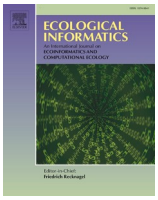
Ecological Informatics 88 (2025) 103140

Contents lists available at ScienceDirect

Ecological Informatics

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



Does machine learning outperform logistic regression in predicting individual tree mortality?

Aitor Vázquez-Veloso [a,*], Astor Toraño Caicoya [b], Felipe Bravo [a], Peter Biber [b], Enno Uhl [b,c], Hans Pretzsch [b]

[a] SMART Ecosystems Group. Departamento de Producción Vegetal y Recursos Forestales. Instituto Universitario de Investigación en Gestión Forestal Sostenible (iuFOR), ETS Ingenierías Agrarias, Universidad de Valladolid, Palencia, Spain
[b] Chair for Forest Growth and Yield Science, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany
[c] Bavarian State Institute of Forestry, Hans-Carl-von-Carlowitz-Platz 1, 85354 Freising, Germany

ARTICLE INFO

Keywords:
Norway spruce
Survival
Artificial intelligence
Supervised learning
Random Forest
Empirical mortality models

ABSTRACT

Tree mortality is a crucial process in forest dynamics and a key component of forest growth models and simulators. Factors like competition, drought, and pathogens drive tree mortality, but the underlying mechanism is challenging to model. The current environmental changes are even complicating model approaches as they influence and alter all the factors involving mortality. However, innovative classification algorithms can go deep into data to find patterns that can model or even explain their relationship. We use Logistic binomial Regression as the reference algorithm for predicting individual tree mortality. However, different machine learning (ML) alternatives already applied to other forest modeling topics can be used for this purpose. Here, we compare the performance of five different ML algorithms (Decision Trees, Random Forest, Naive Bayes, K-Nearest Neighbour, and Support Vector Machine) against Logistic binomial Regression in individual tree mortality classification under 40 different case studies and a cross-validation case study. The data used corresponds to Norway spruce long-term experimental plots, which have a total of 75,522 tree records and a 10.28 % mortality rate on average. Through different case studies, when more variables were used, general performance improved as expected, while more extensive datasets decreased the performance level of the algorithms. Performance was also higher when plots remained without management compared to thinned ones. Random Forest outperformed the other algorithms in all the cases except cross-validation, where it was the weaker one. Our results demonstrate the potential of ML in assessing tree mortality. When the model application is not clearly defined and/or model interpretability is needed, Logistic binomial Regression is still the best tool for evaluating individual tree mortality.

1. Introduction

Forest modeling has undergone continuous development during the last decades in terms of system abstraction, objectives, and methods. Born aimed to estimate timber yields over time, its evolution until our times supposed an increase in model complexity (Shifley et al., 2017) for a wide variety of processes and situations (Pretzsch, 2009), covering local and regional levels (Bravo et al., 2011). This increase in complexity created an opportunity for forest management simulators (Bravo et al., 2025; Mladenoff, 2004; Pretzsch et al., 2002), reducing the efforts and specialized knowledge required when running models. Despite this evolution, an already unsolved weakness of European forest models is the estimation of single-tree mortality (Bugmann et al., 2019). Being a critical point in all single-tree-based models (Boeck et al., 2014), mortality-associated patterns and processes remain poorly understood (Hülsmann et al., 2016). Tree mortality is difficult to associate with a

Abbreviations: ANN, Artificial Neural Network; bai, basal area increment; BAL, basal area larger than the subject tree; dbh, tree diameter at breast height (1.30 m); DT, Decision Trees; G, stand basal area; h, tree height; KNN, K-Nearest Neighbour; LR, Logistic binomial Regression; M, Martonne Aridity Index; ML, Machine Learning; NB, Naïve Bayes; RF, Random Forest; SDI, Stand Density Index; SI, Site Index; SPEI, Standardized Precipitation Evapotranspiration Index; SVM, Support Vector Machine; TI, time elapsed.

* Corresponding author.
E-mail address: aitor.vazquez.veloso@uva.es (A. Vázquez-Veloso).

https://doi.org/10.1016/j.ecoinf.2025.103140
Received 28 May 2024; Received in revised form 5 April 2025; Accepted 5 April 2025
Available online 12 April 2025

single factor (Wang et al., 2012), as different agents alter tree vitality, like competition, pathogens, snow, fires and frosts, and droughts (Waring, 1987). This situation increases mortality unpredictability, which is even more complicated when including future climate conditions (McDowell et al., 2020). For instance, while a clear relationship between competition and mortality (Bravo-Oviedo et al., 2006; Monserud and Sterba, 1999; Pretzsch et al., 2023b) as well as drought and mortality are already known (Senf et al., 2020), prediction uncertainty increases (Bugmann et al., 2019). Inconsistent patterns appear (Thrippleton et al., 2021) under changing climate conditions. In addition, mortality is also affected by spontaneous patterns that cannot be easily tracked, like wildfires or pest infections. In this situation, algorithms that can go deeper in finding patterns in data can improve the quality of mortality models.

In this context, Logistic binomial Regression (LR) has emerged as the prevailing approach among scientists to predict individual tree mortality (Bravo et al., 2001; Bravo-Oviedo et al., 2006; Hülsmann et al., 2016; Hülsmann et al., 2017; Pascual et al., 2022; Pretzsch et al., 2020; Salas-Eljatib and Weiskittel, 2020; Shearman et al., 2019a, 2019b; Shifley et al., 2017), allowing model interpretation and requiring a reasonably low computing power. Currently, computing capacities have expanded exponentially and are more accessible (Shifley et al., 2017), providing the chance to use higher computational demanding alternatives to LR for predicting tree mortality. Some Machine Learning (ML) algorithms are examples of higher computationally demanding alternatives, especially those belonging to the Deep Learning algorithms family, such as Artificial Neural Networks (ANN). ANN has already been used to model tree mortality (da Rocha et al., 2018a, 2018b; Merkl and Hasenauer, 1998; Reis et al., 2018) but returned questionable results when aimed at predicting death trees (da Rocha et al., 2018a, 2018b; Reis et al., 2018). However, Random Forest (RF) has been tested with particular success (McNellis et al., 2021; Shearman et al., 2019b). Although ML algorithms are alternative tools to LR for predicting tree mortality and disentangling the unknown patterns behind the data (Hülsmann et al., 2016), a comparison is needed to assess their efficacy against LR, the reference methodology in this field.

Due to its forestry relevance in Central Europe in terms of wood supply (von Teuffel et al., 2004) and its endangered status in the climate change context (Schlyter et al., 2006), this study was focused on predicting Norway spruce (*Picea abies* (L.) H. Karst) mortality. Using data from Norway spruce long-term experimental plots from Bavaria (Germany), this work compares the individual tree mortality prediction of the mainly used ML classification algorithms under a broad set of study cases. Given this context, this work aims to develop user guidelines for fitting mortality models under different conditions, from using local datasets to using a National Forest Inventory, prioritizing model transferability and/or accuracy. On the technical side, the comparison was carried out under different dataset sizes and the number of variables used to train models, thus imitating different technical situations faced by modelers according to the data available to use. On the biological side, the comparison was developed by splitting the dataset into different groups regarding thinning degrees (control, from above, from below) and inventory record length, aimed to imitate different modeler situations and understand if different dataset structures can affect the model performance due to the patterns found on data. In addition, model performance can be altered depending on the variability of data used to fit and apply models, as Hülsmann et al. (2017) found. To assess that, extra case studies were developed to track model applications under different thinning degrees, thus fitting models and applying them to different data structures. All comparisons were made based on each algorithm's performance, and the training time spent for each algorithm was recorded to show the different computing demands. With this aim, we hypothesized that: HI) more extensive datasets and more variables available will improve all algorithms' performance, where ML will outperform LR when using more information; HII) model performance is similar under different thinning degrees and using different datasets in the cross-validation; HIII) longer inventory records improve model performance.

2. Methods

2.1. Data

Data from long-term experimental plot networks in Bavaria (Germany) were used in this study to predict individual tree mortality. We selected pure stands of Norway spruce under different initial density conditions and thinning experiments, all of them planted with different research purposes but covering similar site quality conditions (Fig. 1). Different densities and silvicultural regimes provide us with an extensive range of local competition to assess tree mortality, while thinning degrees provide different structures on our datasets (Fig. 2).

While some plots included in that study have been measured since 1880, inventory data from different plots were selected from 1975 to the present due to the climatic data available (Table 1), establishing a common study period for all the experiments. As explained later, the entire database was divided into different information availability levels according to each case study. Information about diameter at breast height (dbh) and tree position is available, the last one based on the plantation spacing stablished on each experimental plot. Stand age was assumed to be equal for all the trees in the plot as all the experiments are even-aged stands, and its age was calculated based on the time elapsed since its establishment. Basal area increment (bai) was calculated for each tree and the period between measurements using dbh and age. The main stand characteristics for each plot are detailed in Table 2. For further information on the long-term experiments, see Pretzsch et al. (2023a, 2023b).

Tree height (h) was only recorded for a portion of trees of each size class on each inventory measurement. A log function was fitted (eq. 1) to assess the height-diameter relationship for each plot and tree age, estimating the height of the rest of the trees.

$$h = a + b \cdot ln(dbh) \quad (1)$$

where:

- h is the total tree height (m).
- dbh it the tree diameter at breast height (cm).
- a and b are empirical parameters calculated for each plot and age.

For all the trees, neighborhood variables were calculated by creating virtual subplots around them, considering trees inside the subplot as their main competitors. Each subplot radii were calculated as a third part of the subject tree height, considering competition dependent on the tree size. For those subplots closest to the border, a weight value (w) was also calculated to supply the lack of information on the outside plot area, assuming homogeneity with the trees inside the subplot and applying the weight value directly to the stand estimated variables (Fig. 3) as shown in eq. 2.

$$Nw = N \cdot w, \quad (2)$$

where:

- Nw is the subplot weighted density (trees/ha).
- N is the subplot estimated density (trees/ha).
- w is the weight value estimated according to the subplot area with and without information.

