Urban Forestry & Urban Greening 101 (2024) 128527

Contents lists available at ScienceDirect

Urban Forestry & Urban Greening

journal homepage: www.elsevier.com/locate/ufug

Machine learning-based prediction of tree crown development in competitive urban environments

Hadi Yazdi [a,*], Astrid Moser-Reischl [b], Thomas Rötzer [b], Frank Petzold [a], Ferdinand Ludwig [a]

[a] School of Engineering and Design, Technical University of Munich, Munich, Germany
[b] School of Life Sciences, Technical University of Munich, Freising, Germany

ARTICLE INFO

Keywords:
Tree light competition
Tree light demand
Tree crown geometry
Tree growth
Urban forestry
Urban green infrastructure

ABSTRACT

In urban forestry, managing trees is crucial for sustainable urban environments, especially in the context of climate change and the urban heat island effect. This research explores the complex dynamics of tree crown geometry development by asking the question: how do surrounding objects, such as nearby trees, buildings, and other urban structures, affect the shape of tree crowns? It aims to uncover how competition for light and space influences tree crown development in competitive urban environments. Our study employs machine learning models on six main species in Munich, using the measured data from the LiDAR scans, with the Hist Gradient Boosting Regressor (HGBR) emerging as the most promising performer across various metrics. Notably, the evaluation of 13 models reveals the HGBR's consistent ranking as the best or second-best across all tree crown dimensions assessed, with R^2 values reaching 0.83 for the tree height model and 0.7 on average for crown radiuses in eight directions. Employing SHapley Additive exPlanations (SHAP) values elucidate factors influencing model predictions, emphasising the significant impact of adjacent trees and buildings. After evaluating the models to include additional tree species in Munich, the models show strong predictive capabilities for some additional species. Despite the studies' limitations - the models are only valid for selected species, and there are constraints in predicting tree crown start height - our findings contribute valuable insights for urban forestry management and planning.

1. Introduction

The management of urban trees has become a challenging task in recent decades, as it involves finding the right species for the right location for planting and caring for the trees so that they survive on site and provide sufficient ecosystem services for a healthy and livable urban climate (Young, 2010; McPherson and Peper, 2012; Endreny, 2018). Particularly with regard to ongoing climate change and the urban heat island effect, urban trees are an important, sustainable component for climate adaptation and mitigation in cities Ramyar et al. (2021). As Berglihn and Gómez-Baggethun (2021) stated, urban trees play a crucial role in defining a sense of place, maintaining environmental quality and improving well-being in and around the places where most people now live FAO (2018).

The urban environment is a complex system with multifunctional green, grey and blue structures. Thus, urban trees are often named urban green infrastructures, as they provide a wide range of ecosystem services MEA (2005). Urban trees especially are increasingly recognised for their use in cities due to their multiple ecological and social benefits Li et al. (2017), air purification through pollutant filtering McDonald et al. (2018), temperature regulation through cooling by shading and by transpiration (Zhang et al., 2022; Morakinyo et al., 2018; McPherson et al., 2018; Kroeger et al., 2018), run-off reduction Rahman et al. (2023), noise buffering, promotion of biodiversity Clucas et al. (2018) and recreational effects Korkou et al. (2023).

The tree canopy provides many of these ecosystem services, especially by the crowns of individual trees Franceschi et al. (2022). The geometry of tree crowns plays a central role in determining the extent of these services. Canopy structure, canopy density and canopy size have a direct impact on improving the local climate, e.g. the effect of trees in providing shade and cooling by transpiration or reducing solar radiation is strongly related to the shape and density of the canopy and the overall structure of the trees (Franceschi et al., 2022; Zhu et al., 2021; Shahidan and Jones, 2008). In addition, tree canopies influence wind dynamics and are the primary source of habitat provision. Understanding the nuances of tree growth and canopy geometry is essential for sustainable

* Corresponding author.
E-mail address: hadi.yazdi@tum.de (H. Yazdi).

https://doi.org/10.1016/j.ufug.2024.128527
Received 9 May 2024; Received in revised form 25 September 2024; Accepted 26 September 2024
Available online 30 September 2024

urban planning and design Yazdi et al. (2023).

Research on urban tree allometries such as (Moser-Reischl et al., 2019; Pretzsch et al., 2015; Yang et al., 2023) estimate the crown volume based on the crown shape of a cylinder. However, as Franceschi et al. (2022) showed, this under- or overestimates the crown volume of many tree species tremendously (up to 87 %), which also affects the estimation of the provided ecosystem services of a tree. As the study found out, spherical crown shapes of urban trees provide the greatest shade, while ovoid crown shapes provide the highest shade density Franceschi et al. (2022). Moreover, the crown shape of a tree can even vary over the lifetime of a tree, depending on its age.

Thus, changes in canopy volume and shape can directly affect the extent to which ecosystem services are provided. Canopy variables, especially the crown volume, are often applied as proxies for estimating leaf area, shade provision, transpiration and filtration of fine particles. Crown volume is one of the most important variables to calculate the cooling efficiency of trees Gratani and Varone (2006). Therefore, the precise calculation of a tree's crown volume is an important prerequisite for accurately estimating such growth parameters and ecosystem services in planning and modelling approaches for urban planners and foresters alike Zhu et al. (2021). The crown volume is regarded as a significant variable among crown characteristics when aiming to enhance the credibility of empirical models, such as those concerning growth and yield, and to evaluate the dynamics of forests Bragg (2001).

The development of space occupancy over time is significant for green space planning, i.e. what size a tree will be in a certain period after planting Franceschi et al. (2022). Depending on the tree species, this can vary significantly Pretzsch et al. (2012). In addition, the changing precipitation absorption with increasing tree size (and thus plant surface area) must be taken into account, as this leads to changes in the water balance of a tree (e.g. Baptista et al., 2018), affecting the tree growth and allocation patterns Rötzer et al. (2019).

However, the formation and shape of a tree crown depends not only on the species and age but also on its direct surroundings like light and space competition by other trees, adjacent buildings, pruning for safety reasons or traffic issues and cable connections Pretzsch et al. (2015). The complex relationship between trees and their surroundings, especially the geometry of their crowns, is therefore of particular importance for shaping the city's urban image and thermal comfort.

This research paper delves into the development of tree crown geometry, unravelling its crucial role in enhancing the urban environment. It also addresses the existing gaps in the interactions between trees and their urban context. While previous studies have primarily focused on elucidating light competition between trees in forest stands (Kanjevac et al., 2021; Pretzsch, 2022), the effects of light competition on tree crown geometry between buildings and urban trees, as well as urban tree to urban tree, have been overlooked. This comprehensive study on the interrelationship of trees and the complex dynamics between trees and their surrounding structures is, therefore, an important topic in urban planning. Despite advances in urban forestry research, the prediction of growth and crown geometry of trees competing with each other or with anthropogenic objects still needs to be improved. By investigating these issues, this research aims to provide valuable insights into urban forestry and provide a basis for informed decision-making in urban planning and the management of urban green spaces.

This study attempts to close these gaps by answering the following research questions:

- What is the efficacy of machine learning techniques in forecasting the urban tree crown geometry development?
- In an urban environment characterised by competition with other trees and the surrounding buildings, which features of the surrounding trees and buildings have the most significant impact on the development of tree crowns, and how can these features be quantified and integrated into predictive models?
- To what extent can machine learning models trained on specific tree species accurately predict crown geometry for species for which the models have not been trained?
- How does the predictive performance of a crown geometry development model, trained on data from a specific urban environment, vary when applied to the same tree species in different urban environments and climate conditions?

2. Methods

To predict tree crown geometry, we developed TreeML, a data-driven machine learning approach that leverages an ensemble of ten distinct regression models. Throughout this discussion, we will refer to this novel method as TreeML. These models can forecast tree height, tree crown start, and tree crown radius in eight directions. By combining these predicted measures, a rough visualisation of the tree crown geometry in relation to nearby objects can be achieved. Alongside these models, TreeML incorporates three additional ones for predicting the maximum crown diameter, crown projection area, and total woody volume of trees. These supplementary models contribute to more precise calculations of tree growth biomass and ecosystem services.

Creating a valuable and precise dataset is crucial in developing machine learning models. For this purpose, we utilised TreeML-Data, an accurate open-source urban tree dataset Yazdi et al. (2024). Subsequently, a model for measuring distance and height was developed to enhance the dataset with accurate geometry data for surrounding objects. Following this, the data was prepared, and several new features were added to aid the models in understanding the local geometry around the trees. Finally, the data was trained and tested using 16 different regression machine learning models to compare results and select the most effective model.

2.1. TreeML-data

TreeML-Data Yazdi et al. (2024) serves as this study's foundational material, offering an urban tree dataset. Comprising labelled point clouds from 40 urban scanning projects on streets in Munich, Germany, the dataset includes 3755 point clouds specific to individual trees during winter (without leaves). It encompasses quantitative structure models (QSM), detailed tree structure measurements, and tree graph structure models corresponding to trees in these urban settings.

For this study, we utilised the tree structure measurement data within TreeML-Data. This dataset includes comprehensive tree structure information, such as TreeID, botanical name, Diameter at Breast height (DBH) (m), tree height (m), crown start height (m), maximum crown diameter (m), crown projection area (m^2), and total volume (L). Additionally, the dataset features crown radius (m) in 72 different directions (crown radius definition is the maximum distance of the crown to the trunk location in a plan view - see Fig. 2), spaced every five degrees. While the dataset encompasses various species with different sample sizes, our focus was on six main species during the training step to ensure an adequate sample representation: *'Tilia cordata'* (837 samples), *'Acer platanoides'* (742 samples), *'Platanus x hispanica'* (528 samples), *'Robinia pseudoacacia'* (509 samples), *'Populus nigra var. italica'* (297 samples), and *'Aesculus hippocastanum'* (256 samples). In total, the dataset comprises 3169 tree samples of different species, featuring 80 structural measurement features (eight main measurements and 72 crown radiuses).

