Journal of Oil Palm Research
DOI: https://doi.org/10.21894/jopr.2025.0041

PREDICTION OF OIL PALM BUNCHES PRODUCTION USING ARTIFICIAL NEURAL NETWORK

EMANUELLY CANABRAVA MAGALHÃES[1]*; CARLOS ALBERTO ARAÚJO JÚNIOR[2]; HELIO GARCIA LEITE[1]; GIANMARCO GOYCOCHEA CASAS[1]; JULIANA MACHADO LISBOA[3]; CARLOS HENRIQUE GARCIA[3] AND RODRIGO RABE SALES BICALHO[3]

ABSTRACT

The study aimed to evaluate the ability of Artificial Neural Networks (ANN) to estimate the current monthly production of oil palm bunches using variables from the forest inventory, climatic elements, water deficit, soil and those related to the registration and management of plantations. The ANN estimated current production with a correlation above 0.6 and an average percentage relative error of around 13.0%, with the variables that gave the greatest contribution to modelling being those related to management, soil, genetic material and the accounting of mature bunches. Climatic variables were not as important, however, due to the influence of the climatic element on oil palm productivity, it is necessary to keep them in the modelling. The ANN demonstrated that it is capable of modelling oil palm production, characterised by high variability, opening opportunities for future studies, combining and using new variables to improve the accuracy of estimates using this tool.

Keywords: artificial intelligence, modelling, oil palm.

Received: 12 August 2024**; Accepted:** 26 May 2025**; Published online:** 13 August 2025.

INTRODUCTION

Brazil has the potential to become one of the world's largest producers of palm oil due to its vast availability of land and favourable climatic conditions. However, current production is still much lower than the world's leading producers and cannot meet the growing domestic demand (Murphy *et al*., 2021; Organização Internacional do Trabalho [OIT], 2022). In 2022, Brazil produced around 570,000 t of palm oil, ranking 10th globally. This represents just 0.70% of the world's total production (U.S. Department of Agriculture [USDA], 2022). The oil palm cultivated area in Brazil has expanded rapidly in recent years, and it is expected to continue to grow. The Brazilian Palm Oil Producer Association (ABRAPALMA), estimates that the palm oil area in Pará is 535,493 ha, with 226,834 ha planted with oil palm (Drost *et al*., 2021).

The consumption of palm oil in Brazil has been growing steadily since 2008. In 2023, Brazil is expected to consume about 930,000 t of palm oil, an increase of 15,000 t from the previous year. This growth is being driven by a number of factors, including the increasing popularity of processed foods that contain palm oil, the growing middle class in Brazil, and the expansion of the palm oil industry in the country (Statista, 2023).

To meet increasing demand, it's crucial to optimise production by efficiently utilising available resources. Achieving this optimisation hinges on a deeper understanding of the factors influencing

---

[1] Federal University of Viçosa,
P. H. Rolfs Avenue, Campus UFV,
Viçosa/MG, 36.570-900, Brazil.

[2] Federal University of Minas Gerais,
Universitária Avenue, Universitário,
Institute of Agricultural Sciences,
Montes Claros/MG, 39.404-547, Brazil.

[3] Agropalma S.A., Road PA 150 s/n Km 74 left,
Tailândia/PA, 68.695-000, Brazil.

* *Corresponding author e-mail: emanuellymagalhaes1@gmail.com*

species growth and development, enabling informed decision-making. Consequently, it's imperative for forest managers to employ computational tools capable of modelling production based on production-related variables, aiding in problem-solving (Briassoulis, 2020).

Artificial intelligence (AI), which includes machine learning, deep learning and computer vision, is being actively promoted to improve industrial efficiency and sustainability while minimising environmental impact. Its applications in precision agriculture, production mapping, land use monitoring and remote sensing, which use satellite imagery, unmanned aerial vehicles (UAVs) and ground sensors, are enabling advances in palm oil management, such as yield forecasting, disease detection and land use classification (Akhtar *et al*., 2023).

AI has been shown to be very effective in accurately classifying the maturity of oil palm fruit bunches, leading to significant improvements in harvesting efficiency. The potential benefits of using this AI technology include increased harvesting efficiency and productivity, as well as improved quality control and reduced waste (Harsawardana *et al*., 2020). AI can more accurately and efficiently estimate oil palm yield by using vegetation and moisture indices extracted from satellite imagery, outperforming traditional methods (Watson-Hernández *et al*., 2022). AI can reliably forecast the anthesis stages of oil palm flowers, helping oil palm planters improve assisted pollination operations and overall productivity in oil palm plantations (Yousefi *et al*., 2021). Additionally, AI can predict future yield functions in oil palm plantations using Bayesian networks, with a study showing that AI can accurately forecast these future yield functions (Chapman *et al*., 2018). Furthermore, Ashrafian *et al*. (2023) explored another application, investigating the use of AI to predict the strength of agroecological lightweight concrete with palm oil by-products.

Among AI techniques, Artificial Neural Networks (ANN) have gained prominence in modelling forest products, with a series of studies focusing on species such as eucalyptus, pine and teak, primarily aimed at estimating height and wood volume (Binoti, 2010; Binoti *et al*., 2015; Gorgens *et al*., 2009; Loureiro, 2016). The widespread application of ANN in forest species, combined with robust results and consistent performance, has led to increasing acceptance in both the scientific community and the industry. These techniques have been used to optimise processes, enhance decision-making, and increase the efficiency of natural resource use. Thus, the expansion of ANN for modelling agricultural crops emerges as a promising alternative, offering opportunities to improve productivity and promote sustainable practices in the agricultural sector. This approach not only enhances the prediction of outcomes but also contributes to adaptation to environmental variables and effective management of crops.

ANN can accurately forecast the duration of palm oil production processes. This predictive capability has the potential to help palm oil mills optimise their production efficiency and reduce operational costs (Adizue *et al*., 2020). ANN have also been used to predict oil palm yield using meteorological variables and have been shown to be able to accurately predict oil palm yield (Kartika *et al*., 2016).

AI is a promising technology that has the potential to significantly improve palm oil management. By improving efficiency, sustainability and yield, AI can help the palm oil industry to meet the growing demand for palm oil while minimising its environmental impact. This method could be used by oil palm farmers and plantation managers to improve the yield and productivity of their crops.

Most studies on palm oil have been conducted in other countries, and there is no development of studies in Brazil. Therefore, it is important to conduct national-scale research to model the growth and development of the crop using AI tools. In this sense, the objective of this study was to predict the production of palm oil bunches using inventory and climatic variables through ANN.

MATERIALS AND METHODS

Localisation of the Study Area

The study area was a 107,000 ha property owned by Agropalma in the state of Pará, Brazil. It is in the northern region of the country, as shown in *Figure 1*.

Approximately 39,000 ha of the property are used for commercial seed plantations of the *tenera* variety of *Elaeis guineensis* (Jacq.). The remaining 64,000 ha are occupied by forest reserves and other infrastructure areas (Agropalma, 2019).

The predominant soil type in the region is Argissolo Amarelo Distrófico, which is variable in depth, acidic, well-drained and has a medium to clayey texture. In some areas, there is also Gleissolo Háplico dystrophic, which is shallow, acidic, hydromorphic, greyish in colour and sandy in texture (Rodrigues *et al*., 2005; Santos *et al*., 2018).

The climate in the Tailândia region is of the Am type, which is tropical monsoon. This means that the driest month has less than 60 mm of rainfall. The other three regions (Tomé-Açu, Acará and Moju) have an Af climate, which is tropical with no dry season. For both climate types, the average temperature of the coldest month of the year is above 18ºC (Alvares *et al*., 2013).