On each subplot, local variables concerning competition (N, BAL (Wykoff, 1990), Hegyi index (Hegyi, 1974), G local (Steneker and Jarvis, 1963) and SDI (Reineke, 1933) adapted for Norway spruce (Pretzsch and Biber, 2005)), growth (bai), time elapsed (TI), and tree social position (dbh/dg (Lynch et al., 1999) and h/hm (Alenius et al., 2003)) were estimated as detailed in Table 3. Trees with dbh under 5 cm (trees from regeneration) were not included in our analysis, and trees from different species were just considered to estimate subplot variables, not as subject trees. Thus, 75,522 tree records were available as subject trees, while it

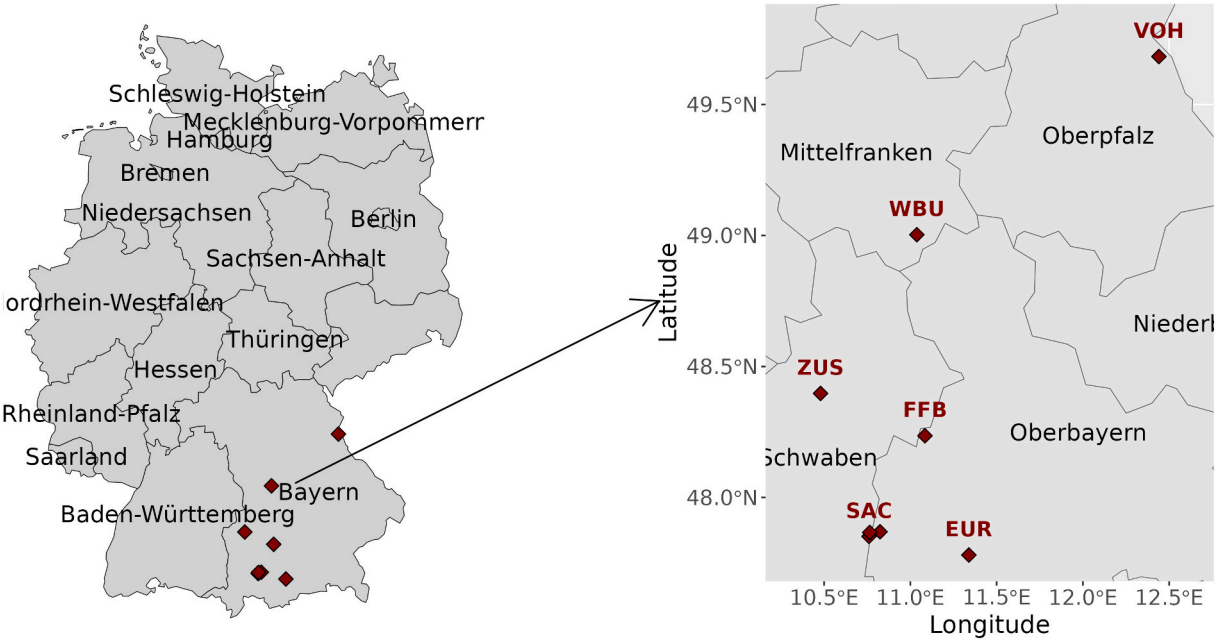


Fig. 1. Norway spruce long-term experimental plots located in Germany (left) and detailed location in Bavaria (right).

was reduced to consider time-elapsed variables between measurements. Finally, 56,628 tree records were studied, including the different measurements of each one in consecutive inventories, with a total of 5824 dead tree records (10.28 %).

Site quality, time elapsed between inventory editions, and climate variables were estimated at plot level due to their homogeneity for each subplot. Productivity was studied through the Site Index (SI), calculated based on the plot dominant height at 100 years using the Bavarian Norway spruce yield tables (Assmann and Franz, 1972). Based on previous work (Caicoya and Pretzsch, 2021), SI was dynamically estimated each year, reflecting site quality changes over time. Time elapsed between inventory editions was calculated using the time between measurements, and the annual bai of each tree and plot was calculated linearly. Historical monthly climate data were taken from CRU-TS 4.06 (Harris et al., 2020) and downscaled with WorldClim 2.1 (Fick and Hijmans, 2017). Thus, average minimum and maximum temperature (°C) and total precipitation (mm) were downloaded monthly for the period 1975–2021 with 2.5 min of spatial resolution. Local data for each experimental site were extracted, and mean temperature (°C) and total precipitation (mm) were calculated each year when measurements were developed. De Martonne Aridity Index (M) (Martonne, 1926) and the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) were estimated, the second one for six months before August. Both indexes were calculated as described by their authors to consider the climate in our study. A categorical variable representing the experiment was not included, as it is indirectly captured through the previous experiment-level covariates. Data curation and all the calculations were developed using R (R Core Team, 2021).

2.2. Algorithms

Logistic binary Regression (LR) was the reference used in this study, as it is a traditional methodology used in tree mortality classification (Bravo-Oviedo et al., 2006; Merkl and Hasenauer, 1998; Shearman et al., 2019a). As Machine Learning (ML) alternatives, five different algorithms were applied in our study, selected based on their popularity and application in different forestry topics. Decision Trees (DT) is a non-parametric supervised learning method that can be used for regression and classification. It tries to predict simple decision rules inferred from data variables to create a model, and it is easy to use but can be unstable due to data variability. Random Forest (RF) is an algorithm that fits several DT on various sub-samples and averages them to improve model accuracy. It can reach better performance levels than DT but is more computing demanding. Naive Bayes (NB) is a supervised learning algorithm based on the application of Bayes' theorem, which assumes a "naive" conditional independence between every pair of variables. It alleviates problems related to the model dimensionality while often failing to produce good estimations for the correct class probability. K-nearest neighbour (KNN) is another supervised model used for classification and regression tasks. In the classification context, KNN predicts the class of a data point based on a majority vote of its k-nearest neighbors in the feature space. For regression, it predicts the value by averaging the values of its k-nearest neighbors. It is easy to use without assumptions but can struggle with imbalanced datasets. Support Vector Machine (SVM) algorithms are used for classification, regression, and outlier detection. They are effective in high-dimensional spaces but have difficulties dealing with big datasets. Artificial Neural Networks (ANN) were initially considered but finally excluded due to the complexity of adapting this algorithm for each case study. All the analyses were conducted using different case studies (detailed in the next section) under a supervised learning approach, as records of dead and alive trees were part of the initial dataset. Analysis was developed on R (R Core Team, 2021) using the caret library (Kuhn, 2008), as it facilitates the model parameters optimization and selects the model through repetitive cross-validation procedures. Through the caret library, LR was performed using the "glm" method; DT using the "rpart" method; RF using the "rf" method; NB using the "naive_bayes" method; KNN using the "knn" method, estimating k value using the square root of the train dataset length; and SVM using "svmLinearWeights2" method. Due to the excessive computational resources required, all the analyses were run on a process server hosted at iuFOR-University of Valladolid, provided with 80 cores and 256 GB RAM.

Algorithms were compared based on the best model performance. Resources needed for training models in terms of time are available in Appendix A as a comparison, as they all were run with the same system. Different metrics were estimated for each model and algorithm to compare their performance, thus prediction accuracy (ACC), also split into alive (ACC_{alive}) and dead (ACC_{dead}) trees prediction accuracy; area under the precision/recall curve (AUPRC) (Saito and Rehmsmeier, 2015), Matthews Correlation Coefficient (MCC) (Matthews, 1975) and

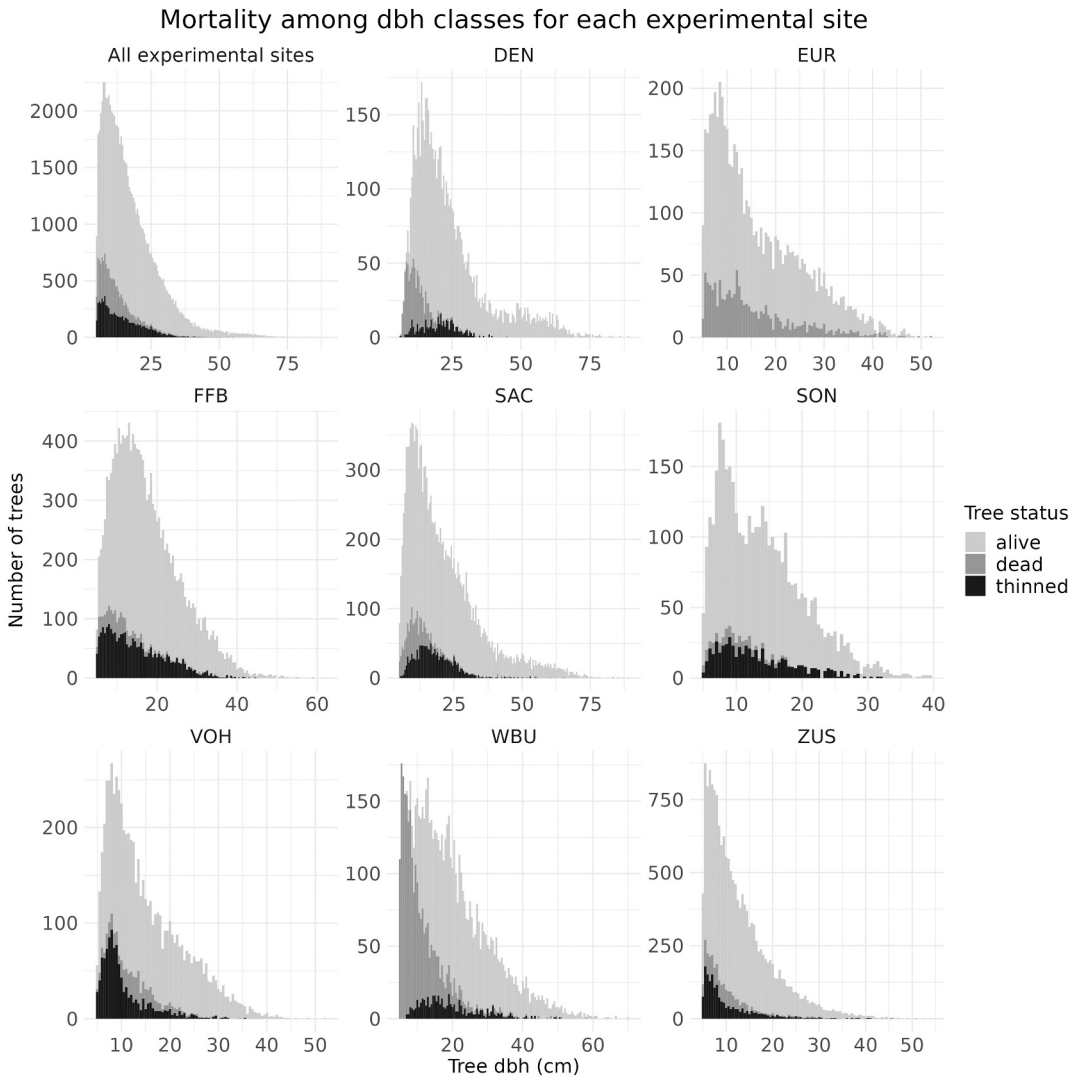


Fig. 2. Alive, dead, and thinned trees are distributed among dbh classes per each experimental site. The first box (up-left) includes all the information about the experimental site.

Cohen's Kappa coefficient (K) (Cohen, 1960). MCC was used in this work to compare algorithms in a single value, as is the one recommended for imbalanced datasets (Chicco, 2017; Chicco and Jurman, 2020). However, the remaining metrics already mentioned can be consulted in the "Data availability" section. These metrics were calculated using the R package ROCR (Sing et al., 2005) and irr (Gamer et al., 2019) in the case of Cohen's Kappa coefficient. Besides metrics, training resources were assessed through the time required to perform the training process per model and algorithm, all conducted using the same HPC service for homogeneity.

A standard workflow was implemented across algorithms and case studies (Fig. 4). First, the data was loaded, the first algorithm was selected, and a timer was activated. For each model proposal (all variable combinations proposed for model fitting), the data was normalized (except the independent variable) and randomly split into train (80 %) and test (20 %) datasets. At this point, another timer was activated, and the model was fitted. Resulting predictions were received as a probability of dead (1) or alive (0) while real inventory data takes two possible values: 1 (dead) or 0 (alive). To deal with that, a threshold value was selected, maximizing the success of predictions, thus testing all the possibilities from 0 and 1 (3 decimals) and choosing the value which maximized the MCC metric. The threshold value was used to convert the probability from predictions to a binary value, classifying trees as dead (1) or alive (0). After that, metrics for comparing models and algorithms were calculated, and the second timer stopped, recording the time needed per model. At the end of this process, a new model proposal was selected. Once the process had run through all model proposals, the first timer was stopped (recording the time spent developing all the previous processes), all output results were saved, and the next algorithm was selected. When no more algorithms were available, the process finished.

This methodology was applied to all the case studies detailed in the following section. For all the models obtained from each algorithm and case study, the best model was selected based on the higher MCC value.

2.3. Case studies description

2.3.1. Data-based case studies

These case studies aimed at representing different data availability configurations of users when fitting tree mortality models. We selected the size of the dataset and the number of available variables as primary drivers, studying both independently and combined.