Additionally, alongside the tabular tree structure measurement data, the study incorporates tree and building point clouds from TreeML-Data. These point clouds, categorised in three classes (building, tree, other), are utilised parallel to the tabular dataset to integrate the features of surrounding objects into the training dataset (see section 2.2).

2.2. Surrounding objects' distance and height measurement

To train a prediction model and examine the impact of surrounding objects on tree crown geometry development, we require information regarding the geometry of these surrounding objects. Consequently, we developed a Python script code specifically designed to measure the distance and height of the surrounding objects in proximity to the tree. This tool necessitates knowledge of the tree's location and the point clouds of nearby buildings and trees to extract these measurements. Illustrated in Fig. 1, the tool gauges the distance to buildings and trees in each 5-degree interval (across 72 directions) and their respective heights in these directions. Moreover, this measurement is repeated for surrounding trees. Ultimately, this process adds 288 new features (columns) to our dataset, including disAdjacentBuilding (m), heightAdjacentBuilding (m), disAdjacentTree (m), and heightAdjacentTree (m) across 72 directions (each 5 degrees).

2.3. Data preparation and preprocessing

In the initial step of data preparation, we randomly set aside 20 % of the data as the test dataset. This was done before applying any augmentation methods or increasing the dataset size. The purpose of this step is to ensure that model evaluation occurs on data that the model has not encountered before. Subsequently, data augmentation and preprocessing were applied to both the training and test datasets.

To increase the accuracy of the training models on the crown geometry, we increased the dataset dimensionally by utilising the novel data augmentation method on the TreeML-Data. Data augmentation can be characterised as a technique to reduce overfitting and increase the dataset size Maharana et al. (2022). In addition to the dimensional increment, we focused on eight main directions (N, W, S, E, NW, SW, SE, NE) to reduce the required training models for visualising the crown geometry. An augmentation method was employed in the preprocessing steps of machine learning to increase the dataset's size dimensionally. Each tree in the dataset was duplicated nine times, and measurements for each main direction were used from − 20° to + 20° (refer to Fig. 2). For instance, in the east direction, all measurements from S + 70° to E + 20° were considered part of the East direction for crown radius and surrounding object measurements. This augmentation method increased the dataset size by a factor of nine, resulting in 28,521 samples. This augmentation benefits the training of machine learning models with larger datasets. Consequently, the dataset's 360 features (crownRadius (m), disAdjacentBuilding (m), heightAdjacentBuilding (m), disAdjacentTree (m), and heightAdjacentTree (m) in 72 directions) were reduced to 40 features for just the eight main geographical directions.

Drawing from research on urban tree allometries Fauk and Schneider (2023) and existing datasets of urban tree measurements Rötzer et al. (2021), we developed double logarithmic equations to approximate the tree structures of primary species in Munich. The equation, ln(y) = a + b * ln(DBH), where y represents tree height, crown diameter, or crown start height, was fitted to the data. Table 1 displays the values of a and b for each equation across various species. The table 1 also includes the Relative Standard Error (RSE) (see equation (1)) and R^2 score (see equation (4)), assessing the accuracy of the fitted allometric equations. The result of these three allometric equations in table 1 were added to the dataset to enrich it with rough estimations of the tree structure. These features are titled equation_treeHeight(m), equation_crownStartHeight(m), and equation_crownRadius(m), as detailed in Table 2.

$$\mathrm{RSE} = \frac{\sqrt{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}}{\sum_{i=1}^{n} y_i} \tag{1}$$

where:

- y_i is the actual value for the i-th observation,
- $\widehat{y}_i$ is the predicted value for the i-th observation,
- n is the number of observations.

In addition to the fundamental structural measurements of trees, information about surrounding objects, and measurements derived from allometric equations, we have extracted additional sub-features based

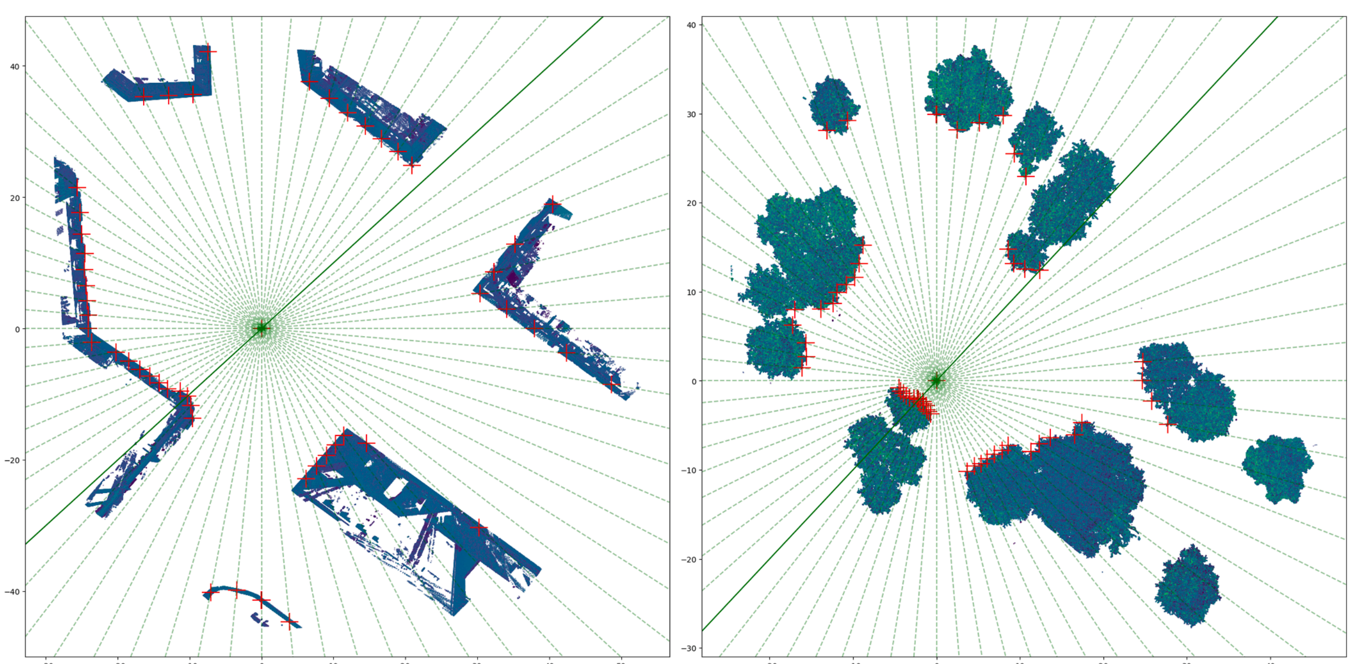

Fig. 1. The distance and height measurement for surrounding objects around a sample tree is conducted at each 5-degree interval, spanning 72 directions. In the left image, the measurements pertain to neighbouring buildings, encompassing the distance from the tree location to the buildings and the height of the building in the specified direction. The right image showcases the analogous measurements for the surrounding trees. The Maximum measuring distance in this tool is set to 100 m based on the urban tree measurement practices in previous studies Rötzer et al. (2019).

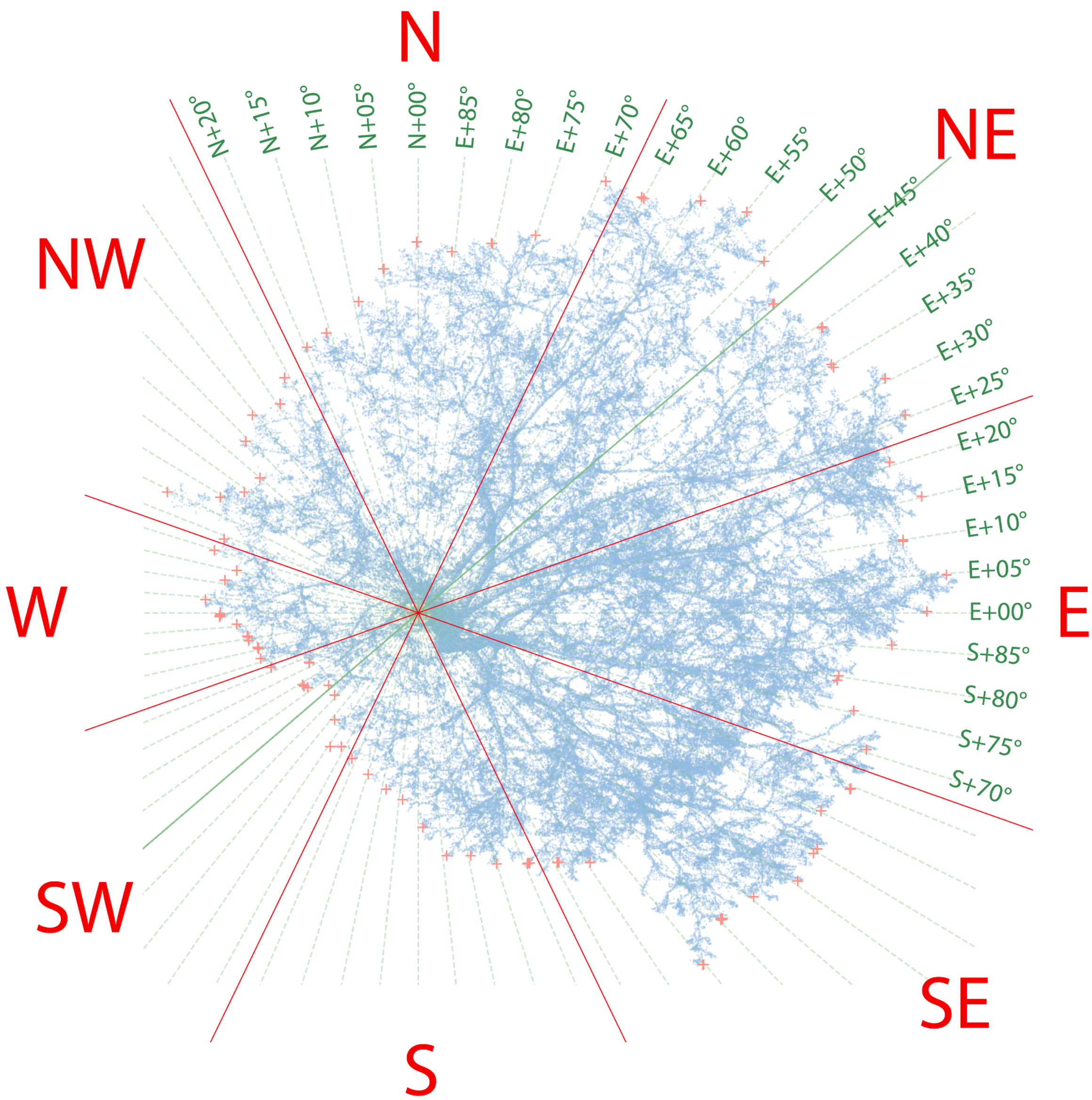

Fig. 2. The dataset size is expanded by duplicating the crown radius measurements in 72 directions nine times for each main direction (N, W, S, E, NW, SW, SE, NE). All measurements ranging from S + 70° to E + 20° (covering nine directions) are considered under the east direction in the dataset.