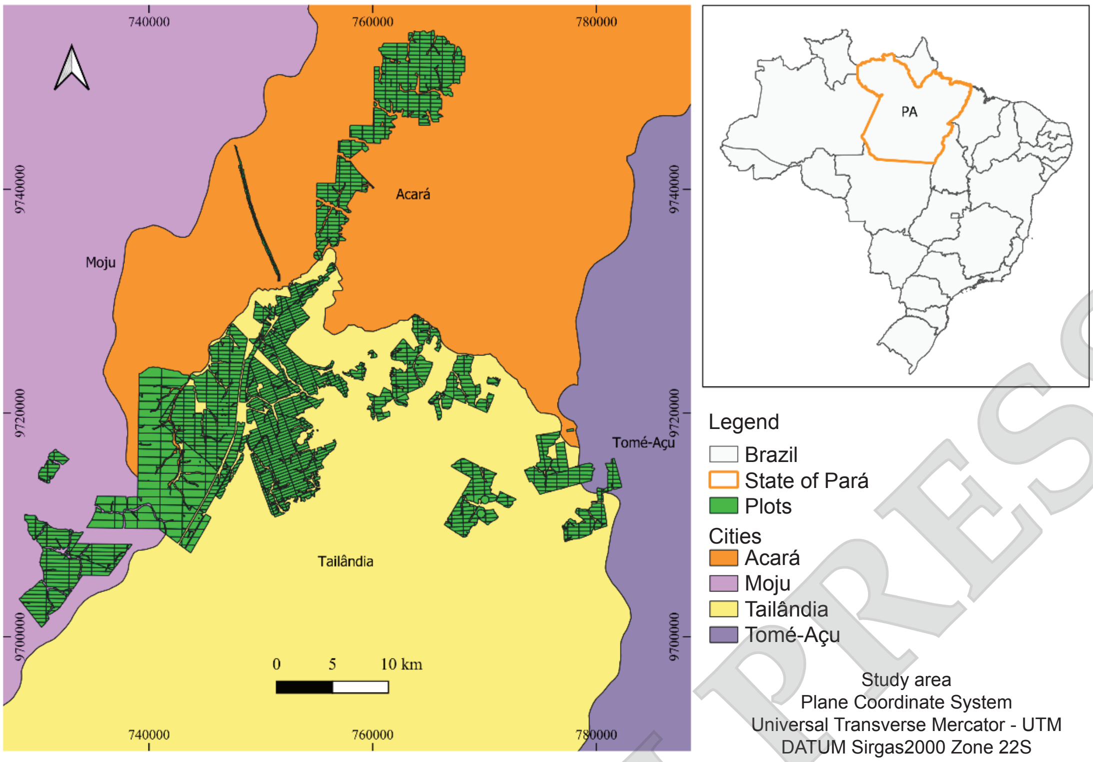

Figure 1. Location of the study area.

Characterisation of the Plantations and Organisation of the Database

Palm oil plantations were stratified by planting years and distributed across the A, B, C, D, E, F, G and H farms. Plantings were established on plots of varying sizes, ranging from 2-118 ha. Different genetic materials were used, with Deli X Lamé being the most common variety. Planting was also carried out with different spacings between plants, with a standard density of 143 plants/ha.

Forest inventories were conducted monthly, sampling 25-40 plants/plot. Inflorescences in anthesis (before fertilisation) and fecund (already fertilised), green bunches and mature bunches were quantified.

Four meteorological stations were installed at the A, C, D and F farms to collect monthly climate data, including temperature, precipitation, solar radiation and relative humidity. Additionally, the monthly water deficit was calculated.

To build the dataset, the forestry company provided historical monthly data from 2014-2022, including information from the forest inventory, total bunch production per plot, meteorological station records and general plantation records. This data was linked and checked, errors were corrected, rows without values for all attributes were deleted and outliers in bunch production, number of inflorescences in anthesis, fecund, green and mature bunches were removed. With the checked dataset, qualitative and quantitative variables were pre-defined for use in modelling the production of palm oil bunches.

Oil Palm Bunch Yield Prediction

The ANN with default configuration were applied to predict the production of oil palm bunches. An ANN consists of three layers: The input layer, where the information is received, the output layer, responsible for generating the outputs/responses and the hidden layer that connects both previous layers (*Figure 2*).

The prediction of the current monthly production of oil palm bunches was based on the values of climatic variables recorded in the months prior to the prediction month. This approach was taken because climatic variables for the current month were not yet available at the time of sampling, and previous climatic conditions have an impact on current bunch production.

Table 1 presents the minimum, mean, maximum, standard deviation (SD) and coefficient of variation (CV%) of the quantitative variables used in the study. These variables include the number of bunches per ha, the weight of bunches and the yield of oil per bunch. Qualitative variables encompass 8 farms, 26 strata based on planting years, 19 genetic materials and 7 soil classes.

TABLE 1. DESCRIPTIVE ANALYSIS OF QUANTITATIVE VARIABLES FOR PREDICTING THE YIELD OF PALM OIL BUNCHES

Quantitative variables	Acronym	Unit	Min	Mean	Max	SD	CV (%)
Anthesis inflorescences	AntHa	nº. ha^{-1}	0.00	4.32	102.96	7.91	182.90
Inflorescences in fecund	FecHa	nº. ha^{-1}	0.00	3.42	205.29	7.26	211.91
Green bunches	GBHa	nº. ha^{-1}	5.24	566.22	1,836.12	324.36	57.29
Mature bunches	MBHa	nº. ha^{-1}	1.79	96.63	1,024.00	95.92	99.26
Average bunch weight	BWCa	kg	2.02	14.81	84.06	5.79	39.11
Sampling percentage	%samp	%	0.50	1.06	14.63	0.77	72.69
Age	-	yr	4.00	15.33	33.00	8.08	52.69
Average temperature of the previous month	Temp1	°C	25.62	26.89	28.19	0.70	2.59
Precipitation of the previous month	PP1	mm	0.40	194.76	670.61	173.36	89.01
Water deficit of the previous month	Def1	mm	0.00	41.29	168.16	50.70	122.78
Relative humidity of the previous month	Humid1	%	67.91	84.8	92.86	5.74	6.77
Insolation of the previous month	Insol1	W m^{-2}	116.54	157.59	205.44	19.35	12.28
Monthly bunch production	ProdHa	t ha^{-1}	1.00	2.26	21.57	1.44	63.66

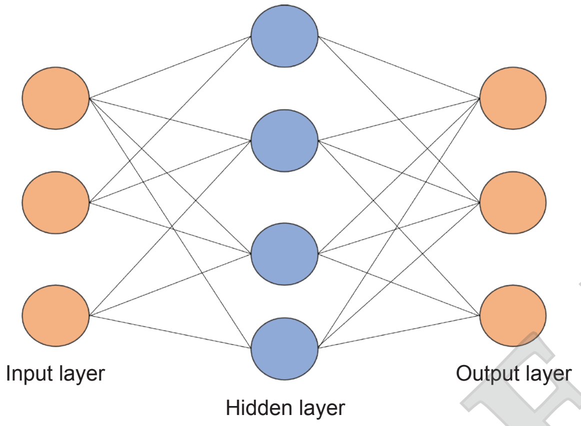

Figure 2. Structure of an artificial neural networks (ANN).

The prediction of the current monthly production of oil palm bunches relied on a comprehensive set of variables. These included general plot registration details such as farm, planting year, genetic material and soil. Also factored in were current month inventory metrics (AntHa, FecHa, GBHa, MBHa, BWCa), climatic conditions, and water deficit from the previous month (Temp1, PP1, Def1, Humid1, Insol1), as well as harvest month, age and sampling percentage, totalling 73 variables. These variables were input into an ANN trained to forecast bunches/ha production.

The output variable was the current monthly production of oil palm bunches per hectare (ProdHa) at the plot level. To achieve this, we employed Neuroforest 4.0 software (Binoti, 2012).