The first analysis involved the size of the dataset. After data curation, the original dataset had a total of 75,522 records. As variables concerning time elapsed between measurements were considered and could not be estimated on the first plot record, the amount of data available for this study was 56,628, thus losing each tree's first record information.

Table 1
Research experiment details and mean climate variables. The coordinates column shows the longitude and latitude in degrees, minutes and seconds using the WGS84 datum. Climate data is expressed as the range of time elapsed between the inventories used from each research experiment.

Research experiment	Abbreviation	Number of plots	Coordinates	Total annual rainfall [mm]			Mean annual temperature [°C]		
				Min	Mean	Max	Min	Mean	Max
Denklingen 05	DEN 05	3	10° 50′ 32″ E 47° 52′ 15” N	1036	1078	1119	6.6	6.9	7.2
Denklingen 606	DEN 606	6	10° 49′ 26″ E 47° 51′ 36” N	903	1066	1181	7.6	8.0	8.8
Eurach 605	EUR 605	7	11° 20′ 20″ E 47° 46′ 48” N	1012	1108	1296	7.7	8.6	9.9
Fürstenfeldbruck 612	FFB 612	21	11° 05′ 04″ E 48° 14′ 07” N	834	894	993	7.3	8.2	9.0
Sachsenried 602	SAC 602	4	10° 45′ 37″ E 47° 51′ 06” N	916	1132	1200	7.5	8.0	8.6
Sachsenried 607	SAC 607	12	47° 52′ 01″ E 10° 49′ 23” N	1088	1223	1348	7.5	7.6	7.8
Sachsenried 67	SAC 67	3	10° 45′ 13″ E 47° 50′ 03” N	1155	1207	1262	6.5	7.1	7.9
Sachsenried 68	SAC 68	3	10° 45′ 18″ E 47° 50′ 03” N	1155	1207	1262	6.5	7.1	7.9
Schongau 623	SON 623	7	10° 45′ 54″ E 47° 51′ 58” N	1044	1239	1487	7.2	7.7	8.7
Vohenstrauß 622	VOH 622	9	12° 26′ 26″ E 49° 40′ 59” N	696	859	1057	6.7	7.3	8.5
Weißenburg 613	WBU 613	7	11° 02′ 16″ E 49° 00′ 11” N	694	832	1038	6.9	7.8	8.8
Zusmarshausen 603	ZUS 603	6	48° 23′ 49″ E 10° 28′ 46” N	843	886	933	7.0	8.2	9.4
Zusmarshausen 604	ZUS 604	4	48° 23′ 56″ E 10° 28′ 59” N	843	892	933	7.0	8.4	9.4

Table 2
Structural characteristics for each research experiment were calculated from the single tree subplots, using 1/3 of the subject tree's height as a subplot radii. SDI denotes the stand density index, and dg is the quadratic mean diameter.

Research experiment	Stand density [tree/ha]			SDI [adm]			Age [years]			dg [cm]			Acquisition year	
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	First	Last
DEN 05	10.0	27.2	57.0	709	1654	3304	138	140	143	41.1	51.2	69.6	1985	1991
DEN 606	4.0	35.1	102.0	282	1647	4699	45	52	62	13.7	22.0	37.8	1998	2015
EUR 605	1.0	14.2	59.0	62	1031	3461	21	36	63	5.0	19.1	42.9	1973	2015
FFB 612	1.0	17.2	74.4	48	1209	3575	27	35	48	5.2	20.1	53.1	1996	2017
SAC 602	5.0	27.9	92.0	302	1672	4136	27	35	51	8.7	15.8	34.3	1989	2013
SAC 607	2.0	34.1	87.6	291	1444	4640	48	53	60	14.9	25.5	40.3	2001	2013
SAC 67	7.7	22.6	55.2	571	1489	3271	126	128	131	42.9	53.8	66.4	1985	1990
SAC 68	1.9	29.0	58.7	57	1536	2778	125	127	130	27.6	49.1	65.6	1985	1990
SON 623	3.0	26.0	71.9	246	1535	3827	27	33	43	8.4	16.6	33.0	1999	2015
VOH 622	1.0	8.0	28.9	27	667	3785	23	30	44	5.2	17.7	42.0	1997	2018
WBU 613	3.0	29.2	104.0	184	1506	4680	55	70	100	11.7	25.5	53.3	1982	2016
ZUS 603	1.0	29.4	118.0	66	1381	4537	17	31	54	5.0	13.9	46.5	1980	2017
ZUS 604	2.0	22.9	87.0	153	1290	4425	21	33	54	6.8	14.6	34.0	1985	2017

Then, the original dataset was randomly split into three different category sizes, hence considering different user data availability situations: small, medium, and big datasets, detailed in Table 4.

The second analysis involved the number of variables proposed for the model. For this, we grouped variables by topics (tree size, productivity, competition, growth, inventory time elapsed, climate, and social position), and different model proposals were made based on the number of variables we wanted to implement in the model, thus representing different situations of data availability when fitting models. Model proposals (with various numbers of variables) were developed by combining one variable from each group to avoid correlation. Variables for model proposals were studied before variable selection to prevent correlation. Variables concerning tree size, growth/time, and climate topics were not correlated so that they could be included separately or grouped in model proposals. Different case studies were proposed by selecting the number of variables available to fit the model. Thus, each case study comprises different model proposals generated by combining variables among topics (I = 10 proposals; II = 30 proposals, III = 90 proposals; IV = 180 proposals) detailed in Table 3.

A third analysis was conducted by crossing both analyses. Thus, each dataset size and variable case study were studied together, presenting 12 different case studies through all the possible combinations (3 dataset sizes x 4 variable groups).

2.3.2. *Mortality under different thinning degrees and inventory records length*

Forestry-related aspects of the data were also considered in additional case studies. Thinning degrees and inventory record lengths were used to represent them.

The used dataset includes three different thinning degrees for each plot (Table 4): unthinned (control), thinning from above (above) and thinning from below (below). As each situation represents a different

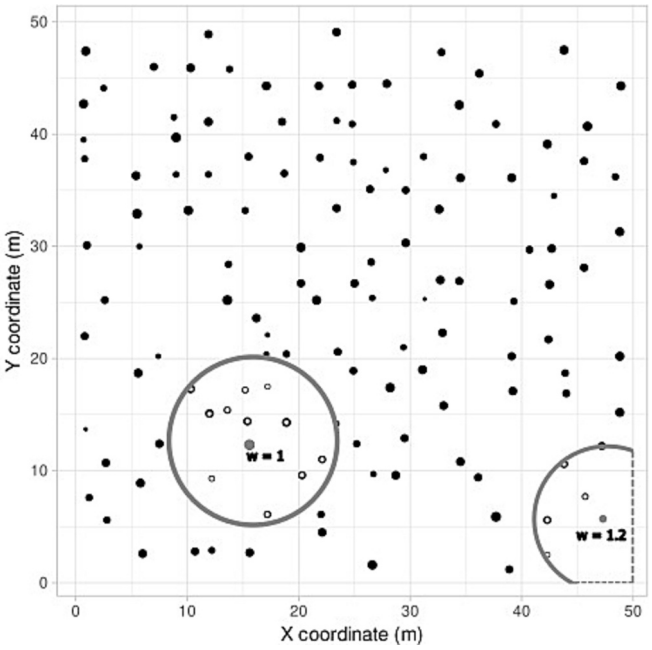


Fig. 3. Example of experimental design using subplots. A circle with a size related to the tree height of the subject tree was used to calculate its stand neighborhood, applying the weight value (w) on their estimation when no information is available due to a close border.

Table 3
Variables used to fit the models by topic and case studies where they were included: dbh is the diameter at breast height; slenderness is the ratio h/dbh; SI is the Site Index; BAL is the basal area larger than the subject tree; G local is the local stand basal area; N is the stand density; SDI is the stand density index; bai is the annual basal area increment; TI is the time elapsed; M is the Martonne Aridity Index; SPEI is the Standardized Precipitation Evapotranspiration Index; dbh/dg is the ratio of the tree diameter at breast height and the stand quadratic mean diameter; h/hm is the ratio of the tree height and the stand mean height.

Topic	Variables	Case study
Tree size	dbh	I, II, III,
	slenderness	IV
Site productivity (plot)	SI (Assmann and Franz, 1972)	I, II, III,
		IV
Tree competition (subplot)	BAL (Wykoff, 1990)	I, II, III,
	Hegyi index (Hegyi, 1974)	IV
	G local (Steneker and Jarvis, 1963)	
	N	
	SDI (Reineke, 1933) adapted for Norway spruce (Pretzsch and Biber, 2005)	
Tree growth and inventory time elapsed (plot)	bai	II, III,
	TI	IV
Climate (plot)	M (Martonne, 1926)	III, IV
	SPEI (Vicente-Serrano et al., 2010)	
Tree social position (subplot)	dbh/dg (Lynch et al., 1999)	IV
	h/hm (Alenius et al., 2003)	

stand management and forest structure affecting mortality patterns, we split the original dataset by thinning groups to assess their differences. Additionally, all possible combinations with each variable case study were carried out. This resulted in 12 case studies (3 thinning degrees x 4 variable groups). A second analysis was carried out by grouping 2 thinning degrees and discarding the third, studying each variable case study combination too and resulting in 12 additional case studies (3 thinning degrees combination x 4 variable groups).

The length of inventory records of each plot was also used to understand the amount of time that the same plot was measured. It aimed to determine if different algorithms could take advantage of more

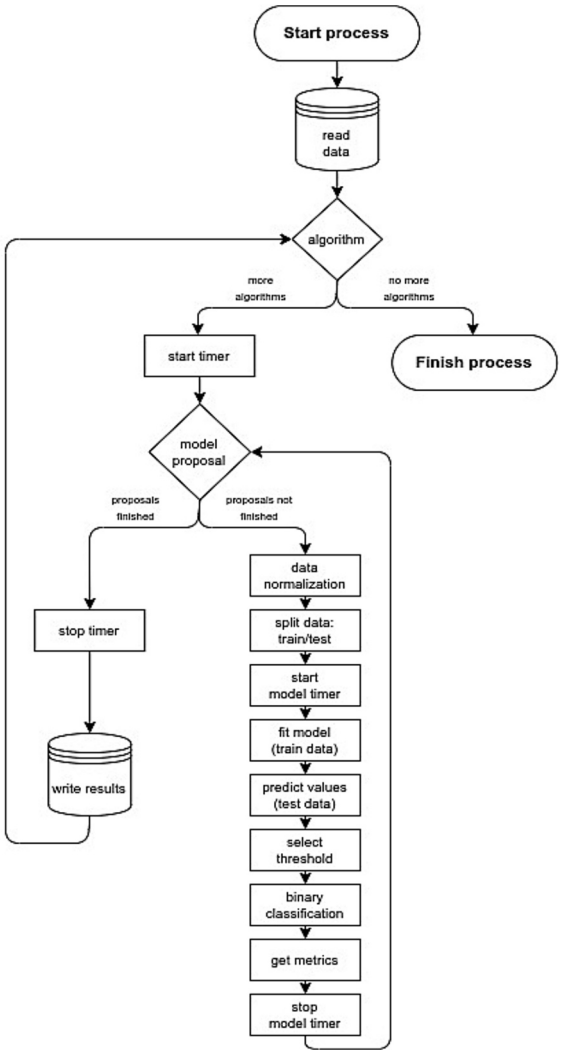


Fig. 4. Flowchart summarizing the model fitting process for each algorithm.

Table 4
Final dataset sizes for each data-based and forest-based case study. The first measurement of each plot was removed from each dataset due to the lack of variables concerning time, so values on that table exclude initial measurements.

