



OPEN ACCESS

EDITED BY

Weimin Ma,
Shanghai Normal University, China

REVIEWED BY

Muhammad Faheem Jan,
Northeast Agricultural University, China
Marko Kolaksazov,
Institute of Forage Crops (Bulgaria), Bulgaria
Zhenfeng Yang,
Ningxia University, China

*CORRESPONDENCE

Isabell Pappert
✉ isabell.pappert@tu-darmstadt.de
Stefan Klir
✉ klir@lichttechnik.tu-darmstadt.de

RECEIVED 10 March 2025

ACCEPTED 21 April 2025

PUBLISHED 19 May 2025

CITATION

Pappert I, Klir S, Jokic L, Ühlein C,
Tran Quoc K and Kaldenhoff R (2025)
Plant photosynthesis in basil (C3) and
maize (C4) under different light conditions
as basis of an AI-based model for PAM
fluorescence/gas-exchange correlation.
Front. Plant Sci. 16:1590884.
doi: 10.3389/fpls.2025.1590884

COPYRIGHT

© 2025 Pappert, Klir, Jokic, Ühlein, Tran Quoc
and Kaldenhoff. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Plant photosynthesis in basil (C3) and maize (C4) under different light conditions as basis of an AI-based model for PAM fluorescence/ gas-exchange correlation

Isabell Pappert^{1*}, Stefan Klir^{2*}, Luca Jokic¹, Celine Ühlein¹,
Khanh Tran Quoc² and Ralf Kaldenhoff¹

¹Department of Applied Plant Sciences, Faculty of Biology, Technical University of Darmstadt, Darmstadt, Germany, ²Laboratory of Adaptive Lighting Systems and Visual Processing, Technical University of Darmstadt, Darmstadt, Germany

Photosynthetic activity can be monitored using pulse amplitude modulated (PAM) fluorescence or gas exchange. While PAM provides insight into the light-dependent reactions, gas exchange reflects CO₂ fixation and water balance. Accurate, non-invasive prediction of photosynthetic performance under varying conditions is highly relevant for phenotyping and stress diagnostics. Despite their physiological link, data from both methods do not always correlate. To systematically investigate this relationship, photosynthetic parameters were measured in maize (*Zea mays*, C4) and basil (*Ocimum basilicum*, C3) under different photon densities and spectral compositions. Maize showed the highest CO₂ assimilation rate of 30.99 ± 1.54 μmol CO₂/(m²s) under 2000 PAR green light (527 nm), while basil reached 10.56 ± 0.92 μmol CO₂/(m²s) under red light (630 nm). PAM-derived electron transport rates (ETR) increased with light intensity in a pattern similar to CO₂ assimilation, but did not reliably reflect its absolute values under all conditions. To improve prediction accuracy, we applied a machine learning model. XGBoost, a gradient-boosted decision tree algorithm, efficiently captures nonlinear interactions between physiological and environmental parameters. It achieved superior performance (R² = 0.847; MSE = 5.24) compared to the Random Forest model. Our model enables accurate photosynthesis prediction from PAM data across light intensities and spectral conditions in both C3 and C4 plants.

KEYWORDS

chlorophyll fluorescence, gas exchange, machine learning, photosynthesis prediction, C3/C4 plants

Introduction

Chlorophyll fluorescence analysis by pulse-amplitude modulated (PAM) fluorometry and gas exchange measurements are essential techniques for investigating the mechanisms underlying photosynthesis. The PAM technique allows for a detailed analysis of parameters such as maximum photosynthetic efficiency (Fv/Fm), photochemical efficiency of Photosystem II (Φ PSII), and related terms (Kornyeyev and Holaday, 2008).

Fv/Fm represents the ratio of variable fluorescence (Fv) to maximum fluorescence (Fm) and indicates the maximum efficiency of Photosystem II (PSII) under optimal light conditions. It shows how effectively light energy is converted into chemical energy. Φ PSII is the quantum yield of the photosynthetic electron transport chain under a given light condition. Accordingly, Φ PSII values reflect the competence with which the plant uses absorbed light for electron transport, which ultimately contributes to the synthesis of NADPH and ATP.

While PAM measurements primarily assess the efficiency of energy conversion in the thylakoid membranes, gas exchange measurements directly assess the diffusion of CO₂ as well as that of water vapor. Key data collected from gas exchange analysis include photosynthetic rate, transpiration rate, and stomatal conductance. These provide information about net carbon gain and stomatal regulation (Long and Bernacchi, 2003). Even though both techniques are related to photosynthesis they do not monitor the same biochemical reaction. PAM focuses on photochemical reactions at the level of PSII, whereas gas exchange reflects the integrated result of downstream metabolic activity, particularly the fixation of carbon through the Calvin cycle. The two methods are functionally linked, as the energy generated in the light reactions drives the dark reactions. Optimal photochemistry is a prerequisite for effective carbon assimilation. Conversely, impaired light reactions can reduce the flux into carbon metabolism. However, discrepancies arise because several intermediate and downstream processes—such as Rubisco activity, mesophyll conductance, or photorespiration—can modulate the link between photochemical and net carbon fixation (Flexas et al., 2012). This decoupling becomes particularly evident under stress conditions such as high light, where PAM fluorescence may overestimate photosynthetic capacity due to ongoing electron transport despite limited CO₂ fixation (Baker and Rosenqvist, 2004; Flexas and Medrano, 2002; Genty et al., 1989). If a calculation of the relationship between PAM and gas exchange measurements would be feasible, these could significantly reduce the time required for photosynthetic assessments because PAM measurements are quick and less time-consuming in comparison to gas exchange measurements. PAM fluorometry enables fast and non-invasive measurements of photosynthetic activity, with fluorescence parameters such as Fv/Fm or Φ PSII typically measured within seconds to a few minutes. In contrast, gas exchange measurements are often more time-intensive, requiring stabilization of measurement conditions and taking several minutes to up to an hour per sample. Portable PAM fluorometers can be used both in the field as well as laboratory and provide data within a few minutes, whereas gas exchange

measurements are more labor-intensive, requiring specialized equipment, which is substantially larger and data observation requires more time (Baker, 2008). In case of a determinable relation between values from PAM and gas exchange, i.e. using Φ PSII to predict the efficiency of CO₂ fixation, PAM measurements can be utilized to determine photosynthetic rates. Under optimal conditions there is a good correlation between Φ PSII and the CO₂ assimilation rate as the electrons produced in PSII are mainly used for CO₂ fixation (Lysenko et al., 2022). This makes PAM fluorescence measurements a meaningful indicator for photosynthetic activity in different plant species (Maxwell and Johnson, 2000) particularly under moderate light intensities. Genty et al., showed that the quantum yield of PSII electron transport correlates well with carbon assimilation when there are no additional active electron sinks. This relationship is more stable under optimal conditions with moderate light (200 to 400 PAR (Genty et al., 1989)). It is the basis for using PAM measurements to estimate the rate of CO₂ fixation. However, it is increasingly evident that these predictions considerably lose accuracy under non-optimal conditions. For example, Genty et al. (1989) also showed that the correlation between Φ PSII and the gas exchange decreases notably at high light intensities (up to 600 PAR) as alternative electron flows are activated that do not contribute to CO₂ fixation. The measured photosynthetic rates do not match with the PAM predicted assimilation rate, particularly under high light irradiation. It is explained by excess energy in the photosystem that can no longer be used efficiently and flows into alternative protection mechanisms (Genty et al., 1989). Similar results were reported by Flexas and Medrano (2002) under water deficit conditions. Stomatal conductance regulates the entry of CO₂ into the leaf. It was reduced by water deficiency more than the capacity to transport electrons. Consequently, a high electron transport rate was measured by PAM, even though the gas exchange was severely restricted. This is due to the water stress induced reduction in stomatal conductance which decreases substomatal CO₂ concentrations. The CO₂ limitation is apparently not detected by PAM (Flexas and Medrano, 2002). Baker and Rosenqvist (2004) showed that under severe stress, such as high temperature, electron flow rates remain high, while the CO₂ exchange is severely restricted. Heat damage can disrupt the function of enzymes such as Rubisco while electron transport in Photosystem II is maintained by alternative mechanisms and detected by PAM. Under these conditions, current CO₂ assimilation is inefficient (Baker and Rosenqvist, 2004). Under conditions of low CO₂ concentrations or low temperatures, conventional models also fail (Murchie and Lawson, 2013). Here the assimilation rate depends on enzyme-activity rather than electron transport.

In summary, PAM parameter and assimilation rates are decoupled to a certain extent under the conditions exemplarily mentioned above (Murchie and Lawson, 2013). The current photosynthetic activity is overestimated, as alternative electron fluxes or limitations in stromal processes can affect the results (Baker, 2008; Flexas and Medrano, 2002; Genty et al., 1989). Suboptimal conditions include extreme light intensities, water scarcity, heat stress or non-ideal CO₂ concentrations.

In addition, the predictive ability of PAM measurements for photosynthetic activity relates to the specific physiology of the plant under investigation due to different photosynthesis pathways (von Caemmerer and Furbank, 2003; Farquhar et al., 1980). C3 plants, such as basil, tend to exhibit higher variations when photorespiratory processes are intensified (Yin and Struik, 2009). It occurs more frequently when light intensity is high (>1000 PAR), and CO₂ availability is low (e.g., 100 ppm). Then a considerable part of the electron transport is directed toward oxygen degradation and can no longer be used for CO₂ fixation. Foyer and Noctor (2000) provide a detailed explanation of the relationship between electron transport, oxygen degradation and photorespiration, especially under high light intensities and low CO₂ availability (Foyer and Noctor, 2000).

In contrast, a linear relationship between PSII activity and CO₂ fixation was initially expected for C4 plants such as maize (Krall and Edwards, 1992). Unlike C3 plants, they utilize a more efficient CO₂ fixation mechanism characterized by the spatial separation of CO₂ uptake and fixation in the Calvin cycle (Sage and Zhu, 2011). This leads to an almost complete elimination of photorespiration and to a stronger correlation between the electron transport rate in PSII and the CO₂ fixation rate. However, under extreme light intensity or stress conditions, such as drought, high temperatures, or nutrient deficiencies, alternative electron fluxes that do not directly contribute to CO₂ fixation are also activated in C4 plants.

Finally, nitrogen availability plays a significant role in restricted photosynthesis rate prediction by PAM parameters, particularly in C4 plants. Kromdijk et al. (2016) showed that under conditions of high nitrogen supply, both C3 and C4 plants show a stronger correlation between PAM parameters and CO₂ fixation, as nitrogen is an important factor for photosynthetic capacity and electron transport rates (Kromdijk et al., 2016). Nitrogen deficiency, on the other hand, significantly weakens this correlation, as photosynthetic capacity is limited under suboptimal nutrient conditions. Thus, C4 plants tend to exhibit a stronger correlation under optimal conditions, while C3 plants are more susceptible to decoupling. As a tool for the model setup, we examined the impact of light on the electron transport rate in PSII obtained by PAM, kept temperature, CO₂ concentration, water supply, as well as humidity constant and measured gas exchange.

