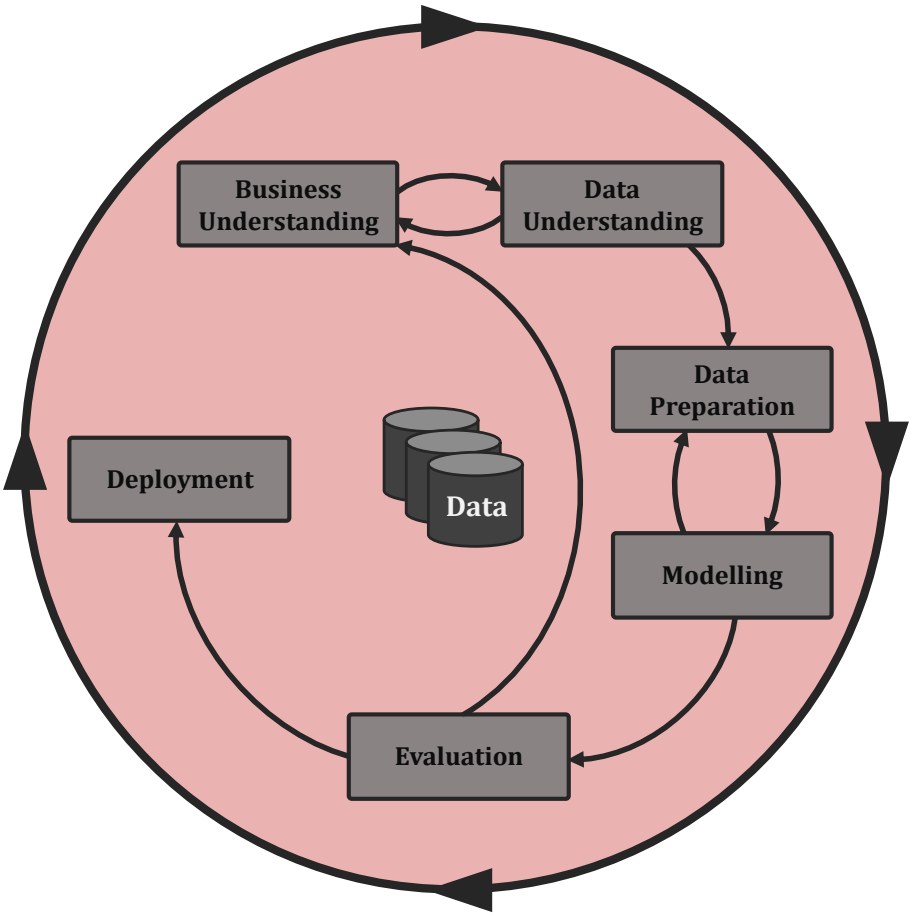


Figure 1.3: Framework illustration for Cross-Industry Standard Process for Data Mining (CRISP-DM), as adapted directly from Wirth and Hipp (2000) (10).

user of the process can be of lower skill level and might also not have as much time available as a dedicated data mining expert (10). Additionally, the suggested methodology is reproducible for almost any project and remains robust for numerous applications. A description of each of the different stages in the CRISP-DM approach follows:

- Business understanding — the starting point of the CRISP-DM life cycle involves determining business objectives and understanding the background of the relevant project or organisation. From this point, it becomes possible to define the scope into a problem statement, while using the appropriate terminologies, by formulating the findings into an initial project plan (10). Aspects that also need to be included are the end goals, objectives and anticipated milestones of the comprehensive project.

- Data understanding — the next phase involves collecting the initial data as well as preliminary review thereof in order to discover insights, screen for integrity issues and detect interesting subsets of the isolated metadata (10). There exists a dependency between "Business Understanding" and "Data Understanding" activities within the CRISP-DM

project life cycle (10) — accordingly, in order to formulate the initial project plan, a firm understanding of the available data is required.

- Data preparation — towards obtaining a final data set which serves as "input" for the modelling approach, the collected data must typically first undergo a data preparation phase (10). Hence, the bi-directional link between the preparation phase and the modelling node in the data mining life cycle, as displayed in Figure 1.3. Activities within this phase are likely to be performed multiple times, and in no particular sequence, which should enable a user to construct a suitable data set for final use (10). These tasks include data cleaning, feature selection and transformations so as to ensure the data is in an appropriate format for algorithmic implementation.

- Modelling — during this phase appropriate modelling techniques should be selected. This selection process ought to consider the specifications of the project's problem statement and objectives (10). In an ML-based project, multiple test designs can be generated on isolated training data in order to consider preliminary results. There exists a dependence between the "Data Preparation" and "Modelling" phase as some techniques require specific data input formats (10) in order to execute correctly. Once these models are assessed, the user could identify problems in the data and iteratively revise ways to construct the input set in a different manner to assist the comprehensive model framework.

- Evaluation — in this phase the specific modelling approaches are evaluated in respect of the relevant data with the aim being to compare how close the produced results are to the desired results (10). Before ultimately deploying the final model(s) in a production environment, however, it is essential to closely review the model and execution steps which led to these outputs — this ensures that the modelling approach achieved the predefined business objectives. From the data mining life cycle diagram in Figure 1.3 it is also noted that the "Evaluation" phase is connected *via* a feedback loop to the "Business Understanding" phase. This interlude is critical for the stakeholders towards determining if there were any important model issues that might not have been considered sufficiently — a process usually facilitated by consulting with the relevant domain experts and stakeholders.

- Deployment — this phase of the data mining endeavour does not necessarily conclude the project, as the life cycle (or parts thereof) can be repeated multiple times. Typically, this phase is not performed by the same person who first created the specific project steps, but rather by the customer or user that can easily reproduce the steps — ascribed to the fact that the approach is adequately notated during the initial phases (10). The deployment of said model(s) can be as straightforward as generating a simple report or running entire data pipelines, depending on the initial project requirements.

An enhanced CRISP-DM process model has been studied by Tavares *et al.* (2018) (11) where a variability-aware design approach is proposed to the data analysis process. In their study, the authors attempted to conduct a pre-assessment screening of input variables that compensate for the variability experienced in downstream models, by means of feature provisioning (11). The research also suggests the design and definition of a structured process which identifies data feature variability pre-emptively. Finally, an evaluation of possibilities pertaining to total process automation was investigated by the research team for the design frameworks.

Kristoffersen *et al.* (2019) (12) addressed the original CRISP-DM framework and highlighted issues such as missing phases which loops back to the business level and users not having any control over value contributed by the process. These problems were seemingly solved by the research, introducing an additional phase of "Data Validation" in between "Data Preparation" and "Modelling" steps. This investigation clearly emphasised the importance of re-incorporating business entities, or its domain experts, to append specific knowledge for the structuring and validating of data sets prior to modelling.

These aspects in data science suggest the proper approach towards completing projects. Furthermore, pitfalls are afforded the necessary attention in order to guide the process user in the correct direction. In this project, the basic CRISP-DM structure is adopted with limited amendments and compensations at certain stages of each desired milestone — listed as chapters in the current thesis outline.

1.5 Problem statement

If forestry enterprises harvest outside of an approved *annual plan of operations* (APO), it may have negative consequences and cause a chain reaction due to failing to reach a sustainable cycle of renewable resource management. The most important input variables when it comes to forestry remains the total tree count, DBH and height distributions or appropriate means and variances for compartments being surveyed. The reasoning behind tree counts is because a single tree could hold a lot of volume and if counts are wrong (*i.e.* limited to only a few), the effects could have a severe impact on volume predictability.

SAFCOL aims to enumerate between 10–20% of its landholdings because an overall greater sample size yields better results, but the industry standard is more in line with a lower threshold of 5–10%. More tree data will likely lead to greater precision, bearing in mind that the sampling design is another huge factor for accurate modelling. A hypothesis is deduced that there are insufficient measurement operations being conducted within forestry practices on a continuous basis. Ideally, obtaining 100% enumeration data of all plantation trees at a frequent

rate is sought — this approach could avoid many challenges and pitfalls stemming from current sampling methodology. A complete biological asset census is the ideal if it could be executed regularly and in a cost-effective manner, the question remains if the informational gain would be worth the actual expense and to what degree it is beneficial.

As the availability of individual tree data becomes an ever-increasing possibility, enabled through technology, the way in which making tree-level management decisions for forest stands will also need to be adapted as reported by Vauhkonen (2020) (13). This research entailed a simulation that was executed to determine the effects of DBH distribution errors that can impede operational options in a scenario of having the data readily availably. The findings points to the effect of management decisions based on erroneous information which can potentially cause losses between of up to 17% of immediate/future income derived from harvesting (13).

Practically, forestry companies in Southern Africa are recording enumeration data from plantations with the organisation's available resources (*i.e.* personnel and instrumentation), whilst considering cost implications and simultaneously being limited by capacity. The quality of these observations is always under scrutiny and having some form of past reference framework may be greatly beneficial (*e.g.* keeping the raw data files such as imagery from video for review if queries ever came up). It is conjectured that following an approach of recording and storing video files, indexed in the correct manner, will provide a reference or archive to revert back to, if the CV or DL modelling accuracy is ever questioned or needs to be recalibrated.

The manner according to which forest compartment data are collected and stored, together with the structure thereof, forms the basis of investigation and analyses carried out in this project. By having basic external tree data (*i.e.* DBH readings and thereby inferring total tree counts), but covering wide-ranging areas, could have positive benefits serving as input features for long-term forest management planning objectives.

Chapter 2

Literature Review

This chapter aims to bridge concepts induced from known challenges experienced by the broader forestry industry and the data science solutions proffered through findings available in current literature. Consequently, a practical application domain (*i.e.* forestry, specifically remote sensing for precision tree measurements) is identified which can be best served through the progressive developments in current data science technologies — utilising ML and DL. In the sections that constitute this chapter, the relevant background information that underpins the analyses carried out in this project are delineated.

2.1 Data science

Wolpert and Macready (1997) (14) formalised the famous "*No Free Lunch*" (NFL) theorem which is based on the principle that no single algorithm will consistently outperform each of its counterparts in respect of all problems. Across different data sets, and separate problems, the individual performances achieved by algorithms tend to differ greatly. The average performance, in terms of computational cost and model accuracy, is often regarded as the desired generalisation ability of the trained algorithms in respect of a particular problem class. Solutions should therefore be developed on an *ad hoc* basis and various ML/DL algorithms must be explored in order to find the best suited model.

Fundamentally, the notion underpinning NFL asserts that there exists no general-purpose algorithm that is capable of outperforming all other algorithms in respect of any given problem. Rather, various incarnations of some modelling approach ought to be fitted to the problem at hand in order to achieve adequate performance. Therefore, it is the responsibility of a data scientist (or analyst) to have the knowledge of which of these modelling approaches to use in order to best solve a certain problem. In the context of ANNs (and DL), this responsibility may be related to the exploration and investigation of different network architectures in respect of the particular problem. Additional to implementing the appropriate ML/DL algorithms, the

developer should explore and fine-tune model hyper parameters that may provide the best fit for mapping input-output abstractions.

2.1.1 Machine learning

The broader field of AI encapsulates ML as a subset which utilises an extensive range of statistical learning tools to explore and analyse data, as described by Kelleher *et al.* (2020) (15). Predominant paradigms of ML include *supervised*, *unsupervised* and *reinforcement learning*, as reported by Zietsman (2021) (8) and Nel (2021) (16) — they are summarised as follow:

- Supervised learning — the approach maps inputs to known and desired outputs, as stated by Ayodele (2010) (17), by means of a functional approximation that was generated by an algorithm typically presented with labelled data.
- Unsupervised learning — the outputs are not known or available beforehand and the model aims to find informative patterns (or structure) embedded in the data. The approach therefore involves unlabelled data and aims to employ methods such as dimensionality reduction and clustering, in order for some analyst to derive meaningful insight from the unstructured data, as discussed by Maimon and Rokach (2005) (18).
- Semi-supervised learning — a combination of labelled and unlabelled data sets are employed to conduct either supervised or unsupervised learning tasks (16). The concept is divided into two distinct learning models, namely *transductive* (*i.e.* attempting to infer accurate labels for unlabelled data) or *inductive* (*i.e.* pursuing to improve input-output mappings by including unlabelled data) learning.
- Reinforcement learning — typically an agent is tasked with navigating an environment during which different states are observed. An appropriate reward & penalty mechanism is devised (expressed as a policy) so as to prescribe the program to improve its performance in respect of achieving some goal (8; 9; 16).

One of the most basic architectures of a *neural network* is the so-called *perceptron* which refers to a single-layer network (*i.e.* all inputs are directly mapped to an output) as displayed in Figure 2.1 obtained from Nel (2021) (19). Additionally, a simple perceptron is described by Nel (2021) (16) as a linear binary classification algorithm.

Within such a network all the nodes (*i.e.* neurons) have *input signals*, as denoted by $[x_1 \ldots x_i \ldots x_n]^T$, which are scaled by so-called weights so as to transmit information signals through the network. Each connection is associated with a *weight* w_i value (considered an adjustable parameter) for the denoted communication link, with the purpose to scale its input variable (16). Artificial networks are able to alter these weight values in accordance with the

target feature, mapping to a pre-determined output in the case of supervised learning — during a process of *training* which is regarded as *model learning*.

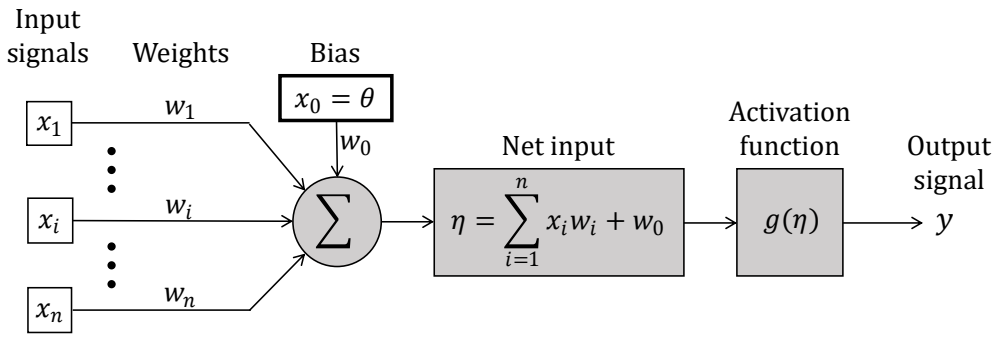


Figure 2.1: A common mathematical model of an artificial neuron, obtained from Nel (19).

The *weighted sum* of the input values (which may or may not include a bias value θ), denoted by η, is simply an aggregation and serves as input to a so-called *activation function*, denoted by $g(\cdot)$, which in turn, calculates the final *output signal* (16). Each input is associated with a weight value in this simple instantiation (20) and these weights are aggregated by means of an appropriate activation function — different activation functions are shown in Figure 2.2.

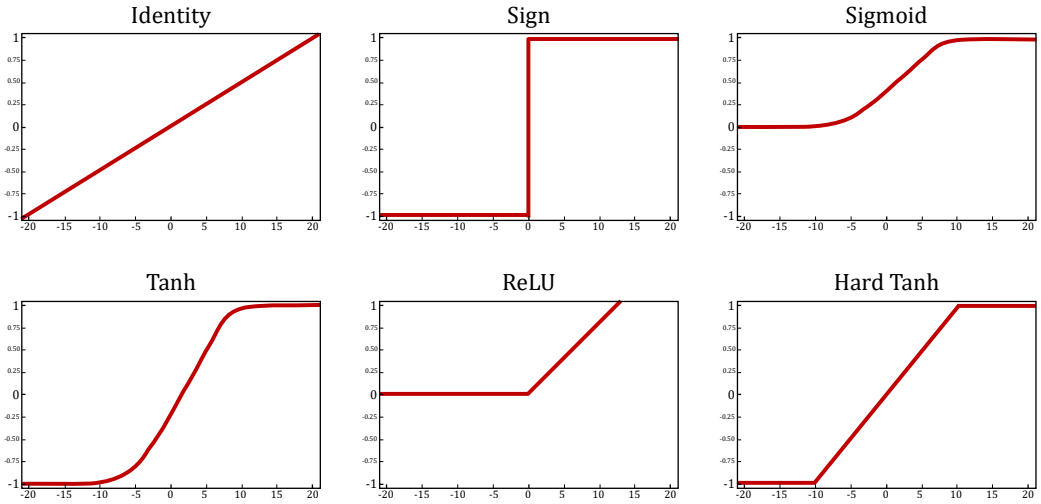


Figure 2.2: Various activation functions (20).

2.1.2 Deep learning

DL is a type of ML and usually operates with a *deep neural network* (DNN) which refers to an ANN architecture that comprises multiple hidden layers of nodes, each containing an activation function and input connections which generate a single output with its corresponding weight value. Often a single bias function node is added to each hidden layer which can be adjusted

either directly or *via* the corresponding weight value. DL will, throughout its network structure and processing of data sets, extract higher levels (*i.e.* dimensions) of features in a progressive manner. Conversely, a "*shallow*" network represents a single hidden layer of activation function nodes, which might include a bias neuron.

According to Goodfellow *et al.* (21), DL enables computers to program itself (similar to ML) from training procedures using many millions of examples as parameters. To date, DL has been proven extremely effective for solving an array of problems when only exposed to the rules or limits of the given problem and when trained with a large number of examples (21), as corroborated by Skiena (6). A depiction of a simplified deep network with N hidden layers is illustrated in Figure 2.3. DL-based approaches represent state-of-the-art solutions toward solving various classes of problems (*e.g.* CV or natural language processing).

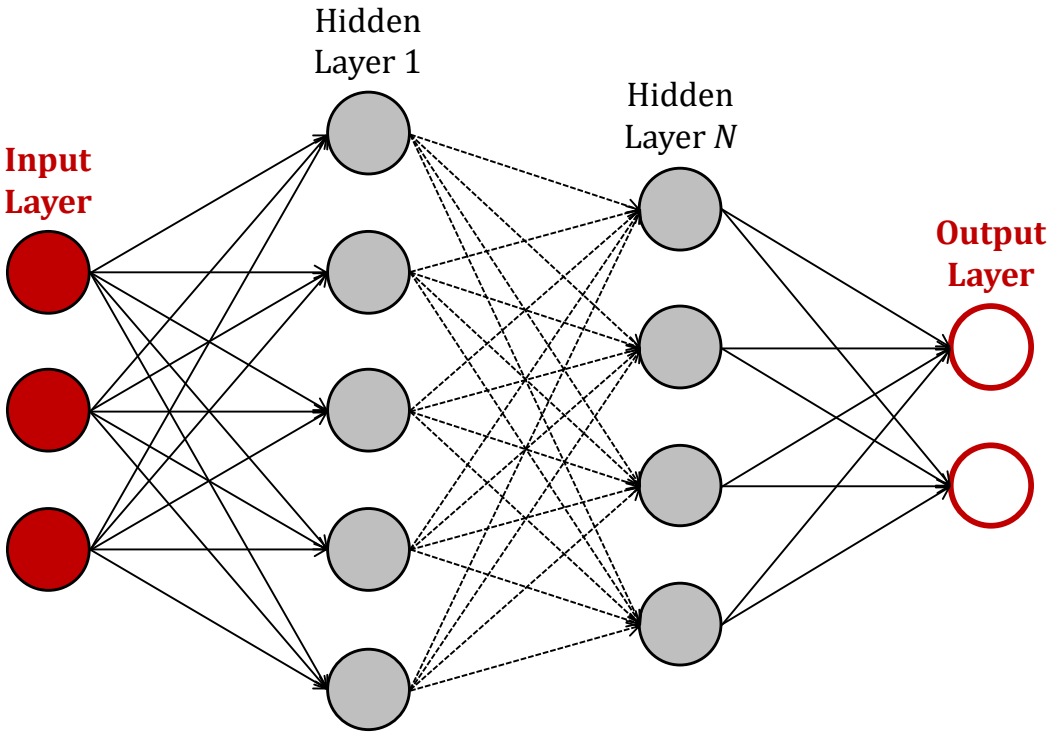


Figure 2.3: A simplified version of a DNN, adapted from Skiena (6).

Gradient descent is a simple, yet powerful, algorithmic approach adopted towards finding the best local optima of the search space and is employed as an optimisation technique that adjusts the model weight values by utilising network derivative information obtained using the method of back-propagation. There are different variations of gradient descent — *e.g.* stochastic (22; 23; 24), Adam (25), and AdaGrad (26), to name but a few — each exhibiting different performance dynamics in respect of various problems.

As an incarnation of dynamic programming, the method of *back-propagation* is instrumental in training ANNs effectively. After the weight values of a network have been established (*e.g.* by means of random initialisation), the derivative of a loss function can be computed with known activation functions at each neuron. The specific outputs are evaluated with respect to all the weights in the multi-layer network during the forward phase. Subsequently, the backward phase aims to learn the gradient step of the loss function (*i.e.* the anticipated degree of influence each update iteration will have over weight values) by using the chain rule in differential calculus. This produces an iterative weight update rule that can be used to fine-tune network weights in order to ensure the network generates an output that coincides with the desired target (19).

2.1.3 Computer vision

Many practical use cases for CV have materialised, as the concept entails utilising DL algorithms to process images or video in order to automate the extraction of useful insights. These advancements include text recognition (*e.g.* natural language processing and text analytics), object detection, photo enhancements and virtual environment reconstruction (*i.e.* localisation mapping) to name a few as reported by Szeliski (2021) (27).

Most notable perhaps is the technology's application towards creating self-driving cars and autonomous drones with obstacle avoidance capabilities. CV techniques along with applied ML are also deemed as a viable alternative towards automating data collection processes for measuring inventory accurately within the forestry industry — a matter discussed in more detail later in this report.

CV algorithms and applications entails mathematical techniques used to emulate the appearance of objects and represent three-dimensional shapes from imagery (27). This literature encompasses most of the computing challenges from CV such as geometric transformations (explaining model fitting and optimisation of scattered data interpolation), instance recognition/classification, object detection, semantic segmentations and feature matching (using edges, contours, lines and vanishing points) (27). Furthermore, the author includes a discussion on computational photography (*e.g.* super-resolution, texture analysis and synthesis) as well as motion estimation through translational alignment and optical flow models — leading to geometric intrinsic calibration that generate structure from motion and *simultaneous localisation and mapping* (SLAM) models.

Depth estimations are obtained by using epipolar geometry in conjunction with sparse or dense correspondence mechanisms, and through the utilisation of deep networks and global optimisation techniques (27). The work in this thesis revolves around a particular interest in MDE, which suggests that depth maps can be generated successfully from single-lens cameras

(*i.e.* monocular) with only input images — thereby avoiding the incorporation of a sequence flow (*i.e.* video logic formats). This favourable outcome is achieved by utilising a pre-trained DL network and its associated structures.

2.1.3.1 Monocular depth estimation

Ming *et al.* (2021) (28) present an overview of DL-based approaches towards MDE, their research summarises publications since the topic gained popularity from 2014 onwards. Monocular images or video does not require complicated equipment or professional operation (compared with binocular/stereo recording devices). It is noted, however, that because MDE trains with stereo vision data sets that there is not enough reference examples available for current DL models to sufficiently produce effective MDE predictions. The same depth results (for monocular imagery) are obtained by using DL networks trained on multi camera data sets with *Recurrent neural networks* (RNNs) (29), *generative adversarial networks* (GANs) (21) and CNNs which are the primary architectures utilised to achieve accurate MDE performance metrics when evaluated against publicly available data sets. The most commonly used data sets are the NYU Depth V2[1] or KITTI[2] reference data sets.

Regarded as a social media phenomenon, the *mannequin challenge* (MC) has made available thousands of viral internet videos during 2016 of people imitating mannequins — *i.e.* freezing in diverse, natural poses, while a hand-held camera tours the scene as stated by Li *et al.* (2019) (32). Their research catalogued these videos into the MC data set and provided a way to learn structure from motion and depths of moving people by watching frozen people (32). The study showed an improvement in performance against other MDE benchmarks and the approach has provided an alternative for recovering depth from non-rigid or dynamic video with examples of people contained in the data (an area of MDE that has not enjoyed overwhelming exposure in terms of training data freely available).

Widely regarded as the initial breakthrough for many subsequent MDE innovations, the work on single image multi-scale DL from Eigen *et al.* (2014) (33) has contributed tremendous value for general depth map predictions. Their research proposes a two-fold stack for delivering

[1] An indoor data set for MDE based on DL, provided by Silberman *et al.* (2012) (30) at the New York University. The repository contains 407 024 frames of *red-green-blue* (RGB) image pairs in conjunction with the Microsoft Kinect depth camera.

[2] An outdoor data set for MDE and object detection and tracking based on DL, jointly developed by the Karlsruhe Institute of Technology in Germany and Toyota Institute of Technology in the United States by Geiger *et al.* (2021) (31). The KITTI data set is captured through a car equipped with 2 greyscale and 2 high-resolution colour cameras, laser scanner and *global positioning system* (GPS).

depth estimations — namely an initial global coarse prediction on the entire image and thereafter another network that refines this output prediction locally. At the time of its publishing their study achieved state-of-the-art results on benchmark data sets such as NYU Depth and KITTI. The DL network uses a *scale-invariant error* loss function and is trained by comparing relationships between pairs of pixels, denoted in tuple form as (i, j), as output. The objective function entails finding the smallest error — each pixel pair in the depth prediction must be different by a magnitude similar to that of the corresponding pair in the ground truth. From Eigen *et al.* (2014) (33) the *scale-invariant mean squared error* (in logarithmic terms) is expressed as D with the following loss function

$$D(y, y^*) = \frac{1}{n^2} \sum_{i,j} ((\log y_i - \log y_j) - (\log y_i^* - \log y_j^*))^2$$

where y denotes the predicted depth map, y^* denotes the ground truth and n is denoted as data points indexed for pairs of i, j pixels.

The work of Eigen *et al.* (2014) (33) naturally resulted in significant interest in MDE because many other authors subsequently attempted to achieve state-of-the-art results using different model approaches and multi-scale structures — as with the research of Chen *et al.* (2016, 2019) (34; 35) where their own DL networks were developed by adding limited stereo vision reference sets and experimenting with novel network architectures (*i.e.* modular fusion and residual decoding). In a similar fashion, Godard *et al.* (2017, 2019) (36; 37) followed suit with their studies on proposed CNN structures incorporating epipolar geometry constraints for assisting their models in terms of generating disparity images. Later, the same authors also suggested the use of self-supervised methods in order to decrease reliance on increasingly complex architectures by means of using a minimum re-projection loss, full-resolution multi-scale sampling and auto-masking (which avoids invalid camera motion assumptions during model training).

Andraghetti *et al.* (2019) (38) propose an innovative vision odometry algorithm which outperformed existing approaches at the time of publishing their research. Their framework initially processes sparse depth predictions using an auto-encoder, then iteratively repeats the process to produce denser depth mappings. A consolidating algorithm then combines these dense outputs with the original RGB images and inserts it into a CNN network. The model employs self-supervised learning as an error checking mechanism on the ground truth stereo images and applies its loss function based on re-projection and consistency, the study evaluated results on the KITTI data set. At the time of writing, Wang *et al.* (2019) (39) have achieved the best performance results to date in regards to RMSE (*i.e.* 2.320) on the KITTI challenge using an RNN and unsupervised learning approach for MDE.

Great strides have been made towards improving MDE predictions, as presented in the work done by Luo *et al.* (2020) (40) who proposed a method to achieve precision and consistency for depth mappings. Their algorithm is able to reconstruct dense and geometrically consistent depth for pixels from monocular video. By means of quantitative validation, the research group could produce higher accuracies than previously known techniques.