Case study		Dataset size (n)	Dead trees (n)	Dead trees proportion (%)
Dataset size	*small*	2830	299	10.66
	medium	19,983	2099	10.50
	big	56,628	5824	10.28
Thinning degrees	*control*	11,380	1663	14.61
	thinning from above	34,319	2826	8.23
	thinning from below	10,929	1355	12.21
Inventory records length	*3 to 5*	12,644	1412	11.17
	6	15,183	883	5.81
	7	11,063	1555	14.06
	8 to 9	17,738	1974	11.12

information available in the same inventory. Thus, the original dataset was split into four groups (Table 4) where the size of the original dataset was similar: plots with three to five records, six records, seven records, and eight to nine records. Those groups were combined with each variable case study, conducting 16 case studies (4 inventory records length x 4 variable groups).

2.3.3. Cross-validation among thinning degrees

While previous case studies used different data subsamples to compare algorithms, this case study wants to test the performance of each algorithm when training and testing datasets are different, simulating a typical situation when a user applies an already existing model to its data. This approach explores the model performance deviations caused by different data origins and structures, quantifying the loss of predicting accuracy. With that in mind, we selected the best model per algorithm fitted for each previous thinning degree case study. In all cases, the number of variables used corresponds to case III for homogeneity. Thus, models fitted on just one type of thinning (i.e., thinning from below) were applied to the three datasets grouped by thinning degree (control, from below, from above), having three different results by model. A second analysis was carried out by grouping 2 thinning degrees on the fitting process (i.e., control + thinning from below) and applied to the three datasets grouped by thinning degree (control, from below, from above) as in the previous case. Performance assessed by MCC was used to compare them among all the possible combinations.

3. Results

3.1. Effect of dataset size and variable number (HI)

Case studies involving different dataset sizes showed various levels of performance on each algorithm, as shown in Fig. 5A. A clear trend is shown among algorithms (except for KNN): performance was lower when more data was provided. In all the cases, RF was the better algorithm. Cases involving variables (Fig. 5B) showed reduced differences among the algorithm's performance with the lower number of variables (case I), increasing in the rest of the cases. That increase was higher in RF than in other algorithms, showing the best performance among all studied cases. All the algorithms consistently improved the performance level in the II and III case studies concerning the previous one. In contrast, performance was similar when comparing III and IV, with smooth variations without a common trend. Analyzing both dataset size and variables used (Fig. 5C) when using a low number of variables (case I), the differences among algorithms' performance were lower independent of the dataset size. When the number of variables increases, then algorithms' performance differences also increase, and when the size of the dataset increases, differences among algorithms are more considerable. RF was the algorithm that performed better in all the case studies, and performance also increased more than others when more variables were provided. An increasing number of variables improved performance for all the algorithms, but cases III and IV showed no common trends.

3.2. Effect of different thinning degrees (HII)

When splitting data among thinning degrees (Fig. 6A) performance was better in control plots compared to those where thinning was applied consistently among algorithms. In all the cases, RF showed better performance and a lower performance reduction for above and below thinning degrees compared to control, which was registered in both RF and SVM. DT and SVM performed better on plots where thinning from above was applied compared to thinning from below, while the rest of the algorithms showed the opposite trend. When combining both thinning degrees and variables used (Fig. 6B) control plots showed better performance among classifiers than thinned ones, consistently among combinations. RF always reached the best performance, while differences comparing RF to other classifiers were higher when more variables were used. The performance improved for most of the cases regarding variables, comparing case I to II and case II to III. Still, no clear improvements were shown when comparing cases III and IV.

3.3. Effect of inventory records length (HIII)

When data was split by inventory record length (Fig. 7A) RF still performed the best in all the case studies. Differences were not

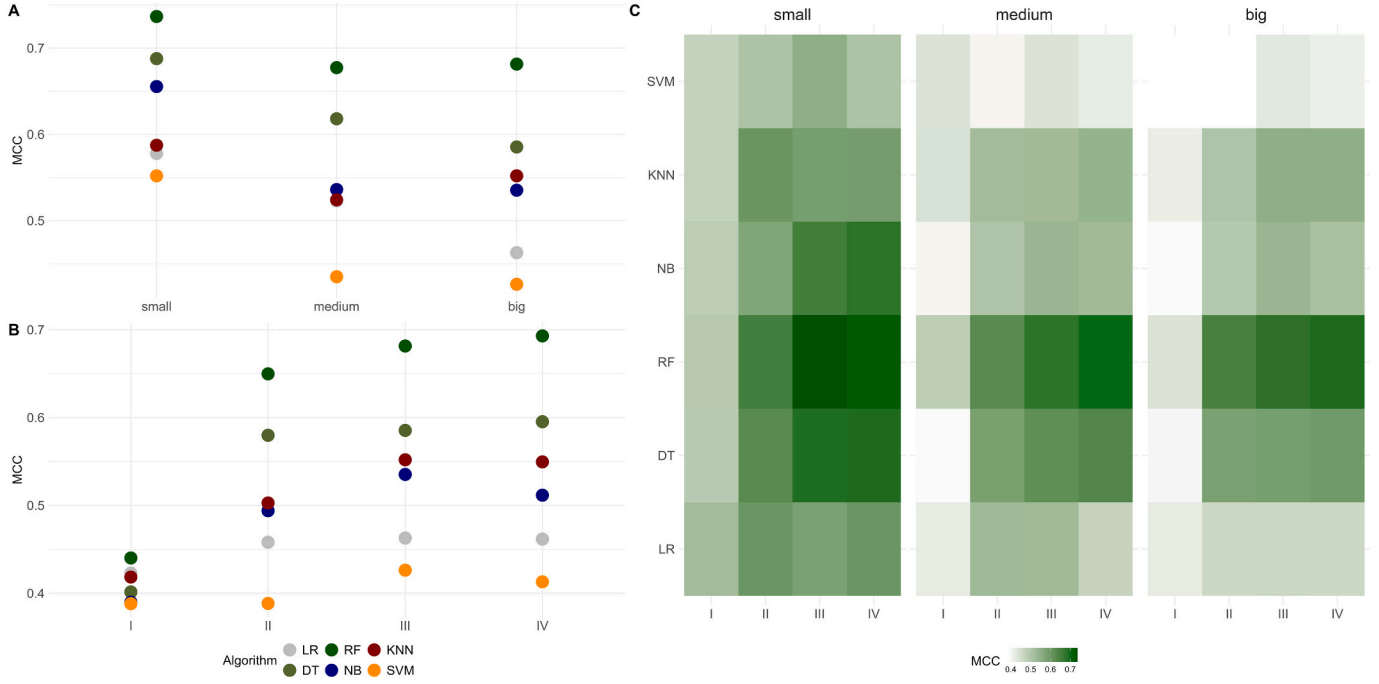


Fig. 5. Algorithm performance for each dataset size (A), number of variables used (B), and both dataset size and number of variables (C) were measured using the MCC metric. On graphs A and B, the x-axis refers to the dataset sizes group (graph A) and the number of variables used (graph B); y-axis values represent the performance of the best model obtained on each case study based on the MCC metric; and each algorithm is shown under a different color according to the legend. In Graph C, dataset sizes group (x-axis upper), the number of variables used per dataset group (x-axis down) and the algorithms used (y-axis) according to its abbreviation (SVM: Support Vector Machine; KNN: K-Nearest Neighbour; NB: Naive Bayes; RF: Random Forest; DT: Decision Trees; LR: Logistic Regression) is shown, while color intensity represents the performance of the best model obtained on each case study based on the MCC metric.

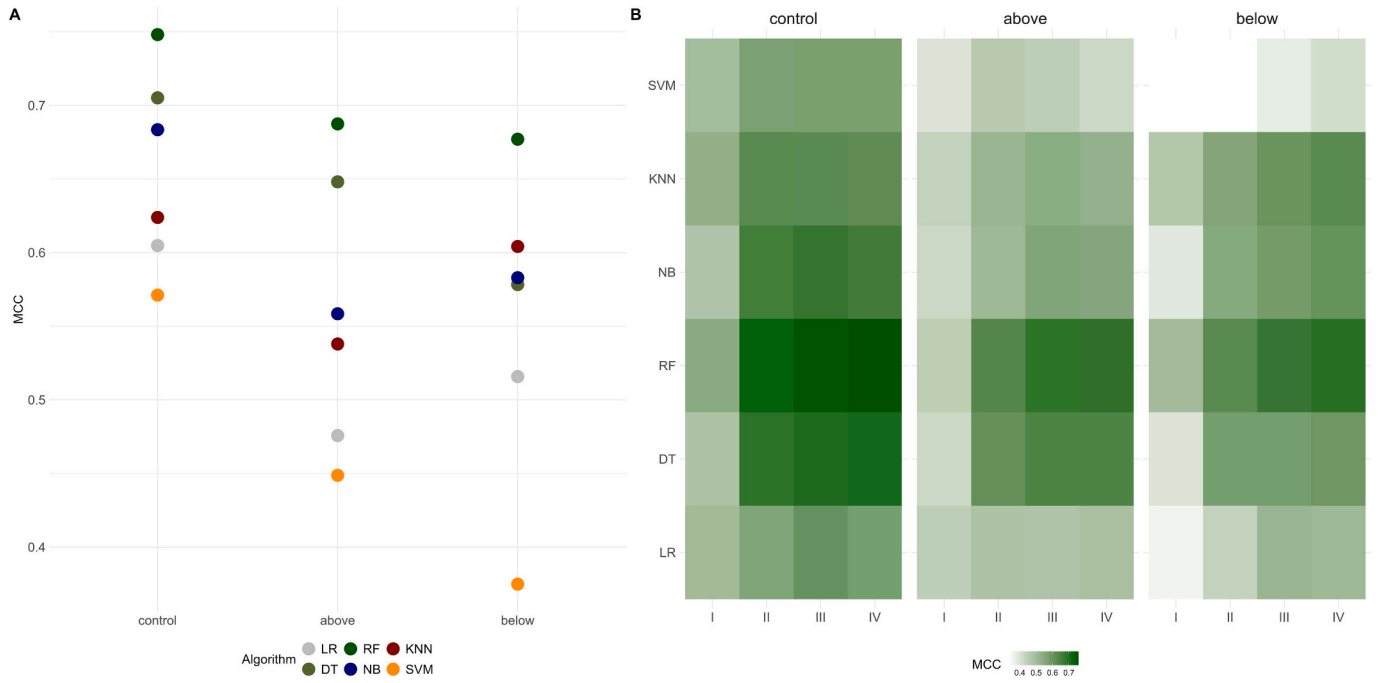


Fig. 6. Algorithm performance for each thinning degree (A) and both thinning degree and number of variables (B) was measured using the MCC metric. In graph A, the x-axis refers to the thinning degree; y-axis values represent the performance of the best model obtained on each case study based on the MCC metric; and each algorithm is shown under a different color according to the legend. In graph B, thinning degrees (x-axis upper), the number of variables used per dataset group (x-axis down) and the algorithms used (y-axis) according to its abbreviation (SVM: Support Vector Machine; KNN: K-Nearest Neighbour; NB: Naive Bayes; RF: Random Forest; DT: Decision Trees; LR: Logistic Regression) is shown, while color intensity represents the performance of the best model obtained on each case study based on the MCC metric.

