Digital Twin Techniques for Managing Natural Regrowth Forests

Stan Clark

4th Year Project Report
Artificial Intelligence and Computer Science
School of Informatics
University of Edinburgh

2025

Abstract

This study examines the effectiveness of varying digital twin techniques in predicting and quantifying the state of natural regrowth forests compares them.

This was done using LiDAR scans of a natural regrowth region of Aberfoyle forest in 2012 and 2021. From this LiDAR scan individual trees were identified and matched across the dataset. This data now in tabular form was preprocessed and cleaned so it could be used in neural networks to predict changes in height and crown area across the snapshots. The models would use the traits of an individual tree and information on surrounding trees to make a prediction. Two methods of vectorizing neighbours were compared, one presenting data in its rawest form using K-nearest-neighbour representation and the other a more abstracted representation showing statistics of trees within a radius. Neural networks were trained to predict tree growth, evaluating varying architectures (block, funnel, hourglass), activation functions (PReLU, ELU, Mish), and varying prediction methods (direct future value vs. normalized difference).

Results showed that neighbour statistics with a funnel architecture and PReLU activation resulted in the lowest mean relative error of 10.25% for height prediction, while crown area predictions were less accurate with 51.74% error. Despite the high error, analysis on features still shed light on important traits to prediction. These results highlight the advantages of data abstraction in improving model performance and identifies challenges in data preprocessing for large data sets of segmented LiDAR scans, such as tree matching and noise filtering. The findings contribute to digital twin applications in forestry by demonstrating actionable, tree-level insights for natural regrowth management.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Stan Clark*)

Acknowledgements

Thank you to my friends and family who supported me while I worked on this project.

Thank you to my supervisor Michael Herrmann who guided me throughout working on this project.

Thank you to Juan Suarez-Minguez of The Forest Research - Forestry Commission who supplied the data used in this project and helped me understand the ecology relevant to the project.

Finally thank you to the Sad Boyz podcast for entertaining me as I listened to it several times through when working on this project

Table of Contents

Chapter 1

Introduction

1.1 Motivation

As climate change and biodiversity loss continue to accelerate ensuring forests stay healthy, diverse and efficient carbon sinks becomes more important [Trumbore et al.]. Plantation style forests do not fulfil these needs however, natural regrowth forests aid in the support of diverse ecology, are more resistant to disease and sequester carbon at a higher rat. For this reason, new policy has been released pushing forest managers to move towards natural regrowth [Raum, 2015] [Malcolm et al., 2001]. These methods of forest management still require oversight especially in a changing climate but these kinds of forests are harder to manage and monitor due to their lack of structure compared to trees evenly distributed in a grid in plantation style forests. A solution to this issue can be found in digital twins, virtual representations of real world systems, that catalogue, monitor and predict traits of the system they represent. This study explores how digital twin techniques can be applied to data of high spatial resolution to gain insights on the state of the forest on a per tree basis and applys machine learning to this data to capture and predict complex ecological patterns.

1.2 Approach

This study uses two LiDAR scans of Aberfoyle forest taken in 2012 and 2021. From these, individual trees were segmented and matched to their corresponding instance across the datasets. This allowed the measurement of changes in height and crown area, key proxies for forest health and carbon sequestration.

Neural networks were trained to predict these target variables based on information on the tree and nearby trees. Two methods of representation were compared to explore the effects of higher and lower levels of abstraction when presenting the data. It was also tested which model architectures, activation functions and predictions method were best suited to this problem. The models created were analysed in depth to give a high amount on details of what methods are applicable.

1.3 Project Contributions

This project contributes the following to the space of digital twins within forestry:

- A pipeline for cleaning, matching, and processing large datasets on individual trees derived from LiDAR data into a clean, suitable format for machine learning.
- A comparison of two contrasting neighbour vectorization methods for encoding the effects of trees on one another.
- An exhaustive evaluation of a range of neural networks with varying structures and prediction methods, examining their effectiveness and identifying the traits that improve performance.
- An analysis of models performance relative to specific forest traits.
- Insights into how more granular data and data processing techniques can be applied to digital twins use in managing natural regeneration forests.

Chapter 2

Background

2.1 Introduction to Digital Twins

2.1.1 History of Digital Twins

Digital Twins, DTs exist as a digital copy of an existing physical system referred to as the physical twin with a flow of data between them known as the physical thread [Grieves]. These have many purposes depending on the context in which they are used. However almost all are used to keep up to date information on individual entities within the system for monitoring, diagnostic and prediction purposes. This is often then used to inform any changes to or actions performed by the physical twin [Kritzinger et al.]. The concept for a digital twin was pioneered by NASA who referred to it as a living model [NAS]. This used internal physics simulations and an incoming stream of data from various sensors to conduct failure analysis in real time.

Upon being adopted by the manufacturing industry the methods for creating digital twins were further refined [Grieves, 2015]. Despite this, exact definition for a DT had not been settled on as the specification and scope would change depending on what was being modelled. However a broadly agreed upon concept would be needed for effective communication. VanDerHorn and Mahadevan identified the lack of a clear definition of the DT and proposed the following as a suitable definition:

> ”a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual systems” [VanDerHorn and Mahadevan]

Within manufacturing, where most DTs resided, they would be used for monitoring efficiency, ensuring certain levels of quality, and managing dynamic variables during production [Grieves, 2015]. As DTs were adopted by the wider world and applied to more abstract physical twins they would also be used for prediction and diagnostics [Grieves].

2.1.2 Types of Digital Twins

As the definition of DTs were defined, classifications were made for varying levels of implementation: Digital Model, Digital Shadow and Digital Twin. These represent increasing levels of integration between the physical twin and digital counterpart, although many use the term interchangeably with the lower levels of implementation. [Kritzinger et al.]. fig 2.1

Digital Model

In this context, a digital model is a digital representation of a real-world object or a plan for a real-world object with no automatic transfer of data between the between the physical and digital components. These are often used for planning, simulation and mathematical modelling [Tagarakis et al.]. Any changes to the physical component would not affect the digital and vice versa.

Digital Shadow

Building on the concept for the digital model a digital shadow has an automatic one directional flow of data from the physical component to the digital. Any changes applied to the physical component would also happen to the digital counter part but not in reverse. These are often used for monitoring, diagnostics and analysis [Kritzinger et al.].

Digital Twin

This is the fully realized implementation of the concept. A fully implemented Digital Twin has a bi-directional flow of data between the physical and digital component. Changes in the physical twin will result in those changes appearing in the digital twin and changes in the digital twin will result in changes to the physical. This is often done to create a controlling digital instance of the physical twin to manage it.

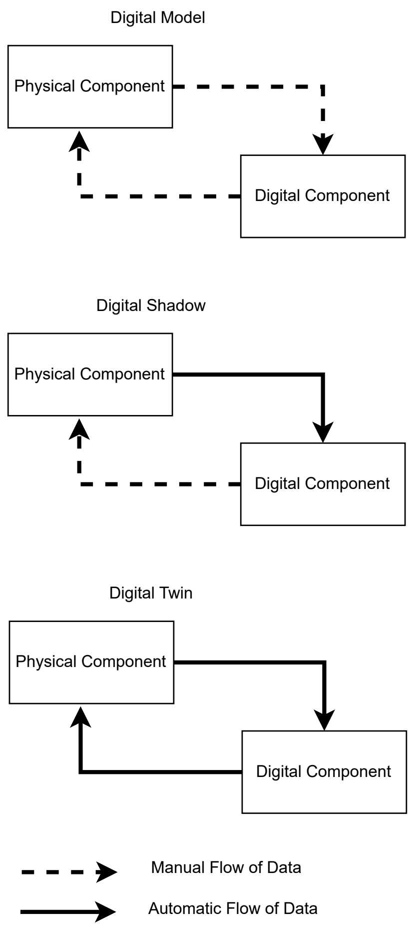

Figure 2.1: Levels of Implementations for DTs

For example, a climate controlled greenhouse could use a DT where the digital twin gets readings of temperature and humidity from the physical twin but can also change these variables itself by controlling components of the physical twin.

2.1.3 Digital Twins use Within Forestry

Within environmental and life science, DTs exist in a different context with a more muddied definition as the standard definition does not account for the open nature of life sciences. A new definition would be made to better fit the context.

> ”Digital twin in the unpredictable environments of agriculture and forestry can be defined as a virtual representation of a physical system, dynamic or static, and its associated environment and processes, both biotic and abiotic, that at full implementation may be updated through the exchange of information between the physical and virtual systems”. [Tagarakis et al.]

In recent years the use of DTs in agriculture and forestry has risen[Tagarakis et al.]. Within Forestry, DTs are used to monitor entire forests. Depending on the scope of what is needed for the digital twin it may model stands of trees, soil condition, weather, humidity, animals within the forest and even individual trees. When doing forestry that prioritises the well being of wildlife, non-intrusive ways of measuring and applying data from the DT is required. Because of the difficulty in this and the size of forests most DT of forests are actually digital shadows [Tagarakis et al.].

Forest Monitoring and Management:

A DT of a forest can be used to keep a record of tree growth, species distribution, and soil conditions. These are conditions relevant to tree growth but also those relevant to drainage and flooding. This will allow for closer monitoring of important variables relevant to tree health, pest infestations, the spread of disease, flooding and wildfire. They can also be used to create scene reconstructions allowing the forest to be viewed digitally as a way of interacting with the DT. These aid in informing decisions around management and enabling proactive interventions to problems.

Predictive Analytics and Decision Support

Using the information stored in the digital twin, predictions can made on growth, carbon sequestration, soil conditions and tree health. These predictions can be used to create simulations of what is likely to change as a result of certain events. This can model what will happen under changes to the climate specific seasons of weather and can also show what will change if certain actions are taken like logging or thinning an area. These predictions will show how these events affect those mentioned in the monitoring section.

By using a DT in this way the repercussions of actions can be more accurately predicted, management can be prepared for a wider range of future scenarios and forest management strategies can be optimized.

2.2 Methods of Digital Twins within Forestry

Creating DTs of forests faces challenges unique to the common use cases of digital twins. It is done over a much larger area than most manufacturing use cases, and methods used to observe data and implement change from the DT to the physical twin are often required to be less intrusive.

2.2.1 Sensing

UAV/Drones

Unmanned aerial vehicles, UAVs are common in gathering data for DTs as they are cost-effective compared to manned surveys. They are also time-saving as they can cover large areas quickly enabling more frequent monitoring and because they can get close enough to gather a high amount of detail [Awais et al.]. In the context of forestry, this is often in the form of overhead imagery and 3D point clouds. This method of collecting data is fairly nonintrusive as does not alter the surroundings as is not a permanent fixture, however, the noise created by drones can disrupt wildlife if scans are performed extremely often [Wallace et al.].

Satellite

The cheapest and least intrusive method of data collection by far is satellite imagery [Madeira et al.]. Satellite imagery can be used to classify areas of vegetation and collect statistics on canopy coverage; the albedo of regions; information on the wavelengths of lights reflected within that pixel or the distance between canopy and ground if the satellite uses radar [Johansen et al.]. The applications of this method is limited by the satellites that pass over the area and the resolution relative to the ground as well as the satellite's radar capabilities.

Soil Sensors

Soil sensor installed throughout the forest can give important data on the ground the trees grow in. These can be used to infer what may be affecting a region's growth or to directly monitor important properties like drainage. Soil sensors often collect data on soil's nitrogen, carbon and organic matter content. In addition to this, they can also collect data on soil moisture, salinity, temperature and even compaction although that requires more intrusive techniques [Pajares et al.].

Weather Stations

Weather stations can be used to gather supplementary data that often has a strong effect on forest health. A history of weather information along with digital twin knowledge can be used to view weather's effect on the specific physical twin that is being viewed so that the effect of future weather events can be predicted. Weather stations collect data on air temperature, relative humidity, dew point, barometric pressure, wind direction and speed, rainfall and solar radiation levels [Zachariassen et al.].