Their approach is notably computationally expensive for the most part, in particular because of the algorithmic insertion step of sampling frame pairs. With these image pairs they first attempt to establish equal pixel correspondence using an optical sequence flow to check for consistency. The resulting deterministic correspondence parameters are then used to extract geometric constraints as a three-dimensional structure. Disparity and spatial attributes are used as a multi-objective loss function and the network weights are adjusted *via* standard back-propagation. After the fine-tuning stage is completed — ascribed as the network minimising geometric inconsistency error over multiple image pairs of the input video — the depth estimation results are delivered back in the original format (40).

The approach proposed in this thesis has taxing computational requirements because of the size of video data sets (further explained in the modelling chapter) and therefore literature was sought that addresses processing capabilities of MDE. Aleotti *et al.* (2021) (41) present research on real-time depth perception from single images using handheld devices in everyday settings. They investigate the resources that are needed to achieve real-time performance and explore various architecture options of MDE when balancing compute requirements, and reported good results. Interestingly, their research constantly refers to Ranftl *et al.* (2020) (42), in turn, as a solution with the best state-of-the-art performance for robust MDE on most common use cases (*i.e.* evaluated against NYU Depth V2 and KITTI).

The research conducted by Ranftl *et al.* (2020) (42) is called MiDaS and they propose an algorithm for mixing data sets during a zero-shot evaluation process. Their innovative zero-shot cross-data set transfer protocol — *i.e.* testing on data sets separate from training sets rather than randomly selecting from overall subsets (which conversely permits models to have some exposure to the final answer). This approach reportedly produced a system that works far better for MDE on real-world images and avoids unintended bias. Their impressive performance findings were also enabled by proportional image augmentation such as the horizontal flipping of frames during training. Furthermore, their models are freely available in the public domain (with the hopes to contribute toward practical applications) and these frameworks were used extensively in this thesis.

2.1.3.2 Vision transformers

With its origins stemming from *natural language processing* (NLP), Vaswani *et al.* (2017) (43) have devised a simple network architecture called the *transformer* which is based on a so-called *attention mechanism*, thereby disregarding the need for RNNs and CNNs entirely. During recent years, using transformer architectures in CV applications have become a more effective alternative — ascribed to notable performance improvements and speed of processing time — as reported by Dosovitskiy *et al.* (2020) (44).

A pre-trained vision transformer model can obtain the same state-of-the-art results as CNNs, whilst using significantly fewer computational resources. Reportedly, Dosovitskiy *et al.* (2020) (44) adopt the transformer framework from Vaswani *et al.* (2017) (43) and also disregard the notion of employing traditional CNNs for CV purposes. The model starts by appropriately segmenting an image into equal and standardised square patches — these smaller separations are subjected to positional transfers and linear embeddings throughout its network layers. The proposed technique solves the inductive bias issue, in proverbial terms, attributable to an overall improved generalisation performance. In the context of many examples, the proposed network learns from the training data and performs markedly well on classification tasks. Their model is most effective when pre-trained at sufficient scales, then transferred and applied to tasks with fewer data points.

The standard transformer framework (adapted from NLP) (43) takes in a one dimensional sequence of token[1] embeddings. To process two-dimensional images the pictures are reshaped according to a scope of real numbers (44), as denoted by $\boldsymbol{x}$ being an element of

$$\boldsymbol{x} \in \mathbb{R}^{H \times W \times C}$$

and serve as a sequence for the flattened two-dimensional patches denoted by $\boldsymbol{x}_p$

$$\boldsymbol{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$$

where (H, W) denotes the resolution of the original input image (*i.e. height* and *width*), C denotes the number of channels (*e.g.* RGB). Furthermore, (P, P) is the resolution of each of the separated image patches, and $N = HW/P^2$ is the number of patches (*i.e.* sequence length of the transformer).

[1] In terms of NLP, tokens can be regarded as the individual building blocks of written language. Tokenisation is common practice in NLP architectures as it is employed to separate a body of text (or corpus) in a format that makes sense logically. Therefore, tokens can either be combined phrases, characters, words or sub-words.

The transformer uses a constant latent vector (size D) throughout all of its layers, patches are flattened and mapped to D dimensions with a trainable linear projection (44). When compared with fully-convolutional networks, a *vision transformer* (ViT) provides better coherence for an entire image along with finer-grained detail as reported by Ranftl *et al.* (2021) (45). Their proposed *dense prediction transformer* (DPT) architecture delivers substantial performance improvements when evaluated against known benchmarks, especially if large training data sets are employed. Testing MDE metrics, a relative performance increase of around 28% is noted compared ot state-of-the-art CNN methods (45). One of the reasons for the superior performance — as opposed to its convolutional counterparts — is because a ViT avoids explicit dimensional downsampling (*i.e.* after initial image embeddings, transformers maintain the constant resolution throughout all processing stages).

Described as set-to-set models based on a self-attention mechanism, ViTs are particularly effective when constructed as high-capacity instances (*i.e.* able to ingest large quantities of inputs) and trained with an abundance of reference examples (45). The architecture is based on an *encoder-decoder* framework which is best suited for dense prediction models (*i.e.* making accurate estimations for the entire image frame as opposed to good localised performance often at the centre of pictures). Other literature has enjoyed similar success in terms of image analysis when utilising attention mechanisms in this manner (46; 47; 48). In the current thesis, DPT was selected for comprehensive modelling requirements because of its superior prediction quality and efficient runtime characteristics.

2.2 Forestry research

The well-known management consulting firm McKinsey & Company published a commercial report compiled by Choudhry and O'Kelly (2018) (49) concerning precision forestry and innovative technology trends across a holistic silviculture value chain. Their report describes current forest management practices as outdated, only more recently has the industry started to adapt to digital technologies.

The slow pace of adoption is attributed to limited private involvement as the majority of forest ownership resides with public entities — *i.e.* 76%, as reported in **The Global Forest Resources Assessment** (2015) (50). The conservative management style of state forests is a consequence of balancing diverse objectives around environmental, social and governmental goals. Nevertheless, with notable productivity improvements in adjacent industries such as agriculture (achieved through the adoption of digital technologies) (49), forestry companies have a renewed focus to implement novel systems in order to enjoy the same benefits.

A number of promising practices in the landscape of precision forestry include advancements in genetics, nurseries, silviculture (*i.e.* forest management), harvesting systems and tools used for efficient wood delivery. Markedly, mechanised harvesting and digital inventory are two areas where the most pronounced impact is being experienced when developing/implementing new technologies in this sector. The measurement of forest inventories, and collection of additional tree metrics such as species classification, log product mix and volume determination by means of remote sensing (49) are listed as specific important factors for precision forestry in the near future.

2.2.1 Enumeration importance

This section considers the various methods in which DBH measurements of trees are obtained using different instruments or devices. As an industry standard, tree diameters are recorded at breast height for an average sized person (*i.e.* 1.30m above ground), as depicted in Figure 2.4.

Calipers are normally employed to take two measurements per tree, rotated perpendicularly and averaged to compensate for ovality. Another common practice is to use flexible measuring tape around a tree in order to measure the circumference and then convert it to diameter using circular geometry. This is often regarded as a better way to generalise for ovality even though some trees might not represent a perfect circle. Clark *et al.* (2000) (51) report that foresters primarily use "contact" diameter tape in the form of measuring bands or rubbery rulers.

The value proposition offered by having precision tree data, mainly entailing external tree properties, is continually being recognised by commercial forestry enterprises, as reported by Drew and Downes (2009) (52). Their study explores devices used for tracking diameters historically, *i.e.* so-called *dendrometers* — these instruments are typically fixed at a certain tree height in order to record expanding ring growth over a long term period. This mechanism collects temporal measurements which can be conjoined with overlapping climate information in order to correlate growth within an array of environmental factors. Such efforts exemplify the need to accurately determine external tree features over time, and showcases the value it can add when combined with other relevant data sets, *e.g.* climate data. Corresponding spatial information, recorded periodically, could potentially produce a "template" (52) for growth predictions linked by surrounding attributes.

DBH is regarded as the industry standard when conducting measurements of tree diameter or basal area, but this type of estimation across vast landholdings is typically characterised by notable variance. A number of contributing factors may be ascribed to this phenomenon — *e.g.* diameters change as a person would measure the diameters of a tree at varying heights because of taper and various terrain factors could affect the exact ground level used as starting base.

Figure 2.4: Tree characteristics and depiction of DBH height relative to an average person.

The aforementioned practices are regarded as "direct" or "contact" measurements because the data collector would need to physically touch the tree and record the answer. Remote sensing, however, aims to infer the same measurements without direct contact whilst situated close-by or at a reasonable distance away from the tree.

2.2.2 Remote sensing

On a markedly bigger scale, *above-ground biomass density* (AGBD) calibration performed by means of space satellite imagery in collaboration with the *Global Ecosystem Dynamics Investigation* (GEDI) programme sponsored by NASA is a form of remote sensing, as reported by Kellner *et al.* (2019) (53). Their study combines GEDI information with high-resolution point cloud data from flying low-altitude drones that possess *light detection and ranging* (LiDAR) scanners. The approach introduces opportunities for remote sensing in forestry through the

validation and quantification of AGBD data used for demographic tree population analysis. Similar arguments are made by Guo *et al.* (2020) (54) according to which the potential of LiDAR for modelling ecological observations at various spatial extents and 3D resolutions is reviewed. This research group notes the commencement of a multi-dimensional big data era which, in conjunction with the consolidation of time-series information, present both challenges and opportunities for a greater ecological understanding.

Apart from boasting its potential benefits for the calculation of biological asset values, by means of calibrated quantification of forest plantations, remote and proximal sensing present even greater utility for operational applications usable in precision forestry, as reported by Talbot *et al.* (2017) (55). Forest accessibility, maintenance, infrastructure planning and construction are only a few areas that can enjoy advantages from having individual tree data and terrain models available. Harvesting systems can also derive value by utilising sensors or CV for autonomous navigation in forest machines (55). The authors do, however, caution that proximity or remotely sensed data are still experimental in forest operations and a number of problems ought to be addressed before wide-ranging adoption of the technology can take place. Continuous sensor information from forest machines is regarded as big data accessible through *internet of things* (IoT) platforms and, in combination with analytical approaches, provides vast opportunities for operational monitoring which can serve as a basis for decision support systems and evaluation.

2.2.2.1 Photogrammetry

The importance of valuing biological assets correctly is underscored by Mulverhill *et al.* (2019) (56) who state: "*Changing resource demands and climatic conditions require quick and inexpensive means of deriving robust and accurate forest inventory measurements.*" Their study explores digital terrestrial photogrammetry as a possible solution for large scale forest enumerations — especially a good fit for forest environments that exhibit changing and complex natural habitats.

In a study by Wells and Chung (2020) (57), stereo vision was employed and analyses were carried out by means of overlaying a virtual ground surface on an image of observed trees as illustrated in Figure 2.5, thereafter elevating the plane to an approximate breast height. Their method evaluated the RMSE in respect of 560 actual observations of tree data — the results of their research produced an RMSE on DBH measurements of 10.20cm and 13.36cm from a stereo camera at distances of 10m and 20m, respectively (57). Their study describes an interesting approach towards overlaying a flat ground-level plane and then elevating it to DBH level, but this procedure might be problematic for areas presenting varying slopes (*e.g.* on mountainous terrain or obscured with thick vegetation and rocks).

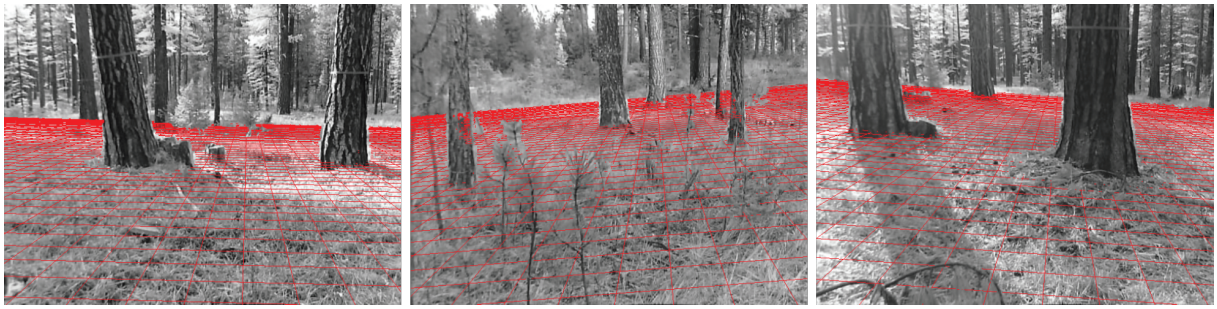


Figure 2.5: Projected ground planes detected with stereo image pairs by Wells and Chung (2020) (57).

In line with the methodology of the present thesis, and suggested by Marzulli *et al.* (2020) (58), cost effective hardware equipment such as smartphones can be used to capture similar tree data which are normally collected manually. Their research illustrates a working timeline comparison between the two methods (*i.e.* traditional forestry enumeration practices *versus* data extraction *via* photogrammetry) and showed around a 40% reduction in the total time required to derive the same information (including image acquisition and point cloud data set processing on 45 trees). They argue that traditional laser scanning technologies are expensive and that the same point cloud structures can be recreated using motion photography which, in turn, reconstructs a three dimensional space. The proposed technique achieved good performance in terms of DBH evaluation with a reported RMSE of 1.90cm (which is in line with industry best practices). Factors such as point cloud density and image scale, however, are mentioned as considerations when attempting this approach.

The findings from (58) are supported by Iglhaut *et al.* (2019) (59) who conducted a similar review using photogrammetry on aerial and ground-based systems. When evaluated against outputs from field and *terrestrial laser scanners* (TLS) in terms of DBH measurements, their study reported an RMSE ranging from 0.88cm to 6.80cm from sampled areas.

The aforementioned research also refers to Liu *et al.* (2018) (60) who reported good relative RMSE (*i.e.* RMSE %) statistics between 3–4.5% with designed observation instruments when combining a *real-time kinematic* (RTK) and *charge-coupled device* (CCD) through continuous photography. Similar performance has been reported by Mikita *et al.* (2016) (61) observing an RMSE less than 1cm by means of *close-range photogrammetry* (CRP) techniques for forest inventories.

Favourable results were achieved by Mokroš *et al.* (2018) (62) where a process of CRP was used around a tree at a radius distance of three metres away from trees, at various specified heights (*i.e.* 0.8m, 1.3m and 1.8m). The least error was obtained with a fisheye camera lens, stating the RMSE varied from 0.386cm to 0.596cm on diameter measurements (which is markedly

good in terms of accuracy). In their study, a scaling and orientation process was employed, enabled by coded markers which calibrated the reported results — it might not be practical, however, on a large scale to employ standardised markers across compartments in this fashion for data collection. The authors also supports the notion of a growing demand for precision forest information pertaining to individual trees (62).

Unmanned aerial vehicles[1] (UAVs) are experiencing greater utility towards measuring forest inventory, as reported by Seifert *et al.* (2019) (63). Their study reviewed changes in video capture parameters from drones — adjusting optical sensor resolution, image overlap and aerial vehicle altitude on the same forest compartment. Commercial software was used to extract and reconstruct a three dimensional point representation in order to understand the effects when changing these variables. The findings from their report concluded that higher resolution video data (achieved by flying at lower altitudes), along with setting a greater image overlap, generate the best reconstruction details and high-quality precision.

Dainelli *et al.* (2021) (64; 65) have performed extensive research with respect to the current advancements of UAV utilisation in forest remote sensing. The first part of their report (64) decomposes the general aspects analytically when using UAV devices in artificial (*i.e.* planted), semi-natural and natural forest ecosystems. The second part of their research (65) addresses the technical challenges of using UAVs for remote sensing in forestry.

A noteworthy observation from their research is that ML techniques are recognised as an important component necessary when processing markedly large image data sets collected by means of drones. Part two (65) addresses some research questions in respect of ML techniques used for remote sensing — particularly employing object detection algorithms and tree crown segmentation methods. A large part of their bibliography, in turn, reference to ML algorithms with applied techniques for automatic detection and segmentation of trees, or classification of plant *species*. ML regression models are often noted in this research (64; 65), in particular using *random forest* (RF) (66; 67), *support vector regression* (SVR) (68), *k-nearest neighbour* (*k*-NN) (69; 70) modelling and ANNs.

Even though classification of different tree *species* is of less relevance in the present thesis, a study by Nevalainen *et al.* (2017) (71) reported outstanding overall accuracy results (*i.e.* 95% and an F-score of 0.93) when classifying various *species* located at the same site. Their data sets were derived from hyper-spectral imaging and photogrammetric point clouds gathered from UAVs. The report also attempted individual tree identification across its eleven tested

[1] Commonly referred to as *drones*.

sites (accumulating 4 151 trees) but with less success ranging between 40–95% correctly, subject to area characteristics. The same results were achieved separately using RF and *multilayer perceptron* (MLP) modelling — MLP is considered a specific class of a *feed-forward neural network* (FNN). Similar research has been conducted by Briechle *et al.* (2021) (72) using a dual-CNN approach for standing tree *species* classification based on remote sensing data.

2.2.2.2 Laser scanning techniques

In terms of hardware, laser scanning is mainly divided into *aerial-* and *terrestrial laser scanners* (*i.e.* ALS and TLS), with some other studies referring to *personal laser scanners* (PLS) which are smaller devices operated by an individual directly. TLS was first introduced during the early 2000s as a means of allometric scaling for basic forest measurements, delivering diameter and tree height as reported by Calders *et al.* (2020) (73). This domain is described as a greater interdisciplinary field because of the on-going development of sensors and advancements made in algorithmic design for automated data processing — in the pursuit of understanding broader forest ecologies.

Interestingly, the authors mention the possibility of realistic virtual forests when radiative transfer modelling is employed for the purpose of monitoring large scale forest ecosystems. The research carried out by Calders (73), in turn, also commend techniques from the realms of CV, ML, and DL as potential solutions for processing large ALS, TLS, and PLS data sets.

Some promising results are reported by Wang *et al.* (2021) (74) who carried out a boreal forest inventory by utilising laser scanning technology on a UAV for close range remote sensing, and integrating under- and above-canopy point cloud data sets. The study encompasses RMSE and relative RMSE (*i.e.* RMSE %) as an accuracy metric on features of DBH and stem curve estimates. The investigation excluded erroneous estimates for DBH measurements and hence, the model could produce a RMSE result ranging between 2–4cm on a plot with an average DBH of 25.41cm (a heterogeneous area with approximately 200 stems/ha).

Hand-held PLS (H-PLS) devices are reviewed by Balenovic *et al.* (2021) (75) according to which they investigate the possibilities of utilising these type of systems for forest inventory applications — *i.e.* estimating the main attributes of trees such as DBH and height. These authors compiled a summary of other research on reported performances of various H-PLS and TLS systems in forest inventory studies, based on inherent *bias* and RMSE evaluation. They tabulated some state-of-the-art relative error results from different multi-scan TLS instrument combinations and tuning parameters, delivering some RMSE percentage results ranging between 3.5–13.4% for observed DBH measurements over a 5cm threshold.

Lightweight H-PLS devices are typically associated with high efficiency when compared with traditional TLS systems, and are less expensive to procure and operate than the aforementioned solutions (75). Even though the quality of data might not meet the high standards of TLS, attributable to hardware selection or inconsistent walking speeds of operators, H-PLS units provide a viable alternative to current TLS practices.

The constraints of data acquisition for TLS systems are addressed by Wilkes *et al.* (2017) (76) and they formulate a good strategy to follow in order to obtain proper geometric modelling metrics. Notably, spatial data sets (*i.e.* point clouds) are effectively derived form a 10m $\times$ 10m sampling grid over forest plots larger than 1 *ha* and required between 3–6 days in total to collect for the stated surface area.

Measurement precision also decrease with forest density as lower-quality results are reported by Wang *et al.* (2019) (77) for increasingly dense forest areas — stands in their study review are classified as: **Easy** (*i.e.* $\pm$700 trees/*ha*), **Medium** (*i.e.* $\pm$900 trees/*ha*) and **Difficult** (*i.e.* $\pm$2 200 trees/*ha*). In their study, comparisons were drawn between terrestrial and aerial point clouds where DBH values were determined during the data processing stage either manually or automatically — *i.e.* using algorithms on the basis of *individual tree detection* (ITD). Vastly different results are reported between the stated terrestrial (better) and aerial (worse) approaches, heavily influenced by the classification of separate stand densities. Denser plots are noted as the most challenging, however, manual intervention for terrestrial scans delivered the best overall performance (77).

Best in-class results are reported by Hyyppä *et al.* (2020) (78; 79) utilising a backpack *mobile laser scanner* (MLS) and UAV separately. At current knowledge, the lowest RMSE evaluation for DBH using ALS found in literature is 0.60cm (2.20%) in a sparse sample plot of 42 trees (79), obtained by flying a UAV with an attached laser scanner under-canopy. In an earlier study by Hyyppä *et al.* (2018) (80), which focussed on an investigation into frame-based depth sensors, the researchers compared DBH estimates against tape measurements and reported a good RMSE of 0.73cm which employed automated circle fitting algorithms — these accuracies are noted as operationally adequate. Fan *et al.* (2018) (81) employed the Levenberg-Marquardt[1] algorithm for point cloud analysis and achieved an RMSE of 1.26cm (6.39%) for observed DBH measurements.

[1]Commonly known as the damped least-squares method, the Levenberg–Marquardt algorithm is employed to solve least square error for non-linear problems. By fitting a least squares curve, the algorithm will attempt to produce the smallest error for minimisation problems (82).

An international benchmarking framework is proposed by Liang *et al.* (2018) (83) for an industry standard single- or multi-scan TLS procedure, on the basis of a relative RMSE which range between 5–10% for *easy* and *medium* plot densities. A slew of approaches are mentioned in this research and most notably the creation of varying algorithms that can be defined as aggressive, conservative or robust — *i.e.* the robust principles induces accurate estimations for parameters such as DBH while maintaining an improved detection rate, however, at the cost of algorithmic complexities.

The formulas for bias, bias %, RMSE and RMSE % have been adapted from all of the aforementioned research papers (58; 59; 60; 61; 62; 74; 75; 78; 79; 80; 81; 83) for use throughout this thesis and are defined as follow

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)$$

$$\text{bias }\% = \frac{\text{bias}}{\frac{1}{n}\sum_{i=1}^{n} x_i} \times 100\% = \frac{\text{bias}}{\text{mean}(x)} \times 100\%$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}}$$

$$\text{RMSE }\% = \frac{\text{RMSE}}{\frac{1}{n}\sum_{i=1}^{n} x_i} \times 100\% = \frac{\text{RMSE}}{\text{mean}(x)} \times 100\%$$

where $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3, \ldots, \hat{x}_n$ is denoted as predicted DBH values, x_1, x_2, $x_3, \ldots, x_n$ is denoted as actual/observed DBH values and n denotes the number of observations.

Liang *et al.* (2015) (84) argue that precision from their **stereoscopy applications** meant for forest inventories is surpassed by TLS and PLS systems because of additional geometric features obtained from laser scanning. The report states that image-based point clouds of photographic measurement are obtained with light-weight, low-cost equipment and is easy to operate but requires additional data processing afterwards. Laser scanning techniques have therefore proven to be superior when evaluating accuracy metrics between the aforementioned alternatives. The potential benefits of TLS systems to scan forest inventories effectively with low DBH error was reiterated by Liang *et al.* during recent years (2016, 2018, 2019) (85; 86; 87) where these reports track the progress of laser scanning technologies adapting to mobile and aerial platforms.

Similar work employing different TLS approaches and technologies has been carried out across forestry academia and industry partnerships as reported by Brolly *et al.* (2021) (88) where they used an automated voxel space algorithm for tree feature extraction, achieving DBH RMSE results in line with state-of-the-art methodologies. Chen *et al.* (2019) (89) also investigates PLS devices in combination with SLAM technology and reported an RMSE of 1.58cm

on their DBH estimations. This research notes increased efficiency when using PLS devices — the study reports a 30-fold productivity improvement compared to normal field survey practices when equal surface area is covered within an allotted time-frame.

2.2.2.3 Simultaneous localisation and mapping

A challenging aspect associated with forest enumerations involves establishing the precise geographical coordinates (*i.e.* geocodes) of individual trees. This is because of poor *global navigation satellite system* (GNSS) coverage under tree canopies as reported by Fan *et al.* (2018) (81). In their study, the MLS system was unable to provide globally-consistent point cloud data, however, a mobile phone with SLAM functionality was utilised in real-time as a viable alternative for the estimation of positioning attributes.

During execution of fieldwork for the present thesis, exactly the same problem was experienced — in that the GNSS sensor initially used to record geographic positions of trees was extremely erratic — ultimately geocode collection was avoided in favour of using an holistic SLAM approach.

In research performed by Pierzchała *et al.* (2018) (90), these authors reported the successful implementation of a graph-SLAM model in a forest environment. The resulting method could generate an accurate map of the sample area and produced stand characteristics such as single tree positions and DBH measurements. Furthermore, the RMSE evaluation on DBH metrics for this study was reported as 2.38cm (*i.e.* 9% relative RMSE). From this literature it is noted that the employment of robust SLAM algorithms is a cost-effective option and permits an acceptable quality standard for geographical forest mappings.

Chapter summary

The literature review commenced with a discussion covering topics prevalent in the domain of data science, including briefly discussing concepts utilised in ML, DL and CV. The field of depth estimation was identified as an area of particular interest for the research objectives pursued. Moreover, specific emphasis was placed on MDE because of the advantages of employing single lens (*i.e.* monocular) camera hardware in order to collect and process video imagery.

Transformers for visual processing represent a novel approach towards producing accurate depth predictions ingesting single image data files. The performance improvements offered from ViTs are notably better (and computationally less expensive) than other popular CV techniques.

The chapter also described the practical application background (*i.e.* forestry, specifically remote sensing utilised for accurate tree measurements). The appropriate performance metrics (*i.e. bias* and RMSE) in line with industry standards for evaluating the proposed methodology were also described in detail. The importance of enumerating trees in commercial forest plantations was high-lighted in order to set the stage for the business case of this project.

Chapter 3

Business Understanding

A good starting point of any data science project is grounded in a firm understanding of the business practices at hand. As per the CRISP-DM framework, the first step is to establish a suitable perspective that will enable and facilitate a deeper comprehension of the commercial domain under investigation. In the following sections, the primary aim is to describe the basic assumptions that underpin operational procedures when carrying out a forest inventory which provides further motivation behind the need for the proposed research. The problem statement addressed in this project is closely related to the requirements set out by the involved company and, naturally, the contributions made by the research conducted represent a matter of great importance to the industry partner.

3.1 Company background

SAFCOL was established in 1992 as a state-owned entity, under the Management of State Forests Act (MSFA) No. 128 of 1992 (91). In terms of the MSFA, the objectives of SAFCOL represent the long-term development of the national forestry industry according to accepted commercial management practices. The entity is an integrated forestry business operating in South Africa and has a smaller footprint in the neighbouring country of Mozambique too.

SAFCOL is controlled by the government of the Republic of South Africa (RSA) and forms part of the Department of Public Enterprises (DPE) — the organisational structure of the state owned entity is illustrated in Figure 3.1. SAFCOL conducts business through the sustainable management of plantation forests and its other assets. Revenue is generated from the sale of logs and lumber, as well as other non-timber related products. Komatiland Forests (KLF) is the main operating entity and generator of income within SAFCOL which oversees fifteen prime timber plantations listed as biological assets across Mpumalanga, Limpopo and KwaZulu Natal. KLF's commercial and non-commercial operations cover a land area of 189 747 hectares (ha), as shown in Table 3.1. The classification of SAFCOL's total 138 347 ha of plantable plantation

Figure 3.1: SAFCOL Organisational Structure (92).

area comprises primarily of pine, eucalyptus and wattle tree *species* (91). KLF also manages a Research and Development (R&D) facility, nursery and training centre within close proximity to Sabie, Mpumalanga, South Africa.