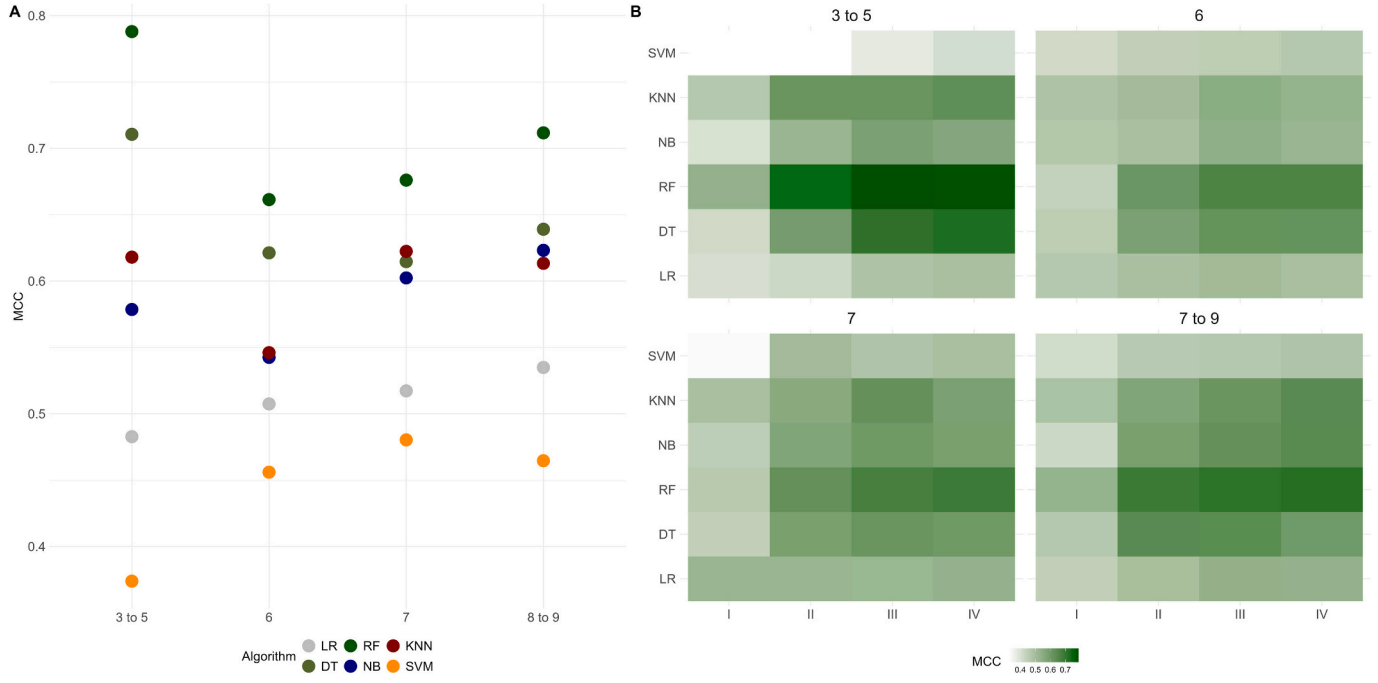


Fig. 7. Algorithm performance for each inventory record length (A) and inventory record length and number of variables (B) is measured using the MCC metric. In graph A, the x-axis refers to the inventory record length; y-axis values represent the performance of the best model obtained on each case study based on the MCC metric; and each algorithm is shown under a different color according to the legend. In graph B, inventory record length (x-axis upper), the number of variables used per dataset group (x-axis down) and the algorithms used (y-axis) according to its abbreviation (SVM: Support Vector Machine; KNN: K-Nearest Neighbour; NB: Naive Bayes; RF: Random Forest; DT: Decision Trees; LR: Logistic Regression) is shown, while color intensity represents the performance of the best model obtained on each case study based on the MCC metric.

significant in most cases, and trends were difficult to find in the case studies. Different behaviors regarding case studies were also shown for each algorithm's performance. LR improved with the record length, while SVM enhanced in the second and third case studies but not in the last one; DT and RF performed better with the smaller record length, and then performance was maintained similar but a bit lower; NB and KNN varied among cases with different trends. Consistently among record-length cases (Fig. 7B) performance was improved when more variables were used to fit the models, while cases III and IV showed no apparent differences. Results were obtained when comparing just the variables used to fit models, and RF was the best algorithm among all the case studies.

3.4. Cross-validation for different thinning degrees (HIII)

Differences were also found among the application case studies for different thinning degrees (Fig. 8). There, results showed how the algorithm's performance decreased when models were applied over data from different thinning degrees, consistently among algorithms and case studies. While LR and DT showed lower performance decreases in some cases, that trend was inconsistent. In some cases, the decrease was lower in NB, SVM, or KNN. Among all case studies, the higher performance loss when using a different thinning degree dataset was registered in RF.

Similar trends were found when using data from two thinning degrees to fit the model (Fig. 9). All the algorithms have reduced their performance level, although LR showed more stability. NB and KNN showed reductions in their performance depending on the dataset used for predictions, having on case A and C similar performance levels when predicting mortality by using the thinning from above dataset. The rest of the algorithms, including RF, showed more unstable behaviors.

4. Discussion

Patterns and dynamics behind tree mortality remain poorly understood, penalizing accurate mortality predictions (Hülsmann et al., 2016). This difficulty increases when mortality tries to be assessed under climate change conditions (Bugmann et al., 2019). While Logistic binomial Regression (LR) is, by far, the most used algorithm to assess tree mortality (Bravo-Oviedo et al., 2006; Hülsmann et al., 2016; Hülsmann et al., 2017; Pretzsch et al., 2020; Shearman et al., 2019a, 2019b; Shifley et al., 2017), the results of this work among all the case studies show how Random Forest (RF) is the algorithm that reaches the overall best performance level. However, RF is the most demanding computing power and the weaker when applying its models to different datasets to those it was trained on, and the interpretability of its results is also a handicap.

4.1. Thinning degrees and cross-validation performance

Algorithms' performance across the different thinning degrees showed consistent results. Algorithm performance in control plots was higher than in those where thinning was applied (both from above and below). This is linked to the forest dynamics, as natural, mainly competition-driven mortality is better represented when no interventions are made, and models can find more robust relationships between covariates provided. When harvests are applied, it was proved that thinning from below reduces the natural mortality in small trees (Dodson et al., 2013; Powers et al., 2010) and thinning from above can increase it (Powers et al., 2010), both altering the short-term stand dynamics after its application. Our data shows that the mortality rate is higher on control plots than on thinned ones, as thinning anticipates mortality. That means that classification rules change when silviculture is applied, which strengthens the alteration of mortality through thinning and can be essential for modeling when trying to find patterns in data. When thinning was used, the algorithm performance was lower, and there were no clear trends among the thinning degrees. The performance of models trained on thinning from below data was expected to be closer to control plots regarding mortality prediction, as trees are more prone to die due to their size, but that trend was not found.

Cross-validation case studies regarding different thinning degrees

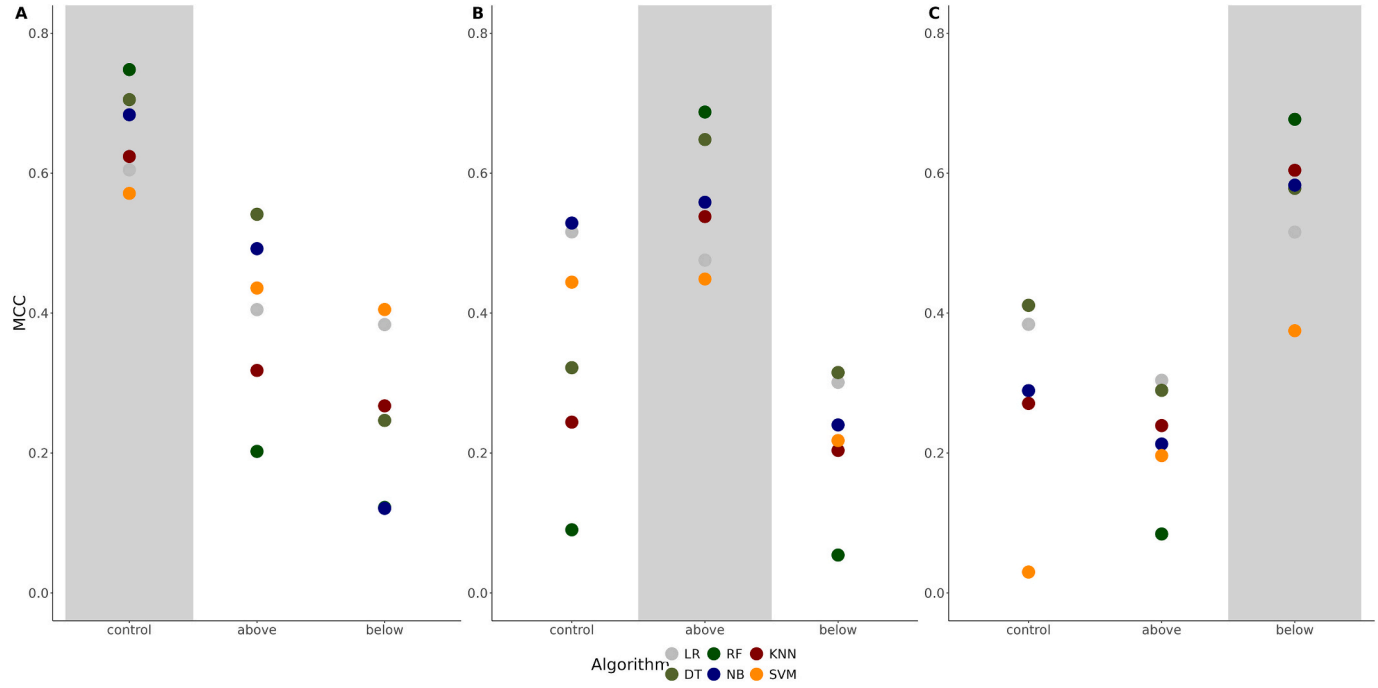


Fig. 8. Algorithm performance for each cross-application case study using different test datasets. On each graph, the x-axis represents the thinning degrees data used; y-axis values represent the performance of the best model obtained on each case study based on the MCC metric; and each algorithm is shown under a different color according to the legend. The shadowed area represents the case when the thinning degree model and data coincide. One RF value is omitted on graph A (thinning from below) as it takes the same value as NB; one RF value is omitted on graph C as it takes a negative value out of the scale.