The ANN models were configured using standard settings provided by the software: Eight neurons in the hidden layer, a sigmoid activation function, resilient propagation as the training algorithm, a stopping criterion of 3,000 epochs and a data split of 70% for training and 30% for validation.

Using these ANN models, we estimated bunches per ha production for individual plots. Subsequently, we calculated total bunches production in t across various levels: Total t/plot/yr and respective harvest month (level 1); total t/yr and harvest month (level 2); total t/harvest month (level 3); and total t/yr of harvest (level 4), without specific plot breakdown, representing aggregate data. This process enabled assessment of the correlation between observed and estimated values, and determination of the average relative percentage error at each level.

Model Performance

After processing, statistical metrics were calculated to evaluate the performance of the artificial neural network defined for this study, considering both the training and validation phases, following the methods described by Campos and Leite (2017). Additionally, graphical representations were generated to visualise the results, including scatter plots illustrating the distribution of relative percentage errors, the relationship between estimated and observed values and histograms showing the frequency of relative percentage errors. These metrics include Linear correlation between predicted and observed values [Equation (1)], bias [Equation (2)], square root of the mean squared percentage error [Equation (3)], mean relative percentage error (Equation 4), percentage bias [Equation (5)] and variance [Equation (6)].

$$r_{y\hat{y}} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y}) \cdot (\hat{Y}_i - \overline{\hat{Y}})}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \cdot \sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})^2}} \qquad (1)$$

$$bias = \tfrac{1}{n} \sum_{i=1}^{n}(Y_i - \hat{Y}_i) \qquad (2)$$

$$RMSE\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{n}}}{\overline{Y}} \,.\, 100 \qquad (3)$$

$$\overline{ER}\% = \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{Y}_i-Y_i)\cdot 100}{Y_i} \qquad (4)$$

$$bias\% = \frac{bias}{\overline{Y}} \,.\, 100 \qquad (5)$$

$$Var = \frac{\sum[\text{bias} - (Y_i-\hat{Y}_i)]^2}{n-1} \qquad (6)$$

Where, Y_i is the observed value, $\hat{Y}_i$ is the estimated value, $\overline{Y}$ is the average of the observed values, $\overline{\hat{Y}}$ is the average of the estimated values and n is the total number of observations.

Identification of the Most Influential Variables in the Modelling Process

To understand the relationship between the input and output variables used in the study, a correlation matrix was generated. The correlation coefficient was used to measure the strength of the relationship between two variables, where a correlation coefficient of 1 indicated a perfect positive relationship, -1 indicated a perfect negative relationship and 0 indicated no relationship.

The correlation matrix was generated using the psych package (Revelle, 2023) in the R software (R Core Team, 2020). This package produced a correlation matrix with histograms of the distribution of each variable on the diagonal, correlations in the upper part of the diagonal and graphs illustrating the relationships between variables in the lower part.

Additionally, the Garson algorithm (Garson, 1991) was employed to assess the importance of each input variable in the model. This algorithm utilised the absolute values of weights between neurons in the input and hidden layers, as well as between the hidden and output layers, to quantify variable importance (Casas *et al*., 2022; Freitas *et al*., 2020).

A neural network, configured identically to the one used in the previous study, was trained using all available qualitative and quantitative variables. The weights derived from this training were then applied in the Garson algorithm. This process was executed using the R software and the Neuralnet package (Fritsch *et al*., 2019).

Additional Modelling Tests

Following the identification of variables most closely associated with bunch production per hectare, two additional modelling tests were conducted. These tests involved removing specific inventory and climatic elements from the dataset. The modified models were subsequently compared against the original model, and statistical analyses were performed to compute the correlation between observed and estimated values, as well as the mean relative percentage error.

RESULTS AND DISCUSSION

Prediction of palm oil bunch production

The ANN exhibited similar error statistics in both the training and validation phases, with a mean relative percentage error of approximately 13%, RMSE% around 40%, bias close to zero and variance around one, representing the prediction variability. However, when evaluating the correlation between observed and estimated values, there is a difference, with a correlation above 0.74 for training and 0.67 for validation (*Table 2*).

The scatter plots of the mean relative percentage error show that most of the errors range between -100% and 100%, particularly for productions between 1 and 5 t/ha. However, for lower production levels, some points begin to deviate from the overall trend, resulting in errors exceeding 100% (*Figure 3b, 3e*). The distribution of relative percentage errors is better visualised when grouped into classes (*Figure 3c, 3f*). Consequently, classes were defined ranging from -100%-100%, encompassing slightly over 80% of the errors falling between -40% and 60%, both in the training and validation datasets (*Figure 3*).

The concentration of the largest error ranges in lower bunch yields per hectare may be related to inadequate agricultural management practices, water stress and soil quality. These variables can significantly impact the productivity of oil palm plantations, resulting in inaccurate predictions. Understanding these factors is fundamental to

TABLE 2. STATISTICS CALCULATED TO ASSESS THE ACCURACY OF THE ARTIFICIAL NEURAL NETWORKS (ANN) IN THE TRAINING AND VALIDATION PHASES

Data splitting	Bias	Bias%	RMSE%	Var	$\overline{ER}\%$	$r_{y\hat{y}}$
Training	0.000001	0.000060	42.29	0.92	12.62	0.74
Validation	-0.009437	-0.420264	47.86	1.15	13.36	0.67