Table 3.1: Outline of operational areas managed by SAFCOL.

Operational Region	Plantable Area (*ha*)	Conservation Area (*ha*)	Total Area (*ha*)
South Africa (KLF)	120 870	68 877	189 747
Mozambique (IFLOMA)	17 477	65 070	82 547
Total SAFCOL Group	**138 347**	**133 947**	**272 294**

3.2 Forestry value chain

Upon considering the comprehensive life cycle of forests under management — from its seed to final products — SAFCOL invests considerable resources into R&D, specifically in the domain of genetics which, in turn, drives overall tree and log properties (91). These factors have an impact on the ultimate timber characteristics and wood features in the long term. For continual

and managed reforestation practices, trees start growing in nurseries where seedlings and cuttings are produced, to supply planting activities in compartments being re-established. Even before the new trees are physically planted, proper silviculture preparations take place (*e.g.* clearing and pitting) so as to achieve the correct espacement specifications.

Further important aspects include planning for the optimal *species* selection beforehand and plantation logistics to deliver new trees within the appropriate planting windows — *i.e.* usually performed during spring and summer months to increase survival rates before being exposed to colder temperatures in winter. For the duration of a forest management cycle the necessary biological protection should take place — weed, vegetation, fire and environmental management as well as control over pests and diseases are essential in order to produce high-quality timber material (91). Two critical management interventions, which add value to the product, are pruning and thinning — described respectively as the processes of cutting off branches up to a certain tree-height and felling selected trees (*e.g.* smallest trees) before compartment maturity in order to enable the remaining trees to grow better.

Finally, at the end of a tree's life cycle (time frames that are specified according to set predefined growing regimes), harvesting operations commence and the timber/biomass is extracted to roadside positions. The logs produced are directed to clients according to sales orders, by utilising the appropriate transportation systems suited for forestry operations. As an auxiliary business model, SAFCOL also permits its clients to pick up logs directly at roadside locations with the appropriate plantation access permits issued. The process of extraction is commonly separated into short- and long-hauling activities which facilitate the deliveries from roadside locations or depots, into saw mills.

3.2.1 Tree enumeration procedures

KLF manages 120 870 *ha* of plantable area within South Africa — comprising mainly pine tree *species* in Mpumalanga through sustainable forestry practices and conducts regular but limited compartment enumerations. Growth and yield models are used in conjunction with harvesting schedules to determine long-term sustainable volumes, and the acceptable annual cut volume available for consumption (*i.e.* raw material demand) by the broader South African saw-milling industry.

A comprehensive document that describes *best operating procedures* (BOP) for enumeration teams was obtained from SAFCOL and provides the outline for all the technical specifications necessary to effectively complete compartment enumerations on *temporary sample plots* (TSPs). Tree data drives one of the critical functions in forestry management, because it forms the basis of all current and future volume projections. Therefore, accuracy of measurements is of utmost

importance and the activity is conducted by specialised teams that are trained to collect tree measurements correctly, without diminishing the quality of work under various field conditions, in order to produce proper growing stock figures. Most of the enumerations are performed in parts of a plantation that are rarely visited by other employees, foresters or contractors. Some compartments may, for example, be heavily infested with weeds which increases the difficulty experienced by the enumeration teams when attempting to gain access to and work effectively in such locations. Essentially, forest inventory needs to be done in a fully randomised manner (by adhering to properly designed statistical methodologies) and challenges in accessing these compartments render the overall objective difficult.

Local forestry companies follow set procedures towards collecting and calculating its forest inventory — operationally, two-person teams are dispatched to pre-determined TSPs located according to a geographical grid system which pinpoints a centre spot for the unique TSP. Plantation managers will generate a working plan or instruction from the forest information system — directing the teams to enumerate only certain necessary compartments. By using a GNSS/GPS device, temporary plots are found in order to measure all trees within standard sample radii with approved enumeration apparatus (*e.g.* calipers for determining diameters). DBH measurements are recorded for every tree in the plot, after which the sample data are normalised (*e.g.* producing average DBH values and mean tree heights per TSP), and the practice is never reconciled on a "tree-by-tree" basis. The stated fixed centre point serves as an approximation for its surrounding trees, rather than the actual tree data that were measured during a previous enumeration round. Sampled data are therefore relative to the centre of the TSP — *i.e.* growth and yield features are tracked as an average dimensionality over time. Currently, this approach of data generalisation is regarded as a pragmatic solution in forestry because of the gargantuan scale of plantations which often comprises millions of trees.

The area of a TSP is typically 500m^2 (*i.e.* corresponding to a radius of 12.62m from the centre point), but plots can also be 300m^2 when enumerating smaller compartments. These aspects are influenced by a number of factors, including stand density and product classes. The number of TSPs in a certain compartment is dependent on the effective area of a designated compartment — usually between 5–20% of the total area is surveyed when considering the given geographical grid system. This standard may be true for SAFCOL but will vary quite markedly depending on management objectives and statistical design, a good practice would take into account sampling plans and population variance.

In Figure 3.2, a high resolution aerial photograph of a managed forest compartment (which was selected for study in the present thesis) is illustrated graphically. Enumeration concepts for the remainder of this exposition are elucidated using the same location's corresponding images. The specific compartment under investigation (shown in Figure 3.2) has an effective area of

Figure 3.2: High resolution aerial photograph of forest compartment under investigation.

2.82 *ha* and, upon taking the standard surface area of a sample plot into consideration, the relative number of TSPs located within the stand is six, as indicated in Figure 3.3. In this example, the specific compartment which holds six plots is derived from an effective area of 2.82 *ha* (28 200m^2) because 10% of the this area is approximately 3 000m^2. Therefore, six units of 500m^2 (*i.e.* set TSP size) will be a sufficient number of plots to represent an appropriate proportion from the samples, by standard forestry practices (to subsequently make inference of figures for the entire location). Generally, statistical theory pertaining to forest sampling methodology is sound, widely adopted and positive feature relationships can be derived when analysing data.

Infield enumeration work ultimately yield a data file containing tree counts for the sample plots — which provides an indication of tree density. The record also states a compartment identifier as well as a DBH reading for each tree measured. At random, a portion of the enumerated areas are audited by internal personnel as part of forest planner's responsibilities. Small deviations from audited data are permitted and are modelled on a sliding scale (*e.g.* larger locations are permitted to "absorb" added variances when compared to smaller areas). There is a difference between (1) percentage sampling area requirements according to the compartment size and (2) the extent of deviation percentage permitted when reviewing audited sample plots.

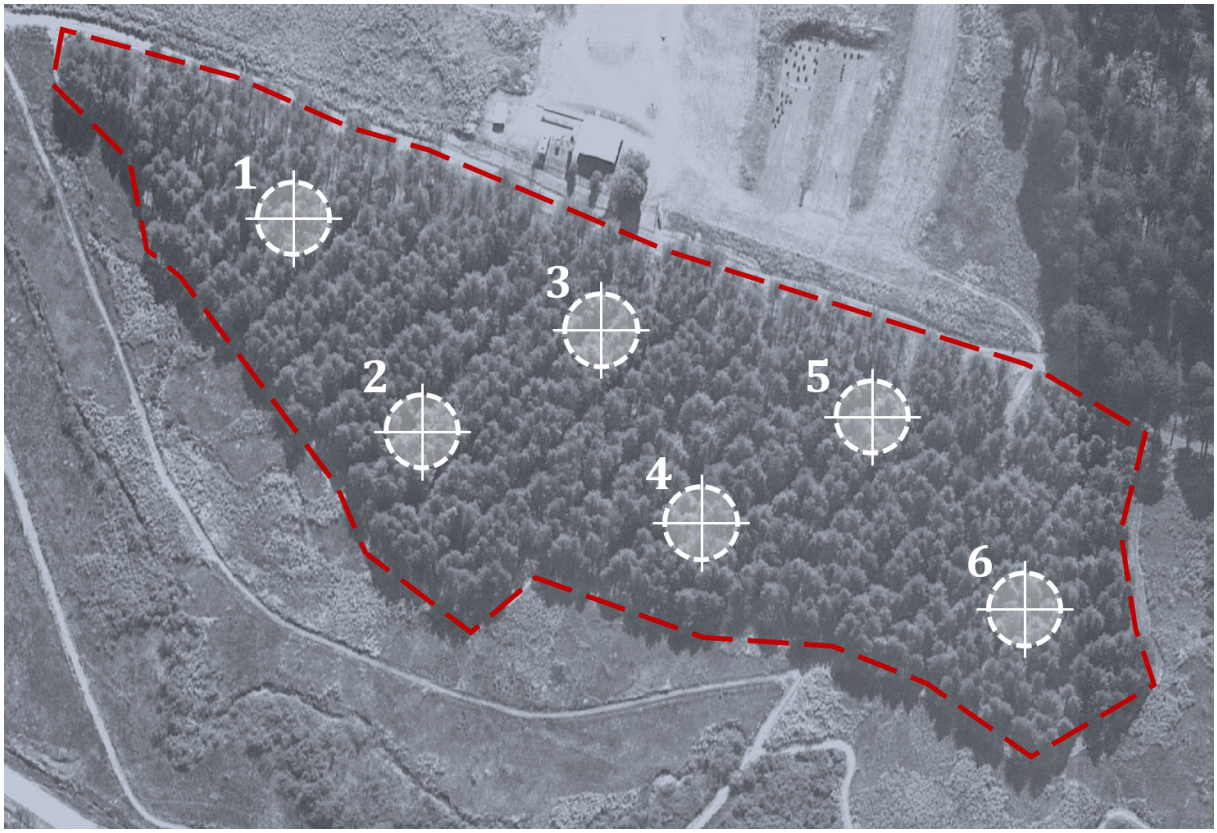


Figure 3.3: High resolution aerial photograph of the total compartment — showing the relative number of TSPs. The specific stand will hold 6 plots derived from an effective area of 2.82 *ha* (*i.e.* 28 200m^2) and 10% of the this area is approximately 3 000m^2 therefore 6 units of 500m^2 (*i.e.* standard TSP size).

Practically, however, plots are adjusted toward the middle of a specific stand — *i.e.* situated away from boundaries in order to simplify TSP placements in view of radii estimations. Consideration for factors such as the known "edge effect" of compartments are therefore noted, as reported by Wise (2013) (93). Normally, boundary trees perform better in terms of volume or biomass growth compared with inner-located trees — attributable to additional sunlight enjoyed by these trees. Edge trees, however, do not necessarily possess higher quality or strength properties — these characteristics are derived from the extent of natural resistance trees are exposed to within the immediate surroundings over its entire lifetime (*e.g.* plant spacings, wind exposure, etc.). Trees located on boundaries are often lopsided because of bigger branches toward open spaces, leading to reaction wood which is undesired by processing facilities.

It is important to realise, however, that sample plots can (and should) fall over boundaries in some cases which cater for an inclusion of edge effect trees. In some practices TSPs are (incorrectly) moved when they fall over the boundary — the mirage technique discussed by Kleinn (2007) (94) has been proposed in this regard to avoid systematic error. This principle allows for border plot corrections and entails "mirroring" the section falling outside of the forest boundary, then "doubling" the tree data available from the true inclusion area (*i.e.* observing

the same trees twice). In the current thesis, all trees from the designated compartment are measured for DBH and therefore avoids the inherent bias that can result from operational TSP placements which often avoids edge/boundary trees.

3.2.2 Compartment inventory valuation

The accuracy of plantable areas improve through many iterations of re-mapping boundaries of plantations from detailed data, such as direct ground observations, usually consolidated on an annual basis (SAFCOL surveys approximately 17% of their total area annually). Forestry activities are reported on the central management planning system, thereby further enhancing the reliability of comprehensive growing stock data. It is apparent that the biological asset data are of utmost importance — as the level of plantation management, degree of harvesting instruction and direction for logistics is derived from the information at hand.

Compartment information of the site selected for the present thesis, gathered from the forest management system, is provided in Table 3.2. According to a set regime, the specified area has already undergone completion of four pruning activities at ages 3.4, 4.5, 5.6 and 6.6 years — *i.e.* cutting off branches up to a height of 2m, 4m, 6m, 7.5m, respectively. Furthermore, the first thinning operation occurred at age 11.3 years which reduced density to approximately 585 trees/*ha*. At the time of writing, an upcoming second thinning is planned which aims to decrease the total remaining trees towards the 450 trees/*ha* threshold. This silviculture activity is a pre-emptive measure towards obtaining optimal expected volumes at clear-felling age.

Table 3.2: Total area information of the selected compartment for the research project.

Compartment Parameter	**System Value**
Species	*Pinus taeda*
Working Circle	Pine Sawlogs
Effective Area	2.82 *ha*
Established Date	2005-01-01
Planted Espacement	3m × 3m
Planted Trees per Hectare	1 111 stems/*ha*
Utilisable Mean Annual Increment	13.4 m^3/ha/year (calculated)
Site Index	27.1 *metres* (enumerated)

Site index (SI) refers to the average height of the top 20% of the thickest trees (*i.e.* DBH) located at a single compartment, gathered from TSP enumeration data. *Utilisable mean annual increment* (UMAI) is a system performance indicator and defined as the expected production volume at the age of measurement — calculated as all the timber that a grower is able to sell,

which excludes the stump and tip of a tree left over after harvesting. Noteworthy remarks from the forest management system, which also ties back to the general enumeration procedures of SAFCOL, indicates an average DBH of 23.87cm for the stand collected from an activity labelled: "*Thinning Control Cruised: 2015-10-12.*" A separate observation in respect of the system's interface is: "*Thinning Control* ***Audit*** *Cruised: 2015-10-21*" (*i.e.* 9 days apart from the initial control with an average DBH of 21.57cm).

These activities exemplify the described audit process applied as a control measure for the operational enumeration teams — the variances in measurement conducted in quick succession are representative of the variability that can result from variable factors such as TSP centre spot determinations by separate teams and error in manual measurements.

These phenomena may be regarded as variances that stem from the process of measuring trees physically and within a short period of time. The corresponding error in this regard is addressed by McRoberts and Westfall (2016) (95) in their research on propagation of individual tree volume and models for predicting large-area volumes. Essentially, their study showed error derived from uncertainty of individual tree data was negligible on large-area volume estimates during random sampling. Conversely, however, modelling parameters (*i.e.* stratified estimators) proved non-negligible and produced significantly different results — thereby reducing sampling variability effects. Model uncertainties are generally ignored in practical forestry applications, resulting in optimistic volume expectations for large-area estimates (95). Their findings support the notion that close attention should be paid during statistical methodology design and model parameter selection, rather than individual tree measurements from sample plots.

A biological asset valuation study has to represent figures as reliably as possible when compiling reports — these statistics are often used by insurance companies and investment firms in order to evaluate risks and returns of a relevant forestry company. They use high-level valuation methodologies, however, it has little to do with operational enumeration activities on the ground, but rely heavily on the data produced by these measurement teams for relevant model inputs. In turn, planning departments should provide accurate information upon requests with updated data sets. As required by invested stakeholders these valuation reports should outline timber prices, compartment volumes and costs associated with extracting these logs (*i.e.* the three input variables that serve as key drivers of the resulting valuation model). Sales prices and costs for harvesting are fairly straight-forward but the "volume" is a significant variable in terms of quantifying an accurate output amount.

3.3 Research methodology

The aim in this section is to describe the approach adopted towards collecting the necessary data and information for performing the analyses in this project. A forest area was first identified which was based on its suitability for executing fieldwork objectives — details pertaining to the compartment under investigation are provided in Table 3.2. The fieldwork entailed conducting measurement activities that aligned with those completed by nominal operational enumeration teams (*i.e.* tree counts, diameter measurements and height recordings).

Additional data points were gathered which represent the spatial outlay of each tree over a designed grid-system — serving as a high-resolution standard of measure (*i.e.* on centimetre scale). An exploratory data analysis and brief informal time study are reported on later in this thesis, during which the research findings are illustrated visually and practical considerations pertaining to the data collection process are described.

3.3.1 Approach

This project centres around a data science approach which involved physical fieldwork at a suitable forest compartment selected according to ease of access, favourable (*i.e.* flat) slope characteristics and other practical factors that contributed towards an effective experimental design. Appropriate access permissions were obtained from SAFCOL in order to conduct the proposed experiment at the desired site. In Figure 3.4, the same aerial photograph of the compartment selected for the study is illustrated graphically and in which the research plot sub-section is contained — highlighted in red on the right side. The site is located close to Sabie, Mpumalanga, South Africa.

Even though the total area of the compartment is 2.82 *ha*, only a smaller sub-section of the stand was utilised for the project. Interestingly, the sub-sectioned area was surveyed to be 0.70 *ha*, which represent 24.8% of the bigger enclosing management unit. A fixed road located between these sections was used as the boundary to separate the workload and informational spread during the experiment. Only a sub-section of the stand was used for research purposes (as displayed in Figure 3.4) and the exact records isolated from the forest information system are unknown for the encompassed study area.

A macro-grid was devised and set up (in-person) at the research site which served as a spatial identification layer during planning and measuring of tree attributes. The diagrams depicted in Figures 3.5 and 3.6 illustrate the approach adopted towards isolating the compartment's sub-section, and subsequent shape reorientation. Respectively, the two figures indicate the planning phase of the fieldwork which entailed forming lanes according to parallel lines adjacent to the

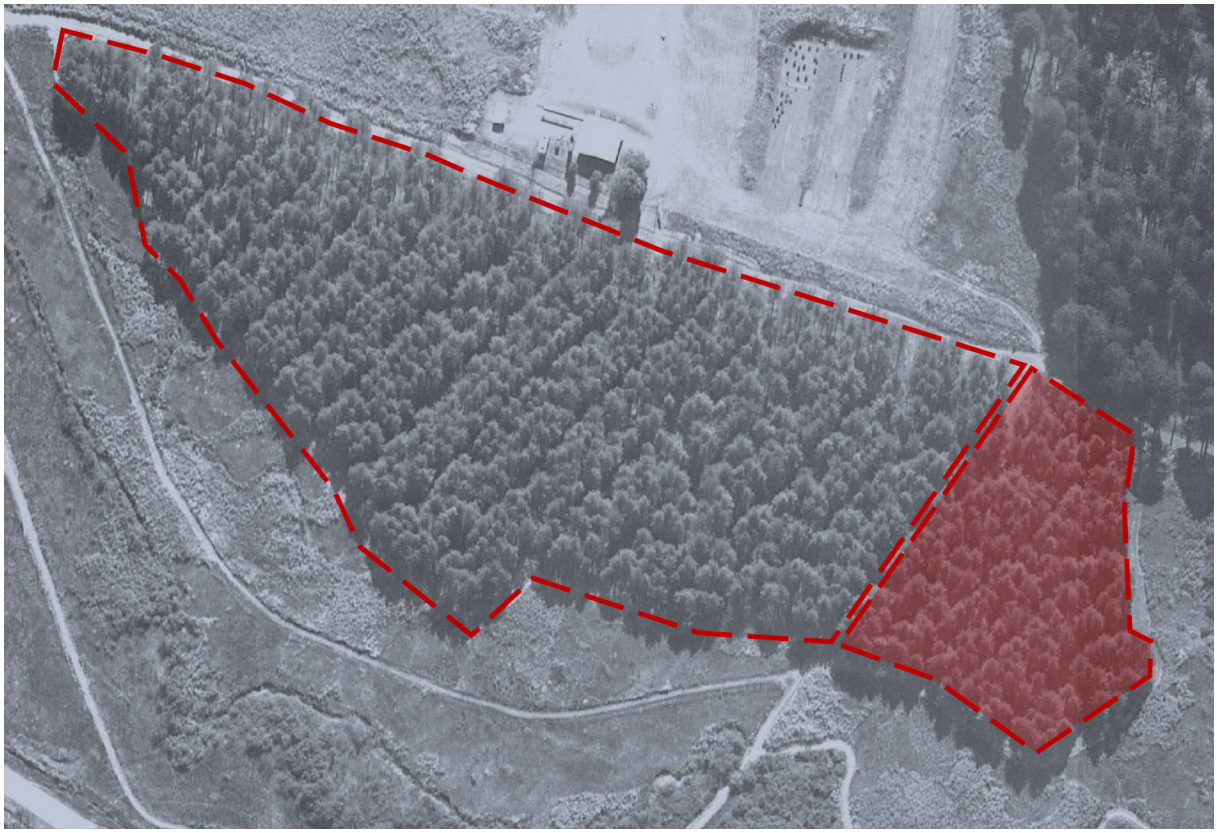


Figure 3.4: High resolution aerial photograph of greater compartment, in which the research plot sub-section is contained — transparently highlighted in red on the right side.

aforementioned forest road that separates the sub-section from the greater encompassing compartment. The lanes were designed to be five metres apart and physically running in parallel, but smaller micro-grids (*i.e.* 5 × 5 metres) are contained within these longer lanes as well. The number of parallel lanes that were able to fit into the research section equated to sixteen — denoted alphabetically from letters A to P, as displayed in Figure 3.6.

3.3.2 Fieldwork

The data set produced based on the fieldwork carried out represents the ground truth — it is envisaged that an appropriate algorithmic approach should be able to approximate a functional representation of the abstractions within the data — recommendations for model deployment are therefore also addressed in the following discourse.

The virtual lanes were translated to physical lanes — a photograph taken of the research plot is shown in Figure 3.7 which illustrates the tape (high-lighted transparently red) that was used to physically demarcate the plot. Upon closer inspection of the pegs that hold up the demarcation tape, it can be observed that each pin is located squarely every five metres apart

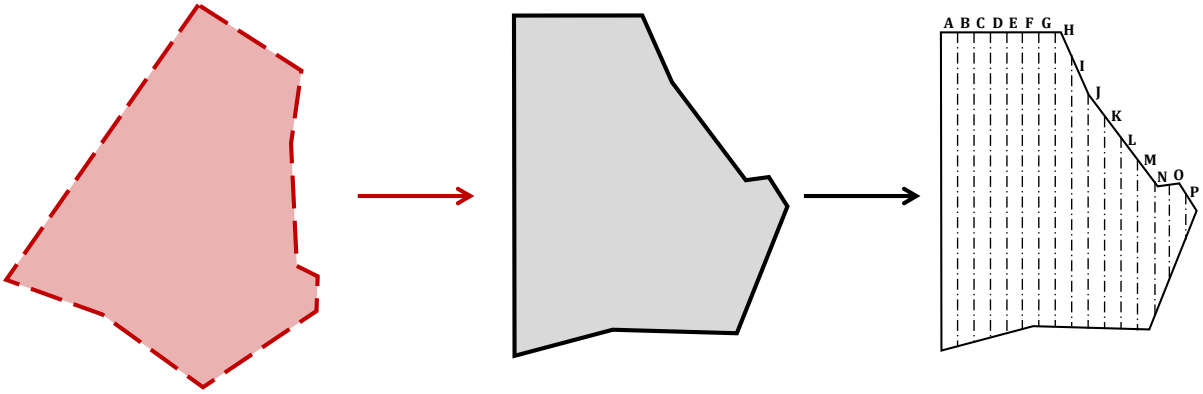


Figure 3.5: Virtual representations of compartment section under review — isolated and re-orientated, thereafter vertical lanes are added every five metres apart in parallel and labelled as illustrated.

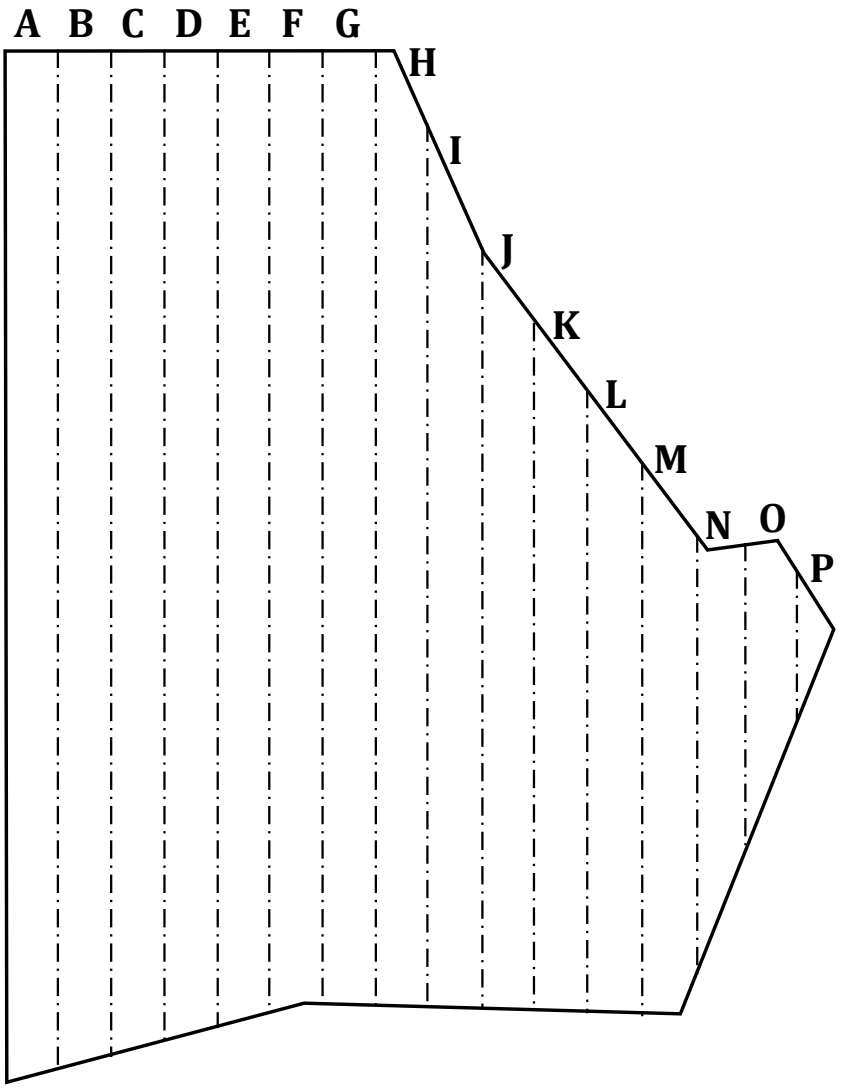


Figure 3.6: The research plot is designed such that it comprises virtual lanes — orientated in vertical formation in parallel at five metres apart and labelled accordingly.

which permitted the accurate placement of the inner micro-grids in turn (attributable to the perpendicularity of the pegs in respect of its nearest corners and constantly five meters away from each other).

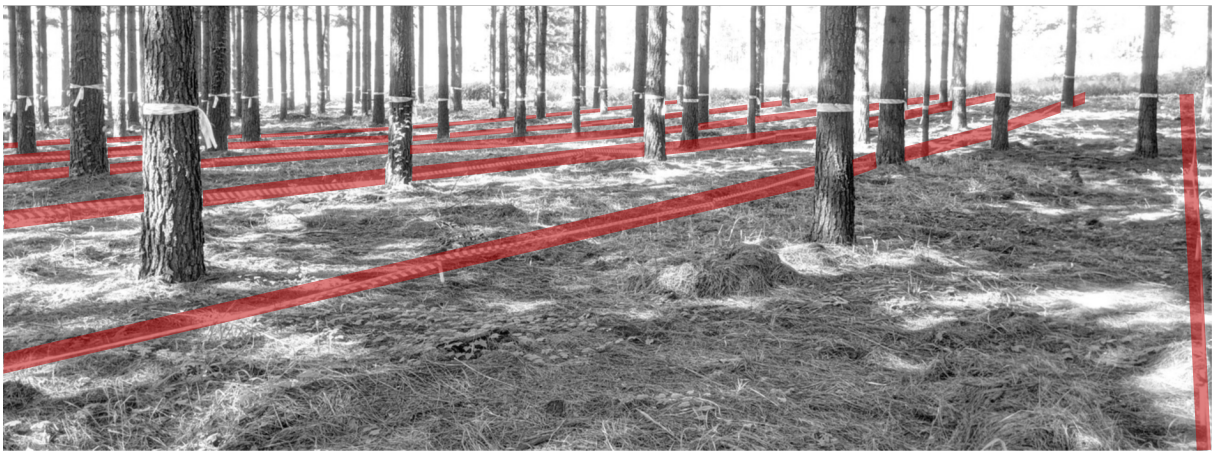


Figure 3.7: Demarcated areas and depiction of macro-grid layout in fieldwork plot of research project.