on the main features to enhance model training. Fig. 3 visually represents the features in grey and green colours. The measurements in grey represent the basic structural measurements and surrounding object information, while the green text signifies features derived from allometric equation estimations. For a detailed explanation of all main features and sub-features in the dataset, refer to Table 2.

Sub-features are computed based on surrounding object measurements and their ratios with main features and allometric equations. All sub-features are recorded in eight main geographic directions (N, W, S, E, NW, SW, SE, NE). In the main features section, aside from DBH, tree height, crown start height, and allometric equation features, all other features are direction-based. This implies eight distinct measurements for each feature corresponding to each main direction.

2.4. Model training and testing

As outlined in the initial step of Section 2.3, we designated 20 % of the dataset as the test data before conducting any data preprocessing and cleaning. Consequently, after the preprocessing step, two distinct datasets were formed, each containing the features listed in Table 2. Based on the features outlined in Table 2, the main and sub-features serve as the input (X) data for the machine learning models, while each prediction feature acts as an individual output (Y). Hence, 13 distinct machine learning models were trained to predict each of the following features: tree height, crown start height, crown radius (in 8 main directions), crown projection area, maximum crown diameter, and total volume. Next, the "botanical.name" feature, encompassing categorical variables, was converted into dummy variables, where the dummy variable is binary (yes = 1 or no = 0).

To gain an overview of performance across a diverse array of machine learning models, we utilised several widely recognised models from the scikit-learn library Pedregosa et al. (2024). A comparison was made among 16 common regression models with their default settings, including Random Forest Regressor, Support Vector Regression, Gradient Boosting Regressor, and Extra Trees Regressor. Additionally, a basic Artificial Neural Network (ANN) model was tested, constructed with Keras Chollet (2018), consisting of two hidden layers with 16 and 8 nodes, respectively.

To evaluate the result of the model on the test dataset, we used three different evaluation metrics on each training model. These are the Mean

Table 1
The variations of the double logarithmic equations (ln(y) = a + b * ln(DBH)) for estimating tree height, tree crown start, and crown diameter (y) based on DBH for six main species in Munich. (TH = Tree height, CSH = Crown start height, CD = Crown diameter).

Species Number of samples (n)	y	a	b	RSE	R^2
Tilia cordata n = 876	TH	0.89751	0.51497	0.1931	0.5
	CSH	0.88474	0.13545	0.3006	0.03
	CD	−0.265	0.66905	0.2021	0.6
Acer platanoides n = 425	TH	0.58282	0.57232	0.1978	0.64
	CSH	0.5752	0.16624	0.2879	0.07
	CD	−0.081	0.63817	0.1828	0.72
Platanus x hispanica n = 237	TH	0.75336	0.55972	0.1192	0.85
	CSH	1.242624	0.004335	0.2685	0.00006
	CD	−0.11448	0.69816	0.1302	0.88
Robinia pseudoacacia n = 200	TH	1.00757	0.45387	0.215	0.55
	CSH	0.8696	0.15307	0.3297	0.05
	CD	−0.07680	0.61171	0.1892	0.74
Populus nigra var. italica n = 73	TH	1.4582	0.43943	0.1404	0.71
	CSH	−0.6253	0.4481	0.8457	0.07
	CD	−0.18672	0.47888	0.2457	0.5
Aesculus hippocastanum n = 290	TH	0.8242	0.47354	0.1756	0.5
	CSH	0.57639	0.12873	0.3706	0.02
	CD	−0.05216	0.59508	0.1612	0.65

Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2. MAE is the average error from every dataset sample and predicted value (equation (2)).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{2}$$

RMSE, or Root Mean Squared Error, is the square root of the average of the squared differences between the actual (observed) and predicted values (equation (3)).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Moreover, R^2, also referred to as the Coefficient of Determination, is a metric regularly ranging from 0 to 1, indicating the goodness of fit of our regression model to the data. However, Chicco et al. (2021) pointed out that R^2 values can be negative, indicating that a regression model's performance may be inadequate. It quantifies the proportion of variance in the dependent variable that can be accounted for by our model. A higher R^2 value, closer to 1 or 100 %, suggests that our model is more adept at predicting the dependent variable Wright (1921) (equation (4)).

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4}$$

At the numerator of the formula, we find the Residual sum of squared errors of the regression model (SS_{RES}), which is essentially equivalent to the Root Mean Squared Error (RMSE). However, unlike RMSE, it is not divided by the number of samples (n). The denominator of the formula comprises the total sum of squared errors (SS_{TOT}). This involves comparing the actual y values to our baseline model, which is the mean. To achieve this, we square the difference between all the actual y values and the mean, summing them together. Negative R^2 values result when the SS_{RES} is higher than SS_{TOT}, which means the regression equation is worse than the mean value.

In addition to evaluating the model's performance using various metrics and visualising predictions, we explored the significance of different features in the prediction process. Assessing the importance of features becomes challenging in large datasets with complex machine learning models, such as the Hist Gradient Boosting Regressor (HGBR), which is an ensemble model. To address this challenge, we employed SHAP (Shapley Additive exPlanations) values Lundberg and Lee (2017) to quantify the importance of various features in model predictions.

In the subsequent phase, we assessed the trained model using additional common tree species in Munich that were not part of the training set. This evaluation aimed to determine the model's applicability to other species and assess whether these species respond to light competition similarly to the main species. The new species and their sample sizes include *Fraxinus excelsior*: 83, *Corylus colurna*: 80, *Prunus species*: 59, *Carpinus betulus*: 37, *Acer pseudoplatanus*: 31, *Sorbus aria*: 25, *Fagus sylvatica*: 11, *Betula pendula*: 10. Since the models were trained with only six main species, the new species were not recognisable for the models.

To test the models' performance on these new species, each new species was individually labelled as a known species., and the models were evaluated using various metrics such as R^2, MAE, and RMSE. Each model was evaluated six times for each new species, referencing the six main species.

Then, we utilised a small sample dataset from the cities of Hamburg and Essen to assess the TreeML models' performance on the primary tree species in other cities with different climate conditions and environments. This point cloud dataset comprises 28 *Tilia europaea* and *T. cordata* trees in Hamburg and 14 *P. x hispanica* trees in Essen, and their local environment. The TreeML-Structure Measurement tool Yazdi et al. (2024) was employed to measure the tree structure and gather information on surrounding objects.

3. Results

3.1. Evaluation of the machine learning models

Following the training of 13 models with 16 common regression methods, we identified the top-performing models among them. The five models that exhibited the best performance are the Hist Gradient Boosting Regressor (HGBR), Extra Trees Regressor (ETR), Gradient Boosting Regressor (GBR), Nu Support Vector Regression (NSVR), Random Forest Regressor (RFR), and Artificial Neural Network model (ANN). Tables 3, 4, and 5 showcase these models' R^2, MAE, and RMSE scores, respectively.

The R^2 score for the tree height model reached approximately 0.83, establishing it as the top-performing model. For the crown radius models in eight directions, the R^2 score averaged around 0.7. However, the crown start model did not perform satisfactorily, with an R^2 score of approximately 0.24. On the other hand, the models for maximum crown diameter, crown projection area, and total volume demonstrated strong performance, achieving R^2 scores of around 0.81, 0.78, and 0.70, respectively.

Gradient Boosting Regressor (GBR) is a robust ensemble technique commonly applied to regression problems. It functions by combining multiple weak learners, usually decision trees, to create a more powerful predictive model. The process begins with the construction of simple trees with a single node, followed by additional trees that focus on correcting the errors made by their predecessors. The impact of each tree is moderated by a learning rate, which ensures that no single tree dominates the prediction. This iterative procedure continues, adding new trees to the ensemble, until the specified number of trees is reached or there is no significant improvement in the model's accuracy Friedman (2002). Furthermore, Hist Gradient Boosting Regression (HGBR) enhances this method by employing histogram-based techniques to speed up the training of decision trees. This involves binning continuous features into discrete intervals, which significantly reduces computation time. The hist gradient boosting approach applies this faster algorithm to input variables, making it more efficient for large datasets. Each additional tree in the ensemble aims to refine the predictions by addressing the errors of the existing models, thus continually improving the overall accuracy of the predictions Gayathri et al. (2022).

Table 2
List of the features in the dataset and their definitions.