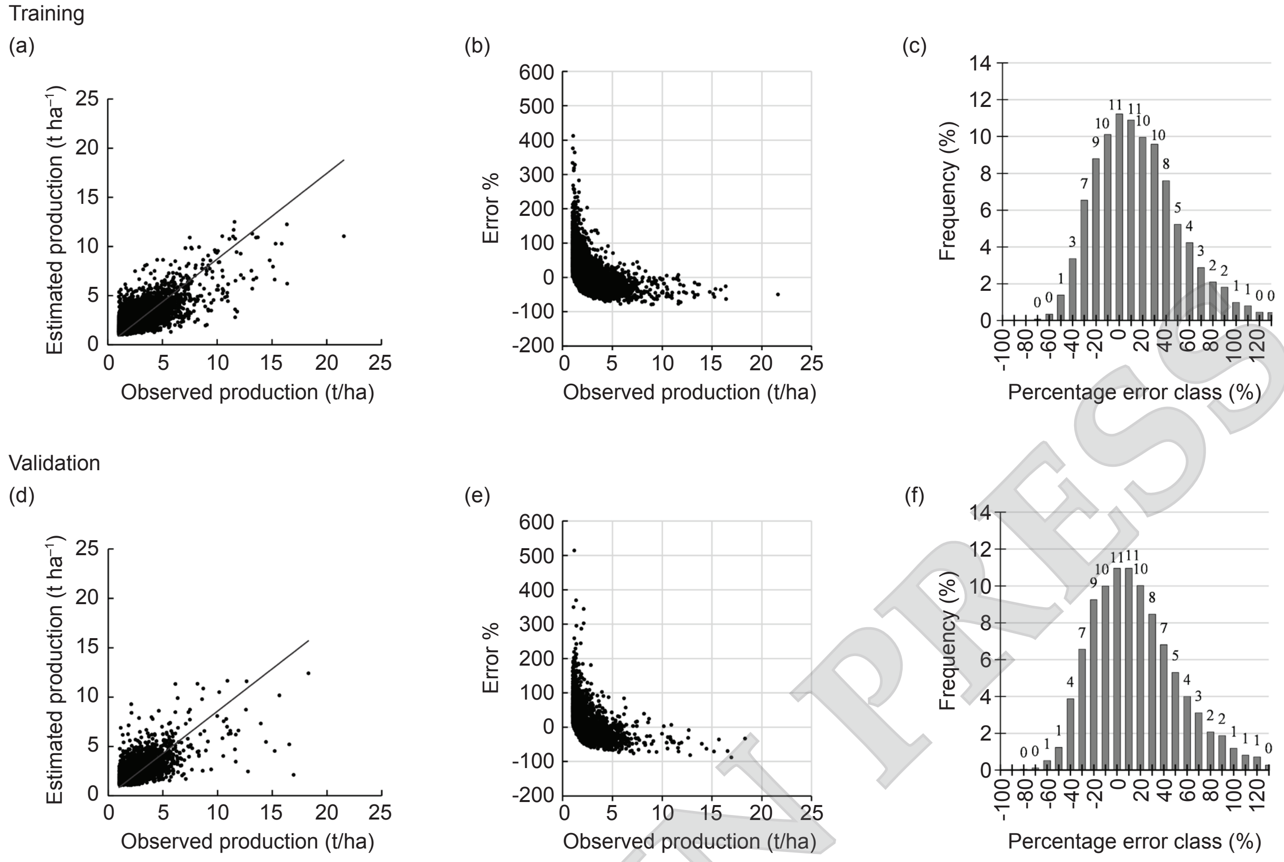

Figure 3. Relationships between (a, d) estimated and observed productions, (b, e) relative percentage errors and observed productions and (c, f) percentage of cases per class of relative percentage error in bunch production estimates.

identifying the limitations of the model, allowing for the analysis of the specific conditions affecting production in different contexts. Additionally, developing mitigation strategies, such as implementing more appropriate management practices, monitoring and improving soil fertility, can help enhance the accuracy of the predictions.

The behaviour of percentage errors concerning qualitative variables reveals the presence of outliers in most categories of each variable. For instance, the planting years 2001, 2009, 2010 and 2012, characterised by the largest planting areas, exhibited a significant number of outliers, meaning errors exceeding 100%. A similar pattern was observed for the farms; larger farms generated outliers, while smaller ones like B and H did not. Genetic materials such as Deli Lamé and Deli Yangambi, commonly found in most plantations, exhibited errors that deviated from the general trend, whereas the less represented Deli X Dami variety displayed errors ranging between 80% and 200% (*Figure 4*).

Regarding soil types, it is evident that the Argissolo (pad1, pad2, pad3, pad4 and pad5), which is predominant in the region, accounted for errors exceeding 100%. For both genetic materials and soil types, there is no consistent error in the behaviour pattern, implying the influence of other factors on these variables (*Figure 4*).

When evaluating the residual percentage dispersion in relation to quantitative variables, the sampling percentage deserves special attention, as it shows the highest errors at lower percentages, specifically between 0% and 1%. Inventory-related variables such as the number of anthesis, fecundation and mature bunches follow a similar pattern, with the ANN displaying greater errors at lower values. When conducting monthly inventories, in many cases, there is a minimal identification of bunches and inflorescences, or sometimes their presence is not even confirmed. However, this does not imply zero production for that month since production is quantified at a specific time. These minimal or null values do not directly represent the actual tonne of bunches delivered to the factory in the current month. Consequently, when relating these variables to the bunch production for the current month, the ANN tends to make more errors because it cannot establish a clear relationship. As for climatic variables, there is no distinct trend (*Figure 5*).

Evaluating the performance of estimates at different levels, there is an increase in correlation and a decrease in percentage errors as the estimates are assessed more broadly. This demonstrates the high accuracy of the ANN in estimating the total production for each month (0.99% error) as well

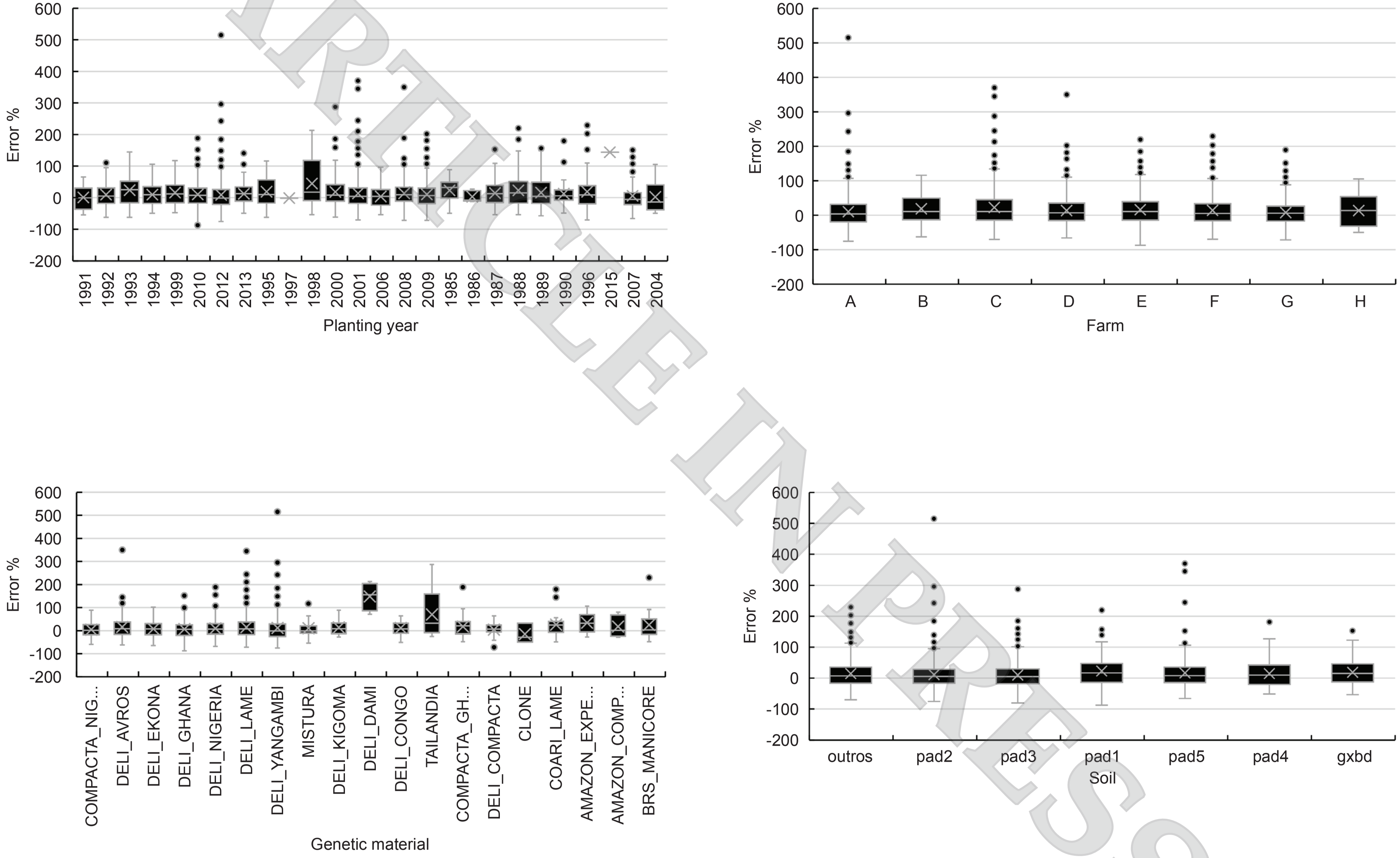

Figure 4. Scatterplot of relative percentage errors assessed by qualitative variable in the validation phase of the artificial neural networks (ANN).