Apart from the designated lanes structured in a vertical formation, as described in Figures 3.5 and 3.6, horizontal rows were also introduced running across the research plot (if the lanes are regarded as virtual columns). This approach enabled the micro-grid strategy according to which individual trees were located spatially within a 5 × 5 metre block — these blocks were also appropriately named and their labels are illustrated in Figure 4.4. According to this structured approach each block could plot all the trees contained within its own boundaries and permitted a high-resolution representation of the spatial characteristics of the study area.

During data capture, the horizontal and vertical coordinates of each tree were recorded and later subjected to an offset (according to the block label) to reconstruct an overall macro-grid localisation map, which incorporated all trees from the study section. The number of trees that forms part of the experiment equated to 298 and each tree's characteristic data points (e.g. DBH and heights) were connected via the unique labels that were attached to each tree. As shown in Figure 3.7 each tree has demarcation tape tied around its bark at DBH level, with an attached tree identification label.

Chapter summary

Ending off the first phase of the CRISP-DM framework (*i.e.* Business Understanding), this chapter commenced with a description of the company background of the relevant industry partner involved with the project (SAFCOL). Sections addressing the general forestry value chain,

BOP for tree enumerations and the inherent commercial problem statement was discussed in order to provide context for the matter under investigation.

Materials and methods pertaining to the devised research methodology were also discussed in this chapter. In particular, the approach toward physical fieldwork was delineated and visually illustrated in an effort to appropriately describe the planned endeavour for collecting the necessary data points. These data are reviewed and visualised further in the next chapter by means of an exploratory data analysis, understanding and subsequent preparation (of separate video files) to serve the modelling phase of the CRISP-DM process.

Chapter 4

Data Understanding and Preparation

In this chapter, the data understanding and preparation phases of the CRISP-DM framework are discussed in detail. Recall that the collected data represent the information collected from physical fieldwork (which is regarded as the ground truth — *i.e.* the DBH measurements from individual trees represent the target feature during calibration modelling). Furthermore, the constructed data set contains video files recorded at the same study area which is to be utilised in a structured manner in order to derive the same DBH results by means of deploying CV, DL, and ML techniques available from the domain of data science.

4.1 Data understanding

The aim in this section is to describe the data that were collected during fieldwork from the designated sub-section of the greater encompassing compartment under consideration. This particular phase in the CRISP-DM framework is derived from the business understanding step and needs to be reconfirmed for corrections with the relevant stakeholders and domain experts before proceeding onwards. Data understanding mainly entails finding some initial insights from the data sets that were gathered, and it also serves as an intermediate process step towards the next phase of data preparation, which in turn is necessary for the appropriate modelling purposes thereafter.

4.1.1 Exploratory data analysis

The DBH measurements for all individual trees located in the study section were recorded — accordingly, 298 trees were measured twice with a standard caliper (by rotating the apparatus perpendicularly and later averaging these two measurements for the final feature data entry) as well as separately with a rubber band (*i.e.* centimetre tape) in order to collect the circumference

readings for each unique tree. Theoretically, these two independent methods of enumeration should closely produce the same measurement result (when the circumference readings are converted to diameter by means of circular geometry). In fact, the correlation was tested between these similar features and the linear relationship was calculated as exactly one (indicating a perfect positive correlation between the separate approaches). For the purposes of exploratory data analysis, the converted circumference measurements producing DBH readings were selected as the functional feature because this aspect is a better representation when considering tree ovality, and as a result of the video recordings observing individual trees from multiple perspectives.

Illustrated in Figure 4.1 is the continual DBH density distribution of the 298 trees with a mean for DBH of 30.59cm (indicated with the vertical dashed line). Rather than using the normal arithmetic mean DBH as a statistic, the quadratic mean is often calculated in forestry practices as a better indication of central tendency as reported by Curtis and Marshall (2000) (96). The formula for *quadratic mean diameter* (QMD) is expressed as follow

$$QMD = \sqrt{\frac{\sum_{i=1}^{n} DBH_i^2}{n}}$$

where DBH_i is the observed DBH values and n is the number of observations. Essentially, the calculated QMD (*i.e.* 31.21cm) sets a greater weight on larger DBH values — Figure 4.1 also indicates the QMD value as a vertical dotted line.

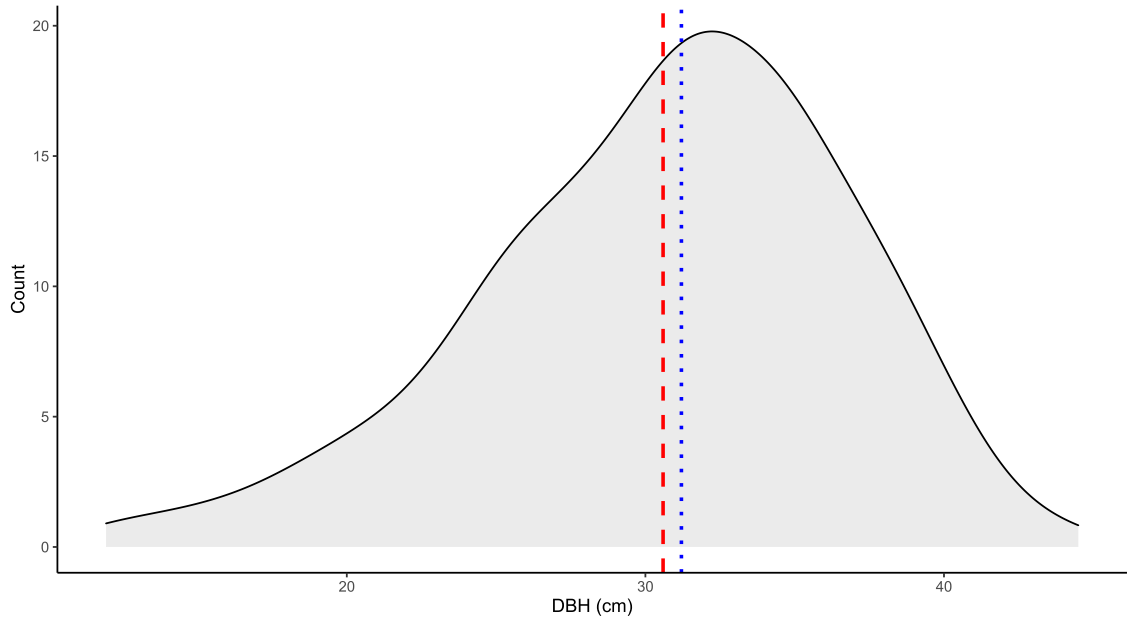


Figure 4.1: Density plot of DBH distribution from research area with a normal arithmetic mean of 30.59cm (red dashed line) and QMD of 31.21cm (blue dotted line).

It should be noted that older forestry procedures typically group DBH readings together within its closest **odd** centimetre (*e.g.* 11cm, 13cm, 15cm, 17cm, 19cm, *etc.*) for the primary reason of practical generalisation and simplifying the feature's cardinality. Modern precision forestry practices, however, considers the exact measurements from trees on a centimetre scale or even on a millimetre scale which is aligned with the reviewed literature pertaining to LiDAR technologies, as reported by Hyyppä *et al.* (2020) (78; 79).

Even though it is of less relevance in this thesis, the height of each individual tree was recorded using a vertex device during fieldwork data collection. The correlation between DBH and heights equates to a positive relationship of 0.60, as shown in Figure 4.2, which is normally regarded as weak. It is also noted that the height measurements are recorded in metres whereas the DBH readings are measured on a centimetre unit scale — this type of distinguishing unit measurements are standard practice in forestry enumeration activities (51). Calculating the *residual standard error* (RSE) for the linear regression model between DBH and height data produced a value of **1.838** as an initial benchmark.

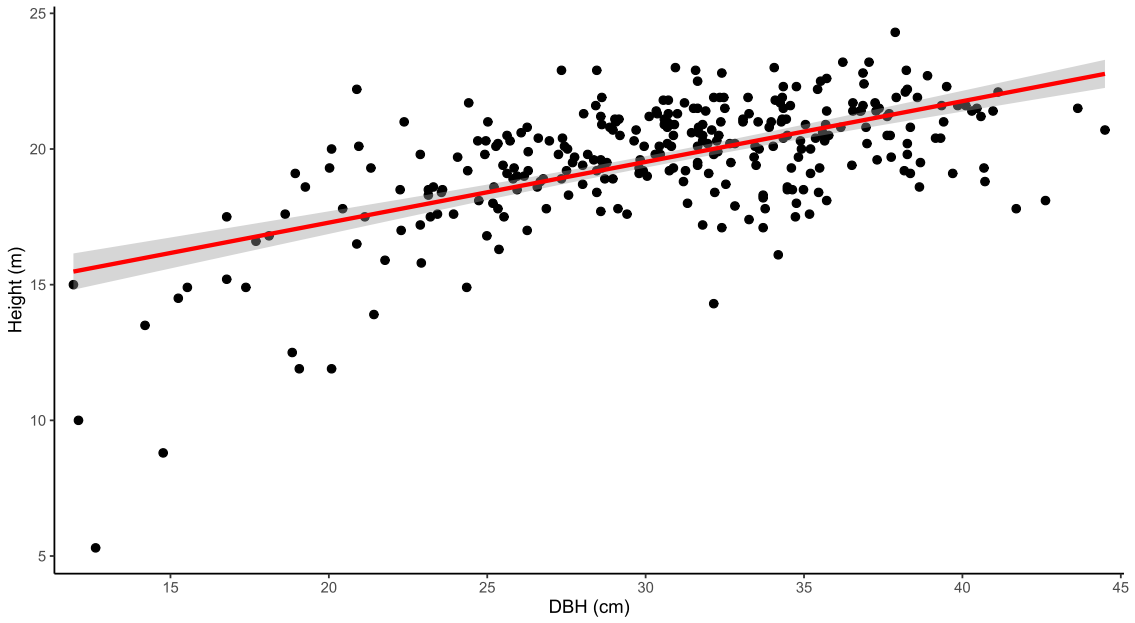


Figure 4.2: Linear regression model scatter plot between DBH and height measurements from the fieldwork data set.

Using a strict linear model might not necessarily be the best fit for a prediction when evaluating DBH measurements against its associated height readings, as reported by Arabatzis and Burkhart (1992) (97). Coincidentally, their research also studied *Pinus taeda* (which is the same type of tree in the current thesis) and included 175 observations of loblolly pine *species*. Their review of model formations estimating DBH-height paired relationships found that utilising a non-linear model performed the best overall (given that the data set was initially subjected to

random sampling). Therefore, a local polynomial regression fitting approach was evaluated on the same DBH and height data, as displayed in Figure 4.3, which delivered a better result with an RSE of **1.638** — a notable improvement when compared with linear regression. It should be noted that this sample of 298 trees represent a small fraction of all the trees in plantation forests. Therefore, the non-linear model cannot be applied for an entire population of trees because other compartments are different in DBH size, age and *species*.

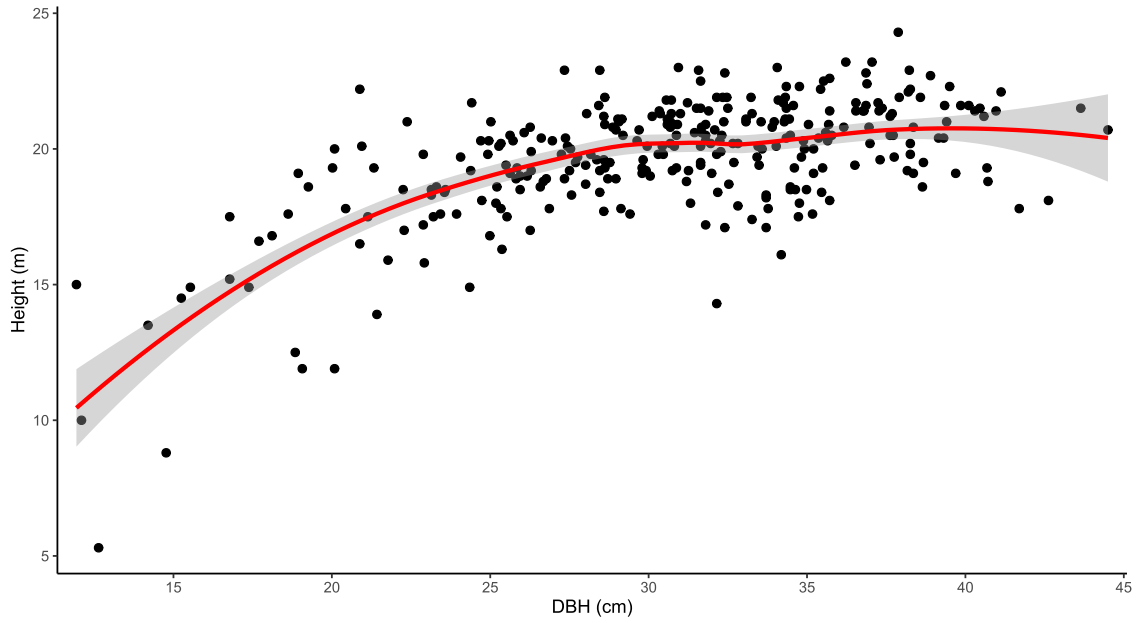


Figure 4.3: Polynomial regression of DBH and height measurements from the fieldwork data set.

4.1.2 Spatial characteristics

Finally, spatial characteristics of the research plot were collected by means of recording the coordinates of each individual tree by using the supporting micro-grid (*i.e.* labelled blocks) structure. These measurements were then subjected to an offset parameter according to the specific block placement over the entire macro-grid.

This enabled an overall reconstruction and visualisation of the top-view spatial diagram of the study area, as displayed in Figure 4.4. Even though this thesis does not elaborate too much on spatial aspects, it is conjectured that the CV toolkit would also be able to produce the same results by way of employing similar data-driven algorithms that can establish the specific tree positions.

Upon visual inspection of Figure 4.4 it is noticeable that the study area has been exposed to a first thinning operation — *i.e.* with some clear pathways running in an oblique direction

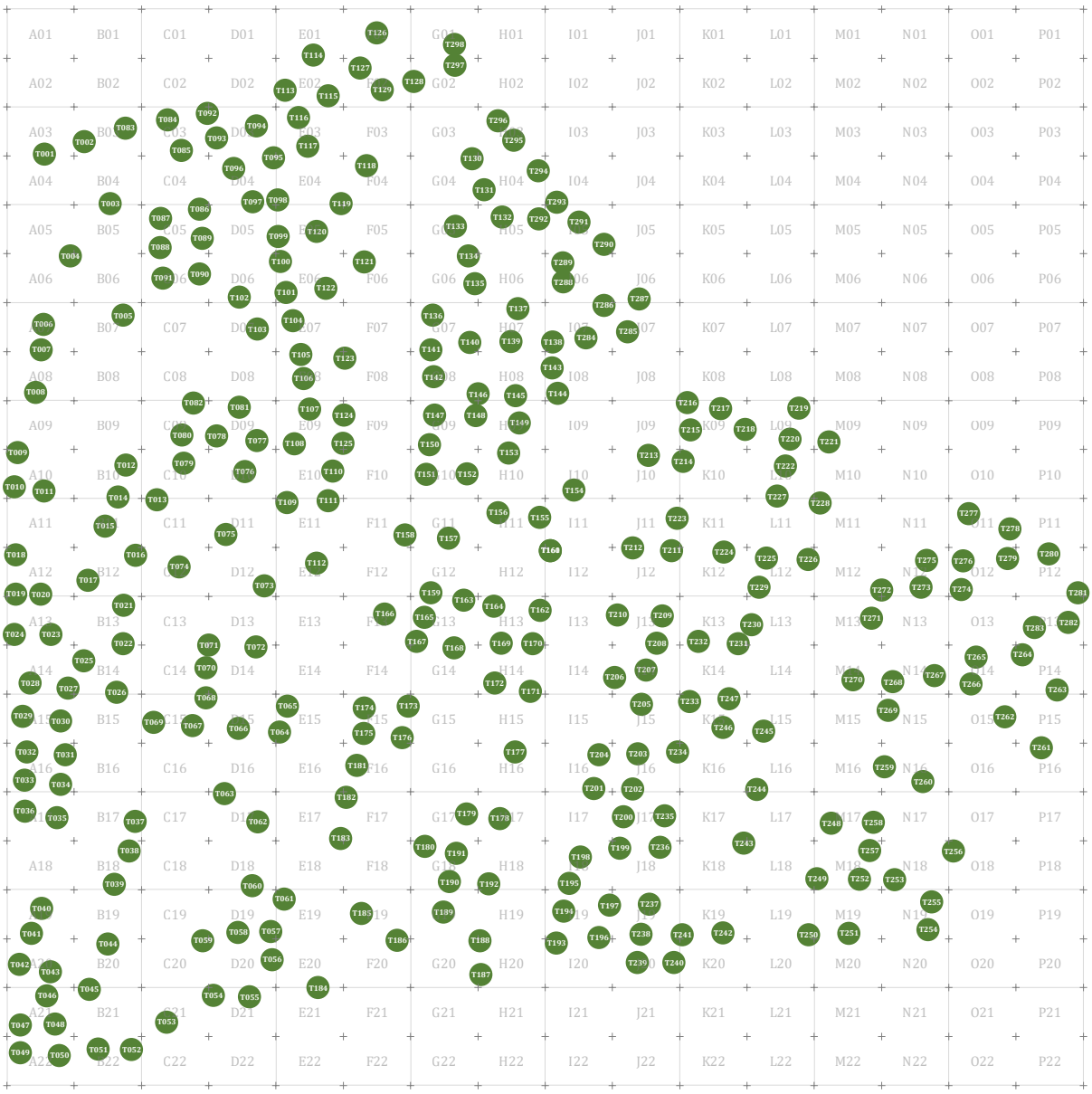


Figure 4.4: Top-view spatial outlay visualisation of all individual trees found in research sub-section.

across the diagram. Furthermore, it does not seem that the original planting of trees occurred exactly every 3 × 3 metres apart, as described in Table 3.2. Visually the trees are clustering together more densely at some spots, when compared with other sparsely separated areas.

Nevertheless, having actual and accurate spatial measurements recorded is beneficial because individual tree areas can be calculated in turn and analysed for growing behaviour according to the extent of ground space at each tree's disposal. Geocodes of the individual trees can also be determined through these spacial features (by reorientation of the localisation diagram and utilising a superimposed geodesic function) which can provide insightful knowledge, especially if this type of information is available for wide-ranging areas.

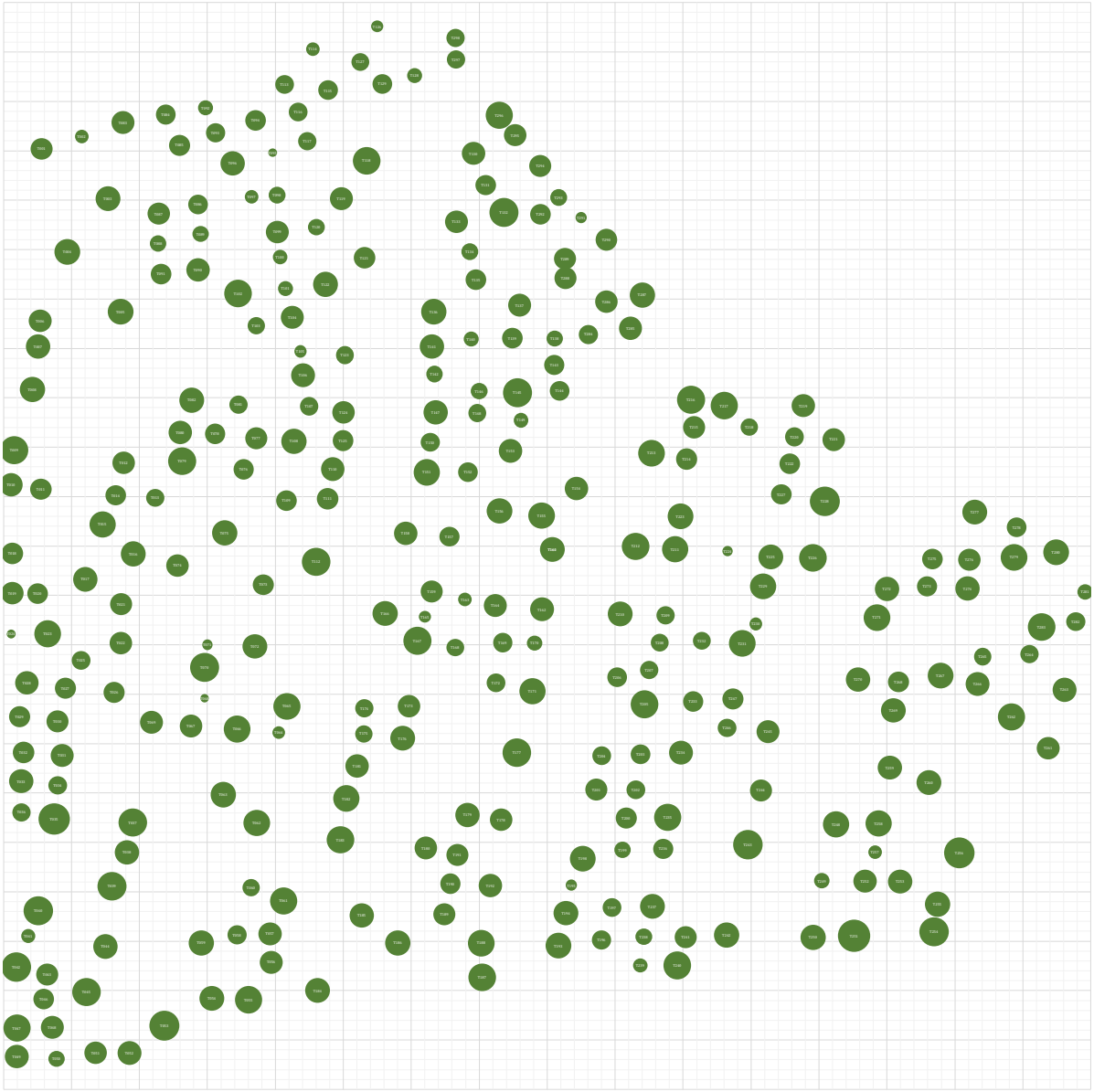


Figure 4.5: Top-view spatial outlay visualisation of all individual trees found in research sub-section, scaled relatively to DBH measurements.

Comparably, the same visualisation of spatial tree positions from the macro-grid can be illustrated with relevant DBH measurements of all trees in the research study area, as displayed in Figure 4.5. This alternative illustration paints an interesting picture when reviewing clusters of trees within close proximity of one another, thereby visually noticing a number of "bigger" trees located at certain positions as well as more predominately around the boundary/edge (on the right side) of the greater compartment.

4.2 Data preparation

This section on data preparation adopts an unconventional approach when compared with the CRISP-DM framework which would normally entail using the very same data sets produced from the business and data understanding phases. The current project sought to record the same environment (*i.e.* fieldwork study area) with a different type of structured data set (*i.e.* video files that were labelled according to separate lanes).

Ultimately the end objective is to obtain the same DBH measurements by the utilisation of CV, DL, and ML techniques which report on the degree of error that was delivered through this adopted alternative solution.

Video recordings

The observer was walking at an average stride throughout the forest compartment with a handheld recording device (*i.e.* smartphone[1]) storing video files in a structured manner — *i.e.* manually labelling the files with its relevant lane names and direction of walking (*i.e.* moving forward or backward, in reverse). A supplementary online form was utilised to capture the different file names and recording timestamps for the purposes of an informal time study.

The smartphone was attached to a so-called gimbal for video frame stabilisation (which later helped the DPT model during processing). The smartphone's built-in video recording application also had sophisticated stabilisation functionality already incorporated, which in turn contributed significantly toward smooth video recordings. The aspect ratio of frames favoured width above height (*i.e.* shot in landscape mode) which in some instances accounted for other trees falling outside of the demarcated lane boundaries.

Table 4.1 indicates the video recordings that were made during fieldwork — it shows the lane and direction of the recording as well as other important file parameters of the data set (*e.g.* frame width, height, rate, etc.). Adding the file sizes together yield a total of **4 645 MB** (*i.e.* approximately **4.6 GB**), which can be regarded as the initial storage requirement for the video recordings.

[1]The smartphone utilised in this project was a Google Pixel 3a released in 2019 and runs on the Android operating system. The specific model has 64 GB of storage capacity and possesses a 12 MP front-facing camera. The device has 4 GB of onboard read access memory and is powered by a 3 000 milliamp hour lithium polymer battery. The video functionality allows for a horizontal display resolution of approximately 4 000 pixels (*i.e.* so-called 4K) recorded at 30 frames per second.

Furthermore, the total time it took for all the recordings equated to almost **32** minutes, although neglecting a short set-up time and data label capture step before proceeding to the next lane in the series. The amount of time spent infield recording video files is an important indicator when comparing the execution time against normal tree enumerations. A time study comparison will be concluded later on — between physically measuring trees and the time it takes to record and process videos needed to collect the same information.

Table 4.1: Parameter descriptions of video data set collected during fieldwork.

Lane	Direction	File type	Frame width	Frame height	Frame rate	File size	Duration	Frames
A	Forward	MP4	1 920	1 080	30 frames/sec	186 MB	77 sec	2 340
A	Reverse	MP4	1 920	1 080	30 frames/sec	190 MB	79 sec	2 394
B	Forward	MP4	1 920	1 080	30 frames/sec	186 MB	77 sec	2 343
B	Reverse	MP4	1 920	1 080	30 frames/sec	189 MB	79 sec	2 389
C	Forward	MP4	1 920	1 080	30 frames/sec	187 MB	78 sec	2 360
C	Reverse	MP4	1 920	1 080	30 frames/sec	186 MB	77 sec	2 341
D	Forward	MP4	1 920	1 080	30 frames/sec	182 MB	76 sec	2 299
D	Reverse	MP4	1 920	1 080	30 frames/sec	180 MB	75 sec	2 272
E	Forward	MP4	1 920	1 080	30 frames/sec	188 MB	78 sec	2 372
E	Reverse	MP4	1 920	1 080	30 frames/sec	199 MB	83 sec	2 504
F	Forward	MP4	1 920	1 080	30 frames/sec	180 MB	75 sec	2 263
F	Reverse	MP4	1 920	1 080	30 frames/sec	186 MB	78 sec	2 351
G	Forward	MP4	1 920	1 080	30 frames/sec	181 MB	75 sec	2 280
G	Reverse	MP4	1 920	1 080	30 frames/sec	174 MB	73 sec	2 197
H	Forward	MP4	1 920	1 080	30 frames/sec	165 MB	69 sec	2 074
H	Reverse	MP4	1 920	1 080	30 frames/sec	178 MB	74 sec	2 243
I	Forward	MP4	1 920	1 080	30 frames/sec	151 MB	63 sec	1 905
I	Reverse	MP4	1 920	1 080	30 frames/sec	158 MB	66 sec	1 998
J	Forward	MP4	1 920	1 080	30 frames/sec	146 MB	61 sec	1 841
J	Reverse	MP4	1 920	1 080	30 frames/sec	139 MB	58 sec	1 746
K	Forward	MP4	1 920	1 080	30 frames/sec	113 MB	47 sec	1 432
K	Reverse	MP4	1 920	1 080	30 frames/sec	117 MB	48 sec	1 471
L	Forward	MP4	1 920	1 080	30 frames/sec	105 MB	44 sec	1 326
L	Reverse	MP4	1 920	1 080	30 frames/sec	115 MB	48 sec	1 459
M	Forward	MP4	1 920	1 080	30 frames/sec	109 MB	45 sec	1 373
M	Reverse	MP4	1 920	1 080	30 frames/sec	105 MB	43 sec	1 321
N	Forward	MP4	1 920	1 080	30 frames/sec	87 MB	35 sec	1 079
N	Reverse	MP4	1 920	1 080	30 frames/sec	88 MB	36 sec	1 095
O	Forward	MP4	1 920	1 080	30 frames/sec	79 MB	32 sec	980
O	Reverse	MP4	1 920	1 080	30 frames/sec	78 MB	32 sec	976
P	Forward	MP4	1 920	1 080	30 frames/sec	55 MB	22 sec	669
P	Reverse	MP4	1 920	1 080	30 frames/sec	63 MB	25 sec	764

It may be argued that the storage requirement is rather taxing when reviewing the total recording times — attributable to high dimensionality of frame sizes (1 920 pixels horizontally by 1 080 pixels vertically). The algorithmic process includes a resizing step in order to address storage issues after extracting picture frames from video files. The combined number of pictures that were extracted from the video data set equated to **58 457** frames in total.