Knowledge Databases

The most common source of data for DT twins is knowledge databases [Tagarakis et al.]. These allow even new digital twins to have a historic record of information and forest dynamics to work off before higher granularity digital twin data is collected over a large time span. This data is often used to fill in gaps in datasets, create artificial environments and leverage expert knowledge in predictive methods.

2.2.2 Metrics

Protecting forest health is a priority of DTs within forestry, however ”forest health” is an ambiguous term and not strictly defined. Organisations take metrics to represent forest health, these being: soil fertility, total biomass, new growth of wood, leaf area, tree mortality rates and ecological biodiversity among trees and the wildlife they support [Trumbore et al.]. The ”robustness” of the forest is also important, this is how quickly these values return to normal after disturbance [Trumbore et al.]. These values are often prioritised differently depending on forest management especially between conservation forests and lumber forests.

This study will primarily focus on the metrics of total biomass, new wood growth and leaf area via the proxy variables of height and crown area. These selections have been made due to their relevance to the dataset.

2.2.3 Prediction Methods

Historic methods used to predict growth and other important factors in forestry fall into two categories: individual tree level models and stand level models, where a stand refers to an area of forest. Most use the attributes of an individual tree or the average of the stand to make predictions based on regression developed on a per species basis [Dale et al.]. These are useful and trusted as they have been used for a long amount of time, however, these regressions do not account for the effects of neighbouring trees and is unlikely to be bespoke to the specific climate where they are being applied.

Other methods incorporate competition indices to capture the effect that nearby trees have on the growth of an individual. Among trees planted at the same time and or of the same species simpler models performed just as well as ones incorporating competition indices however in more natural arrangements of trees these additions had a positive affect on accuracy [Biging and Dobbertin].

2.2.4 Challenges Faced by Digital Twins in Forestry

Digital twins’ original purpose was for managing and diagnosing problems in highly controlled closed-off environments, like the systems on a space craft [NAS] or machinery in a semi autonomous factory [Grieves, 2015]. When applying this technique to forestry the normal framework must be modified to account for more objects. Instead of 1000s of machines at most, there are often 100,000s of trees in a digital twin of a forest. It must also be modified to handle less predictable variables as the natural system of a forest has far more noise and the relationship between variables is highly interconnected. To solve this, more sophisticated methods of prediction would be needed as cause and effect is far easier to quantify in the closed-off controlled system compared to in forestry. Forest Digital Twins must also address the volume of data by breaking down regions into reasonable sizes and using feature extraction along with abstraction to avoid being bogged down by the volume of information while still deriving meaning.

Forest Digital twins must also account for an inflow of data that occurs less often as data must be taken over a large area and permanent infrastructure cannot be used to gather all data as it is likely to intrusive on the natural environment. The inflow of data is less compared to digital twins in manufactured structures as sensors can be directly installed and give the digital twin real time updates on the physical twin.

2.3 Related Works and Critiques

With the rise in fidelity of remote sensing [Zachariassen et al.], many studies have been done investigating the feasibility of a digital twin to be used within forestry [Tagarakis et al.].

One such study, ”Forestry Digital Twin With Machine Learning in Landsat 7 Data” [Jiang et al., 2022] focused on developing methods for projecting how a region of forest would appear in satellite imagery into the future using past images. This makes good use of the most accessible data for use in forestry digital twins [Madeira et al.] and uses it to its fullest. However, this method is far less granular compared to the methods used in this study as it only returns predictions as satellite imagery whereas the methods in this study use more detailed data to make predictions on a per tree basis. Because of this, their study can only make predictions on the general state of an area. In contrast the models created in this study's more granular approach allows them to identify individual trees that are predicted to struggle and can be ran on edited datasets to show what would happen if certain trees were removed or new ones planted. This makes the more granular approach much better suited to existing within a forestry digital twin as its outputs are more applicable to informing decisions in forest management.

The paper ”Artificial neural networks as an alternative method to nonlinear mixed-effects models for tree height predictions” [Skudnik and Jevšenak, 2022] is similar to this study in how it examines the way in which higher granularity data and neural network methods can be used to outperform traditional techniques to predict tree height that operate on a stand level. Their study used a higher density of samples and competition indexing to create a model that did better than the traditional method they compared it to. Although competition indexing helped their models use the higher density of samples to predict tree growth, more features could be taken from the higher density of samples using such methods as used in this study. This study made full use of the higher density of samples in order to learn more complex relationships between trees.

Chapter 3

Methods

Tabular data on the trees within Aberfoyle forest was collected by the Research and Forestry Commission. This data will be used to train neural networks which will measure and predict how the traits of surrounding trees affect the growth of an individual tree's height and crown area. Two snapshots were taken at different times so the change in variables could measured. These traits were shown as they can be used as markers for forest health and carbon sequestration [Pugh et al., 2019]. Multiple methods of presenting data and predicting growth will be created so they can be analysed.

3.1 Data

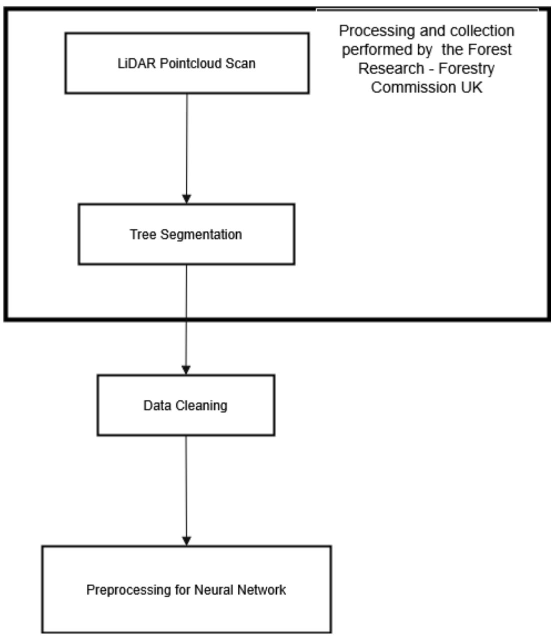

Figure 3.1: Data processing pipeline showing portion done by The Forest Research - Forestry Commission

3.1.1 Data Collection

The raw data from Aberfoyle forest was collected by UAVs performing LiDAR scans that produced point clouds. Scans were performed at an angle of 10 degrees flying at 1000 metres above ground level. The scan taken in 2012 was performed with a point density of 40 per cubic metre by The Environment Agency whereas the scan taken in 2021 was performed with a point density of 60 per cubic metre by Fugro.

This point cloud data was then preprocessed by the Forest Research - Forestry Commission UK using Lidar360 to perform tree segmentation on the point clouds [Jones et al., 2022] and extract the variables shown in table 3.3.

Variable	**Unit**	**Description**
X position	metres	Distance east from origin to the top of the tree[1]
Y Position	metres	Distance north from origin to the top of the tree[1]
Height	metres	Distance from ground level to top of tree
Crown Diameter	metres	Average diameter of crown
Crown Diameter North to South	metres	Distance from north most point of crown to south most
Crown Diameter East to West	metres	Distance from east most point of crown to west most
Crown Area	square metres	Area of crown
Crown Volume	cubic metres	Volume within boundary of crown

Table 3.1: List of variables derived via tree segmentation from point cloud data.

In addition to this, further data was supplied by the Forest Research - Forestry Commission. This data described the traits of 30 by 30 meter cells of forest over the same area as the LiDAR scans. These traits are shown in Table 3.2.

Variable	**Unit**	**Description**
X Position	metres	Distance east centre of cell is from origin
Y Position	metres	Distance north centre of cell is from origin
Canopy Coverage	fraction	Proportion of area covered by canopy
Leaf Area Index	m^2/m^2	The total estimated area of all leaves over area of cell
Top Height	metres	Height of tallest tree in cell
Age	years	Years since the trees in cell were planted
Yield Class	m^3/ha/year	Estimated volume of growth per year over the area
Mean Diameter at Breast Height	metres	Average trunk diameter at breast height for cell
Basal Area	square metres	Total area of all trunks at breast height for area
Volume per Hectare	m^3/ha	Total volume of trees in cell

Table 3.2: List of variables describing 30 by 30 meter cells of forest in Aberfoyle Forest.

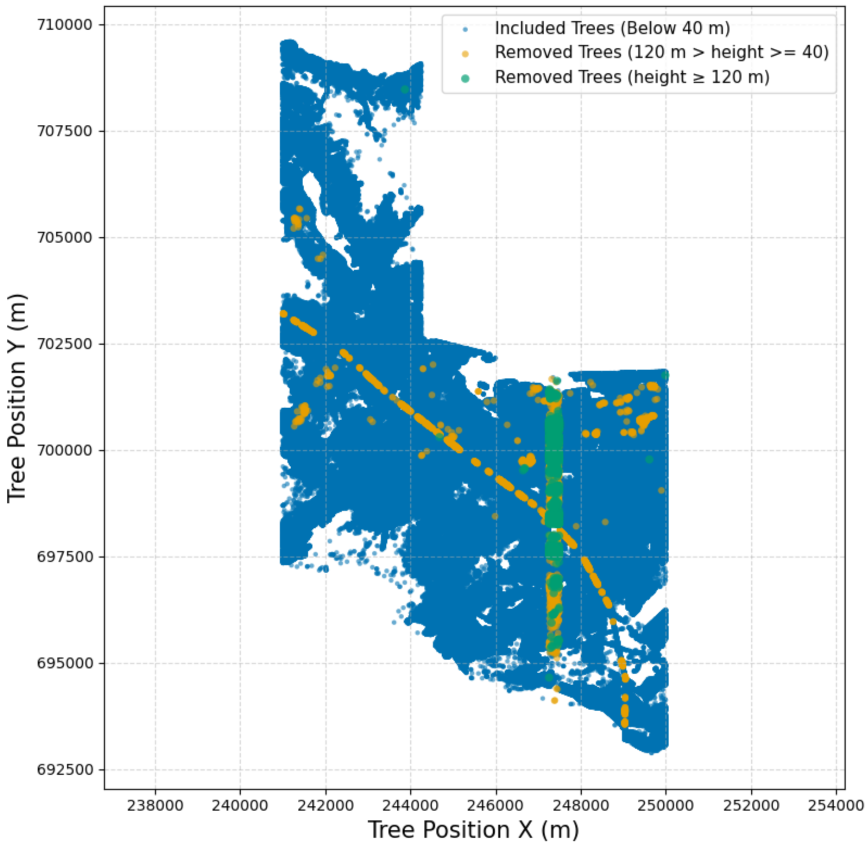

Figure 3.2: Spatial Distribution of Trees by Height Class

3.1.2 Data Pre-processing

Before use in training, the data was first explored and cleaned. Some discrepancies in the heights of trees were found, with some having listed heights near kilometres tall. This is likely an issue caused by how the height of a tree is determined. When calculating the height for a tre, the highest point in the segment of tree is selected. This is expressed as distance above sea level and it is then compared to the distance above sea level for the ground at that location. This means unusual heights are likely to be the result of either inaccurate topological data or by the LiDAR scans detecting non tree objects like power lines. This is evident in the way unexpected heights appear in clusters as shown in figure 3.2. To stop inaccurate data disrupting the training processes all trees with a height above 40 metres were removed. The height was chosen because the most common and highly prioritised tree for the Forest Research - Forestry Commission within Aberfoyle forest is scots pine which tends to grow to a maximum of 36 metres [Forestry and Land Scotland, 2024]. The cut off for valid trees was raised to 40 metres to allow for abnormally tall trees. Similarly trees with height less than 4 metres were considered bushes or other objects and were discounted. Removing trees with a height less than 4 metres led to an 8% decrease in data points going from 2,126,536 data points to 1,956,451. Removing trees with a height more than 40 metres caused a further decrease of 0.8% bringing the total to 1,945,823.