	Features	Definitions
Main Features	DBH(m)	the diameter of the circle fitted to the height 1.36-1.33 m
	botanical.name	the botanical name (species) of the tree
	disAdjacentBuilding(m)*	distance to the adjacent building in 8 main directions
	heightAdjacentBuilding(m)*	height of the adjacent building in 8 main directions
	disAdjacentTree(m)*	distance to the adjacent tree in 8 main directions
	heightAdjacentTree(m)*	height of the adjacent tree in 8 main directions
	equation_treeHeight(m)	estimation of the tree height based on the allometric equation
	equation_crownStartHeight(m)	estimation of the crown start height based on the allometric equation
	equation_crownRadius(m)	estimation of the crown radius based on the allometric equation
Sub-features	sin_building*	Sin(x) (see figure 3)
	sin_tree*	Sin(y) (see figure 3)
	tan_building*	Tan(x) (see figure 3)
	tan_tree*	Tan(y) (see figure 3)
	ratio_radius/dist_building*	equation_crownRadius(m) / disAdjacentBuilding(m)
	ratio_radius/dist_tree*	equation_crownRadius(m) / disAdjacentTree(m)
	ratio_dist/equ-height_building*	disAdjacentBuilding(m) / equation_treeHeight(m)
	ratio_dist/equ-height_tree*	disAdjacentTree(m) / equation_treeHeight(m)
	ratio_height/equ-height_building*	heightAdjacentBuilding(m) / equation_treeHeight(m)
	ratio_height/equ-height_tree*	heightAdjacentTree(m) / equation_treeHeight(m)
Predictions	treeHeight(m)	Height (m) of the tree
	crownStartHeight(m)	Crown's base height (m) from the ground
	crownRadius(m)*	crown radius in 8 main directions
	crownProjectionArea(m2)	Area (m2) of the crown's planar projection's convex hull
	crownDiameterMax(m)	Maximum horizontal crown diameter (m)
	totalVolume(L)	Total volume (L) of the tree
*These features are direction-based, which means there are eight different measurements of each feature for each main direction.		

*These features are direction-based, which means there are eight different measurements of each feature for each main direction. The features are divided into three groups: main features, sub-features, and predictions (see Fig. 3)). The blue and red colours of the main features and sub-features are hints for the readability of the Fig. 6 on the importance of the features during the model training.

Across all three evaluation metrics, the Hist Gradient Boosting Regressor (HGBR) consistently emerged as the top or second-best performer in all models. Consequently, we selected the Hist Gradient Boosting Regressor (HGBR) as the best-performing model for further analysis and evaluation. In addition to the evaluation results in the Tables 3, 4, and 5 for the six best performing models, Fig. 4 shows the residual diagram ($y_i - \hat{y}_i$) of the ten crown geometry prediction models of the Hist Gradient Boosting Regressor (HGBR) on the test dataset.

To visually assess the TreeML model's efficacy in predicting crown geometry development, we selected a sample *"Tilia cordata"* tree from the test dataset in Munich, illustrated in Fig. 5. This tree is situated near a building in the north direction, with neighbouring trees on the east and west sides. The green points depict the actual crown shape in both plan and section views. The red line represents the estimated crown shape based on allometric equations from Table 1, derived from tree height, crown start height, and crown radius estimations based on the measured DBH and species.

In contrast, the TreeML model incorporates surrounding object geometries, enhancing the accuracy of crown geometry predictions by considering the competition with neighbouring objects. The blue line in Fig. 5 reflects the TreeML model's prediction for this specific tree in Munich. Tree crowns tend to grow higher and extend through the south and southeast directions.

Fig. 6 presents the SHAP values for two sample models from our 13 models. The diagram on the left corresponds to the tree height prediction model, while the diagram on the right illustrates the results for the tree crown radius in the east direction. Analysing the SHAP values for the Tree Height model reveals that, in addition to DBH and allometric equation features, the height of adjacent trees in eight different directions significantly influences the main tree's height. Subsequently, features related to adjacent buildings impact the main tree's height. Additionally, it is observed that the height of adjacent trees on the northern side has a more pronounced effect on the main tree's height than those on the southern side (see Fig. 6 left).

Examining the SHAP value chart for the crown radius in the east direction, it becomes evident that the third most important feature is the distance to the adjacent tree in the east direction. Alongside this distance measure, other sub-features related to the adjacent tree in the east direction exhibit substantial influence, following the allometric equations and species. Similar to the Tree Height prediction model, adjacent buildings demonstrate lesser importance than adjacent trees in crown radius prediction (see Fig. 6 right). The SHAP value diagram for all 13 models can be found in the paper's supplementary material.

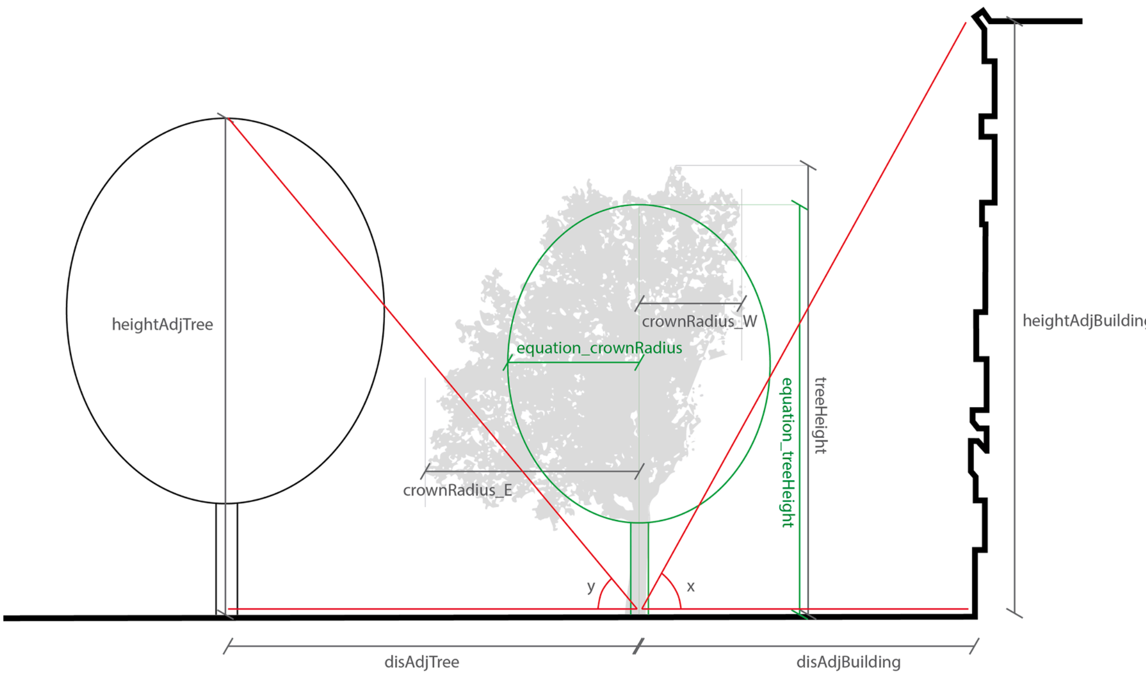

Fig. 3. Visual representation of the main features within the dataset, distinguishing them in grey and green colours. The grey colour denotes the basic structural measurements of the tree and surrounding object information. In contrast, the green colour represents measures derived from the estimation of the allometric equation. Additionally, the red colour indicates the degree of surrounding trees and buildings' height from the tree location, contributing to the extraction of sub-features.

Table 3
The R^2 metrics for the five best-performing models on the train and test dataset.

	R^2-train/R^2-test					
Models	HGBR	ETR	GBR	NSVR	RFR	ANN
Tree Height (m)	0.95/ **0.83**	1/ 0.79	0.88/ 0.81	0.85/ 0.80	0.99/ 0.80	0.68/ 0.78
Crown Start Height (m)	0.83/ 0.22	1/ **0.24**	0.64/ 0.23	0.37/ 0.23	0.98/ 0.22	0.20/ 0.22
crown_E (m)	0.88/ **0.71**	1/ 0.70	0.76/ 0.69	0.70/ 0.63	0.99/ 0.69	0.62/ 0.65
crown_NE (m)	0.97/ **0.71**	1/ 0.70	0.75/ **0.71**	0.70/ 0.63	0.99/ 0.68	0.48/ 0.62
crown_NW (m)	0.85/ 0.67	1/ **0.68**	0.72/ 0.67	0.67/ 0.58	0.99/ 0.65	0.70/ 0.61
crown_N (m)	0.87/ **0.68**	1/ 0.66	0.75/ 0.67	0.69/ 0.57	0.99/ 0.65	0.45/ 0.58
crown_SE (m)	0.89/ **0.68**	1/ 0.66	0.75/ **0.68**	0.70/ 0.59	0.99/ 0.65	0.64/ 0.59
crown_SW (m)	0.86/ **0.67**	1/ **0.67**	0.74/ 0.65	0.68/ 0.59	0.99/ 0.65	0.61/ 0.63
crown_S (m)	0.87/ **0.68**	1/ 0.66	0.75/ 0.67	0.69/ 0.61	0.99/ 0.67	0.70/ 0.60
crown_W (m)	0.86/ 0.68	1/ **0.69**	0.73/ 0.68	0.67/ 0.59	0.99/ 0.68	0.26/ 0.61
Crown Diameter Max (m)	0.93/ **0.81**	1/ **0.81**	0.85/ 0.80	0.82/ 0.78	0.99/ 0.79	0.79/ 0.78
Crown Projection Area (m2)	0.94/ **0.78**	1/ 0.77	0.85/ 0.77	0.52/ 0.54	0.99/ 0.76	0.79/ 0.75
Total Volume (L)	0.93/ **0.70**	1/ 0.68	0.80/ 0.69	−0.07/ −0.04	0.99/ 0.65	0.65/ 0.64

These five models are the Hist Gradient Boosting Regressor (HGBR), Extra Trees Regressor (ETR), Gradient Boosting Regressor (GBR), Nu Support Vector Regression (NSVR), Random Forest Regressor (RFR), and Artificial Neural Network model (ANN). (Bold: best-performing models)

3.2. Evaluation of the models on other species in Munich (eight extra species)

In this phase, we extended the evaluation of our trained models to an additional eight tree species beyond the six primary species utilised during the training process. Tables 6 and 7 illustrate the evaluation of two sample models (Tree height and Crown radius_E) for eight new species using the R^2 evaluation metric. The paper's supplementary material evaluates all 13 TreeML models with all metrics.