Chapter summary

In conclusion of the CRISP-DM phases on data understanding and preparation, this chapter included an explanatory data analysis in which collected fieldwork features were scrutinised. These attributes included the DBH and height measurements as well as spatial features that can be considered as the "ground truth" from the study area of the project. Ultimately, the subsequent modelling phase attempts to reproduce the same measurement results using an alternative data set (*i.e.* structured video files).

The data preparation phase might be considered unconventional for CRISP-DM projects, in the sense that it is a totally separate approach in order to gather similar results with the incorporation of a different data format as an experiment to serve as proof of concept. The modelling phase endeavours to describe the processes that were followed in order to extract data features which made it possible to produce predictions close to DBH measurements.

Chapter 5

Modelling

The following chapter describes the foundational processes for comprehensive system modelling in order to extract precision tree measurements using concepts and techniques from the domain of applied ML. The main intuition is to ingest labelled video recordings and then subject the files to various algorithms in order to extract feature data at sequential modelling stages. These outputs represent input to subsequent steps which, in turn, drives sub-processes and identifies single clusters (*i.e.* tree segment properties) from multiple vantage points (*i.e.* sequenced picture frames). Linking these multiple data entries directly to a specific tree allows for an overarching ML model to train its mapping values toward the fixed actual measurements (*i.e.* DBH), which were collected per tree separately during fieldwork.

5.1 Workflow

In general terms, a ML workflow entails data ingestion, cleaning, preprocessing, transformations, modelling (which might include insight analysis), evaluation, and ultimately deployment (possibly into production systems). When designing the conceptualised precision tree measurements from video model for the current project, the input files are regarded as semi-structured (*i.e.* it cannot be recognised within a relational database, but is labelled in such a manner that its inherent structure can be inferred). Therefore any cleaning, preprocessing and transformations were not directly needed for the labelled video files — apart from extracting picture frames, resizing and staging these output images for the DPT network ingestion.

Standard visual icons used for *business process modelling notation* (BPMN), as described by Chinosi and Trombetta (2012) (98) and White (2004) (99), were used and are presented in Figure 5.1 — it illustrates the entire CV, DL and ML workflow for this project. This modelling notation was selected as the most appropriate visual representation for illustrating the data workflow of the comprehensive algorithm from start to end. Throughout the project, the appropriate modelling methods and coding environments — which help to achieve the desired

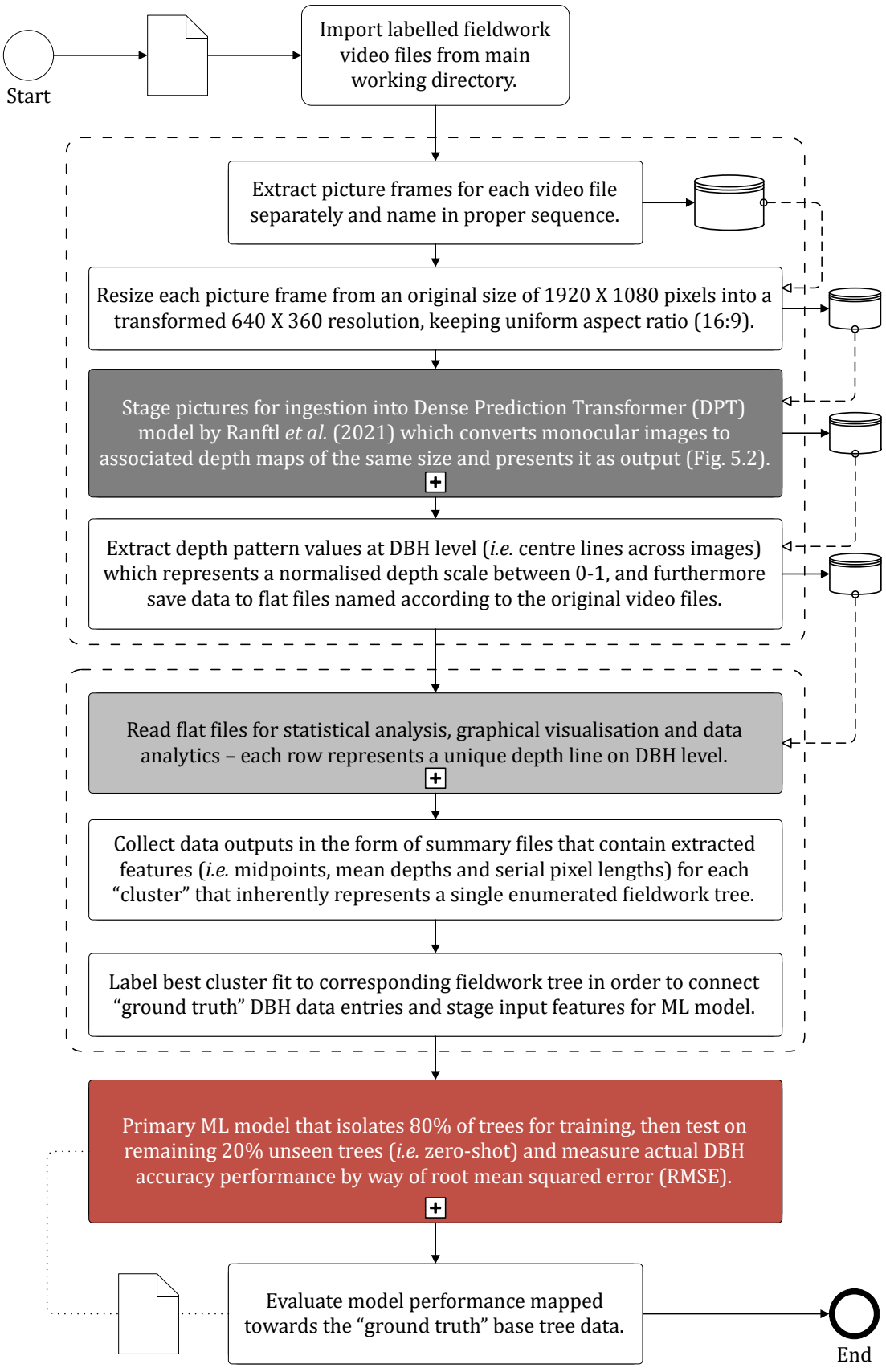


Figure 5.1: Flowchart of the ML model workflow that underpins the algorithmic approach adopted.

objectives best — were used throughout an iterative development process. An initial working directory is set and utilised as the starting point where the original labelled video files reside and the appropriate scripts will identify all the files listed within the said directory (according to its naming convention).

The next step in the algorithm entails a loop procedure which covers the list of files located in the main working directory and, upon execution, processes each video separately which extracts picture frames that make up the isolated recording — these images are renamed according to their numerical sequence in the respected video. At each step of the process a newly created data set is produced and regarded as the output from the specific step, which in turn serves as the input for the next step in the algorithmic series. After picture frame extraction of each video has occurred, these frames are adjusted in size by keeping an uniform aspect ratio of 16:9 and resized to 640×360 pixels.

An inherent advantage of using ViTs is that it can produce excellent MDE predictions for any picture frame size or ratio — the architectural nature permits these frameworks to ingest manageable chunks (*i.e.* tokens) which are subjected to individual convolutions. The reason for resizing images in this thesis was purely based on concerns of storage capacity (which also later assisted model processing speeds). It is conjectured that if the resizing step was skipped it can possibly provide higher resolution data patterns, but for the objectives of this project (*i.e.* concept demonstration) the option of downscaling pictures was selected.

Different architectures for performing MDE were evaluated which included testing the MC framework proposed by Li *et al.* (2019) (32) and also MiDaS designed by Ranftl *et al.* (2020) (42). These models can both generate depth estimations, however, the MC algorithm (32) was trained on dynamic videos containing people, as opposed to the better performing MiDaS model (42) trained on a variety of scenes and also being exposed to image augmentations. In Figures 5.3, 5.4, 5.5, 5.6 and 5.7 examples of these tests at various timestamps are indicated visually, according to which comparisons are made with the original input images collected during the project's fieldwork.

5.1.1 Dense prediction transformer

The forest video data set was processed by the DPT neural network of Ranftl *et al.* (2021) (45) which employs ViTs for performing MDE. Significant progress has been made with respect to MDE experiments — increases in performance of 28% are reported when compared with other state-of-the-art CNN based networks (45). Attention mechanisms (*i.e.* derived from a tokenisation procedure) in the DPT framework, trained on large and relevant data sets, can realise its full potential for solving depth challenges as well as object segmentation problems separately

(45). For its architecture, an encoder-decoder structure is utilised — the model incorporates a multi-objective optimisation algorithm[1] with the Adam method (25), during its reassembly and fusion stages, as illustrated in Figure 5.2.

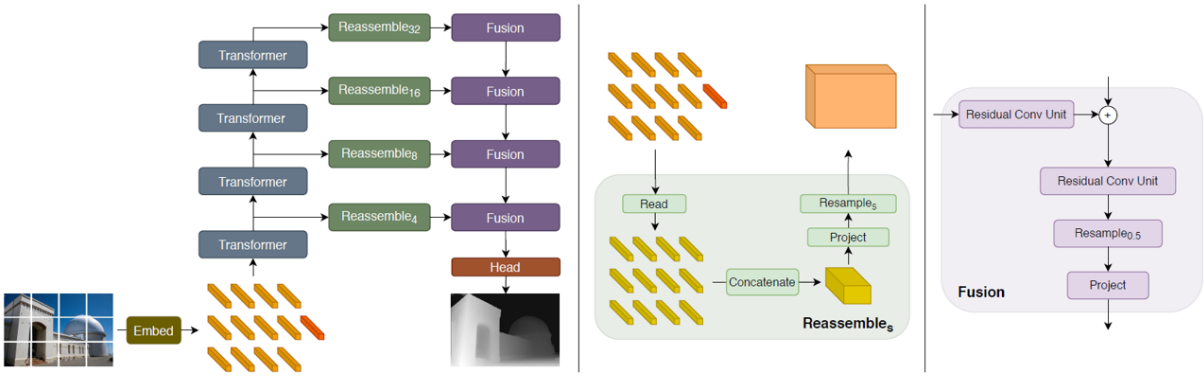


Figure 5.2: Architecture overview of ViTs by Ranftl *et al.* (2021) (45), the illustration is an excerpt from the research paper. *Left:* Input image being embedded into tokens (orange) and these are extracted linear projections of flattened representations (DPT) or generated by applying another feature extractor model (*e.g.* ResNet-50 and DPT-Hybrid). The embedding process identifies and labels the tokens in an ordered sequence and augments tokens with its positional attributes. These tokens are patch-independent and subjected to several transformer stages before it gets reassembled into an image-like representation, generated at various resolutions (green). Progressively fusion occurs through set modules and upsamples the input representations in order to generate a fine-grained depth prediction. *Center:* Transformed tokens are reassembled using sequence labelling, concatenated and projected to its original structure. *Right:* Fusion from different resolutions along with residual convolutions, is upsampled back to an intended depth prediction feature map.

Ultimately the DPT architecture developed by Ranftl *et al.* (2021) (45) was selected due to its superior depth estimation quality and noticeable processing performance for image conversions. The programming language Python 3 was used to program some of the procedures of the algorithmic implementation — Figure 5.1 illustrates the different procedures by means of the upper dashed border grouping executed solely in Python.

The last step within this aforementioned sub-procedure enclosing is to extract data patterns according to the vertical middle line from picture frames — described in more detail in the tree recognition section — *i.e.* representing DBH pixel width segments because the observer was holding the smartphone recording device constantly on DBH level. These pattern values are normalised on a scale ranging from 0 to 1 which represents the horizontal depth estimations (*i.e.* side view recording device perspective).

[1]For multi-task learning, multi-objective optimisation is reported on by Sener and Koltun (2018) (100) in order to find an optimal Pareto solution as its overall objective. Gradient-based algorithms are employed in their research as a conventional enumerative technique for optimisation challenges. A local search method is utilised in this literature for its objective functions, classified for non-linear and continuous problems.

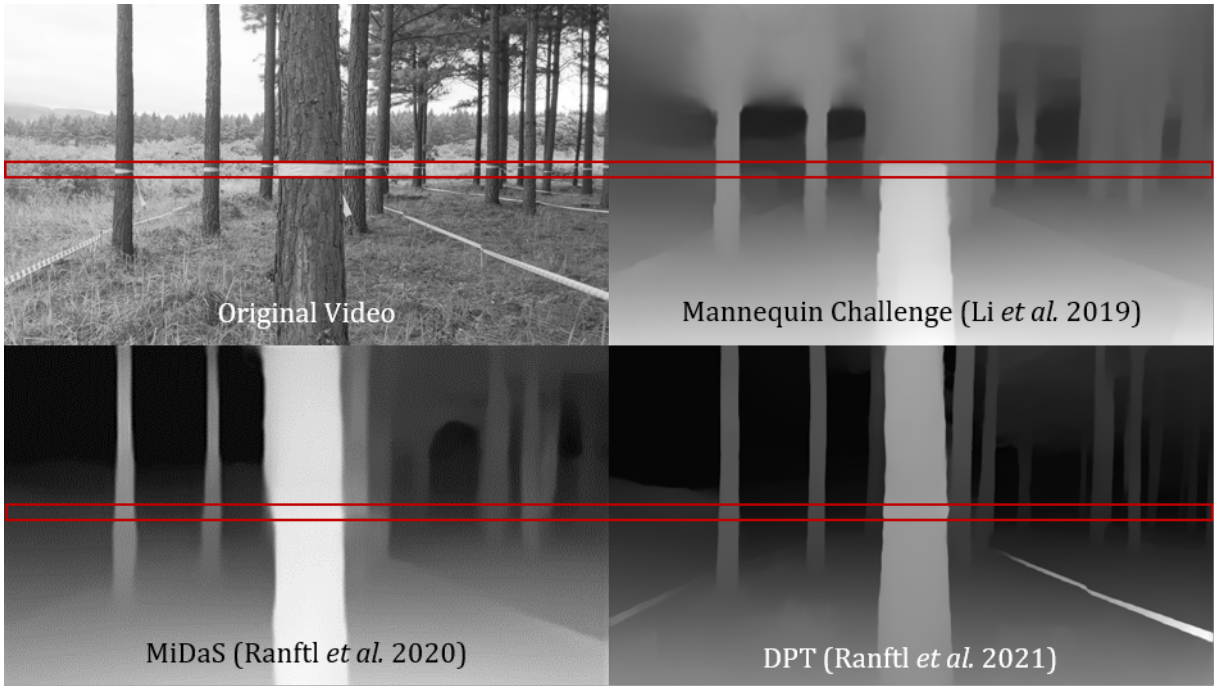


Figure 5.3: Illustration includes video converted to picture frames at t = 0 sec of the original video (top left) and, then using neural network toolkits, processed by the Mannequin Challenge (top right), MiDaS (bottom left) and DPT (bottom right) trained data sets.

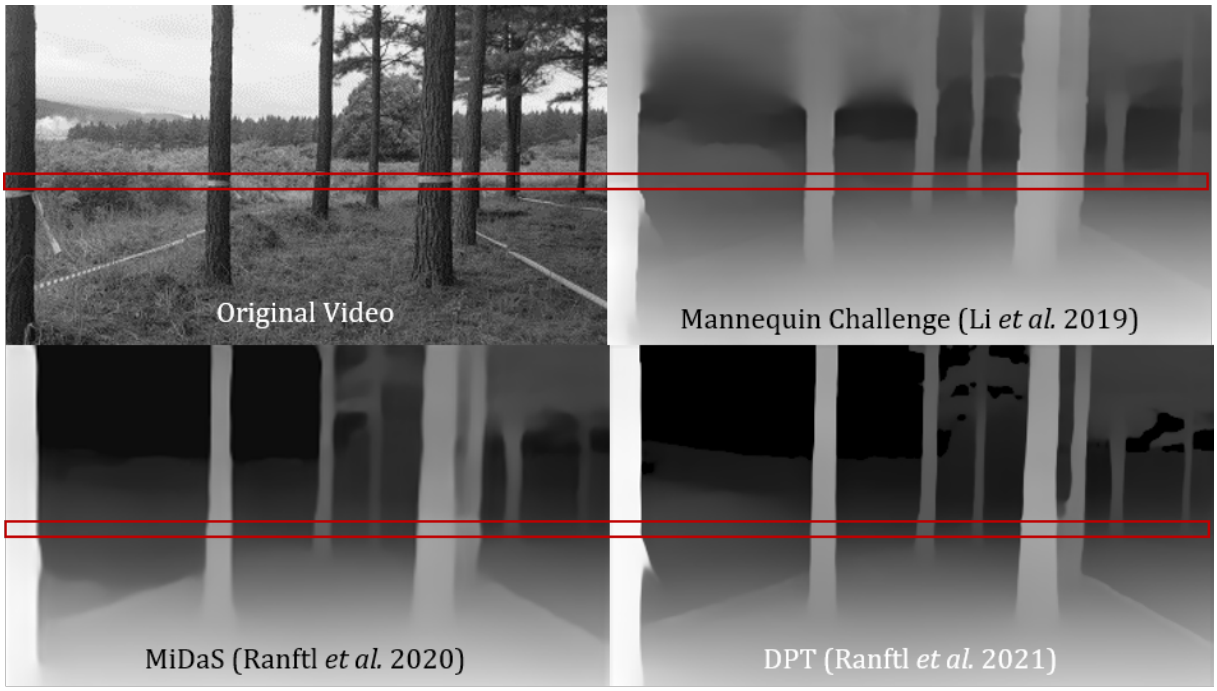


Figure 5.4: Illustration includes video converted to picture frames at t = 5 sec of the original video (top left) and, then using neural network toolkits, processed by the Mannequin Challenge (top right), MiDaS (bottom left) and DPT (bottom right) trained data sets.

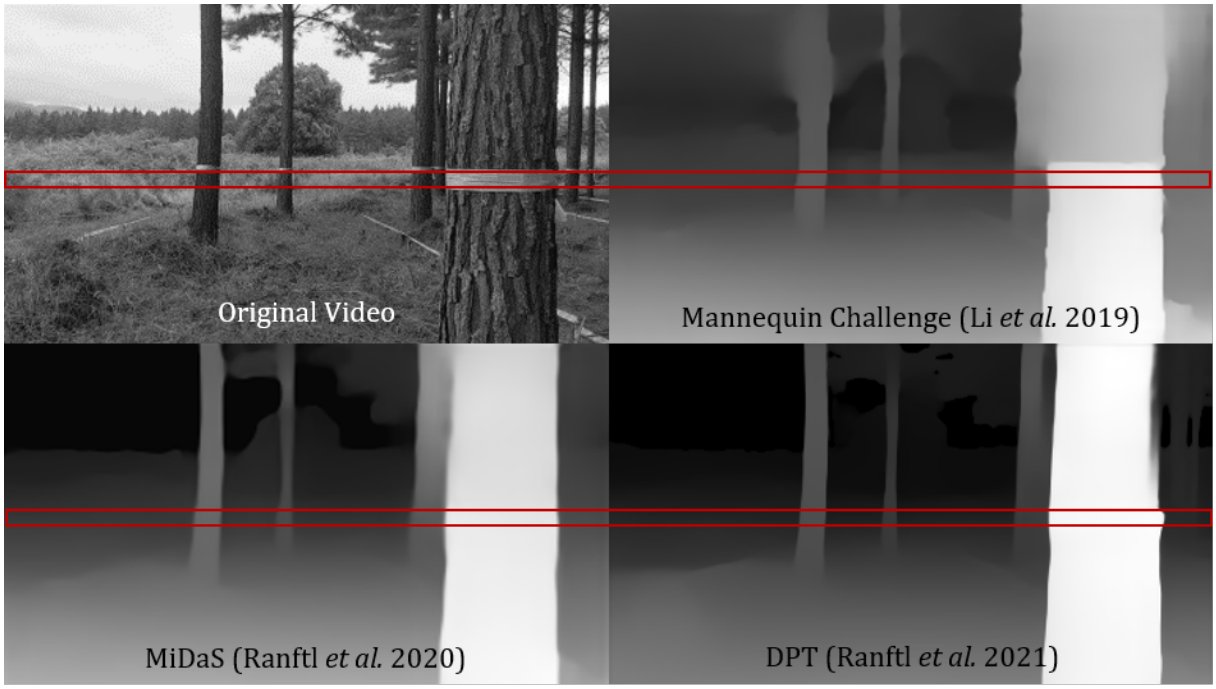


Figure 5.5: Illustration includes video converted to picture frames at t = 10 sec of the original video (top left) and, then using neural network toolkits, processed by the Mannequin Challenge (top right), MiDaS (bottom left) and DPT (bottom right) trained data sets.

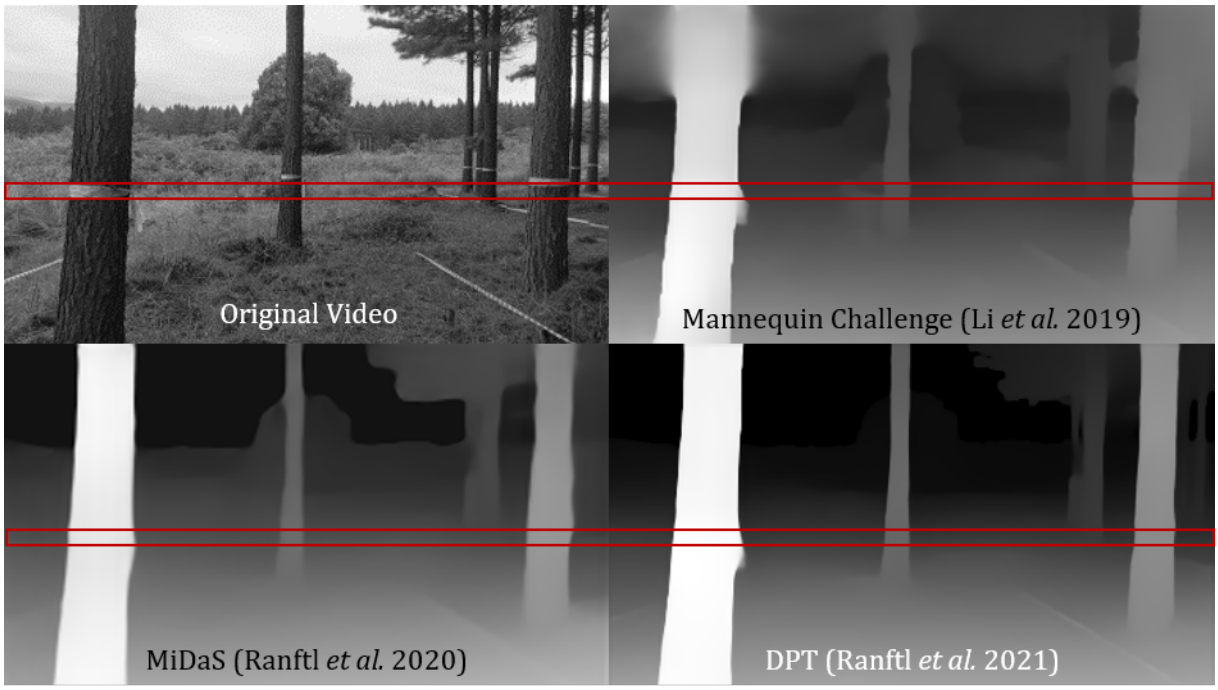


Figure 5.6: Illustration includes video converted to picture frames at t = 15 sec of the original video (top left) and, then using neural network toolkits, processed by the Mannequin Challenge (top right), MiDaS (bottom left) and DPT (bottom right) trained data sets.

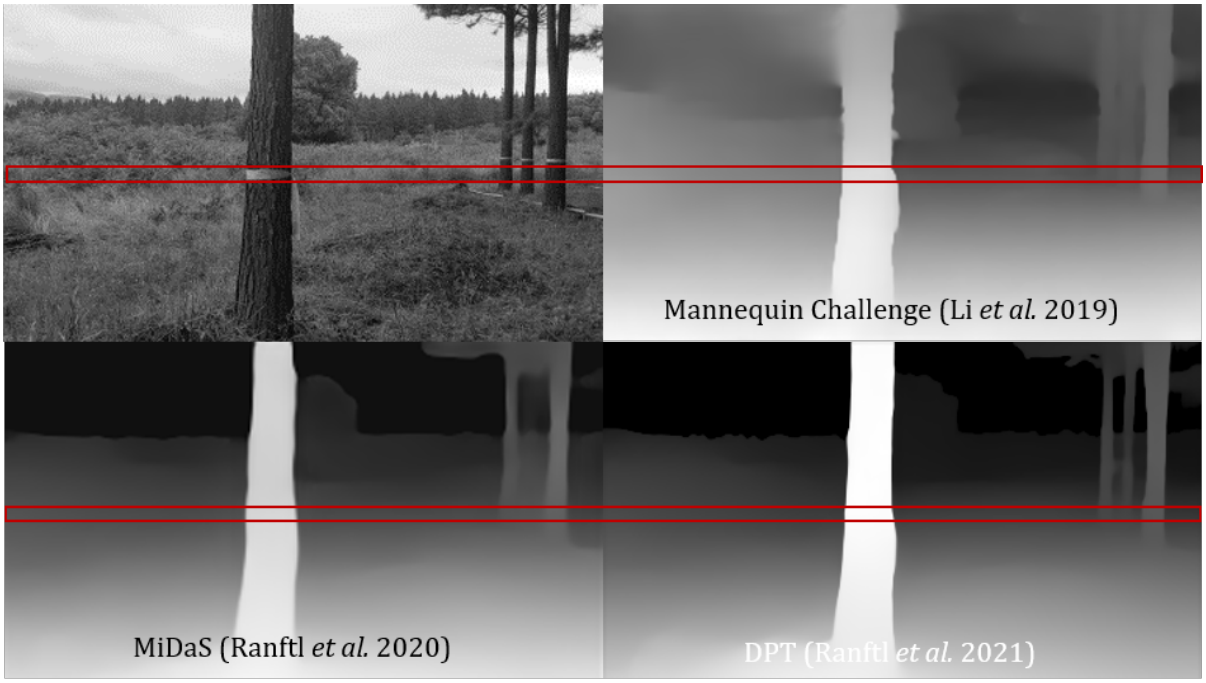


Figure 5.7: Illustration includes video converted to picture frames at t = 20 sec of the original video (top left) and, then using neural network toolkits, processed by the Mannequin Challenge (top right), MiDaS (bottom left) and DPT (bottom right) trained data sets.