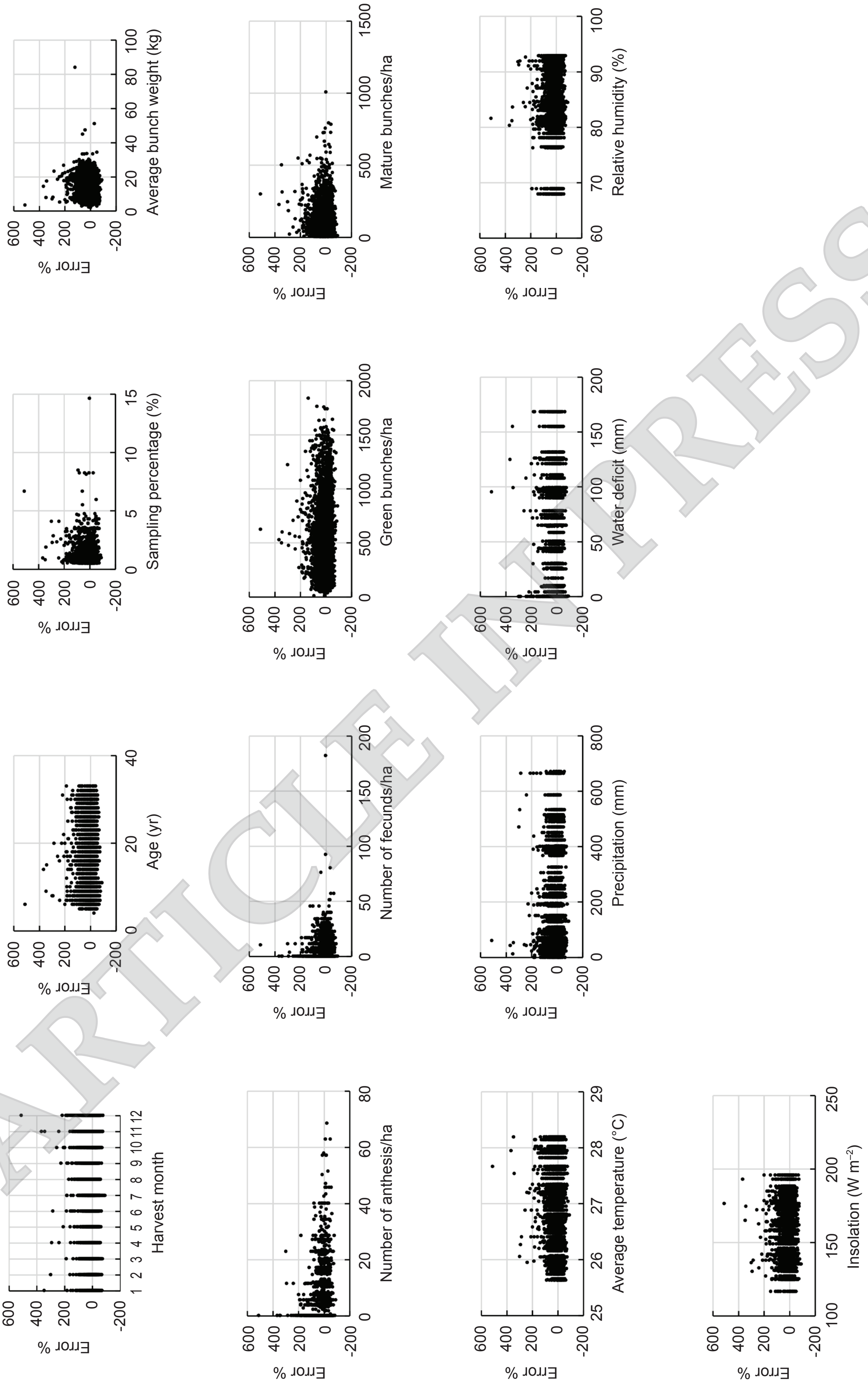

Figure 5. Scatterplot of relative percentage errors assessed by the quantitative variable in the validation phase of the artificial neural networks (ANN).

TABLE 3. ANALYSIS OF ESTIMATED TOTAL BUNCH PRODUCTION BY HARVEST YEAR AND MONTH

Total bunch production (t)	Level	Stratification	$r_{y\hat{y}}$	$\overline{ER}\%$
Harvest yr and month	1	Plot	0.75228	13.30%
Harvest yr and month	2	General	0.99103	6.30%
Harvest month	3	General	0.99725	1.90%
Harvest yr	4	General	0.99615	1.50%

as for each harvest year (1.50% error). Considering that one of the objectives is to obtain more reliable monthly bunch production estimates, the analysis at level 3 confirms the effectiveness of the modelling using ANN (*Table 3*).

Modelling at the parcel level is influenced by the sampling database, an operational procedure where the plot undergoes management and represents the main interfering factor in the estimation of precision (Khan *et al*., 2021; Nain *et al*., 2022). The lower precision of parcel level estimates is due to high variability, which becomes evident when analysing variables such as anthesis, fecundation, green bunches, mature bunches, sampling percentage and bunch production values over the years.

The data collection process for inventory is a significant challenge in the field due to the difficulty in visualising bunches and inflorescences, especially in the case of larger plants, leading to counting errors. This results in reduced reliability in the data collected in the field for consolidation and processing in the office. This issue has been discussed in a study by Chapman *et al*. (2018), which associated lower accuracy in fresh fruit bunches (FFB) yield estimates with errors in database records. Reducing the impact of sampling errors on modelling depends on changes in the planning of operational procedures, such as applying higher sampling percentages in areas with high heterogeneity (Soares *et al*., 2011). This enables the ANN to gain a deeper understanding of production variability and the factors that influence it.

Most Influential Variables in the Modelling

Based on a correlation coefficient threshold of 0.2 bunch yield/ha (ProdHa) has the strongest relationship with the following variables: Harvest month (Month), sampling percentage (%samp), mature bunches (MBHa), average temperature (Temp1), water deficit (Def1), precipitation (PP1) and relative humidity (Humid1) (*Figure 6*).

The Garson algorithm determines that the variables with the greatest contribution to the modelling are planting year, genetic material, soil, farm (qualitative), anthesis and fecundation inflorescences, mature bunches, average bunch weight and sampling percentage (quantitative).

The behaviour of the estimates in each additional modelling test is represented in *Table 4*, which summarises the variables used (marked with an X), the calculated correlation and mean relative error statistics. It is observed that the two new tests, Test 1 and Test 2, which excluded some input variables from modelling, exhibit similar performance to Test 0, which utilises all input variables, demonstrating that certain variables, such as climatic variables, did not exert a significant influence on the model's performance.

TABLE 4. INPUT VARIABLES USED IN EACH TEST AND THEIR RESPECTIVE MODELING STATISTICS (CORRELATION, MEAN RELATIVE PERCENTAGE ERROR) FOR TRAINING (T) AND VALIDATION (V)

Variables		Test 0	Test 1	Test 2
Planting year		X	X	X
Farm		X	X	X
Genetic material		X	X	X
Soil		X	X	X
Month		X	X	X
Age		X		
%samp		X	X	X
AntHa		X		
FecHa		X		
GBHa		X		
MBHa		X	X	X
BWCa		X		
Temp1		X	X	
PP1		X		
Def1		X	X	
Humid1		X		
Insol1		X	X	
$r_{y\hat{y}}$	T	0.74	0.71	0.71
	V	0.67	0.67	0.62
$\overline{ER}\%$	T	12.62	13.38	13.68
	V	13.36	14.66	15.92

Both the Garson algorithm and the conducted tests indicated that climatic variables did not exert a significant influence on the estimates. Test 2, which excluded all climatic variables, obtained statistically similar results to the other