[1] As scans were performed over a region less than 100 km the curvature of the earth was not considered

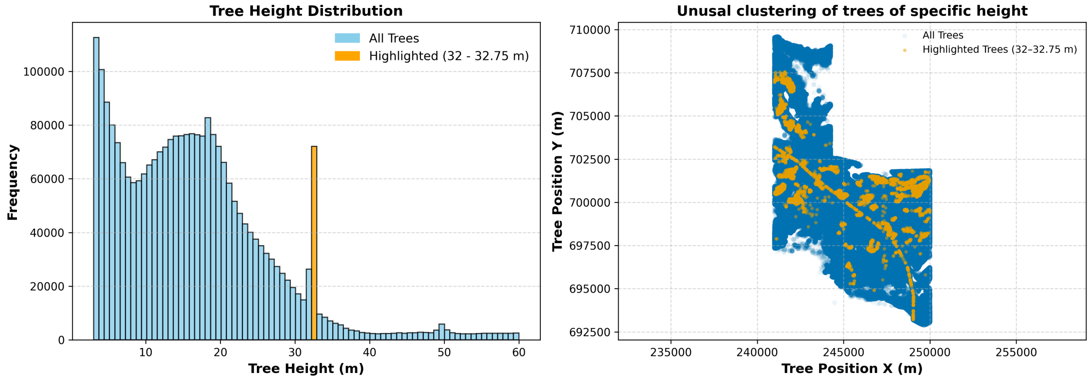

Figure 3.3: Distribution of Height Values by Frequency(left) and Location(right)

The crown diameter can also be used to identify misclassified trees. Data points with a crown diameter less than 0.5 metres were removed as they are unlikely to be trees. In addition, data points with a crown diameter of more than 10 metres were removed as these are likely to be multiple trees that are misclassified as one or a different object altogether being registered as a tree. This maximum crown diameter lead to a reduction of 0.9%. The minimum threshold caused a reduction of 0.1%.

It is shown that 40 metres is a fitting cut off for the data as expected as shown in 3.3(left) as the tail of the distribution tapers close to 0 at this value. However it is also shown that there is an an unusual up-tick in the frequency of heights between 32 and 32.75 metres. One may assume this is the result of the anomaly in the height data shown in 3.3(right) of an unnatural continuous line of heights at this value. These are likely the result of a power line being interpreted as the tip of a tree or another similar discrepancy. And so entries within this small range were discarded. This action removed a high proportion of data that was likely erroneous while only reducing the total dataset by 0.4%.

Next trees had to be matched across the snapshots so that it could be clearly stated how much a tree had grown in the 9 years span. This was done using the X Y coordinates to match the trees, however, this wouldn't be as simple a process as matching these values one to one. Not all trees appear in both datasets as the segmentation does not capture every tree scanned so in some cases a true match will not be possible. In addition to this, there is some noise in the measurements of the coordinates due to error in the LiDAR scans and how the drone is set up relative to the origin, as well as the fact that the coordinate is measured from the top off the tree which can sway in the wind. To account for this matching trees across snapshots would use the closest tree within a threshold of 1.5 metres. This action had the largest impact on the data but was vital to the method and caused a reduction of 87.2%.

Any set of multiples trees from 2012 that referenced the same tree as their future counterpart had all but one removed. This was done arbitrarily keeping the tree with the lowest index. Filters that could be performed on unmatched trees were applied before this to reduce the competition and therefore number of trees removed in this stage. This caused a reduction of 0.2%.

Finally the differences in height and crown area were calculated for each tree across the snapshots. for any trees that was matched to their future counterpart if the future tree was shorter than the past one by 0.5 metres or more they were deemed mismatches as trees are known not to shrink and removed from the dataset. This caused a reduction of 27.7%.

Filter Applied	% Reduction	Data Points Remaining
Trees shorter than 4 m	8%	1,956,451
Trees taller than 40 m	0.5%	1,945,823
Crown diameter more than 10 m	0.9%	1,928,261
Crown diameter less than 0.5 m	0.1%	1,926,761
Crown diameter between 32 and 32.75	0.4%	1,919,838
No match found	87.2%	246,532
Duplicate matches	0.2%	246,004
Tress that had a recorded reduction in size	27.7%	177,928

Table 3.3: Filters applied to datasets and their affect on total number of data points

3.2 Prediction Neural Network

3.2.1 Vectorizing Neighbouring Trees

In natural regrowth forestry one of the largest differences between it and the more predictable plantation planting is the arrangement of trees relative to the subject tree. Because of this, an important input for the predictive model is information on surrounding trees. Two ways of doing this were proposed and compared. Both methods can use any trees from the cleaned 2012 set to inform decision making as only the subject tree must have a match found in the 2021 dataset.

Neighbour Statistics

This method collated all data from trees within a certain radius into the following statistics:

number of neighbouring trees; neighbour mean height; neighbour minimum height; neighbour max height; neighbour mean crown area; neighbour minimum crown area; neighbour mean distance; and neighbour mean angle.

Neighbour mean distance and neighbour mean angle refer to the average location of surrounding trees relative to the subject tree as a polar coordinate.

This method was chosen because it abstracts the high volume of information on surrounding trees into relevant details. Most importantly, this data is always the same shape, 8 variables, so it is suitable for training. When creating models to be evaluated radii of 5, 10 and 20 were separately used to create training sets.

K Nearest Neighbours

The second method collected the exact parametres of the closest K trees within a certain distance. These parametres were:

distance from subject tree; angle from subject tree; height; crown diameter; crown area; crown volume. These values are then concatenated together to form one large vector representing the neighbouring trees.

This method does not abstract any information from the nearby trees and presents the full amount of detail on the trees it collects from and as there is an upper bound on the amount of trees in the vector any without the full number of neighbours can be padded so all entries have the same shape of $6 \cdot K$. Varying K values of 3, 5 and 15 with a distance upper bound of 10, 15, and 30 metres respectively were used to create separate training sets to compare models on.

Inputs were normalized to be used in training as this prevented variables with larger scales from over powering the affects of other inputs during back propagation [Shao et al., 2020], values with defined bounds like the X Y coordinates, angle and canopy coverage used min max normalizing to have values between -1 and 1. whereas values without defined terms were normalized using standardized scaling also known as z-score normalizing.

3.2.2 Model Evaluation Metrics

Some models will directly predict the values for a tree's variable after the time span, some will predict the estimated change in these values and others will predict the normalized values for each of these. In training different evaluation metrics are used depending on what the model is predicting.

So these models could be compared, a unified evaluation method was used to measure the effectiveness of a model. This would be the mean relative error for predictions across the test set, as shown in Equation 3.1 with $\hat{y}_i$ representing the true value and y_i the prediction

$$\text{Mean Relative Error} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{|y_i|} \tag{3.1}$$

Relative error was chosen as it can be applied to all methods. Any methods that predict the difference in a variable create full predictions by adding the predicted difference to the original value. Any that created normalized outputs denormalize them before they are used to calculate predictions. Another reason relative error was selected is as it is grounded in reality and easier to interpret compared to the mean square error on a normalized variable. This metric is also well suited to our problem as it captures the fact that an error of 1 meter means far less when the true value is a 30 meter tall tree compared to a 5 metre tree. MSE does not capture this. The primary scenario in which relative error fails is when true values are close to zero [2] however as the predicted values of height and crown area have filters discounting small values from the dataset this issue is avoided.

[2]Epsilon is often added to the denominator of this function to avoid error when dividing by 0, however there is still an explosion in loss near 0

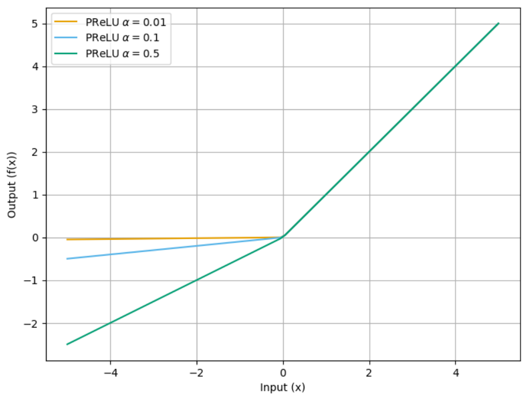

Figure 3.4: PReLU activation with varying alpha

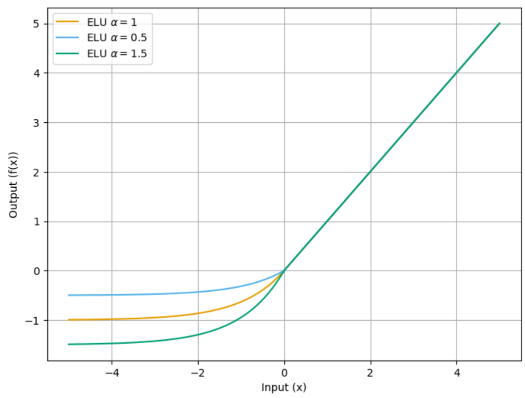

Figure 3.5: ELU activation with varying alpha

3.2.3 Model architectures

Four methods for predicting the traits of a tree after the 9 year span were evaluated and compared. The first outright predicts the target values of height and crown area. The next predicts these values but normalized. Were this method to be used to calculate real values the mean and standard deviation would need to be stored to denormalize these outputs. The other methods are trained to predict the difference in these values over the span, these are then added to the original values to create the full prediction. Two methods work this way, one directly estimating the difference, the other the normalized difference. The normalized method is subject to the same caveats as the other normalized method.

The models that directly predict the values for a tree will use relative error for training as this matches the evaluation method, however, the transformation applied when standardizing variables already accounts for differing scales. It also means that some of the entries are close to 0. For this reason relative error is not applicable for training therefore mean square error, mean absolute error and Huber loss will be tested with these methods. Similarly, the models that estimate difference will have some true values close to 0 and so relative error will not be applicable to this method either meaning the three alternatives loss functions will be tested instead.

Varying activation functions were used and compared when creating models. The functions chosen where:

- **PReLU** (parametric rectified linear unit)3.4 was chosen as it forms a superset of the commonly used and applicable ReLU, rectified linear unit. ReLU suffers from the ”dying ReLU” [Ding et al., 2018] problem as all negative values are set to zero whereas PReLU learns a small slope for the negative range of inputs and the unmodified values for positive inputs. This solves the ”dying Relu” problem while letting the model learn more complexity. An issue with this activation method is that the increase in trainable parametres leads to slower training. This was not an issue as training still was able to finish within 2 hours.
- **ELU** (Exponential Linear Unit) 3.5 uses an exponetial component for negative

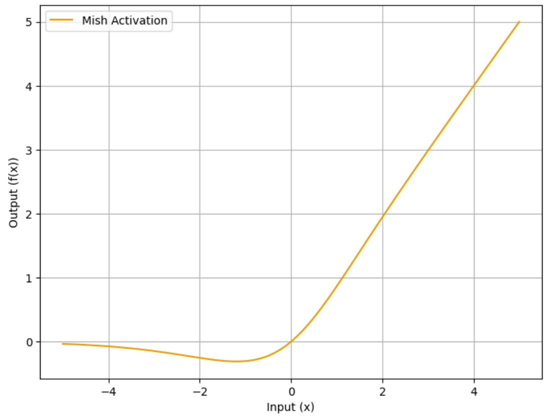

Figure 3.6: Mish activation

inputs, and the unmodified value for positive inputs. It is defined as $\alpha(e^x - 1)$ for $x < 0$ and x for $x \geq 0$. This alpha is also a trainable parameter. This introduces nonlinearity to the model's activation allowing it to capture different patterns. This function also brings mean activation closer to zero, which speeds up learning [Ding et al., 2018].