Finally, the model saves appended data patterns with a file name corresponding to the video file — *i.e.* each row in the comma separated format represents the DBH line from individual pictures in an ordered sequence, whereas the columns are regard as the *field of view* (FOV) and constantly derived as 640 pixels in width. In a manner of speaking, each row in the output can be regarded as the sequential spatial progression of the observer operating the recording device.

It should be noted that Figures 5.3, 5.4, 5.5, 5.6 and 5.7 only display a very limited number of extracted picture frames for the sake of showcasing the various MDE algorithms in respect of a single forestry lane (*e.g.* **P** in the **forward** direction). For the entire model, however, the adopted approach involved processing every image from the complete video data set (*i.e.* **58 457** frames) by using the DPT framework for MDE prediction. Each image has an associated filename which indicates the relevant fieldwork lane, combined with its appropriate sequence number that places the picture at a certain spatial position — *i.e.* considering the specific offset parameter, the model can superimpose the localisation back to the devised macro-grid.

After the capturing of video recordings took place, the effect of fluttering demarcation tape (*i.e.* tied around trees for unique identification purposes) manifests when inspecting the MDE output images in Figures 5.3, 5.4, 5.5, 5.6 and 5.7. Even though the tape served a well-intended purpose to determine the DBH level visually, the practice unintentionally distorted some of the generated line pattern data. During the subsequent clustering process (described in more detail

in a succeeding section), it is apparent that some tree segment "circles" are drastically different in size within close spatial proximity of one another (*i.e.* basically overlapping) upon review. It is noted that these anomalies are few for the overall output data set and will therefore not have a significant impact on comprehensive modelling attempts.

Nevertheless, the issue is recognised as a minor problem in the current project and this lesson learnt should be considered in possible future work. Under normal field conditions, however, natural trees will not exhibit any artificial markings such as the plastic demarcation tape tied around it. The matter will therefore likely be of little to no concern if the concept is applied to other forest compartments. For this project, an attempt was made to record the video files in relative quick succession after the collection of tree measurements from fieldwork were completed. Only a couple of days lapsed between the gathering of the separate data sets, as trees are always growing wider even if growth happens over very long periods of time.

The aforementioned fact (*i.e.* growing trees), along with the insight that marking tape at DBH level did not bode well for model processing, prompted a speculation that the same study can be executed in a more controlled manner. For future work, it is conjectured that a better clinical approach, exhibiting an artificial man-made environment with "fake" trees along with known cylindrical diameters, can be adopted for the purposes of increased accuracy in model calibration. This approach could also provide stable surroundings (*i.e.* none of the diameters growing larger as time passes), which could theoretically be utilised over many iterations in order to sharpen tree measurement predictions.

Additionally, this proposal could also survey a wider uniform spread of DBH measurements in order to include more training examples. Conversely, this thesis later describes good performance on tree data that had many training examples which centred around the mean diameter (*i.e.* approximately between 25–35cm) of the research study area. Worse performance metrics were evaluated for data points falling outside of this range.

5.1.2 Tree recognition

Inherently, a lot of noise was produced from the raw output data gathered from the DBH line patterns/strings — *e.g.* such as distant trees or obscuring landscapes in the immediate background of picture frames derived from original video files. This challenge is controlled by setting a fixed threshold parameter (*i.e.* only accepting depth estimation values larger than **0.5** from the normalised scale for entire tree segments) in order to only extract sensible information. This approach renders certain objects "invisible" (proverbially speaking) as the observations only accepts subjects within a close proximity of the recording device.

The flat files for forest lanes are read into a programming environment (*i.e.* R procedures represented with the second dashed grouping in Figure 5.1) suitable for statistical analysis, visualisations and data analytics. The appropriate indexing matches the row numbers (*i.e.* picture frame sequences) and permitted the correct processing of the DBH line patterns which were extracted from DPT depth estimations. An algorithmic approach was developed in R which identifies sharp and sudden inclines or declines in the DBH pattern slopes which is regarded as the tree segments recognition procedure in the overall model, as illustrated in Figure 5.8. It is noted that incomplete tree segments, such as patterns exiting the FOV on either sides of frames, will be disregarded as it is not representative of an entire tree.

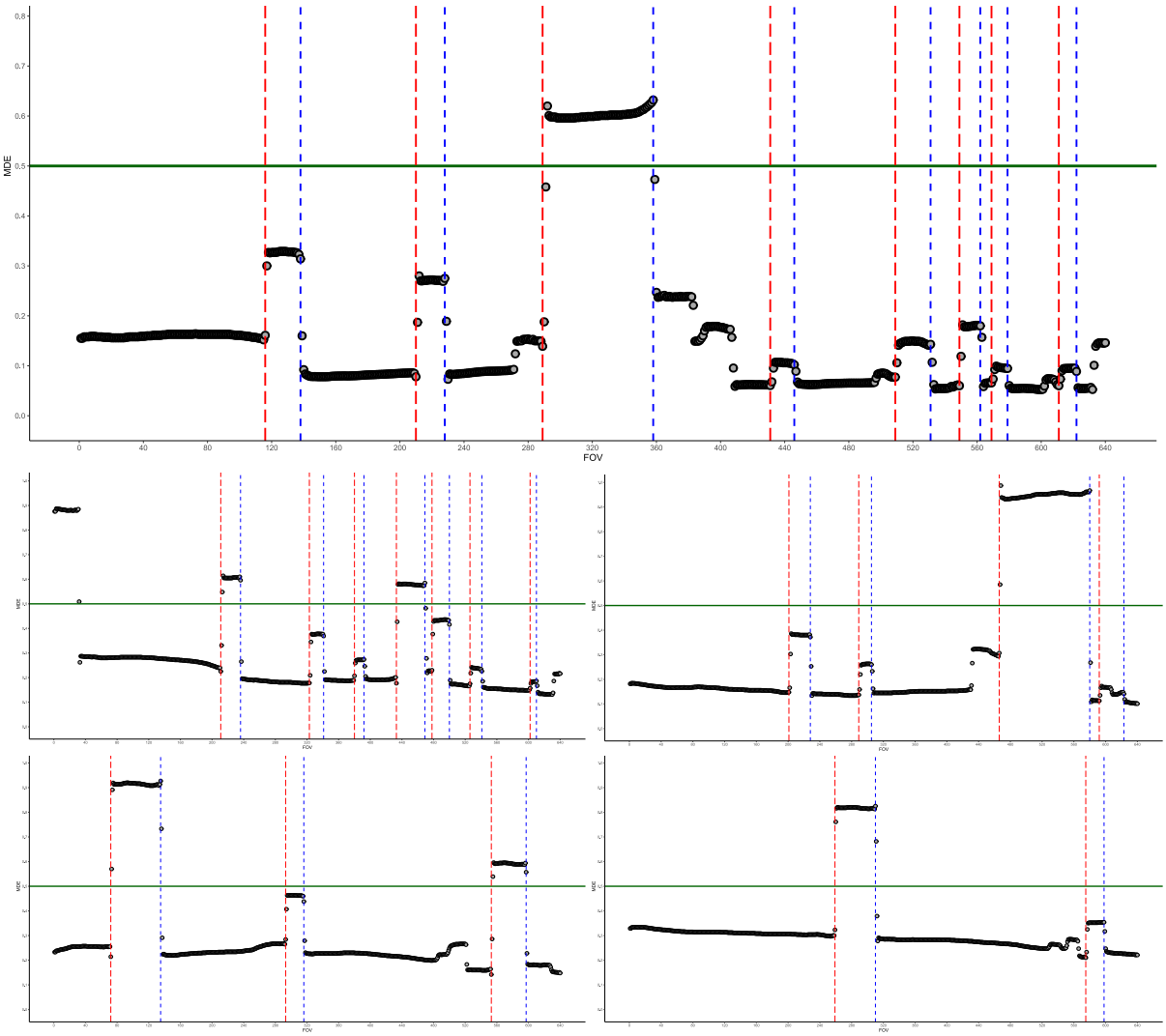


Figure 5.8: MDE outputs for forest lane "**P Forward**" of DBH line patterns at various timestamps (*i.e.* from top to bottom and left to right representing picture frames at 0 sec, 5 sec, 10 sec, 15 and 20 sec corresponding to Figures 5.3, 5.4, 5.5, 5.6 and 5.7), respectively with all plots indicating the set acceptance threshold at value **0.5** (*i.e.* green solid horizontal line) as well as **starting** tree segments (*i.e.* red dashed vertical lines on recognised lefts) and **ending** tree segments (*i.e.* blue dashed vertical lines on recognised rights).

Even though the data points acceptance threshold is indicated on the plots of Figure 5.8, the program will first determine the number of pixels that are contained within the segments and produce a relevant average of its depth estimates, before neglecting the segment in its entirety.

There is a lot of meta data produced during this stage — particularly the picture frame sequence number and filename from the relevant video recording (*i.e.* virtual spatial lane) which are tied to each DBH line pattern and permits additional dimensionality for the entire model. These aspects are considered as the driving factors which will dictate the spatial offsets when positioning multiple data points (*i.e.* tree segments accepted above the controlled threshold setting) graphically. The data should be regarded as the most basic features that are necessary to execute the prediction model — in many ways the required input information is flat in terms of what is being processed from observations — with some added logic these insights are transformed toward an elevated level of understanding.

Ultimately, the developed model will only consider features relating to segment **pixel** lengths (*i.e.* closely resembling the DBH measurement of a tree, and the **midpoint** of each segment record will indicate the observed angle in the FOV). The overarching ML model will incorporate the aforementioned features, along with the average **depth** value and train a prediction algorithm mapping toward the target features (*i.e.* actual DBH from fieldwork), after clustering and connection to uniquely identified tree numbers have been executed. Structuring the final ML model in this manner (*i.e.* only requiring a small number of input variables as described) enables dimensionality reduction and generalisation when (in future) new data of a similar format are inserted in the developed model. It is envisioned that these limited input variables will be sufficient to predict DBH measurements with a fair degree of accuracy.

5.1.3 Clustering

The next step in the algorithmic process is to establish similar cluster formations (relating to individual trees) from the recognised MDE segments. This output data set is derived from the segment properties illustrated in Figure 5.8 and isolated from every forest study area lane for each separate walking direction. In its most basic form, the data records will produce features for horizontal DBH lines of batched **pixels** (*i.e.* string length of tree segments) and its associated **midpoint** (*i.e.* centre point of the segment derived from the length of a certain pixel series and placement in the observing FOV). The mean **depth** estimates for the specific string of pixels (*i.e.* a number ranging from **0.5 to 1.0** which represents an average closeness to the observer of each tree) will also be included in the penultimate output data set.

By utilising all the data points that were extracted as feature values during tree recognition and segmentation, along with the relevant spatial offset parameters, it is possible to graphically

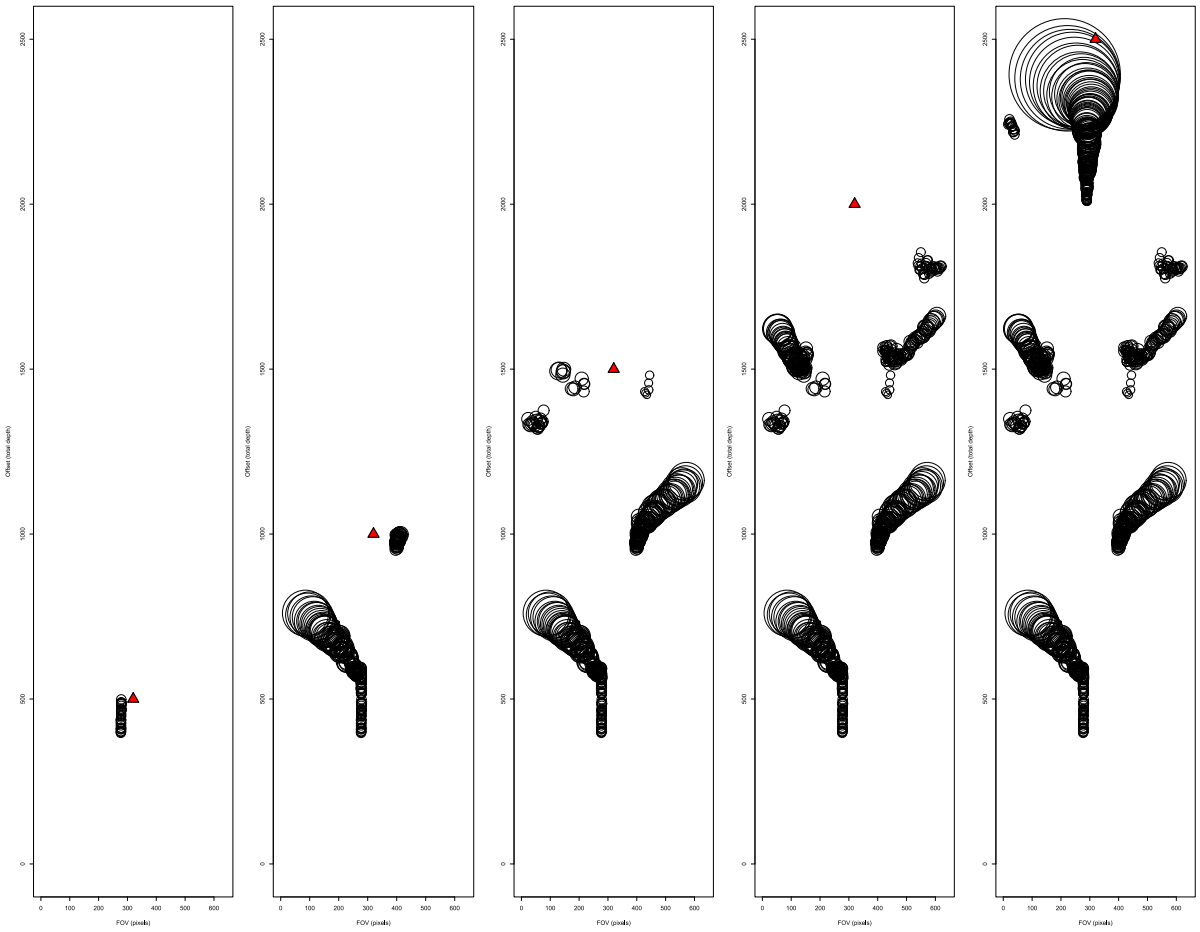


Figure 5.9: From left to right, model outputs for top-view spatial analysis at total depth offsets of 500cm, 1 000cm, 1 500cm, 2 000cm and 2 500cm respectively as illustrated by the red moving arrow (observer/video recording device) pointing in the walking direction.

illustrate the representations roughly as displayed in Figure 5.9. Essentially, this visualisation can be regarded as the penultimate step before completing the project's comprehensive workflow for modelling — all that remains afterwards is to cluster the same tree data points together (by means of conventional algorithmic approaches in the literature) and finally linking the clustered records toward unique tree numbers to create a connection with fieldwork DBH measurements (producing a many-to-one relationship between thousands of data points from MDE imagery and a single tree DBH reading respectively).

A combination of applied ML techniques designed for cluster analysis — *e.g.* interactive k-means clustering, originally devised by MacQueen (1967) (101), using an euclidean distance metric on multivariate features in the data set — and manual intervention was utilised in order to isolate individual tree clusters correctly for this project. This approach proved successful and it is seemingly evident when inspecting the separating cluster formations visually in Figure 5.10 which illustrates individual trees with different colours. The final step in this process is to

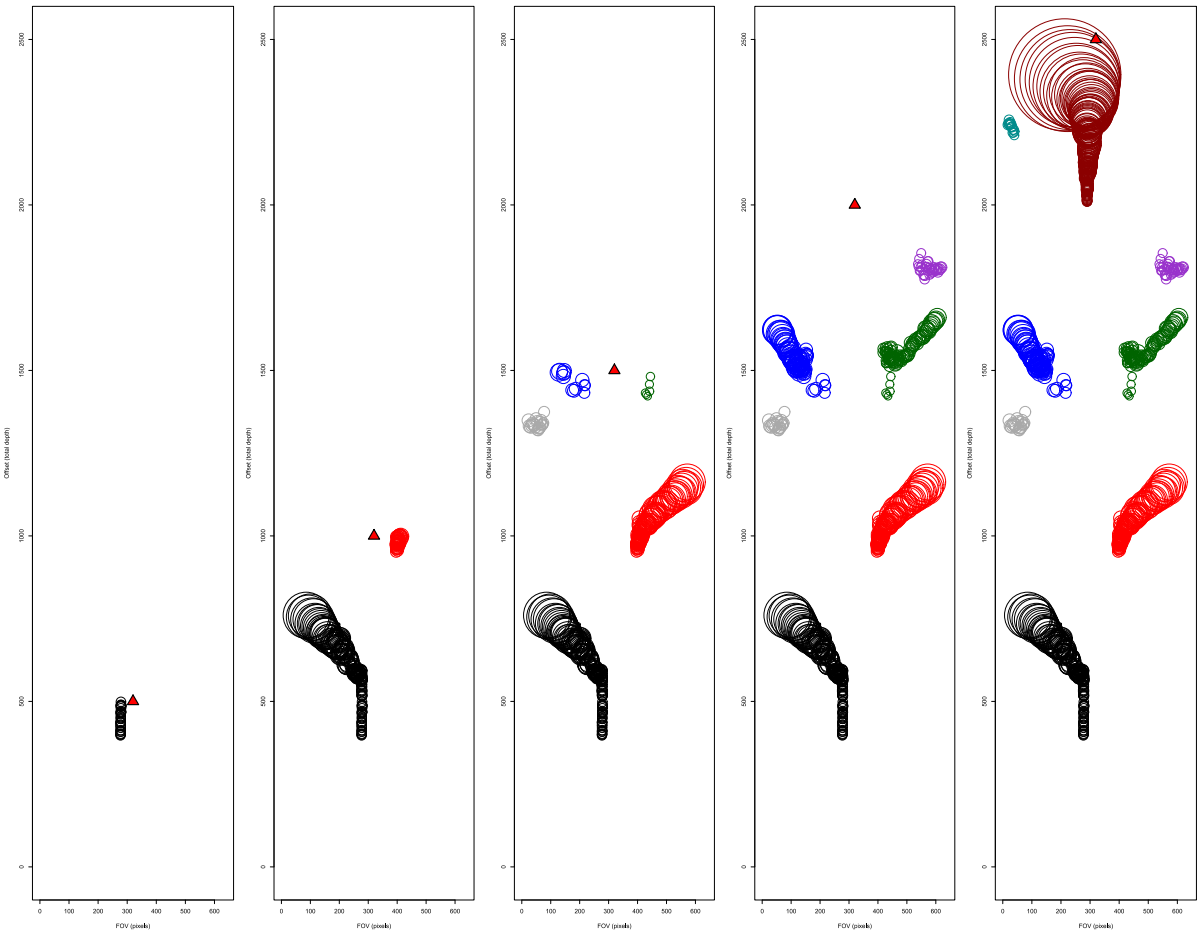


Figure 5.10: From left to right, model outputs for top-view spatial analysis at total depth offsets of 500cm, 1 000cm, 1 500cm, 2 000cm and 2 500cm respectively as illustrated by the red moving arrow (observer/video recording device) pointing in the walking direction, individual tree clusters according to separate colours.

connect the multiple data points for a single unique tree identified from fieldwork — this task was performed manually but it is envisaged that under normal field conditions (*i.e.* dealing with unlabelled trees) this will not be necessary.

5.2 Output

Once the clustering and linking of actual DBH measurements have been concluded for all 298 trees in the research plot, a final flat file is produced which saves the many data points with relevant features of data records. This flat file will only include a limited number of features including midpoints, mean depths, pixel lengths and absolute FOV angle (*i.e.* an additional logic field assisting the overall ML model, calculated with the midpoint value positioned over a total width of 640 pixels). The tree number and its associated DBH measurement is compiled

in the flat file — the DBH column will serve as the primary target feature and continuing with the present thesis, be used in conjunction with various data science modelling techniques for successfully mapping an accurate representation of the findings.

5.2.1 Spatial representation

Additionally, the forest lane name and ordered picture frame sequence numbers are available in the aforementioned complete flat file which serve as specifications for spatial analysis. With this insightful information at hand it is possible to represent relative location characteristics of all the data points graphically. The tree segment data entries are subjected to the appropriate spatial offset parameters according to lane names (*i.e.* horizontal placements) and picture frame sequence numbers (*i.e.* vertical positioning).

The top-view representation is scaled roughly according to actual distances on a centimetre unit of measure as illustrated in Figure 5.11. Interestingly, a familiar pattern emerges with this approach closely resembling the silhouette of physical fieldwork tree positions as described in Figures 4.4 and 4.5 from the data understanding phase of this project.

From this illustration, it is noted that some tree ring data in Figure 5.11 are much more prevalent than other clusters. This is not necessarily an indication of the DBH size of a tree, but rather the total number of data points of individual clustered segments (*i.e.* tending to overlap multiple times at the determined placement points upon visual inspection). The reason for this is because the observer is walking directly toward some trees during video recordings in the constructed forest lanes, as opposed to trees that were located away from the middle guiding line. Effectively, tree segment data at the sides of the recording device (*i.e.* located at the boundaries of demarcated forest lanes) will be produced less frequently after model workflow completion. Even though less examples of these trees are recorded, it can be regarded as a good generalisation ability of the concept, as the model trains on various examples of placement angles in the FOV of video files.

The relevant clustering information per unique tree was effectively utilised retroactively in Figures 5.11 and 5.12 after clusters were already established for all trees. This means that segment data for individual trees are not exactly aligned on top of a centre placement point as displayed in these figures. Conversely, the visual spatial data actually appears more similar to the representations of Figures 5.9 and 5.10 if a close-up review is conducted. An extra logic algorithm was added in order to centre the same clusters according to its average horizontal and vertical coordinates. This approach permitted the orderly visual representations for model outputs serving as spatial analysis, which closely resembles the holistic research study area.

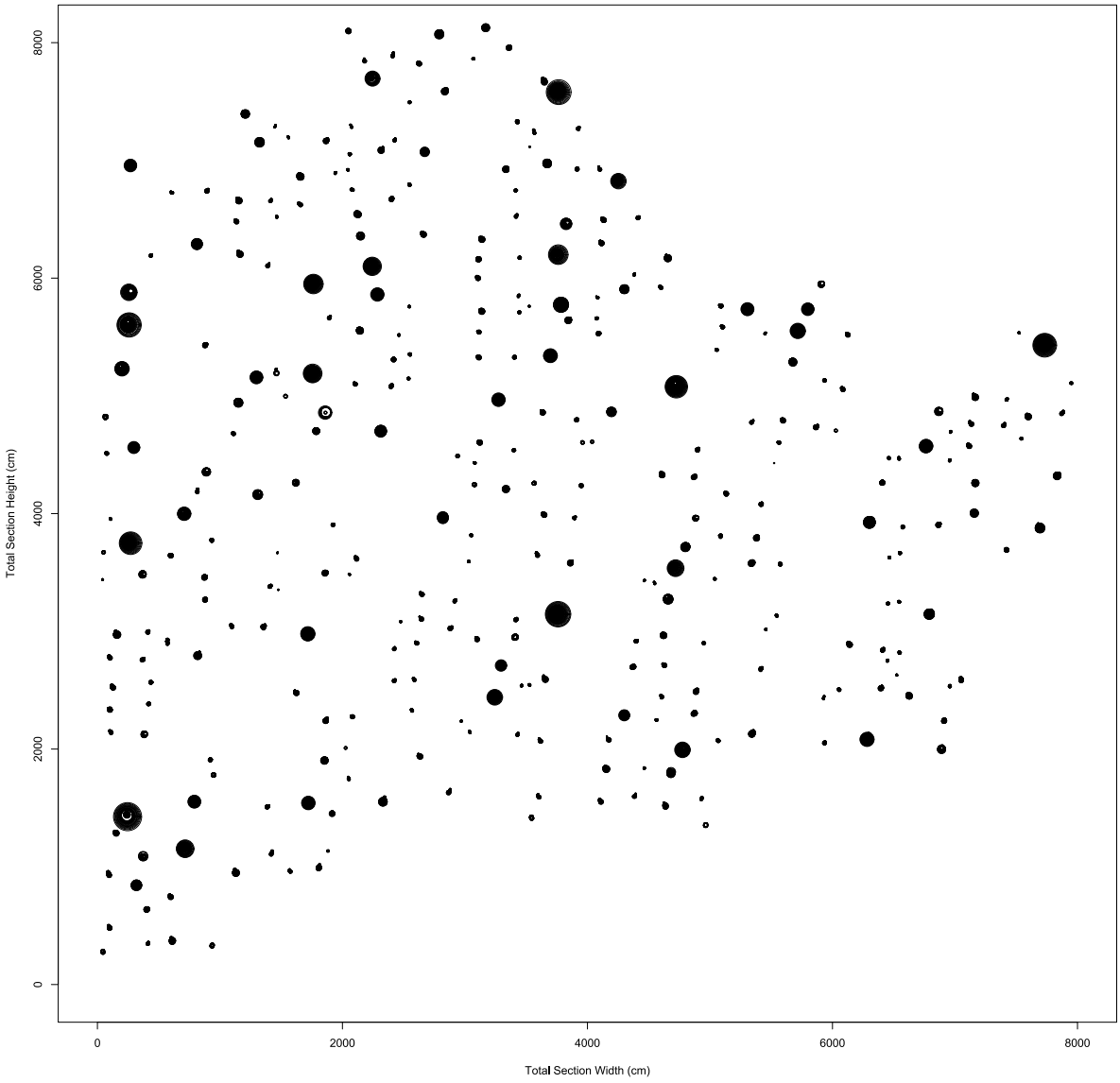


Figure 5.11: Model Output for Spatial Analysis (All Trees).

5.2.2 Runtime analysis

The computing architecture for this project consisted of an Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz processor and 16.0 GB of installed RAM on a 64-bit Microsoft Windows 10 Pro operation system. A graphics card manufactured by NVIDIA (GeForce GTX 1650) was used in conjunction with the CUDA parallel computing platform, also developed by NVIDIA (102) which facilitated parallelisation of the DPT processing network.