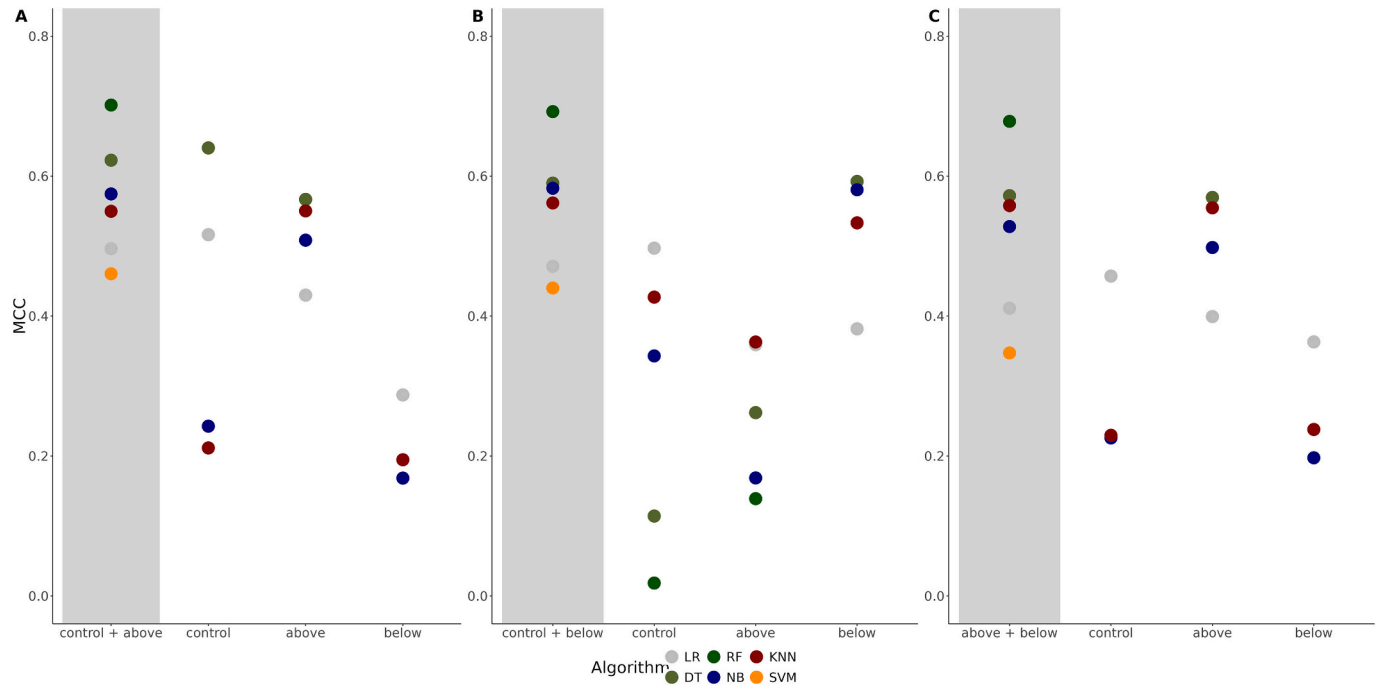


Fig. 9. Algorithm performance for each cross-application case study when using 2 grouped datasets as training data. On each graph, the x-axis represents the thinning degrees data used; y-axis values represent the performance of the best model obtained on each case study based on the MCC metric; and each algorithm is shown under a different color according to the legend. The shadowed area represents the grouped thinning degrees model and its performance. Missing values represent cases when performance is under 0.

showed interesting results related to model applicability. While performance patterns among algorithms were similar when training and test data came from the same thinning degree, their reaction to different test datasets showed different patterns. RF was the algorithm that significantly reduced its performance when applying its models to different thinning datasets. The other Machine Learning (ML) algorithms reacted differently, drastically decreasing their performance in some cases without any common trend. LR was also affected when modifying the test dataset, but performance reductions were smoother. When models were trained on datasets combining two thinning intensities, RF continued to exhibit similar trends, whereas Naive Bayes (NB) and K-Nearest Neighbors (KNN) maintained their performance levels more effectively. LR showed a remarkable performance decline when trained on control data combined with one thinning dataset (e.g., thinning from below) and tested on the other thinning dataset (e.g., thinning from above). However, when trained on both thinning datasets, its performance predicting with the control dataset improved. This may reflect the challenges posed by mortality dynamics under different management regimes, where thinning disturbances increase prediction difficulty for some algorithms. While it is expected to develop species-specific models to predict mortality (Franklin et al., 1987; Hülsmann et al., 2017) and also regional models due to its variation in different geographical areas (Hülsmann et al., 2017; Monserud and Sterba, 1999), our results suggest the idea of developing models to predict mortality under different thinning degrees, as proposed by Bravo-Oviedo et al. (2006). In addition, in long-term simulations, the performance of mortality models often declines following thinning events due to shifts in mortality behavior within stands. Our results suggest that these models can recover their predictive accuracy over time, but their handicap to taking into account the changes immediately after thinning highlights a critical area for improvement.

A similar approach did not report differences in tree mortality between managed and unmanaged forest (natural reserves) patterns, but it stressed the importance of the similar ecological processes between forests used for calibration and validation (Hülsmann et al., 2017). That ecological processes also interact with mortality rates after thinning (Dodson et al., 2013; Powers et al., 2010), triggering different forest dynamics that alter the forest behavior before thinning. Results shown here related to the model performance reduction on cross-validation support the necessity to go deeper into the study of thinning effects on tree mortality. In addition, the effects of applying different data structures on fitting and predicting would be an interesting topic to address when using forest simulators. Frequently, the origin of the data used to fit the models differs from the data used for predictions, which can compromise the reliability of the predictions. When simulating long study periods, the error propagation is unknown, and the use of non-appropriate data structures can alter the reliability of predictions.

4.2. *Size, variables, and inventory records length performance*

The size of the dataset was directly related to the accuracy of predictions, which was consistent with all the algorithms. While it was expected that a more extensive dataset would provide higher accuracy on predictions as many authors expected (Bugmann et al., 2019; Wunder et al., 2008), we found the opposite results, even when small and medium-sized datasets were randomly selected from the bigger ones, thus avoiding deviations in data selection. The higher performance of the algorithms in smaller datasets can be related to the increase of data heterogeneity of the bigger ones (Hülsmann et al., 2017; Sheil and May, 1996), as different data structures (i.e., thinning degrees) are studied together.

The number of variables used to fit the models ordered from case I to IV, as more variables are included, was an essential factor when determining prediction accuracy. Hülsmann et al. (2018, 2017) and Zhang et al. (2009) reported the importance of tree size, competition, and growth as covariates in predicting mortality, all included in our study. The competition was proved to have higher explanatory power when applied to the individual tree neighborhood instead of the plot (Rohner et al., 2012), as we performed in this work. Our results verify the importance of growth as a covariate as predictions improve in case II

compared to case I among all algorithms. Consistently among algorithms, an increase in the number of variables until case III enhances the accuracy of predictions (with exceptions), while an excess of variables in case IV shows different trends among algorithms with minor changes. That can be related to the high explanatory power of covariates already included in the model compared with the social position of the tree included in case IV. While LR used to be fitted using a low number of variables with high explanatory power, algorithms like RF can support a higher number of variables even when their explanatory power is lower. In this context, RF can minimize overfitting by averaging multiple decision trees, reducing the impact of variance caused by overfitting in individual trees (Breiman, 2001). This capability, combined with its ability to capture nonlinear relationships, increases the overall performance of the model (Breiman, 2001), which might explain the results obtained in our study. Alternative variables candidates like tree age or height were excluded from the analysis, and variable transformations and interactions were not tested.

Different trends regarding the number of records per inventory were found for each algorithm. While LR improves its performance when more records of the same inventory are available on the data provided, RF and Naive Bayes (NB) follow the same trend except in one case study. Decision Trees (DT), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM) showed different behaviors. As shown in previous works (Hülsmann et al., 2017). However, that information may already be captured indirectly through other covariates included in the model, suggesting that its influence is embedded within the existing predictors.

4.3. ML modeling implications

Differences among algorithms were observed for each case study, providing helpful information to reference in future studies. RF was the algorithm that better performed in all study cases except on the cross-application with different thinning degrees. Differences among the other algorithms' performance are hard to rate, as no clear results are obtained from their comparison. LR, the reference algorithm, always remains in between in terms of performance, while the other ML algorithms got fewer stable scores in performance depending on the case study. The higher performance of RF compared to other algorithms under different conditions of data size, number of variables, and data type (thinning and records classes) enhances the idea of selecting RF to assess tree mortality in future works. Other advantages were noticed in previous works related to forestry (Cutler et al., 2007; McNellis et al., 2021; Shearman et al., 2019b; Zhao et al., 2019), like its performance to estimate forest quality (Zhao et al., 2019) and to find complex interactions on predictor variables for ecological classification (Cutler et al., 2007). So, in those cases, RF is a good alternative when a higher predictivity rate is needed, i.e., when working with National Forest Inventories. However, its application should be made on data with a similar structure to the training one, as performance in the cross-validation case studies reported the worst predictions among algorithms. Thus, performance reduction can be related to the differences in covariate interactions, as RF creates more complex relationships than other algorithms (Cutler et al., 2007). In that case, while our results did not provide clear trends for each algorithm, LR can be considered the more robust algorithm as it remained with a more consistent performance when it was drastically reduced in some cases on ML algorithms. Furthermore, while computer power limitations are lower each time (Shifley et al., 2017), the time spent training each algorithm (and computer requirements not studied here) should be considered. While RF was by far the algorithm with the highest computing time demanded, SVM also shows higher requirements than LR and the other ML algorithms. In addition, when more data is used to train the algorithms and more variables are provided, the time requirements increase for all the algorithms, with some exceptions.

A point to consider in all the ML algorithms is the difficulty of interpreting their results (Shearman et al., 2019b), which is a critical aspect in every research field. LR offers simplicity in this regard, providing an equation where the effect of each covariate can be easily understood. In contrast, the interpretation of ML algorithms requires different approaches; rather than offering a clear equation, they rely on methods like feature importance or partial dependence plots, which many users may not yet be familiar with. While some algorithms like DT are easily interpretable, others like RF are more complex and are evaluated primarily through predictive performance metrics. This can obscure the relationships between covariates that lead to a result. Consequently, when understanding causality is crucial, LR provides a more straightforward and interpretable framework. However, when the goal is to obtain a better prediction rate, they have been proven to be good alternatives for forestry (McNellis et al., 2021) and even in other topics, like health, where comparisons were also made to find the better algorithm for a particular application (Aldossary et al., 2022; Maydanchi et al., 2023). That debate about the prediction level and the understandability of the models has been extensively discussed in the literature related to this topic.

Deep Learning alternatives like ANN were initially considered but were finally excluded due to their architectural complexity (Reis et al., 2018) and adaptation requirements to each case study. In addition, even though ANN was applied in different works assessing tree mortality (Bayat et al., 2019; da Rocha et al., 2018a; Reis et al., 2018), a comparison with other algorithms could prove if they are a better alternative than RF to reach better performance levels. In that field, metrics selected when assessing model predictions must be appropriate, considering data structure and objectives (Ferri et al., 2009). Different metrics were estimated in this work and can be consulted on the complementary data. Metrics related to the overall classification accuracy of the model are not appropriate as a performance measurement when working with imbalanced data (Chen and Breiman, 2004). MCC (Matthews, 1975) was selected to compare algorithms because, as explained by Chicco (2017) and Chicco and Jurman (2020), MCC is a good choice when dealing with imbalanced datasets as it properly considers the ratio of the confusion matrix size. Examples of estimating tree mortality showed how selecting a non-proper metric to assess model performance provides confusing results. In some cases, models showed a high significance on prediction but provided inferior dead tree prediction rates (da Rocha et al., 2018a, 2018b; Reis et al., 2018), evaluated separately. In other cases, the prediction of dead tree was not even considered (Bayat et al., 2019). Finally, model applicability must be considered depending on the user's purpose. While LR models can be applied through an equation, ML models must be used through the file containing the previously trained model. That fact reduces their applicability for some users due to the knowledge required and the higher complexity of the model transferability, while its importance is lower for advanced users. In addition, the inclusion of mortality models into Decision Support Systems like SILVA (Pretzsch et al., 2002) or SIMANFOR (Bravo et al., 2025) implies that tree death predictions will be estimated using simulated covariates (Hülsmann et al., 2017), thus increasing the error in predictions.

5. Conclusion

This study compared Machine Learning algorithms' performance by assessing individual tree mortality. A performance improvement was found consistently among algorithms when more variables were used to fit the model, while more extensive datasets decreased the performance level of the algorithms. Datasets from control plots provided better performance levels than those acquired on thinned ones, while cross-validation decreased model performance, consistent with all algorithms. No clear trends were found when using different inventory record lengths. Random Forest was the algorithm with higher performance levels in all the study cases. At the same time, cross-validation drastically reduced its performance, while Logistic binomial Regression seemed to be more robust in those cases. Our results suggest that Random Forest is a good choice if higher prediction levels are required,

while other Machine Learning alternatives are inferior. However, when the model application is not clearly defined and/or model interpretability is needed, Logistic binomial Regression is still the best tool for assessing individual tree mortality.