Taken together, a mathematical model is required to predict photosynthetic rates by PAM data under non-optimal conditions. Two machine learning tree-algorithms — Random Forest Regressor (Breiman, 2001) and XGBoost (=Extreme Gradient Boosting) Regressor (Chen and Guestrin, 2016) — are further on validated in this research to predict the relationship between chlorophyll fluorescence (PAM) and gas exchange data. Traditional linear or empirical models often fail to capture the nonlinear and multifactorial nature of photosynthetic responses under variable environmental conditions. Machine learning approaches, particularly tree-based algorithms such as Random Forest and XGBoost, are well suited to model such complex interactions and have shown high predictive accuracy in various plant-related applications (Boulesteix et al., 2012; Lysenko et al., 2022; Saleem et al., 2024; Singh et al., 2016, 2023; Varghese et al., 2023; Wu et al.,

2022; Zhang et al., 2020). These models were selected since the underlying processes are non-linear and comprise large sets of input features to accurately predict the assimilation rate. Another advantage of tree-based algorithms is the explainability due to the accessible split points of the input data. Almeida et al. (2020) revealed the capability of tree algorithms for plant classification and identifying different vegetative and floral traits out of a set of 16 traits (Almeida et al., 2020). An accuracy of 89% was achieved on average. Additionally, XGBoost has reached in multiple biological research questions a high accuracy and low error rate and is therefore utilized in this paper for further investigation (Babajide Mustapha and Saeed, 2016). To the best of our knowledge, this is the first study to apply explainable tree-based models to predict CO₂ uptake from fluorescence-based input data in C3 and C4 plants under diverse light spectra.

The goal of this research is to combine PAM and gas exchange data using machine learning algorithms, to achieve more accurate photosynthesis predictions by PAM acquired data under varying light spectra.

Materials and methods

Plant growth conditions

Seeds were pre-soaked in water for a duration of 3 hours, and 20 seeds were subsequently placed in 1 L pots for germination. The plants were maintained in a greenhouse with a 12-hour light/dark cycle, a light intensity of 300 $\mu\text{mol}/(\text{m}^2\text{s})$, a controlled temperature of 21°C, and 68% humidity. Illumination was provided by SANlight LED lamps from 8:00 am to 8:00 pm, ensuring a constant photon flux density of 300 PAR throughout the photo period. This setup allowed for reproducible growth conditions independent of natural daylight. Plants were watered daily at 12:00 pm. To ensure uniform hydration before measurement, all plants received an additional standardized watering of 100 ml on the day before their introduction into the experimental setup. After a 5-day germination period for maize or a 7-day germination period for basil, the seedlings were transferred into 1 L pots with a 14 cm diameter for maize or a 110 ml pot with a 7 cm diameter for basil and grown for 20 days to reach the appropriate age. The basil plants are repotted in pots with a diameter of 12 cm and a volume of 750 ml after 7 days. The plants used to study the influence of plant age on photosynthesis were cultivated under the same conditions for 6 weeks. Shade plants were cultivated under a light intensity of 100 $\mu\text{mol}/(\text{m}^2\text{s})$, respectively. All measurements and plant cultivation procedures were carried out between March 2023 and September 2024.

Experimental conditions and illumination

In this experiment, illumination was provided by Sevengines chips from Chips4Light GmbH (Regensburg, Germany, Figure 1), an advanced LED system with a narrow beam angle of $\pm 10^\circ$. These



FIGURE 1
Sevenengine LED modules with total-internal-reflection (TIR) lens technology used for precise illumination in photosynthesis experiments. The modules deliver highly focused light with minimal loss, ensuring efficient targeting of the experimental area (Chips4Light GmbH, Regensburg, Germany).

state-of-the-art modules utilize total-internal-reflection (TIR) lens technology, which minimizes light losses and delivers exceptional light focus. In contrast to conventional LEDs with standard lenses, the TIR lenses achieve an efficiency of 90%–95%, ensuring that nearly all generated light is directed precisely to the target area.

The LEDs used represent the latest generation of chips, among the brightest available worldwide. The system's advanced thermal management and optical design guarantee optimal performance for each wavelength. The available modules cover a wide wavelength range between 367 nm and 940 nm, providing flexibility in selecting wavelengths critical for specific experimental requirements.

Measurements and calculations

Steady state gas exchange measurements

Gas exchange was assessed using a portable gas-exchange system (GFS-3000, Heinz Walz GmbH, Germany). Measurement conditions were set to 400 ppm CO₂, 18000 ppm humidity, a cuvette temperature of 25°C, and light intensities ranging from 0 μmol/(m²s) to 8500 μmol/(m²s). These conditions were chosen to reflect a wide range from darkness to full sunlight conditions, including stress levels under extreme irradiance up to 8500 μmol/(m²s), thereby allowing analysis of both optimal and non-optimal photosynthetic responses (Farquhar et al., 1980). Plants were dark-adapted overnight (16 hours) and then exposed to 100 PAR white light (4000 K) for 75 minutes to initiate photosynthesis before increasing the light intensity. At each light intensity level, plants were allowed to acclimate for 12 minutes to reach a steady-state photosynthetic rate, based on prior observations that this duration is sufficient to stabilize their gas exchange parameters after a change

in irradiance. Data from the final 2 minutes of this acclimation period were averaged to ensure steady state conditions before analysis.

Steady state chlorophyll fluorescence measurements

Prior to treatment, the plants were dark-adapted for 60 minutes, after which the lowest light intensity (100 PAR) was applied for 60 minutes. Each subsequent light level was applied for 15 minutes, allowing the plants to acclimate and reach a steady state.

Chlorophyll fluorescence was assessed using two PAM fluorometers from Heinz Walz GmbH (Effeltrich, Germany). For image-based measurements, the Imaging-PAM MAXI system (IMAG-MAX/L) was used. Saturation pulses were applied every 30 seconds at a wavelength of 650 nm, with an intensity of 5800 μmol/(m²s) and a pulse duration of 720 ms. For portable measurements, the Junior-PAM fluorometer (CFMG 0225) was employed. Saturation pulses were generated using a 445 nm LED at a typical intensity of approximately 6000 μmol/(m²s). Chlorophyll fluorescence parameters were calculated, including the maximum quantum efficiency of PSII (F_v/F_m), quantum yield of photosynthetic electron transport [$\Phi_{PSII} = (F_m' - F)/F_m'$], electron transport rate [ETR = $\Phi_{PSII} * PAR * 0.84 * 0.5$], and the quantum yield of energy loss related to non-photochemical quenching [YNPQ = (F/F_m) - (F/F_m)] from the Chl fluorescence measurements.

Measurement repetitions

A minimum of ten plants were tested per condition for all chlorophyll fluorescence and gas exchange experiments. When multiple conditions were assessed on the same plant, the order of these conditions was randomized. In total, 360 plants were analyzed in this study.

Machine learning

Two machine learning models, Random Forest and XGBoost, were adapted to link chlorophyll fluorescence measurements to plant gas exchange measurements. These tree-based models were selected due to their robustness in handling non-linear relationships, high interpretability, and successful application in plant physiological modeling (Singh et al., 2023; Zhang et al., 2020).

The hyperparameters of the utilized models are optimized to enhance their predictive performance. The Randomized Search Cross-Validation (Bergstra and Bengio, 2012) technique was implemented to improve the hyperparameters. Due to the high number of interacting features, randomized hyperparameter tuning was used to avoid overfitting and ensure generalizability of the models. This approach employs a random selection process to choose hyperparameter combinations and subsequently performs cross-validation. The optimized hyperparameter types for the Random Forest Regressor are: the number of trees in the forest (*n_estimators*), the maximum number of features considered for

splitting at each node (*max_features*), the maximum depth of the tree (*max_depth*), the minimum number of samples required to split an internal node (*min_samples_split*), and the minimum number of samples required to be at a leaf node (*min_samples_leaf*). We implemented the Random Forest model and conducted 100 iterations, sampling the specified hyperparameters. The following search ranges were applied:

- *n_estimators* = [100,200,300,400,500,800,1000],
- *max_features* = [None, "sqrt", "log2"],
- [5 <= *max_depth* <= 20] with step size 1,
- [0.1 <= *min_samples_split* <= 0.3] with step size 0.02 and.
- [0.05 <= *min_samples_leaf* <= 0.2] with step size 0.02.

for the Random Forest Regressor. The hyperparameter search interval for the XGBoost Regressor are:

- *n_estimators* = [100, 200, 350, 500, 600, 700, 1000],
- *max_depth* = [4, 6, 8, 10, 12],
- *learning_rate* = [0.01, 0.1],
- *min_child_weight* = [50, 100, 150, 200, 250, 300, 350],
- *gamma* = [100, 200, 500, 1000, 1500, 2000, 2500, 3000, 3500] and
- *colsample_bytree* = [0.8, 0.9, 1.0].

Whereas the *learning_rate* specifies the step size shrinkage used in updates to prevent overfitting, the *min_child_weight* describe the minimum sum of instance hessian-weight needed in a child, *gamma* is the minimum loss reduction required to make a further partition on a leaf node of the tree and *colsample_bytree* is a family of parameters for subsampling of columns.

To provide a reliable evaluation with hyperparameter optimization a 5-fold cross-validation was employed.

Results

The machine learning model input parameters must comprise factors with significant impact on photosynthesis rates, as these are critical for generating accurate predictions. It comprises environmental conditions such as photon density, temperature, and humidity, as well as plant-specific traits such as species, developmental stage, or leaf position, all contributing to the overall photosynthetic performance. The following sections are structured in two parts: The first part (Input-Parameter) elucidates the need and the significance of the features whereas the second part (Machine Learning) employs data of these features in the machine learning models to predict the assimilation rate by Random Forest or XGBoost.