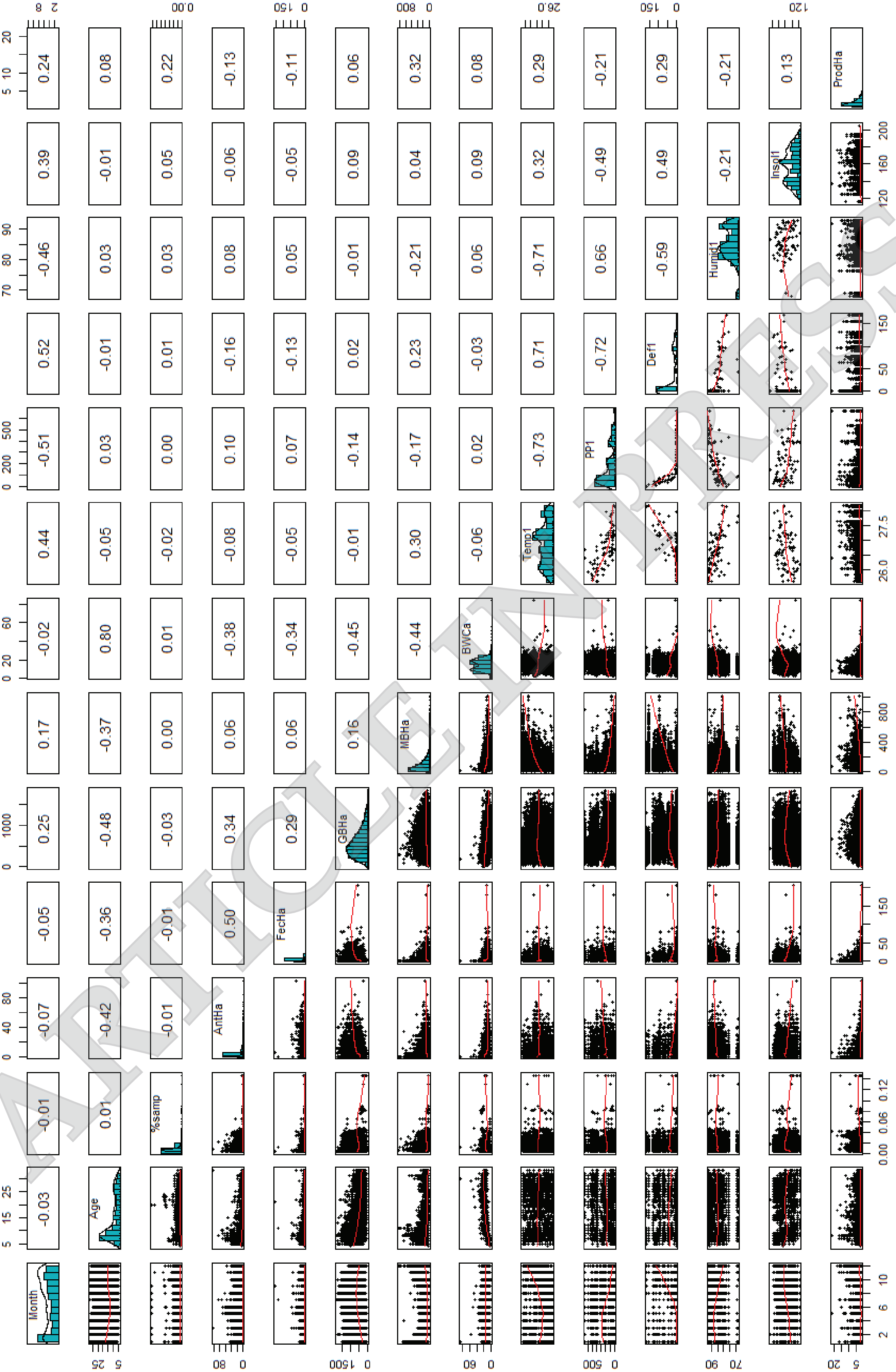

Figure 6. Pearson correlation among the quantitative variables used in the modeling.

tests (*Table 4*). This result may suggest that the climatic variation between months was not significant enough to affect the ANN learning or that the quality of the climatic database was suboptimal. This is because meteorological stations experienced technical issues during certain periods, reducing the reliability of the collected variables. Therefore, it is essential to invest in equipment and techniques to enhance the quality of available climatic data (Hoffmann *et al*., 2014, 2017; Kartika *et al*., 2016). Additionally, using the climatic factor from only the month preceding the prediction month was not sufficient to express the climate's influence on oil palm bunch production. The development of inflorescences occurs 32 months after seed germination, and during this period, the climatic influence accumulates (Adam *et al*., 2005). It would be valuable in future studies to apply climatic variables with lags and cumulative periods of up to 32 months before the prediction month.

The influence of climatic variables on the productivity of oil palm plantations is complex, as different ranges of climate conditions can have varying effects. While adequate levels of rainfall can promote healthy growth, excessive or insufficient precipitation may cause stress, leading to negative impacts on productivity. This duality highlights the need for more sophisticated approaches in analyses that consider the variability of crop responses to climate change. Understanding these nuances is crucial for optimising oil palm production under changing climatic conditions (Fleiss *et al*., 2022).

Maintaining the collection and use of climatic variables is crucial for oil palm modelling studies, given the climate's influence on the crop development. In the case of random events, the inclusion of variables such as temperature, solar radiation, precipitation, or water deficit becomes indispensable. Water deficit, for instance, is the primary limiting factor for bunch production as it delays floral initiation, increases abortion rates and consequently reduces fruit production (Corley & Tinker, 2016; Monzon *et al*., 2022; Murugesan *et al*., 2017; Pirker *et al*., 2016; Rhebergen *et al*., 2016). A similar effect is observed during periods of high precipitation, which has a negative correlation with FFB production and oil production in the bunch. Excessive rainfall during inflorescence development leads to flower abortion and subsequently lower bunch and oil production, affecting the industrial extraction rate (Silva, 2006).

Oil palm is a crop that requires specific climatic conditions, and if not met, it results in low annual yields. The Amazon region, where this study is conducted, is characterised by annual average temperatures between 24°C and 28°C, water deficit below 200 mm, relative humidity between 75% and 90%, precipitation exceeding 100 mm, and high levels of solar radiation (1,600 hr/yr), conditions favourable for oil palm establishment in the region. However, constant climate change tends to alter normal conditions and consequently, crop yields.

Using strata such as planting years and farms in modelling provides specific information about plantation management and reflects environmental changes directly affecting these strata. These strata generally exhibit a pattern of production peaks throughout the harvest year, which occurs in January, November and December, coinciding with peaks in average temperature and insolation. These climatic factors, along with precipitation, significantly influence oil palm productivity (Corley & Tinker, 2016; Hoffmann *et al*., 2014; Monzon *et al*., 2022). Monzon *et al*. (2022) assessed the variation in oil palm productivity in terms of space (plantation strata) and time (year and month) in the Indonesian region. In their study, monthly bunch yield peaks differed in the four regions evaluated, occurring at the beginning (January), middle (May-August), or end of the year (November). These yields were closely associated with water stress. Thus, the monthly bunch production of oil palm depends on specific aspects of the region and the local climatic behaviour.

The importance of categorical variables such as soil type and genetic material aligns with research conducted outside Brazil, as discussed in the review by Nain *et al.* (2022). They categorised variables into three categories: Environment, phenotype and genotype. They highlighted genotype as a widely used factor with soil and fertilisation as the second most addressed environmental factor in the studies, second only to climate.

Among the inventory variables, the identification and counting of mature bunches are essential for modelling due to their direct relationship with oil palm yield (Monzon *et al*., 2022). After harvesting the bunches, they must be processed as quickly as possible to avoid high acidity in the extracted oil and consequently, a loss of final product quality (Corley & Tinker, 2016; Muller *et al*., 2006; Teles, 2014). Typically, the classification of fresh fruit bunch is done manually based on the subjective judgment of the responsible person, influenced by the palm tree's height, fruit position and direction of sunlight. Additionally, the number of fruits found on the ground is observed to confirm bunch maturity. This method is not very efficient as it cannot capture the internal quality of the fruit, resulting in a high error rate that directly affects the prediction of palm oil yield (Alfatni *et al*., 2020, 2018; Ishak *et al*., 2019; Makky & Cherie, 2021). Therefore, the use of alternative and accurate methods for classifying mature bunches is crucial for estimating the t of bunches/planted area, such as image-based classification.