Table 6 displays the R^2 evaluation results of the tree height model on eight new species, referencing the six main species each time. It indicates that the TreeML model performs well in predicting tree height for *F. excelsior*, *C. colurna*, *C. betulus*, *A. pseudoplatanus*, and *Sorbus aria*, with R^2 metrics ranging from 0.49 to 0.77. However, the model did not exhibit strong performance in predicting tree height for *Prunus species*, *Fagus sylvatica*, and *Betula pendula*.

Table 7 presents the R^2 results of the Crown radius_East model for the eight new species. The findings show that the TreeML model performs well in predicting the crown radius in the east direction for most species, with R^2 values ranging from 0.43 to 0.78. However, the prediction of the crown radius_East for Prunus species is not accurate, similar to the Tree height prediction for this species.

3.3. Evaluating the model performance on two tree species in the cities of Hamburg and Essen

In this phase, we assessed our trained models' adaptability and robustness using two smaller datasets from different cities to explore the impact of regional variations on model performance. Tables 8 and 9 present the evaluation metrics for the performance of the TreeML models on these trees, excluding the Total volume model due to limited access to the quantitative structure modelling (QSM) Raumonen et al. (2013) info of these trees for estimating their total volume.

Table 8 displays the results of the models on 28 trees in Hamburg. The evaluation indicates that the performance of the crown radius models in all directions, maximum crown diameter and crown

Table 4
The MAE metrics for the five best-performance models.

Models	Mean Absolute Error (MAE) HGBR	ETR	GBR	NSVR	RFR	ANN
Tree Height (m)	**1.32**	1.38	1.40	1.48	1.38	1.47
Crown Start Height (m)	0.77	**0.76**	0.80	0.79	0.77	0.79
crown_E (m)	**0.76**	0.75	0.77	0.88	0.76	0.83
crown_NE (m)	**0.75**	0.75	**0.75**	0.86	0.77	0.84
crown_NW (m)	0.72	**0.72**	0.72	0.86	0.74	0.78
crown_N (m)	**0.71**	0.72	0.73	0.87	0.73	0.80
crown_SE (m)	**0.73**	0.76	**0.75**	0.86	0.77	0.83
crown_SW (m)	**0.77**	**0.78**	0.80	0.90	0.81	0.84
crown_S (m)	**0.75**	0.78	0.76	0.87	0.77	0.82
crown_W (m)	0.72	**0.71**	0.73	0.86	0.72	0.80
Crown Diameter Max (m)	**1.14**	**1.12**	1.18	1.27	1.16	1.25
Crown Projection Area (m2)	**14.47**	14.26	14.87	26.05	14.68	16.01
Total Volume (L)	**1137.00**	1165.48	1173.86	2811.43	1199.34	1272.63

These five models are the Hist Gradient Boosting Regressor (HGBR), Extra Trees Regressor (ETR), Gradient Boosting Regressor (GBR), Nu Support Vector Regression (NSVR), Random Forest Regressor (RFR), and Artificial Neural Network model (ANN). (Bold: best-performing models)

Table 5
The RMSE metrics for the five best-performance models.

Models	Root Mean Squared Error (RMSE) HGBR	ETR	GBR	NSVR	RFR	ANN
Tree Height (m)	**1.94**	2.16	2.03	2.12	2.12	2.19
Crown Start Height (m)	1.75	**1.72**	1.74	1.73	1.74	1.74
crown_E (m)	**1.04**	1.05	1.07	1.17	1.07	1.19
crown_NE (m)	**1.00**	1.04	1.03	1.16	1.08	1.16
crown_NW (m)	**0.99**	1.02	1.03	1.16	1.06	1.11
crown_N (m)	**1.02**	1.04	1.03	1.17	1.06	1.19
crown_SE (m)	**1.02**	1.05	1.03	1.15	1.07	1.11
crown_SW (m)	**1.08**	1.10	1.13	1.22	1.12	1.17
crown_S (m)	**1.06**	1.08	1.07	1.17	1.07	1.15
crown_W (m)	**1.00**	1.01	1.03	1.15	1.02	1.10
Crown Diameter Max (m)	**1.66**	**1.66**	1.69	1.77	1.74	1.78
Crown Projection Area (m2)	**22.86**	23.10	23.28	33.09	23.98	24.35
Total Volume (L)	**1877.48**	1961.10	1910.05	3493.29	2022.63	2058.29

These five models are the Hist Gradient Boosting Regressor (HGBR), Extra Trees Regressor (ETR), Gradient Boosting Regressor (GBR), Nu Support Vector Regression (NSVR), Random Forest Regressor (RFR), and Artificial Neural Network model (ANN). (Bold: best-performing models)

projection area needs more improvement for better results, with R^2 values ranging from 0.43 to 0.59. However, the tree height model demonstrates weak performance with an R^2 value of around 0.36. Additionally, the crown start height performance is deemed unacceptable. In contrast to the model evaluation on trees in Hamburg, all TreeML models exhibit unacceptable performance on the *P. x hispanica* trees in Essen (Table 9).

4. Discussion

This study delves into the intricate dynamics of tree crown geometry development, aiming to unravel key factors influencing tree development in urban settings. Previous studies emphasised the significance of canopy structure, density, and size in influencing local climate. However, the existing literature revealed a notable gap in the availability of precise models capable of predicting changes in tree crown geometry. To address this gap as the primary goal of this study, it employs machine learning techniques to develop a robust model for urban tree crown geometry development. To identify the best-performing machine learning model, we initially compared various methods. However, a detailed comparison of different model settings was not the primary objective of this study. Therefore, we only present the initial comparison results without delving into the specific hyperparameters of each machine learning method.

Our comprehensive evaluation of 13 machine learning models highlighted the Hist Gradient Boosting Regressor (HGBR) as the most promising performer across various evaluation metrics, including R^2 (Table 3), Mean Absolute Error (MAE) (Table 4), and Root Mean Squared Error (RMSE) (Table 5). This model consistently ranked the best or second-best across all tree crown dimensions assessed. Notably, the R^2 values for the tree height and maximum crown diameter models demonstrated the efficacy of the HGBR model, reaching values as high as 0.83. Furthermore, visualising the TreeML model's performance on a specific *T. cordata* tree in Munich provided tangible evidence of its accuracy in predicting crown geometry growth (Fig. 5). The model, which considers surrounding objects' geometries for more accurate predictions, exhibited a nuanced understanding of how light competition with neighbouring trees and buildings influences tree crown shape. This level of detail is crucial for predicting changes in tree projection area and volume, ultimately affecting the tree's shading and ecosystem services. Notably, our results echo the views of Franceschi et al. Franceschi et al. (2022), who stressed the significance of precise crown geometry predictions in assessing tree shading and ecosystem services.

Next, we employed SHAP values to elucidate the factors influencing the model predictions, revealing insights into the importance of different features in TreeML models. The influence of adjacent trees and buildings emerged as a significant factor in both sample bar charts (Fig. 6). Moreover, expanding the evaluation to include eight additional tree species in Munich provided valuable insights into the generalisability of our models. While the TreeML model strongly predicted tree height and crown radius for some species, limitations were observed for others, such as *Prunus species*, *F. sylvatica*, and *B. pendula*. We assume that the evaluations fit some species because they are physiologically similar to the species we used for developing the model. However, this