Input-parameter

Cultivation under different light conditions

Cultivation of plants under varying light intensities results in the development of so-called sun and shade plants, with distinct physiological characteristics. When maize was exposed to 2000

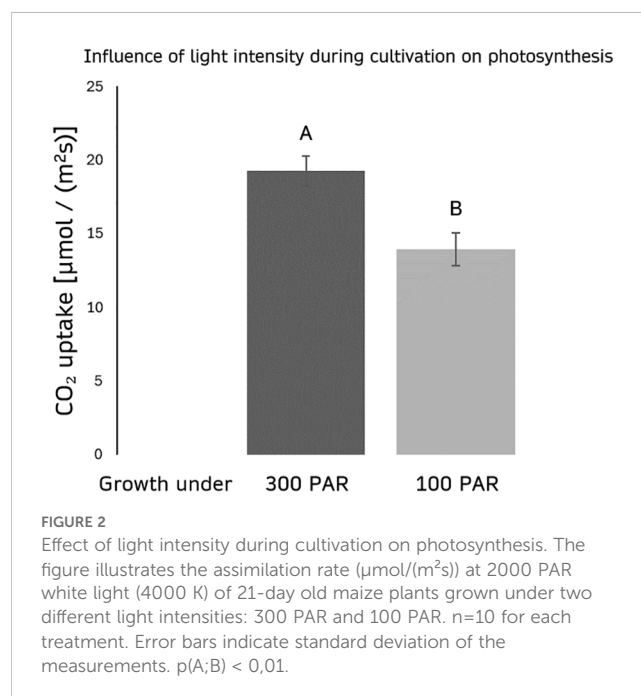
PAR, plants grown under 300 PAR white light (3500K) showed an assimilation rate of $19.7 \pm 1.007 \mu\text{mol}/(\text{m}^2\text{s})$ whereas plants grown under 100 PAR white light exhibit a significantly lower assimilation rate of $13.9 \pm 1.10 \mu\text{mol}/(\text{m}^2\text{s})$ under the same conditions (Figure 2).

Sun leaves, cultivated under high light intensities, exhibit significantly higher CO_2 assimilation rates compared to shade leaves, which developed under lower light intensities. This disparity in assimilation rates under identical light conditions highlights the critical influence of light intensity during cultivation. Therefore, it is essential to include light intensity during plant growth as a variable in any model predicting CO_2 assimilation, as it directly impacts the current photosynthetic performance of sun or shade leaves even if data were obtained under significantly higher light irradiation than those used for plant cultivation.

Plant species: diverse pathways and their impact

Incorporating both C3 and C4 plants into a model examining the relationship between PAM and gas exchange is supportive for capturing a broader range of photosynthetic responses across different environmental conditions. C3 plants, such as basil, rely on the Calvin cycle for carbon fixation, where CO_2 is directly incorporated into a three-carbon compound. These plants perform better under cooler, moist conditions with moderate light. In contrast, C4 plants, such as maize, utilize a specialized carbon fixation pathway that concentrates CO_2 in specific cells, reducing photorespiration. This adaptation makes C4 plants more efficient under high light intensity, elevated temperatures, and dry conditions.

The photosynthetic rates of maize (C4) and basil (C3) under varying light intensities (PAR, white light) clearly reflect their distinct photosynthetic efficiencies. At 500 PAR, maize exhibits an assimilation rate of $12.35 \pm 1.309 \mu\text{mol}/(\text{m}^2\text{s})$, while basil reaches $7.613 \pm 0.916 \mu\text{mol}/(\text{m}^2\text{s})$. As light intensity increases to 1000 PAR,



maize achieves $16.981 \pm 1.439 \mu\text{mol}/(\text{m}^2\text{s})$ compared to basil's $8.376 \pm 0.961 \mu\text{mol}/(\text{m}^2\text{s})$. At the highest light intensity (2000 PAR), maize reaches $19.635 \pm 1.362 \mu\text{mol}/(\text{m}^2\text{s})$, whereas basil only shows a slight increase to $8.702 \pm 0.969 \mu\text{mol}/(\text{m}^2\text{s})$ (Figure 3).

By including data from both C3 and C4 plants, the model can more accurately simulate photosynthetic responses under varying environmental conditions. This broader approach enables a more comprehensive understanding of the factors influencing the relationship between PAM measurements and gas exchange, accounting for the distinct physiological strategies of these plant types.

Plant age: evaluating its impact on photosynthetic rate

We conducted experiments to assess the effect of maize plant age on photosynthesis rates. To ensure consistency, photosynthesis was measured at the same location for six consecutive weeks.

The assimilation rates increased from week 1 ($17.937 \pm 0.376 \mu\text{mol}/(\text{m}^2\text{s})$) to a peak in week 2 ($22.805 \pm 2.480 \mu\text{mol}/(\text{m}^2\text{s})$) and week 3 ($21.862 \pm 0.942 \mu\text{mol}/(\text{m}^2\text{s})$), before gradually declining in week 4 ($17.337 \pm 1.767 \mu\text{mol}/(\text{m}^2\text{s})$) and week 5 ($17.972 \pm 0.721 \mu\text{mol}/(\text{m}^2\text{s})$). Notably, a significant reduction in assimilation rates was observed in week 6 ($6.695 \pm 0.579 \mu\text{mol}/(\text{m}^2\text{s})$), marking a sharp decline in photosynthesis rates (Figure 4).

This suggests that while photosynthesis rates remain relatively stable during early developmental stages a notable reduction occurs as the plant matures beyond the fifth week.

Measurement location: influencing photosynthetic performance

The photosynthetic rate is not uniform across different locations in the leaf or plant. Our data indicate that older leaf tissues exhibit higher photosynthetic rates compared to younger leaf areas. Measurements taken at different locations on the maize plant (Figure 5A) reveal significant variation. The photosynthetic rate at

the leaf tip was $19.760 \pm 2.337 \mu\text{mol}/(\text{m}^2\text{s})$, while the center of the leaf showed a lower rate of $13.251 \pm 0.990 \mu\text{mol}/(\text{m}^2\text{s})$, and the base had the lowest rate of $12.322 \pm 1.108 \mu\text{mol}/(\text{m}^2\text{s})$ (Figure 5B). This variability suggests that age-related changes within individual leaves influence photosynthetic performance, with older tissues being more efficient in carbon assimilation. Consequently, spatial differences across the leaf must be considered when assessing and comparing photosynthetic rates of different plants, as tissue age can significantly impact the overall measurement.

Photosynthesis as a function of leaf position

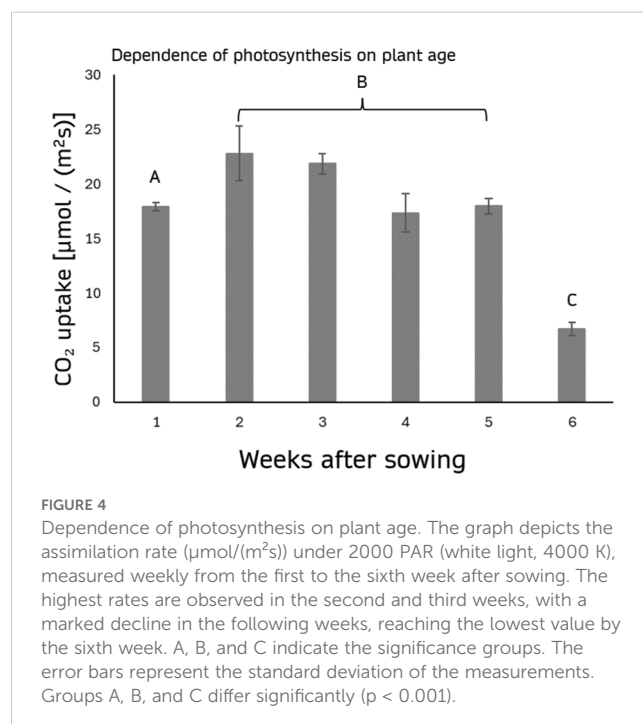
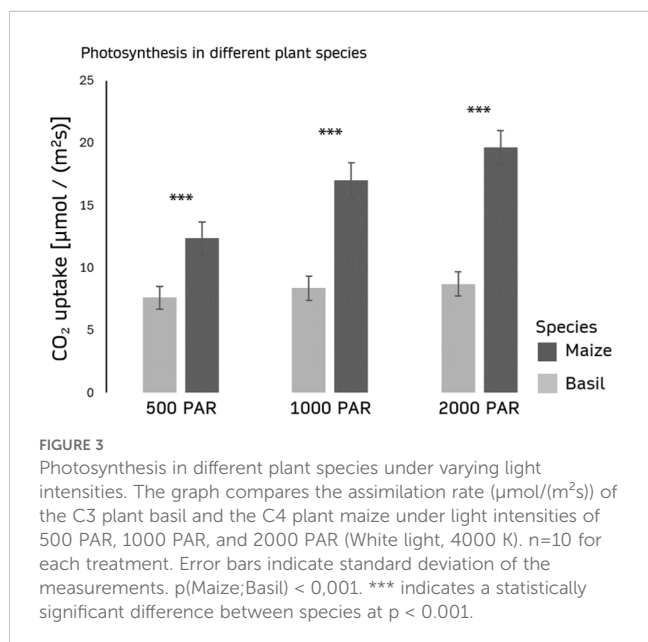
The photosynthetic rate is also not uniform along the shoot axis in different leaf layers of a single plant, and this variability is species dependent (Figures 6A, B).

For maize, the middle leaf-layer exhibited the highest photosynthetic rate at $21.795 \pm 0.421 \mu\text{mol}/(\text{m}^2\text{s})$, followed by the upper layer at $19.247 \pm 1.004 \mu\text{mol}/(\text{m}^2\text{s})$, with no significant difference between these two layers. The lowest photosynthetic rate was observed in the lower leaf layer, with a mean rate of $13.941 \pm 1.050 \mu\text{mol}/(\text{m}^2\text{s})$.

In contrast, basil showed a different pattern, with the upper leaf layer having the highest photosynthetic rate at $11.618 \pm 0.923 \mu\text{mol}/(\text{m}^2\text{s})$, followed by the middle layer at $8.614 \pm 0.760 \mu\text{mol}/(\text{m}^2\text{s})$. The lower leaf layer exhibited the lowest rate, with $5.104 \pm 0.513 \mu\text{mol}/(\text{m}^2\text{s})$ (Figure 6C). This data highlights species-specific differences in the spatial distribution of photosynthetic activity within the canopy.