- **Mish** 3.6 is a smooth non-monotonic function defined as $x \cdot \sigma(x)$ where $\sigma(x)$ is the tanH function. This function also introduces non-linearity. The primary difference this activation function introduces is that it is non-monotonic meaning it is not either strictly increasing or decreasing. The mish function was added because it has portions with a negative and positive gradients which allows it to capture different patterns compared to the other activation functions [Ding et al., 2018].

A learning rate of 1e-6 was found to be reasonable as an initial value for the learning rate scheduler. The learning rate scheduler multiplies the learning rate by a factor of 0.3 when a plateau in validation loss is reached. This allows the model to quickly converge onto an effective model before using finer tuning to optimize it. In addition to this, a early stopping function was used to prevent any unnecessarily long training times. This function was given ample patience to allow for multiple learning rate reductions before stopping learning.

Differing structures of architecture were used to create models so their effectiveness in capturing the patterns within the dataset as well as preventing over-fitting could be compared. The 3 following architectures were used.

- **Block:** This is the simplest architecture consisting of 5 dense layers all with the same width of 512 neurons. Dropout layers with a rate of 0.01 were placed after the first two layers to reduce over-fitting while having little effect on the model's ability to make predictions. This was chosen as its simple structure makes it a good baseline.

- **Funnel:** This architecture consists of multiple dense layers starting with a width of 512 neurons and decreasing gradually to 64 throughout the model. This reduction means the model must extract important features it can generalize to make predictions. The broader start to the model allows it to still be able to capture the relationship between input variables. Dropout layers with rate 0.01 were also applied after the first two dense layers to reduce over fitting further. This was chosen as the shape encourages feature extraction making over fitting more difficult

- **Hourglass:** This architecture consists of multiple dense layers similar to the funnel architecture it has an initial width of 512 decreasing to 128 however this then expands out back to 512 after this. The larger initial width allows the model to capture the effects of relationships between models, the shrinking intermediary layers force the model to extract only the important features and the increasing layers after allow it to rebuild how these features interact for its prediction. This was chosen as it is more complicated to contrast the other structures in how it extracts features and then rebuilds relationships.

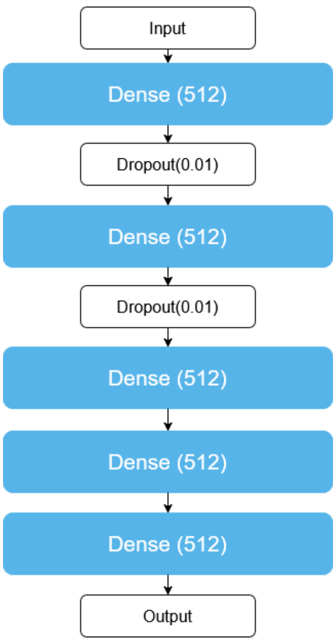

Figure 3.7: Block Architecture

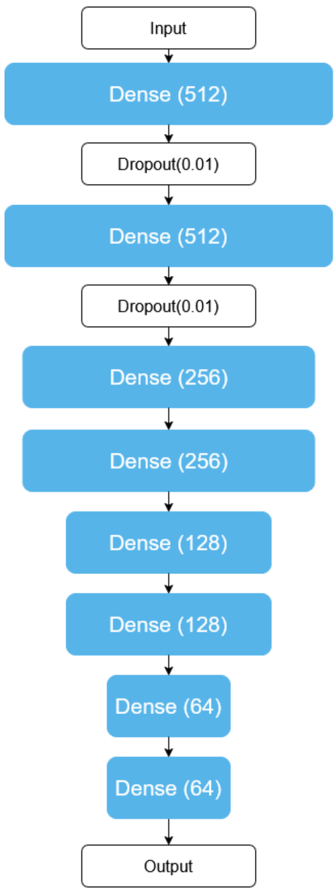

Figure 3.8: Funnel Architecture

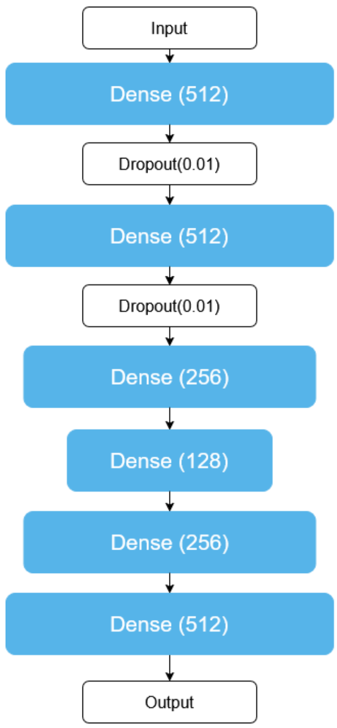

Figure 3.9: Hourglass Architecture

Chapter 4

Results

4.1 Effectiveness of Varying Architectures

Prediction Target	Neighbour Statistics		KNN Representation	
	Future Value	Norm Diff in Value	Future Value	Norm Diff in Value
Block PReLU	H.Err 10.27%	H.Err 10.64%	H.Err 10.89%	H.Err 11.08%
	C.Err 52.09%	C.Err 73.57%	C.Err 52.54%	C.Err 74.55%
Funnel PReLU	H.Err 10.25%	H.Err 10.40%	H.Err 14.53%	H.Err 10.82%
	C.Err 51.76%	C.Err 76.07%	C.Err 52.76%	C.Err 79.24%
Hourglass PReLU	H.Err 10.27%	H.Err 10.46%	H.Err 11.07%	H.Err 10.99%
	C.Err 52.33%	C.Err 73.99%	C.Err 52.66%	C.Err 75.62%
Block ELU	H.Err 10.37%	H.Err 10.78%	H.Err 11.01%	H.Err 11.03%
	C.Err 52.05%	C.Err 73.37%	C.Err 52.51%	C.Err 73.56%
Funnel ELU	H.Err 10.27%	H.Err 10.46%	H.Err 10.79%	H.Err 10.81%
	C.Err 52.73%	C.Err 73.18%	C.Err 52.15%	C.Err 74.32%
Hourglass ELU	H.Err 10.27%	H.Err 10.65%	H.Err 10.81%	H.Err 10.91%
	C.Err 52.33%	C.Err 75.13%	C.Err 51.74%	C.Err 74.15%
Block Mish	H.Err 10.31%	H.Err 10.77%	H.Err 10.91%	H.Err 11.08%
	C.Err 51.93%	C.Err 74.00%	C.Err 52.76%	C.Err 75.91%
Funnel Mish	H.Err 10.38%	H.Err 10.61%	H.Err 10.77%	H.Err 10.87%
	C.Err 51.83%	C.Err 73.79%	C.Err 52.35%	C.Err 74.88%
Hourglass Mish	H.Err 10.29%	H.Err 10.79%	H.Err 10.75%	H.Err 11.02%
	C.Err 52.38%	C.Err 75.12%	C.Err 53.09%	C.Err 74.43%

Table 4.1: Model relative across architectures and activation functions when predicting Height (H.Err) or Crown Area (C.Err) of the test set

Table 4.1 describes the effectiveness of varying techniques on predicting the height or crown area of a tree measured in relative error. Each row shows results for the range of possible outputs used for training described in the methods section across the range of architectures and activation functions described. These combinations were run exhaustively so they could all be compared. Two representation were used: neighbour statistics (using a radius of 10 metres) and the KNN method (using a K value of 5). The methods that directly predict the future value used relative error in training, the ones using the normalized difference in value use mean square error in training. Both however were evaluated using relative error after predictions had been unnormalized and added to initial values if needed. Within each cell the result for mean relative error on height is preceded by H.Err and the mean relative error on crown area is preceded by C.Err. The lowest errors for each prediction method of H.Err and C.Err have been highlighted with an underline.

4.2 Analysis of Results and Techniques

4.2.1 Neighbouring Statistics

As can be seen in the table of results 4.1, when using the neighbour statistics vector as additional input directly predicting the future value consistently outperforms predicting the difference and adding that to the initial value. This approach is almost equally effective when predicting height but it is far worse when used for crown area. When using neighbour statistics to predict future value of height the relative error of models ranged from 10.25% to 10.38% with the best architecture for height being the funnel using PReLU activations. This architecture also performed best when predicting crown area, however, all predictions had a high error ranging from 51.76% to 52.38%.

Direct Prediction:

Within each activation function either funnel or hourglass had the lowest error among architectures with the block architecture underperforming when directly predicting height. However when directly predicting crown area, the best performing architecture was unique to each activation function. Among the same architectures PReLU has the best error levels when predicting height having either the lowest relative error or tied lowest. When grouping by architecture, models with the lowest error predicting crown area for their structure were evenly spread among activation functions.

Difference Prediction:

Similar patterns appear when viewing how height prediction performs across different architectures with the same activation function. With the funnel architecture most commonly having the lowest error, the hourglass structure consistently underperformed in these groups. This pattern was repeated when predicting crown area with hourglass consistently being the architecture with the highest error among the three. Within models with the same architecture and differing activation function PRelU resulted in the lowest error when predicting height for all architectures. The suitability of activation function is less clear cut when predicting crown area with PReLU and ELU contributing to the best error values for some architectures however Mish underperforms.

4.2.2 K Nearest Neighbours Representation

Similar to the neighbour statistics representation when using KNN representation directly predicting values resulted in lower errors in all architectures with estimating height having slightly higher error and estimating crown area having far higher error. When using KNN representation to directly predict values of height the lowest relative error was achieved by the hourglass architecture using Mish activation with an average of 10.75%. The most effective architecture using KNN to predict crown area was the hourglass architecture using ELU activations which resulted in a mean relative error of 51.17% making it the best performing model of all run for crown area however it is only lower than funnel architecture using PReLU and neighbour statistics by a very small margin of 0.02%

Direct Prediction:

Within each activation function the best performing architecture was evenly spread with none being more common than the other when predicting height. The same was true differing architectures of the same activation function, with each architecture being the best for one of the functions each with no one architecture standing out as better overall. Comparing best performing activation functions among architectures we see that when predicting height Mish is most commonly the best performing only bettered by PReLU in the block architecture with ELU performing worse of the three. Conversely when predicting crown area ELU performs best only bettered by PReLU in the block architecture with Mish getting the highest error of the three.

Difference Prediction:

When indirectly predicting values via difference the best performing activation within each architecture was ELU. Comparing any models of the same architecture the one using ELU will always have the least error in either predicting height or crown area. Among models with the same activation and differing architectures, the funnel architecture was stand out when predicting height having the best performance of the 3 architectures for every activation function. However when predicting crown area the funnel architecture did not perform best for any activation function. The block architecture had the lowest relative error in two of the activations with hourglass being best for the other.