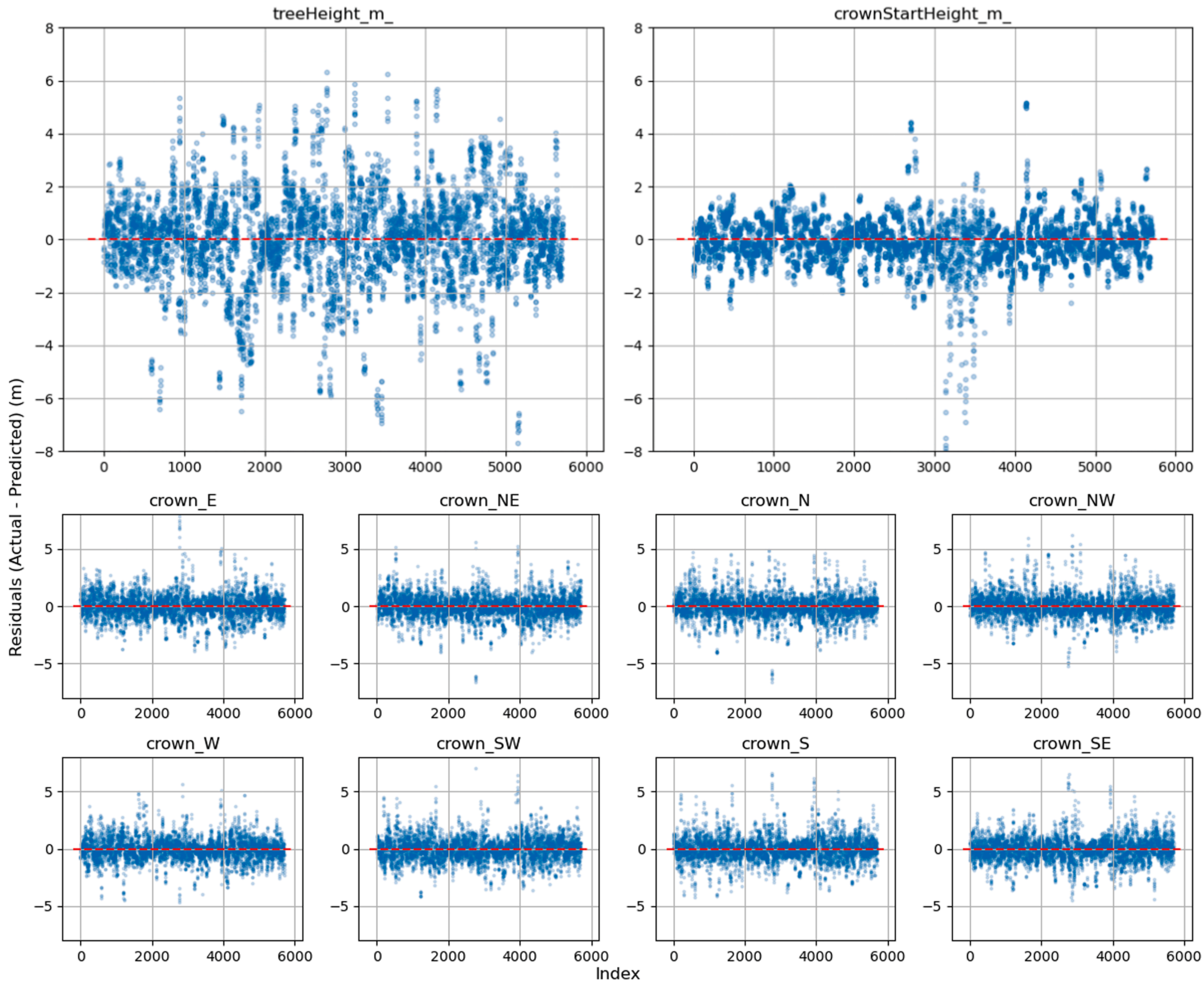

Fig. 4. Residual diagrams (Actual - Predicted) ($y_i - \hat{y}_i$) of the ten crown geometry prediction models of the Hist Gradient Boosting Regressor (HGBR) on the test dataset.

suggests a need for species-specific model development with more species to increase accuracy.

In the last step, further evaluation in Hamburg and Essen revealed varying performance across different urban environments. The models demonstrated acceptable performance on *T. europaea* and *T. cordata* trees in Hamburg, particularly in predicting crown radius, while exhibiting weaknesses in predicting tree height. Conversely, all models performed inadequately on *P. x hispanica* trees in Essen, underscoring the importance of considering regional variations and environmental factors in model training or the need for a large amount of test data.

Our study addresses significant gaps in understanding the interactions between trees and their urban context. While previous research has primarily focused on light competitions in forest stands, the effects of tree-to-tree and building-to-tree competitions in urban environments still need to be studied. The intricate relationship between trees and their surroundings, especially in the context of competition and tree crown structures, underscores the need for a comprehensive examination. The TreeML model should not be regarded as a tree growth prediction model. Instead, it predicts the crown geometry of tree species based on their DBH. The model uses DBH as an input variable, which is influenced by various tree growth factors such as climate, soil conditions, underground characteristics, and open surface area Rötzer et al. (2019). However, it has some limitations that should be considered in the usage and future development:

- The TreeML models are tailored to specific tree species. Although we obtained acceptable results for various species in Munich, this may not universally apply to all cases and diverse species.
- Factors influencing tree crown geometry prediction, such as the urban environment, city layout, climate, and location, can vary. The TreeML models were developed within the context of the city of Munich, and the study's outcomes cannot guarantee the model's applicability in other cities, even with the same tree species.
- The unique tree maintenance and management practices on Tree in Munich, or common issues such as powerlines and vandalism, influence the training of the TreeML models and could reduce its accuracy in a particular situation. It may not align with different urban tree care workflows in other locations. Retraining the models is recommended for different contexts.
- Urban tree manipulation, like pruning branches below 3–4 m for traffic and mobility needs, significantly impacts the accuracy of predicting crown start height, as evident in the evaluation metrics. Thus, due to high human intervention, the TreeML model is limited in predicting start height.

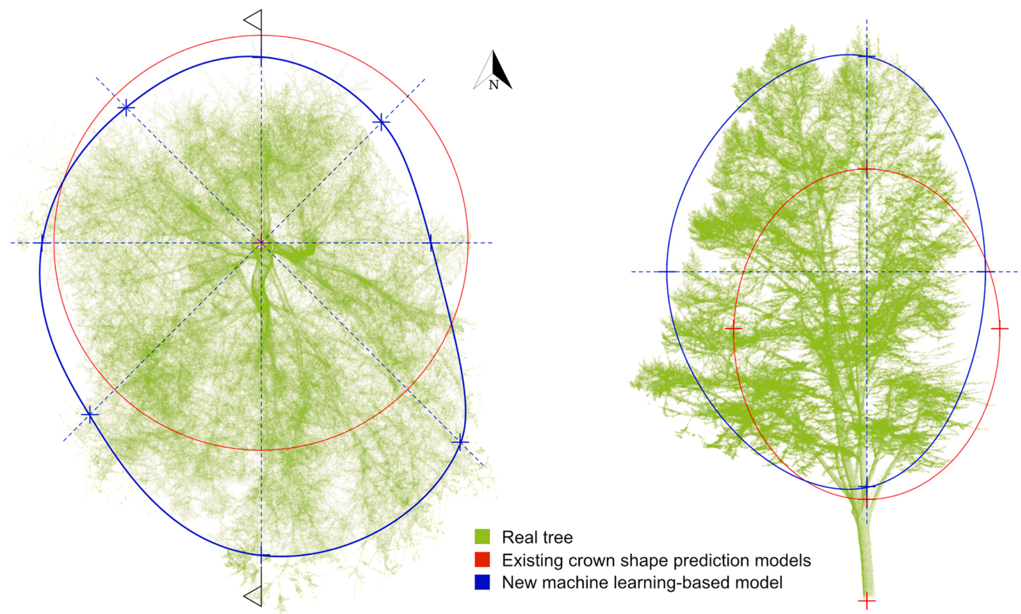

Fig. 5. A *Tilia cordata* tree from the test dataset is positioned near a building in the north direction, with two neighbouring trees in the east and west directions. The green point cloud represents the actual tree configuration, and the red line depicts the crown shape based on the allometric equations outlined in Table 1. In contrast, the blue line showcases the TreeML's predicted crown shape, visible in both the plan (left) and section view (right). The TreeML models predicted ten crown geometry points (marked in blue +): the crown start height, crown height, and the crown radius in eight geographical directions (as illustrated by the dashed line in the left figure). The crown shape is delineated by a curved line that must pass through these points in both the plan and section views.

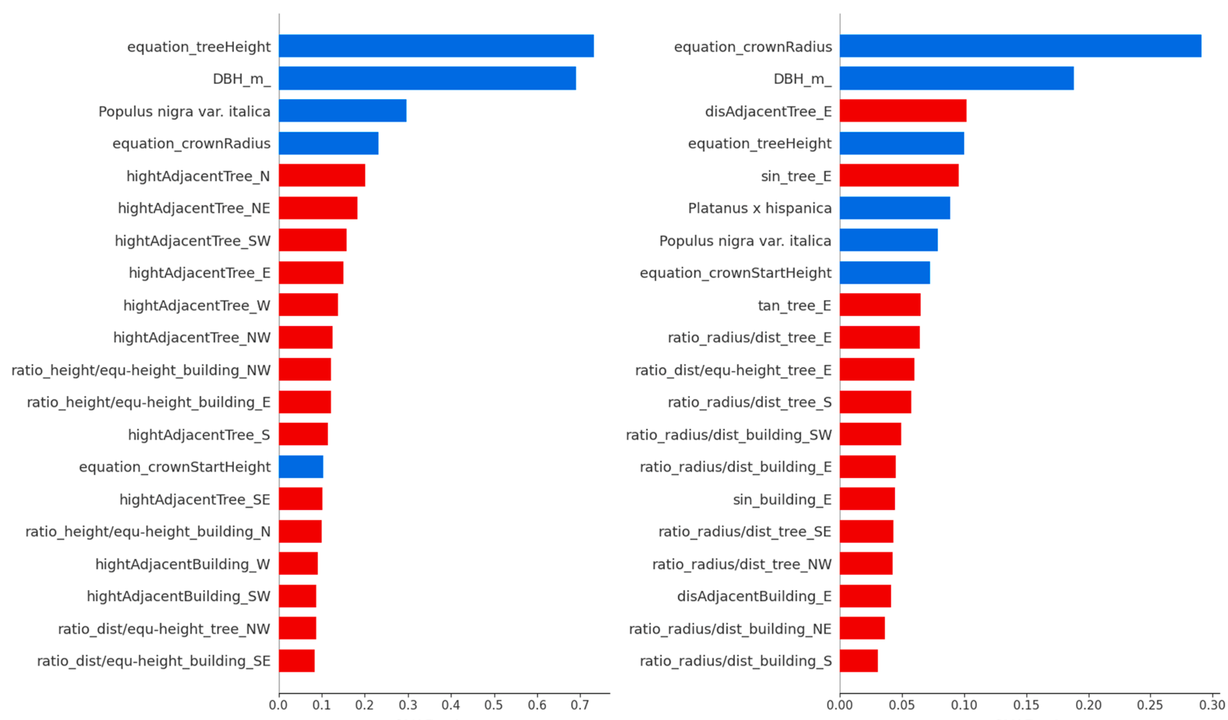

Fig. 6. The bar charts illustrate the importance of features in two sample models based on the SHAP values. The order of the SHAP values for the Tree Height prediction model is depicted in the left figure. The SHAP values, presented on the right, predict the tree crown radius in the east direction. These features are separated into two categories by colour: tree-dependent features (blue) and local environment-dependent features (red). See Table 2 for the list of features and their definitions.

- TreeML models rely on the tree's DBH. While estimating DBH for various tree species based on age is limited and inaccurate, TreeML is restricted to predicting crown geometry growth based on tree age.
- TreeML predicts rough crown geometry based on tree height, crown start height, crown radius in different directions, and the typical shape of the specific species (see Fig. 5). It is limited to providing an accurate 3D model of the crown shape.
- The TreeML model was developed based on a dataset of single-stem trees. Therefore, it needs to be validated on multi-stem trees for crown geometry prediction.