Photosynthetic rates of corn and basil under various intensities of white and monochromatic light

We collected data to analyze the photosynthetic rates of maize (*Zea mays*) and basil (*Ocimum basilicum*) under different intensities



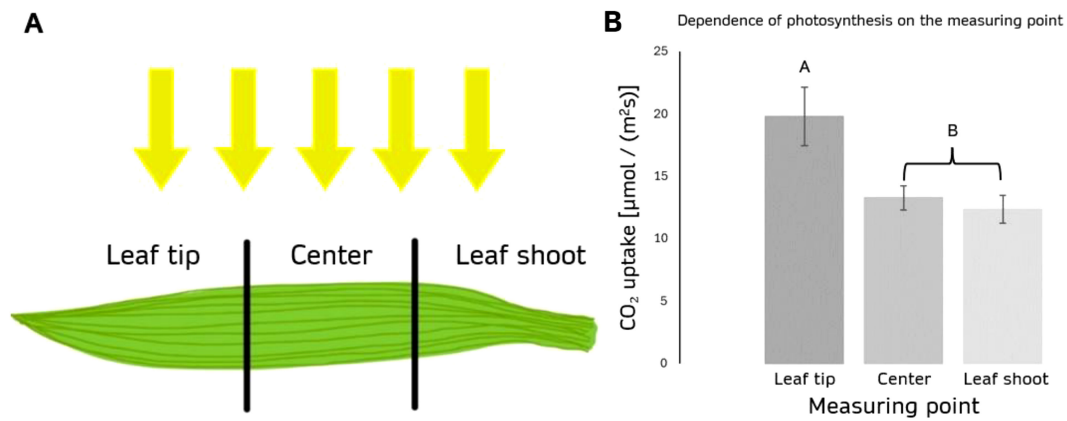


FIGURE 5 Dependence of photosynthesis on the measuring point of 3 weeks old maize plants. **(A)** Schematic representation of the three analyzed leaf segments relative to the light source during maize cultivation. **(B)** The graph illustrates the assimilation rate ($\mu\text{mol}/(\text{m}^2\text{s})$) under 2000 PAR (white light, 4000 K), measured at three locations: the tip, center, and leaf base. $n=10$ for each treatment. The error bars indicate the standard deviation of the measurements. $p(\text{Leaf};\text{Center}) < 0.01$, $p(\text{Leaf};\text{Shoot}) < 0.01$, $p(\text{Shoot};\text{Center}) = 0.534$. Groups A and B differ significantly ($p < 0.001$).

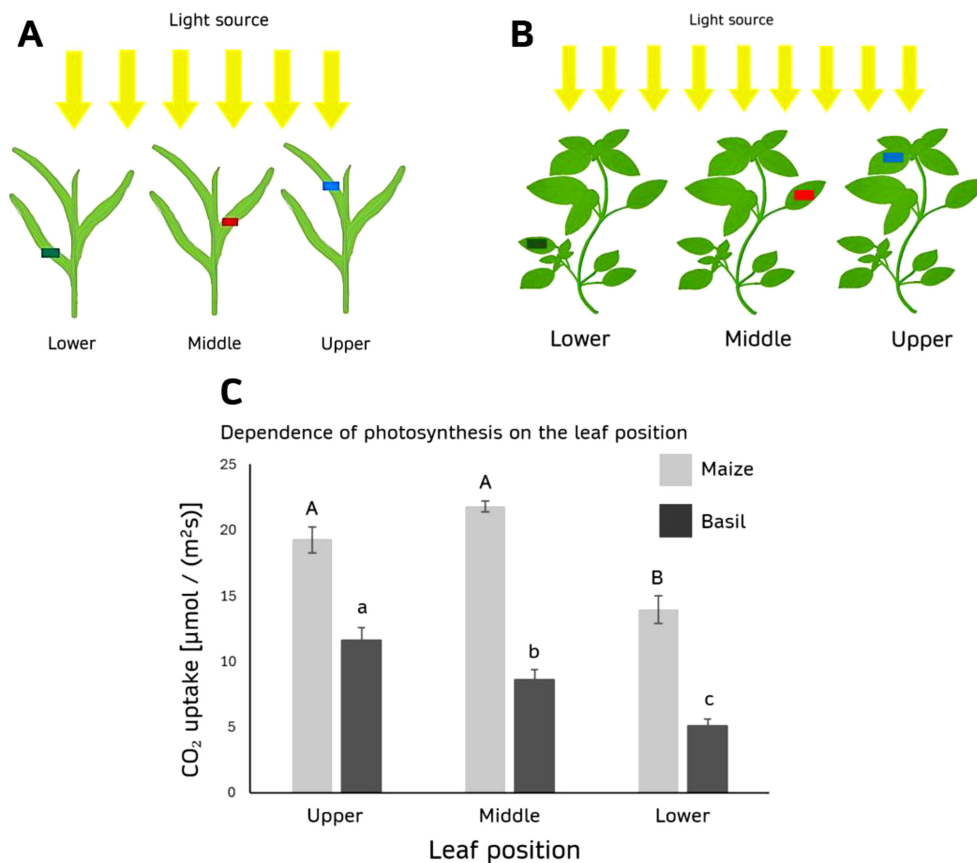


FIGURE 6 Dependence of photosynthesis on the leaf position of 3 weeks old plants. **(A)** Schematic representation of the three analyzed leaf positions relative to the light source during maize cultivation and **(B)** during basil cultivation. **(C)** The graph illustrates the assimilation rate ($\mu\text{mol}/(\text{m}^2\text{s})$) under 2000 PAR (white light, 4000 K), measured at three leaf positions: the lowest, middle, and upper for basil and maize. Maize: $p(\text{upper};\text{middle}) = 0.015$, $p(\text{upper};\text{lower}) = 0.003$, $p(\text{middle};\text{lower}) < 0.001$. Groups A and B differ significantly ($p < 0.001$). Basil: Groups a, b, and c differ significantly ($p < 0.001$).

of white light and monochromatic light (Figure 7). We investigated 500 PAR, 1000 PAR, and 2000 PAR under white (2600 K), red (630 nm), blue (450 nm), and green light (527 nm).

At a PAR of 500 $\mu\text{mol}/(\text{m}^2\text{s})$, the photosynthetic rates of basil were relatively consistent across all light treatments, while maize showed the highest photosynthetic rate under red light (16.620 $\mu\text{mol}/(\text{m}^2\text{s})$), which was notably higher than the rates under blue, green, or white light. At 1000 $\mu\text{mol}/(\text{m}^2\text{s})$, basil exhibited higher rates under red and green light, with the red light reaching 9.076 $\mu\text{mol}/(\text{m}^2\text{s})$, compared to lower rates under blue and white light. Similarly, maize showed improved assimilation under green light (23.655 $\mu\text{mol}/(\text{m}^2\text{s})$) and red light, which were both higher than the blue and white light conditions. At the highest light intensity of 2000 $\mu\text{mol}/(\text{m}^2\text{s})$, basil continued to show enhanced photosynthetic rates under red and green light, with the highest rate of 10.557 $\mu\text{mol}/(\text{m}^2\text{s})$ under red light. For maize, green light produced the highest photosynthetic rate (30.992 $\mu\text{mol}/(\text{m}^2\text{s})$), followed closely by red light, with both outperforming blue and white light significantly.

Therefore, it is essential to incorporate both light spectrum and light intensity as important variables in the predictive model, as they significantly influence photosynthetic performance and thus impact the accuracy of CO_2 assimilation rate predictions.

Feature definition

To establish a connection between chlorophyll fluorescence and gas exchange parameters, machine learning algorithms were applied. These algorithms operate on datasets composed of individual features (Table 1).

Machine learning - input feature set

The input data processing for the tree algorithms were structured into the following four steps: (1) data cleaning, (2) data transformation, (3) model training and (4) model evaluation.

Initially, in the data cleaning step the input dataset is prepared for further evaluation. First the non-numeric features that were only

provided for the human description of the data were removed. This includes the textual described condition and the name of the light source. Furthermore, duplicated data as well as rows which contain missing values were dropped (Lee et al., 2021).

In the remaining data each feature with less than three distinct values was also removed since this attribute would provide only little valuable information (Ilyas and Chu, 2019). In this stage, the initial dataset of 1783 data with 36 features was optimized to 1739 and 26 features.

Following the initial data preparation, additional functions were developed in the second step - the data transformation - to improve the strength of the prediction models. This included the creation of additional 20 interaction features, to map the domain knowledge of the researchers into data. This was achieved by combining the features for specifying the light with the recorded intensities per wavelength in pairs (Hornung and Boulesteix, 2022). The following wavelengths were considered: 405 nm, 430 nm, 450 nm, 465 nm, 485 nm, 500 nm, 527 nm, 550 nm, 590 nm, 630 nm, 660 nm, 730 nm and are pairwise multiplied. By capturing the precise interaction between two wavelengths, the model can gain a better understanding, that these features are related and can thus make more accurate predictions (Hornung and Boulesteix, 2022). Furthermore, the depth of the tree can be reduced since this correlation must not be learned by the tree.

As the evaluation date is a non-numeric feature, a new feature was created by counting the months from the year of the first measurement with the equation: $((\text{year} - \text{first_year}) * 12 + \text{month})$ (Heaton, 2016). This was necessary to analyze the time dependencies and seasonal patterns in greater detail. In combination with the month the temporal course and seasonal fluctuations that could affect the physiological processes in a plant are considered. Since tree-based algorithms will be utilized as models the absolute features were further processed and no standardization or normalization process is required (Breiman, 2001).

Several feature sets were created utilizing different models to determine the optimal set of input parameters for the model. During the analysis, the development of so called minimal and interaction-based feature sets are considered and are subsequently explained.

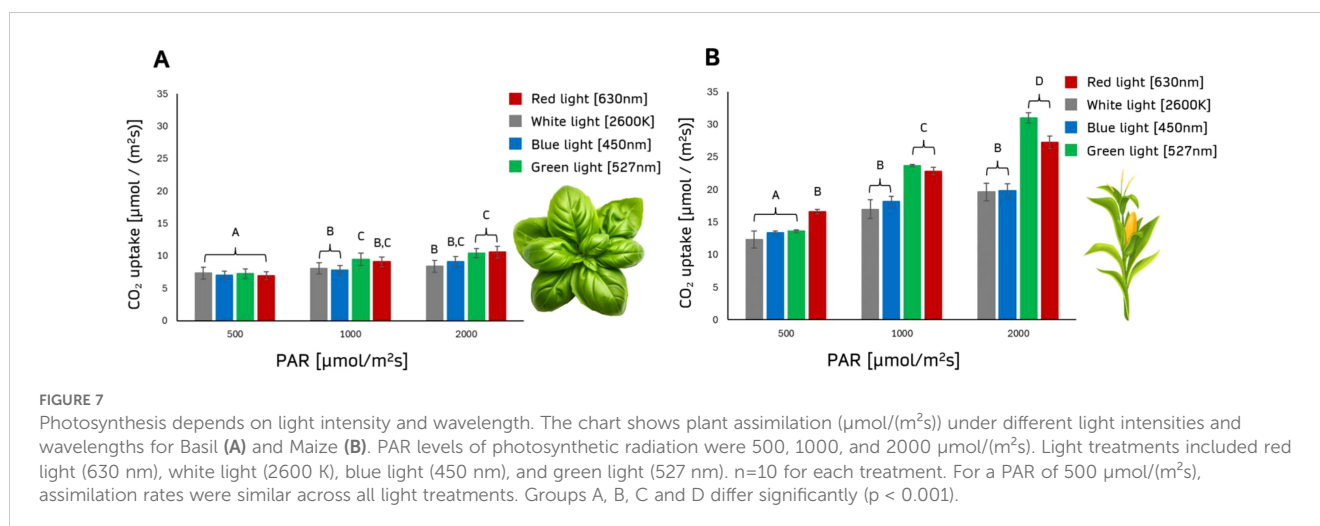


TABLE 1 Overview of the features used for model training and evaluation, including environmental conditions, physiological parameters, and spectral components.