Study Limitations

The developed prediction model relies heavily on the quality and quantity of available historical data. The lack of accurate data related to climatic conditions, management practices and soil characteristics may compromise the model's effectiveness, resulting in less reliable predictions. This dependence on historical data is a significant limitation, as the absence of comprehensive information may lead to either underestimation or overestimation of production.

Another point to highlight is the complexity of the environmental factors influencing oil palm production. The interaction between these factors is intricate, and the model may not fully capture these interactions. Consequently, the predictions generated may not adequately reflect the dynamics of the cultivation system, limiting the accuracy of the estimates.

Furthermore, the generalisation of the model to different regions or cultivation conditions may be limited. Although the study has shown satisfactory results, local factors, such as microclimatic variations and specific agricultural practices, present high variability, which may affect the model's applicability in different contexts.

CONCLUSION

This study demonstrated that ANN are effective in estimating the current monthly production of oil palm bunches. By integrating variables from forest inventory, climatic elements, water deficit, soil characteristics and plantation management records, the methodology proved to be efficient and achieved the proposed objective of accurately predicting production. These findings highlight the practical potential of ANN as a reliable tool for inventory and production planning in oil palm crops. For future work, it is recommended to adopt advanced techniques that ensure proper data treatment, particularly concerning outliers and missing information. Incorporating lagged climatic variables over extended periods is an excellent strategy for improving model performance, as oil palm productivity is highly influenced by climatic factors. Additionally, incorporating data related to soil fertilisation and management practices is crucial for a more comprehensive modelling approach. In the context of Brazil, where the study was conducted and which has significant potential for oil palm expansion, this study represents a significant advancement in the development of management techniques for oil palm plantations. Moreover, it opens up opportunities for other researchers to explore more advanced investigations in the sector, testing and incorporating new variables that could explain variations in palm productivity, thereby contributing to the robustness and applicability of the model.

ACKNOWLEDGEMENT

The present work was carried out with support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Financing Code 001 and the company Agropalma, which provided the production data for the crop under study.

REFERENCES

Adam, H., Jouannic, S., Escoute, J., Duval, Y., Verdeil, J. L., & Tregear, J. W. (2005). Reproductive developmental complexity in the African oil palm (*Elaeis guineensis*, Arecaceae). *American Journal of Botany, 92*(11), 1836–1852. https://doi.org/10.3732/ajb.92.11.1836

Adizue, U. L., Nwanya, S. C., & Ozor, P. A. (2020). Artificial neural network application to a process time planning problem for palm oil production. *Engineering and Applied Science Research, 47*(2), 161–169. https://ph01.tci-thaijo.org/index.php/easr/article/view/211130

Agropalma. (2019). *Relatório de sustentabilidade* [Sustainability report]. Helikonia.

Akhtar, M. N., Ansari, E., Alhady, S. S. N., & Abu Bakar, E. (2023). Leveraging on advanced remote sensing- and artificial intelligence-based technologies to manage palm oil plantation for current global scenario: A review. *Agriculture, 13*(2), 504. https://doi.org/10.3390/agriculture13020504

Alfatni, M. S. M., Mohamed Shariff, A. R., Bejo, S. K., Ben Saaed, O. M., & Mustapha, A. (2018). Real-time oil palm FFB ripeness grading system based on ANN, KNN and SVM classifiers. *IOP Conference Series: Earth and Environmental Science, 169*, 012067. https://doi.org/10.1088/1755-1315/169/1/012067

Alfatni, M. S. M., Mohamed Shariff, A. R., Ben Saaed, O. M., Albhbah, A. M., & Mustapha, A. (2020). Colour feature extraction techniques for real time system of oil palm fresh fruit bunch maturity grading. *IOP Conference Series: Earth and Environmental Science, 540*(1), 012092. https://doi.org/10.1088/1755-1315/540/1/012092

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Ashrafian, A., Panahi, E., Salehi, S., Karoglou, M., & Asteris, P. G. (2023). Mapping the strength of agro-ecological lightweight concrete containing oil palm by-product using artificial intelligence techniques. *Structures, 48*, 1209–1229. https://doi.org/10.1016/j.istruc.2022.12.108

Binoti, M. L. M. S. (2010). *Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto* [Artificial neural networks for predicting the yield of unthinned eucalyptus stands] (Master's thesis). Universidade Federal de Viçosa.

Binoti, M. L. M. S. (2012). *Use of artificial neural networks in measurement and forest management* (Ph.D thesis). Universidade Federal de Viçosa.

Binoti, M. L. M. S., Leite, H. G., Binoti, D. H. B., & Gleriani, J. M. (2015). *Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais* [Stand-level growth prediction of eucalyptus clones using artificial neural networks]. *Cerne, 21*(1), 97–105. https://doi.org/10.1590/01047760201521011153

Briassoulis, H. (2020). *Analysis of land use change: Theoretical and modelling approaches* (2nd ed.). WVU Research Repository.

Campos, J. C. C., & Leite, H. G. (2017). *Forest measurement: Questions and answers* (5th ed.). UFV.

Casas, G. G., Fardin, L. P., Silva, S., Oliveira Neto, R. R., Binoti, D. H. B., Leite, R. V., Domiciano, C. A. R., Sousa Lopes, L. S., Cruz, J. P., Reis, T. L., & Leite, H. G. (2022). Improving yield projections from early ages in eucalypt plantations with the Clutter model and artificial neural networks. *Pertanika Journal of Science & Technology, 30*(2), 1257–1272. https://doi.org/10.47836/pjst.30.2.22

Chapman, R., Cook, S., Donough, C., Lim, Y. L., Vun Vui Ho, P., Lo, K. W., & Oberthür, T. (2018). Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: A proof of concept analysis. *Computers and Electronics in Agriculture, 151*, 338–348. https://doi.org/10.1016/j.compag.2018.06.006

Corley, R. H. V., & Tinker, P. B. (2016). *The oil palm* (5th ed.). BlackWell Science Ltd.

Drost, S., Rijk, G., & Piotrowski, M. (2021). *Palm oil production in Brazilian Amazon threatens NDPE commitments*.Chain Reaction Research.

Fleiss, S., McClean, C. J., King, H., & Hill, J. K. (2022). Limited impacts of climatic conditions on commercial oil palm yields in Malaysian plantations. *CABI Agriculture and Bioscience, 3*(59), 1–18. https://doi.org/10.1186/s43170-022-00127-1

Freitas, E. C. S. de, Paiva, H. N., Neves, J. C. L., Marcatti, G. E., & Leite, H. G. (2020). Modeling of eucalyptus productivity with artificial neural networks. *Industrial Crops and Products, 146*, 112149. https://doi.org/10.1016/j.indcrop.2020.112149

Fritsch, S., Guenther, F., & Wright, M. N. (2019). *neuralnet: Training of neural networks* [R package]. https://CRAN.R-project.org/package=neuralnet

Garson, G. D. (1991). Interpreting neural network connection weights. *AI Expert, 6*(4), 46–51.

Gorgens, E. B., Leite, H. G., Santos, H. N., & Gleriani, J. M. (2009). *Estimação do volume de árvores utilizando redes neurais artificiais* [Estimation of tree volume using artificial neural networks]. *Revista Árvore, 33*(6), 1141–1147.