CRediT authorship contribution statement

Aitor Vázquez-Veloso: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Astor Toraño Caicoya:** Supervision, Resources, Methodology, Data curation. **Felipe Bravo:** Supervision, Methodology, Funding acquisition, Conceptualization. **Peter Biber:** Supervision, Conceptualization. **Enno Uhl:** Supervision, Resources, Funding acquisition, Conceptualization. **Hans Pretzsch:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Funding

Funds were directly received from the European Union and Junta de Castilla y León Education Council (ORDEN EDU/842/2022), and from the University of Valladolid through "MOVILIDAD DE ESTUDIANTES DE DOCTORADO Uva 2023". This study has also been subsidized by the Junta de Castilla y León through the projects "CLU-2019-01 and CL-EI-2021-05 – iuFOR Institute Unit of Excellence" of the University of Valladolid, co-financed by the European Regional Development Fund (ERDF "Europe drives our growth"), and from the projects "778322/CARE4C/H2020/MSCA/RISE/2017 – Carbon smart forestry under climate change" and the IMFLEX Grant PID2021-126275OB-C22 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper and its research would not have been possible without all the work developed over time by the Chair of Forest Growth and Yield Science, which collected data over several years and shared the database used in this study. Thanks are also due to the Bayerische Staatsforsten (BaySF) for providing the experimental plots in Bavaria and to the Bavarian State Ministry for Nutrition, Agriculture, and Forestry for the permanent support of the project W 07"Long-term experimental plots for forest growth and yield research"(#7831-26625-2017). In addition, we would like to thank Rafael Calama for his suggestions and the ideas provided at the beginning of the study.

Appendix A. Appendix

The differences are very clear when comparing the time spent on train models per algorithm and case studies. As expected, more data requires more time, consistently among algorithms (Fig. A.1.Fig. A.1A). By default, RF time requirements are more extensive, and the time necessities increase more than other algorithms among study cases. SVM also shows higher time demand than other algorithms, but differences with RF are more significant. Regarding the number of variables (Fig. A.1.B), in this case, they were split by the number of covariates used in the model instead of the variables case study to extract more information. RF algorithm needs more time among case studies, followed by SVM. The increment in time needed when more variables are included in the model is inconsistent among algorithms. RF needs more time to fit models with six than with 7 and 8 covariates, while SVM reduces the time required with more than seven covariates. Comparing among algorithms, the necessity of RF grows considerably as more variables are included in the model. Combining dataset size and number of variables (Fig. A.1.C), it looks like the differences in time required are bigger when more data is provided, and there are some peaks (appreciable just in RF and SVM) of time needed when fitting models with 6 and 7 covariates, respectively.

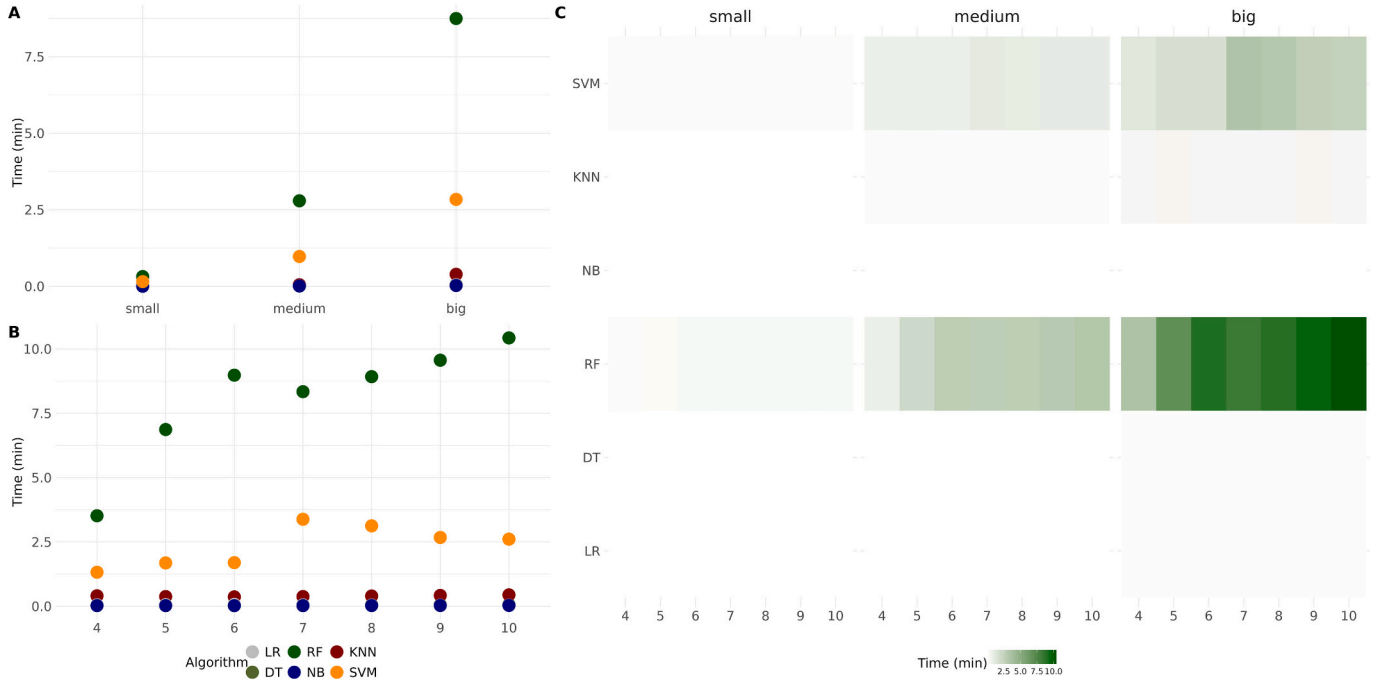


Fig. A.1. The average time used to train a single model per algorithm across dataset size (A), number of variables (B), and both dataset size and number of variables (C) measured by the time needed to train a single model in minutes. On graphs A and B, the x-axis refers to the dataset sizes group (graph A) and the number of variables used (graph B); y-axis values represent the time (minutes) needed to fit a single model; and each algorithm is shown under a different color according to the legend. In Graph C, the dataset sizes group (x-axis upper), the number of variables used per dataset group (x-axis down) and the algorithms used (y-axis) according to its abbreviation (SVM: Support Vector Machine; KNN: K-Nearest Neighbour; NB: Naive Bayes; RF: Random Forest; DT: Decision Trees; LR: Logistic Regression) are shown, while color intensity represents the time needed to fit a single model in minutes.

Data availability

File name: ML_individual_tree_mortality.zip.
Developer: Aitor Vázquez Veloso.
Contact information: aitor.vazquez.veloso@uva.es
Software required: R.
Availability: The full raw dataset and intermediate curation/analysis data used in that publication, the code and the results are available at this DOI on Zenodo: doi: https://doi.org/10.5281/zenodo.14970488
Cost: Free.

References

Aldossary, Y., Alhaddad, S., Ebrahim, M., Zeki, A.M., 2022. Comparing K-nearest neighbors, random Forest and naïve Bayes models to classify fetal health using resampling methods. In: 2022 international conference on data analytics for business and industry (ICDABI) Presented at the 2022 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 250–254. https://doi.org/10.1109/ICDABI56818.2022.10041489.

Alenius, V., Hökkä, H., Salminen, H., Jutras, S., 2003. Evaluating estimation methods for logistic regression in modelling individual-tree mortality. In: Model. For. Syst. Workshop Interface real. Model. Parameter Estim. Process. Sesimbra port. 2-5 June 2002, pp. 225–236. https://doi.org/10.1079/9780851996936.0225.

Assmann, E., Franz, F., 1972. Vorläufige Fichten-Ertragstafel für Bayer 1963. Forstl. Forschungsanst., Inst. für Ertragskunde.

Bayat, M., Ghorbanpour, M., Zare, R., Jaafari, A., Thai Pham, B., 2019. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. Comput. Electron. Agric. 164, 104929. https://doi.org/10.1016/j.compag.2019.104929.

Boeck, A., Dieler, J., Biber, P., Pretzsch, H., Ankerst, D.P., 2014. Predicting tree mortality for European beech in southern Germany using spatially explicit competition indices. For. Sci. https://doi.org/10.5849/forsci.12-133.

Bravo, F., Hann, D.W., Maguire, D.A., 2001. Impact of competitor species composition on predicting diameter growth and survival rates of Douglas-fir trees in southwestern Oregon. Can. J. For. Res. 31, 2237–2247. https://doi.org/10.1139/cjfr-31-12-2237.

Bravo, F., Alvarez-Gonzalez, J.G., Del Rio Gatelurrutia, M.G., Barrio, M., Bonet, J.A., Bravo-Oviedo, A., Calama Sainz, R.A., Castedo-Dorado, F., Crecente-Campo, F., Condes, S., Others, 2011. Growth and yield Models in Spain Historical overview, Contemporary Examples and Perspectives.

Bravo, F., Ordóñez, C., Vázquez-Veloso, A., Michalakopoulos, S., 2025. SIMANFOR cloud decision support system: structure, content, and applications. Ecol. Model. 499, 110912. https://doi.org/10.1016/j.ecolmodel.2024.110912.

Bravo-Oviedo, A., Sterba, H., Del Río, M., Bravo, F., 2006. Competition-induced mortality for Mediterranean Pinus pinaster Ait. And P. Sylvestris L. For. Ecol. Manag. 222, 88–98. https://doi.org/10.1016/j.foreco.2005.10.016.

Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. 16, 199–231. https://doi.org/10.1214/ss/1009213726.

Bugmann, H., Seidl, R., Hartig, F., Bohn, F., Brůna, J., Cailleret, M., François, L., Heinke, J., Henrot, A., Hickler, T., Hülsmann, L., Huth, A., Jacquemin, I., Kollas, C., Lasch-Born, P., Lexer, M.J., Merganič, J., Merganičová, K., Mette, T., Miranda, B.R., Nadal-Sala, D., Rammer, W., Rammig, A., Reineking, B., Roedig, E., Sabaté, S., Steinkamp, J., Suckow, F., Vacchiano, G., Wild, J., Xu, C., Reyer, C.P.O., 2019. Tree mortality submodels drive simulated long-term forest dynamics: assessing 15 models from the stand to global scale. Ecosphere 10, e02616. https://doi.org/10.1002/ecs2.2616.

Caicoya, A.T., Pretzsch, H., 2021. Stand density biases the estimation of the site index especially on dry sites. Can. J. For. Res. https://doi.org/10.1139/cjfr-2020-0389.

Chen, C., Breiman, L., 2004. Using Random Forest to Learn Imbalanced Data. Univ. Calif, Berkeley.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. BioData Min. 10, 35. https://doi.org/10.1186/s13040-017-0155-3.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21, 6. https://doi.org/10.1186/s12864-019-6413-7.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. https://doi.org/10.1177/001316446002000104.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1.

da Rocha, S.J.S.S., Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K. M., Schettini, B.L.S., Villanova, P.H., da Silva, L.F., Reis, L.P., Zanuncio, J.C., 2018a. Artificial neural networks: modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Sci. Total Environ. 645, 655–661. https://doi.org/10.1016/J.SCITOTENV.2018.07.123.

da Rocha, S.J.S.S., Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K. M., Schettini, B.L.S., Villanova, P.H., da Silva, L.F., Reis, L.P., Zanuncio, J.C., 2018b. Artificial neural networks: modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Sci. Total Environ. 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123.

Dodson, E.K., Puettmann, K.J., Ares, A., 2013. Thinning effects on tree mortality and snag recruitment in Western Oregon. In: Presented at the density management in the 21st Century: West Side Story. Corvallis, Oregon, pp. 71–78.