Feature	Meaning
par_total	Total light intensity (400–730 nm), measured in $\mu\text{mol}/(\text{m}^2\text{s})$.
2600K, 3200K, 4000K	Light intensity contributed by the respective white light (2600K, 3200K, 4000K), measured in $\mu\text{mol}/(\text{m}^2\text{s})$.
405_nm – 730_nm	Light intensity contributed by the respective monochromatic light at each wavelength, measured in $\mu\text{mol}/(\text{m}^2\text{s})$.
plant_age	Age of the plant since sowing, measured in days.
plant_id	Unique sequential identifier assigned to each plant to ensure traceability of measurements.
month	The month in which the measurements were conducted.
year	The year in which the measurements were conducted.
e	Transpiration rate measured using the gas exchange system, expressed in $\text{mmol H}_2\text{O}/(\text{m}^2\text{s})$.
g	Stomatal conductance to water vapor, measured in $\mu\text{mol H}_2\text{O}/(\text{m}^2\text{s})$.
tleaf	Leaf temperature measured using the gas exchange system, expressed in $^{\circ}\text{C}$.
Y(II), ETR, NPQ	Quantum yield of photosystem II, Electron transport rate, Non photochemical quenching, measured using the PAM fluorometer.
leaf_level	Indicates the leaf layer within the plant where the measurements were conducted.
m_segment	Specifies the measured segment of the leaf: proximal (near the stem), middle, or distal (leaf tip).
species	Categorical variable indicating whether the plant is <i>Zea mays</i> (C4) or <i>Ocimum basilicum</i> (C3)

The minimal feature set comprised a core selection of characteristics believed to be relevant. The crucial feature subset comprised Y(II), ETR, par_total and plant age. The Interaction data set extends the Minimal set by incorporating additional variables, including plant_id, month and year. Furthermore, specific spectral light components (430 nm, 450 nm, 485 nm, 500 nm, 527 nm, 550 nm, 590 nm, 630 nm, 660 nm, as well as various white light sources at 2600 K, 3200 K, and 4000 K) are considered. To enhance predictability, a feature set is developed based on interactions between two wavelengths at a time as explained before. Additional factors also include tleaf, e, $g\text{H}_2\text{O}$, leaf_level, and m_segment.

Using these two different sets, the influence of diverse combinations of features on the model was systematically evaluated. The minimal set provided insight into the basic and essential predictors, while the interaction term-based set focused on capturing high-order complex interactions and determining their additional value and included all provided features.

These features are further evaluated by two machine learning tree-algorithms: Random Forest Regressor (Breiman, 2001) and XGBoost Regressor (Chen and Guestrin, 2016).

The utilized data set was split into two subsets for training and testing with 80% train and 20% test data. The divide assures that the model's accuracy and ability to generalize is determined by evaluating its performance on data that were not used for training. This ensures that the performance of the trained models is evaluated on data that has not been previously observed during the training phase, thereby yielding a transferable and generalized model.

Machine learning - hyperparameter search

The previous section stated the importance of light intensity, plant age, measuring point, leaf position as well as different wavelengths. Further on, the results of the tree-based machine

learning models XGBoost and Random Forest are stated. Overall, 1739 data points with a maximum of 46 features are utilized as input.

The goal of this model is to achieve the highest level of performance as well as a minimal difference in the training and test data set for predicting the assimilation rate based on the input features. As a first step of the model creation process, the impact of modifications in a solitary hyperparameter on the accuracy of the model were closely examined. The model performance was assessed involving two primary metrics: the Mean Squared Error (MSE) and the R-squared (R^2) Score whereas the R^2 Score is the main metric.

The mentioned hyperparameter ranges were therefore obtained by a visual evaluation of these parameters over the R-squared (R^2) score. A selected number of plots can be inspected in Figure 8. These plots were selected since they illustrate the greatest change over the parameter in the R^2 Score. For a generalized model the R^2 difference between the train and test data set must be minimized.

The final parameters for Random Forest are:

- Interaction data set:
 - max_depth=6, min_sample_leafs=0.05, min_sample_split=0.1, n_estimators=800.
- Minimal data set:
 - max_depth=6, min_sample_leafs=0.05, min_sample_split=0.1, n_estimators=800.

and the parameters for XGBoost are:

- Interaction data set:
 - colsamples_bytree=0.9, gamma=100, learning_rate=0.1, max_depth=4, min_child_weight=50, n_estimators=500.

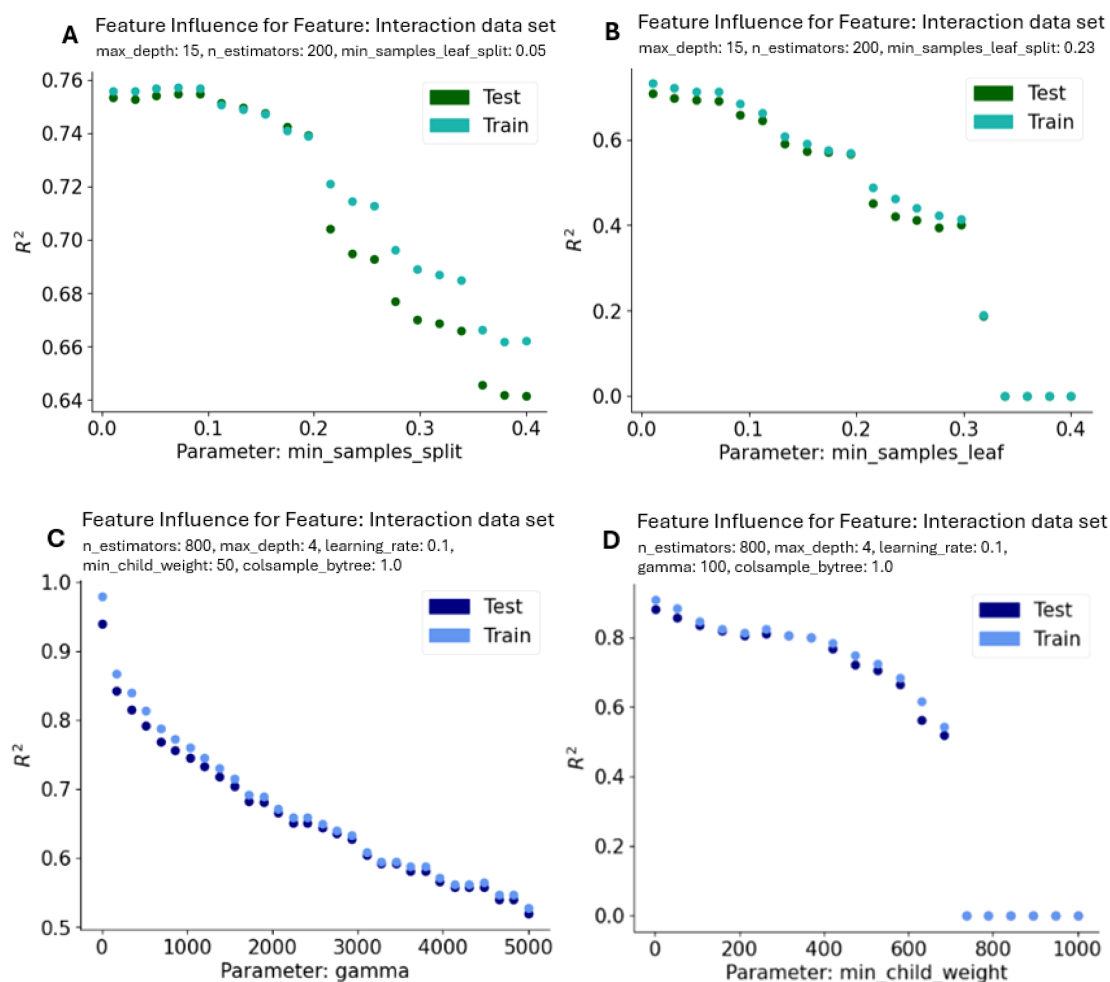


FIGURE 8

Evaluation of two parameters of the Random Forest Regressor (A, B) and two parameters for XGBoost (C, D) with the largest spread and decrease between training and test data set with regards to the R-squared Score R^2 . All diagrams show the interaction (X_interactions) data set. A similar course is revealed for the minimal data set. The used non-default values are displayed on the top of each plot.

- Minimal data set:
 - colsamples_bytree=1.0, gamma=100, learning_rate=0.1, max_depth=8, min_child_weight=50, n_estimators=600.

Machine learning – feature importance

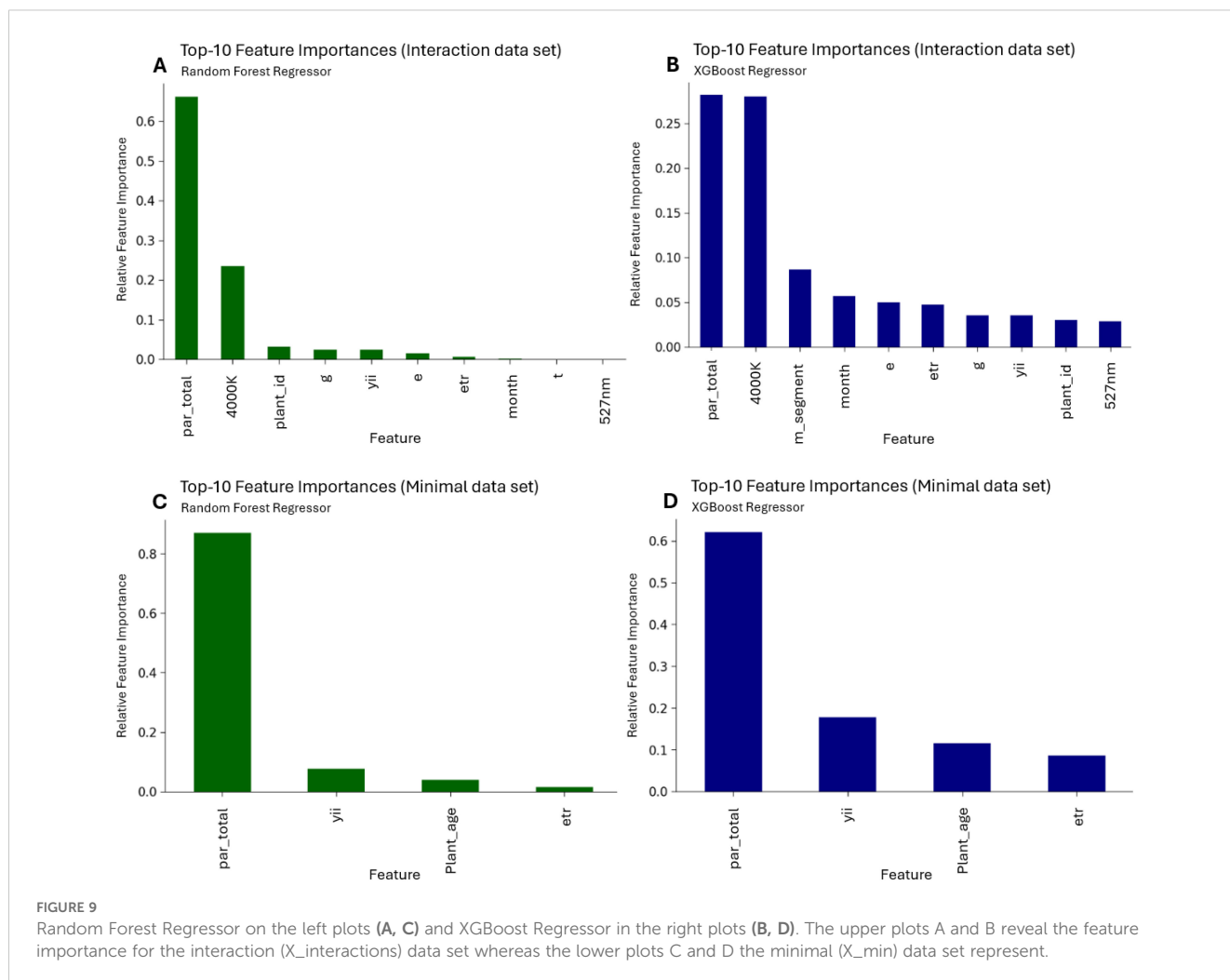
In addition to assessing the overall performance of the model, it is crucial to evaluate the significance of a feature in predicting the target variable. The analysis in Figure 9 highlights the key features that have a significant impact on the model's predictions, offers insight into their biological interpretation and were obtained by the best model after hyperparameter optimization. The importance of a feature in the model indicates its contribution to the final prediction and the sum of the relative importance is 1.0.