Harsawardana, Rahutomo, R., Mahesworo, B., Cenggoro, T. W., Budiarto, A., Suparyanto, T., Atmaja, D. B. S., Samoedro, B., & Pardamean, B. (2020). AI-based ripeness grading for oil palm fresh fruit bunch in smart crane grabber. *IOP Conference Series: Earth and Environmental Science, 426*(1), 012147. https://doi.org/10.1088/1755-1315/426/1/012147

Hoffmann, M. P., Donough, C. R., Cook, S. E., Fisher, M. J., Lim, C. H., Lim, Y. L., Cock, J., Kam, S. P., Mohanaraj, S. N., Indrasuara, K., Tittinutchanon, P., & Oberthür, T. (2017). Yield gap analysis in oil palm: Framework development and application in commercial operations in Southeast Asia. *Agricultural Systems, 151*, 12–19. https://doi.org/10.1016/j.agsy.2016.11.005

Hoffmann, M. P., Vera, A. C., van Wijk, M. T., Giller, K. E., Oberthür, T., Donough, C., & Whitbread, A. M. (2014). Simulating potential growth and yield of oil palm (*Elaeis guineensis*) with PALMSIM: Model description, evaluation and application. *Agricultural Systems, 131*, 1–10. https://doi.org/10.1016/j.agsy.2014.07.006

Ishak, H., Shiddiq, M., Fitra, R. H., & Yasmin, N. Z. (2019). Ripeness level classification of oil palm fresh fruit bunch using laser induced fluorescence imaging. *Journal of Aceh Physics Society, 8*(3), 84–89. https://doi.org/10.24815/jacps.v8i3.14139

Kartika, N. D., Astika, I. W., & Santosa, E. (2016). Oil palm yield forecasting based on weather variables using artificial neural network. *Indonesian Journal of Electrical Engineering and Computer Science, 3*(3), 626–633. https://doi.org/10.11591/ijeecs.v3.i3.pp626-633

Khan, N., Kamaruddin, M. A., Sheikh, U. U., Yusup, Y., & Bakht, M. P. (2021). Oil palm and machine learning: Reviewing one decade of ideas, innovations, applications, and gaps. *Agriculture, 11*(9), 832. https://doi.org/10.3390/agriculture11090832

Loureiro, G. H. (2016). *Redes neurais artificiais e modelos de regressão para estimação de variáveis dendrométricas em Pinus taeda L: Análise comparativa* [Artificial neural networks and regression models for estimating dendrometric variables in *Pinus taeda* L: A comparative analysis] (PhD thesis). Universidade Federal do Paraná.

Makky, M., & Cherie, D. (2021). Pre-harvest oil palm FFB nondestructive evaluation technique using thermal-imaging device. *IOP Conference Series: Earth and Environmental Science, 757*(1), 012003. https://doi.org/10.1088/1755-1315/757/1/012003

Monzon, J. P., Jabloun, M., Cock, J., Caliman, J. P., Couëdel, A., Donough, C. R., Vui, P. H. V., Lim, Y. L., Mathews, J., Oberthür, T., Prabowo, N. E., Edreira, J. I. R., Sidhu, M., Slingerland, M. A., & Sugianto, H. (2022). Influence of weather and endogenous cycles on spatiotemporal yield variation in oil palm. *Agricultural and Forest Meteorology, 314*, 108789. https://doi.org/10.1016/j.agrformet.2021.108789

Muller, A. A., & Furlan, J. J. (2006). *A Embrapa Amazônia Oriental e o agronegócio do dendê no Pará* [Eastern Amazon Embrapa and the oil palm agribusiness in Pará]. Belém.

Murphy, D. J., Goggin, K., & Paterson, R. R. M. (2021). Oil palm in the 2020s and beyond: Challenges and solutions. *CABI Agriculture and Bioscience, 2*(1). https://doi.org/10.1186/s43170-021-00058-3

Murugesan, P., Aswathy, G. M., Kumar, K. S., Masilamani, P., Kumar, V., & Ravi, V. (2017). Oil palm (*Elaeis guineensis*) genetic resources for abiotic stress tolerance: A review. *Indian Journal of Agricultural Sciences, 87*(5), 571-579. https://doi.org/10.56093/ijas.v87i5.70087

Nain, F. N. M., Malim, N. H. A. H., Abdullah, R., Rahim, M. F. A., Mokhtar, M. A. A., & Fauzi, N. S. M. (2022). A review of an artificial intelligence framework for identifying the most effective palm oil prediction. *Algorithms, 15*(6), 218. https://doi.org/10.3390/a15060218

Organização Internacional do Trabalho (OIT). (2022). *Cadeia produtiva do óleo de palma – Avanços e desafios rumo à promoção do trabalho decente: Análise situacional* [Palm oil value chain – Advances and challenges toward promoting decent work: Situational analysis].

Pirker, J., Mosnier, A., Kraxner, F., Havlík, P., & Obersteiner, M. (2016). What are the limits to oil palm expansion? *Global Environmental Change, 40*, 73–81. https://doi.org/10.1016/j.gloenvcha.2016.06.007

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research* (R package version). https://CRAN.R-project.org/package=psych

Rhebergen, T., Fairhurst, T., Zingore, S., Fisher, M., Oberthür, T., & Whitbread, A. (2016). Climate, soil and land-use based land suitability evaluation for oil palm production in Ghana. *European Journal of Agronomy, 81*, 1–14. https://doi.org/10.1016/j.eja.2016.08.004

Rodrigues, T. E., Silva, J. M. L., Silva, B. N. R., Valente, M. A., Gama, J. R. N. F., Santos, E. S., Rollim, P. A. M., & Ponte, F. C. (2005). *Caracterização e classificação dos solos do município de Tailândia, Estado do Pará* [Characterization and classification of soils in the municipality of Tailândia, Pará State]. Embrapa Amazônia Oriental.

Santos, H. G., Jacomine, P. K. T., Anjos, L. H. C., Oliveira, V. A., Lumbreras, J. F., Coelho, M. R., Almeida, J. A., Araújo Filho, J. C., Oliveira, J. B., & Cunha, T. J. F. (2018). *Sistema brasileiro de classificação de solos* [Brazilian soil classification system]. Embrapa Solos.

Silva, J. S. O. (2006). *Produtividade de óleo de palma na cultura do dendê na Amazônia Oriental: Influência do clima e do material genético* [Palm oil productivity in oil palm cultivation in the Eastern Amazon:

Influence of climate and genetic material] (Master's thesis). Universidade Federal de Viçosa.

Soares, C. P. B., Paula Neto, F., & Souza, A. L. (2011). *Dendrometria e inventário florestal* [Forest mensuration and inventory] (2nd ed.). Universidade Federal de Viçosa.

Statista. (2023). *Brazil: Palm oil consumption 2008–2023*. https://www.statista.com/statistics/876584/palm-oil-consumption-brazil/

Teles, D. A. A. (2014). *Características físicas e rendimento mensal em óleo de cachos de duas cultivares de dendezeiro cultivadas, sob irrigação, no Cerrado do Distrito Federal* [Physical characteristics and monthly oil yield of bunches from two oil palm cultivars grown under irrigation in the Cerrado of the Federal District] (Master's thesis). Universidade de Brasília.

U.S. Department of Agriculture. (2022). *Brazil palm oil: Potential and pitfalls* (Commodity Intelligence Report). Foreign Agricultural Service. https://ipad.fas.usda.gov/highlights/2022/12/Brazil_Palm_Oil/

Watson-Hernández, F., Gómez-Calderón, N., & Silva, R. P. (2022). Oil palm yield estimation based on vegetation and humidity indices generated from satellite images and machine learning techniques. *AgriEngineering, 4*(1), 279–291. https://doi.org/10.3390/agriengineering4010019

Yousefi, M. D. B., Mohd Rafie, A. S., Abd Aziz, S., Azrad, S., Mohd Masri, M. M., Shahi, A., & Marzuki, O. F. (2021). Classification of oil palm female inflorescences anthesis stages using machine learning approaches. *Information Processing in Agriculture, 8*(4), 537–549. https://doi.org/10.1016/j.inpa.2020.11.007