4.2.3 Comparison of Neighbouring Tree Representations

Table 4.1 shows that models that use the neighbour statistics method of representation outperform those using KNN in most scenarios. The core difference between these two representations is the lower shape of neighbour statistics that is achieved by abstracting the data into important features like mean height, mean crown area, etc. In contrast, the KNN representation shows the data in full detail with a larger shape. The lower overall error of models using neighbour statistics shows that the abstraction it provides is helpful in learning. Evidence of this can be seen in figure 4.1 (graph KNN Relative Error Method - Height) which shows a protrusion of high error points following a roughly logarithmic line leaving the main distribution. It is likely an artifact either of erroneous data that the neighbour statistics method was able to abstract away to help the model generalize or the artifact could represent a relationship that is harder to learn for the model when the data is presented in a more raw format. The impact of the representation is more evident when viewing the distribution of relative error vs true

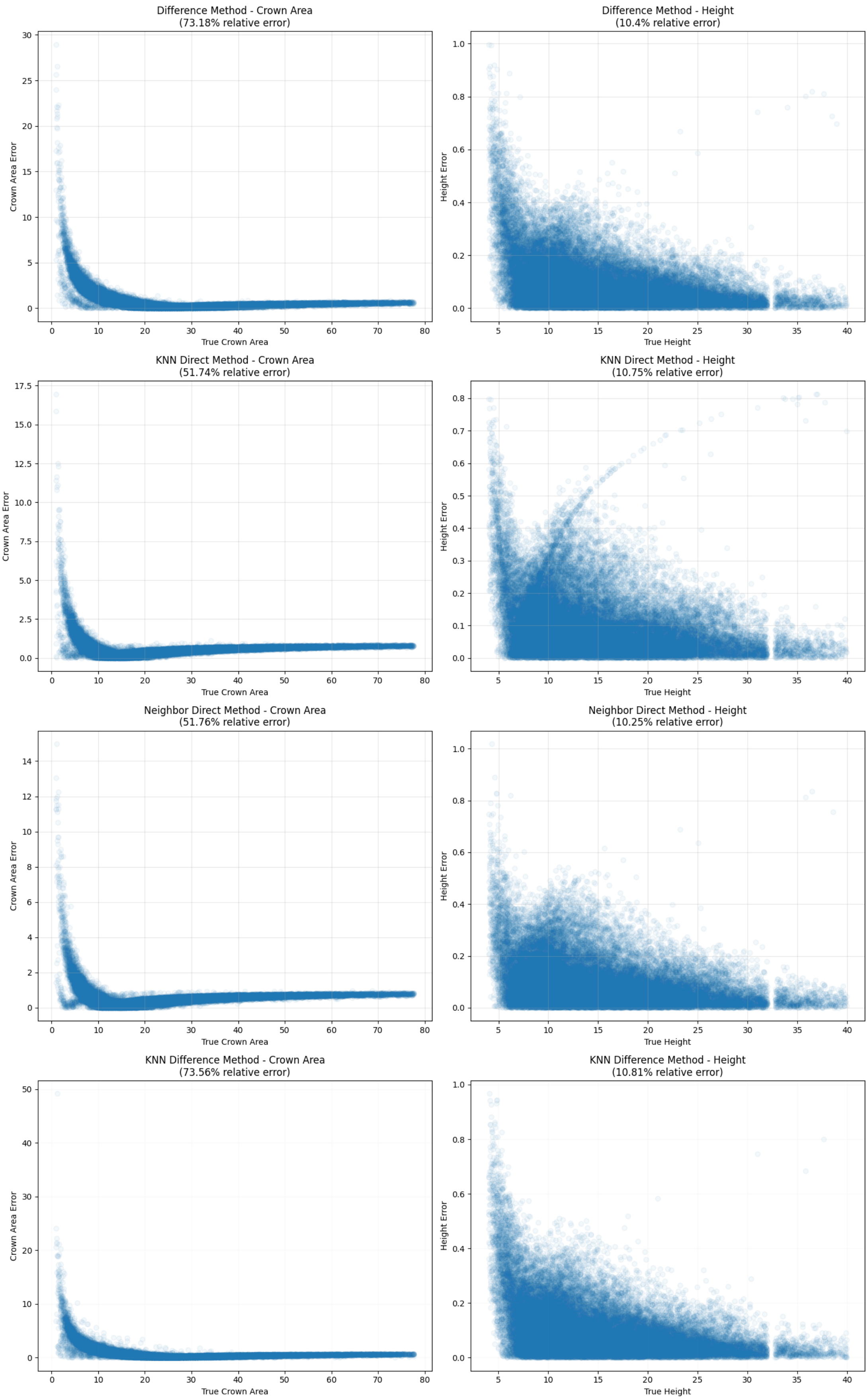

Figure 4.1: True Values vs. Relative Errors for Best Methods in Each Category

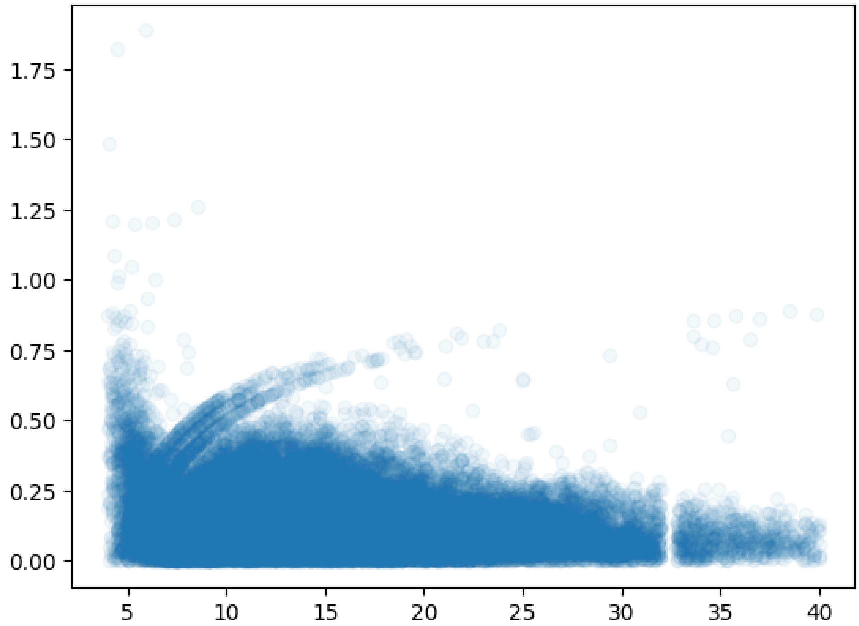

Figure 4.2: Artifacts in Funnel PReLU for KNN future Height Plot of True Value vs Relative Error of Prediction

height for the models using a PReLU funnel to directly predict height. When using neighbour statistics this is the model with the lowest error of all attempted however when performing the same prediction but training on the KNN representation it results in an error of 14.53% the worst of all models attempting to predict height. In figure 4.2 showing the relative error of different true values for this method it is seen that the issues caused by this artifact have an even stronger affect especially compared to 4.1 Neighbour Relative Error Method - Height where this has little affect and the change in representation leads to a large change of 4.28%. KNN Difference Method - Height also shows that predicting the difference eliminates these artifacts which is likely to improve the model however the fact this method performs worse overall mitigates any benefits of this.

4.2.4 Analysis of Models

Trends in Relative Error vs True Value:

It can be seen from figure 4.1 that all models struggle to make accurate predictions for smaller trees. This fits in with the general wisdom that smaller trees are more sensitive to their environment with external factors having the potential to greatly affect growth when a larger tree would have the backup resources to be less affected [Coomes and Allen, 2007]. This effect is incredibly prominent when predicting crown area with most data points being below the mean particularly after $10m^2$. However, the spike in error at low values skews the error up to the mean value shown.

Consistency of Models

From figure 4.3 it is shown that each model has a similar distribution of error when predicting height. This shows that each model is roughly equally consistent in how accurate it is. The figure also shows the skewed nature of the data which is expected as the absolute difference is used to calculate relative error so it has a natural lower bound of zero whereas outliers have no upper bound. However in figure 4.4 it is shown that there are differences in the consistency of the models when predicting crown area. Both models using direct prediction have smaller standard deviations and closer grouping around the mean. This shows that predictions made by these models are more likely to have an error near the mean error, or less, as opposed to the models predicting the difference in values which have far larger standard deviations and less grouping. This shows that, as well as these models having worse average error, they are also more likely to return predictions with relative error far higher than their mean.

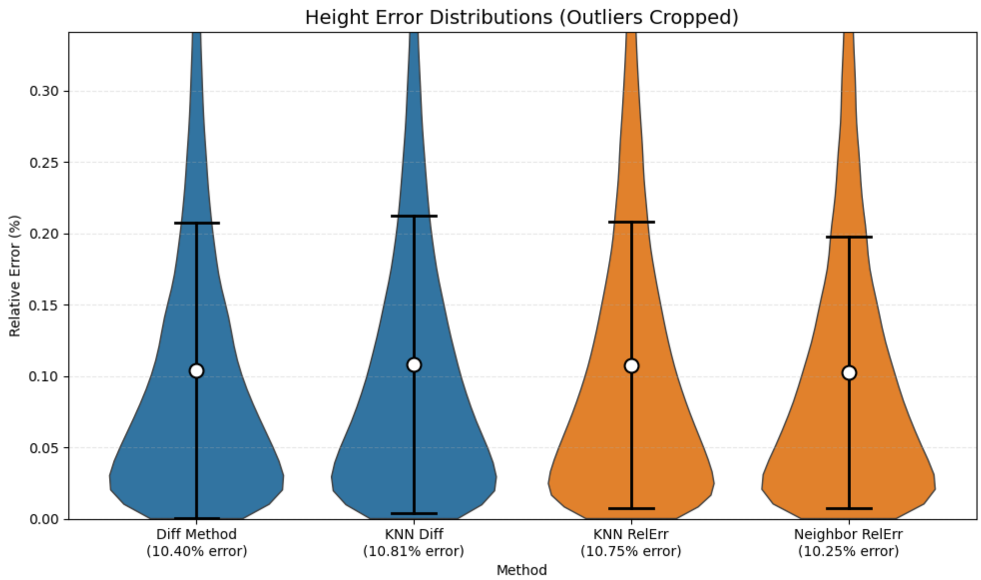

Figure 4.3

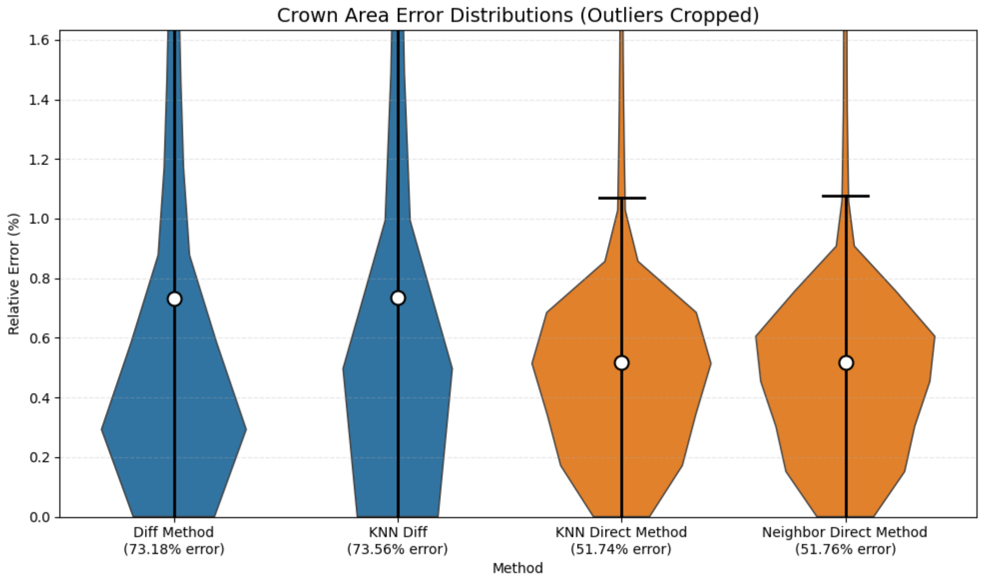
Crown Area Error Distributions (Outliers Cropped)
Relative Error (%)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Diff Method
(73.18% error)
KNN Diff
(73.56% error)
KNN Direct Method
(51.74% error)
Neighbor Direct Method
(51.76% error)
Method