From Figure 9 and the interaction data set the features g, e, par_total, etr and yii have the highest importance, whereas in the minimal data set likewise par_total, etr and yii are revealed to have the highest impact on the model to predict the chlorophyll fluorescence measurements.

Machine learning – model performance

The model performance results for Random Forest (A and B) and XGBoost (C and D) algorithms are evaluated based on predicted vs. true values scatter plots and error frequency distributions (Figure 10).

The scatter plot indicates that the Random Forest (Figure 10A) demonstrates a correlation between predicted and true values. The dispersion around the regression line is considerable and occasionally reaches a maximum for elevated assimilation rates ($>15 \mu\text{mol}/(\text{m}^2\text{s})$). The R^2 score of 0.752, and an MSE of 8.5 indicates favorable outcomes for the test dataset. Nevertheless, the fitted values for elevated assimilation rates tend to be slightly underestimated, resulting in systematic discrepancies. The error frequency figure (Figure 10B) exhibits a somewhat centralized distribution with a peak near 0, indicating that the model predominantly yields unbiased predictions for most observations. Nevertheless, the distribution of errors is extensive, particularly on the positive side, suggesting occasional underestimation of higher values.



The scatter plot for the XGBoost model (Figure 10C) demonstrates a significantly tighter clustering around the regression line, particularly for values below 20 $\mu\text{mol}/(\text{m}^2\text{s})$. The enhanced alignment is evidenced by the elevated R^2 score of 0.847 and the reduced MSE of 5.24 for the test dataset. The variances are reduced, and the model predictions align more closely with the true values throughout the range. Most of the errors are concentrated near zero, signifying that this model more successfully minimizes both bias and variation (Figure 10D).

XGBoost surpasses Random Forest in both datasets, producing superior R^2 scores and reduced MSE values (Table 2).

Discussion

Photosynthesis key insights from feature importance analysis

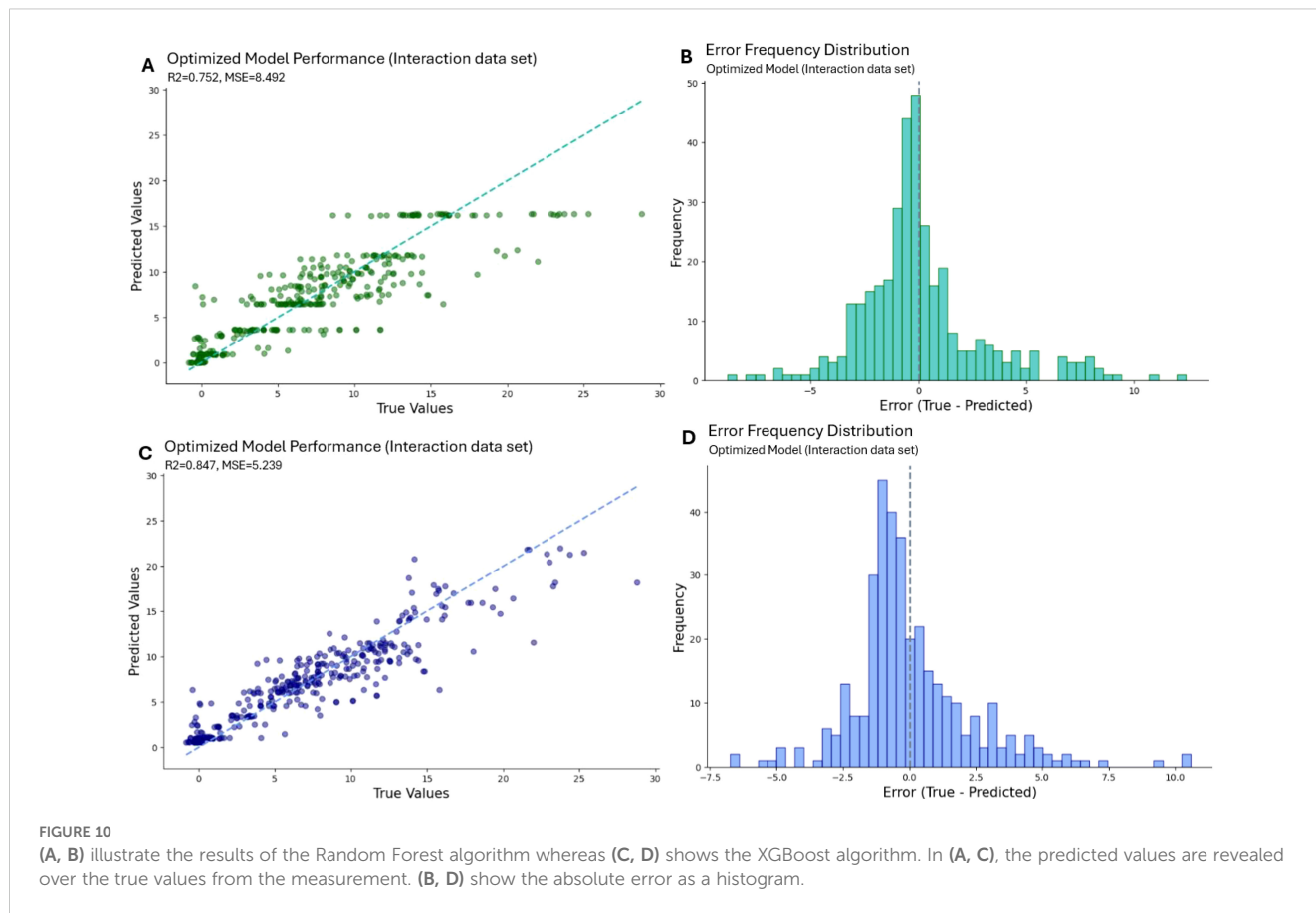
Light intensity is the dominant factor influencing photosynthetic assimilation rates in both datasets. In the minimal dataset, it accounts for over 60% of the total importance, while in the interaction-based dataset, it remains among the most significant

features, particularly in combination with 4000 K. This is consistent with the fundamental role of light as the energy source driving photosynthesis, initiating chlorophyll excitation, electron transport, and carbon fixation (Genty and Harbinson, 2004).

The importance of spectral properties further underscores that photosynthesis is not solely governed by light quantity but also by spectral quality. The spectral composition influences chlorophyll and accessory pigment absorption, light capture efficiency, and energy conversion (Simkin et al., 2022). This highlights the necessity of incorporating both total light intensity and spectral data into predictive models to enhance accuracy.

Chlorophyll fluorescence parameters, particularly Y(II) and ETR, also emerge as key predictors. Y(II) reflects the photochemical efficiency of Photosystem II (PSII), while ETR represents electron transport activity, making them critical indicators of photosynthetic function under varying light conditions (Yousef et al., 2021). Their prominence underscores the value of fluorescence-based metrics in assessing the physiological state of the photosynthetic apparatus (Kalaji et al., 2016).

In the interaction-based dataset, 4000 K and specific wavelength interactions, such as 527 nm, further emphasize the role of spectral



composition. The significance of these variables suggests that certain wavelengths enhance assimilation efficiency, supporting the notion that different pigment systems selectively respond to specific spectral ranges (Kuleshova et al., 2018). Notably, green light (e.g., 527 nm) is not merely reflected but actively contributes to photosynthesis, challenging traditional assumptions about its limited role (Arsenault et al., 2020; Brodersen and Vogelmann, 2010; Chen et al., 2024; Evans and Anderson, 1987; Kaiser et al., 2019; Nishio, 2000; Smith et al., 2017; Sun et al., 1998; Terashima et al., 2009; Virtanen et al., 2022).

Plant-specific traits, including plant age and leaf segment (m_segment), also play a role in both models. Plant age is particularly relevant in the minimal dataset, indicating age-related variations in photosynthetic capacity. This confirms that photosynthesis is not static but changes over a plant's lifespan. The importance of m_segment suggests that photosynthetic efficiency varies within individual leaves, with older tissues generally exhibiting

higher assimilation rates than younger ones (Dwyer and Stewart, 1986; Jahan et al., 2023; Pick et al., 2011; Prochazkova et al., 2001; Suzuki et al., 1987; Trouwborst et al., 2011). Consequently, spatial variation within leaves must be considered in photosynthesis measurements to avoid confounding effects (Kutík et al., 2001). A more physiological level of interpretation is provided by comparing the intrinsic photosynthetic pathways of C3 and C4 species. Maize, a C4 species, possesses a CO₂-concentrating mechanism that minimizes photorespiration and results in a more linear, efficient relationship between light-driven electron transport (as assessed by PAM) and net CO₂ assimilation (Hatch, 1987). In contrast, C3 plants like basil show a less direct correlation between these two processes, especially at high light, due to higher photorespiratory losses (Lee and Kim, 2024). Despite the described differences in the photosynthetic pathways of the studied species, the feature indicating the photosynthetic type (C3 vs. C4) does not appear among the top ten predictors of assimilation rate in the machine learning model.

TABLE 2 The results for the two tree-algorithms Random Forest and XGBoost with the R2 score as well as with the MSE.