Programming environments such as Python 3 and R was used separately for the CV, DL and ML processing steps for implementing the DPT by Ranftl *et al.* (2021) (45) along with designed algorithms in R meant for data pattern recognition as well as feature extractions. These

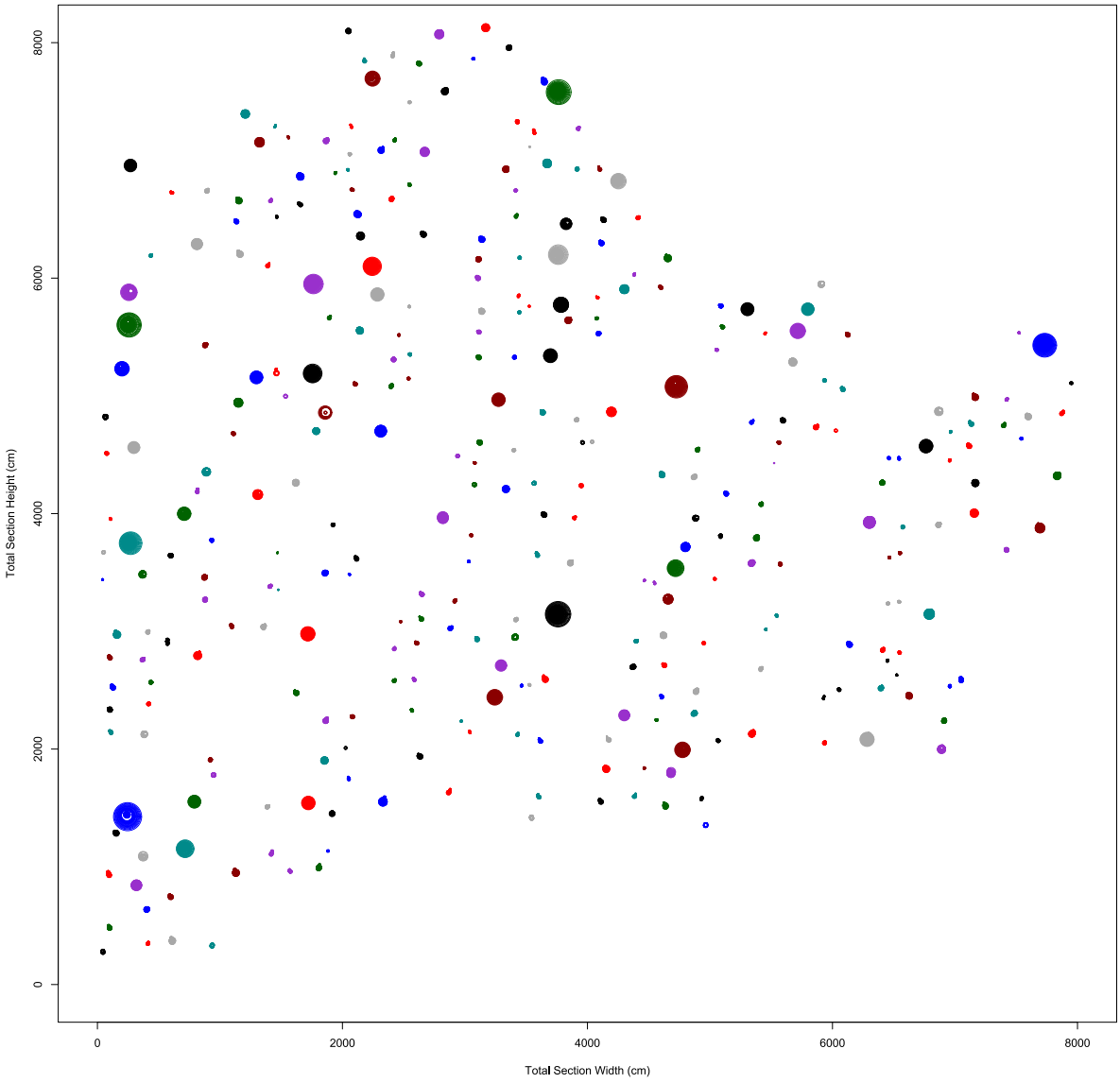


Figure 5.12: Model Output for Spatial Analysis (All Trees Clustered).

process steps were timed and are indicated sequentially in Table 5.1 for each video recording. Table 5.2 indicates the total time required for each process step, converted to **minutes** and **hours**.

Referring back to the physical fieldwork done in-person in line with standardised enumeration procedures (*i.e.* utilising the caliper method), the operational time that lapsed to collect DBH measurements for 298 trees added up 2 hours 58 minutes (*i.e.* approximately three hours). When compared with the approximately 18 hours (Table 5.2) that were required to record infield videos as well as to process the various algorithmic steps — the digital alternative technique took 6 times longer to generate the same information.

Table 5.1: Processing times of each algorithmic component for every file in video data set.

Lane	Direction	Frame extraction	Frame resizing	DPT Ranftl *et al.* (2021)	Feature extraction	Total processing time
A	Forward	202.65 sec	277.43 sec	1 801.10 sec	23.95 sec	2 305.13 sec
A	Reverse	212.06 sec	275.14 sec	1 838.14 sec	24.69 sec	2 350.03 sec
B	Forward	203.41 sec	272.40 sec	1 786.28 sec	24.88 sec	2 286.97 sec
B	Reverse	212.79 sec	265.65 sec	1 825.33 sec	25.21 sec	2 328.98 sec
C	Forward	203.28 sec	281.62 sec	1 804.46 sec	24.83 sec	2 314.19 sec
C	Reverse	208.59 sec	264.13 sec	1 792.04 sec	24.82 sec	2 289.58 sec
D	Forward	198.61 sec	270.58 sec	1 762.19 sec	24.23 sec	2 255.61 sec
D	Reverse	203.98 sec	251.01 sec	1 757.86 sec	24.34 sec	2 237.19 sec
E	Forward	206.94 sec	277.30 sec	1 809.10 sec	25.40 sec	2 318.74 sec
E	Reverse	232.19 sec	284.18 sec	1 910.55 sec	26.75 sec	2 453.67 sec
F	Forward	194.68 sec	265.76 sec	1 831.45 sec	24.21 sec	2 316.10 sec
F	Reverse	211.69 sec	256.95 sec	1 800.91 sec	25.06 sec	2 294.61 sec
G	Forward	195.96 sec	270.63 sec	1 753.90 sec	24.27 sec	2 244.76 sec
G	Reverse	198.43 sec	243.98 sec	1 686.32 sec	23.25 sec	2 151.98 sec
H	Forward	180.35 sec	241.59 sec	1 591.21 sec	23.73 sec	2 036.88 sec
H	Reverse	202.86 sec	250.98 sec	1 717.72 sec	24.32 sec	2 195.88 sec
I	Forward	168.64 sec	218.03 sec	1 454.02 sec	20.37 sec	1 861.06 sec
I	Reverse	183.19 sec	221.83 sec	1 524.84 sec	21.52 sec	1 951.38 sec
J	Forward	159.61 sec	210.99 sec	1 404.87 sec	19.64 sec	1 795.11 sec
J	Reverse	157.97 sec	195.96 sec	1 337.37 sec	19.04 sec	1 710.34 sec
K	Forward	126.11 sec	157.48 sec	1 092.86 sec	15.58 sec	1 392.03 sec
K	Reverse	133.62 sec	160.18 sec	1 124.42 sec	16.34 sec	1 434.56 sec
L	Forward	114.88 sec	142.88 sec	1 011.47 sec	14.22 sec	1 283.45 sec
L	Reverse	133.57 sec	156.26 sec	1 115.19 sec	15.66 sec	1 420.68 sec
M	Forward	118.00 sec	149.69 sec	1 054.52 sec	14.81 sec	1 337.02 sec
M	Reverse	120.84 sec	139.14 sec	1 008.76 sec	14.52 sec	1 283.26 sec
N	Forward	91.51 sec	116.47 sec	823.83 sec	11.36 sec	1 043.17 sec
N	Reverse	100.02 sec	113.54 sec	836.70 sec	11.84 sec	1 062.10 sec
O	Forward	82.80 sec	101.93 sec	748.61 sec	10.28 sec	943.62 sec
O	Reverse	89.38 sec	100.06 sec	745.88 sec	10.56 sec	945.88 sec
P	Forward	55.63 sec	70.50 sec	512.91 sec	4.23 sec	643.27 sec
P	Reverse	70.03 sec	79.18 sec	590.91 sec	4.79 sec	744.91 sec

Table 5.2: Total processing times, aggregated for all the individual video files, of each separate algorithmic component in the data set — converted to minutes and hours.

Frame extraction	Frame resizing	DPT Ranftl *et al.* (2021)	Feature extraction	Total processing time
86.24 min	109.72 min	747.60 min	10.31 min	953.87 min
1.44 hrs	1.83 hrs	12.46 hrs	0.17 hrs	15.90 hrs

It is conjectured that the processing requirements can be decreased significantly (*i.e.* made faster by many orders of magnitude) through adding more compute capabilities as well as structuring processes in parallel. This improvement might not even be necessary because automated algorithmic processes can be left to their own devices and executed during periods outside of normal working hours until completed.

Chapter summary

The modelling phase of the CRISP-DM framework refers to the solution discovery process for the data sets being analysed. For this project, a comprehensive model workflow is devised

and was depicted in Figure 5.1 as a proposal for the ingestion and algorithmic processing steps required to deliver tree segment features. These output data provide multiple perspectives of a single tree which, in turn, can be linked back to the actual DBH measurement for the relevant tree (regarded as the target feature for ML modelling).

The chapter also described the programming environments, CV and DL toolkits as well as developed algorithms that produce MDE predictions and tree recognition methodologies which extracts the appropriate feature data. From this point, a clustering algorithm was employed to identify single tree clusters — by means of manual intervention assistance, any misrepresentations are rectified thereafter which did not make logical sense.

The spatial analysis of the entire research study area was discussed in order to illustrate the potential of this program to position individual tree coordinates relatively toward its surrounding counterparts. Upon visual inspection, it was evident that the same plot pattern emerges which coincides with the spatial attributes collected during fieldwork. Finally, the runtime for the modelling approach was reported on in order to compare how long the method needed for processing against normal enumerated data gathering.

Chapter 6

Evaluation and Deployment

This chapter is devoted to the design and review of working ML models, evaluated on the basis of common metrics as discussed in the literature review (*i.e.* calculating *bias* and RMSE scores), assessed between actual DBH readings from fieldwork and predicted DBH values gathered from the CV/MDE toolkit. The proposed evaluation approach, intentionally adopted for the present thesis objectives, is similar to the "*zero-shot cross-dataset transfer*" methodology inspired by Ranftl *et al.* (2020) (42).

The chapter also describes the last phase of the CRISP-DM framework, namely how the deployment of such a design model can be instituted. Once all stakeholders of the data project are in agreement that the suggested architecture has reached a sufficient level of operability, the workflow can likely be implemented within a production setting. In this scenario, continual monitoring and management of the model is required for the effective runtime of the system.

6.1 Evaluation

The ML design entails isolating 20% of single tree data records in order for the models not to be exposed to already incorporated training examples. Afterwards, testing is conducted on these isolated entries — therefore a good indication for overall generalisation ability and robustness is obtained. Essentially, the project requires an appropriate ML model that can independently predict DBH measurements by means of ingesting only a few features (*i.e.* midpoints, mean depths and pixel lengths) obtained from the comprehensive CV and MDE framework.

6.1.1 Training

A broad variety of regression ML models were explored, trained and tested but ultimately only five were selected (attributable to their good performance results) for deeper investigation in this thesis — namely *linear regression* (LR), *k-nearest neighbours* (*k*-NN), *decision tree* (DT),

random forest (RF) and an MLP. These specific ML models are described in more detail in the next couple of sections, executed through a data science software program called **Orange Data Mining**, developed by Demšar *el al.* (2013) (103).

Before presenting the relevant data to **Orange**, the original and complete flat input file (with features for midpoints, pixels, depths, FOV angles and unique tree numbers deriving DBH) are appropriately separated to produce training and testing subsets that are structured in a similar format. This entails randomly identifying sixty individual trees (*i.e.* 20% of the total 298 tree records) and isolating its associated data entries for the testing data set. The ML models will train on the remaining 238 trees (*i.e.* 80%) and prediction performance evaluations are made on the aforementioned separate testing data set. Furthermore, the models train on thousands of feature examples for every tree, but these estimations will eventually use the average DBH prediction per unique tree number. Essentially creating an equal weighting for every data point prediction, tied back to the individual tree number.

Additionally, the entire experiment was executed ten times — thereby effectively producing a random testing data set with 600 records in total (independent from training data at each iteration). These results were analysed on the basis of *bias* and RMSE along with the *relative* performance of the stated metrics (*i.e.* percentage deviation) against the direct DBH measurement for the unique tree number. Simplified, this multitude of results on *bias* is visualised by using box plots, as illustrated in Figures 6.1 and 6.2. Comparatively, the two best performing ML models were RF and MLP as reported on in Table 6.1 .

6.1.1.1 Linear regression

LR is regarded as the most common modelling approach in statistics for determining relationships between a single target objective and other data features. LR can be defined as either **simple** LR (*i.e.* one explanatory variable) or **multiple** LR (*i.e.* two or more explanatory variables), as reported by Maulud and Abdulazeez (2020) (104). This algorithm can be modified by means of an L1 (LASSO), L2 (ridge) or L1/L2 (elastic) regularisation technique (104).

For this thesis, an elastic regularisation approach was adopted, therefore a combination of the alpha value (*i.e.* regularisation strength) was set to 0.0001 and an equal elastic net mixing (*i.e.* 50:50) between L1 and L2 was dictated for training. These hyper parameters were chosen after sufficiently exploring the range of its inputs, according to which it tended towards better overall performance. The model produces a learner after the training steps are completed, with associated coefficients for the multivariate input features needed for mapping to the best fitting linear relationship.

6.1.1.2 k-Nearest neighbours

Considered as a simple and supervised ML model, the k-NN algorithm can be utilised to solve regression (as well as classification) problems — it proved effective in this study. The ingested data set displays only a few column features but represents thousands of samples which can slow down the overall process of k-NN training. After execution, the trained data is kept in the compute memory (as there is no generalisation of data points taking place), as reported by Song *et al.* (2017) (105). This ML technique can also employ various distance metrics considering power models (*e.g.* Manhattan, Euclidean, Chebyshev to name a few), between numeric feature values in the case of regression problems.

Nevertheless, the model was capable of structuring nearest neighbour nodes (*i.e.* predicting regression outputs according to the available input features) by utilising the particular data set at hand. Within the **Orange** environment (103), the k-NN ML model aims to predict values according to the nearest training instances and continually considered only two of its neighbours on the basis of a Mahalanobis[1] metric as an evaluated distance weight because it resulted in the best performance delivery.

6.1.1.3 Decision tree

Named appropriately, the DT model in ML builds tree-like structures (with its terminology stemming from, and describing phrases such as "*branches*" or "*leaves*") for either classification or regression challenges. Essentially, the DT model breaks up a given dataset into ever smaller subsets incrementally, according to feature cardinalities — *i.e.* maximum likelihood classifiers, as reported by Xu *et al.* (2005) (107). In the case of regression problems, numeric breakpoints are established by systematically decreasing the entropy of input feature data (*i.e.* nominating the data field with the most explaining power will be divided/split for its succeeding sub-sets).

In the current thesis, the DT was induced as a binary tree with forward pruning and a minimum number of instances in its leaves set to two. The model was also directed not to split subsets smaller than five, and to limit the maximum tree depth to one hundred. An additional criteria was employed so as to stop tree building when the majority of its classifications reaches a 95% inclusivity threshold.

[1]For the purpose of good multi-dimensional generalisation ability, the Mahalanobis distance metric considers correlations between multivariate feature distributions (regarded as the centre point of a principal component analysis), as report by De Maesschalck *et al.* (2000) (106) and is calculated with the inverse variance–covariance matrix. It is conjectured that the Mahalanobis distance metric works better for the data set in the current thesis as numerous inherent cluster formations are established for multiple feature dimensions.

6.1.1.4 Random forest

Regarded as a supervised ML algorithm, an RF utilises an ensemble method of building several randomly induced decision trees, a technique developed by Breiman (2001) (66). The predictive estimation results are delivered as an average of all the estimated classes from separate trees, or mean numeric values (in the case of regression problems). Tree predictors will be combined to form the same distribution of all trees in the forest, such that each tree depends on its randomly sampled induction. RF is more robust toward data noise and this aspect provides the ML model with good generalisation capabilities, however, inference error is dependent on the particular strength of individual trees.

Within the **Orange** working environment, the basic properties of the RF entailed setting the number of trees in each forest to 25. The amount of attributes that were considered at each split was set to three, without a replicable training procedure. On the matter of growth controls, the limit to various depths of individual trees was set to 25, and a constraint to not split subsets smaller than five was dictated.

6.1.1.5 Multilayer perceptron

An MLP algorithm together with back-propagation is available as a learning model in **Orange** (103), labelled as a neural network (*i.e.* NN) in Figures 6.1 and 6.2, which only requires an input data set with an associated target feature field. The number of hidden neurons were set to one hundred, and the ReLU activation function was utilised for the design. This MLP incorporated an Adam method optimiser with regularisation rate set to 0.0001, the maximum number of iterations (with a replicable training method) was set to 200 iterations. These hyper parameters were initially loaded as program defaults, the performance output from the MLP proved to be in line with the other ML models and therefore the default settings were selected for consistency over all the testing experiments.

6.1.2 Testing

The next section refers to an evaluation of the overarching ML model that connects actual DBH measurements from fieldwork with predictions from the CV/MDE approach. Table 6.1 indicates the five best performing ML models that were selected, trained and tested along with its associated performance results. Overall the RF ML model performed best with an RMSE score of 4.365 which is a direct indication of the error distance from target of the DBH readings that were estimated through MDE. The relative *bias* is best for the DT model and second best for an RF instantiation, but overall *bias* is good for all the trained ML models. When these two metrics are regarded in combination, as a final means of deciding on model skill, RMSE

can be viewed as the distance of error away from the actual target observation. Whereas *bias* is the degree of error within this predictive scatter placement — therefore the ML models for this thesis produced a predictive ability which is "tightly" contained (*i.e.* low bias), but rather far off from the desired target feature (*i.e.* RMSE), compared to the reviewed literature.

Table 6.1: Performance metrics tested over applied machine learning models.

Training Model	*bias*	*Relative bias (%)*	RMSE	Relative RMSE (%)
Linear Regression (LR)	0.349	1.126%	5.664	18.284%
k-Nearest Neighbours (kNN)	0.114	0.370%	4.413	14.245%
Decision Tree (DT)	−0.024	−0.078%	4.394	14.184%
Random Forest (RF)	0.052	0.169%	4.365	14.091%
Multilayer Perceptron (MLP)	0.346	1.117%	4.381	14.142%

The bias metric of the different training models is a direct deviation indicator in line with the actual DBH readings over 600 sampled testing records, as displayed in box plot format in Figure 6.1. Moreover, these predictions can be represented relatively (*i.e.* percentage wise) on DBH measurements as illustrated in Figure 6.2 — visually it is noted that the relative bias of the RF ML model exhibits smaller variation.

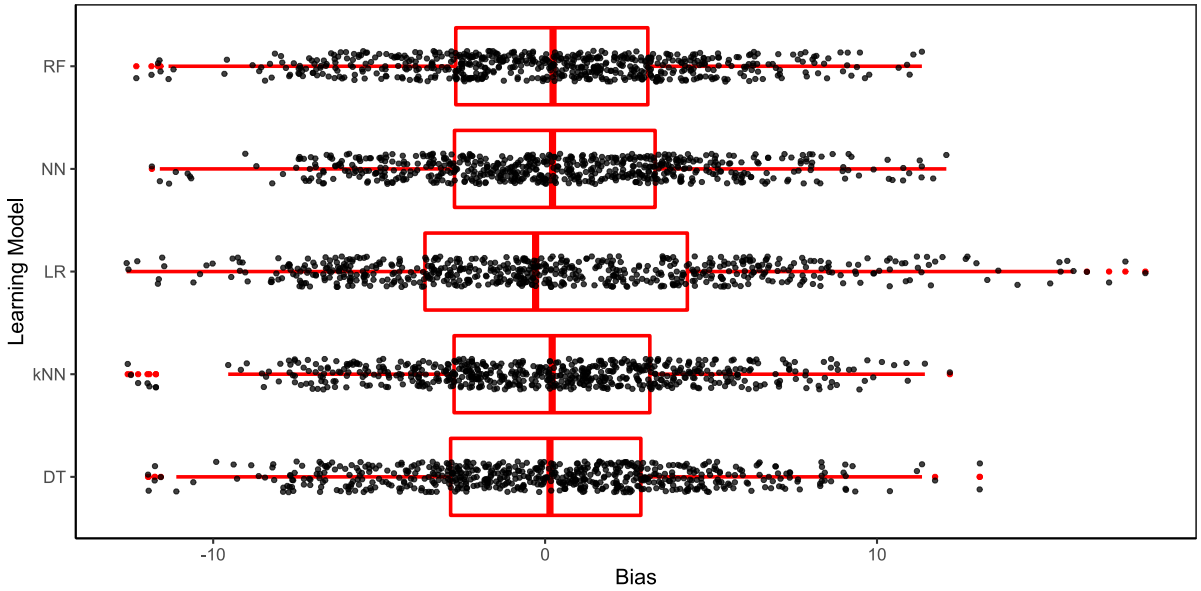


Figure 6.1: Learning Model Results (Bias).

Interestingly, the estimations from predictive learning models largely covers only a close range (*i.e.* DBH 25–35cm) with a much better success rate. To test this hypothesis properly the outlier predictions outside of the aforementioned range was excluded (*i.e.* removing 249 entries from the available 600 test set). This negatively influences the *bias* calculations but

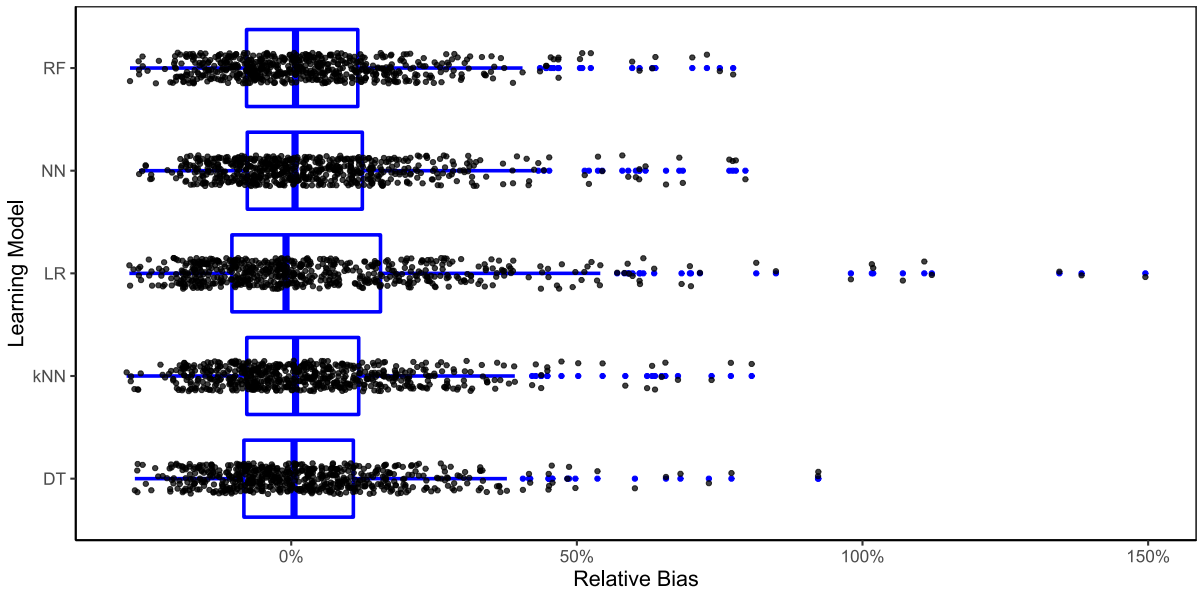


Figure 6.2: Learning Model Results (Relative Bias).

decreases the percentage RMSE scores in a positive manner to an average across all learning models of 3.026% (*i.e.* relative bias) and 9.205% (*i.e.* relative RMSE) respectively. The remaining test samples (*i.e.* 351 entries) holds a mean DBH of 30.54cm which is closely in line with the original fieldwork mean DBH. With this finding it is evident that the developed ML models can properly serve a narrow range of tree DBH data which converges around the mean DBH of the research plot. This is likely due to many learning examples available within this narrow range of training, as opposed to additional outlier data from the study.

In similar fashion, the RMSE and relative RMSE results are visualised in box plot format in Figure 6.3 in order to illustrate the degree of error over the ten experiments that included 60 testing trees each, evaluated against their DBH measurements.

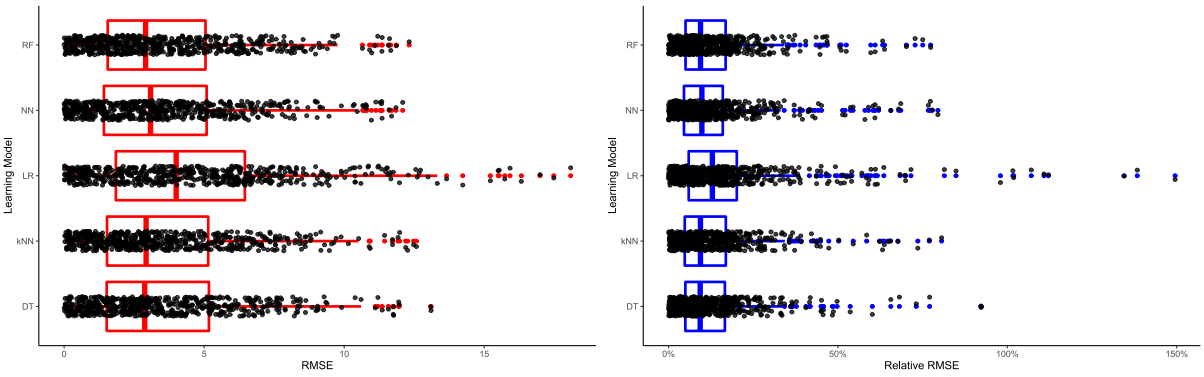


Figure 6.3: Learning Model Results (RMSE and Relative RMSE).

6.2 Deployment

The final phase of the CRISP-DM framework involves addressing matters around deployment of the developed models — which should sufficiently comply with the business requirements and solve the initial problem statement conceived at the start of the project. For the present thesis, the proposed modelling architecture consists of the complete list of activities that will produce the desired outputs, intended for insight analysis and decision making by interested stakeholders or relevant users. The general deployment process entails following the steps in sequential order as conjectured in Figure 5.1 which broadly explains the conceptualised approach.

The deployment requirements will also entail providing written procedures of the devised CRISP-DM phases, describing the necessitated processes of several interrelated components and transitions to achieve objectives effectively. The comprehensive life cycle of the overarching ML model within a production environment also entails the constant management, monitoring and evaluation of its results — this is regarded as a form of continuous calibration by means of retraining models for better accuracies or even encompassing a wider scope of examples.

There is an element of big data with the proposed project — if the solution is ever scaled for larger forestry areas it is evident that the volume and velocity of data produced from fieldwork video data sets will be immense. The production system should be ready and have enough storage capacity to absorb this level of information ingestion — as an example, if the entire compartment had to be recorded (*i.e.* 2.82 *ha*) it would take up **four** times the space as well as processing consumption (because the present study area was approximately 25% of the greater stand). This endeavour would immediately take up about **18 GB** of storage capacity for the video files and would likely result in the total algorithm processing time requiring at least **65 hours** to complete (if computing tasks are not parallelised).

Numerous challenges are prevalent in the domain of big data technologies, but fortunately the appropriate software tools and techniques have been developed to address these issues as reported by Chen and Zhang (2014) (108). They published an article discussing a survey on data-intensive applications and techniques for processing information of extensive magnitudes. Even though the article was published in 2014, considered a long time ago in regards to recent technological developments, their research discussed many of the known software tools still used for large distributed file systems, and techniques for processing certain data types for purposes of insight analysis and visualisation. Furthermore, a general knowledge discovery process is illustrated in their study and mention is made toward big data which yield major opportunities, with the caveat of challenging aspects such as storage and compute to consider when actually constructing such systems.

The value of big data is described by Erl *et al.* (2016) (109) as a greater process — starting with *hindsight* (*i.e.* descriptive analysis), then *insight* (*i.e.* diagnostic analytics and predictive analysis) and progressing towards *foresight* (*i.e.* prescriptive analytics), which also increases systemic complexity dramatically. In their book, the authors address the benefits for *business intelligence*, which is inherently connected to an organisation's *key performance indicators*, as some of the motivating factors for big data adoption, as it leads companies to acquire valuable knowledge and ultimately wisdom for competence in business understanding (109).

Similar views are shared by Loshin (2013) (110) of big data analytics used as an impactful tool for enterprise integration and strategic planning. The author considers that "market conditions (*i.e.* business drivers) have enabled a broad acceptance of big data analytics, including the normalisation of hardware and software, increased data volumes, growing variation in types of data assets for analysis, different methods for data delivery, and increased expectations for real-time integration of analytical results into operational processes." Many of the toolkits described by Chen and Zhang (2014) (108) is also referred to in this textbook and alludes to a common approach for configuring big data systems with concepts spanning across various industries.