Figure 4.4

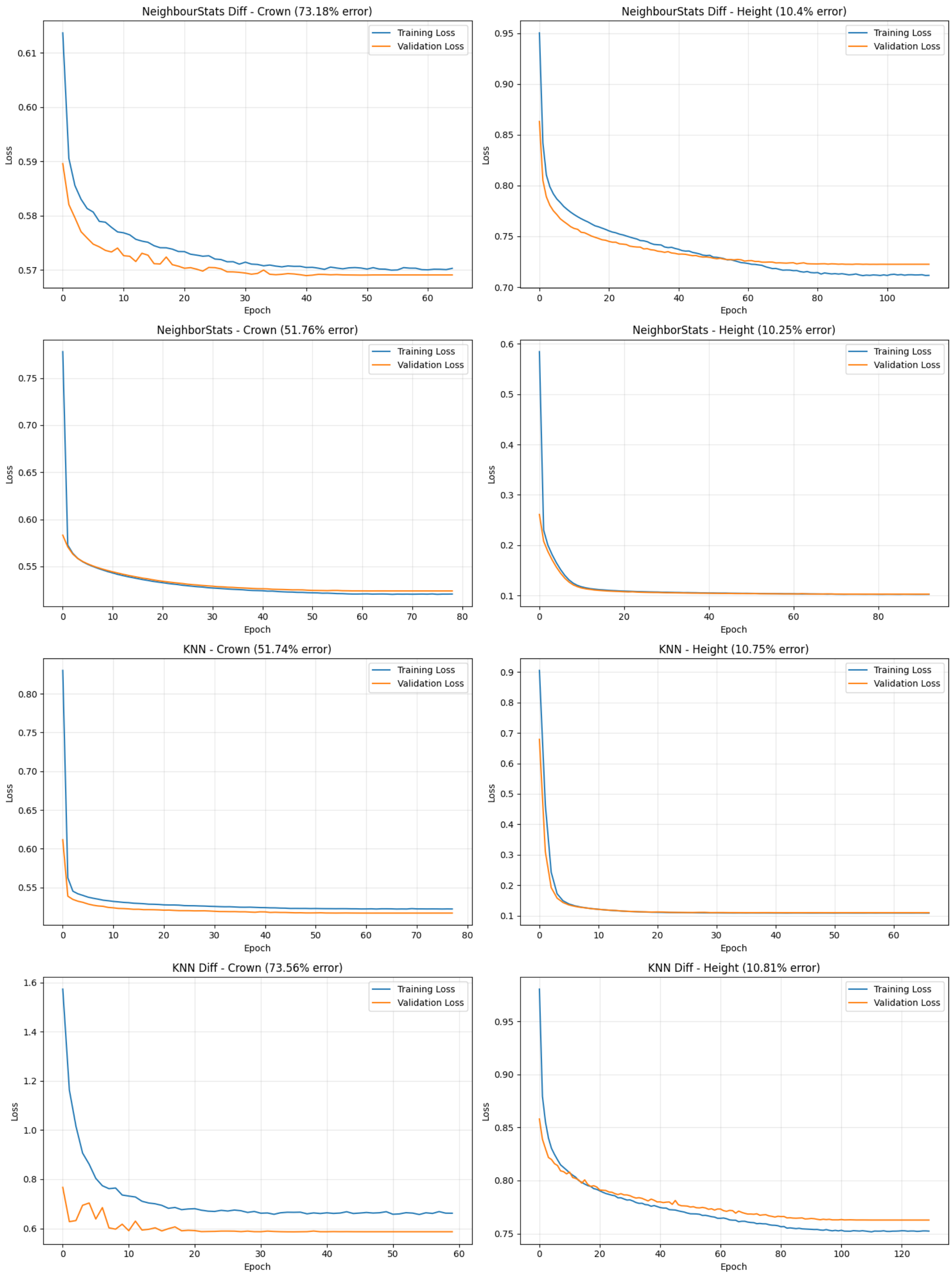

Figure 4.5: Loss Over Training for Best Models

Training and Validation Loss:

Figure 4.5 shows the changing loss across epochs during training. No models suffer severely from over fitting. If a model was over fitting it would be shown as the validation loss being far higher than the training loss possibly even rising while training loss falls. Of all the models, ones predicting height via the prediction method had some over fitting but only to a minor extent as validation is close to training loss and follows the same trends. From figure 4.5 it is shown that models directly predicting values of height or crown area quickly converge to a loss near their lowest found. After this there is little decrease and the learning rate scheduler cycles through lower learning rates to performer finer tuning but a better solution is not found. This means that for these models the loss forms quite a sharp corner, which would normally be a sign the model had too high a learning rate or was over fitting. However, as validation loss still follows with training loss after the corner and the learning rate scheduler reduced the learning rate multiple times this wasn't a concern.

Viewing the models that predict the difference in value, their training has more conventional loss values across epochs with the model slowly converging on a solution gradually decelerating. When validation loss is viewed, it is often near or below the values for training loss but when for models predicting difference in crown area its changes are quite erratic with values going up and down instead of staying in one direction across epochs. This is often a sign of over fitting or too small a batch size however it is solved as the model is trained more and becomes better at generalizing.

Spatial Distribution of Relative Error

The spatial distribution of relative error per model is shown in figure 4.6 it shows that the error when predicting crown area is equally distributed with noise that doesn't correspond to X or Y around the area of the forest scanned meaning that the models are equally effective in any region of the forest. The models that used difference over the span to predict crown area shows a rougher texture with sporadic spikes in error of higher intensity, this is a sign that these models experience higher variance in their error and is backed up by the values in fig 4.4 which explicitly show that these models have higher variance in error. Note that these plots use different scales in their gradient, this was done so the trends spatial could be viewed and compared despite the fact that some have very different ranges and averages.

Viewing the the spatial distribution of error in models that predict height in fig 4.6 it is shown that they are distinctive regions where relative error is higher or lower meaning the model performs better in some areas compared to other. These areas are the same for each of the models.

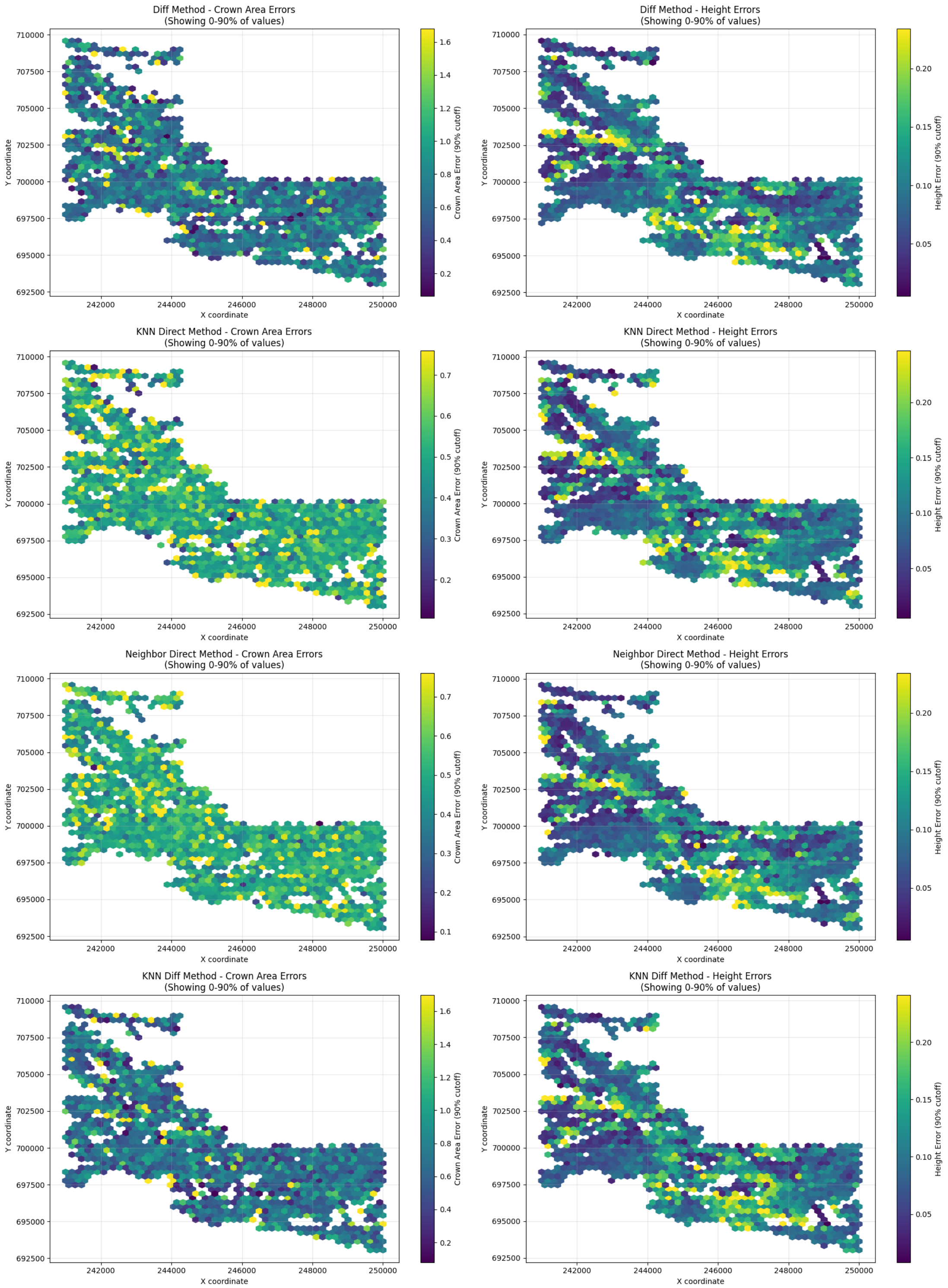

Figure 4.6: Spatial Distribution of Relative Error Across Models

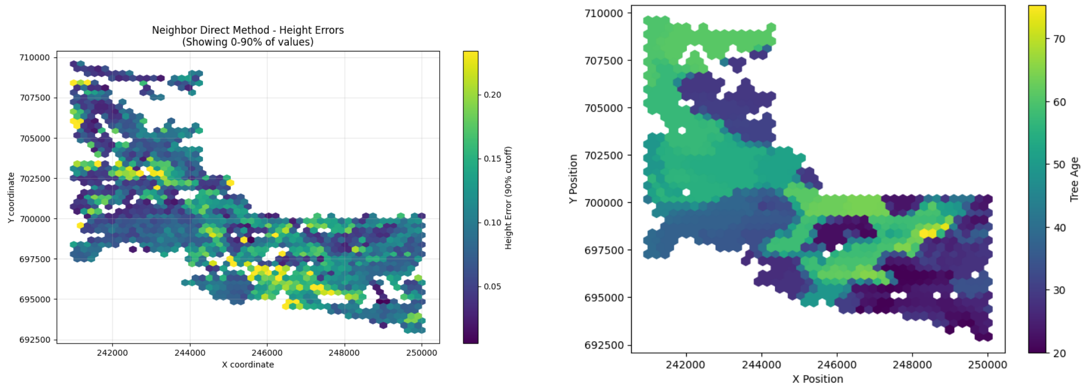

Figure 4.7: Spatial distribution of relative error(left) compared to age(right)

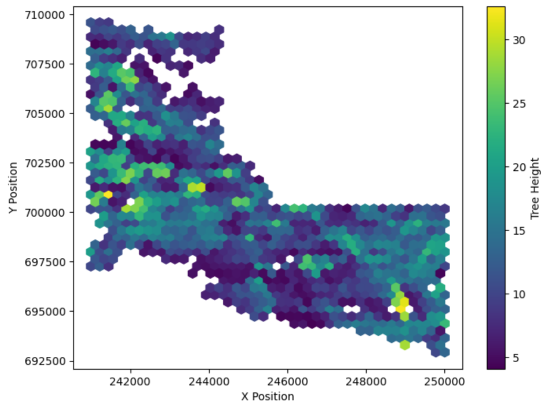

Figure 4.8: Spatial distribution of tree height

Comparing the distribution of error to the distribution of certain tree traits, more can be learned about these regions of higher and lower error. Viewing fig 4.8, it shows that despite the fact that when viewing one to one relative error vs true value, lower tree heights result in higher error. Areas with shorter trees do not correspond to areas with higher error. Comparing this pattern of errors to the distribution of a stand of trees' age in fig 4.7 shows similarities in the pattern. This shows that the model performed better in the regions with younger trees.