Table 6
the evaluation result (R^2-test) of the Tree height model for the eight new species by referencing each main species.

	R^2-**test** for each reference species for the **Tree height model**					
New species (No. of samples)	**Acer platanoides**	**Aesculus hippocastanum**	**Platanus x hispanica**	**Populus nigra**	**Robinia pseudoacacia**	**Tilia cordata**
Fraxinus excelsior (83)	**0.67**	0.67	0.68	0.67	0.69	0.67
Corylus colurna (80)	0.63	**0.63**	0.63	0.63	0.62	0.62
Prunus species (59)	-0.17	-0.26	-0.15	-0.15	-0.17	-0.14
Carpinus betulus (37)	**0.77**	0.75	**0.77**	0.77	0.76	0.77
Acer pseudoplatanus (31)	0.66	**0.65**	0.65	0.65	0.65	0.65
Sorbus aria (25)	**0.47**	0.45	0.47	0.47	0.48	0.49
Fagus sylvatica (11)	0.11	0.11	0.11	0.11	0.12	0.09
Betula pendula (10)	-1.49	-1.55	-1.50	-1.50	-1.64	-1.70

(**Bold:** best-performing models, Red: poor-performing models)

Table 7
The evaluation result (R^2-test) of the Crown radius_E model for the eight new species by referencing each main species.

	R^2-**test** for each reference species for the **Crown radius_E** model					
New species (No. of samples)	**Acer platanoides**	**Aesculus hippocastanum**	**Platanus x hispanica**	**Populus nigra**	**Robinia pseudoacacia**	**Tilia cordata**
Fraxinus excelsior (83)	**0.47**	0.46	0.46	**0.47**	**0.47**	**0.47**
Corylus colurna (80)	0.45	0.46	**0.47**	**0.47**	**0.47**	**0.47**
Prunus species (59)	0.10	-0.09	0.04	0.06	0.04	0.02
Carpinus betulus (37)	**0.62**	0.58	0.61	**0.62**	0.61	0.59
Acer pseudoplatanus (31)	0.62	**0.69**	0.67	0.67	0.67	**0.69**
Sorbus aria (25)	0.64	0.64	0.64	0.64	**0.65**	0.62
Fagus sylvatica (11)	0.34	0.37	**0.43**	0.37	0.38	0.37
Betula pendula (10)	0.73	**0.78**	0.77	0.76	**0.78**	0.77

(**Bold:** best-performing models, Red: poor-performing models)

Table 8
Evaluation of the TreeML models on 28 sample *Tilia europaea* and *Tilia cordata* trees in Hamburg.

	Evaluation metrics		
Models	R^2-**test**	**MAE**	**RMSE**
Tree Height (m)	0.36	1.73	2.01
Crown Start Height (m)	-0.13	0.61	0.81
crown_E (m)	0.49	0.80	1.02
crown_NE (m)	0.48	0.81	1.09
crown_NW (m)	0.55	0.84	1.07
crown_N (m)	0.58	0.74	0.95
crown_SE (m)	0.52	0.73	0.97
crown_SW (m)	0.59	0.82	1.02
crown_S (m)	0.57	0.85	1.07
crown_W (m)	0.53	0.71	0.93
Crown Diameter Max (m)	0.43	1.46	1.85
Crown Projection Area (m2)	0.59	16.1	21.41

(Red: poor-performing models)

Table 9
Evaluation of the TreeML models on 14 sample *Platanus x hispanica* trees in Essen.

	Evaluation metrics		
Models	R^2-test	MAE	RMSE
Tree Height (m)	-35.85	3.87	4.05
Crown Start Height (m)	-0.13	2.12	4.98
crown_E (m)	-0.23	3.32	3.77
crown_NE (m)	-0.68	3.45	3.87
crown_NW (m)	-0.23	2.54	3.36
crown_N (m)	0.22	1.95	2.64
crown_SE (m)	-0.03	1.69	2.55
crown_SW (m)	-0.24	2.96	3.49
crown_S (m)	0.18	1.64	2.41
crown_W (m)	-1.26	3.93	4.48
Crown Diameter Max (m)	-1.58	7.45	7.76
Crown Projection Area (m2)	-2.50	147.93	163.53

(Red: poor-performing models)

Despite the mentioned limitations, our findings have important implications for urban forestry management and planning. While the TreeML model shows promise in predicting the growth patterns of certain tree species in specific urban environments, caution is warranted when applying it to different species or locations. Future research endeavours should focus on refining models for specific species and exploring the integration of additional locational variables to enhance predictive accuracy.

5. Conclusions

In conclusion, this study provides significant insights into the complex dynamics of urban tree growth and crown geometry, addressing crucial gaps in our understanding of the interplay between trees, competition, and urban structures. Our research illuminates the nuanced factors influencing tree crown geometry development, considering not only species and DBH but also the direct surroundings, including light and space competition and adjacent buildings.

By posing key questions about the impacts of competition and urban structures on tree growth and crown geometry, this study charts a course for future studies. The findings presented here can contribute to the broader field of urban forestry, offering valuable insights for informed decision-making in city planning and the sustainable management of urban green spaces.

As cities evolve, grasping the intricate relationships between trees and their surroundings becomes increasingly crucial. The ability to predict canopy geometry with greater precision marks a significant advancement, particularly for the planning and design of urban green infrastructure. This newfound capability allows us to integrate factors such as tree competition and the intricate interplay between trees and buildings into planning and design decisions with unprecedented accuracy. Our research lays the groundwork for a more sustainable and resilient urban future, where the planting and management of urban trees are guided by a thorough understanding of their dynamic interactions in various urban environments.

Code Availability

The TreeML-SM script “TreeML-SM.py”, and the surrounding objects’ distance and height measurement “distance_measurment.py” are published in the GitHub repository (https://github.com/hadi-yazdi/TreeML-Data). Please refer to the Readme file in the Github repository for further information.

Declaration of generative AI in scientific writing

While preparing this work, the authors used ChatGPT 3.5 to improve the language and quality of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

Funding information

This work was carried out within the research training group Urban Green Infrastructure, funded by the German Research Foundation under grant 437788427 - RTG 2679.

CRediT authorship contribution statement

Thomas Rötzer: Writing – review & editing, Supervision, Methodology. **Astrid Moser-Reischl:** Writing – review & editing, Writing – original draft, Methodology. **Ferdinand Ludwig:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Frank Petzold:** Writing – review & editing, Supervision. **Hadi Yazdi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

Great credit goes to LRZ Compute Cloud for providing the cloud computing resources for this research. Moreover, we thank Josafat-Mattias Burmeister and Dr. Rico Richter from the University of Potsdam, Digital Engineering Faculty, for providing sample tree data from the cities of Hamburg and Essen for testing the model (section 3.3). And, we also thank the AllTerra Deutschland GmbH and the Department for Geoinformation, Surveying and Cadastre of the City of Essen for providing point cloud data from the cities of Hamburg and Essen (section 3.3). The authors thank Lumi Kirk for proofreading this paper. And the final great thanks to the RTG UGI organisers, Prof. Dr. Stephan Pauleit, who supports this study. This study would not have been possible without their dedication to scientific research and support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ufug.2024.128527. The bar chart of the SHAP value for all models (section 3.1) and the evaluation results for other species (section 3.2) are provided in the supplementary data.

References

Baptista, M.D., Livesley, S.J., Parmehr, E.G., Neave, M., Amati, M., 2018. Variation in leaf area density drives the rainfall storage capacity of individual urban tree species. Hydrol. Process. 32 (25), 3729–3740. ⟨https://eprints.whiterose.ac.uk/184483/⟩.

Berglihn, E.C., Gómez-Baggethun, E., 2021. Ecosystem services from urban forests: the case of Oslomarka, Norway. Ecosyst. Serv. 51, 101358. https://doi.org/10.1016/j.ecoser.2021.101358. ⟨https://www.sciencedirect.com/science/article/pii/S2212041621001169⟩.

Bragg, D.C., 2001. A local basal area adjustment for crown width prediction. North. J. Appl. For. 18 (1), 22–28. https://doi.org/10.1093/njaf/18.1.22.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623. https://doi.org/10.7717/peerj-cs.623. ⟨https://peerj.com/articles/cs-623⟩.

F. Chollet, Keras: The Python Deep Learning library, Astrophysics Source Code Library (2018)ascl:1806.022.⟨https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C/abstract⟩.

Clucas, B., Parker, I.D., Feldpausch-Parker, A.M., 2018. A systematic review of the relationship between urban agriculture and biodiversity. Urban Ecosyst. 21 (4), 635–643. https://doi.org/10.1007/s11252-018-0748-8.

Endreny, T.A., 2018. Strategically growing the urban forest will improve our world. Nat. Commun. 9 (1), 1160. https://doi.org/10.1038/s41467-018-03622-0. ⟨https://www.nature.com/articles/s41467-018-03622-0⟩.

FAO, Forests and Sustainable Cities: Inspiring stories from around the world, FAO, Rome, Italy, 2018.⟨https://www.fao.org/documents/card/en/c/I8838EN⟩.

Fauk, T., Schneider, P., 2023. Modeling urban tree growth as a part of the green infrastructure to estimate ecosystem services in urban planning. Front. Environ. Sci. 11. https://doi.org/10.3389/fenvs.2023.1090652.

Franceschi, E., Moser-Reischl, A., Rahman, M.A., Pauleit, S., Pretzsch, H., Rötzer, T., 2022. Crown shapes of urban trees-their dependences on tree species, tree age and local environment, and effects on ecosystem services. Forests 13 (5), 748. https://doi.org/10.3390/f13050748. ⟨https://www.mdpi.com/1999-4907/13/5/748⟩.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2. ⟨https://www.sciencedirect.com/science/article/pii/S0167947301000652⟩.