The path to production for the current project proposal comprise of varying systematic components — *i.e.* operational data collection, processing algorithms and delivering the information in such a manner which is regarded as user-friendly and easily accessible *via* the appropriate platforms. Toolkits from the domain of data science lies at the core of the conceptualised design, but model inputs are inevitably dependent on the business insights from relevant stakeholders as well as ***data engineering*** practices that are required for efficient system operability. The anticipated ML models are likely to be constructed as part of a bigger data pipeline, leading to training and calibration of the performance of the overall system.

Chapter summary

The final chapter of this thesis can be regarded as the capstone which connects a practical application in forestry and bridging it towards conceptualised solutions in the domain of data science. The evaluation phase of the CRISP-DM framework entailed predicting tree DBH measurements by means of ML models, by incorporating input feature data which maps to a desired target output (*i.e.* actual DBH readings from fieldwork).

This section also described numerous ML techniques and high-lighted the best performing model for achieving DBH estimations on the basis of *bias* and RMSE scoring. The reviewed ML models produced very similar results, where the RMSE was off target by 4–6cm compared

with the actual DBH measurement. The *bias*, however, performed relatively well which can be regarded as repeatedly resulting in the same error at an RMSE distance from the target feature.

The chapter further endeavoured to describe the potential for deploying such a solution for larger forestry areas and addressed the known challenges in terms of big data technologies. Effectively scaling the conceptualised workflow will require immense storage capacity and compute capabilities, but the necessary tools and programming techniques are available in current technological offerings to make this approach a reality.

Chapter 7

Conclusions

The primary objectives of this thesis was to investigate the possibility of collecting tree data and extracting pertinent information (such as DBH measurements) by means of alternative enumeration practices. The motivation for the current project stems from the fact that compartment statistics are determined manually and are only conducted on a per sample plot basis which only represent a fraction of the larger encompassing area. Forest inventories are an important aspect for wood processing enterprises as the information is used for volume estimations and future growth projections of biological assets.

Recent progress in technology and novel data-driven algorithmic processes have spurred on a renaissance for solving practical challenges in new and innovative ways. The physical activity of capturing tree data through advanced hardware, such as LiDAR and UAVs, has been around for the last couple of decades but these devices are often expensive and cannot be carried out effectively in respect of large forest areas. The approach proposed in this thesis focussed on investigating the viability of adopting an ML-based solution to the problem of collecting tree compartment data at scale.

An experiment on a smaller sample plot was therefore conducted at an actual forest area (with access granted by SAFCOL — the industry partner to the project), utilising only the video recording functionality of an ordinary smartphone device. Based on the findings in this project it may be proffered that video processing techniques (*i.e.* CV, DL and MDE using ViT) can be utilised as a more cost effective alternative towards collecting DBH measurements of trees.

The study commenced with the adoption of the well-known data project reference framework (CRISP-DM) and the meticulous execution of the constituent phases — a structured approach toward achieving the proposed research aims was therefore formulated. This document also includes an appropriate literature review which addressed data science concepts as a solution for the practical application domain (*i.e.* remote sensing for tree enumerations in forestry). This section described the industry standard for evaluating proposed techniques (*i.e.* metrics such

bias and RMSE) to measure the accuracy of predictions against actual observations. State-of-the-art techniques have been developed in the relevant literature (which arguably performs much better than the proposed method in the present research), but often this reporting lacks to strike a good balance between its foremost accuracy deliverables and the requirement for surveying vast forest landholdings, which usually encompasses millions of trees.

Initially, the fieldwork data included aspects of less relevance toward DBH readings such as the heights of trees and spatial characteristics of the study area (gathered in order to be thorough during the data exploration stage). The data understanding phase of CRISP-DM in this thesis described these properties in the data set but could not determine a good linear (or non-linear) relationship between DBH and heights of all 298 trees. During forest fieldwork and by carrying out virtual planning, a physical grid outlay was erected — according to which the spatial coordinates of each tree in the research plot could be established. The discussion on data preparation addressed the video files that were recorded at the study area, along with the specifications of each file. By assigning labels to each video recording, an inherent structure could be inferred which assisted the remainder of the algorithm processing steps.

The following chapter focussed on modelling, in which a workflow for data set ingestion and iterative transformations was explained visually through BPMN. The devised workflow included picture frame extraction, resizing of images, and feeding it into a DNN developed by Ranftl *et al.* (2021) (45), called DPT. Subsequently the DBH line pattern data features were extracted by means of script created in the Python programming environment. Furthermore, the DBH strings associated with each sequentially ordered picture was loaded in the R working directory for additional exposure to algorithmic steps including tree segment recognition and clustering.

Ultimately, the output derived from these steps produced a flat file with a small number of data features (*i.e.* midpoints, mean depths and pixel lengths) which could be connected to actual tree data and its associated DBH measurements. It was also possible to generate a relative spatial representation by using the meta data from file names (*i.e.* forest lanes for horizontal placements and picture frame sequence number for vertical positioning). A brief description of the complete runtime for the entire video data set was analysed in order to compare it with the consumption time of collecting the same measurement through normal enumeration practices.

Finally, model evaluation and deployment phases of the CRISP-DM framework were discussed as conclusion to this thesis. The evaluation entailed setting up various ML models which received the same input data — partitioned on the basis of training (80%) and testing (20%) sub-sets. These data sets where isolated on the notion of "zero-shot" learning — accordingly, performances are evaluated against completely unseen tree data in order to improve its generalisation capabilities and robustness for related data ingestion and processing. The best

collective performance was achieved by the RF model in terms of relative *bias* (*i.e.* 0.169%) and RMSE (*i.e.* 14.091%). Furthermore, it was conjectured that if a narrower range of predicted DBH measurements were considered (*i.e.* between 25–35cm), then the RMSE would decrease to approximately 9.20% which is more in line with industry standards for remote sensing. It was argued that a controlled recording environment (*i.e.* artificially constructed) would permit exposure to outlier DBH sizes if repeated for a large number of iterations.

Nevertheless, the comprehensive study of this thesis was interesting and insightful, especially in respect of the challenges and limitations of the particular approaches adopted. Similar future work will build on the foundations that were established in this research. The objectives initially devised for this project were achieved — it is indeed possible to collect tree data by means of simple video recordings, thereby offering a cheaper solution for forest enumerations of larger areas. This can be illustrated in Figure 7.1 which indicates the relative DBH readings of each tree along with its spatial coordinates from actual fieldwork (left), compared with the outputs of the proposed model workflow (right) which showcases the ability to produce relatively similar results upon visual inspection.

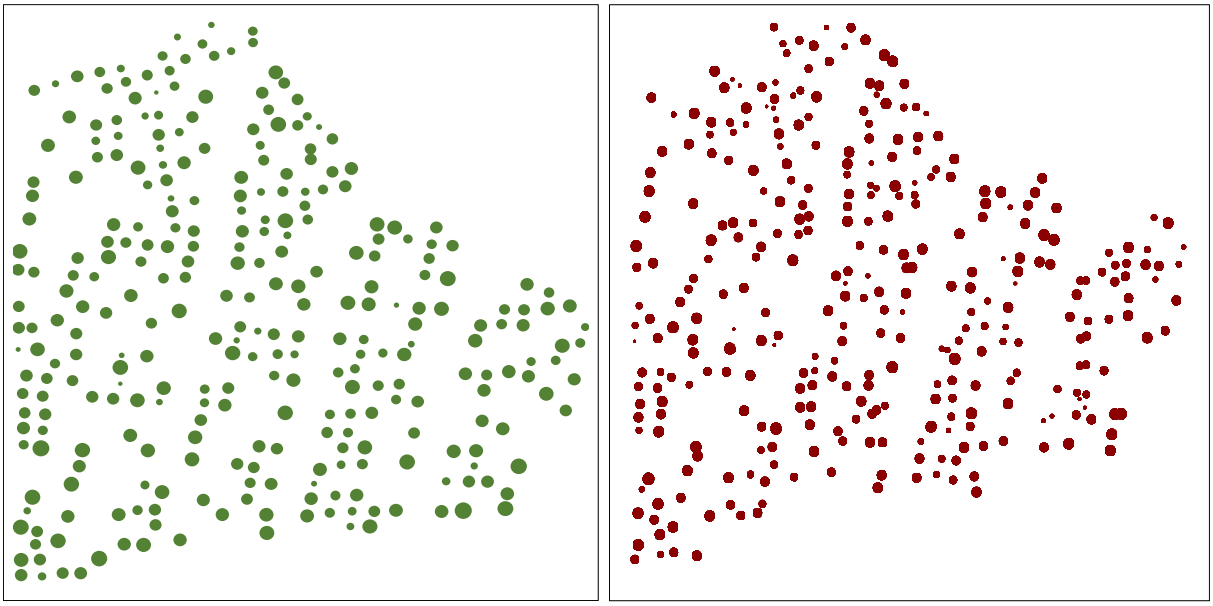


Figure 7.1: Top-view spatial outlay visualisation of all individual trees found in research sub-section from actual fieldwork data, scaled relatively to DBH measurements (left). Spatial reconstruction from model outputs, scaled relatively according to DBH predictions produced from ML workflow (right).

It is conjectured that improved accuracy can be achieved through a controlled environment which should calibrate toward better DBH predictions. Spatial properties could also be determined through a similar process of error reduction — these perspectives serve as conceptualised ideals for applications in precision forestry.

References

[1] S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*. Prentice Hall, 2009. 2

[2] A. P. Engelbrecht, *Computational intelligence: An introduction*. John Wiley & Sons, 2007. 3

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. 3

[4] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014. 3

[5] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden technical debt in machine learning systems," *Advances in neural information processing systems*, vol. 28, pp. 2503–2511, 2015. 4

[6] S. S. Skiena, *The data science design manual*. Springer, 2017. 6, 14

[7] S. S. Skiena, *The algorithm design manual*. Springer International Publishing, 2020. 6

[8] H. J. Zietsman, "The data science process — from data preparation to model evaluation and deployment." Unpublished, 2021-01. 6, 12

[9] L. Cao, "Data science: A comprehensive overview," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, pp. 1–42, 2017. 6, 12

[10] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1, Springer-Verlag London, UK, 2000. 6, 7, 8

[11] M. C. V. Tavares, P. Alencar, and D. Cowan, "A variability-aware design approach to the data analysis modeling process," in *2018 IEEE International Conference on Big Data*, pp. 2818–2827, IEEE, 2018. 9

[12] E. Kristoffersen, O. O. Aremu, F. Blomsma, P. Mikalef, and J. Li, "Exploring the relationship between data science and circular economy: An enhanced crisp-dm process model," in *Conference on e-Business, e-Services and e-Society*, pp. 177–189, Springer, 2019. 9

[13] J. Vauhkonen, "Effects of diameter distribution errors on stand management decisions according to a simulated individual tree detection," *Annals of Forest Science*, vol. 77, no. 2, pp. 1–21, 2020. 10

[14] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997. 11

[15] J. D. Kelleher, B. Mac Namee, and A. D'arcy, *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies*. MIT press, 2020. 12

[16] G. S. Nel, "A brief history and overview of machine learning and artificial intelligence." Unpublished, 2021-02. 12, 13

[17] T. O. Ayodele, "Types of machine learning algorithms," *New advances in machine learning*, vol. 3, pp. 19–48, 2010. 12

[18] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*. Springer, 2005. 12

[19] G. S. Nel, "Feed-forward neural networks." Unpublished, 2021-04. 12, 13, 15

[20] C. C. Aggarwal, "Neural networks and deep learning," *Springer*, vol. 10, pp. 978–3, 2018. 13

[21] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016. 14, 16

[22] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*, pp. 421–436, Springer, 2012. 14

[23] M. Zinkevich, M. Weimer, A. J. Smola, and L. Li, "Parallelized stochastic gradient descent," in *NIPS*, vol. 4, p. 4, Citeseer, 2010. 14

[24] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," *Advances in neural information processing systems*, vol. 26, pp. 315–323, 2013. 14

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 14, 56

[26] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research*, vol. 12, no. 7, 2011. 14

[27] R. Szeliski, *Computer vision: Algorithms and applications*. Springer, 2021. 15

[28] Y. Ming, X. Meng, C. Fan, and H. Yu, "Deep learning for monocular depth estimation: A review," *Neurocomputing*, 2021. 16

[29] D. E. Rumelhart and J. L. McClelland, "On learning the past tenses of english verbs," tech. rep., California University, 1985. 16

[30] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European conference on computer vision*, pp. 746–760, Springer, 2012. 16

[31] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving: The kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354–3361, IEEE, 2012. 16

[32] Z. Li, T. Dekel, F. Cole, R. Tucker, N. Snavely, C. Liu, and W. T. Freeman, "Learning the depths of moving people by watching frozen people," 2019. 16, 55

[33] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," 2014. 16, 17

[34] W. Chen, Z. Fu, D. Yang, and J. Deng, "Single-image depth perception in the wild," *Advances in neural information processing systems*, vol. 29, pp. 730–738, 2016. 17

[35] X. Chen, X. Chen, and Z. J. Zha, "Structure-aware residual pyramid network for monocular depth estimation," *arXiv preprint arXiv:1907.06023*, 2019. 17

[36] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 270–279, 2017. 17

[37] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow, "Digging into self-supervised monocular depth estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3828–3838, 2019. 17

[38] L. Andraghetti, P. Myriokefalitakis, P. L. Dovesi, B. Luque, M. Poggi, A. Pieropan, and S. Mattoccia, "Enhancing self-supervised monocular depth estimation with traditional visual odometry," in *2019 International Conference on 3D Vision (3DV)*, pp. 424–433, IEEE, 2019. 17

[39] R. Wang, S. M. Pizer, and J.-M. Frahm, "Recurrent neural network for un-supervised learning of monocular video visual odometry and depth," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5555–5564, 2019. 17

[40] X. Luo, J.-B. Huang, R. Szeliski, K. Matzen, and J. Kopf, "Consistent video depth estimation," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 71–1, 2020. 18

[41] F. Aleotti, G. Zaccaroni, L. Bartolomei, M. Poggi, F. Tosi, and S. Mattoccia, "Real-time single image depth perception in the wild with handheld devices," *Sensors*, vol. 21, no. 1, 2021. 18

[42] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," 2020. 18, 55, 70

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017. 19

[44] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. 19, 20

[45] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," 2021. 20, 55, 56, 66, 80

[46] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3286–3295, 2019. 20

[47] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," in *International Conference on Machine Learning*, pp. 4055–4064, PMLR, 2018. 20

[48] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *arXiv preprint arXiv:1906.05909*, 2019. 20

[49] H. Choudhry and G. O'Kelly, "Precision forestry: A revolution in the woods," *McKinsey & Company*, 2018. 20, 21

[50] Food and Agricultural Organization (FAO), "Global forest resources assessment 2015," *Desk Reference*, 2015. 20

[51] N. A. Clark, R. H. Wynne, and D. L. Schmoldt, "A review of past research on dendrometers," *Forest Science*, vol. 46, no. 4, pp. 570–576, 2000. 21, 46

[52] D. M. Drew and G. M. Downes, "The use of precision dendrometers in research on daily stem size and wood property variation: A review," *Dendrochronologia*, vol. 27, no. 2, pp. 159–172, 2009. 21

[53] J. R. Kellner, J. Armston, M. Birrer, K. Cushman, L. Duncanson, C. Eck, C. Falleger, B. Imbach, K. Král, *et al.*, "New opportunities for forest remote sensing through ultra-high-density drone lidar," *Surveys in geophysics*, vol. 40, no. 4, pp. 959–977, 2019. 22

[54] Q. Guo, Y. Su, T. Hu, H. Guan, S. Jin, J. Zhang, X. Zhao, K. Xu, D. Wei, M. Kelly, *et al.*, "Lidar boosts 3d ecological observations and modelings: A review and perspective," *IEEE Geoscience and Remote Sensing Magazine*, no. 99, pp. 0–0, 2020. 23

[55] B. Talbot, M. Pierzchała, and R. Astrup, "Applications of remote and proximal sensing for improved precision in forest operations," *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, vol. 38, no. 2, pp. 327–336, 2017. 23

[56] C. Mulverhill, N. C. Coops, P. Tompalski, C. W. Bater, and A. R. Dick, "The utility of terrestrial photogrammetry for assessment of tree volume and taper in boreal mixedwood forests," *Annals of Forest Science*, vol. 76, no. 3, pp. 1–12, 2019. 23

[57] L. A. Wells and W. Chung, "Evaluation of ground plane detection for estimating breast height in stereo images," *Forest Science*, vol. 66, no. 5, pp. 612–622, 2020. 23, 24

[58] M. I. Marzulli, P. Raumonen, R. Greco, M. Persia, and P. Tartarino, "Estimating tree stem diameters and volume from smartphone photogrammetric point clouds," *Forestry: An International Journal of Forest Research*, vol. 93, no. 3, pp. 411–429, 2020. 24, 28

[59] J. Iglhaut, C. Cabo, S. Puliti, L. Piermattei, J. O'Connor, and J. Rosette, "Structure from motion photogrammetry in forestry: A review," *Current Forestry Reports*, vol. 5, no. 3, pp. 155–168, 2019. 24, 28

[60] J. Liu, Z. Feng, L. Yang, A. Mannan, T. U. Khan, Z. Zhao, and Z. Cheng, "Extraction of sample plot parameters from 3d point cloud reconstruction based on combined rtk and ccd continuous photography," *Remote Sensing*, vol. 10, no. 8, p. 1299, 2018. 24, 28

[61] T. Mikita, P. Janata, and P. Surovỳ, "Forest stand inventory based on combined aerial and terrestrial close-range photogrammetry," *Forests*, vol. 7, no. 8, p. 165, 2016. 24, 28

[62] M. Mokroš, J. Výbošt'ok, J. Tomaštík, A. Grznárová, P. Valent, M. Slavík, and J. Merganič, "High precision individual tree diameter and perimeter estimation from close-range photogrammetry," *Forests*, vol. 9, no. 11, p. 696, 2018. 24, 25, 28

[63] E. Seifert, S. Seifert, H. Vogt, D. Drew, J. Van Aardt, A. Kunneke, and T. Seifert, "Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images," *Remote sensing*, vol. 11, no. 10, p. 1252, 2019. 25

[64] R. Dainelli, P. Toscano, S. F. Di Gennaro, and A. Matese, "Recent advances in unmanned aerial vehicle forest remote sensing — a systematic review. part i: A general framework," *Forests*, vol. 12, no. 3, p. 327, 2021. 25

[65] R. Dainelli, P. Toscano, S. F. D. Gennaro, and A. Matese, "Recent advances in unmanned aerial vehicles forest remote sensing — a systematic review. part ii: Research applications," *Forests*, vol. 12, no. 4, p. 397, 2021. 25

[66] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. 25, 73

[67] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, pp. 278–282, IEEE, 1995. 25

[68] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, V. Vapnik, *et al.*, "Support vector regression machines," *Advances in neural information processing systems*, vol. 9, pp. 155–161, 1997. 25

[69] L. E. Peterson, "k-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009. 25

[70] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967. 25

[71] O. Nevalainen, E. Honkavaara, S. Tuominen, N. Viljanen, T. Hakala, X. Yu, J. Hyyppä, H. Saari, I. Pölönen, N. N. Imai, *et al.*, "Individual tree detection and classification with uav-based photogrammetric point clouds and hyperspectral imaging," *Remote Sensing*, vol. 9, no. 3, p. 185, 2017. 25

[72] S. Briechle, P. Krzystek, and G. Vosselman, "Silvi-net — a dual-cnn approach for combined classification of tree species and standing dead trees from remote sensing data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 98, p. 102292, 2021. 26

[73] K. Calders, J. Adams, J. Armston, H. Bartholomeus, S. Bauwens, L. P. Bentley, J. Chave, F. M. Danson, M. Demol, M. Disney, *et al.*, "Terrestrial laser scanning in forest ecology: Expanding the horizon," *Remote Sensing of Environment*, vol. 251, p. 112102, 2020. 26

[74] Y. Wang, A. Kukko, E. Hyyppä, T. Hakala, J. Pyörälä, M. Lehtomäki, A. El Issaoui, X. Yu, H. Kaartinen, X. Liang, *et al.*, "Seamless integration of above-and under-canopy unmanned aerial vehicle laser scanning for forest investigation," *Forest Ecosystems*, vol. 8, no. 1, pp. 1–15, 2021. 26, 28

[75] I. Balenović, X. Liang, L. Jurjević, J. Hyyppä, A. Seletković, and A. Kukko, "Hand-held personal laser scanning — current status and perspectives for forest inventory application," *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, vol. 42, no. 1, pp. 165–183, 2021. 26, 27, 28

[76] P. Wilkes, A. Lau, M. Disney, K. Calders, A. Burt, J. G. de Tanago, H. Bartholomeus, B. Brede, and M. Herold, "Data acquisition considerations for terrestrial laser scanning of forest plots," *Remote Sensing of Environment*, vol. 196, pp. 140–153, 2017. 27

[77] Y. Wang, J. Pyörälä, X. Liang, M. Lehtomäki, A. Kukko, X. Yu, H. Kaartinen, and J. Hyyppä, "In situ biomass estimation at tree and plot levels: What did data record and what did algorithms derive from terrestrial and aerial point clouds in boreal forest," *Remote Sensing of Environment*, vol. 232, p. 111309, 2019. 27

[78] E. Hyyppä, A. Kukko, R. Kaijaluoto, J. C. White, M. A. Wulder, J. Pyörälä, X. Liang, X. Yu, Y. Wang, H. Kaartinen, *et al.*, "Accurate derivation of stem curve and volume using backpack mobile laser scanning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 246–262, 2020. 27, 28, 46

[79] E. Hyyppä, J. Hyyppä, T. Hakala, A. Kukko, M. A. Wulder, J. C. White, J. Pyörälä, X. Yu, Y. Wang, J.-P. Virtanen, O. Pohjavirta, X. Liang, M. Holopainen, and H. Kaartinen, "Under-canopy uav laser scanning for accurate forest field measurements," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 41–60, 2020. 27, 28, 46

[80] J. Hyyppä, J.-P. Virtanen, A. Jaakkola, X. Yu, H. Hyyppä, and X. Liang, "Feasibility of google tango and kinect for crowdsourcing forestry information," *Forests*, vol. 9, no. 1, p. 6, 2018. 27, 28

[81] Y. Fan, Z. Feng, A. Mannan, T. U. Khan, C. Shen, and S. Saeed, "Estimating tree position, diameter at breast height, and tree height in real-time using a mobile phone with rgb-d slam," *Remote Sensing*, vol. 10, no. 11, p. 1845, 2018. 27, 28, 29

[82] J. J. Moré, "The levenberg-marquardt algorithm: Implementation and theory," in *Numerical analysis*, pp. 105–116, Springer, 1978. 27

[83] X. Liang, J. Hyyppä, H. Kaartinen, M. Lehtomäki, J. Pyörälä, N. Pfeifer, M. Holopainen, G. Brolly, P. Francesco, J. Hackenberg, *et al.*, "International benchmarking of terrestrial laser scanning approaches for forest inventories," *ISPRS journal of photogrammetry and remote sensing*, vol. 144, pp. 137–179, 2018. 28

[84] X. Liang, Y. Wang, A. Jaakkola, A. Kukko, H. Kaartinen, J. Hyyppä, E. Honkavaara, and J. Liu, "Forest data collection using terrestrial image-based point clouds from a handheld camera compared to terrestrial and personal laser scanning," *IEEE transactions on geoscience and remote sensing*, vol. 53, no. 9, pp. 5117–5132, 2015. 28

[85] X. Liang, V. Kankare, J. Hyyppä, Y. Wang, A. Kukko, H. Haggrén, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, *et al.*, "Terrestrial laser scanning in forest inventories," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 63–77, 2016. 28

[86] X. Liang, A. Kukko, J. Hyyppä, M. Lehtomäki, J. Pyörälä, X. Yu, H. Kaartinen, A. Jaakkola, and Y. Wang, "In-situ measurements from mobile platforms: An emerging approach to address the old challenges associated with forest inventories," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 97–107, 2018. 28

[87] X. Liang, Y. Wang, J. Pyörälä, M. Lehtomäki, X. Yu, H. Kaartinen, A. Kukko, E. Honkavaara, A. E. Issaoui, O. Nevalainen, *et al.*, "Forest in situ observations using unmanned aerial vehicle as an alternative of terrestrial measurements," *Forest Ecosystems*, vol. 6, no. 1, pp. 1–16, 2019. 28

[88] G. Brolly, G. Király, M. Lehtomäki, and X. Liang, "Voxel-based automatic tree detection and parameter retrieval from terrestrial laser scans for plot-wise forest inventory," *Remote Sensing*, vol. 13, no. 4, p. 542, 2021. 28

[89] S. Chen, H. Liu, Z. Feng, C. Shen, and P. Chen, "Applicability of personal laser scanning in forestry inventory," *Plos One*, vol. 14, no. 2, p. e0211392, 2019. 28

[90] M. Pierzchała, P. Giguère, and R. Astrup, "Mapping forests using an unmanned ground vehicle with 3d lidar and graph-slam," *Computers and Electronics in Agriculture*, vol. 145, pp. 217–225, 2018. 29

[91] "SAFCOL annual integrated report 2019-2020." http://www.safcol.co.za/wp-content/uploads/2017/09/SAFCOL-INTERGRATED-REPORT-2019-2020.pdf. Accessed: 2021-05-11. 31, 32, 33

[92] "SAFCOL business / organisational structure." http://www.safcol.co.za/who-we-are/our-structure/. Accessed: 2021-05-14. 32

[93] A. Wise, "A spatial approach to edge effect modelling for plantation forestry," Master's thesis, Stellenbosch: Stellenbosch University, 2013. 36

[94] C. Kleinn, *Lecture notes for the teaching module forest inventory.* University of Göttingen, 2007. 36

[95] R. E. McRoberts and J. A. Westfall, "Propagating uncertainty through individual tree volume model predictions to large-area volume estimates," *Annals of Forest Science*, vol. 73, no. 3, pp. 625–633, 2016. 38

[96] R. O. Curtis and D. D. Marshall, "Why quadratic mean diameter?," *Western Journal of Applied Forestry*, vol. 15, no. 3, pp. 137–139, 2000. 45

[97] A. A. Arabatzis and H. E. Burkhart, "An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations," *Forest science*, vol. 38, no. 1, pp. 192–198, 1992. 46

[98] M. Chinosi and A. Trombetta, "Bpmn: An introduction to the standard," *Computer Standards & Interfaces*, vol. 34, no. 1, pp. 124–134, 2012. 53

[99] S. A. White, "Introduction to bpmn," *IBM Cooperation*, vol. 2, 2004. 53

[100] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," *arXiv preprint arXiv:1810.04650*, 2018. 56

[101] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967. 63

[102] C. Nvidia, "Nvidia cuda c programming guide, version 4.2," *NVIDIA: Santa Clara, CA*, 2010. 66

[103] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, *et al.*, "Orange: Data mining toolbox in python," *the Journal of machine Learning research*, vol. 14, no. 1, pp. 2349–2353, 2013. 71, 72, 73

[104] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, 2020. 71

[105] Y. Song, J. Liang, J. Lu, and X. Zhao, "An efficient instance selection algorithm for k-nearest neighbor regression," *Neurocomputing*, vol. 251, pp. 26–34, 2017. 72

[106] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000. 72

[107] M. Xu, P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, "Decision tree regression for soft classification of remote sensing data," *Remote Sensing of Environment*, vol. 97, no. 3, pp. 322–336, 2005. 72

[108] C. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information sciences*, vol. 275, pp. 314–347, 2014. 76, 77

[109] T. Erl, W. Khattak, and P. Buhler, *Big data fundamentals: Concepts, drivers & techniques*. Prentice Hall Press, 2016. 77

[110] D. Loshin, *Big data analytics: From strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Elsevier, 2013. 77