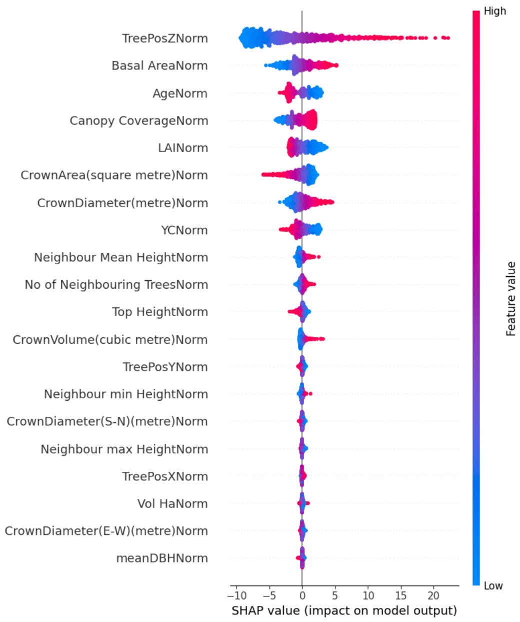

Figure 4.9: Feature Importance for best performing model directly predicting height by neighbour stats

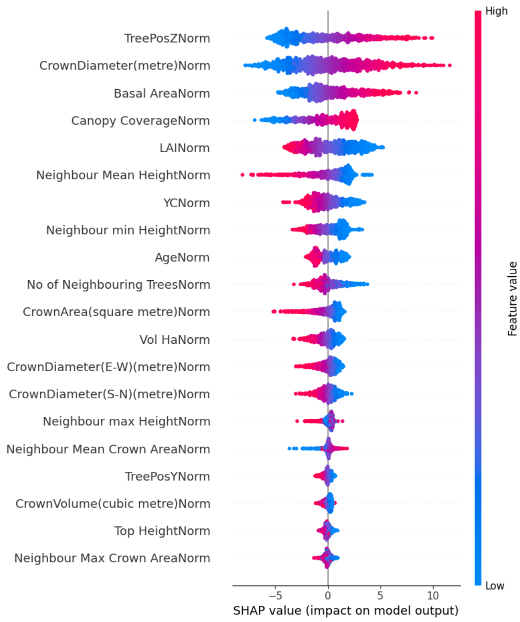

Figure 4.10: Feature Importance for best performing model directly predicting crown area by neighbour stats

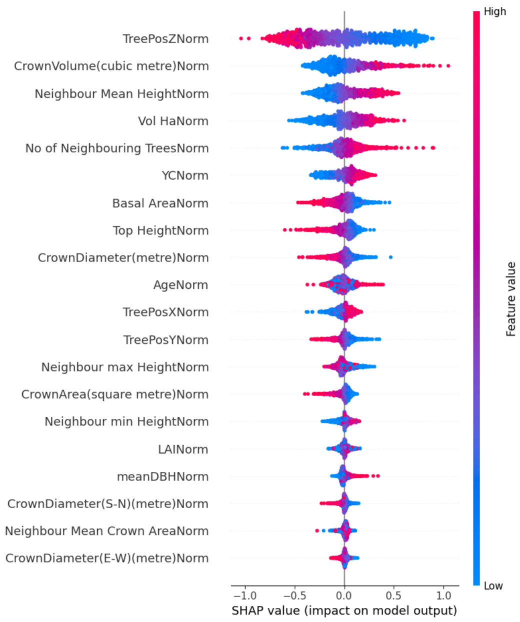

Figure 4.11: Feature Importance for best performing model predicting difference in height by neighbour stats

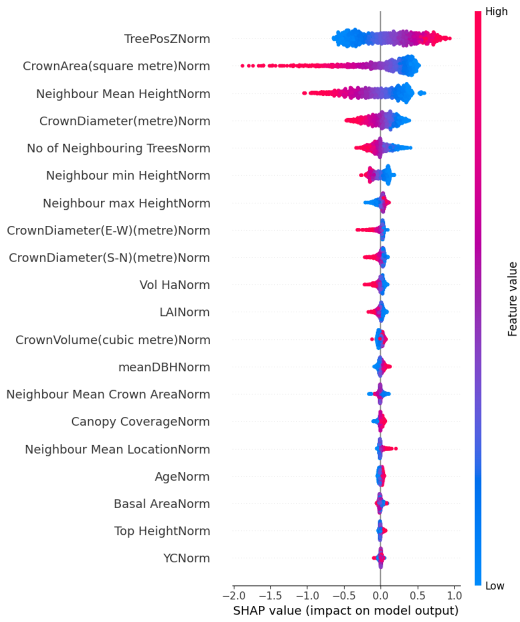

Figure 4.12: Feature Importance for best performing model predicting difference in crown area by neighbour stats

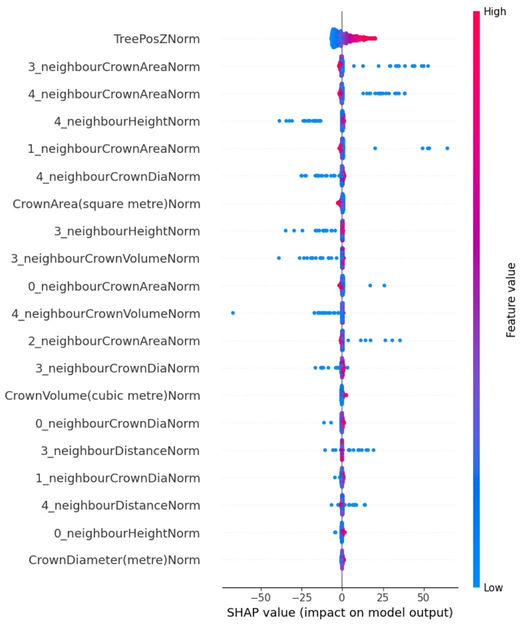

Figure 4.13: Feature Importance for best performing model directly predicting height by KNN

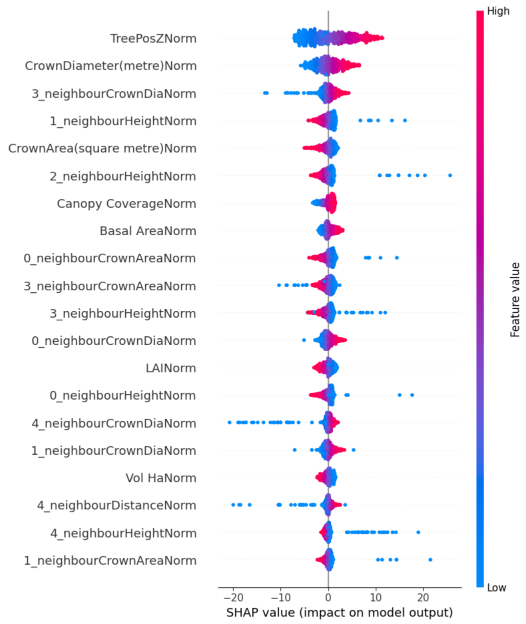

Figure 4.14: Feature Importance for best performing model directly predicting crown area by KNN

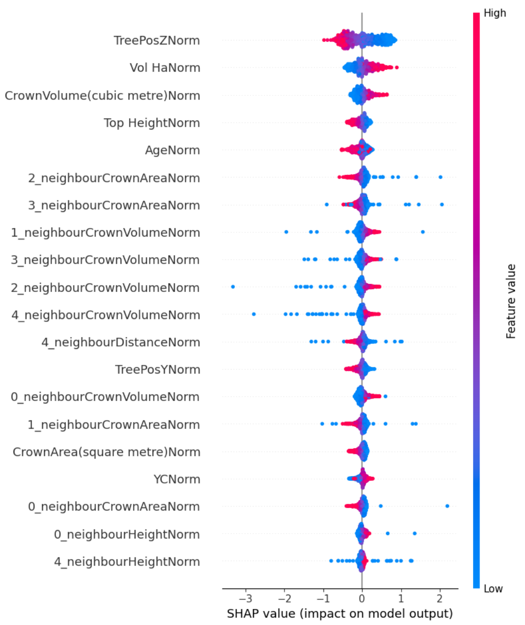

Figure 4.15: Feature Importance for best performing model predicting difference in height by KNN

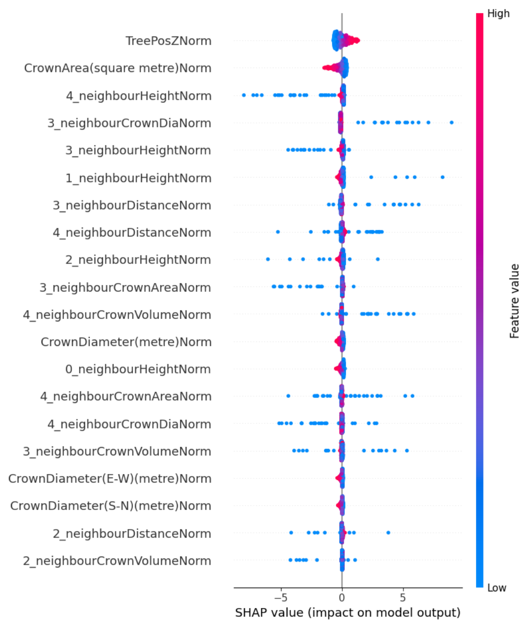

Figure 4.16: Feature Importance for best performing model predicting difference in crown area by KNN

Feature importance in best performing Models:

Figures 4.9 to 4.16 show the 20 inputs with the highest impact on the models predictions calculated by SHAP values using permutated inputs to models [Marcílio and Eler, 2020]. These plots show that the initial height of a tree is very important to the models outcome. This is expected for models predicting future height but is also the most impactful feature for those predicting future crown area.

For models predicting future crown area the second most important feature was either crown diameter fig 4.10 and 4.14 or area fig 4.12 and 4.16 with models using the direct prediction method favouring diameter and indirect favouring area. The prominence of one representation of the trees' crown having a very high impact with the other having much lower in all models and the preferred representation changing across models shows that during training the models work off one representation and manage to infer the information it needs from that one only. This shows that only one of these representations is needed and one could be removed to reduce the shape of the input without negatively affecting new models.

Among models using neighbour statistics fig 4.9, 4.10 4.11, 4.12 the most valuable feature taken from the neighbour statistics is neighbour mean height followed by number of neighbours in all but one case were minimum neighbouring tree height sits between the two fig 4.10.

For models using KNN representation, figures 4.13, 4.14, 4.15 and 4.16, the index of the neighbour did not have a noticeable affect on how impactful a feature was, with all of the five neighbours used having very similar impact on the prediction with some being place higher as a result of noise. This is seen in how sparse the distribution of points is for features referring to neighbours. The most impactful neighbour statistic for models using the KNN representation was height and crown diameter or area with distance from subject tree appearing less often.

Chapter 5

Discussion

5.1 Evaluation

This study proposed and implemented models using varying architecture, input representation, activation functions and methods of prediction so that the suitability of these could be compared in the context of existing within a digital twin. Specifically, the volume and nature of data for the input and the needs of the output were considered. 36 models for predicting future height and crown area were created and trained on the dataset converging on a solution with reduced loss without harmful over fitting.

The study identified challenges encountered when using data acquired from LiDAR scans of real world natural forests that have been passed through LiDAR 360's tree segmentation software and what techniques can be used to identify and clean erroneous data.

There were many challenges in matching trees from the 2012 dataset to the same tree in the 2021 dataset due to two factors. Not all trees in the scanned area were recognized by the tree segmentation algorithm, many were missed entirely or two trees were counted as one. Because of this, not all trees would have a match in the corresponding dataset even if they were in the area of both scans. The other issue was that noise in the coordinates of trees meant that coordinates could not be used as a unique identifier for trees. Some leniency was needed for matching because of this, but this would also introduce the opportunity for trees to be mismatched especially if their true match was missing from the future dataset. Not finding matches caused a large reduction in the amount of useable data bringing down the count by an order of magnitude giving the models less information to train on. Mismatching also caused many issues, for example, instances where a tree was mismatched to a shorter tree in the future were identified, removed and were not uncommon in the data set. Mismatches with that identifying feature were only a subset of the true amount however, some could not be identified in this way and so would exist in the dataset as noise that the model would have to compensate for.