Dataset	Random Forest		XGBoost	
	Train R ² (MSE)	Test R ² (MSE)	Train R ² (MSE)	Test R ² (MSE)
minimal data set	0.737 (8.94)	0.741 (9.45)	0.776 (8.64)	0.762 (8.22)
interaction data set	0.756 (8.31)	0.752 (8.5)	0.87 (4.44)	0.847 (5.24)

Bold values highlight the best test performance (highest R², lowest MSE) for each dataset.

XGBoost's superiority in capturing complex biological interactions

The XGBoost model outperforms Random Forest in predicting photosynthetic assimilation rates, achieving a higher R^2 score of 0.847 and a lower mean squared error (MSE) of 5.24. These results indicate that XGBoost captures a greater proportion of the variability in photosynthesis rates and provides more accurate predictions. The model did not regard the interaction features being significant. Only particular spectral components, specifically 4000 K and 527 nm, were recognized as significantly relevant. This highlights that individual wavelengths, rather than complex spectrum relationships, directly affect photosynthetic performance.

A key advantage of XGBoost lies in its ability to model complex feature interactions, particularly those associated with specific wavelengths, which may be overlooked by Random Forest. This is crucial, as photosynthesis involves nonlinear relationships between light, temperature, gas exchange parameters, and plant-specific traits. The ability of XGBoost to integrate diverse features, including *par_total*, chlorophyll fluorescence parameters, spectral data, and plant-specific characteristics, ensures biologically meaningful predictions closely aligned with actual photosynthetic processes.

Additionally, XGBoost effectively manages feature interactions, reducing both bias and variance, as evidenced by its tighter distribution of values along the regression line compared to Random Forest. This effect is particularly pronounced at higher assimilation rates ($>15 \mu\text{mol}/(\text{m}^2\text{s})$), where Random Forest exhibits systematic deviations. XGBoost's improved accuracy and robustness make it the preferred model for photosynthesis prediction (Burdett and Wellen, 2022; Ha et al., 2024; Li et al., 2021; Singh et al., 2023; Zhang et al., 2020).

Evaluation of our model

While the model demonstrates high predictive power ($R^2 = 0.847$), limitations remain. Studies by Ogren and Sjöström (1990) and Sharma et al. (2015) indicate that *Fv/Fm* correlates well with quantum yield under low to moderate light conditions (220 PAR – 500 PAR (Ogren and Sjöström, 1990; Sharma et al., 2015). Lakowicz (2006) further demonstrated that fluorescence intensity reflects photosynthesis quantum yield and chlorophyll concentration (Lakowicz, 2006). This correlation is particularly strong in C4 plants due to their reduced photorespiration.

However, Bucher et al. (2018) found that *Fv/Fm* was not a reliable predictor of assimilation rates in field conditions (Bucher et al., 2018). This discrepancy may stem from the influence of competing electron-consuming processes, such as photorespiration, nitrogen metabolism, the Mehler reaction, and oxygen photoreduction, which are not directly accounted for in fluorescence-based measurements (Heber, 2002; Kalaji et al., 2017; Maxwell and Johnson, 2000).

Another limitation arises from the depth of PAM fluorescence measurements, which only capture the outermost leaf layers. If measurements are restricted to the adaxial leaf surface, data reflect

only chloroplast activity within the palisade parenchyma, neglecting potential differences in spongy mesophyll layers (Troeng and Linder, 1982).

Furthermore, photosynthesis and electron transport rates fluctuate seasonally within a species (Holland et al., 2014; Lazár, 2006; Neuner and Pramsohler, 2006; Oquist and Huner, 2003). These temporal variations must be considered in future model refinements to enhance predictive accuracy under natural conditions. Despite these limitations, our findings provide a powerful tool for further optimizing photosynthesis modeling using machine learning.

The integration of machine learning techniques into photosynthesis modeling offers promising perspectives for future research. A particularly valuable direction could involve combining mechanistic models with machine learning approaches to leverage the strengths of both. Specifically, the development of hybrid models that integrate the physico-chemical foundations of photosynthesis with the pattern-recognition capabilities of machine learning algorithms would allow for the analysis of complex relationships within large datasets.

Conclusion

The study demonstrates a significant advancement in data-driven prediction of photosynthetic assimilation rates using machine learning. We identified a biologically consistent hierarchy of predictors. The results validate the selection of features and their alignment with fundamental biological processes. The XGBoost algorithm demonstrated superior accuracy and reliability compared to Random Forest, particularly in managing difficult relationships and interactions between features. This highlights the potential of machine learning in advancing our understanding of photosynthetic processes and paving the way for further integration of data-driven approaches into plant science. In future applications, such models could support precision agriculture by enabling rapid, non-invasive monitoring of plant performance under field conditions. Especially under abiotic stress scenarios such as drought or high temperatures, this approach may help to identify physiological limitations and optimize crop management strategies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/KlirS/Model-for-PAM-Fluorescence-Gas-Exchange-Correlation>.

Author contributions

IP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation,

Visualization, Writing – original draft, Writing – review & editing. SK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. LJ: Data curation, Investigation, Methodology, Validation, Writing – review & editing. CÜ: Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing – review & editing. KQ: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. RK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by the German Research Foundation (DFG), Project Number 500805487.

References

- Almeida, B. K., Garg, M., Kubat, M., and Afkhami, M. (2020). Not that kind of tree: Assessing the potential for decision tree-based plant identification using trait databases. *Appl. Plant Sci.* 8, e11379. doi: 10.1002/aps3.11379
- Arsenault, E. A., Yoneda, Y., Iwai, M., Niyogi, K. K., and Fleming, G. R. (2020). The role of mixed vibronic Qy-Qx states in green light absorption of light-harvesting complex II. *Nat. Commun.* 11, 6011. doi: 10.1038/s41467-020-19800-y
- Babajide Mustapha, I., and Saeed, F. (2016). Bioactive molecule prediction using extreme gradient boosting. *Mol. (Basel Switzerland)*. 21 (8). doi: 10.3390/molecules21080983
- Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis *in vivo*. *Annu. Rev. Plant Biol.* 59, 89–113. doi: 10.1146/annurev.arplant.59.032607.092759
- Baker, N. R., and Rosenqvist, E. (2004). Applications of chlorophyll fluorescence can improve crop production strategies: an examination of future possibilities. *J. Exp. Bot.* 55, 1607–1621. doi: 10.1093/jxb/erh196
- Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 281–305. doi: 10.5555/2503308.2188395
- Boulesteix, A.-L., Janitza, S., Kruppa, J., and König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs. Data Min. Knowledge. Discov.* 2, 493–507. doi: 10.1002/widm.1072
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi: 10.1023/A:1010933404324
- Brodersen, C. R., and Vogelmann, T. C. (2010). Do changes in light direction affect absorption profiles in leaves? *Funct. Plant Biol.* 37, 403. doi: 10.1071/FP09262
- Bucher, S. F., Bernhardt-Römermann, M., and Römermann, C. (2018). Chlorophyll fluorescence and gas exchange measurements in field research: an ecological case study. *Photosynthetica* 56, 1161–1170. doi: 10.1007/s11099-018-0809-5
- Burdett, H., and Wellen, C. (2022). Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precis. Agric.* 23, 1553–1574. doi: 10.1007/s11119-022-09897-0
- Chen, Y., Bian, Z., Marcellis, L. F. M., Heuvelink, E., Yang, Q., and Kaiser, E. (2024). Green light is similarly effective in promoting plant biomass as red/blue light: a meta-analysis. *J. Exp. Bot.* 75, 5655–5666. doi: 10.1093/jxb/erae259
- Chen, T., and Guestrin, C. (2016). “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA (ACM, New York, NY, USA), 785–794. doi: 10.1145/2939672.2939785
- Dwyer, L. M., and Stewart, D. W. (1986). Effect of leaf age and position on net photosynthetic rates in maize (*Zea Mays* L.). *Agric. For. Meteorol.* 37, 29–46. doi: 10.1016/0168-1923(86)90026-2
- Evans, J. R., and Anderson, J. M. (1987). Absolute absorption and relative fluorescence excitation spectra of the five major chlorophyll-protein complexes from spinach thylakoid membranes. *Biochim. Biophys. Acta (BBA) - Bioenerget.* 892, 75–82. doi: 10.1016/0005-2728(87)90249-0
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149, 78–90. doi: 10.1007/BF00386231
- Flexas, J., Barbour, M. M., Brendel, O., Cabrera, H. M., Carriqui, M., Diaz-Espejo, A., et al. (2012). Mesophyll diffusion conductance to CO₂: an unappreciated central player in photosynthesis. *Plant Sci.* 193–194, 70–84. doi: 10.1016/j.plantsci.2012.05.009
- Flexas, J., and Medrano, H. (2002). Drought-inhibition of photosynthesis in C₃ plants: stomatal and non-stomatal limitations revisited. *Ann. Bot.* 89, 183–189. doi: 10.1093/aob/mcf027
- Foyer, C. H., and Noctor, G. (2000). Oxygen processing in photosynthesis: regulation and signalling. *New Phytol.* 146, 359–388. doi: 10.1046/j.1469-8137.2000.00667.x
- Genty, B., Briantais, J.-M., and Baker, N. R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. Acta (BBA) - Gen. Subj.* 990, 87–92. doi: 10.1016/S0304-4165(89)80016-9
- Genty, B., and Harbinson, J. (2004). “Regulation of light utilization for photosynthetic electron transport,” in Photosynthesis and the Environment. (*Advances in Photosynthesis and Respiration*). Ed. N. R. Baker (Kluwer Academic Publishers, Dordrecht), 67–99.
- Ha, Y., Ma, H. R., Wu, F., Weiss, A., Duncker, K., Xu, H. Z., et al. (2024). Data-driven learning of structure augments quantitative prediction of biological responses. *PLoS Comput. Biol.* 20, e1012185. doi: 10.1371/journal.pcbi.1012185
- Hatch, M. D. (1987). C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta (BBA) - Rev. Bioenerget.* 895, 81–106. doi: 10.1016/S0304-4173(87)80009-5
- Heaton, J. (2016). “An empirical analysis of feature engineering for predictive modeling,” in *SoutheastCon 2016* (IEEE, Norfolk, VA, USA), 1–6. 30 March - 3 April. doi: 10.1109/SECON.2016.7506650
- Heber, U. (2002). Irrungen, Wurrungen? The Mehler reaction in relation to cyclic electron transport in C₃ plants. *Photosynth. Res.* 73, 223–231. doi: 10.1023/A:1020459416987