Ferri, C., Hernández-Orallo, J., Modroiu, R., 2009. An experimental comparison of performance measures for classification. Pattern Recogn. Lett. 30, 27–38. https://doi.org/10.1016/j.patrec.2008.08.010.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/JOC.5086.
Franklin, J.F., Shugart, H.H., Harmon, M.E., 1987. Tree death as an ecological process. BioScience 37, 550–556. https://doi.org/10.2307/1310665.
Gamer, M., Lemon, J., Puspendra Singh, I.F., 2019. Irr: Various Coefficients of Interrater Reliability and Agreement.
Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci. Data 71 (7), 1–18. https://doi.org/10.1038/s41597-020-0453-3.
Hegyi, F., 1974. A simulation model for managing jack-pine standssimulation. In: J fried Ed growth models tree stand Simul. R. Coll. For. Stockh, pp. 74–90.
Hülsmann, L., Bugmann, H., Commarmot, B., Meyer, P., Zimmermann, S., Brang, P., 2016. Does one model fit all? Patterns of beech mortality in natural forests of three European regions. Ecol. Appl. 26, 2465–2479. https://doi.org/10.1002/eap.1388.
Hülsmann, L., Bugmann, H., Brang, P., 2017. How to predict tree death from inventory data — lessons from a systematic assessment of European tree mortality models. Can. J. For. Res. 47, 890–900. https://doi.org/10.1139/cjfr-2016-0224.
Hülsmann, L., Bugmann, H., Cailleret, M., Brang, P., 2018. How to kill a tree: empirical mortality models for 18 species and their performance in a dynamic forest model. Ecol. Appl. 28, 522–540. https://doi.org/10.1002/eap.1668.
Kuhn, M., 2008. Building predictive models in *R* using the **caret** package. J. Stat. Softw. 28. https://doi.org/10.18637/jss.v028.i05.
Lynch, T.B., Hitch, K.L., Huebschmann, M.M., Murphy, P.A., 1999. An individual-tree growth and yield prediction system for even-aged natural shortleaf pine forests. South. J. Appl. For. 23, 203–211. https://doi.org/10.1093/SJAF/23.4.203.
Martonne, E., 1926. L'indice d'aridité. Bull. L'Association Géographes Fr. 3, 3–5.
Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta BBA - Protein Struct. 405, 442–451. https://doi.org/10.1016/0005-2795(75)90109-9.
Maydanchi, M., Ziaei, A., Basiri, M., Azad, A.N., Pouya, S., Ziaei, M., Haji, F., Sargolzaei, S., 2023. Comparative study of decision tree, AdaBoost, random Forest, naïve Bayes, KNN, and perceptron for heart disease prediction. In: SoutheastCon 2023. Presented at the SoutheastCon, 2023, pp. 204–208. https://doi.org/10.1109/SoutheastCon51012.2023.10115189.
McDowell, N.G., Allen, C.D., Anderson-Teixeira, K., Aukema, B.H., Bond-Lamberty, B., Chini, L., Clark, J.S., Dietze, M., Grossiord, C., Hanbury-Brown, A., Hurtt, G.C., Jackson, R.B., Johnson, D.J., Kueppers, L., Lichstein, J.W., Ogle, K., Poulter, B., Pugh, T.A.M., Seidl, R., Turner, M.G., Uriarte, M., Walker, A.P., Xu, C., 2020. Pervasive shifts in forest dynamics in a changing world. Science 368, eaaz9463. https://doi.org/10.1126/science.aaz9463.
McNellis, B.E., Smith, A.M.S., Hudak, A.T., Strand, E.K., 2021. Tree mortality in western U.S. forests forecasted using forest inventory and random Forest classification. Ecosphere 12, e03419. https://doi.org/10.1002/ecs2.3419.
Merkl, D., Hasenauer, H., 1998. Using neural networks to predict individual tree mortality. In: Proceedings of the int'l Conference on Engineering Applications of Neural Networks, pp. 10–12.
Mladenoff, D.J., 2004. LANDIS and forest landscape models. In: Ecol. Model., Modelling disturbance and succession in forest landscapes using LANDIS, 180, pp. 7–19. https://doi.org/10.1016/j.ecolmodel.2004.03.016.
Monserud, R.A., Sterba, H., 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manag. 113, 109–123. https://doi.org/10.1016/S0378-1127(98)00419-8.
Pascual, A., Tupinambá-Simões, F., Guerra-Hernández, J., Bravo, F., 2022. High-resolution planet satellite imagery and multi-temporal surveys to predict risk of tree mortality in tropical eucalypt forestry. J. Environ. Manag. https://doi.org/10.1016/j.jenvman.2022.114804.
Powers, M.D., Palik, B.J., Bradford, J.B., Fraver, S., Webster, C.R., 2010. Thinning method and intensity influence long-term mortality trends in a red pine forest. For. Ecol. Manag. 260, 1138–1148. https://doi.org/10.1016/j.foreco.2010.07.002.
Pretzsch, H., 2009. Forest Dynamics, Growth, and Yield. Springer.
Pretzsch, H., Biber, P., 2005. A re-evaluation of Reineke's rule and stand density index. For. Sci. 51, 304–320. https://doi.org/10.1093/FORESTSCIENCE/51.4.304.
Pretzsch, H., Biber, P., Ďurský, J., 2002. The single tree-based stand simulator SILVA: construction, application and evaluation. For. Ecol. Manag. https://doi.org/10.1016/S0378-1127(02)00047-6.
Pretzsch, H., Grams, T., Häberle, K.H., Pritsch, K., Bauerle, T., Rötzer, T., 2020. Growth and mortality of Norway spruce and European beech in monospecific and mixed-species stands under natural episodic and experimentally extended drought. Results of the KROOF throughfall exclusion experiment. Trees - Struct. Funct. 34, 957–970. https://doi.org/10.1007/S00468-020-01973-0/TABLES/4.
Pretzsch, H., del Río, M., Arcangeli, C., Bielak, K., Dudzinska, M., Forrester, D.I., Klädtke, J., Kohnle, U., Ledermann, T., Matthews, R., Nagel, J., Nagel, R., Ningre, F., Nord-Larsen, T., Biber, P., 2023a. Forest growth in Europe shows diverging large regional trends. Sci. Rep. https://doi.org/10.1038/s41598-023-41077-6.
Pretzsch, H., del Río, M., Arcangeli, C., Bielak, K., Dudzinska, M., Ian Forrester, D., Kohnle, U., Ledermann, T., Matthews, R., Nagel, R., Ningre, F., Nord-Larsen, T., Szeligowski, H., Biber, P., 2023b. Competition-based mortality and tree losses. An essential component of net primary productivity. For. Ecol. Manag. 544, 121204. https://doi.org/10.1016/j.foreco.2023.121204.
R Core Team, 2021. R: A Language and Environment for Statistical Computing.
Reineke, L.H., 1933. Perfection a stand-density index for even-aged forest. J. Agric. Res. 46, 627–638.
Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Torres, Miquelino Eleto, da Silva, C.M., Ruschel, L.F., Rêgo, A.R., L.J.S., Leite, H.G, 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecol. Eng. 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014.
Rohner, B., Bigler, C., Wunder, J., Brang, P., Bugmann, H., 2012. Fifty years of natural succession in Swiss forest reserves: changes in stand structure and mortality rates of oak and beech. J. Veg. Sci. 23, 892–905. https://doi.org/10.1111/j.1654-1103.2012.01408.x.
Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10, e0118432. https://doi.org/10.1371/journal.pone.0118432.
Salas-Eljatib, C., Weiskittel, A.R., 2020. On studying the patterns of individual-based tree mortality in natural forests: a modelling analysis. For. Ecol. Manag. 475, 118369. https://doi.org/10.1016/j.foreco.2020.118369.
Schlyter, P., Stjernquist, I., Bärring, L., Jönsson, A.M., Nilsson, C., 2006. Assessment of the impacts of climate change and weather extremes on boreal forests in northern Europe, focusing on Norway spruce. Clim. Res. 31, 75–84. https://doi.org/10.3354/cr031075.
Senf, C., Buras, A., Zang, C.S., Rammig, A., Seidl, R., 2020. Excess forest mortality is consistently linked to drought across Europe. Nat. Commun. 111 (11), 1–8. https://doi.org/10.1038/s41467-020-19924-1.
Shearman, T.M., Varner, J.M., Hood, S.M., Cansler, C.A., Hiers, J.K., 2019a. Modelling post-fire tree mortality: can random forest improve discrimination of imbalanced data? Ecol. Model. 414, 108855. https://doi.org/10.1016/j.ecolmodel.2019.108855.
Shearman, T.M., Varner, J.M., Hood, S.M., Cansler, C.A., Hiers, J.K., 2019b. Modelling post-fire tree mortality: can random forest improve discrimination of imbalanced data? Ecol. Model. 414, 108855. https://doi.org/10.1016/j.ecolmodel.2019.108855.
Sheil, D., May, R.M., 1996. Mortality and recruitment rate evaluations in heterogeneous tropical forests. J. Ecol. 84, 91–100. https://doi.org/10.2307/2261703.
Shifley, S.R., He, H.S., Lischke, H., Wang, W.J., Jin, W., Gustafson, E.J., Thompson, J.R., Thompson, F.R., Dijak, W.D., Yang, J., 2017. The past and future of modeling forest dynamics: from growth and yield curves to forest landscape models. Landsc. Ecol. 32, 1307–1325. https://doi.org/10.1007/s10980-017-0540-9.
Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 3940–3941. https://doi.org/10.1093/bioinformatics/bti623.
Steneker, G.A., Jarvis, J.M., 1963. A preliminary study to assess competition in a white spruce - trembling Aspen stand. For. Chron. 39, 334–336. https://doi.org/10.5558/tfc39334-3.
Thrippleton, T., Hülsmann, L., Cailleret, M., Bugmann, H., 2021. An evaluation of multi-species empirical tree mortality algorithms for dynamic vegetation modelling. Sci. Rep. 11, 19845. https://doi.org/10.1038/s41598-021-98880-2.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.
von Teuffel, K., von Heinrich, B., Baumgarten, M., 2004. Present Distribution of Secondary Norway Spruce in Europe. In: Norway Spruce Conversion. Brill, pp. 63–96. https://doi.org/10.1163/9789047412908_007.
Wang, W., Peng, C., Kneeshaw, D.D., Larocque, G.R., Luo, Z., 2012. Drought-induced tree mortality: ecological consequences, causes, and modeling. Httpsdoiorg 20, 109–121. https://doi.org/10.1139/A2012-004.
Waring, R.H., 1987. Characteristics of trees predisposed to die. Bioscience 37, 569–574. https://doi.org/10.2307/1310667.
Wunder, J., Brzeziecki, B., Żybura, H., Reineking, B., Bigler, C., Bugmann, H., 2008. Growth–mortality relationships as indicators of life-history strategies: a comparison of nine tree species in unmanaged European forests. Oikos 117, 815–828. https://doi.org/10.1111/j.0030-1299.2008.16371.x.
Wykoff, W.R., 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. For. Sci. 36, 1077–1104. https://doi.org/10.1093/FORESTSCIENCE/36.4.1077.
Zhang, J., Hao, Z., Sun, I.-F., Song, B., Ye, J., Li, B., Wang, X., 2009. Density dependence on tree survival in an old-growth temperate forest in northeastern China. Ann. For. Sci. 66 (2), 1–9. https://doi.org/10.1051/forest/2008086.
Zhao, Q., Yu, S., Zhao, F., Tian, L., Zhao, Z., 2019. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. For. Ecol. Manag. 434, 224–234. https://doi.org/10.1016/j.foreco.2018.12.019.