Digital twins of forests struggle with the volume of data in the system. This study views solutions to this by comparing methods of making data digestible for training over a range of options, with some using the least amount of abstraction for the data to be of regular shape (K nearest neighbours) and the other using a high amount of abstraction (neighbour statistics).

In addition to this, the study implements architectures that differently address problems encountered by neural networks in how they capture complex relationships, extracting important features and avoiding over fitting. This was done by varying the composition of neurons in each layer, using activation functions with unique natures to one another and by using different methods of creating the prediction: directly predicting and predicting the difference.

The way in which these different methods affect the suitability of models was compared holistically by exhaustively implementing and training all combinations of models and then identifying trends in what techniques led to lower loss. In-depth analysis was also performed on the best performing models of each category that identified what issues certain architecture solved better than others and how this was the case. In this analysis the importance of features was quantified to show possible ways in which they could be streamlined and what features were integral to the models.

Despite the fact that the nature of models and inputs varied, many of them still had mean relative errors within a very narrow range. The information inferred from one model outperforming others was therefore less meaningful as it may be the result of noise.

Of the models predicting height the best performing achieved a relative error of 10.25%. This is a common and acceptable level of error when predicting natural systems [Skudnik and Jevšenak, 2022] [Biging and Dobbertin]. Evaluating this model identified factors highly relevant to predicting tree height. The best performing model on crown area however had a mean relative error of 51.74% making its predictions too unreliable to be useful, however the analysis on what factors affect performance sheds light on what trends may be true for a better performing model.

5.2 Future Work

More consistent recognition of trees in the LiDAR scan would aid greatly in the training process of the model as more trees would have matches found for them and there wouldn't be such a drastic loss in useable data due to matches not being found for trees. This would mean less mismatching as stricter rules could be applied. New functionality could also be achieved if matching is so reliable that a trees presence in the earlier dataset without it being in the later dataset is evidence that the tree has died in the physical twin. This would allow new models to predict tree mortality and wind through damage.

The input data relating to surrounding trees is likely to have different effects depending on other data on surrounding trees. A tall tree nearby is more relevant than one far away and a tree that would be in the shade of another is less affected if it is taller than the one casting the shadow. For these reasons, attention layers would be useful in helping the model recognize how the context changes the effects of certain inputs because of its dynamic weightings that change with context.

Were these techniques to be used in a fully realised digital twin, scans of the area would be taken at regular intervals. This could be used to create similar models that use temporal data from multiple scans to make predictions with higher accuracy.

5.3 Conclusion

This study demonstrated how digital twin techniques could be applied to the chaotic and data rich context of managing and predicting natural regrowth forest. Using LiDAR derived datasets and neural network models, this study explored the best ways of preparing this data and how the heterogeneous spatial arrangement of trees in natural regrowth forests, primarily their proximity and traits relative to one another can influence changes in height and crown area over a nine year timespan.

Two methods of encoding neighbouring tree information were used: one with a high level of abstraction being encoded as the statistics of other trees within a certain radius and the other preserving as much information raw information as possible using k-nearest-neighbours. These were used to train a wide range of neural networks of varying structure, activation function and prediction method. It was found that a funnel structure using PReLU activations with neighbour statistic lead to a relative error of 10.25%, which was within acceptable range for ecological models, performed the best. This finding underlined that the abstraction provided aided in reducing error and over-fitting in the digital twin model.

The best performing model for predicting crown area had less promising results with the lowest error of models made being outside of the acceptable range. However, analysis of both models shed light on what features were important to the problem and how the different representation affected this importance. This analysis also showed that error in predicting height was not evenly distributed across space with higher error instead correlating visually with regions of higher age. This finding showed areas that had been planted longer ago had more complex relationships and were harder to predict.

In the broader context of digital twins in forestry this project contributed an exploration of how recent developments in LiDAR drone technology could be used to acquire more granular, individual tree focused data and the effectiveness of varying techniques in handling and processing this data as opposed to the more common stand level models. This individual tree focused approach means results are more actionable as they pertain to specific trees and can also be used to create projections of what will happen in scenario were new trees were added or removed.

Bibliography

Digital twins and living models at NASA. URL https://ntrs.nasa.gov/citations/20210023699.

Muhammad Awais, Wei Li, Haoming Li, Muhammad Jehanzeb Masud Cheema, Saddam Hussain, and Chenchen Liu. Optimization of intelligent irrigation systems for smart farming using multi-spectral unmanned aerial vehicle and digital twins modeling. In *The 1st International Precision Agriculture Pakistan Conference 2022 (PAPC 2022)—Change the Culture of Agriculture*, page 13. MDPI. doi: 10.3390/environsciproc2022023013. URL https://www.mdpi.com/2673-4931/23/1/13.

Gregory S. Biging and Matthias Dobbertin. Evaluation of competition indices in individual tree growth models. 41(2):360–377. ISSN 0015-749X, 1938-3738. doi: 10.1093/forestscience/41.2.360. URL https://academic.oup.com/forestscience/article/41/2/360/4627255.

David A. Coomes and Robert B. Allen. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology*, 95(1):27–40, January 2007. ISSN 0022-0477, 1365-2745. doi: 10.1111/j.1365-2745.2006.01179.x. URL https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2745.2006.01179.x.

V.H. Dale, T.W. Doyle, and H.H. Shugart. A comparison of tree growth models. 29(1):145–169. ISSN 03043800. doi: 10.1016/0304-3800(85)90051-1. URL https://linkinghub.elsevier.com/retrieve/pii/0304380085900511.

Bin Ding, Huimin Qian, and Jun Zhou. Activation functions and their characteristics in deep neural networks. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 1836–1841, 2018. doi: 10.1109/CCDC.2018.8407425.

Forestry and Land Scotland. Scots pine. https://forestryandland.gov.scot/learn/trees/scots-pine, 2024. Accessed: 2025-04-04.

Michael Grieves. Digital twin: Manufacturing excellence through virtual factory replication. 03 2015. URL https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication.

Michael W. Grieves. Digital twins: Past, present, and future. In Noel Crespi, Adam T. Drobot, and Roberto Minerva, editors, *The Digital Twin*, pages 97–121. Springer International Publishing. ISBN 978-3-031-21343-4. doi: 10.1007/978-3-031-21343-4_4. URL https://doi.org/10.1007/978-3-031-21343-4_4.

Xuetao Jiang, Meiyu Jiang, YuChun Gou, Qian Li, and Qingguo Zhou. Forestry Digital Twin With Machine Learning in Landsat 7 Data. *Frontiers in Plant Science*, 13: 916900, June 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.916900. URL `https://www.frontiersin.org/articles/10.3389/fpls.2022.916900/full`.

Kasper Johansen, Nicholas C. Coops, Sarah E. Gergel, and Yulia Stange. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. 110(1):29–44. ISSN 00344257. doi: 10.1016/j.rse.2007.02.014. URL `https://linkinghub.elsevier.com/retrieve/pii/S0034425707000818`.

Caren E. Jones, Angeline Van Dongen, Jolan Aubry, Stefan G. Schreiber, and Dani Degenhardt. Use of Mobile Laser Scanning (MLS) to Monitor Vegetation Recovery on Linear Disturbances. *Forests*, 13(11):1743, October 2022. ISSN 1999-4907. doi: 10.3390/f13111743. URL `https://www.mdpi.com/1999-4907/13/11/1743`.

Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital twin in manufacturing: A categorical literature review and classification. 51(11): 1016–1022. ISSN 24058963. doi: 10.1016/j.ifacol.2018.08.474. URL `https://linkinghub.elsevier.com/retrieve/pii/S2405896318316021`.

Rui Neves Madeira, Pedro Albuquerque Santos, Oskars Java, Torsten Priebe, Eduardo Graça, Eszter Sárközi, Bernward Asprion, and Raquel Pinto-Bello Gómez. Towards digital twins for multi-sensor land and plant monitoring. 210:45–52. ISSN 18770509. doi: 10.1016/j.procs.2022.10.118. URL `https://linkinghub.elsevier.com/retrieve/pii/S1877050922015745`.

D.C. Malcolm, W.L. Mason, and G.C. Clarke. The transformation of conifer forests in Britain — regeneration, gap size and silvicultural systems. *Forest Ecology and Management*, 151(1-3):7–23, October 2001. ISSN 03781127. doi: 10.1016/S0378-1127(00)00692-7. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378112700006927`.

Wilson E. Marcílio and Danilo M. Eler. From explanations to feature selection: assessing shap values as feature selection mechanism. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–347, 2020. doi: 10.1109/SIBGRAPI51738.2020.00053.

Gonzalo Pajares, Andrea Peruzzi, and Pablo Gonzalez-de Santos. Sensors in agriculture and forestry. 13(9):12132–12139. ISSN 1424-8220. doi: 10.3390/s130912132. URL `https://www.mdpi.com/1424-8220/13/9/12132`.

Thomas A. M. Pugh, Mats Lindeskog, Benjamin Smith, Benjamin Poulter, Almut Arneth, Vanessa Haverd, and Leonardo Calle. Role of forest regrowth in global carbon sink dynamics. *Proceedings of the National Academy of Sciences*, 116(10):4382–4387, March 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1810512116. URL `https://pnas.org/doi/full/10.1073/pnas.1810512116`.

Susanne Raum. Forestry paradigms and policy change: The evolution of forestry policy in britain in relation to the ecosystem approach. *Land Use Policy*, 49:462–470, 08 2015. doi: 10.1016/j.landusepol.2015.08.021.

Jie Shao, Kai Hu, Changhu Wang, Xiangyang Xue, and Bhiksha Raj. Is normalization indispensable for training deep neural network? In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 13434–13444. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/9b8619251a19057cff70779273e95aa6-Paper.pdf.

Mitja Skudnik and Jernej Jevšenak. Artificial neural networks as an alternative method to nonlinear mixed-effects models for tree height predictions. *Forest Ecology and Management*, 507:120017, March 2022. ISSN 03781127. doi: 10.1016/j.foreco.2022.120017. URL https://linkinghub.elsevier.com/retrieve/pii/S0378112722000111.

Aristotelis C. Tagarakis, Lefteris Benos, George Kyriakarakos, Simon Pearson, Claus Grøn Sørensen, and Dionysis Bochtis. Digital twins in agriculture and forestry: A review. 24(10):3117. ISSN 1424-8220. doi: 10.3390/s24103117. URL https://www.mdpi.com/1424-8220/24/10/3117.

S. Trumbore, P. Brando, and H. Hartmann. Forest health and global change. 349 (6250):814–818. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aac6759. URL https://www.science.org/doi/10.1126/science.aac6759.

Eric VanDerHorn and Sankaran Mahadevan. Digital twin: Generalization, characterization and implementation. 145:113524. ISSN 01679236. doi: 10.1016/j.dss.2021.113524. URL https://linkinghub.elsevier.com/retrieve/pii/S0167923621000348.

Pip Wallace, Ross Martin, and Iain White. Keeping pace with technology: drones, disturbance and policy deficiency. 61(7):1271–1288. ISSN 0964-0568, 1360-0559. doi: 10.1080/09640568.2017.1353957. URL https://www.tandfonline.com/doi/full/10.1080/09640568.2017.1353957.

John Zachariassen, Karl F. Zeller, Ned Nikolov, and Tom McClelland. A review of the forest service remote automated weather station (RAWS) network. URL https://www.fs.usda.gov/treesearch/pubs/6118.