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Holland, V., Koller, S., and Brüggemann, W. (2014). Insight into the photosynthetic apparatus in evergreen and deciduous European oaks during autumn senescence using OJIP fluorescence transient analysis. *Plant Biol. (Stuttgart, Germany)*. 16, 801–808. doi: 10.1111/plb.12105
- Hornung, R., and Boulesteix, A.-L. (2022). Interaction forests: Identifying and exploiting interpretable quantitative and qualitative interaction effects. *Comput. Stat Data Anal.* 171, 107460. doi: 10.1016/j.csda.2022.107460
- I. F. Ilyas and X. Chu (Eds.) (2019). *Data Cleaning* (New York, NY, USA: ACM).
- Jahan, E., Sharwood, R. E., and Tissue, D. T. (2023). Effects of leaf age during drought and recovery on photosynthesis, mesophyll conductance and leaf anatomy in wheat leaves. *Front. Plant Sci.* 14, 1091418. doi: 10.3389/fpls.2023.1091418
- Kaiser, E., Weerheim, K., Schipper, R., and Dieleman, J. A. (2019). Partial replacement of red and blue by green light increases biomass and yield in tomato. *Sci. Hortic.* 249, 271–279. doi: 10.1016/j.scienta.2019.02.005
- Kalaji, H. M., Jajoo, A., Oukarroum, A., Brestic, M., Zivcak, M., Samborska, I. A., et al. (2016). Chlorophyll a fluorescence as a tool to monitor physiological status of plants under abiotic stress conditions. *Acta Physiol. Plant.* 38 (4). doi: 10.1007/s11738-016-2113-y
- Kalaji, M. H., Goltsev, V. N., Zuk-Golaszewska, K., Zivcak, M., and Brestic, M. (2017). *Chlorophyll Fluorescence, Understanding Crop Performance* (6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742: CRC Press).
- Kornyejev, D., and Holaday, A. S. (2008). Corrections to current approaches used to calculate energy partitioning in photosystem 2. *Photosynthetica* 46, 170–178. doi: 10.1007/s11099-008-0028-6
- Krall, J. P., and Edwards, G. E. (1992). Relationship between photosystem II activity and CO₂ fixation in leaves. *Physiol. Plant.* 86, 180–187. doi: 10.1111/J.1399-3054.1992.TB01328.X
- Kromdijk, J., Glowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., et al. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Sci. (New York, N.Y.)* 354, 857–861. doi: 10.1126/science.aai8878
- Kuleshova, T. E., Likhachev, A. I., Pavlova, E. S., Kuleshov, D. O., Nashchekin, A. V., and Gall, N. R. (2018). Interrelation of absorption spectra of plant pigments and LED lighting with different spectral compositions. *Tech. Phys.* 63, 1243–1247. doi: 10.1134/S1063784218090104
- Kutik, J., Holá, D., Vičánková, A., Šmidová, M., Kočová, M., Körnerová, M., et al. (2001). The heterogeneity of structural and functional photosynthetic characteristics of mesophyll chloroplasts in various parts of mature or senescing leaf blade of two maize (*Zea mays* L.) genotypes. *Photosynthetica* 39, 497–506. doi: 10.1023/A:1015639609268
- J. R. Lakowicz (Ed.) (2006). *Principles of Fluorescence Spectroscopy* (Boston, MA: Springer US).
- Lazár, D. (2006). The polyphasic chlorophyll a fluorescence rise measured under high intensity of exciting light. *Funct. Plant Biol.: FPB*. 33, 9–30. doi: 10.1071/FP05095
- Lee, G. Y., Alzamil, L., Doskenov, B., and Termehchy, A. (2021). *A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance*.
- Lee, G., and Kim, J. (2024). Sufficient light intensity is required for the drought responses in sweet basil (*Ocimum basilicum* L.). *Agronomy* 14, 2101. doi: 10.3390/agronomy14092101
- Li, G., Zrimec, J., Ji, B., Geng, J., Larsbrink, J., Zelezniak, A., et al. (2021). Performance of regression models as a function of experiment noise. *Bioinf. Biol. Insights*. 15, 11779322211020315. doi: 10.1177/11779322211020315
- Long, S. P., and Bernacchi, C. J. (2003). Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *J. Exp. Bot.* 54, 2393–2401. doi: 10.1093/jxb/erg262
- Lysenko, V., D. Rajput, V., Kumar Singh, R., Guo, Y., Kosolapov, A., Usova, A., et al. (2022). Chlorophyll fluorometry in evaluating photosynthetic performance: key limitations, possibilities, perspectives and alternatives. *Physiol. Mol. Biol. Plants* 28, 2041–2056. doi: 10.1007/s12298-022-01263-8
- Maxwell, K., and Johnson, G. N. (2000). Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* 51, 659–668. doi: 10.1093/jxb/51.345.659
- Murchie, E. H., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64, 3983–3998. doi: 10.1093/jxb/ert208
- Neuner, G., and Pramsohler, M. (2006). Freezing and high temperature thresholds of photosystem 2 compared to ice nucleation, frost and heat damage in evergreen subalpine plants. *Physiol. Plant.* 126, 196–204. doi: 10.1111/j.1399-3054.2006.00605.x
- Nishio, J. N. (2000). Why are higher plants green? Evolution of the higher plant photosynthetic pigment complement. *Plant. Cell Environ.* 23, 539–548. doi: 10.1046/j.1365-3040.2000.00563.x
- Ogren, E., and Sjöström, M. (1990). Estimation of the effect of photoinhibition on the carbon gain in leaves of a willow canopy. *Planta* 181, 560–567. doi: 10.1007/BF00193011
- Oquist, G., and Huner, N. P. A. (2003). Photosynthesis of overwintering evergreen plants. *Annu. Rev. Plant Biol.* 54, 329–355. doi: 10.1146/annurev.arplant.54.072402.115741
- Pick, T. R., Bräutigam, A., Schlüter, U., Denton, A. K., Colmsee, C., Scholz, U., et al. (2011). Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. *Plant Cell* 23, 4208–4220. doi: 10.1105/tpc.111.090324
- Prochazkova, D., Sairam, R. K., Srivastava, G. C., and Singh, D. V. (2001). Oxidative stress and antioxidant activity as the basis of senescence in maize leaves. *Plant Sci.* 161, 765–771. doi: 10.1016/S0168-9452(01)00462-9
- Sage, R. F., and Zhu, X.-G. (2011). Exploiting the engine of C4 photosynthesis. *J. Exp. Bot.* 62, 2989–3000. doi: 10.1093/jxb/err179
- Saleem, M. A., Harrou, F., and Sun, Y. (2024). Explainable machine learning methods for predicting water treatment plant features under varying weather conditions. *Results. Eng.* 21, 101930. doi: 10.1016/j.rineng.2024.101930
- Sharma, D. K., Andersen, S. B., Ottosen, C. O., and Rosenqvist, E. (2015). Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. *Physiol. Plant.* 153, 284–298. doi: 10.1111/pp1.12245
- Simkin, A. J., Kapoor, L., Doss, C. G. P., Hofmann, T. A., Lawson, T., and Ramamoorthy, S. (2022). The role of photosynthesis related pigments in light harvesting, photoprotection and enhancement of photosynthetic yield in plants. *Photosynth Res.* 152, 23–42. doi: 10.1007/s11120-021-00892-6
- Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015
- Singh, B., Kumar, S., Elangovan, A., Vasht, D., Arya, S., Duc, N. T., et al. (2023). Phenomics based prediction of plant biomass and leaf area in wheat using machine learning approaches. *Front. Plant Sci.* 14, 1214801. doi: 10.3389/fpls.2023.1214801
- Smith, H. L., McAusland, L., and Murchie, E. H. (2017). Don't ignore the green light: exploring diverse roles in plant processes. *J. Exp. Bot.* 68, 2099–2110. doi: 10.1093/jxb/erx098
- Sun, J., Nishio, J. N., and Vogelmann, T. C. (1998). Green light drives CO₂ fixation deep within leaves. *Plant Cell Physiol.* 39, 1020–1026. doi: 10.1093/oxfordjournals.pcp.a029298
- Suzuki, S., Nakamoto, H., Ku, M. S., and Edwards, G. E. (1987). Influence of leaf age on photosynthesis, enzyme activity, and metabolite levels in wheat. *Plant Physiol.* 84, 1244–1248. doi: 10.1104/pp.84.4.1244
- Terashima, I., Fujita, T., Inoue, T., Chow, W. S., and Oguchi, R. (2009). Green light drives leaf photosynthesis more efficiently than red light in strong white light: revisiting the enigmatic question of why leaves are green. *Plant Cell Physiol.* 50, 684–697. doi: 10.1093/pcp/pcp034
- Troeng, E., and Linder, S. (1982). Gas exchange in a 20-year-old stand of Scots pine. *Physiol. Plant.* 54, 7–14. doi: 10.1111/j.1399-3054.1982.tb00569.x
- Trouwborst, G., Sander, W. H., Harbinson, J., and Wim Van, I. (2011). The influence of light intensity and leaf age on the photosynthetic capacity of leaves within a tomato canopy. *J. Hortic. Sci. Biotechnol.* 86, 403–407. doi: 10.1080/14620316.2011.11512781
- Varghese, R., Cherukuri, A. K., Doddrell, N. H., Doss George, C. P., Simkin, A. J., Ramamoorthy, S., et al. (2023). Machine learning in photosynthesis: Prospects on sustainable crop development. *Plant Sci.* 335, 111795. doi: 10.1016/j.plantsci.2023.111795
- Virtanen, O., Constantinidou, E., and Tyystjärvi, E. (2022). Chlorophyll does not reflect green light – how to correct a misconception. *J. Biol. Educ.* 56, 552–559. doi: 10.1080/00219266.2020.1858930
- von Caemmerer, S., and Furbank, R. T. (2003). The C(4) pathway: an efficient CO(2) pump. *Photosynth Res.* 77, 191–207. doi: 10.1023/A:1025830019591
- Wu, T., Zhang, W., Wu, S., Cheng, M., Qi, L., Shao, G., et al. (2022). Retrieving rice (*Oryza sativa* L.) net photosynthetic rate from UAV multispectral images based on machine learning methods. *Front. Plant Sci.* 13, 1088499. doi: 10.3389/fpls.2022.1088499
- Yin, X., and Struik, P. C. (2009). C3 and C4 photosynthesis models: An overview from the perspective of crop modelling. *NJAS: Wageningen. J. Life Sci.* 57, 27–38. doi: 10.1016/j.njas.2009.07.001
- Yousef, A. F., Ali, M. M., Rizwan, H. M., Tadda, S. A., Kalaji, H. M., Yang, H., et al. (2021). Photosynthetic apparatus performance of tomato seedlings grown under various combinations of LED illumination. *PLoS One* 16, e0249373. doi: 10.1371/journal.pone.0249373
- Zhang, X.-Y., Huang, Z., Su, X., Siu, A., Song, Y., Zhang, D., et al. (2020). Machine learning models for net photosynthetic rate prediction using poplar leaf phenotype data. *PLoS One* 15, e0228645. doi: 10.1371/journal.pone.0228645