Gayathri, R., Rani, S.U., Čepová, L., Rajesh, M., Kalita, K., 2022. A comparative analysis of machine learning models in prediction of mortar compressive strength. Processes 10 (7), 1387. https://doi.org/10.3390/pr10071387. ⟨https://www.mdpi.com/2227-9717/10/7/1387⟩.

Gratani, L., Varone, L., 2006. Carbon sequestration by Quercus ilex L. and Quercus pubescens Willd. and their contribution to decreasing air temperature in Rome. Urban Ecosyst. 9 (1), 27–37. https://doi.org/10.1007/s11252-006-5527-2.

Kanjevac, B., Krstić, M., Babić, V., Govedar, Z., 2021. Regeneration dynamics and development of seedlings in sessile oak forests in relation to the light availability and competing vegetation. Forests 12 (4), 384. https://doi.org/10.3390/f12040384. ⟨https://www.mdpi.com/1999-4907/12/4/384⟩.

Korkou, M., Tarigan, A.K.M., Hanslin, H.M., 2023. The multifunctionality concept in urban green infrastructure planning: A systematic literature review, 9 Accepted: 2023-12-29T09:58:48Z. Elsevier Ltd. ⟨https://uis.brage.unit.no/uis-xmlui/handle/11250/3109082⟩, 10.1016/j.ufug.2023.127975.

Kroeger, T., McDonald, R.I., Boucher, T., Zhang, P., Wang, L., 2018. Where the people are: Current trends and future potential targeted investments in urban trees for PM10 and temperature mitigation in 27 U.S. Cities. Landsc. Urban Plan. 177, 227–240. https://doi.org/10.1016/j.landurbplan.2018.05.014. ⟨https://www.sciencedirect.com/science/article/pii/S0169204618303542⟩.

Li, F., Liu, X., Zhang, X., Zhao, D., Liu, H., Zhou, C., Wang, R., 2017. Urban ecological infrastructure: an integrated network for ecosystem services and sustainable urban systems. J. Clean. Prod. 163, S12–S18. https://doi.org/10.1016/j.jclepro.2016.02.079. ⟨https://www.sciencedirect.com/science/article/pii/S0959652616002560⟩.

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc.. In: ⟨https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html⟩, 10.48550/arXiv.1705.07874.

Maharana, K., Mondal, S., Nemade, B., 2022. A review: data pre-processing and data augmentation techniques. Glob. Transit. Proc. 3 (1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020. ⟨https://www.sciencedirect.com/science/article/pii/S2666285X22000565⟩.

McDonald, B.C., de Gouw, J.A., Gilman, J.B., Jathar, S.H., Akherati, A., Cappa, C.D., Jimenez, J.L., Lee-Taylor, J., Hayes, P.L., McKeen, S.A., Cui, Y.Y., Kim, S.-W., Gentner, D.R., Isaacman-VanWertz, G., Goldstein, A.H., Harley, R.A., Frost, G.J., Roberts, J.M., Ryerson, T.B., Trainer, M., 2018. Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Sci. (N. Y., N. Y.) 359 (6377), 760–764. https://doi.org/10.1126/science.aaq0524.

McPherson, E.G., Peper, P.J., 2012. Urban tree growth modeling. Arboric. Urban For. (AUF) 38 (5), 172–180. https://doi.org/10.48044/jauf.2012.026. ⟨https://auf.isa-arbor.com/content/38/5/172⟩.

McPherson, E.G., Berry, A.M., van Doorn, N.S., 2018. Performance testing to identify climate-ready trees. Urban For. Urban Green. 29, 28–39. https://doi.org/10.1016/j.ufug.2017.09.003. ⟨https://www.sciencedirect.com/science/article/pii/S1618866716305076⟩.

MEA, Millennium Ecosystem Assessment, Millennium Ecosystem Assessment. Living beyond Our Means: Natural Assets and Human Well-Being, Washington D.C, 2005.

Morakinyo, T.E., Lau, K.K.-L., Ren, C., Ng, E., 2018. Performance of Hong Kong's common trees species for outdoor temperature regulation, thermal comfort and energy saving. Build. Environ. 137, 157–170. https://doi.org/10.1016/j.buildenv.2018.04.012. ⟨https://www.sciencedirect.com/science/article/pii/S0360132318302166⟩.

Moser-Reischl, A., Rahman, M.A., Pauleit, S., Pretzsch, H., Rötzer, T., 2019. Growth patterns and effects of urban micro-climate on two physiologically contrasting urban tree species. Landsc. Urban Plan. 183, 88–99. https://doi.org/10.1016/j.landurbplan.2018.11.004. ⟨https://www.sciencedirect.com/science/article/pii/S0169204618313367⟩.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: in Python, in Python.2024.

Pretzsch, H., Matthew, C., Dieler, J., 2012. Allometry of Tree Crown Structure. Relevance for Space Occupation at the Individual Plant Level and for Self-Thinning at the Stand Level. In: Matyssek, R., Schnyder, H., Oßwald, W., Ernst, D., Munch, J.C., Pretzsch, H. (Eds.), Growth and Defence in Plants: Resource Allocation at Multiple Scales. Springer, Berlin, Heidelberg, pp. 287–310. https://doi.org/10.1007/978-3-642-30645-7_13.

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Rötzer, T., Caldentey, J., Koike, T., van Con, T., Chavanne, A., Seifert, T., Toit, B.D., Farnden, C., Pauleit, S., 2015. Crown size and growing space requirement of common tree species in urban centres, parks, and forests. Urban For. Urban Green. 14 (3), 466–479. https://doi.org/10.1016/j.ufug.2015.04.006. ⟨https://www.sciencedirect.com/science/article/pii/S1618866715000473⟩.

Pretzsch, H., 2022. Facilitation and competition reduction in tree species mixtures in Central Europe: consequences for growth modeling and forest management. Ecol. Model. 464, 109812. https://doi.org/10.1016/j.ecolmodel.2021.109812. ⟨https://www.sciencedirect.com/science/article/pii/S0304380021003574⟩.

Rötzer, T., Rahman, M.A., Moser-Reischl, A., Pauleit, S., Pretzsch, H., 2019. Process based simulation of tree growth and ecosystem services of urban trees under present and future climate conditions. Sci. Total Environ. 676, 651–664. https://doi.org/10.1016/j.scitotenv.2019.04.235. ⟨https://www.sciencedirect.com/science/article/pii/S0048969719317644⟩.

T. Rötzer, A. Reischl, M. Rahman, H. Pretzsch, S. Pauleit, Leitfaden zu Stadtbäumen in Bayern. Handlungsempfehlungen aus dem Projekt Stadtbäume - Wachstum, Umweltleistungen und Klimawandel. (2021). ⟨https://www.zsk.tum.de/fileadmin/w00bqp/www/PDFs/Leitfaeden/leitfaden_stadtbaeume_in_bayern_einzelseiten_web.pdf⟩.

Rahman, M.A., Pawijit, Y., Xu, C., Moser-Reischl, A., Pretzsch, H., Rötzer, T., Pauleit, S., 2023. A comparative analysis of urban forests for storm-water management. Sci. Rep. 13 (1), 1451. https://doi.org/10.1038/s41598-023-28629-6. ⟨https://www.nature.com/articles/s41598-023-28629-6⟩.

Ramyar, R., Ackerman, A., Johnston, D.M., 2021. Adapting cities for climate change through urban green infrastructure planning. Cities 117, 103316. https://doi.org/10.1016/j.cities.2021.103316. ⟨https://www.sciencedirect.com/science/article/pii/S026427512100216X⟩.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M., Lewis, P., 2013. Fast automatic precision tree models from terrestrial laser scanner data. Remote Sens. 5 (2), 491–520. https://doi.org/10.3390/rs5020491. ⟨https://www.mdpi.com/2072-4292/5/2/491⟩.

M.F. Shahidan, P.J. Jones, Plant Canopy Design in Modifying Urban Thermal Environment: Theory and Guidelines, Dublin, Ireland, 2008.⟨http://architecture.ucd.ie/Paul/PLEA2008/content/papers/oral/PLEA_FinalPaper_ref_179.pdf⟩.

S. Wright, Correlation and causation (1921).

Yang, J., Zhang, M., Zhang, J., Lu, H., Hauer, R.J., 2023. Allometric growth of common urban tree species in Qingdao City of Eastern China. Forests 14 (3), 472. https://doi.org/10.3390/f14030472. ⟨https://www.mdpi.com/1999-4907/14/3/472⟩.

Yazdi, H., Shu, Q., Ludwig, F., 2023. A target-driven tree planting and maintenance approach for next generation urban green infrastructure (UGI). Wichmann 8, 178–185. https://doi.org/10.14627/537740019.

Yazdi, H., Shu, Q., Rötzer, T., Petzold, F., Ludwig, F., 2024. A multilayered urban tree dataset of point clouds, quantitative structure and graph models. Sci. Data 11 (1), 28. https://doi.org/10.1038/s41597-023-02873-x. ⟨https://www.nature.com/articles/s41597-023-02873-x⟩.

Young, R.F., 2010. Managing municipal green space for ecosystem services. Urban For. Urban Green. 9 (4), 313–321. https://doi.org/10.1016/j.ufug.2010.06.007. ⟨https://www.sciencedirect.com/science/article/pii/S1618866710000543⟩.

Zhang, X., Lei, Y., Li, R., Ackerman, A., Guo, N., Li, Y., Yang, Q., Liu, Y., 2022. Research on thermal comfort of underside of street tree based on LiDAR point cloud model. Forests 13 (7), 1086. https://doi.org/10.3390/f13071086. ⟨https://www.mdpi.com/1999-4907/13/7/1086⟩.

Zhu, Z., Kleinn, C., Nölke, N., 2021. Assessing tree crown volume—a review. For.: Int. J. For. Res. 94 (1), 18–35. https://doi.org/10.1093/forestry/cpaa037.